

ΠΑΝΕΠΙΣΤΗΜΙΟ ΠΕΙΡΑΙΑ

ΑΝΑΛΥΣΗ ΝΑΥΤΙΚΩΝ ΑΤΥΧΗΜΑΤΩΝ  
ΜΕ ΧΡΗΣΗ ΤΕΧΝΙΚΩΝ ΟΡΥΞΗΣ ΔΕΔΟΜΕΝΩΝ

ΚΟΚΟΤΟΣ Χ. ΔΗΜΗΤΡΙΟΣ

Πανεπιστήμιο Πειραιώς

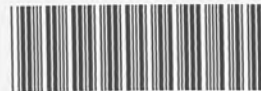
## ΠΑΝΕΠΙΣΤΗΜΙΟ ΠΕΙΡΑΙΑ

ΔΙΔΑΚΤΟΡΙΚΗ ΔΙΑΤΡΙΒΗ

Ανάλυση Ναυτικών Ατυχημάτων  
με χρήση τεχνικών Ορυξης Δεδομένων.

ΚΟΚΟΤΟΣ Χ. ΔΗΜΗΤΡΙΟΣ

ΠΑΝΕΠΙΣΤΗΜΙΟ ΠΕΙΡΑΙΩΣ	
ΑΡ. ΕΙΣ.	51232
COMP	35134
ΤΑΞΗ	519 S Kok
ΒΙΒΛΙΟΘΗΚΗ	



00151232

ΠΕΙΡΑΙΑΣ 2002



Πανεπιστήμιο Πειραιώς

Αφιερωμένο στην Αγγελική και τον Αλέξανδρο που μου πρόσφεραν συναισθηματική στήριξη και συμπαράσταση, όλα αυτά τα χρόνια, χωρίς τις οποίες δεν θα μπορούσε να ολοκληρωθεί η παρούσα διατριβή.

Η διατριβή αυτή εκπονήθηκε στο Εργαστήριο Ναυτικών Ατυχημάτων του Τμήματος Ναυτλιακών Σπουδών του Πανεπιστημίου Πειραιώς, στο χρονικό διάστημα από το Μάρτιο του 1997 μέχρι το Μάιο του 2002, υπό την επίβλεψη του Καθηγητή κ. Κωνσταντίνου Γκιζιάκη. Το θέμα της διατριβής εντάσσεται στο πλαίσιο ευρύτερης έρευνας που αφορά την ανάπτυξη μεθοδολογιών ανάλυσης για την μελέτη των Ναυτικών Ατυχημάτων.

Ευχαριστώ το Τμήμα Ναυτλιακών Σπουδών, που αποδεχόμενο την υποψηφιότητα μου ως διδάκτορα, μου έδωσε την δυνατότητα να αξιοποιήσω το υπόβαθρο γνώσεων μου σε ερευνητικές εφαρμογές σχετικές με αντικείμενα του Τμήματος.

Ευχαριστώ το Εργαστήριο Ναυτικών Ατυχημάτων για την παροχή του χώρου και της Βιβλιοθήκης του.

Θα ήθελα να εκφράσω τις θερμότερες ευχαριστίες μου στον Καθηγητή κ. Κωνσταντίνο Γκιζιάκη για την ανάθεση του θέματος, την καθοδήγηση, το συνεχές και αμειώτο ενδιαφέρον του, την συμπαράσταση και την ηθική βοήθεια που μου προσέφερε κατά την διάρκεια της εκπόνησης και συγγραφής της διατριβής.

Ευχαριστώ θερμά τον Αναπληρωτή Καθηγητή του Τμήματος Πληροφορικής του Πανεπιστημίου Πειραιώς κ. Δημήτρη Δεσπότη, μέλος της τριμελούς συμβουλευτικής επιτροπής, για τις εύστοχες υποδείξεις του και την συμβολή του στην διαμόρφωση του μέρους της διατριβής που αφορά ειδικά θέματα Πληροφορικής.

Ευχαριστώ ιδιαίτερα τον Επίκουρο Καθηγητή κ. Ερνέστο Τζαννάτο, μέλος της τριμελούς συμβουλευτικής επιτροπής, για τις εύστοχες παρατηρήσεις και τις συμβουλές του.

Ευλικρινέστερες ευχαριστίες θα ήθελα να εκφράσω προς όλα τα μέλη Δ.Ε.Π του Τμήματος Ναυτλιακών Σπουδών για την συνεργασία, το θαυμάσιο εργασιακό περιβάλλον και την ενθάρρυνση τους.

## ΠΕΡΙΛΗΨΗ

Ο καθορισμός των παραγόντων που επιδρούν στα ναυτικά ατυχήματα παρουσιάζει ιδιαίτερες δυσκολίες, γεγονός αποδεκτό από την επιστημονική κοινότητα. Σκοπός της παρούσας διατριβής, όπως αναφέρεται στο πρώτο κεφάλαιο, είναι η διερεύνηση του συνδυασμού των παραγόντων ο οποίος οδηγεί στην ολική απώλεια με δεδομένο ότι για ένα πλοίο έχουμε την πληροφορία ότι έχει καταγραφεί ολική απώλεια. Επίσης είναι η εξέταση του συνδυασμού παραγόντων που παρόλο που έχει συμβεί κάποιο ατύχημα, δεν κατέληξε σε ολική απώλεια.

Διερευνάται με βάσει την συνεπαγωγή «αίτιο» → «αποτέλεσμα» η ολική απώλεια των πλοίων (αποτέλεσμα) σε σχέση με ένα σύνολο παραγόντων (αίτιο) και στη συνέχεια ελέγχεται η χρησιμότητά της διερεύνησης. Προς τούτο εφαρμόστηκαν τεχνικές Όρυξης Δεδομένων (Data mining). Μετατράπησαν τα πρωτογενώς ακατέργαστα δεδομένα, όπως αυτά συλλέχθηκαν από τις διάφορες πηγές, σε ένα δομημένο αρχείο πληροφοριών στο οποίο κάθε ατύχημα θα συνδέεται με τους παράγοντες προς διερεύνηση, όπως η γεωγραφική περιοχή, οι καιρικές συνθήκες, η χωρητικότητα, ο τύπος πλοίου κ.ο.κ. Η εφαρμογή των τεχνικών Όρυξης Δεδομένων με χρήση Δένδρων απόφασης (Decision Trees) παρήγαγε κανόνες διερεύνησης της ολικής απώλειας ενός πλοίου. Ουσιαστικά λοιπόν, έγινε μετασχηματισμός μη δομημένης πληροφορίας σε σύστημα που βοηθά την λήψη αποφάσεων.

Στο δεύτερο κεφάλαιο γίνονται αναφορές σε Ελληνική και Διεθνή βιβλιογραφία σχετικές με το θέμα της διατριβής που παρουσιάζουν:

- Τον κατά τον Ι.Μ.Ο ορισμό του ναυτικού ατυχήματος καθώς και μια ιδεατή μεθοδολογία ανάλυσης ναυτικών ατυχημάτων.
- Τεχνικές όρυξης δεδομένων με τις οποίες έγινε επεξεργασία στοιχείων για ανάλυση και πρόβλεψη ατυχημάτων και αναλύσεις με κλασσικές μεθοδολογίες που έγιναν για τα ίδια στοιχεία, καθώς και συγκρίσεις μεταξύ των δύο τεχνικών.
- Μεθοδολογίες οι οποίες είναι σχετικές με αυτές που εφαρμόστηκαν στη παρούσα διατριβή.

- Αναλύσεις δεδομένων για Ναυτικά Ατυχήματα με κλασσικές μεθοδολογίες, οι οποίες επιπλέον δείχνουν την αναγκαιότητα και τα οφέλη από την δημιουργία Βάσης Δεδομένων για αυτά

Στη συνέχεια, στο τρίτο κεφάλαιο, λαμβάνοντας υπόψη τις προαναφερθείσες βιβλιογραφικές αναφορές γίνεται επιλογή της μεθοδολογίας για την απάντηση των ερευνητικών ερωτημάτων και αναφέρονται οι λόγοι επιλογής της συγκεκριμένης μεθοδολογίας. Το αρχείο δεδομένων περιλαμβάνει 38536 εγγραφές και 74 μεταβλητές και για να αντιμετωπιστεί το πρόβλημα εφαρμογής αναλύσεων σε μεγάλα σύνολα δεδομένων απαιτείται μία νέα γενιά τεχνικών, μεθόδων και εργαλείων, όπως η Όρυξη Δεδομένων για ανακάλυψη και εξαγωγή χρήσιμης πληροφορίας.

Με τις τεχνικές Όρυξης αναζητείται η εύρεση των περισσότερων δυνατών σχέσεων χωρίς την προϋπόθεση ύπαρξης κάποιου πειραματικού σχεδιασμού (experimental design). Ειδικότερα στην παρούσα διατριβή με βάση ένα σύνολο παρατηρήσεων ζητείται να εξαχθούν ορισμένοι κανόνες και συμπεράσματα από τα οποία θα μπορούμε να κάνουμε υποθέσεις και εικασίες για το πείραμα που «παράγει» αυτά τα δεδομένα.

Η Όρυξη Δεδομένων αποτελεί ουσιαστικά μία μη παραμετρική, χωρίς ιδιαίτερες προϋποθέσεις μέθοδο ανάλυσης, απαλλαγμένη από επιλογή δείγματος από την οποία εξαρτώνται τα αποτελέσματα. Άρα αποτελεί περισσότερο γενική και αξιόπιστη μέθοδο ανάλυσης δεδομένων η οποία έχει την ικανότητα να βρίσκει «κοινές συμπεριφορές» (patterns) μέσα σε μεγάλους όγκους δεδομένων.

Τα Δέντρα Απόφασης (Decision Trees), που χρησιμοποιήθηκαν για την ανάλυση, ουσιαστικά είναι ταξινομητές / ελκυστές με τη μορφή δέντρου. Έχουν υψηλή δυνατότητα για ταξινόμηση και πρόβλεψη. Αποκλείουν από τη φύση τους, όλους τους παράγοντες που δεν επιδρούν σημαντικά στο σχηματισμό τους, άρα παρέχουν αναλυτική εικόνα για τη διαδικασία δράσης των πεδίων / μεταβλητών. Τα δέντρα προσφέρουν και ιεράρχηση των παραγόντων σε σχέση με τα Σύνολα κανόνων τα οποία παρακολουθούν το στόχο της ερευνητικής εργασίας. Σε σχέση μάλιστα με τα Νευρωνικά δίκτυα, των οποίων η πολυπλοκότητα της δομής καθιστά την ερμηνεία τους δύσκολη, τα δέντρα παρέχουν και τη διαισθητική αντίληψη του συνόλου των δεδομένων με ένα τρόπο βάση του οποίου αντλαμβάνεται κανείς το στόχο σε μία σχέση αιτίου – αποτελέσματος. Τα πολύπλοκα δένδρα δύσκολα γίνονται κατανοητά,



όμως η λογική που διέπει τον κάθε κλάδο του δένδρου χωριστά, γίνεται εύκολα κατανοητή.

Ο αλγόριθμος C5.0, που χρησιμοποιήθηκε για να παράγει τα δένδρα, προσπαθεί να βρει τον καλύτερο δυνατό κανόνα που επαληθεύεται στο μεγαλύτερο τμήμα του συνόλου δεδομένων. Συνοπτικά αυτός διαλέγει εκείνο το πεδίο / μεταβλητή το οποίο διαφοροποιεί με τον καλύτερο δυνατό τρόπο το σύνολο των εγγραφών, χωρίζοντάς το σε υποσύνολα, ως προς το εξαρτημένο πεδίο / μεταβλητή. Πέρα από το βασικό σώμα του αλγόριθμου που σχολιάζεται, αναφέρονται και μερικές σημαντικές ακόμα ιδιότητες και βελτιώσεις που έχουν ενσωματωθεί σε αυτό.

Οι ταξινομητικοί κανόνες που παρήχθησαν προσφέρουν έναν εναλλακτικό τρόπο να διαβαστεί ένα δέντρο αποφάσεων. Οι κανόνες ταξινόμησης δίνουν την δυνατότητα να φαίνεται και το πλήθος των τελικών κανόνων. Ο σημαντικότερος αλγόριθμος, ειδική περίπτωση του C5.0, που χρησιμοποιήθηκε στην παρούσα διατριβή λόγω της μορφής των δεδομένων είναι ο G.R.I. Αυτός ανιχνεύει την ύπαρξη όλων των σημαντικών κανόνων και πλεονέκτημά του είναι ότι παράγει μικρά σε βάθος δένδρα.

Το αρχείο δεδομένων περιλαμβάνει τα ατυχήματα που συνέβησαν τα έτη 1992 έως 1999. Η παρουσίαση του αρχείου δεδομένων είναι ιδιαίζουσας σημασίας λόγω της μορφής των δεδομένων, τα οποία πρέπει να μετασχηματιστούν σύμφωνα με κάποιο δομημένο σύστημα, πράγμα απαραίτητο για την επεξεργασία του. Στο τέταρτο κεφάλαιο αναλυτικά αναφέρονται οι πηγές άντλησης των στοιχείων, περιγράφεται η δημιουργία, η γραμμογράφηση και η κωδικοποίηση του αρχείου δεδομένων, για την ανάλυση με το στατιστικό πακέτο S.P.S.S. Εξηγούνται, σύμφωνα με τους στόχους της παρούσας εργασίας, οι λόγοι επιλογής των πεδίων / μεταβλητών και σχολιάζεται το πεδίο ορισμού των. Τέλος δίνονται στοιχεία για τις εγγραφές του αρχείου.

Το στάδιο προετοιμασίας των δεδομένων αφορά την απομάκρυνση λανθασμένων εγγραφών, την διαχείριση τιμών που λείπουν, την τυποποίηση των πεδίων / μεταβλητών, την δημιουργία σχεσιακών δομών και γενικά την ορθή αποκατάσταση της μορφής των πεδίων / μεταβλητών και των εγγραφών. Ο καθαρισμός του αρχείου δεδομένων (data cleaning), που γίνεται στο πέμπτο κεφάλαιο, θεωρήθηκε απαραίτητο βήμα από τον τρόπο δημιουργίας και σχηματισμού του αρχείου με σκοπό την δημιουργία μοναδικής εγγραφής του ατυχήματος πλοίου. Ο καθαρισμός δεν έχει ακόμα εξελιχθεί σε τυποποιημένη εργασία για να υπάρχει

κάποια σχετική μεθοδολογία. Η όλη διαδικασία καθαρισμού του παρόντος αρχείου δεδομένων βασίζεται στην ημερολογιακή κατασκευή του αρχείου και τη δυναμική του φύση. Έτσι επινοήθηκαν αλγόριθμοι, προσαρμοσμένοι στις ιδιαιτερότητες του, οι οποίοι αντικατέστησαν τις ελλείπουσες τιμές, εντόπισαν και αποκατέστησαν στο ορθόν τις διαφορετικές τιμές και έκαναν απαλοιφή των διυλεγγραφών όπου υπήρχε η δυνατότητα. Τέλος, εξαιτίας της σημαντικότητας του πεδίου / μεταβλητής της ολικής απώλειας αυτό ελέγχθηκε ως προς την ποιότητα εγγραφής του, και επιβεβαιώθηκε η αξιοπιστία των πληροφοριών του αρχείου δεδομένων χρησιμοποιώντας και τα δεδομένα μίας δεύτερης πηγής πληροφοριών των Lloyd's Register "World Casualty Statistics". Στη συνέχεια, παρουσιάστηκαν τα βήματα, για την εφαρμογή του αλγόριθμου C5.0 και των λοιπών μεθοδολογιών που χρησιμοποιήθηκαν για την ανάλυση του αρχείου δεδομένων. Ορίστηκαν όλες οι παράμετροι καθώς και οι αναγκαίες συνθήκες και καθορίστηκαν τα ανεξάρτητα και εξαρτημένα πεδία / μεταβλητές.

Στον κεφάλαιο έξι και επτά παρουσιάζονται τα αποτελέσματα και τα συμπεράσματα που προέκυψαν από το σύνολο των διαδικασιών και αναλύσεων που εφαρμόστηκαν στο αρχείο.

Με χρήση μεθοδολογίας δέντρων απόφασης, με ποσοστό αξιοπιστίας ταξινόμησης 92,44 % στο σύνολο των εγγραφών, εντοπίστηκαν τα χαρακτηριστικά που έχουν την μεγαλύτερη επιρροή στην ολική απώλεια και μάλιστα ο αλγόριθμος:

- ομαδοποίησε τις κατηγορίες ατυχήματος άλλες σε ιδιαίτερα επικίνδυνες (CAPSIZED/ FOUNDERED) και άλλες σε λιγότερο (FIRE/COLLISSION κ.α.).
- διαχώρισε τα πλοία ανάλογα με την χωρητικότητά τους και υπέδειξε σαν περισσότερα επικίνδυνα για ολική απώλεια αυτά με χωρητικότητα μικρότερη από 16518 grt, τα οποία είναι και περισσότερα ευάλωτα.
- χώρισε τις γεωγραφικές περιοχές του ατυχήματος σε περιοχές υψηλού κινδύνου, που ανήκουν στην ευρύτερη ζώνη Ερυθράς Θάλασσας, Περσικού Κόλπου, Αφρικανικών ακτών, Κόλπου της Βεγγάλης, Ινδικού, Θάλασσας Ιαπωνίας και Νότιας Κίνας, που έχουν σαν επικρατούν γεγονός την ολική απώλεια, αναμενόμενο για τις περιοχές αυτές, λόγω μεγάλης κυκλοφορίας και ιδιόμορφων καιρικών φαινομένων.
- ανέδειξε την ανεξαρτησία, κατά κύριο λόγο, της εποχικότητας και μάλιστα του μήνα σε σχέση με την ολική απώλεια πλοίου.



- υπέδειξε εν τούτοις για την περίοδο των Μουσώνων, ότι τα πλοία τα εγγεγραμμένα σε νηογνώμονες, πλην αυτών του Διεθνούς Συνδέσμου Νηογνώμωνων (I.A.C.S) είναι περισσότερο επικίνδυνα και έχουμε για αυτά σαν επικρατούσα τιμή την ολική απώλεια.

- δεν διαχώρισε τα πλοία ανάλογα με την ηλικία, τον τύπο τους και τις καιρικές συνθήκες που επικρατούσαν την χρονική στιγμή του ατυχήματος διότι τα δένδρα από την φύση τους ιεραρχούν τους παράγοντες και επιτρέπουν να διαφαιίνεται κατά προτεραιότητα η σημαντικότητα κάθε παράγοντα.

Έγινε δυνατή η δημιουργία κανόνων ταξινόμησης που έχουν συγκριτικό πλεονέκτημα την ιδιαίτερα εύκολη χρήση πράγμα απαραίτητο για τις αναλύσεις άρα και για ένα σύστημα λήψης αποφάσεων. Από μια σύγκριση που γίνεται μεταξύ των συμπληρωματικών κανόνων, συμπεραίνουμε ότι η πρώτη καταγραφή ενός ατυχήματος πλοίου στο αρχείο δεδομένων δε συνοδεύεται αυτόματα με ολική απώλεια. Στην περίπτωση όμως ύπαρξης και δεύτερης καταγραφής ατυχήματος για το ίδιο πλοίο αυξάνει η πιθανότητα για ολική απώλεια του.

Εξειδικεύσαμε, στην συνέχεια, την ανάλυση της επίδρασης των παραγόντων, με τον αποκλεισμό μερικών, με βάση τις γενικότερες παρατηρήσεις στο αρχικό δένδρο, ώστε να μελετήσουμε με εφαρμογή τεχνικών Boosting την επίδραση ενός υποσυνόλου παραγόντων στην ολική απώλεια. Ανεξαρτήτως ατυχήματος λοιπόν, ο αλγόριθμος του δένδρο διαχώρισε καταρχήν τους ομαδοποιημένους νηογνώμονες σε δύο ομάδες. Η μια ομάδα, που περιλαμβάνει σε τρεις κλάσεις, τους νηογνώμονες AB-GL-NK-BV-CC-KR-NV-RI-PR-RS και LR που ανήκουν στον Διεθνή Σύνδεσμο Νηογνώμωνων (I.A.C.S) και δεν έχει σαν επικρατούν γεγονός την ολική απώλεια. Η δεύτερη ομάδα περιέχει τους υπόλοιπους νηογνώμονες (OTHERS) και παρατηρείται σε αυτήν σαν επικρατούν γεγονός η ολική απώλεια. Η μη διάκριση των νηογνώμωνων που ανήκουν στον Σύνδεσμο μεταξύ τους, για την συγκεκριμένη ομαδοποίηση νηογνώμωνων, αναδεικνύει την γενικότερη αξιοπιστία των προσφερομένων υπηρεσιών του συγκεκριμένου Συνδέσμου στο σύνολό του. Μετά τον πρώτο διαχωρισμό, όμως συμπεραίνουμε ότι η αποτελεσματικότητα των υπηρεσιών που παρέχει ο Διεθνής Σύνδεσμος Νηογνώμωνων ελέγχεται για τα παλιά πλοία με έτος ναυπήγησης παλιότερο του 1974 στα οποία το έργο επιθεώρησης και εντοπισμού βλαβών καθίσταται επίπονο και ιδιαίτερα δύσκολο. Δευτερευόντως ο διαχωρισμός που έκανε ο αλγόριθμος σε συγκεκριμένες ομάδες δείχνει την ευαισθησία, τις

δυνατότητες την αξιοπιστία του άρα και την ορθότητα επιλογής του. Σε συνέχεια των ανωτέρω σε μια εύχρηστη εφαρμογή του αλγόριθμου G.R.I. αναπτύσσονται όλα τα σχετικά δέντρα με περιορισμό του βάθους τους με ευδιάκριτα αποτελέσματα.

Εξήχθησαν πολύτιμα συμπεράσματα, με σκοπό την αποφυγή ατυχημάτων και ειδικότερα της ολικής απώλειας με βάση τα οποία πρέπει:

- να γίνονται οι απαραίτητοι έλεγχοι με έμφαση σε ορισμένους παράγοντες.

Ένα από τα πλεονεκτήματα της συγκεκριμένης μεθοδολογία είναι το ότι μας δίνει την δυνατότητα να παρατηρούμε την ολική απώλεια στην τομή των ανεξαρτήτων μεταβλητών / παραγόντων που οδηγούν σε αυτήν. Για τις προαναφερθείσες, λοιπόν, γεωγραφικές περιοχές, στα μικρού μεγέθους πλοία, πρέπει να δοθεί προσοχή στα ατυχήματα CAPSIZED, διότι αυτός ο συνδυασμός παραγόντων οδηγεί σε ολική απώλεια. Προκύπτει λοιπόν, μια άμεση ωφέλεια για φορείς όπως οι Λιμενικές Αρχές οι οποίες μπορούν να κατευθύνουν τις Υπηρεσίες Ελέγχου που διαθέτουν προς τα πλοία με το μεγαλύτερο κίνδυνο. Επίσης ωφέλεια για τους Πλοιοκτήτες που, εκτός από την γνώση του ιστορικό του πλοίου από την επεξεργασία δεδομένων πρωτογενώς, πρέπει να δίνουν περισσότερη προσοχή σε ορισμένους παράγοντες που υποδεικνύονται από τους παραγόμενους κανόνες.

- να θεσμοθετήσουν κατάλληλα μέτρα.

Αναλυτικότερα και ενδεικτικά αναφέρεται να ληφθούν μέτρα παρέμβασης κατά προτεραιότητα, για αποτροπή CAPSIZED (FOUNDERED) λόγω απώλειας της ευστάθειας των πλοίων, που μπορεί να προκύψει από μετατόπιση φορτίου, εισροή υδάτων κ.λ.π., με απόδοση έμφασης στα μικρού μεγέθους πλοία και για συγκεκριμένες γεωγραφικές περιοχές κυκλοφορίας αυτών των πλοίων. Οι Επιτροπές Ασφαλείας των Νηογυμνώνων, ο I.M.O, το N.E.E, η E.E.E μπορούν να χρησιμοποιήσουν τα συμπεράσματα από την ανάλυση προκειμένου να βελτιώσουν τους κανόνες τους και να λάβουν τα αποτελεσματικότερα μέτρα όσον αφορά την μείωση του αριθμού και του κόστους των ατυχημάτων και μάλιστα της ολικής απώλειας και γενικότερα των συνεπειών που προκύπτουν από αυτά.

- να γίνονται προβλέψεις.

Τα δένδρα προσφέρονται για πρόβλεψη της ολικής απώλειας, πολύ καλύτερη από ότι αν χρησιμοποιήσουμε τις υπόλοιπες μεθόδους ανάλυσης. Αυτή η ικανότητα πρόβλεψης είναι ιδιαίτερα χρήσιμη για τις Κρατικές Υπηρεσίες και τους Διεθνείς Οργανισμούς καθώς και για τους Οργανισμούς τους σχετικούς με την προστασία του

Περιβάλλοντος, οι οποίοι μπορούν να οργανωθούν ως προς τον σχεδιασμό, την προετοιμασία και αντιμετώπιση εκτάκτων αναγκών και μάλιστα μπορούν να επέμβουν γρήγορα και προς την σωστή κατεύθυνση. Επίσης οι Ασφαλιστικές εταιρίες είναι σε θέση να διαμορφώσουν σωστή Τιμολογιακή Πολιτική όταν έχουν εντοπισθεί οι παράγοντες που χαρακτηρίζονται σαν «υψηλού κινδύνου» και να αξιολογήσουν συγκριτικά αυτούς τους παράγοντες ώστε να γνωρίζουν τα απαιτούμενα κεφάλαια προς αποζημίωση.

Γενικότερα υποστηρίζεται ότι μία συμβολή της παρούσας διατριβής στην απάντηση των γενικότερων ερευνητικών ερωτημάτων που τίθενται, είναι να δώσει τις πρώτες βάσεις για ένα σύστημα υποστήριξης αποφάσεων

- είτε μέσω των πρωτογενώς επεξεργασμένων δεδομένων
- είτε με τη μεθοδολογία που ακολουθήθηκε
- είτε με τους παραγόμενους κανόνες και συμπεράσματα

με αποτέλεσμα να βοηθηθούν οι Ευρωπαϊκοί και οι Παγκόσμιοι Οργανισμοί ως προς την αύξηση της ασφάλειας στη Ναυτιλία στα ύδατα της Ευρωπαϊκής Ένωσης και σε παγκόσμιο επίπεδο.

Η αποκτηθείσα ήδη εμπειρία από την δημιουργία του αρχείου, τα νέα προγράμματα, οι νέοι αλγόριθμοι και τα συμπεράσματα από την ερμηνεία των αποτελεσμάτων με την εφαρμογή των τεχνικών Όρυξης Δεδομένων, αναμένεται να συνεισφέρουν στην επίλυση συναφών επιστημονικών προβλημάτων.



## ΠΙΝΑΚΑΣ ΠΕΡΙΕΧΟΜΕΝΩΝ

	Κεφάλαια	Σελ.
	Ευχαριστίες	3
	Περίληψη	4
	Πίνακας Περιεχομένων	11
	Συνοπτομογραφίες	16
	<b>Κεφάλαιο 1</b>	
1	Σκοπός και δομή της διατριβής	
1.1	Εισαγωγή	17
1.2	Σκοπός της διατριβής	18
1.3	Δομή της διατριβής	19
	<b>Κεφάλαιο 2</b>	
2	Αναφορές σε ερευνητικές εργασίες	
2.1	Εισαγωγή	22
2.2.	Αναφορές και σχετικές παραπομπές, γενικά	23
2.2.1	Ορισμοί και βασικές θεωρητικές προσεγγίσεις	25
2.2.2	Ανάλυση ναυτικών ατυχημάτων με τεχνικές Όρυξης δεδομένων (Data mining)	28
2.2.3	Ερευνητικές εργασίες σχετικές με την εφαρμοσθείσα μεθοδολογία	34
2.2.4	Ερευνητικές εργασίες για τα ναυτικά ατυχήματα.	39
	<b>Κεφάλαιο 3</b>	
3	Επιλογή Μεθοδολογίας	
3.1	Εισαγωγή	47
3.2	Περιγραφή της μορφής του αρχείου δεδομένων	48

	Κεφάλαια	Σελ.
3.3	Επιλογή μεθοδολογίας ανάλυσης δεδομένων	49
3.3.1	Δέντρα απόφασης (Decision trees)	53
3.3.2	Ο Αλγόριθμος C5.0	58
3.3.2.1	Απόδοση του αλγόριθμου σε διάγραμμα	60
3.3.2.2	Βελτιώσεις του βασικού αλγόριθμου	63
3.3.2.3	Το διάγραμμα αποτελέσματος	64
3.3.3	Ταξινομητικοί κανόνες	65
3.3.4	Ο αλγόριθμος G.R.I	66
	<b>Κεφάλαιο 4</b>	
4	Τα Δεδομένα	
4.1	Εισαγωγή	68
4.2	Δημιουργία του αρχείου δεδομένων	68
4.3	Γραμμογράφηση και κωδικοποίηση του αρχείου δεδομένων	69
4.3.1	Ατύχημα	71
4.3.2	Χωρητικότητα	72
4.3.3	Ηλικία	72
4.3.4	Νηογνώμονας	72
4.3.5	Είδος περιοχής	73
4.3.6	Έτος ναυπήγησης	73
4.3.7	Μήνας	73
4.3.8	Γεωγραφική περιοχή	73
4.3.9	Χρονολογία ατυχήματος	74
4.3.10	Πλήθος αλλαγής ιδιοκτητών	74
4.3.11	Τύπος πλοίου	74
4.3.12	Μη επιτυχές ταξίδι	75
4.3.13	Σημεία	75
4.3.14	Καιρικές συνθήκες	76
4.3.15	Ολική απώλεια	76
4.3.16	Νηογνώμονες ομαδοποιημένοι	77

	Κεφάλαια	Σελ.
4.3.17	Χρονική μεταβλητή ατυχήματος	78
4.3.18	Χρονική μεταβλητή καταγραφής ατυχήματος	78
4.3.19	Σχετικό μέγεθος	79
4.3.20	Ακολουθία εγγραφών για κάθε πλοίο	79
4.3.21	Πλήθος ατυχημάτων	79
4.3.22	Σειρά ατυχήματος	80
4.3.23	Πλήθος αλλαγής ιδιοκτητών ομαδοποιημένο	80
4.3.24	Πεδίο ταξινόμησης	80
4.4	Εγγραφές και στοιχεία του αρχείου δεδομένων	80
	<b>Κεφάλαιο 5</b>	
5	Εφαρμογή της Μεθοδολογίας στα δεδομένα	
5.1	Εισαγωγή	82
5.2	Διαδικασία καθαρισμού του αρχείου δεδομένων	82
5.2.1	Το αρχείο δεδομένων	83
5.2.2	Καθορισμός ομάδων μεταβλητών	84
5.2.3	Ορισμός νέων μεταβλητών για τον καθαρισμό του αρχείου δεδομένων	84
5.2.4	Έλεγχος των μεταβλητών μονοδιάστατα	67
5.2.5	Έλεγχος των μεταβλητών με διασταυρώσεις	85
5.2.6	Επιπλέον έλεγχοι εγγραφών	85
5.2.7	Μοναδική εγγραφή του πλοίου, γενικές διαδικασίες	86
5.2.8	Αντικατάσταση των ελλειπουσών (missing) τιμών	86
5.2.9	Αντικατάσταση των διαφορετικών τιμών	88
5.2.10	Μοναδική εγγραφή του Ατυχήματος	89
5.2.11	Διόρθωση των υπολοίπων μεταβλητών	91
5.2.12	Βαθμός αξιοπιστίας κάθε εγγραφής	92
5.3	Επιλογή παραμέτρων για την εφαρμογή του αλγόριθμου C5.0 και των λοιπών τεχνικών	93



	Κεφάλαια	Σελ.
	<b>Κεφάλαιο 6</b>	
6	Αποτελέσματα	
6.1	Εισαγωγή	98
6.2	Αποτελέσματα από την ανάλυση των μεταβλητών	99
6.3	Αποτελέσματα από την ανάλυση του αρχείου δεδομένων με δέντρα αποφάσεων	112
6.4	Αποτελέσματα από την ανάλυση του αρχείου δεδομένων με ταξινομητικούς κανόνες	118
6.5	Αποτελέσματα από την ανάλυση του αρχείου δεδομένων με άλλες μεθόδους	122
	<b>Κεφάλαιο 7</b>	
7	Συμπεράσματα	
7.1	Εισαγωγή	129
7.2	Συμπεράσματα και σχόλια από την ανάλυση των μεταβλητών	130
7.3	Συμπεράσματα από την ανάλυση του αρχείου με τις επιλεγείσες μεθοδολογίες	132
7.4	Προτάσεις για μελλοντικές ερευνητικές εργασίες	140
	<b>Βιβλιογραφία</b>	143
	<b>Αναφορές στο Διαδίκτυο</b>	151
	<b>Παραρτήματα</b>	152
Σ.	Συμβολισμοί μεταβλητών του αρχείου δεδομένων	153
Κ.	Κωδικοί που χρησιμοποιήθηκαν	154

	Κεφάλαια	Σελ.
	Αρχείο αποτελεσμάτων SPSS	
Π <sub>1</sub> .	Πίνακες μονοδιάστατης ανάλυσης	168
Π <sub>2</sub> .	Πίνακες δυσδιάστατης ανάλυσης	174
Π <sub>3</sub> .	Περιγραφικά στοιχεία	179
Δ.	Δέντρα αποφάσεων για την ολική απώλεια	184
T <sub>Yes</sub> .	Ταξινομητικοί κανόνες για την ολική απώλεια	191
T <sub>No</sub> .	Ταξινομητικοί κανόνες που δεν οδηγούν σε ολική απώλεια	199
Π.	Παράγοντες που ο συνδυασμός τους οδηγεί σε ολική απώλεια	208
	<b>Παράρτημα Α</b>	
N.	Η μεθοδολογία των νευρωνικών δικτύων	230
	<b>Παράρτημα Β</b>	
A.	Νέοι αλγόριθμοι και προγράμματα	231
B.	Ο αλγόριθμος C5.0. στα δεδομένα του αρχείου σε γλώσσα C	249

## ΣΥΝΤΟΜΟΓΡΑΦΙΕΣ

D.M. :	Data mining
E.T.A. :	Event Tree Analysis
F.M.E.A. :	Failure Mode and Effects Analysis
F.S.A. :	Formal Safety Assessment
G.R.I. :	Generalized rule induction
GR.T :	Gross Tonnage
H.A.Z.O.P.:	Hazard and Operability studies
I.A.C.S :	International Association of Classification Societies
I.M.O :	International Maritime Organization
I.M.R :	lower Mississippi River
O.L.A.P. :	On line analytical processing
R.O.C. :	Receiver Operating Characteristic
S.O.M. :	Self Organizing Map
S.Q.L. :	Structured Query Language
SAFECO :	Safety of Shipping in Coastal Waters
Υ.Ε.Ν. :	Υπουργείο Εμπορικής Ναυτιλίας
[Όνομα, έτος]:	Παραπομπή στην αλφαβητικά παρατιθέμενη Βιβλιογραφία
[(1)-web]:	Παραπομπή με αριθμηση σε Αναφορά web

## ΚΕΦΑΛΑΙΟ 1

### 1.1 Εισαγωγή

Για την κατανόηση των παραγόντων που επιδρούν στην ασφάλεια του “Ναυτιλιακού Συστήματος” θα πρέπει να μελετηθούν τα Ναυτικά Ατυχήματα. Η μελέτη τους απαιτεί τη δημιουργία Αρχείου Δεδομένων που να περιλαμβάνει όλα τα Ναυτικά Ατυχήματα συμπεριλαμβανομένων και των ολικών απωλειών.

Η σημασία της κατασκευής του αρχείου δεδομένων για τη διερεύνηση των ναυτικών ατυχημάτων είναι θεμελιώδης, διότι από την επεξεργασία των πληροφοριών του μπορούμε να παράγουμε χρήσιμα συμπεράσματα πέρα από το πλαίσιο της εμπειρικής παρατήρησης. Με τη βοήθεια μάλιστα δομημένου αρχείου δεδομένων είναι εφικτό να παραχθούν προβλέψεις καθώς και προτάσεις για μελλοντικές ερευνητικές εργασίες.

Για την ανάλυση του αρχείου δεδομένων χρησιμοποιήθηκαν οι μεθοδολογίες και διαδικασίες που ταιριάζουν καλύτερα στη μορφή των δεδομένων και στο στόχο της παρούσας διατριβής. Πιο συγκεκριμένα :

- Οι διαδικασίες κατασκευής και καθαρισμού του αρχείου δεδομένων είναι αυτές που επιτρέπουν το μετασχηματισμό των ακατέργαστων δεδομένων σε υλικό από το οποίο μπορούν να ανιχνευτούν σχέσεις και συσχετίσεις.
- Οι μεθοδολογίες ανάλυσης των πληροφοριών του αρχείου δεδομένων αναζητούν απαντήσεις στα βασικά ερευνητικά ερωτήματα. Οι τεχνικές όμως της Όρυξης Δεδομένων (One of the ten emerging technologies that will change the world) [\*] είναι εκείνες που προχωρούν πέρα από την απλή αναζήτηση και παρέχουν δυνατότητες συσχέτισης των δεδομένων, εξαγωγής ποσοτικών δεικτών που να εκφράζουν ποιοτικές έννοιες, ταξινόμησης και κατάταξης οντοτήτων σύμφωνα με σύνθετα και ασαφή κριτήρια, εντοπισμού τυποποιημένων μορφών (patterns). Ο όρος Όρυξη Δεδομένων (Data mining) εκφράζει αυτή ακριβώς τη διαδικασία Εφαρμογής μεθόδων ανάλυσης σε μεγάλο όγκο δεδομένων και της ανακάλυψης γνώσεων με συσχετισμούς.

[\*] MIT Technology Review, Jan/Feb 2001



Με χρήση του αλγορίθμου C5.0, παράγονται δέντρα απόφασης (Decision Trees) τα οποία ταξινομούν ιεραρχικά, το σύνολο των μεταβλητών του αρχείου, σε αντίθεση με την επιλογή μεταβλητών όταν διατυπώνεται αρχική υπόθεση, και για το σύνολο των εγγραφών του αρχείου και όχι μόνο δείγματος.

Στην παρούσα διατριβή, με στόχο – οδηγό, η συλλογή των πληροφοριών να είναι αξιόπιστη, η μεθοδολογία η πλέον αρμόζουσα για την περίπτωση και τα αποτελέσματα να μπορούν να αξιοποιηθούν κατάλληλα, συμπληρώθηκε, τροποποιήθηκε (Giziakis K. P.h.d, 1987) και μετασχηματίστηκε η μη δομημένη πληροφορία σε σύστημα που βοηθά την λήψη αποφάσεων. Η εφαρμογή της συγκεκριμένης μεθόδου ανάλυσης παρήγαγε δένδρα απόφασης και ταξινομητικούς κανόνες με σκοπό την διερεύνηση, μετά από κάποιο ατύχημα, της ολικής απώλειας ενός πλοίου.

## 1.2. Σκοπός της διατριβής

Η μελέτη των περιπτώσεων ατυχημάτων στο ναυτιλιακό τομέα μπορεί να γίνει με κατά περίπτωση (case-to-case) ανάλυση για κάθε ένα πλοίο και κάθε ένα ατύχημα. Αυτό που έμενε στους ερευνητές στην περίπτωση αυτή ήταν ο συνδυασμός των εμπειρικών γνώσεων σε μία διαδικασία αιτιότητας « Το ατύχημα  $x$  συνέβη επειδή  $y$  ».

Ο σκοπός της ερευνητικής αυτής εργασίας είναι να γίνει προσπάθεια υπέρβασης της κατά περίπτωση ανάλυσης και να μελετηθούν τα ατυχήματα ανεξαρτήτως του πλοίου. Να ακολουθηθεί δηλαδή η αντίστροφη πορεία. Με δεδομένο ότι γνωρίζουμε ότι έχουμε ολική απώλεια για ένα πλοίο, ζητείται να βρεθεί ποιος είναι αυτός ο συνδυασμός χαρακτηριστικών ή άλλων παραγόντων ο οποίος να οδηγεί στο παραπάνω ενδεχόμενο. Και αντίστοιχα, ποιος είναι ο συνδυασμός παραγόντων που παρόλο που έχει συμβεί κάποιο ατύχημα, δεν κατέληξε σε ολική απώλεια αυτό το συμβάν.

Πρακτικά αυτό σημαίνει ότι θα γίνει διερεύνηση της ολικής απώλειας των πλοίων (σαν αποτέλεσμα) με βάση ένα σύνολο παραγόντων (σαν αίτιο), η οποία θα μπορούσε να τοποθετηθεί σε μια αλληλουχία της μορφής « αίτιο  $\rightarrow$  αποτέλεσμα » και θα μπορεί στη συνέχεια να ελεγχθεί για την χρησιμότητά της.

Για να επιτευχθεί ο παραπάνω στόχος είναι απαραίτητο να μετατραπούν τα πρωτογενώς ακατέργαστα δεδομένα, όπως αυτά βρίσκονται σε διάφορα σχετικά

έντυπα και πηγές, σε ένα δομημένο αρχείο πληροφοριών στο οποίο κάθε ατύχημα θα συνδέεται με τους προς διερεύνηση παράγοντες, όπως η χωρητικότητα, ο τύπος του πλοίου κ.ο.κ. Πρέπει λοιπόν, να κατασκευαστεί αρχείο δεδομένων που να περιλαμβάνει όλα τα δεδομένα για τα ατυχήματα και να είναι εφικτή η ανάλυση των πληροφοριών του. Το παραπάνω στάδιο αποτελεί το μεγαλύτερο και πιο επίτονο μέρος της διατριβής κυρίως λόγω της έλλειψης ετοιμών για στατιστική ανάλυση σχετικών στοιχείων.

Αφού σκοπός της παρούσας διατριβής είναι η διερεύνηση των ναυτικών ατυχημάτων και μάλιστα της ολικής απώλειας, το προς κατασκευή και επεξεργασία αρχείο δεδομένων θα πρέπει να περιέχει όλα τα καταγεγραμμένα ατυχήματα της προς εξέταση περιόδου 1992-1999, ανεξάρτητα από το αν συνδέονται με την ολική απώλεια του πλοίου, έτσι ώστε να ανιχνευτεί η ύπαρξη συγκεκριμένων συνδυασμών και ταξινομητικών κανόνων, που οδηγούν στην ολική απώλεια του πλοίου.

Από το παραγόμενο αρχείο δεδομένων μπορεί να παραχθεί απεριόριστος αριθμός σεναρίων, ενώ οι παραγόμενες σχέσεις μπορούν να αποτελέσουν την πρώτη ύλη για ένα σύστημα λήψης αποφάσεων.

### 1.3. Δομή της διατριβής

Στην συνέχεια παρουσιάζεται η βασική δομή της διατριβής και μάλιστα στο πρώτο κεφάλαιο δίνεται η περιλήψη, ο πίνακας περιεχομένων, οι συντομογραφίες, και ο σκοπός της.

Στο δεύτερο κεφάλαιο (Αναφορές σε σχετική βιβλιογραφία) παρουσιάζονται σχετικές με τη διατριβή έρευνες. Αφού προταθούν ο ορισμός του ναυτικού ατυχήματος καθώς και μια θεωρητική προσέγγιση μεθοδολογίας ανάλυσης ναυτικών ατυχημάτων έπονται:

- Ερευνητικές εργασίες στις οποίες περιγράφονται τεχνικές όρυξης δεδομένων οι οποίες εφαρμόστηκαν σε αναλύσεις Ναυτικών ατυχημάτων. Επίσης, αναλύσεις με κλασσικές μεθοδολογίες που έγιναν συγχρόνως για τα ίδια δεδομένα και συγκρίσεις μεταξύ των δύο τεχνικών.

- Ερευνητικές εργασίες που είναι σχετικές με τις μεθοδολογίες που εφαρμόστηκαν στη παρούσα διατριβή.



- Ερευνητικές εργασίες που είναι σχετικές με αναλύσεις δεδομένων για τα Ναυτικά Ατυχήματα με χρήση κλασσικών μεθοδολογιών, οι οποίες επιπλέον δείχνουν την αναγκαιότητα και τα οφέλη από την δημιουργία Βάσης Δεδομένων για αυτά

Στο τρίτο κεφάλαιο (Επιλογή μεθοδολογίας) παρουσιάζονται λαμβάνοντας υπόψη και τις αναφορές του δεύτερου κεφαλαίου, οι αρμόζουσες μεθοδολογίες για την απάντηση των ερευνητικών ερωτημάτων. Στόχος για την επιλογή της μεθοδολογίας είναι η εύρεση του καλύτερου εργαλείου, ή συνδυασμού εργαλείων για ανάλυση σε μεγάλα σύνολα δεδομένων. Επιλέχθηκε η τεχνική της Όρυξης δεδομένων και με εφαρμογή του αλγορίθμου C.5.0, καθώς και βελτιώσεων του και ειδικών περιπτώσεων του, παρήχθησαν δένδρα και ταξινομητικοί κανόνες.

Στο τέταρτο κεφάλαιο (Τα δεδομένα) παρουσιάζεται το αρχείο δεδομένων πάνω στο οποίο εξετάστηκαν οι ερευνητικές υποθέσεις της εργασίας. Το κεφάλαιο αυτό είναι ιδιάζουσας σημασίας λόγω της μορφής των δεδομένων, τα οποία πρέπει να μετασχηματιστούν σύμφωνα με κάποιο δομημένο σύστημα, μετασχηματισμοί απαραίτητοι για την επεξεργασία τους. Αναφέρονται οι πηγές, περιγράφεται η δημιουργία, η γραμμογράφηση και η κωδικοποίηση του αρχείου δεδομένων. Επίσης εξηγούνται οι λόγοι επιλογής των πεδίων / μεταβλητών και σχολιάζεται το πεδίο ορισμού των.

Στο πέμπτο κεφάλαιο (Εφαρμογή Μεθοδολογίας) γίνεται η προετοιμασία των δεδομένων. Ο καθαρισμός του αρχείου δεδομένων (data cleaning) είναι απαραίτητο βήμα οφειλόμενο στον τρόπο δημιουργίας και σχηματισμού του αρχείου. Γίνεται με σκοπό την δημιουργία μοναδικής εγγραφής του ατυχήματος πλοίου. Επινόηθηκαν αλγόριθμοι, προσαρμοσμένοι στις ιδιαιτερότητες του, μια και ο καθαρισμός δεν έχει ακόμα εξελιχθεί σε τυποποιημένη εργασία και δεν υπάρχει σχετική μεθοδολογία. Αυτοί αντικατέστησαν τις ελλείπουσες τιμές, όπου υπήρχε η δυνατότητα, καθώς και εντόπισαν και αποκατέστησαν στο ορθόν τις διαφορετικές τιμές. Στη συνέχεια, παρουσιάζονται τα βήματα, για την εφαρμογή του αλγόριθμου C5.0 και των λοιπών τεχνικών που χρησιμοποιήθηκαν για την ανάλυση του αρχείου δεδομένων. Ορίζονται όλες οι παράμετροι καθώς και οι αναγκαίες συνθήκες και καθορίζονται τα ανεξάρτητα και εξαρτημένα πεδία / μεταβλητές για να εφαρμοστεί η μεθοδολογία που παρουσιάστηκε στο τρίτο κεφάλαιο.

Στο έκτο κεφάλαιο (Αποτελέσματα) παρουσιάζονται τα αποτελέσματα από το σύνολο των αναλύσεων και διαδικασιών που εφαρμόστηκαν στην παρούσα διατριβή.

Στο έβδομο κεφάλαιο (Συμπεράσματα) παρουσιάζονται τα συμπεράσματα των μελετών των επί μέρους κεφαλαίων καθώς και συνολικά της διατριβής. Στη συνέχεια αναφέρονται τα οφέλη που μπορεί να έχουν φορείς σχετικοί με την ναυτιλία, Ευρωπαϊκοί και Παγκόσμιοι και προτείνεται:

- να γίνονται οι απαραίτητοι έλεγχοι με έμφαση σε ορισμένους παράγοντες που εντοπίστηκαν,

- να θεσμοθετηθούν κατάλληλα μέτρα αντιμετώπισης των παραγόντων υψηλής επικινδυνότητας που υποδείχθηκαν,

- να γίνονται οι απαραίτητες προβλέψεις,

με σκοπό την αύξηση της ασφάλειας στη Ναυτιλία στα ύδατα της Ευρωπαϊκής Ένωσης και σε παγκόσμιο επίπεδο. Επίσης παρουσιάζονται προτάσεις για μελλοντικές ερευνητικές εργασίες από την αποκτηθείσα εμπειρία από την δημιουργία του αρχείου, από τα νέα προγράμματα, τους νέους αλγόριθμους καθώς και τα συμπεράσματα από την ερμηνεία των αποτελεσμάτων.

Η διατριβή τελειώνει με την παράθεση Ξένων και Ελληνικών βιβλιογραφικών αναφορών που χρησιμοποιήθηκαν.

Στο τέλος της διατριβής στα Παραρτήματα παρουσιάζονται οι συμβολισμοί γραμμογράφησης του αρχείου δεδομένων και οι κωδικοί που χρησιμοποιήθηκαν. Δίνεται αρχείο αποτελεσμάτων του S.P.S.S με στοιχεία που τεκμηριώνουν τα αναφερθέντα αποτελέσματα στα οποία βλέπουμε την επίδραση διαφόρων παραγόντων στην ολική απώλεια του πλοίου. και ειδικότερα εκτός από τους Πίνακες μονοδιάστατης ανάλυσης, τους Πίνακες δυοδιάστατης ανάλυσης και τα Περιγραφικά στοιχεία, παρουσιάζονται τα Δέντρα απόφασης, οι Ταξινομητικοί κανόνες και οι συνδυασμοί παραγόντων που οδηγούν στην ολική απώλεια. Τέλος δίνονται στοιχεία για την μεθοδολογία των νευρωνικών δικτύων, τους νέους αλγόριθμοι και τα προγράμματα που εφαρμόστηκαν καθώς και ο αλγόριθμος C5.0. στα δεδομένα του αρχείου σε γλώσσα C.

## ΚΕΦΑΛΑΙΟ 2

### 2.1 Εισαγωγή

Το αρχείο δεδομένων στο οποίο θα εφαρμοσθούν αναλύσεις περιλαμβάνει πλήθος εγγραφών και μεταβλητών και για να αντιμετωπιστεί το πρόβλημα αυτό, απαιτείται μία νέα γενιά μεθόδων και εργαλείων ανάλυσης. Η τεχνική Όρυξης Δεδομένων εντάσσεται στην ερευνητική περιοχή που αφορά την διαδικασία ανακάλυψης γνώσης και εξαγωγής χρήσιμης πληροφορίας από μεγάλα συστήματα βάσεων δεδομένων

Με χρήση αλγορίθμων όπως ο ID3 ο C4.5 ο C5.0, ο CART καθώς και ο CHAID παράγονται δέντρα αποφάσεων. Η σπουδαιότητα της Όρυξης Δεδομένων με χρήση δέντρων αποφάσεων, σαν μέθοδο ανάλυσης, έχει αναγνωρισθεί από διάφορους επιστημονικούς τομείς όσο και από ευρύτερους τομείς παροχής υπηρεσιών στους οποίους αυτή εφαρμόζεται, όπως ιατρική, μάρκετινγκ, ασφαλιστικές εταιρίες, τράπεζες, τηλεπικοινωνίες. Απομένει να αποδειχθεί η δυνατότητα εφαρμογής της καθώς και η χρησιμότητά της σε άλλους τομείς όπως η ναυτιλία, οι συνδυασμένες μεταφορές κ.α.

Οι πλέον πρόσφατες εφαρμογές τεχνικής Όρυξης Δεδομένων με χρήση δέντρων αποφάσεων καλύπτουν ένα ευρύ φάσμα ερευνητικών εφαρμογών όπως Γεωμορφολογίας - Geomorphology (Thoms et al, 2002), Χημείας - Chemistry (Tong et al, 2002), Ιατρικής - Medicine (Andrews et al, 2002) και (Kosuda et al, 2002) και (Ogan et al, 2002), Διαχείριση στις Βιομηχανίες - Industrial management (Durvasula et al, 2002) και (Lian et al, 2002), Τηλεπικοινωνίες - Telecommunications (Wei et al, 2002).

Αξιοσημείωτη είναι η μέχρι τώρα σχετικά μικρή συγκριτικά συχνότητα εμφάνισης ερευνητικών εργασιών ανάλυσης δεδομένων των ναυτιλιακών ατυχημάτων με εφαρμογή αντιστοίχων με της παρούσας διατριβής τεχνικών. Η μόνη στη διεθνή βιβλιογραφία ερευνητική εργασία του Καθηγητή Κ. Γκιζιάκη, που αποτέλεσε το ερέθισμα για την δημιουργία της παρούσας διατριβής, ανέδειξε και την αναγκαιότητα ύπαρξης βάσης στην οποία να καταγράφονται τα Ναυτικά ατυχήματα (Giziakis K. P.h.d, 1987). Με χρήση τεχνικών Όρυξης σε ερευνητικές εργασίες (Hashemi R. R. et al, 1995) και (Blanc A. L. et al, 2001) σχετικές με ναυτιλιακά



ατυχήματα δείχθηκε αφενός στην έρευνα των Hashemi R. R. et al ότι η πρόβλεψη με τεχνικές Όρυξης Δεδομένων είναι περισσότερο αποτελεσματική σε σχέση με άλλες μεθόδους Στατιστικής ανάλυσης, αφετέρου στην έρευνα των Blanc A. L. et al εξετάστηκε η βελτιστοποίηση ταξινόμησης με σκοπό την πρόβλεψη, εφαρμοσμένης σε δεδομένα ναυτικών ατυχημάτων με τεχνικές Όρυξης Δεδομένων (Νευρωνικά δίκτυα Kohonen).

## 2.2. Γενικά για τις αναφορές και σχετικές παραπομπές

Στις επόμενες παραγράφους του παρόντος υποκεφαλαίου παρουσιάζονται έρευνες οι οποίες εντάσσονται σε δύο διακριτές μεταξύ τους ομάδες, είναι οι έρευνες που αφορούν αναλύσεις δεδομένων σχετικών με τα ναυτικά ατυχήματα και οι έρευνες που είναι σχετικές με τη μεθοδολογία που εφαρμόστηκε στη παρούσα διατριβή.

Καταρχήν παρουσιάζεται ο κατά τον I.M.O ορισμός του ναυτικού ατυχήματος και μια θεωρητική προσέγγιση ανάλυσης του [Kristiansen, 1995]. Παρουσιάζεται και σχολιάζεται αρχικά, η ερευνητική εργασία των [Giziakis et al, 1997], στην οποία καθορίστηκε το γενικό πλαίσιο το οποίο αποτέλεσε οδηγό για τη δημιουργία της βάσης με την παρούσα μορφή.

Αναφέρονται έρευνες [Vijayshankar R. et al, 2001], [Bonchi F. et al, 1999] και [Fawcett T., 2001] που είναι σχετικές με μεθοδολογίες που εφαρμόστηκαν στη διατριβή. Στην ερευνητική εργασία των Vijayshankar R. et al παρουσιάστηκε η υλοποίηση ενός συστήματος καθαρισμού δεδομένων το οποίο εφαρμόστηκε σε μία βάση δεδομένων η οποία είχε κατασκευαστεί από διαφορετικές πηγές σε διαφορετικά χρονικά διαστήματα. Προτάθηκε ένα αλληλεπιδρόν σύστημα καθαρισμού δεδομένων που κάνει μετασχηματισμούς και συγχρόνως ανιχνεύει τα δεδομένα για πιθανά λάθη. Από τους Bonchi F. et al παρουσιάστηκε πως από ιστορικά δεδομένα γίνεται να εκτιμηθεί αν θα συμβεί ή όχι ένα γεγονός. Χρησιμοποιήθηκαν δέντρα αποφάσεων φτιαγμένα με την μέθοδο C5.0 για να κάνουν πρόβλεψη ενός γεγονότος με βάση ορισμένους άλλους παράγοντες. Τονίστηκε ότι η ανεύρεση γνώσης σε βάσεις δεδομένων αντιμετωπίζει δύο κυρίαρχα ζητήματα την επιλογή της μεθοδολογίας και την ικανότητα να ανιχνεύονται τα συστατικά στοιχεία της βάσης. Τέλος ο συγγραφέας Fawcett T. αναφέρθηκε στην εργασία του ουσιαστικά στα

πλεονεκτήματα από τη χρήση των κανόνων ταξινόμησης. Είναι μία εργασία που αναφέρεται κυρίως ως σημείο αναφοράς στην περίπτωση παραγωγής ορισμένων σεναρίων.

Με τεχνικές όρυξης δεδομένων έγινε αξιόπιστη ανάλυση ναυτικών ατυχημάτων και από τους [Hashemi R. R. et al. ,1995] και [Le Blanc A. L. et al. 2001]. Ανιχνεύτηκε η σχετική ακρίβεια στην πρόβλεψη ατυχημάτων, στην ερευνητική εργασία των Hashemi R. R. et al, με τεχνική όρυξης δεδομένων και με στατιστικές αναλύσεις. Παρουσιάστηκε μία μεθοδολογία νευρωνικών δικτύων και σαν αποτέλεσμα οι συγγραφείς υποστηρίζουν ότι ο αλγόριθμος τους ταξινομεί τις εγγραφές καλύτερα από τις κλασικές μεθόδους της στατιστικής. Σχολιάζεται μία ακόμα εργασία, των Le Blanc A. L. et al, που αντιμετώπισε το θέμα της τμηματοποίησης των ατυχημάτων με στατιστική ανάλυση και όρυξη δεδομένων (δίκτυα Kohonen). Τονίζεται η σημασία της χρήσης των προσδιοριστικών μεταβλητών στην προσπάθεια ομαδοποίησης (clustering) των ναυτικών ατυχημάτων. Επίσης σημειώνεται ότι η χρήση διαφορετικών τεχνικών στο ίδιο σύνολο δεδομένων μπορεί να παράγει διαφορετικά αποτελέσματα.

Από τις λοιπές σχετικές ερευνητικές εργασίες σχολιάζονται των [Psraftis Ch. et al. 1998] και [Karydis P. et al, 1998]. Στην εργασία των Psraftis Ch. et al, βάση των πληροφοριών σ' ένα περιορισμένο πληθυσμό ατυχημάτων, για τα οποία υπήρχαν όμως διαθέσιμες λεπτομερείς πληροφορίες, εντοπίστηκε η αιτία καθενός από αυτά τα ατυχήματα. Επίσης έγινε απολογισμός του τι θα πετύχαινε ένα συγκεκριμένο σχέδιο μείωσης του κινδύνου. Στην εργασία των Karydis P. et al οι συγγραφείς έδωσαν μία σύντομη περιγραφή της Μεθόδου Δικτύων Σφαλμάτων και εφάρμοσαν την Μέθοδο των Δικτύων Γεγονότων σε μια συγκεκριμένη περίπτωση.

Επειδή, τα τελευταία χρόνια, από τους Διεθνείς οργανισμούς έχει δοθεί έμφαση στη διερεύνηση των ναυτικών ατυχημάτων έγιναν από αυτούς σχετικές μελέτες με σκοπό να υιοθετήσουν την πλέον κατάλληλη πολιτική και τα πλέον αρμόζοντα μέτρα για την ενίσχυση της ασφάλειας στο χώρο της Ναυτιλίας. Ενδεικτικά λοιπόν αναφέρονται μελέτες που έγιναν από τους Bureau of Transport and Communications Economics Australia καθώς και από Γενική Διεύθυνση Μεταφορών DGVII της Ευρωπαϊκής Ένωσης (SAFECO). Στην ανάλυση που έκαναν οι ερευνητές Gentle N, Wheatstone S, Salmi J, και Breusch T. δείχθηκε ότι με έλεγχο του 34% των Bulk Carriers που επισκέπτονται τα λιμάνια της Αυστραλίας, μπορούν να προλάβουν

το 93% των πλοίων που θα έχουν ατυχήματα. Στο SAFECO I διερευνήθηκαν τυχόν υπάρχουσες σχέσεις ανάμεσα στις διάφορες πιθανές αιτίες πρόκλησης ατυχήματος και στο τελικό αποτέλεσμα αυτού του ατυχήματος αναλύοντας τους παράγοντες που πιθανά συνεισφέρουν στη δημιουργία ατυχημάτων. Στόχος του SAFECO II ήταν να παρουσιάσει νέες και βελτιωμένες τεχνολογικά λύσεις. Σκοπός των SAFECO I και II ήταν να αυξήσουν την ασφάλεια της Ναυσιπλοίας στα Ευρωπαϊκά παράκτια ύδατα

### 2.2.1. Ορισμοί και βασικές θεωρητικές προσεγγίσεις

Κρίθηκε σκόπιμο, πριν παρουσιασθούν αντίστοιχες έρευνες, να παρουσιαστεί ο κατά τον Ι.Μ.Ο ορισμός του ναυτικού ατυχήματος.

Ο Ι.Μ.Ο υιοθέτησε τον Κώδικα για τη Διερεύνηση Ναυτικών Ατυχημάτων και Συμβάντων [Ι.Μ.Ο, 1997]. Σκοπός του Κώδικα είναι να προωθήσει μία κοινή προσέγγιση στη διερεύνηση ατυχημάτων και τη συνεργασία μεταξύ κρατών για την αναγνώριση των παραγόντων που συμβάλλουν στα ναυτικά ατυχήματα. Αποτέλεσμα της κοινής προσέγγισης και συνεργασίας θα είναι η υποβοήθηση λήψης διορθωτικών μέτρων και αύξηση της ασφάλειας.

Ο Κώδικας ορίζει σαν ναυτικό ατύχημα (marine casualty) κάθε συμβάν το οποίο έχει σαν αποτέλεσμα:

1. το θάνατο ή το σοβαρό τραυματισμό ατόμου που προκλήθηκε από, ή σε σχέση, με τις λειτουργίες του πλοίου,
2. την απώλεια προσώπου από το πλοίο που προκλήθηκε από, ή σε σχέση, με τις λειτουργίες του πλοίου,
3. την απώλεια, τεκμαρτή απώλεια ή εγκατάλειψη πλοίου,
4. την υλική ζημία στο πλοίο,
5. την προσάραξη ή ανικανότητα πλοίου ή την εμπλοκή του σε σύγκρουση,
6. την υλική ζημία που προήλθε από, ή σε σχέση με, τις λειτουργίες του πλοίου,
7. ζημία στο περιβάλλον που συνέβη από τη βλάβη πλοίου ή πλοίων και που προκλήθηκε από, ή σε σχέση με, τις λειτουργίες του πλοίου ή των πλοίων.

Δίνεται στην συνέχεια χαρακτηρισμός των ατυχημάτων. “Πολύ σοβαρό ατύχημα” θεωρείται εκείνο που έχει σαν αποτέλεσμα την απώλεια του πλοίου, απώλεια ζωής ή σοβαρή ρύπανση. “Σοβαρό ατύχημα” θεωρείται το μη πολύ σοβαρό ατύχημα και



περιλαμβάνει πυρκαγιά, έκρηξη, προσάραξη, πρόσκρουση, βαριά ζημιά λόγω καιρού και ρήγματα, που έχουν σαν αποτέλεσμα ζημιές στο πλοίο που το καθιστούν αναξιόπλοο, ρύπανση, βλάβη ή ανάγκη ρυμούλκησης ή βοήθειας από την ξηρά.

Ο Κώδικας περιγράφει τις αρχές που διέπουν τη διερεύνηση των ατυχημάτων, πως πρέπει να εφαρμοσθούν αυτές και συνοψίζει τους αντικειμενικούς στόχους της διερεύνησης, με πρώτο στόχο αυτόν της πρόληψης ατυχημάτων στο μέλλον.

Κατά τον Kristiansen μια ιδεατή μεθοδολογία ανάλυσης ναυτικών ατυχημάτων θα πρέπει να ικανοποιεί τις ακόλουθες απαιτήσεις [Kristiansen, 1995]:

1. Να αντικατοπτρίζει τη χρονική συνέχεια και αλληλουχία ενός πιθανού αριθμού γεγονότων ή δυσλειτουργιών.
2. Να επιτρέπει τον εντοπισμό των εργασιών ή διαδικασιών που δεν πραγματοποιήθηκαν ή που πραγματοποιήθηκαν σε βαθμό κάτω του προσδοκώμενου.
3. Να διακρίνει τα αίτια σε τεχνικές αστοχίες, σφάλματα, και ακραίες περιβαλλοντικές επιδράσεις.
4. Να επιτρέπει τον συσχετισμό των αστοχιών με τα ακόλουθα βασικά είδη αιτιών: Τεχνικός παράγοντας, περιβάλλον χρήστη-μηχανής, διαδικασίες, οργάνωση υποστήριξης και περιβάλλον.
5. Να είναι ρεαλιστική, να προτείνει δηλαδή προληπτικά μέτρα καθώς και μέτρα μείωσης συνεπειών που είναι εφικτά.

Σε εργασία των ερευνητών του Εργαστηρίου Ναυτικών Ατυχημάτων του Πανεπιστημίου Πειραιά [Giziakis et al, 1997] διαπιστώθηκε η ύπαρξη γενικής αναγνώρισης της πολυπλοκότητας και διεπιστημονικότητας της ασφάλειας στο χώρο της Ναυτιλίας.

Αναφέρθηκε σε αυτήν ότι :

- τα Ναυτιλιακά Ατυχήματα είναι απόρροια της αδυναμίας κατανόησης του τρόπου λειτουργίας του "Ναυτιλιακού Συστήματος",
- το πρόβλημα της διερεύνησης των αιτιών των Ναυτικών Ατυχημάτων είναι πολυπαραμετρικό και δεν έχει νόημα να μιλάμε για κάποιες απλές αιτίες ατυχημάτων,
- υπάρχει δυσκολία στον καθορισμό των παραγόντων που επιδρούν στα ατυχήματα,

- είναι εξαιρετικά δύσκολο να απομονωθεί ένα χαρακτηριστικό του “Ναυτιλιακού Συστήματος” που φαίνεται να επηρεάζει την εμφάνιση των ατυχημάτων.

Η ικανότητά μας να αντιμετωπίσουμε τα προαναφερθέντα εξαρτάται:

- από το βαθμό κατανόησης της λειτουργίας του “Ναυτιλιακού Συστήματος” και των παραγόντων που επιδρούν στην ασφάλειά του,
- την ικανότητά μας να διαλέγουμε και να εφαρμόζουμε αποδοτικά προγράμματα βελτίωσης.

Το ζητούμενο είναι να μπορούν να παρθούν κάποια μέτρα, ώστε να επιτευχθεί η πρόληψη, η αποφυγή και άρα η μείωση του αριθμού των ατυχημάτων.

Η δυσκολία στην μείωση του αριθμού των Ναυτικών Ατυχημάτων έκανε φανερή την έλλειψη επιστημονικής γνώσης όσον αφορά τους παράγοντες που επιδρούν σε αυτά. Για την επιστημονική ανάλυση και εξαγωγή τεκμηριωμένων συμπερασμάτων, ώστε να καλυφθεί όσο το δυνατόν ευρύτερα και διεπιστημονικά το θέμα, έγινε πρόταση για δημιουργία Βάσης Δεδομένων η οποία ήταν βασική ιδέα του Καθηγητή Κ. Γκιζιάκη που άρχισε να την επεξεργάζεται από το 1978 (Giziakis K. P.h.d, 1987).

Οι ερευνητές πιστεύουν ότι η Βάση Δεδομένων πρέπει να αναπτυχθεί και να αναλυθεί διεπιστημονικά. Υποστηρίχθηκε ότι υπάρχει ανάγκη να προσεγγίζεται το αντικείμενο της ασφάλειας της Ναυσιπλοΐας από διαφορετικές πλευρές και αυτό ακριβώς αποτελεί και το συγκριτικό πλεονέκτημα από την δημιουργία βάσης με αυτή την διεπιστημονική προσέγγιση. Διαπιστώθηκε η απαίτηση να εκφραστούν οι ανάγκες της ανάλυσης από την πλευρά πολλών επιστημών, Οικονομικών, Κοινωνικών, Τεχνικών. Χρειάζεται η εφαρμογή επιστημών της Πληροφορικής και των Μαθηματικών, πράγμα που αποδεικνύεται και στην παρούσα ερευνητική εργασία.

Διαπιστώνεται τέλος πως η κατανόηση της λειτουργίας του “Ναυτιλιακού Συστήματος”. θα επιτευχθεί με την ανάλυση των πληροφοριών της βάσης που θα περιέχει όλα τα Ναυτικά ατυχήματα, μέσα από την διερεύνηση των παραγόντων που επιδρούν στα ατυχήματα, ούτως ώστε να βοηθηθούν αυτοί που παίρνουν αποφάσεις να λάβουν τα κατάλληλα μέτρα για τη μείωση του αριθμού τους.

### 2.2.2. Ανάλυση ναυτικών ατυχημάτων με τεχνικές Όρυξης δεδομένων

Στην ερευνητική εργασία των Hashemi R. R. et al. τα ατυχήματα που καταγράφηκαν για την περιοχή των εκβολών του ποταμού Μισισσιππή (Lower Mississippi River) είχαν γίνει κάτω από διαφορετικές συνθήκες πλοήγησης.

Η Όρυξη δεδομένων παρέχει ένα σύνολο εργαλείων τα οποία μπορούν να χρησιμοποιηθούν για την αντιμετώπιση ενός προβλήματος ταξινόμησης όπως είναι αυτό που παρουσιάζεται στην παρούσα διατριβή.

Στο συγκεκριμένο άρθρο [Hashemi R. Ray et al., 1995] παρουσιάστηκε μία μέθοδος Όρυξης δεδομένων με χρήση νευρωνικών δικτύων. Αφού έκαναν την θεωρητική εισαγωγή στη μεθοδολογία παρουσιάζοντας τον αλγόριθμο και ορισμένες εφαρμογές του στην αντιμετώπιση προβλημάτων, συνέκριναν τον αλγόριθμο με βάση τις παραδοσιακές μεθόδους ελέγχου υποθέσεων σε περισσότερα από ένα σημεία όπως

- τα δεδομένα των αλγορίθμων
- το παραγόμενο αποτέλεσμα
- τη συναρτησιακή μορφή του μοντέλου
- τη βαρύτητα των συντελεστών

Στη συνέχεια εφάρμοσαν την θεωρία τους για τον έλεγχο ικανότητας ταξινόμησης των δεδομένων στις εξής κατηγορίες ατυχημάτων:

1. COLLISION.
2. CONTACT.
3. GROUNDING.

Χρησιμοποιήθηκαν και εξωτερικές πληροφορίες όπως αν μετείχε στο συμβάν και άλλο πλοίο, η θέση του πλοίου στην περιοχή του I.M.R, η κίνηση στο ποταμό, ο χρόνος, οι καιρικές συνθήκες κ.ο.κ.

Η όλη ερευνητική εργασία απέδειξε όπως φαίνεται και από τον Πίνακα 2.1 ότι με την τεχνική της Όρυξης δεδομένων έχουμε μεγαλύτερη ακρίβεια στην πρόβλεψη από ότι με τις κλασσικές μεθόδους Στατιστικής. Με την τεχνική αυτή έγινε πρόβλεψη σωστή κατά 80% ενώ με την εφαρμογή τεχνικών Διαχωριστικής ανάλυσης (Discriminant) 53% και εφαρμογή τεχνικών Παλινδρόμησης (Regression) 56%.

Όσον αφορά το ποσοστό ταξινόμησης, με την συγκεκριμένη μεθοδολογία ταξινομήθηκαν σωστά το 90% των στοιχείων που είχαν στη διάθεσή τους. Μάλιστα το 80% των σωστά ταξινομημένων περιπτώσεων έχει υπολογιστεί θεωρώντας ότι το



10% των περιπτώσεων τις οποίες ο αλγόριθμος δεν μπόρεσε να ταξινομήσει είναι λάθος ταξινομημένες. Κατάφερε δηλαδή ο αλγόριθμος να παρουσιάσει και τις ειδικές περιπτώσεις όπου ο γενικός κανόνας δεν εφαρμόζεται άμεσα.

Παρατηρώντας τις στήλες του Πίνακα 2.1 για τα ατυχήματα “COLLISION” παρατηρούμε ότι την καλύτερη πρόβλεψη κάνει η τεχνική της Όρυξης Δεδομένων με ποσοστό 67% έναντι 33% για την Παλινδρόμηση και 25% για την Διαχωριστική Ανάλυση. Ανάλογα ποσοστιαία αποτελέσματα έχουμε για τα ατυχήματα “CONTACT” με την τεχνική της Όρυξης Δεδομένων ποσοστό 77% έναντι 62% με την Παλινδρόμηση και 62% με την Διαχωριστική Ανάλυση και για τα ατυχήματα “GROUNDING” με την τεχνική της Όρυξης Δεδομένων ποσοστό 90% έναντι 65% με την Παλινδρόμηση και 65% με την Διαχωριστική Ανάλυση.

ΠΙΝΑΚΑΣ 2.1

Ποσοστά πρόβλεψης ανά ατύχημα και τεχνική ανάλυσης

	TOTAL	COLLISION	CONTACT	GROUNDING
Όρυξη Δεδομένων	80%	67%	77%	90%
Διαχωριστική Ανάλυση.	53%	25%	62%	65%
Παλινδρόμηση	56%	33%	62%	65%

Το όφελος από την χρήση των μεθοδολογιών Όρυξης Δεδομένων στην ερευνητική εργασία των Hashemi R. R. et al οφείλεται και στο ότι οι μεθοδολογίες αυτές βασίζονται σε μεγάλο βαθμό στο ότι δεν κάνουν a priori προϋποθέσεις για τη φύση της σχέσης στα δεδομένα. Μπορούν και επεξεργάζονται πολλαπλούς γραμμικούς συνδυασμούς παράλληλα αναζητώντας ταυτόχρονα οποιαδήποτε μορφής σχέση, όχι μόνο την απλή γραμμική. Αντίθετα οι κλασικές μέθοδοι ζητούν την ύπαρξη της γνώσης εκ των προτέρων της προς διερεύνηση σχέσης.

Πιστεύουν οι συγγραφείς της εργασίας ότι μελλοντικά με τη χρήση του Αλγόριθμου ID3 (Algorithm ID3) (Quinlan, 1979), των Ασαφών Συνόλων (Rough Sets) (Pawlak, 1984) θα μπορούν να έχουν καλή πρόβλεψη των ατυχημάτων.

Εδώ θα πρέπει να σημειωθεί ότι ο αλγόριθμος C5.0 που θα χρησιμοποιηθεί στην παρούσα διατριβή είναι η εξέλιξη του αλγορίθμου C4.5 ο οποίος είναι εξέλιξη του αλγορίθμου ID3 και στηρίζεται στα Τροποποιημένα Ασαφή Σύνολα και στους

Γενετικούς Αλγόριθμους (Modified Rough Sets and Genetic Algorithms) [Hashemi et al. 1997]. Και οι τρεις αλγόριθμοι είναι υλοποιημένοι από τον Quinlan και την ερευνητική ομάδα του και έχουν εφαρμοσθεί με επιτυχία σε πάρα πολλά σύνολα δεδομένων.

Σε άλλη ερευνητική εργασία [Le Blanc A. L. et al, 2001] μελετήθηκε δείγμα με πλήθος ατυχημάτων περισσότερων των 936 στην ευρύτερη περιοχή της Ν. Ορλεάνης (New Orleans - Gulf of Mexico).

Σε αυτή την ερευνητική εργασία ομαδοποίησαν τα δεδομένα, σε κατηγορίες σαφώς διαφορετικές μεταξύ τους, αφ' ενός με Όρυξη Δεδομένων και συγκεκριμένα με χρήση του Νευρωνικού δικτύου Kohonen (S.O.M.- Self Organizing Map) και αφ' ετέρου με μεθοδολογία Ομαδοποίησης (Cluster analysis).

Όπως τονίζεται το να ταξινομηθούν τα δεδομένα σε ομάδες με τη μέθοδο της τμηματοποίησης είναι σα να «ψάχνεις βελόνα στα' άχυρα». Είναι εξαιρετικά δύσκολο να βρεις ομάδες με βάση τα κοινά χαρακτηριστικά των προς διερεύνηση πεδίων / μεταβλητών. Και όπως αναδεικνύουν τα ευρήματά τους μπορούν να κατασκευαστούν πολλές διαφορετικές σε αριθμό ομάδες από το ίδιο σύνολο δεδομένων.

Περισσότερο αναλυτικά για χρονική διάρκεια περίπου δέκα ετών κατέγραψαν τα ναυτικά ατυχήματα για την συγκεκριμένη περιοχή. Για το σκοπό αυτό χρησιμοποίησαν τα εξής πεδία / μεταβλητές:

- K1. Μεταβλητή που αναφέρεται στη στάθμη των υδάτων.
- K2. Μεταβλητή για το πλήθος μετακινήσεων των πλοίων.
- K3. Μεταβλητή για το ποσοστό κυκλοφορίας στην περιοχή.
- K4. Μεταβλητή για τον τύπο ατυχήματος (σύγκρουση με πλοίο, πρόσκρουση με σταθερό αντικείμενο, προσάραξη, κ.α.).
- K5. Μεταβλητή για τον χρόνο που συνέβη το ατύχημα.
- K6. Μια διχοτομική μεταβλητή που δηλώνει αν το πλοίο που είχε ατύχημα είχε επικοινωνία με το NOLA-VTS's control system.
- K7. Μεταβλητή για τις καιρικές συνθήκες.
- K8. Μεταβλητή για τις περιοχές που συνέβησαν τα ατυχήματα.

Τα σχόλια που έγιναν στη συνέχεια αφορούσαν τις τέσσερες ομάδες και μάλιστα οι ομάδες που μελετήθηκαν με βάση την τεχνική της Ομαδοποίησης ήταν :

- Cgr1. Ατυχήματα πλοίων σε επικίνδυνα νερά.
- Cgr2. Ατυχήματα πλοίων με κακές συνθήκες και καλούς πλοηγούς.
- Cgr3. Ατυχήματα που θα μπορούσαν πιθανά να αποφεύγονταν.

Cgr4. Ατυχήματα που θα έπρεπε να είχαν προβλεφθεί.

Στις ομάδες με την χρήση της Ομαδοποίησης για την ομάδα Cgr1 οι εμπλεκόμενοι ναυτικοί θεωρήθηκαν επαρκώς ενημερωμένοι, για την ομάδα Cgr2 και για την ομάδα Cgr3 χαρακτηρίστηκαν από διαφορές στις συνθήκες πλοήγησης και στην ομάδα Cgr4 οι εμπλεκόμενοι θεωρήθηκαν ανεπαρκώς ενημερωμένοι.

Οι ομάδες που μελετήθηκαν με χρήση του S.O.M. ήταν :

Dgr1. Ατυχήματα πλοίων με μεγάλη δυσκολία πλοήγησης.

Dgr2. Ατυχήματα πλοίων με μέτρια δυσκολία πλοήγησης.

Dgr3. Ατυχήματα σε πλοία που πλοηγούνται εύκολα.

Dgr4. Ατυχήματα σε πλοία που βρέθηκαν σε γρήγορο ρεύμα, αλλά μικρή κυκλοφορία.

Στις ομάδες με την χρήση της S.O.M. για την ομάδα Dgr1 και Dgr2 θεωρήθηκαν μικρές διαφορές στις συνθήκες πλοήγησης, για την ομάδα Dgr3 τα ατυχήματα έγιναν όλα όταν υπήρχε αργό ρεύμα, μικρή κυκλοφορία και κανένα ατύχημα δεν έγινε με κακές καιρικές συνθήκες, και για την ομάδα Dgr4 έγιναν όταν υπήρχε γρήγορο ρεύμα και μικρή κυκλοφορία.

Για την πρώτη τεχνική της Ομαδοποίησης και για τις δύο ομάδες, Cgr1 για 224 περιπτώσεις και Cgr4 για 345 περιπτώσεις ταξινομήθηκαν τέλεια οι περιπτώσεις με ποσοστό 100%, για την ομάδα Cgr2 για 223 περιπτώσεις ταξινομήθηκαν με ποσοστό 89.7%, και για την Cgr3 για 134 περιπτώσεις ταξινομήθηκαν με ποσοστό 88.1%.

Για τη δεύτερη τεχνική και για την ομάδα Dgr3 ταξινομήθηκαν τέλεια οι περιπτώσεις με ποσοστό 100%, ενώ για την ομάδα Dgr1 ταξινομήθηκαν με ποσοστό 96.7% για 608 περιπτώσεις, για την Dgr2 ταξινομήθηκαν με ποσοστό 80,5%. για 118 περιπτώσεις και για την Dgr4 ταξινομήθηκαν με ποσοστό 94,6%. για 202 περιπτώσεις.

Ένα άλλο χαρακτηριστικό μεταξύ των τμημάτων των ομάδων ήταν η τοποθεσία του ατυχήματος όπως αυτές ορίστηκαν στα δεδομένα του προς διερεύνηση Προβλήματος (Χάρτης 2.2). Με χρήση της τεχνικής της Ομαδοποίησης για όλες τις ομάδες υπήρχαν ατυχήματα στις γεωγραφικές περιοχές III IV στο χάρτη 2.2 με ελάχιστο άθροισμα ποσοστών 40%. Με την χρήση τεχνικής S.O.M. στην γεωγραφική περιοχή II οι τρεις ομάδες Dgr2, Dgr3, Dgr4 είχαν άνω του 25% των ατυχημάτων για κάθε μια τους. Στις γεωγραφικές περιοχές III IV δεν υπήρχε ποσοστό περισσότερο από 27% των ατυχημάτων για τις ομάδες Dgr1, Dgr2. Επίσης στους πιο



απομακρυσμένους από την ανοικτή θάλασσα τομείς III IV η ομάδα Dgr4 δεν είχε κανένα ατύχημα.

Χάρτης 2.2



Οι συγγραφείς εκτέλεσαν ανάλυση ομάδων με πέντε, τέσσερις και με τρεις ομάδες ως ζητούμενο. Για την ανίχνευση της ισχύος της ομαδοποίησης οι συγγραφείς χρησιμοποίησαν Διαχωριστική Ανάλυση για να προσδιορίσουν με τις ίδιες ανεξάρτητες μεταβλητές το αποτέλεσμα της ανάλυσης των ομάδων. Η Διαχωριστική Ανάλυση δίνει αποδεκτά ποσοστά ταξινόμησης του αποτελέσματος της ανάλυσης ομάδων με τα δεδομένα που χρησιμοποιήθηκαν. Ελαφρώς καλύτερα ήταν τα αποτελέσματα για την ανάλυση με τέσσερις ομάδες, γεγονός που σημαίνει ότι τα ναυτικά ατυχήματα περιγράφονται καλύτερα με τέσσερις διαφορετικές ομάδες. Η

τεχνική της Ομαδοποίησης ταξινομήσε σωστά το 95,7% των 936 ατυχημάτων ενώ με χρήση της S.O.M. ταξινομήθηκαν σωστά το 94,3% των 936 ατυχημάτων.

Συγκριτική εικόνα των αποτελεσμάτων δίνεται στον Πίνακα 2.3 που ακολουθεί ανά πέντε, ανά τέσσερες και ανά τρεις ομάδες.

### ΠΙΝΑΚΑΣ 2.3

Συγκριτικός πίνακας πλήθους ταξινομηθεισών περιπτώσεων ανά ομάδες και ανά τεχνική ανάλυσης

#### Ανά πέντε ομάδες

	Group 1	Group 2	Group 3	Group 4	Group 5	Hit Ratio
Ομαδοποίηση	258	139	185	206	148	94.4%
Όρυξη Δεδομένων	112	118	257	441	4	93.3%

#### Ανά τέσσερες ομάδες

Ομαδοποίηση	224	223	134	345		95.7%
Όρυξη Δεδομένων	608	118	4	202		94.3%

#### Ανά τρεις ομάδες

Ομαδοποίηση	222	348	366			94.8%
Όρυξη Δεδομένων	366	557	9			97.6%

Στα τμήματα με την χρήση τεχνικής Ομαδοποίησης δεν υπάρχουν σημαντικές διαφορές αναφορικά με τον αριθμό των μελών τους. Στα τμήματα με χρήση της S.O.M. υπάρχουν δύο ομάδες με 4 περιπτώσεις δύο φορές στα Group 5 και Group 3

αντίστοιχα και μια ομάδα στο Group 3 με 9 περιπτώσεις. Για τα τμήματα με τέσσαρες και πέντε ομάδες είναι οι ίδιες τέσσαρες εγγραφές με τις ακραίες τιμές μερικών χαρακτηριστικών τους. Οι 9 περιπτώσεις στην ομάδα Group 3 δημιουργημένες με χρήση της S.O.M. περιέχουν 4 περιπτώσεις εγγραφών συν επιπλέον 5 εγγραφές.

Με την μεθοδολογία που θα αναπτυχθεί στην παρούσα διατριβή θα επαληθευθούν οι παρατηρήσεις των συγγραφέων για την εγκυρότητα χρήσης της ανάλυσης ομάδων και της προτίμησης των κανόνων και της ταξινομητικής ανάλυσης ως μεθόδου ανάλυσης δεδομένων.

Τέλος, οι συγγραφείς τονίζουν τη σημασία της χρήσης των προσδιοριστικών μεταβλητών στην ανάλυση ομάδων ναυτικών ατυχημάτων. Επίσης σημειώνουν ότι η χρήση διαφορετικών τεχνικών στο ίδιο σύνολο δεδομένων μπορεί να παράγει διαφορετικά αποτελέσματα. Οι συνέπειες αυτής της παρατήρησης είναι όπως ακριβώς δηλώνουν και οι συγγραφείς "a voluntary tracking service may be subject to debate". Τα ανωτέρω παραπέμπουν στο γεγονός ότι έχει ήδη δημιουργηθεί μία νέα επιστήμη (Meta-analysis) βασισμένη σε ανάλογες με την παραπάνω παρατηρήσεις.

### 2.2.3. Ερευνητικές εργασίες σχετικές με την μεθοδολογία

Ακολουθεί αναφορά σε εργασία σχετική με χρήση μεθοδολογιών όρυξης δεδομένων και ειδικά στην εφαρμογή διαδικασιών καθαρισμού (data cleaning).

Στην εργασία αυτή [Vijayshankar R. et al, 2001], επισημαίνεται ότι ο καθαρισμός των δεδομένων αποτελεί μία από τις σημαντικότερες εργασίες στη βάση δεδομένων. Σκοπός της είναι να απομακρύνει τις προβληματικές εγγραφές αναφορικά με τη δομή της βάσης, τη μορφή των δεδομένων και τις εξωτερικές συνθήκες που προσδιορίζουν τη δημιουργία της βάσης.

Για ένα σύστημα καθαρισμού σύμφωνα με την εργασία αυτή απαιτείται:

- η συνεχής διαδικασία να γίνεται σε βήματα,
- η γνώση των συγκεκριμένων δομών της βάσης,
- η αλληλεπίδραση μεταξύ του ερευνητή και των στοιχείων για να εκτιμάται η

ποιότητα αυτών.

Τα στάδια καθαρισμού δεδομένων και η αρχιτεκτονική γενικότερα ενός συστήματος καθαρισμού δεδομένων διαμορφώνεται σε σχέση με:



1. τις πηγές δεδομένων,
2. την δυνατότητα άμεσης επαφής με τα δεδομένα,
3. τον καθορισμό μετασχηματισμών στα δεδομένα,
4. τον μηχανισμό αυτόματης ανίχνευσης διαφοροποιήσεων,
5. την δυνατότητα εφαρμογής των επιθυμητών μετασχηματισμών,
6. την επανάληψη της διαδικασίας έως ότου φθάσουμε στο επιθυμητό αποτέλεσμα.

Στην συνέχεια προτείνεται από τους συγγραφείς να γίνεται :

- Ανίχνευση της βάσης.

Στο στάδιο αυτό ανιχνεύονται οι ιδιότητες της βάσης καθώς και η δομή της. Ειδικότερα θα πρέπει να επισημαίνονται οι ιδιαιτερότητες κάθε βάσης.

- Κανονικοποίηση των τιμών των πεδίων / μεταβλητών.

Ενδογενής στη φάση του καθαρισμού των δεδομένων είναι η ικανότητα να οριοθετηθεί η φύση των δεδομένων. Με αυτό τον τρόπο μπορεί να γίνει εφικτή η ανίχνευση των «λογικών» σφαλμάτων στη βάση δεδομένων. Για κάθε ένα πεδίο / μεταβλητή αναζητούνται κανόνες στη σύνταξη των τιμών τους.

- Αντιπαράταξη (matching) εγγραφών.

Σε αυτήν ζητείται να αποσαφηνιστεί η ύπαρξη ή όχι διπλοεγγραφών. Η αντιπαράταξη προβλέπει τη χρήση κάποιου μετασχηματισμού (transformation function) για την κατασκευή δεικτών (index keys) οι οποίοι στη χρησιμεύουν στην σύγκριση με τις εγγραφές που έχουν επίσης ίδιους δείκτες ως προς κάποια κριτήρια. Από τη σύγκριση των συναφών τιμών παράγεται συμπέρασμα για τη σχέση των εγγραφών.

Το στάδιο αυτό είναι το παραγωγικό μέρος του καθαρισμού δεδομένων αφού εδώ διαμορφώνονται όλες οι υπορουτίνες απαραίτητες για την εφαρμογή της διαδικασίας. Βασικό στοιχείο των ρουτινών ανίχνευσης της βάσης είναι η ικανότητά τους να μετασχηματίζουν τις τιμές των πεδίων / μεταβλητών και να μπορούν να απεικονίσουν τις εγγραφές :

είτε σε σχέση «ένα προς ένα» όπου αν επιχειρήσουμε να κάνουμε αναλογικά εφαρμογή στα δεδομένα μας απεικονίζουμε ένα συγκεκριμένο πλοίο σε ένα άλλο συγκεκριμένο πλοίο

είτε σε σχέση της μορφής «πολλά με πολλά» όπου για τα δεδομένα μας στην περίπτωση που πρόκειται να συγκρίνουμε τα ατυχήματα ενός πλοίου με τα ατυχήματα ενός άλλου πλοίου για να αποφανθούμε σχετικά.

- Ανίχνευση διαφοροποιήσεων.

Η διαδικασία ανίχνευσης διαφοροποιήσεων θα πρέπει να έχει ορισμένες ιδιότητες με σημαντικότερες την ευκολία στη χρήση και τη δυνατότητα να ανακαλείται η υλοποίηση του μετασχηματισμού, αφού ο χρήστης διαπιστώνει ότι δεν εφαρμόζεται σε όλες τις εγγραφές του αρχείου

- Εφαρμογή μετασχηματισμών.

Η εφαρμογή των υπορουτινών στη βάση των δεδομένων αποτελεί το σημαντικότερο σημείο κατά το οποίο ο ερευνητής αποφασίζει υπέρ ή όχι κάποιου μετασχηματισμού των δεδομένων.

- Έλεγχος και εκτίμηση των αποτελεσμάτων.

Το τελικό στάδιο του καθαρισμού προσδιορίζει και εκείνες τις συνθήκες με τις οποίες είμαστε ικανοποιημένοι από την ποιότητα των εγγραφών. Στην καλύτερη περίπτωση η ύπαρξη δεύτερης ανεξάρτητης πηγής δεδομένων βάσει της οποίας μπορεί να γίνει η διασταύρωση των αποτελεσμάτων βοηθάει στη εγκυρότητα της διαδικασίας.

Συνοψίζοντας, οι συγγραφείς τονίζουν την «κατά περίπτωση» εφαρμογή του καθαρισμού μίας βάσης και επισημαίνουν ότι αυτό πρέπει να γίνεται σε συνδυασμό με:

1. τις υπολογιστικές τεχνικές που είναι ικανές να ανιχνεύσουν λάθη,
2. την κρίση και αξιολόγηση του ερευνητή σε όσα σημεία δεν μπορούν να ελεγχθούν αλγοριθμικά,
3. την εγκυρότητα ή όχι της κάθε εγγραφής της βάσης με ένα σύστημα το οποίο να βοηθάει τον ερευνητή να συνεκτιμήσει όλους τους σχετικούς παράγοντες.

Τα ανωτέρω υποστηριζόμενα από τους Vijayshankar R. et al βρίσκονται σε πλήρη συμφωνία με τα λογικά βήματα που ακολουθήθηκαν κατά τον καθαρισμό του Αρχείου δεδομένων που χρησιμοποιείται στην παρούσα διατριβή.

Άλλη ερευνητική εργασία [Bonchi F. et al, 1999] σχολιάζεται στη συνέχεια σχετική με χρήση τεχνικών όρυξης δεδομένων και χρήση δένδρων σε διάφορες περιπτώσεις. Αυτή παρόλο που δεν διερευνά τα ναυτικά ατυχήματα ανήκει τεχνικά στην ίδια κατηγορία γιατί αντιμετωπίζει ένα παρόμοιο πρόβλημα.

Πιο συγκεκριμένα η αναζήτηση περιπτώσεων όπου πρέπει να αναλυθούν δεδομένα με σκοπό να ελέγξει αν ένας συγκεκριμένος οργανισμός, μία ασφαλιστική εταιρεία στην περίπτωση αυτή, έχει προβεί σε απάτη, δηλαδή έχει συμβεί κάποιο

συγκεκριμένο και «ακραίο» γεγονός. Είναι επιθυμητό να μπορείς να ανιχνεύσεις το γεγονός εκ των προτέρων (a priori) και να εφαρμόσεις στρατηγικές και κανόνες ώστε εκ των υστέρων (a posteriori) να μπορείς να γνωρίζεις περιοχές του δείγματος με υψηλές συγκεντρώσεις / εμφανίσεις του γεγονότος.

Στη παρούσα διατριβή σε σχέση με τα Ναυτικά ατυχήματα μελετάται, τηρουμένων των αναλογιών κάτι αντίστοιχο, όπου η ολική απώλεια ενός πλοίου μπορεί να θεωρηθεί σαν ένα «ακραίο» γεγονός. Για τον σκοπό αυτό αναλύονται ιστορικά δεδομένα για να μπορούν να σχεδιαστούν στρατηγικές ώστε να προβλέπουν με τον καλύτερο δυνατό τρόπο την μελλοντική εμφάνιση του γεγονότος. Στη συγκεκριμένη ερευνητική εργασία επρόκειτο για το γεγονός της απάτης, ενώ στην παρούσα διατριβή πρόκειται για την ύπαρξη ολικής απώλειας.

Για την υλοποίηση του συγκεκριμένου έργου οι συγγραφείς της εργασίας ακολούθησαν τα εξής σημαντικά για αυτούς βήματα:

- Σχεδιασμός του ζητούμενου της εργασίας.

Με βάση ιστορικά δεδομένα ζήτησαν να βρουν κανόνες και σχέσεις οι οποίες μπορεί να τους βοηθήσουν για τις μελλοντικές ενέργειες.

- Επιλογή της μεθοδολογίας.

Προτιμήθηκε η αναζήτηση γνώσεων σε βάσεις δεδομένων. Κατά τους συγγραφείς αυτή η μεθοδολογία επιτρέπει την κατασκευή και εύρεση μεθόδων που μπορούν να χρησιμοποιηθούν σε παρόμοιες περιπτώσεις. Για την δική τους περίπτωση χρησιμοποιήθηκαν δέντρα αποφάσεων παραγόμενα με την αλγόριθμο C5.0 για να εφαρμόσουν επαρκείς στρατηγικές στη πρόβλεψη του υπό διερεύνηση γεγονότος, αυτό της οικονομικής απάτης.

- Επιλογή των κατάλληλων ιδιοτήτων του αλγορίθμου για το συγκεκριμένο πρόβλημα.
- Κατασκευή της επιθυμητής βάσης δεδομένων.
- Διαδικασία αναζήτησης περισσότερη πληρέστερη και για να το επιτύχουν ενώσανε ορισμένα εξωτερικά στοιχεία στη βάση (όπως τηλέφωνα κ.α.).
- Δημιουργία το προς διερεύνηση μεγέθους και στη συνέχεια υπολογισμός του κόστους σφάλματος ανίχνευσης της απάτης.
- Εκκαθάριση από το αρχείο δεδομένων ορισμένων εγγραφών, στο στάδιο του καθαρισμού των δεδομένων.



- Επιλογή του συνόλου των πεδίων / μεταβλητών που θα συμμετάσχουν στην ανάλυση καθώς και οι ιδιότητές τους.

- Κατασκευή ορισμένων εναλλακτικών μοντέλων με τον C5.0

Το πρώτο δεν περιείχε καμία υπόθεση / συνθήκη. Το δεύτερο είχε μεταβλητά κόστη σφάλματος ταξινόμησης στο “false negative” (πίνακας 2.4) δηλαδή στις περιπτώσεις που, ενώ στην πραγματικότητα δεν θα είναι, θα ταξινομούνται ως «συμβάν».

- Επιλογή για κάθε ένα μοντέλο των παραμέτρων του.

- Εκτίμηση των αποτελεσμάτων με βάση την «ωφέλεια» και την «συνάφεια».

Συνοψίζοντας, οι συγγραφείς με την κατάλληλη χρήση του παραγωγικού συλλογισμού όπως αυτός μπορεί να υλοποιηθεί με ερωτήματα αναζήτησης (queries) και του επαγωγικού όπως αυτός υλοποιείται μέσω των δέντρων αποφάσεων προσεγγίζουν το πρόβλημα της ταξινόμησης κάθε μίας συγκεκριμένης εγγραφής.

Τονίζουν ότι η ανεύρεση γνώσης σε βάσεις δεδομένων αντιμετωπίζει, γεγονός που αντιμετωπίστηκε και στην παρούσα διατριβή, δύο σημαντικά ζητήματα :

1. Η μεθοδολογία είναι κεντρικό τμήμα και αποτελεί μια ανάγκη να χρησιμοποιείται η πρόβλεψη, ακόμα και αν πρέπει να προσαρμοστεί, στην αντίστοιχη ή ανάλογη περίπτωση.
2. Αποτελεί σημαντικό ρόλο να ανιχνεύονται τα συστατικά στοιχεία σε όλα τα στάδια της ερευνητικής διαδικασίας σε κάθε στάδιο της χωριστά.

Ο συγγραφέας του άρθρου που παρουσιάζεται στη συνέχεια [Fawcett Tom, 2001] εξηγεί πως από ένα δέντρο αποφάσεων οδηγούμαστε σε ένα σύνολο κανόνων οι οποίοι είναι αυτόνομοι, εύκολοι στην κατασκευή και «έξυπνοι». Όπως ανέφερε ο συγγραφέας οι κανόνες έχουν την ιδιότητα να «βελτιστοποιούν» την ταξινομητική ικανότητα ενός συνόλου δεδομένων κατασκευάζοντας μικρούς, λακωνικούς ταξινομητικούς κανόνες οι οποίοι επιτυγχάνουν ισχυρή ακρίβεια πρόβλεψης.

Στη συνέχεια με εμπειρικό τρόπο έδειξε με ποια διαδικασία η χρήση των κανόνων μπορεί να βοηθήσει στη μεγιστοποίηση της απόδοσης (scoring performance) ενός κανόνα ταξινόμησης ή πρόβλεψης και παρουσιάζει ορισμένα προβλήματα για τις περιπτώσεις όπου το κόστος σφάλματος ταξινόμησης δεν είναι σταθερό.

Αλλά τόνισε και την προβληματική της εφαρμογής των κανόνων όταν γίνεται

προσπάθεια ταξινόμησης με διαφορετικά κόστη σφάλματος ταξινόμησης (πίνακας διασταύρωσης 2.4).

ΠΙΝΑΚΑΣ 2.4

Πίνακας διασταύρωσης πραγματικής και εκτιμηθείσας ταξινόμησης

	Εκτιμηθείσα ταξινόμηση	
	p	n
Πραγματική κατάσταση p	True Positive	False Negative
n	False Positive	True Negative

Συνολικά στην εργασία αυτή συνόψισε τη θεωρία και τα πλεονεκτήματα από τη χρήση των κανόνων ταξινόμησης, παρουσίασε ορισμένα πειραματικά προβλήματα που παρατηρούνται σε ειδικές περιπτώσεις και πρότεινε τρόπους αντιμετώπισής των.

Το αντικείμενο αυτό της εργασίας παρουσιάζεται εδώ ως απαραίτητο εφόδιο για τα πιθανά μετέπειτα βήματα στην παρούσα διατριβή.

#### 2.2.4. Ερευνητικές εργασίες για τα ναυτικά ατυχήματα

Ο σκοπός της ερευνητικής εργασίας που ακολουθώς σχολιάζεται [Psraftis Ch. et al, 1998] ήταν να παρουσιάσει μία ανάλυση του ανθρώπινου παράγοντα σαν μία παράμετρο των ναυτικών ατυχημάτων. Αυτή η ανάλυση διερεύνησε τυχόν υπάρχουσες σχέσεις ανάμεσα στις διάφορες πιθανές αιτίες πρόκλησης ατυχήματος και στο τελικό αποτέλεσμα αυτού του ατυχήματος.

Δημιούργησαν καταρχήν μια βάση δεδομένων και όρισαν τους τύπους των πλοίων που θα μελετήσουν καθώς και πιο περιστατικό θεωρείται «Σοβαρό». Σαν «Σοβαρό» περιστατικό ορίστηκε κάθε περιστατικό που συμπεριλαμβάνει απόλεια πλοίου, μεγάλη ζημία στο πλοίο και / ή στο φορτίο του, συμπεριλαμβανομένης και της απόλειας ανθρώπινης ζωής.

Πηγή των στοιχείων που χρησιμοποιήθηκε για την ανάλυση ήταν τα στοιχεία ναυτικών ατυχημάτων που συλλέχθηκαν από το Υ.Ε.Ν. Καταγράφηκαν 1355 ναυτικά ατυχήματα για πλοία με ελληνική σημαία για το χρονικό διάστημα 1978 - 1984. Εξαιρέθηκαν τα ατυχήματα που αφορούσαν γενικά αλιευτικά σκάφη καθώς επίσης και πλοία μικρότερης χωρητικότητας από 1000 τόρους και με αυτό τον τρόπο

δημιουργήθηκε μία ξεχωριστή βάση δεδομένων που περιελάμβανε αποκλειστικά «Σοβαρό» περιστατικά.

Αποτέλεσμα ήταν ότι εντοπίστηκαν συνολικά 94 περιπτώσεις ατυχημάτων και τελικά απέμεινε ένας αριθμός 75 σοβαρών ατυχηματικών περιστατικών για μελέτη.

Για κάθε μία από τις 75 εγγραφές ατυχημάτων ξεχωριστά και, για κάθε μία από αυτές, κατέγραψε ο καθένας από τους ερευνητές τα ακόλουθα:

- την υποκειμενική του άποψη για τις πιθανές αιτίες του ναυτικού ατυχήματος
- την υποκειμενική του κρίση για ποια πιθανά μέτρα θα μπορούσαν να είχαν ελαττώσει τον κίνδυνο ή αποτρέψει εντελώς το ατύχημα.

Παρατηρήθηκαν από την ανάλυση τα εξής συμπεράσματα :

- Το σύστημα της κωδικοποίησης των αιτιών των ναυτικών ατυχημάτων δημιούργησε πρόβλημα.
- Η διαπίστωση από όλους της ύπαρξης πολυπλοκότητας των αιτιών των ναυτικών ατυχημάτων.
- Η διαφορά γνώμης και άποψης ανάμεσα στους συγγραφείς αυτής της εργασίας που μελέτησαν τα ιστορικά των ατυχημάτων.

Προκύπτει και από την παρούσα διατριβή ότι το πρόβλημα της διερεύνησης των αιτιών των Ναυτικών Ατυχημάτων είναι πολυπαραμετρικό, πράγμα που υποστηρίζεται και στην προαναφερθείσα εργασία των Ψαραυτή Χ.Ν. κ.α.

Οι πέντε πλέον σημαντικές αιτίες ατυχημάτων ήταν, ταξινομημένες κατά φθίνουσα συχνότητα, οι διαδικασίες για έλεγχο ασφαλείας υπάρχουσες και γνωστές που δεν ακολουθήθηκαν σε ποσοστό 8,2 %, η ανεπαρκής πραγματική ικανότητα-αντίδραση σε ποσοστό 7,9 %, ο πολύ άσχημος καιρός και οι φυσικές καταστροφές σε ποσοστό 6,6 %, η ανεπαρκής παρατήρηση θέσης μη σχεδιασμένη στους χάρτες σε ποσοστό 5,0 % και η κακή κρίση των κινήσεων του σκάφους (ρεύματα, άνεμος κ.α) σε ποσοστό 4,0 %.

Έγινε ομαδοποίηση ανάμεσα στις διάφορες γενικότερες κατηγορίες των αιτιών και οι πλέον σημαντικές κατά τους συγγραφείς ήταν:

Α. Καταστάσεις που δεν συσχετίζονται με το πλοίο. Με ποσοστό 49,5 % πολύ άσχημος καιρός, φυσικές καταστροφές κ.α., με ποσοστό 16,2 % λειτουργικό σφάλμα άλλου πλοίου, όπως λανθασμένες μανούβρες κ.α. και τέλος με ποσοστό 7,1 %



ρεύματα, άνεμος κτλ που οδηγούν σε ισχυρή ολίσθηση ή σε άλλες δυσκολίες κινήσεων. Συνολικά για την κατηγορία Α ποσοστό 72,8 % των περιπτώσεων.

Β. Κατασκευή του πλοίου και θέση των μηχανημάτων σε αυτό. Με ποσοστό 49,1 % μη επαρκής δομική αντοχή του πλοίου και με ποσοστό 30,9 % εξασθένηση της δομικής αντοχής του πλοίου από μεταγενέστερες συγκολλήσεις σε αυτό, διάβρωση κ.α. Συνολικά για την κατηγορίας Β ποσοστό 80 % των περιπτώσεων.

Γ. Τεχνικές συνθήκες που αφορούν μηχανήματα πάνω στο πλοίο. Με ποσοστό 34 % τεχνικό πρόβλημα του εξοπλισμού-μηχανημάτων.

Δ. Καταστάσεις που έχουν να κάνουν με την χρήση και τον σχεδιασμό των μηχανημάτων. Δεν υπάρχει καμία στατιστικά σημαντική αιτία.

Ε. Φορτίο, ασφάλιση και συμπεριφορά φορτίου και καυσίμων. Με ποσοστό 50 % αυτανάφλεξη φορτίου η καυσίμων.

Ζ. Επικοινωνίες, οργάνωση, διαδικασίες και ρουτίνες. Με ποσοστό 31,4 % διαδικασίες που δεν ακολουθήθηκαν για έλεγχο ασφαλείας αλλά γνωστές και με ποσοστό 11,3 % αποτυχία των ρουτινών επιθεώρησης και συντήρησης πάνω στο πλοίο.

Η. Μεμονωμένο πρόσωπο πάνω στο πλοίο, καταστάσεις, κρίση, αντιδράσεις. Με ποσοστό 22 % ανεπαρκής πραγματική ικανότητα εξάσκησης κ.α., με ποσοστό 13,8 % επαρκής παρατήρηση και εξακρίβωση της θέσης, μη σχεδιασμένη στους ναυτικούς χάρτες και με ποσοστό 11,2 % όχι καλή κρίση των κινήσεων του σκάφους ρεύματα, άνεμος κ.α.

Παρουσιάστηκε πως υπήρχε ισχυρή συγκέντρωση στατιστικά σημαντικών αιτιών στις γενικότερες κατηγορίες Ζ και Η.

Με μεθόδους Στατιστικής ελέγχθηκε η ύπαρξη εξάρτησης και συσχέτισης ανάμεσα σε αυτές τις επτά κατηγορίες αιτιών, και παρατηρήθηκε ότι η κατηγορία Α εξαρτάται από τις κατηγορίες Β και Η, η κατηγορία Β εξαρτάται από τις κατηγορίες Α και Η, η κατηγορία Γ εξαρτάται από την κατηγορία Ε, η κατηγορία Ζ εξαρτάται από τις κατηγορίες Γ, Η και η κατηγορία Η εξαρτάται από τις κατηγορίες Α, Β και Ζ.

Στα συμπεράσματα για μείωση του ρίσκου των ατυχημάτων, τονίσθηκε ότι θα πρέπει να δοθεί ιδιαίτερη προσοχή κατά τους συγγραφείς του άρθρου στα εξής:

- Ανθρώπινο παράγοντα.
- Εκπαίδευση και εξάσκηση,
- Πρακτικές και διαδικασίες για την ασφάλεια και τον έλεγχο,

- Εξωτερικούς παράγοντες,
- Τεχνικούς παράγοντες.

Τελειώνοντας στα συμπεράσματα σημειώνεται η συσχέτιση της κατηγορίας Γ με την κατηγορία Ζ, όπου δεδομένης της μεγαλύτερης σημαντικότητας της κατηγορίας Ζ σαν παράγοντα αιτίας, ισχύει πως μειώνοντας το ρίσκο ναυτικών ατυχημάτων αλλά με την «διόρθωση» της κατηγορίας Γ δεν θα έχουμε κανένα αποτέλεσμα αν δεν διορθωθεί εξίσου και η κατηγορία Ζ και κατ' επέκταση η κατηγορία Η.

Για την εργασία που σχολιάστηκε μπορεί να διατυπωθεί η παρατήρηση ότι ο πληθυσμός του δείγματος που αναλύθηκε δεν ήταν ικανοποιητικά μεγάλος και ότι η συγκέντρωση στατιστικά σημαντικών αιτιών στις γενικότερες κατηγορίες Ζ και Η ήταν αναμενόμενη αφού αυτές οι κατηγορίες βασικά περιέχουν αιτίες σχετικές με τον ανθρώπινο παράγοντα.

Σε άλλη ερευνητική εργασία [Karydis P. et al, 1998] οι συγγραφείς κάνουν ανάλυση των ναυτικών ατυχημάτων προτείνοντας μια άλλη προσέγγιση με την Μέθοδο Δικτύων Γεγονότων (Event Tree Analysis). Αναφέρεται ώστε να υπάρχει σφαιρική άποψη για τις μεθόδους ανάλυσης τις σχετικές με τα ναυτικά ατυχήματα.

Στην εργασία τους παραπέμπουν, στην αρχή, στην Αξιολόγησης Τυπικής Ασφάλειας (F. S. A.), που με συστηματική διαδικασία μπορεί να εκτιμήσει την ριψοκινδυνότητα που αφορά τις ναυτικές δραστηριότητες καθώς επίσης και τα κόστη και οφέλη των εναλλακτικών λύσεων που προτείνονται για τον περιορισμό των κινδύνων.

Η διαδικασία αυτή θα μπορούσε για παράδειγμα να εφαρμοσθεί, αφού επιτρέπει την εκτίμηση πιθανών κινδύνων πριν προκύψει κάποιο σοβαρό ατύχημα, στον προσδιορισμό της διάταξης διαύλων έτσι ώστε να περιορισθεί η πιθανότητα συγκρούσεως σε προκαθορισμένα σημεία κάποιας γεωγραφικής περιοχής της υδρογείου, όπως για παράδειγμα στα στενά του Γιβραλτάρ.

Αναφέρθηκε στη συνέχεια ότι αυτή συνίσταται στην εφαρμογή μίας σειράς βημάτων τα οποία είναι:

- Προσδιορισμός κινδύνων (hazard identification),
- Αξιολόγηση ριψοκινδυνότητας (risk assessment),
- Επιλογή μέτρων περιορισμού ριψοκινδυνότητας (risk control options),

- Εκτίμηση κόστους-οφέλους (cost-benefit assessment),
- Προτάσεις λήψης αποφάσεων (recommendations for decision making).

Για κάθε ένα από τα βήματα μπορούν να εφαρμοσθούν διάφορες μέθοδοι, οι οποίες έχουν ήδη κριθεί ως αξιόπιστες από τον Ι.Μ.Ο.

Στη συνέχεια επικέντρωσαν το ενδιαφέρον τους οι συγγραφείς στον Προσδιορισμό κινδύνων και αναφέρθηκαν στις παρακάτω μεθόδους, αποδεκτές από τον Ι.Μ.Ο στο στάδιο αυτό. Την Μέθοδο Δικτύου Σφαλμάτων (Fault Tree Analysis), την Μέθοδο Δικτύων Γεγονότων (Event Tree Analysis), την Ανάλυση Τρόπου Αστοχίας και Αποτελεσμάτων (Failure Mode and Effects Analysis), και τις Μελέτες κινδύνων και λειτουργικότητας (Hazard and Operability studies).

Ακολούθως εστίασαν το ενδιαφέρον τους στην εφαρμογή της Μεθόδου Δικτύων Γεγονότων (E.T.A.) η οποία συνίσταται στη σύνταξη των παρακάτω:

A. Περιγραφή ατυχήματος και μάλιστα αντικειμενική παράθεση όλων των λεπτομερειών.

B. Δημιουργία Δικτύου Γεγονότων (event tree): Για να δημιουργηθεί το Δίκτυο Γεγονότων πρέπει να καταγραφούν με χρονική σειρά όλα τα γεγονότα που οδήγησαν στο ατύχημα.

Γ. Ανάλυση των αιτιών σε βασικές κατηγορίες όπως, λάθη στην περάτωση εργασιών, εσωτερικές δυσλειτουργίες, βασικά αίτια και μέτρα πρόληψης.

Δ. Συνοπτική καταγραφή των γεγονότων και συνθηκών που επικρατούσαν.

Κατά την E.T.A. το ατύχημα αξιολογείται ως επακόλουθο των ακόλουθων χρονικών φάσεων (σταδίων):

1. Η λανθάνουσα φάση ορίσθηκε ως η περίοδος η οποία είναι δυνατό να σχετίζεται με τυχαίες καταστάσεις ή ενέργειες που θέτουν το πλοίο σε κίνδυνο.
2. Σαν εναρκτήρια φάση ορίσθηκε η φάση που σηματοδοτείται από το τεχνικό ή ανθρώπινο σφάλμα, το οποίο οδηγεί στην εξέλιξη του ατυχήματος.



3. Σαν κλιμάκωση ορίσθηκε η επιδείνωση της κατάστασης με διαδοχικά σφάλματα ή αστοχίες και η αδυναμία καταπολέμησης των αρχικών προβλημάτων.
4. Σαν κρίσιμη φάση ορίσθηκε η τελευταία ευκαιρία για να αποφευχθεί το ατύχημα.

Ο διαχωρισμός των γεγονότων στις τέσσερες φάσεις είναι υποκειμενικός. Αλλά υποστηρίζουν οι συγγραφείς ότι αυτό που έχει σημασία είναι να γίνει η καταγραφή με σωστή χρονική σειρά όλων των γεγονότων, ώστε να είναι δυνατή η σωστή ανάλυση και η εξαγωγή των συμπερασμάτων.

Τέλος εφάρμοσαν την Ε.Τ.Α. σε ένα συγκεκριμένο ατύχημα που παρουσιάζει ενδιαφέρον, για την Πειραιϊκή ναυτική κοινότητα. Το συγκεκριμένο ατύχημα αφορά πυρκαϊά-έκρηξη σε δεξαμενόπλοιο κατά την διάρκεια επισκευής του στην επισκευαστική βάση Περάματος. Αφού περιέγραψαν με λεπτομέρεια το ιστορικό του γεγονότος οι συγγραφείς διατύπωσαν τα συμπεράσματά τους χωριστά για κάθε μία από τις τέσσερες φάσεις της μεθόδου για το συγκεκριμένο ατύχημα.

Σε πρόγραμμα του Bureau of Transport and Communications Economics, Australia, έγινε ανάλυση Επικινδυνότητας σε ένα τύπο πλοίων τα large Bulk Carriers.

Στην ανάλυση που έκαναν οι ερευνητές Gentle N, Wheatstone S, Salmi J, και Breusch T. [Bureau of Transport and Communications Economics Australia, 1994] δείχθηκε ότι με έλεγχο μόλις του 34% των Bulk Carriers που επισκέπτονται τα λιμάνια της Αυστραλίας, μπορούν να προλάβουν το 93% των πλοίων που θα έχουν ατυχήματα. Σχολιάζεται σαν σημείο αναφοράς για τα οφέλη που προκύπτουν, τουλάχιστον για τους αρμόδιους φορείς, από την ανάλυση των ναυτικών ατυχημάτων. Η παρούσα διατριβή θα διερευνήσει την δυνατότητα πρόβλεψης έχοντας επεκτείνει την μελέτη για όλες τις γεωγραφικές περιοχές και για όλους τους τύπους πλοίων.

Το SAFECO ήταν ένα ερευνητικό πρόγραμμα οποίο χρηματοδοτήθηκε από την Γενική Διεύθυνση Μεταφορών (DGVII) της Ευρωπαϊκής Ένωσης. Στην ανάλυση διερευνήθηκαν τυχόν υπάρχουσες σχέσεις ανάμεσα στις διάφορες πιθανές αιτίες πρόκλησης ατυχήματος και στο τελικό αποτέλεσμα αυτό καθαυτό.

Σκοπός του SAFECO I (1996-1998) ήταν να αυξήσει την ασφάλεια της Ναυσιπλοΐας στα Ευρωπαϊκά παράκτια ύδατα αναλύοντας τους παράγοντες που πιθανά συνεισφέρουν στη δημιουργία ατυχημάτων. Σχολιάζεται συνοπτικά και

αναφέρεται μόνο για να τονισθεί η έμφαση που έχει δοθεί από τους Διεθνείς Οργανισμούς στην διερεύνηση παραγόντων που επιδρούν στα Ναυτικά ατυχήματα.

Επιχείρησαν να διαμορφώσουν μοντέλο που να δίνει συνολική θεώρηση του ρίσκου επικινδυνότητας. Η μεθοδολογία, αλλά και τα αποτελέσματα από την εφαρμογή του μοντέλου ελπίζουν οι ερευνητές ότι μπορούν να βοηθήσουν ολόκληρη την ναυτιλιακή κοινότητα να κατασταλάξει στο ποιοι είναι οι σημαντικοί παράγοντες που επηρεάζουν την θαλάσσια ασφάλεια και σαν αποτέλεσμα αυτού να αναπτύξουν και να υιοθετήσουν την πλέον κατάλληλη πολιτική και τα πλέον αρμόζοντα μέτρα για την ενίσχυσή της.

Το SAFECO II (1998-1999), διάρκειας 22 μηνών, ήταν συνέχεια του SAFECO I. Σκοπός του ήταν να παρουσιάσει νέες και βελτιωμένες τεχνολογικά λύσεις, όπως για εσωτερική και εξωτερική επικοινωνία στο πλοίο. Στόχος του ήταν να αποδειχθεί η ορθότητα των εφαρμογών των στατιστικών αναλύσεων, ώστε να επιτευχθούν οικονομικά οφέλη και βελτιώσεις της ασφάλειας στην Ναυτιλία.

Απομένει, με την υλοποίηση των συμπερασμάτων και την εφαρμογή των προταθέντων λύσεων των συγκεκριμένων προγραμμάτων, να διαπιστωθεί αν επιτεύχθηκε μείωση του αριθμού των ατυχημάτων και άρα αύξηση της ασφάλειας στα παράκτια ύδατα της Ευρωπαϊκής Ένωσης.

Οι συγγραφείς [Psarrafis Ch., Panagakos G. et al, 1998] στην ερευνητική τους εργασία έκαναν μια ευρεία και αθροιστική ανάλυση ενός μεγάλου πλήθους ατυχημάτων, έτσι ώστε να εντοπιστεί ποιοι παράγοντες επηρεάζουν τον κίνδυνο.

Δεν έγινε καμία προσπάθεια να γίνει μία σε βάθος ανάλυση, σε ότι αφορά στις λεπτομέρειες ανεξάρτητων ατυχημάτων. Τέτοιου είδους πληροφόρηση είναι δύσκολη και σπάνια διαθέσιμη στις βάσεις δεδομένων. Μια τέτοια εργασία, υποστηρίζουν οι συγγραφείς, είναι δυνατή εάν υπάρχει και λεπτομερής καταγραφή και για μεγάλο πλήθος περιπτώσεων.

Μία από τις ερωτήσεις κλειδιά που τελικά διατυπώθηκαν ήταν το κατά πόσο μπορούσαν να εντοπιστούν παράμετροι όπως: το μέγεθος του πλοίου, ο τύπος του, η ηλικία του, ο καιρός, η γεωγραφική περιοχή του ατυχήματος κ.α. που να μπορούν να κάνουν την στατιστικά σημαντική διαφορά σε ότι αφορά στο ρίσκο.

Αναγνωρίστηκε τέλος σαν αιτία ατυχημάτων η σπουδαιότητα του ανθρώπινου παράγοντα και του ρόλου του τύπου του πλοίου.

Γενικά, πιστεύουν οι συγγραφείς ότι μία ανάλυση στατιστικής σημαντικότητας δεν θα αποδείξει μία σχέση αιτίας και αποτελέσματος, αλλά θα αποκαλύψει το κατά πόσο οι μεταβολές στους ρυθμούς των ατυχημάτων είναι συστηματικές ή οφείλονται αποκλειστικά στην τύχη.

Στην παρούσα διατριβή, όπως αναφέρθηκε και στο προηγούμενο κεφάλαιο, γίνεται προσπάθεια να απαντηθούν με χρήση τεχνικών Όρυξης δεδομένων και ερωτήματα που διατυπώθηκαν στην εργασία των Ψαραύτης Χ. Παναγάκος Γ. κ.α..

Πανεπιστήμιο Πειραιώς



## ΚΕΦΑΛΑΙΟ 3

### 3.1. Εισαγωγή

Η επιλογή μεθοδολογίας για την υλοποίηση της παρούσας διατριβής εξαρτάται από δύο κυρίαρχα σκέλη:

- Τον τρόπο κατασκευής του αρχείου δεδομένων.
- Τον τρόπο ανάλυσης του αρχείου δεδομένων.

Τα δύο αυτά σκέλη αλληλοσυμπληρώνονται αφού πάνω σε ένα αρχείο πρέπει να εκτελεστούν συγκεκριμένοι αλγόριθμοι για να παραχθούν ορισμένα αποτελέσματα ώστε να διατυπωθούν συμπεράσματα που να απαντούν στις ερευνητικές υποθέσεις. Οι αλγόριθμοι απαιτούν τα δεδομένα που θα χρησιμοποιηθούν να είναι συγκεκριμένης μορφής. Αντίστροφα, στο σχεδιασμό του αρχείου δεδομένων θα πρέπει να ληφθούν υπόψη οι συγκεκριμένοι αλγόριθμοι που θα εκτελεστούν.

Παρ' ότι η τεχνολογία των βάσεων δεδομένων μας παρέχει βασικά εργαλεία για την αποδοτική αποθήκευση και ανάκτηση κάθε είδους δεδομένων, ωστόσο το θέμα της εύκολης κατανόησης και ανάλυσης μεγάλων συνόλων δεδομένων παραμένει ένα από τα βασικότερα προβλήματα. Προκειμένου λοιπόν να αντιμετωπιστεί το πρόβλημα εφαρμογής αναλύσεων σε μεγάλα σύνολα, απαιτούνται νέες τεχνικές με νέες γενιές μεθόδων και εργαλείων. Η ανάγκη εύκολης κατανόησης και ανάλυσης σε μεγάλα συστήματα βάσεων δεδομένων οδήγησε σε μία σχετικά νέα ερευνητική περιοχή που αφορά την διαδικασία ανακάλυψης γνώσης και εξαγωγής χρήσιμης πληροφορίας από αυτά. Στην ερευνητική αυτή περιοχή εντάσσεται και η Όρυξη δεδομένων (Data mining). Γενικά, η Όρυξη δεδομένων είναι η εξερεύνηση και η ανάλυση με αυτοματοποιημένες ή ημι-αυτοματοποιημένες μεθόδους μεγάλου όγκου δεδομένων, προκειμένου να ανακαλυφθούν σημαντικά πρότυπα (patterns) και κανόνες. Η Όρυξη δεδομένων σχετίζεται και με άλλους συναφείς χώρους όπως η ανάλυση δεδομένων, η στατιστική, οι βάσεις δεδομένων. Οι ανωτέρω τομείς έχουν πολύχρονη αυτόνομη πορεία στην επιστήμη αλλά μόλις τα τελευταία χρόνια συνδυάστηκαν για την δημιουργία του νέου κλάδου αυτού της Όρυξης δεδομένων.

Τα τελευταία χρόνια έχουν αναπτυχθεί και είναι διαθέσιμος ένας μεγάλος αριθμός μεθόδων Όρυξης δεδομένων οι βασικότερες από τις οποίες είναι τα Δένδρα Αποφάσεων (Decision Trees) τα Νευρωνικά Δίκτυα (Neural Networks) και τα

Σύνολα Κανόνων (Rule Sets), [Quinlan J.R., 1986] και [Hand D. et al, 2001]. Προκειμένου όμως οι μέθοδοι Όρυξης δεδομένων να επιτύχουν την εξαγωγή και περιγραφή της γνώσης από σύνολα δεδομένων, εκτελούν κάποιες εργασίες (tasks). Οι σημαντικότερες από αυτές είναι: Ομαδοποίηση (Clustering), Ταξινόμηση (Classification), Εκτίμηση και Πρόβλεψη (Estimation & Prediction), Παλινδρόμηση (Regression) [Hand D. et al. 2001] και [Quinlan J.R., 1986.].

Η αναζήτηση συναφειών μεταξύ των πεδίων / μεταβλητών του αρχείου δεδομένων είναι ένας από τους δυσκολότερους τομείς της ερευνητικής εργασίας γιατί υποχρεωτικά αυξάνει την πολυπλοκότητα της. Έτσι η ανίχνευση παραγόντων που συσχετίζονται στην περίπτωση μας με την ολική απώλεια του πλοίου και η αναζήτηση σχέσεων που οδηγούν σε αυτήν μπορεί να είναι «κρυμμένες» σε ένα λογικό κανόνα με συγκεκριμένο αριθμό παραγόντων. Μόνο με τη χρήση των τεχνικών Όρυξης δεδομένων μπορεί να ανιχνευτεί η ύπαρξη τέτοιων σχέσεων. Η αναζήτηση των οποίων θα γίνει με εφαρμογή στα δεδομένα του αλγορίθμου C5.0 ο οποίος παράγει δένδρα αποφάσεων. Θα παραχθούν επίσης ταξινομητικοί κανόνες οι οποίοι προσφέρουν έναν εναλλακτικό τρόπο να διαβαστεί ένα δέντρο αποφάσεων καθώς και θα εφαρμοσθεί μια ειδική περίπτωση του C5.0, λόγω της μορφής των δεδομένων, ο αλγόριθμος G.R.I. που παράγει μικρά σε βάθος δέντρα.

### 3.2. Περιγραφή της μορφής του αρχείου δεδομένων

Επειδή ένας από τους στόχους της διατριβής είναι η μετατροπή των ακατέργαστων δεδομένων σε πληροφορία, και μάλιστα σε ένα ανοιχτό σύστημα συλλογής πληροφοριών, αποφασίστηκε να κατασκευαστεί η απλούστερη μορφή αρχείου δεδομένων. Αυτό έχει τις εξής ιδιότητες :

- έχει μορφή πίνακα με τις στήλες του πίνακα να είναι τα πεδία / μεταβλητές του αρχείου δεδομένων, τα μεγέθη δηλαδή πάνω στα οποία μετρούνται οι εγγραφές, οι οποίες αποτελούν τις γραμμές του πίνακα.
- έχει αναλυτική μορφή ώστε μέσο διαφορετικών μετασχηματισμών να μπορεί να μετατρέπει τα δεδομένα σε πληροφορία.
- μπορεί στη συνέχεια να ενταχθεί σε μία άλλη πολυδιάστατη βάση [Smyth, P. & Goodman, R., 1992].

### 3.3. Επιλογή μεθοδολογίας ανάλυσης δεδομένων

Η επιλογή συγκεκριμένης μεθοδολογίας για την επεξεργασία των δεδομένων πρέπει να συνάδει με τη φύση τους, καθώς και τη μορφή της ερευνητικής υπόθεσης.

Στην επεξεργασία των δεδομένων μιας βάσης και ενός αρχείου δεδομένων υπάρχουν τρεις κυρίαρχες τάσεις :

- Η ανάλυση των δεδομένων με πινακοποιήσεις και O.L.A.P.
- Η ανάλυση των δεδομένων με στατιστικές αναλύσεις.
- Η ανάλυση των δεδομένων με μεθοδολογίες της Όρυξης δεδομένων

Κάθε μία από τις παραπάνω τεχνικές, οι οποίες αλληλοσυμπληρώνονται, έχει τα πλεονεκτήματά της και τα μειονεκτήματά της. Ακόμα σημαντικότερο είναι ότι ουσιαστικά κάθε μία αντιμετωπίζει ένα διαφορετικό σύνολο προβλημάτων και υποθέσεων.

Για την επίτευξη του σκοπού της παρούσας διατριβής η περισσότερο κατάλληλη μεθοδολογία ανάλυσης είναι η Όρυξη δεδομένων [Arning A. et al, 2000] και [Holsheimer M. et al, 1996] για τους λόγους που αναφέρονται στην συνέχεια.

Η Όρυξη δεδομένων αποτελεί επιστημονικό πεδίο σχετιζόμενο με τη στατιστική, τη μηχανική μάθηση (machine learning), τη διαχείριση βάσεων δεδομένων, την αναγνώριση προτύπων (pattern recognition), την τεχνητή νοημοσύνη (artificial intelligence), και άλλες επιστημονικές περιοχές. Στόχος αυτής της σχετικά νέας τεχνολογίας είναι η απόκτηση γνώσης για το πεδίο στο οποίο εφαρμόζεται, η στήριξη των αποφάσεων μέσω της εξαγωγής συμπερασμάτων και η επίτευξη εγκύρων προβλέψεων.

Η Όρυξη δεδομένων στην αρχή είχε σχεδιαστεί ως μία εφαρμογή των εμπειρών συστημάτων (expert systems) έτσι ώστε να επιλύονται συγκεκριμένα προβλήματα / υποθέσεις. Για τον λόγο αυτό δε δίνεται απόλυτη σημασία στην κατανόηση του αλγορίθμου και της τεχνικής αλλά στο αποτέλεσμα. Η αναζήτηση σχέσεων στα δεδομένα γίνεται με οποιονδήποτε αλγόριθμο εξόρυξης μετασχηματίζει τα δεδομένα σε γνώση, ακόμη και με απλή οπτική παρατήρηση (data understanding / data visualization κ.ο.κ)

Από πλήθος ερευνητικών εργασιών, μερικές από τις οποίες αναφέρθηκαν στο προηγούμενο κεφάλαιο, διαπιστώθηκε ότι δεν είναι εύκολο να προσδιοριστούν οι



παράγοντες στους οποίους οφείλονται τα ατυχήματα. Αυτή η διαπίστωση σε συνδυασμό με την ανάγκη πρόληψης και μείωσης αυτών, μας υποχρεώνει να κατασκευάσουμε ένα αρκετά εκτενές αρχείο δεδομένων με πλήθος 38536 εγγραφών και 74 μεταβλητών. Με αυτό τον τρόπο είναι δυνατή η ανίχνευση λεπτομερών και σπάνιων σχέσεων οι οποίες απαιτούν σημαντικό όγκο εγγραφών για να αποκτήσουν στατιστική εγκυρότητα. Το μέγεθος και μόνο του αρχείου υπαγορεύει υπολογιστικά τη χρήση των βαθμωτών (scalable) αλγορίθμων της μεθοδολογίας Όρυξης δεδομένων, σε αντίθεση με τις τεχνικές στατιστικής ανάλυσης των οποίων σημαντικός περιορισμός είναι ότι δε λειτουργούν το ίδιο αξιόπιστα σε μεγάλους όγκους δεδομένων, όπως φαίνεται και από τις εργασίες που αναφέρθηκαν και σχολιάστηκαν στο 2.2.2 υποκεφάλαιο.

Η Όρυξη δεδομένων αναφέρεται κυρίως στην εξερεύνηση των δεδομένων χωρίς προηγουμένως να έχει προσδιοριστεί τι ακριβώς ζητείται, πέραν από την εύρεση των περισσότερων δυνατών σχέσεων χωρίς να έχουν ορισθεί ιδιαίτερες προϋποθέσεις.

Η χρήση στατιστικής ανάλυσης των δεδομένων προϋποθέτει την ύπαρξη κάποιου πειραματικού σχεδιασμού (experimental design) βάσει του οποίου έχουν προέλθει οι παρατηρήσεις για τον έλεγχο συγκεκριμένων υποθέσεων. Για το λόγο αυτό βασίζεται σε μια σειρά υποθέσεων και συνθηκών καθώς και στην αντίστοιχη χρήση τεχνικών που είναι απαραίτητες στη γνωμοδότηση υπέρ ή όχι μιας υπόθεσης.

Στην παρούσα διατριβή, το ενδιαφέρον εντοπίζεται στην εκτίμηση του τελικού αποτελέσματος με μία λογική διαδικασία της μορφής ( AN και EAN.... τότε θα συμβεί, θα παρατηρηθεί το γεγονός ή όχι.)

Στην περίπτωση μας η όλη προσέγγιση είναι η εξής :

Με βάση ένα σύνολο παρατηρήσεων:

- Να εξαχθούν ορισμένοι κανόνες και κάποια συμπεράσματα.
- Από τα οποία θα μπορούμε να κάνουμε υποθέσεις και εικασίες για το πείραμα που «παράγει» αυτά τα δεδομένα.

Η έννοια της Όρυξης δεδομένων είναι απόλυτα συνδεδεμένη με μεγάλες και πολλές φορές ετερογενείς βάσεις δεδομένων ή αρχεία καταχωρημένα από πηγές με διαφορετική μορφή και λογική διάταξη. Προσφέρει η τεχνική αυτή πλεονέκτημα διότι δίνει την δυνατότητα να γίνει διαχείριση των δεδομένων με εφαρμογές της επιστήμης της Πληροφορικής.

Στην περίπτωση μας, με βάση ιστορικά δεδομένα ζητήθηκε να βρεθούν κανόνες και σχέσεις για πρόβλεψη οι οποίες μπορεί να βοηθήσουν σε επιλογές για μελλοντικές ενέργειες. Οι τεχνικές ανακάλυψης κανόνων και προτύπων εφαρμόζονται για τον εντοπισμό συνδυασμών στοιχείων που εμφανίζονται συχνά μαζί σε ένα γεγονός ή στοιχείων που παρουσιάζουν μεγάλη απόκλιση στη συμπεριφορά τους σε σχέση με την πλειονότητα των στοιχείων μιας βάσης δεδομένων.

Η Όρυξη αποτελεί ουσιαστικά μία μη παραμετρική, χωρίς ιδιαίτερες προϋποθέσεις μέθοδο ανάλυσης, απαλλαγμένη από επιλογή δείγματος από το οποίο εξαρτώνται τα αποτελέσματα. Άρα περισσότερο γενική και αξιόπιστη, μέθοδο ανάλυσης δεδομένων η οποία έχει την ικανότητα να βρίσκει «κοινές συμπεριφορές» (pattern) μέσα σε μεγάλους όγκους δεδομένων.

Η προετοιμασία των δεδομένων αφορά την απαλοιφή των σφαλμάτων και την ορθή αποκατάσταση της μορφής των πεδίων / μεταβλητών και των εγγραφών διότι οι βάσεις δεδομένων συχνά περιέχουν κυρίως:

- κενές εγγραφές,
- λανθασμένους τύπους,
- σφάλματα κωδικοποιήσεων
- σφάλματα λογικών συνδέσμων
- πιθανές όμοιες εγγραφές που δεν προσφέρουν επιπλέον πληροφορία
- πολλαπλές εγγραφές που αλληλοσυμπληρώνονται

Τα δεδομένα που συλλέγονται, σε συνοπτική ή αναλυτική μορφή, ελέγχονται με χρήση αλγορίθμων προσαρμοσμένων στις ιδιαιτερότητες κάθε βάσης, όπως θα γίνει στο πέμπτο κεφάλαιο της διατριβής. Ελέγχονται επίσης, ως προς την ποιότητά τους, όπου είναι απαραίτητο διορθώνονται και τυποποιούνται σύμφωνα με την μορφή στην οποία τηρούνται τα υπόλοιπα ιστορικά ή άλλα δεδομένα. Η διόρθωση των δεδομένων περιλαμβάνει την απομάκρυνση λανθασμένων εγγραφών, την τυποποίηση των πεδίων / μεταβλητών, την δημιουργία σχεσιακών δομών και την διαχείριση δεδομένων που λείπουν, τα οποία και θα γίνουν στο πέμπτο κεφάλαιο. Εδώ αξίζει να σημειωθεί ότι εκτιμάται ότι επί της συνολικής μεθοδολογίας όρυξης δεδομένων, μεθοδολογίας που θα αναπτυχθεί στην συνέχεια στο παρόν κεφάλαιο, το μεγαλύτερο μέρος αφορά την προετοιμασία των δεδομένων.

Τέλος, βασικό σημείο είναι ο έλεγχος της εγκυρότητας και αξιοπιστίας των στοιχείων του αρχείου δεδομένων, απαραίτητος ώστε τα συμπεράσματα από την ανάλυση να είναι έγκυρα και αξιόπιστα.

Τα δένδρα απόφασης που θα χρησιμοποιηθούν είναι ένας τρόπος αναπαράστασης μιας σειράς κανόνων που οδηγούν σε μια κατηγορία - κλάση περιπτώσεων η σε μια τιμή. Με άλλα λόγια, τα δένδρα απόφασης αναπαριστούν τον τρόπο με τον οποίο ένα σύνολο περιπτώσεων (πληθυσμός), που αντιστοιχεί σε ένα κόμβο του δένδρου, διασπάται σε δύο ή περισσότερα υποσύνολα, όσο το δυνατόν πιο διαφορετικά μεταξύ τους.

Η ομαδοποίηση αφορά περισσότερο την ανακάλυψη φυσικών ομάδων σε μεγάλες βάσεις δεδομένων. Ο διαχωρισμός στην περίπτωση αυτή γίνεται σε ομάδες που ο αριθμός τους προσδιορίζεται από την ίδια την διαδικασία, είναι κατά κάποιο τρόπο ο βέλτιστος, ώστε οι ομάδες να διαφέρουν όσο το δυνατόν πιο πολύ μεταξύ τους και τα στοιχεία κάθε ομάδας να είναι όσο το δυνατόν όμοια μεταξύ τους. Τα χαρακτηριστικά - μεταβλητές που θα χρησιμοποιηθούν για την ομαδοποίηση δεν είναι εκ των προτέρων γνωστά.

Για την παραγωγή των δέντρων απαιτείται η χρήση αλγορίθμων. Στην παρούσα διατριβή ο αλγόριθμος που χρησιμοποιήθηκε είναι ο C5.0. Η σωστή εκτέλεση του αλγόριθμου απαιτεί τα δεδομένα να χωρίζονται σε δύο διαφορετικές ομάδες [Hand D. et al, 2001]. Το πρώτο υποσύνολο είναι το σύνολο υπολογισμού ή η ομάδα εκπαίδευσης, του οποίου στόχος είναι να εκπαιδεύσει κατάλληλα το δέντρο, έτσι ώστε να δίνει σωστά αποτελέσματα με βάση το σύνολο των στοιχείων που του δίνουμε στην είσοδο. Το άλλο υποσύνολο είναι το σύνολο ελέγχου ή η ομάδα της δοκιμής, το οποίο χρησιμοποιείται, για να επιβεβαιώσει ότι το δέντρο κατασκευάστηκε και δουλεύει σωστά. Το δέντρο θα μπορεί να χρησιμοποιείται για περιπτώσεις όπου είναι γνωστή η είσοδος και άγνωστη η έξοδος μόνο όταν θα έχει ολοκληρωθεί σωστά η εκπαίδευση και η δοκιμή του.

Ένας από τους σημαντικούς λόγους που χρησιμοποιήθηκαν τα δέντρα αποφάσεων είναι η δυνατότητα που έχουν να δίνουν κανόνες για μελλοντική ανάπτυξη ώστε στη συνέχεια με κλασικές μεθόδους πειραματικού σχεδιασμού μπορούν να επαληθευτούν και στατιστικά οι παραπάνω κανόνες.

Τα δέντρα αποφάσεων προτιμούνται από άλλους αλγόριθμους Όρυξης δεδομένων από τη μορφή του παραγόμενου αποτελέσματος το οποίο είναι οι κανόνες. Αυτοί οι κανόνες μπορούν εύκολα να γίνουν αντιληπτοί και με τη βοήθεια σειράς



εντολών, όπως (γλώσσα SQL κ.α.) να απεικονιστούν στο αρχείο δεδομένων όλοι οι δυνατοί κανόνες με τη μορφή συναρτησοειδών.

Εξάλλου τα νευρωνικά δίκτυα, άλλη μέθοδος Όρυξης δεδομένων, αποτελούν ένα ισχυρό εργαλείο μοντελοποίησης σύνθετων προβλημάτων πρόβλεψης με μεγάλο αριθμό ανεξάρτητων μεταβλητών και πολλές αλληλεπιδράσεις. Η πολυπλοκότητα της δομής ενός νευρωνικού δικτύου καθιστά την ερμηνεία του δύσκολη σε σχέση με τα δένδρα απόφασης.

Η πολυπλοκότητα όμως των δένδρων απόφασης αυξάνεται με το πλήθος των ανεξαρτητών μεταβλητών αλλά και με το πλήθος των τιμών που μπορεί να λάβει η εξαρτημένη μεταβλητή. Η πολυπλοκότητα συνιστά μειονέκτημα που όμως αίρεται από το ότι η λογική ερμηνεία του κάθε κλάδου του δένδρου χωριστά γίνεται εύκολα, γεγονός το οποίο και συνιστά συγκριτικό πλεονέκτημα των δένδρων απόφασης έναντι των νευρωνικών δικτύων.

Τέλος, τα δέντρα προσφέρουν ιεράρχηση των παραγόντων σε σχέση με τα σύνολα κανόνων τα οποία παρακολουθούν το στόχο της ερευνητικής εργασίας. Σε σχέση μάλιστα με τα νευρωνικά δίκτυα τα δέντρα παρέχουν και τη διαισθητική αντίληψη του συνόλου των δεδομένων με ένα τρόπο βάση του οποίου αντιλαμβάνεται το στόχο σε μία σχέση αιτίου – αποτελέσματος.

### 3.3.1. Δέντρα απόφασης (Decision Trees)

Η Όρυξη δεδομένων αποτελείται από μεθόδους και αλγόριθμους, μέσω των οποίων αυτή πραγματοποιείται. Ένα λογισμικό Όρυξης δεδομένων, συνήθως χρησιμοποιεί περισσότερους από έναν αλγόριθμους, τον έναν μετά τον άλλον, για να γίνει η Όρυξη και πολλά από αυτά τα προϊόντα αφήνουν την επιλογή του αλγόριθμου στον χρήστη, δηλαδή ο χρήστης επιλέγει από διάφορους τον αλγόριθμο που θέλει να χρησιμοποιήσει. Τέλος, πολλά από αυτά τα προϊόντα έχουν υβριδικά συστήματα [Ismail M.A. et al, 1989] τα οποία συνδυάζουν έξυπνα τα πλεονεκτήματα κάποιων αλγορίθμων για να είναι ακόμα πιο ακριβές το τελικό αποτέλεσμα.

Ειδικότερα τα δέντρα αποφάσεων, που πολλές φορές αναφέρονται και ως «μαθητευόμενα δέντρα» (learning trees) χρησιμοποιούν πάντα συμβολική αναπαράσταση. Παίρνουν μια κατάσταση σαν είσοδο και βγάζουν μια απόφαση, η οποία συνήθως είναι τύπου Boolean της μορφής Yes / No, σαν έξοδο. Με άλλα λόγια κάνουμε μια ερώτηση στο δέντρο αποφάσεως και αυτό με την σειρά του θα μας

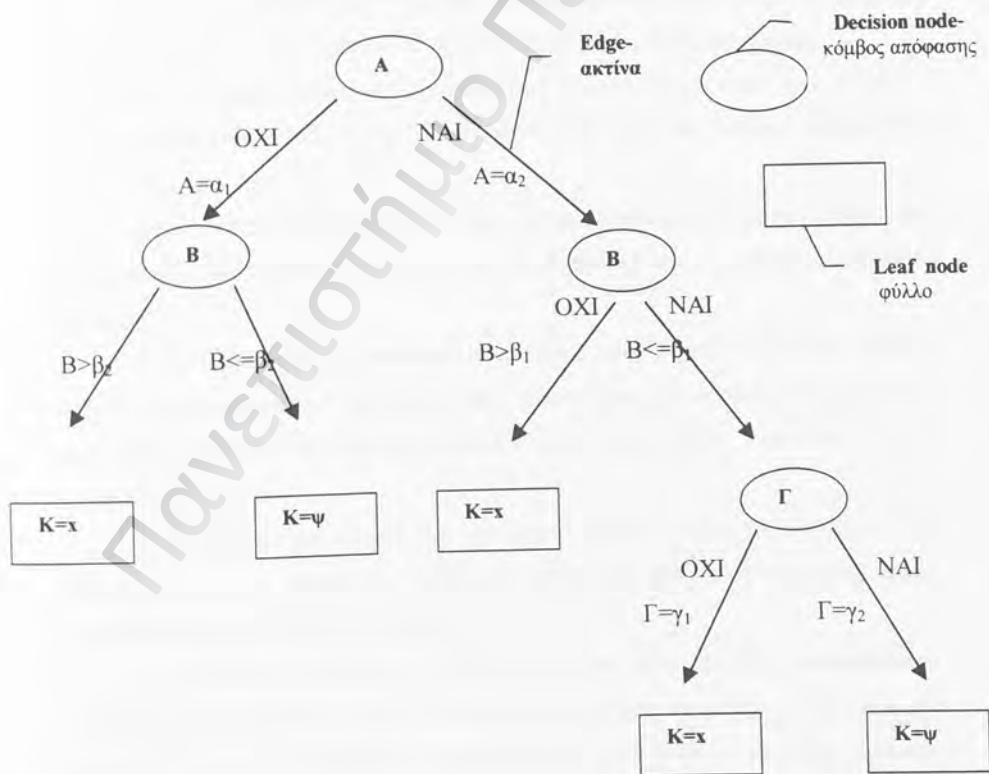
δώσει μια απάντηση της μορφής Yes ή No. Η παραπάνω ερώτηση είναι γνωστή πολλές φορές και ως στόχος του δέντρου.

Δέντρο απόφασης είναι, στην πράξη, ένας ταξινομητής / ελκυστής με τη μορφή δέντρου.

Για ένα δέντρο όπως αυτό που δίνεται στο διάγραμμα 3.1 παρατηρούμε τα εξής βασικά :

- το Φύλλο (leaf node) που αναπαριστά τις τιμές του προς διερεύνηση μεγέθους,
- τον Κόμβο απόφασης (decision node) που αναπαριστά τις τιμές των άλλων πεδίων / μεταβλητών και οι οποίες οδηγούν σε συγκεκριμένα φύλλα,
- την Ακτίνα - Δεσμό (edge).

ΔΙΑΓΡΑΜΜΑ 3.1  
Απεικόνιση δέντρου



Υπάρχουν δύο κύριοι τρόποι να διαβαστεί ένα δέντρο αποφάσεων αναλόγως με την κατεύθυνση

- Διαβάζοντας το δέντρο από τη ρίζα του έως τα φύλλα (δέντρα αποφάσεων).

Τα συμπεράσματα που προκύπτουν από αυτή τη διαδικασία αφορούν την κατανόηση των κανόνων που διαμορφώνουν περισσότερο την προς εξέταση μεταβλητή.

- Διαβάζοντας το δέντρο από τα φύλλα έως τη ρίζα του (σύνολα κανόνων).

Έτσι προσδιορίζονται όλοι οι τελικοί κανόνες για κάθε δυνατό ενδεχόμενο.

Στον διάγραμμα 3.1 δίνεται παράδειγμα απλού δέντρου αποφάσεων στο οποίο μπορούμε να παρατηρήσουμε τα εξής :

- Η ρίζα του συνολικού δένδρου είναι ο κόμβος A.
- Με ρίζα τον κόμβο A οι δυνατές επιλογές είναι δύο «Ναι» και «Όχι» για τιμές των πεδίων / μεταβλητών  $A=a_1$  και  $A=a_2$  μέσω των ακτίνων αντίστοιχα.
- Με ρίζα τον κόμβο B μέσω των ακτίνων με τιμές των πεδίων / μεταβλητών  $B \leq \beta_1$  και  $B > \beta_1$  καταλήγουμε στις επιλογές Γ και Κ=x αντίστοιχα.
- Για το Γ παρατηρούμε ότι μπορεί να θεωρηθεί φύλλο με ρίζα τον κόμβο B καθώς και ότι το Γ μπορεί να θεωρηθεί ρίζα για τα επόμενα επίπεδα.
- Αναλογικά μπορούμε να σχολιάσουμε για το Β με τιμές των πεδίων / μεταβλητών  $B \leq \beta_2$  και  $B > \beta_2$  και για το Γ με τιμές των πεδίων / μεταβλητών  $\Gamma = \gamma_1$  και  $\Gamma = \gamma_2$ .

Σημαντικά χαρακτηριστικά για ένα δέντρο αποφάσεων είναι το πλήθος των περιπτώσεων που καλύπτει ο κανόνας (coverage) καθώς και η ακρίβεια (confidence) του κανόνα.

- Το πλήθος των περιπτώσεων στις οποίες εφαρμόζεται ο κανόνας είναι το πλήθος των στοιχείων της τομής όλων των υποσυνόλων του συνόλου δεδομένων για τις οποίες εφαρμόζονται οι συνθήκες «ΕΑΝ συμβεί το γεγονός j». Αυτό υπολογίζεται σε απόλυτο αριθμό.

- Η ακρίβεια του κανόνα έχει ως παρονομαστή το παραπάνω πλήθος των περιπτώσεων και ως αριθμητή το πλήθος των εγγραφών του υποσυνόλου στις οποίες παρατηρείται το συγκεκριμένο γεγονός.

Με βάση το συγκεκριμένο σύνολο κανόνων, που αποτελεί ουσιαστικά το υπολανθάνων θεωρητικό μοντέλο / πλαίσιο που παράγει τα δεδομένα, σε κάθε μία εγγραφή του αρχείου δεδομένων προσδιορίζεται η, με βάση το μοντέλο, τιμή του



παρατηρούμενου φαινομένου και εκτιμάται η αξιοπιστία αυτής της πρόβλεψης σε μία κλίμακα από το 0 έως το 1.

Στην εικόνα 3.2 δίνεται ισοδύναμος μαθηματικός τρόπος γραφής δένδρου τον οποίο θα χρησιμοποιήσουμε στη συνέχεια για να μελετήσουμε τα αποτελέσματα.

### EIKONA 3.2

Απεικόνιση δένδρου

$A=[\alpha_1]$  [Mode: N]  $(\alpha, \beta) \rightarrow N$

$A=[\alpha_2]$  [Mode: Y]  $(\gamma)$

$B \leq \beta_1$  [Mode: Y]  $(\delta)$

$\Gamma=[\gamma_1]$  [Mode: N]  $(\eta, \theta) \rightarrow N$

$\Gamma=[\gamma_1]$  [Mode: Y]  $(\iota, \kappa) \rightarrow Y$

$B > \beta_1$  [Mode: N]  $(\epsilon, \zeta) \rightarrow N$

Στη μορφή αυτή παρατηρούμε ότι είναι N, Y οι επικρατούσες τιμές. Η συνθήκη  $A=[\alpha_1]$  πληρούται για α πλήθος εγγραφών και το πλήθος των εγγραφών X που έχουν τιμή N δίνεται από την σχέση  $X / \alpha = \beta$ . Σχολιασμός ανάλογος μπορεί να γίνει και για τους υπόλοιπους διαχωρισμούς στο δένδρο.

Επιπλέον για τα δέντρα αποφάσεων αναγκαίο είναι να σχολιάσουμε [Quinlan J. R., 1986] και [(3)-web] και τα εξής:

1. Έχουν υψηλή δυνατότητα για ταξινόμηση και πρόβλεψη.
2. Προσφέρουν επεξήγηση στην κατανόηση του φαινομένου που παράγει το προς μελέτη συγκεκριμένο γεγονός. Αναλόγως με την αύξηση του αριθμού των ανεξαρτήτων μεταβλητών αλλά και με το πλήθος των τιμών της εξαρτημένης μεταβλητής – μεταβλητή πρόβλεψης αυξάνεται και η πολυπλοκότητα ενός δένδρου. Αν και τα πολύπλοκα δένδρα δύσκολα γίνονται κατανοητά, εν τούτοις η λογική που διέπει ξεχωριστά τον κάθε κλάδο είναι εύκολα κατανοητή και μπορεί κανείς να ερμηνεύει, κατά κάποιο τρόπο, τις προβλέψεις που κάνει.
3. Προσφέρουν ιεράρχηση των παραγόντων που επιλέχθηκαν για την ανάλυση σε σχέση με το προς εξέταση γεγονός.
4. Αποκλείουν από τη φύση τους, όλους τους παράγοντες που δεν επιδρούν σημαντικά στο σχηματισμό τους, με αποτέλεσμα να παρέχουν αναλυτική εικόνα για τη διαδικασία δράσης του πεδίου / μεταβλητής.

## 5. Αποτελούν χαρακτηριστικό παράδειγμα επαγωγής από ένα σύνολο δεδομένων.

Τα κυριότερα χαρακτηριστικά των συνόλων δεδομένων στα οποία εφαρμόζονται δέντρα είναι :

- Ο αριθμός των διαφορετικών τιμών που παίρνουν τα πεδία / μεταβλητές πρέπει να είναι γνωστός εκ των προτέρων, δηλαδή δεν μπορεί να θεωρηθεί ότι υπάρχουν άλλοι εξωγενείς παράγοντες που δεν έχουμε παρατηρήσει. Στατιστικά όλες οι συμβάσεις είναι εκ των προτέρων γνωστές.

- Οι τιμές των πεδίων / μεταβλητών πρέπει να είναι προκωδικοποιημένες έχοντας συγκεκριμένες τιμές οι οποίες και θα αποτυπώνονται στο δέντρο.

- Μία εγγραφή θα ανήκει σε μόνο μία συγκεκριμένη ομάδα και θα συνδέεται με μόνο ένα γεγονός.

- Απαιτείται τέλος, μεγάλο πλήθος περιπτώσεων για να είναι αποτελεσματικός ο αλγόριθμος.

Για την κατασκευή αυτών των δέντρων χρησιμοποιήθηκε ο αλγόριθμος C5.0 που παράγει δέντρα με βάση τη γνώση του τελικού συμβάντος. Η ερμηνεία και κατανόηση κάθε ενός δέντρου καθώς και της οικογένειας δίνει πληρέστερη και καλύτερη αντίληψη για τη διαδικασία που παράγει ένα γεγονός αποτελώντας ένα ισχυρό όπλο για τον ερευνητή. Ακόμα πιο σημαντικό είναι ότι τα δέντρα που παράγονται με το C5.0 είναι και τα περισσότερο σημαντικά δέντρα από το σύνολο των αναρίθμητων δέντρων που μπορούμε να παράγουμε. Αυτά τα δέντρα μας βοηθούν στην ακόμα καλύτερη κατανόηση του φαινομένου. Η μεθοδολογία αυτή χρησιμοποιείται για να περιγράφεται η συμπεριφορά των δεδομένων αναφορικά με την τιμή ενός συγκεκριμένου πεδίου / μεταβλητής. Η δομή του δέντρου ακολουθεί τη λογική στο σχηματισμό της διαδικασίας αποφάσεων. Επίσης η δομή του δέντρου είναι αυτή που βοηθά στην κατανόηση της διαδικασίας που παράγει το προς διερεύνηση γεγονός.

Από την εννοιολογική προσέγγιση του δέντρου αποφάσεων, οι αλγόριθμοι προσπαθούν να βρουν το καλύτερο δυνατό κανόνα που επαληθεύεται στο μεγαλύτερο τμήμα του συνόλου δεδομένων. Κάτι τέτοιο δεν ευνοεί την εφαρμογή του κανόνα σε ένα νέο σύνολο δεδομένων άρα και στην αξιοπιστία του κανόνα. Η ικανότητα του C5.0 να αντισταθμίζει αυτό το φαινόμενο (parsimony) [Iba W. et al, 1988] μας οδήγησε στην επιλογή του συγκεκριμένου αλγορίθμου ανάμεσα από αρκετούς

άλλους. Η διερεύνηση αυτής της ικανότητας ταξινόμησης του αλγορίθμου συνεισφέρει σημαντικά στην ανάλυση των αποτελεσμάτων.

### 3.3.2. Ο Αλγόριθμος C5.0

Ο σκοπός του αλγορίθμου είναι να κατασκευαστεί ένας δέντρο  $T$  από ένα σύνολο εγγραφών και στη συνέχεια. περιγράφεται αλγοριθμικά η βάση του C5.0.

Δεδομένα:  $R$ : το σύνολο των ανεξάρτητων πεδίων / μεταβλητών.

$C$ : το εξαρτημένο πεδίο / μεταβλητή.

$S$ : το σύνολο δεδομένων / σύνολο εγγραφών.

Με τον συμβολισμό  $\{ R, C, S \}$  έχουμε μια συνοπτική απόδοση των δεδομένων.

Ζητούμενο: Κατασκευή δέντρου απόφασης  $T$ .

Βήμα 1. Αν το  $S = \emptyset$  να παραχθεί ένα μόνο φύλλο με τιμή FAILURE.

Βήμα 2. Αν το  $S = \{c\}$  δηλαδή όλες οι εγγραφές έχουν την ίδια τιμή για το εξαρτημένο πεδίο να παραχθεί ένα μόνο φύλλο με αυτή τη συγκεκριμένη τιμή.

Βήμα 3. Αν το  $R = \emptyset$  να παραχθεί ένα μόνο φύλλο με την περισσότερο συχνά παρατηρούμενη τιμή του  $S$ . Αυτόματα έτσι παράγονται σφάλματα τα οποία είναι οι μη σωστά ταξινομημένες περιπτώσεις του συνόλου  $S$ .

Βήμα 4. Υπολόγισε για όλα τα στοιχεία του  $R$  τις τιμές  $GAIN(S,*)$ , όπου η συνάρτηση  $GAIN(S,*)$  ορίζεται στην συνέχεια στον επεξηγηματικό σχολιασμό 1 του αλγορίθμου.

Έστω  $A$  είναι εκείνο το πεδίο / μεταβλητή (υποσύνολο του  $R$ ) που κάνει μέγιστη τη συνάρτηση  $Gain(S, A)$  και  $\{a_j \mid j=1,2, \dots, m\}$  είναι οι δυνατές τιμές του  $A$  από όλα τα υπόλοιπα πεδία / μεταβλητές και όπου  $\{S_j \mid j=1,2, \dots, m\}$  είναι το πλήθος στοιχείων του  $S$  που αντιστοιχούν στις εγγραφές του  $a_j$  για το  $A$ .

Βήμα 5. Κατασκεύασε ένα δέντρο με ρίζα το  $A$  και ακτίνες τα  $a_1, a_2, \dots, a_m$  που καταλήγουν στα  $m$  υπο-δέντρα:

$$\{R-[A], C, S_1\}, \{R-[A], C, S_2\}, \dots, \{R-[A], C, S_m\}.$$



Βήμα 6. Αναδρομικά, εφάρμοσε την παραπάνω διαδικασία στα υποσύνολα  $\{S_j \mid j=1,2,\dots,m\}$  έως ότου αυτά γίνουν κενά ή γενικά προσδιοριστεί άλλο κριτήριο.

Επεξηγηματικός σχολιασμός I του αλγορίθμου

Συνοπτικά το δέντρο αποφάσεων διαλέγει εκείνο το πεδίο / μεταβλητή A το οποίο διαφοροποιεί με τον καλύτερο δυνατό τρόπο το σύνολο των εγγραφών, χωρίζοντάς το σε υποσύνολα, ως προς το εξαρτημένο πεδίο / μεταβλητή. Βασικό στοιχείο στο παραπάνω σκεπτικό είναι να κατασκευαστεί το κριτήριο αυτό που ποσοτικοποιεί το παραπάνω αίτημα. Για το σκοπό αυτό χρησιμοποιείται η έννοια της εντροπίας από τη θεωρία Πληροφοριών [Cover T. et al, 1991] και [MacKay D., 1998] η οποία ορίζεται [MacKay D., 1998] και [(1)-web] ως εξής :

$$Entropy(S) = \sum_{i=1}^c - p_i \log_2 p_i$$

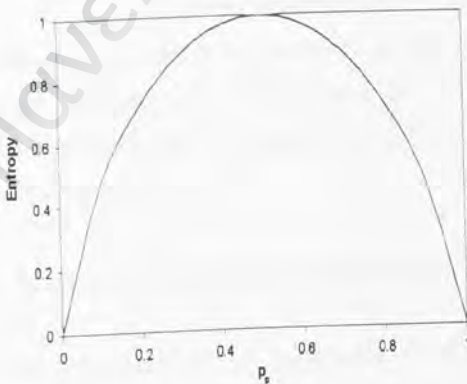
όπου  $p_i$  είναι το ποσοστό των περιπτώσεων που ανήκουν στην τάξη  $i$ ,

όπου  $c$  είναι οι διαφορετικές τιμές του εξαρτημένου πεδίου / μεταβλητής.

Η εντροπία E είναι το αντίθετο της πληροφορίας I του συνόλου των δεδομένων, όπου από το πιθανοθεωρητικό αντίθετο ισχύει  $E+I=1$ . [MacKay D., 1998] και [(7)-web], και πρακτικά αυτή καταγράφει τον ελάχιστο αριθμό πληροφορίας απαραίτητης να τμηματοποιηθεί το σύνολο S.

### ΓΡΑΦΗΜΑ 3.3

Μεταβολής εντροπίας



Το γράφημα 3.3 δείχνει την συνάρτηση εντροπίας για ένα συγκεκριμένο γεγονός καθώς η μεταβλητότητα της πιθανότητας εμφάνισης του παίρνει τιμές από το 0 στο 1. Η συνάρτηση της εντροπίας έχει μια θεμελιώδη ιδιότητα, για ισοπίθανα αποτελέσματα χρειαζόμαστε περισσότερη πληροφορία I, από ότι για σχεδόν βέβαια γεγονότα, για να διαφοροποιήσουμε το σύνολο δεδομένων.

Όσο λοιπόν εξελίσσεται αναδρομικά το δέντρο τόσο μεταβάλλονται στα φύλλα του δέντρου οι πιθανότητες εμφάνισης του γεγονότος προς ακραίες τιμές, απαιτώντας έτσι και σχετικά μικρότερη πληροφορία I [Yang H. and Amari S., 1997] για να διαφοροποιηθεί το δέντρο επαρκώς.

Με βάση την εντροπία σαν μέτρο έλλειψης πληροφορίας για το σύνολο S μπορούμε να κατασκευάσουμε ένα ποσοτικό μέγεθος το οποίο να δείχνει πόσο βοηθάει η επιπλέον πληροφορία του συνόλου A στην ταξινόμηση του S.

Η συνάρτηση Gain (S, A) που αποτελεί το «όφελος» από τη γνώση του A, μπορεί να οριστεί και σαν κέρδος πληροφορίας. Το κέρδος πληροφορίας Gain (S, A) ενός χαρακτηριστικού A, αναφορικά με το σύνολο S, ορίζεται με τον ακόλουθο λογικό τύπο :

$$Gain(S, A) = Entropy(S) - \sum_{v \in Values(A)} \frac{|S_v|}{|S|} Entropy(S_v)$$

όπου :

Values(A) είναι όλες οι επιτρεπτές τιμές του A,

$S_v$  υποσύνολο του S για το οποίο το A έχει τιμή v

δηλαδή,  $S_v = \{s \in S \mid A(s) = v\}$ .

Gain (S,A) είναι η σχετική μείωση στην εντροπία του S δεδομένου ότι γνωρίζουμε τις τιμές του A [Markham J. M et al., 1999]. Αφαιρεί δηλαδή την παραγόμενη εντροπία των  $S_v$ , λαμβάνοντας υπόψιν το υποσύνολο των περιπτώσεων  $|S_v| / |S|$  που ανήκουν στο  $S_v$ .

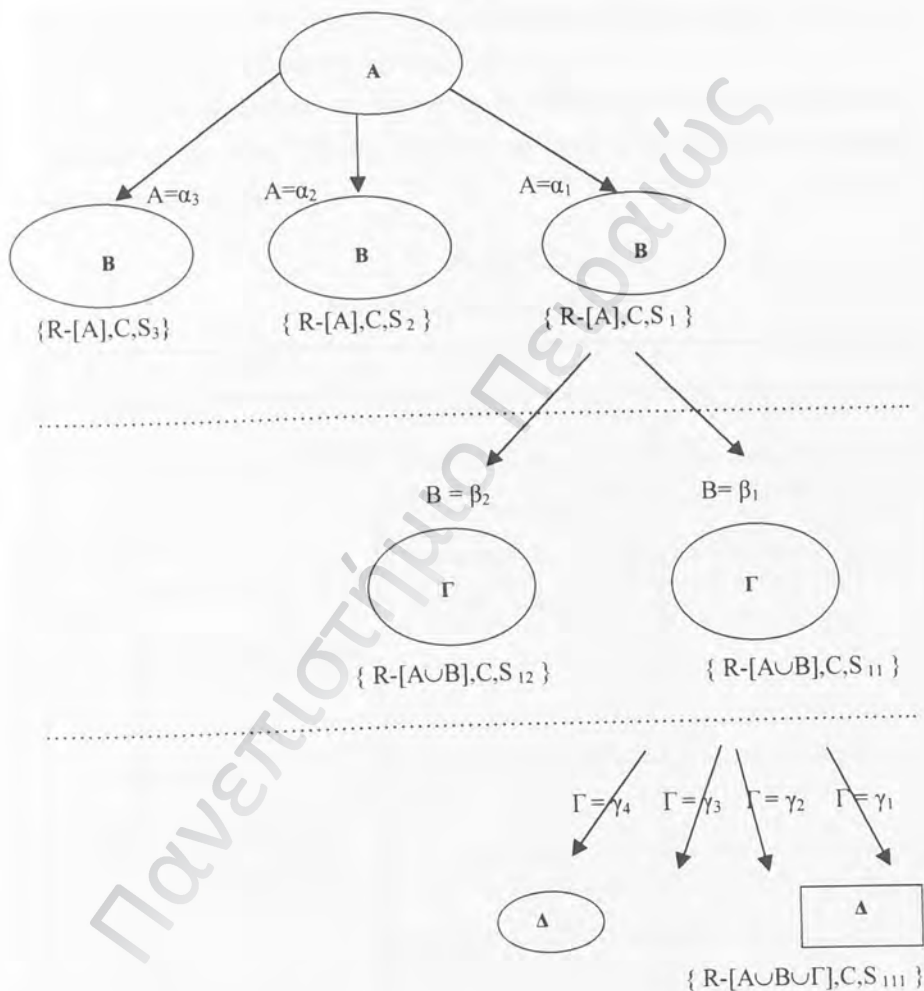
### 3.3.2.1. Απόδοση του αλγόριθμου σε διάγραμμα

Στο διάγραμμα 3.4 παρουσιάζονται οι διαχωρισμοί που πραγματοποιεί ο αλγόριθμος και μπορούμε να παρατηρήσουμε ότι:

- Ρίζα είναι το A με τρεις για την περίπτωση δυνατές τιμές του  $A = \{a_1, a_2, a_3\}$ . Το S διαμοιράστηκε σε τρία υποσύνολα με πλήθος στοιχείων το καθένα  $S_1, S_2, S_3$  που αντιστοιχούν στις εγγραφές με τιμές  $a_1, a_2$  και  $a_3$ .

ΔΙΑΓΡΑΜΜΑ 3.4

Απεικόνιση των διαχωρισμών με δένδρο



- Στο κόμβο B, που προέκυψε για την τιμή  $A = a_1$ , για το πεδίο / μεταβλητή B έχουμε σαν σύνολο δεδομένων το  $\{R-[A], C, S_1\}$ . Οι δυνατές τιμές του B στον



συγκεκριμένο κλάδο είναι  $B = \{ \beta_1, \beta_2 \}$  ενώ  $S_{11}, S_{12}$  είναι το πλήθος στοιχείων του  $S_1$  που αντιστοιχούν στις εγγραφές με τιμές  $\beta_1, \beta_2$ .

• Κατ'αναλογία σχολιασμός μπορεί να γίνει για τους κόμβους  $B$  που προέκυψαν για τιμές  $A=\alpha_2$  και  $A=\alpha_3$ .

• Στο κόμβο  $\Gamma$ , που προέκυψε για την τιμή  $B=\beta_1$ , για το πεδίο / μεταβλητή  $\Gamma$  έχουμε σαν σύνολο δεδομένων το  $\{ R-[A \cup B], C, S_{11} \}$ . Οι δυνατές τιμές του  $\Gamma$  είναι  $\Gamma = \{ \gamma_1, \gamma_2, \gamma_3, \gamma_4 \}$  ενώ  $S_{111}, S_{112}, S_{113}, S_{114}$  είναι το πλήθος στοιχείων του  $S_{11}$  που αντιστοιχούν στις εγγραφές με τιμές  $\gamma_1, \gamma_2, \gamma_3, \gamma_4$ .

Αναλυτικά μπορεί να αποδοθούν οι διαχωρισμοί που πραγματοποιεί ο αλγόριθμος και στον Πίνακα 3.5, όπου το σύνολο των δεδομένων συνοπτικά απεικονίζεται σαν  $\{R, C, S\}$ .

### ΠΙΝΑΚΑΣ 3.5

Απεικόνιση των διαχωρισμών

A	B	$\Gamma$	$\Delta$
$\{R, C, S\}$ / $A=\alpha_1$	$\{ R-[A], C, S_1 \} / B=\beta_1$	$\{ R-[A \cup B], C, S_{11} \} / \Gamma=\gamma_1$	$\{ R-[A \cup B \cup \Gamma], C, S_{111} \} /$
		$\{ R-[A \cup B], C, S_{11} \} / \Gamma=\gamma_2$	$\{ R-[A \cup B \cup \Gamma], C, S_{112} \} /$
		$\{ R-[A \cup B], C, S_{11} \} / \Gamma=\gamma_3$	$\{ R-[A \cup B \cup \Gamma], C, S_{113} \} /$
		$\{ R-[A \cup B], C, S_{11} \} / \Gamma=\gamma_4$	$\{ R-[A \cup B \cup \Gamma], C, S_{114} \} /$
	$\{ R-[A \cup B], C, S_{12} \} /$		
$\{R, C, S\}$ / $A=\alpha_2$	$\{ R-[A], C, S_2 \} /$		
	$\{ R-[A], C, S_2 \} /$		
$\{R, C, S\}$ / $A=\alpha_3$	$\{ R-[A], C, S_3 \} /$		
	$\{ R-[A], C, S_3 \} /$		

### 3.3.2.2. Βελτιώσεις του βασικού αλγόριθμου

Ο αλγόριθμος που περιγράφηκε παράγει το καλύτερο δυνατό δέντρο για ένα συγκεκριμένο όμως σύνολο δεδομένων. Για να είναι το δέντρο χρήσιμο και σε νέα διαφορετικά δεδομένα παραγόμενα από την ίδια διαδικασία, πρέπει να προστεθούν τρία ακόμα βήματα στον αλγόριθμο [Berry M.J.A. et al, 1997].

Βήμα 1. Το σύνολο των δεδομένων χωρίζεται στα δύο – όχι απαραίτητα ίσα – μέρη, σε υποσύνολο υπολογισμού  
σε υποσύνολο ελέγχου

Βήμα 2. Βρίσκεται ένας κανόνας στο πρώτο υποσύνολο δεδομένων και αυτός ο κανόνας ελέγχεται στο δεύτερο υποσύνολο δεδομένων.

Βήμα 3. Εφόσον ο κανόνας αυτός ισχύει και στο δεύτερο υποσύνολο δεδομένων καταγράφεται ο κανόνας για το συγκεκριμένο τμήμα του αρχείου.

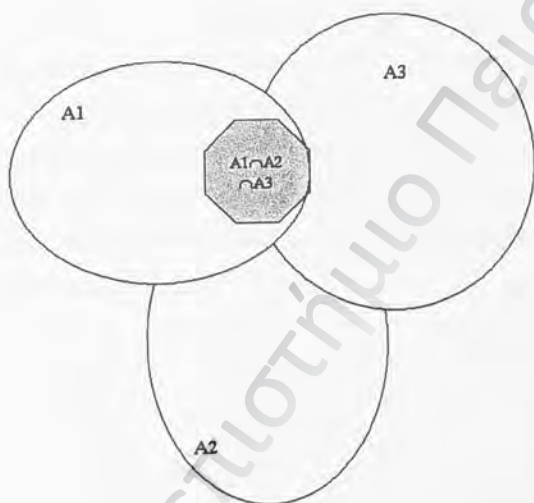
Πέρα από το βασικό σώμα του αλγόριθμου διαχρονικά έχουν ενσωματωθεί και μερικές ακόμα ιδιότητες / βελτιώσεις του αλγόριθμου [(4)-web] και [(9)-web] και [(10)-web] οι σημαντικότερες των οποίων είναι :

1. Ενώνει τα τελικά κλαδιά που έχουν λίγες και σχετικά όμοιες εγγραφές και υπολογίζει νέα.
2. Διαχειρίζεται θόρυβο [Quinlan J.R., 1986a] στα δεδομένα, δηλαδή περιπτώσεις που έχουν σχετικά μικρή απόσταση μεταξύ τους αλλά αντιστοιχούν σε διαφορετική τιμή του προς διερεύνηση πεδίου / μεταβλητής
3. Επεξεργάζεται σημαντικό αριθμό εγγραφών και παράγει διακριτά αποτελέσματα χρησιμοποιώντας τη μεθοδολογία Boosting [Druker H., 1999], επιτρέποντας την ταχύτερη ανακάλυψη περιπτώσεων οι οποίες εμφανίζουν υψηλά ποσοστά θορύβου.
4. Διαχειρίζεται σημαντικά αριθμητικά προβλήματα όπως οι ελλείπουσες τιμές και πεδία / μεταβλητές με αφηρημένους χαρακτήρες. Σημειώνεται εδώ, ότι επιτρέπει την ταυτόχρονη αντιμετώπιση της πληροφωρίας και ποσοτικά και ποιοτικά. Χειρίζεται πολύ καλά τόσο ποσοτικά όσο και κατηγορικά δεδομένα γεγονός το οποίο ελαχιστοποιεί την ανάγκη για μετασχηματισμό των δεδομένων.
5. Μπορεί να διαχειρίζεται συνεχείς μεταβλητές τις οποίες διακριτοποιεί σε  $n$ -συνεχόμενα υποσύνολα  $a_1, a_2, \dots, a_n$  τα οποία διαφοροποιούν καλύτερα το  $S$ .

6. Ορίζεται, αν είναι επιθυμητό, το μέγεθος του τελικού φύλλου ώστε να έχει ένα ελάχιστο αριθμό περιπτώσεων, κάνοντας την μεθοδολογία περισσότερο λειτουργική.

### 3.3.2.3. Το διάγραμμα αποτελέσματος

Θεωρούμε ότι δίνονται το σύνολο των ανεξάρτητων πεδίων / μεταβλητών  $R$ , το εξαρτημένο πεδίο / μεταβλητή  $C$ , το σύνολο δεδομένων - σύνολο εγγραφών  $S$ , και για παράδειγμα ώστε είναι εύκολη η απεικόνιση με διάγραμμα, τρία σύνολα  $A_1$ ,  $A_2$ , και  $A_3$ . Για τα σύνολα  $A_1$ ,  $A_2$ ,  $A_3$  ισχύει ότι  $A_1 = \{R, C, S\}$ ,  $A_2 = \{R-[A], C, S_1\}$  και  $A_3$



ΔΙΑΓΡΑΜΜΑ 3.6

Διάγραμμα αποτελέσματος

$= \{R-[A \cup B], C, S_{11}\}$ . Ζητούνται τα  $A_1, A_2, A_3$  τέτοια ώστε η τομή τους  $A_1 \cap A_2 \cap A_3$  να διαχωρίζει καλύτερα σε κατηγορίες τις τιμές του  $C$ . Το σύνολο  $A = \{A_1 \cap A_2 \cap A_3\}$  αντιστοιχεί στο γεγονός  $C_j$  που θα εμφανισθεί με την μεγαλύτερη πιθανότητα από ότι αν χρησιμοποιηθεί μία καθαρά τυχαία προσέγγιση. Με την βοήθεια διαγραμμάτων του Venn σχηματικά παριστάνεται η περιγραφείσα διαδικασία στο 3.6 διάγραμμα αποτελέσματος.



Για τα σύνολα  $A_1, A_2, A_3$  τα οποία στην περίπτωση μας είναι πεδία / μεταβλητές, με αναφορά στη θεωρία Συνόλων, μπορούμε να υπολογίσουμε το πλήθος των περιπτώσεων του  $\{ A_1 \cap A_2 \cap A_3 \}$  στις οποίες ισχύει ο κανόνας, που είναι το πλήθος των στοιχείων του  $\{ A_1 \cap A_2 \cap A_3 \}$ .

Επίσης μπορούμε να υπολογίσουμε την ακρίβεια του κανόνα, η οποία ορίζεται σαν το ποσοστό του το πλήθους των περιπτώσεων του  $\{ A_1 \cap A_2 \cap A_3 \}$  στο σύνολο των περιπτώσεων για τις οποίες παρατηρείται το συγκεκριμένο γεγονός  $C_j$ .

### 3.3.3. Ταξινομητικοί κανόνες

Οι ταξινομητικοί κανόνες [(15)-web] και [(17)-web] αποτελούν έναν εναλλακτικό τρόπο να διαβαστεί ένα δέντρο αποφάσεων που έχει παραχθεί με την εφαρμογή του αλγορίθμου C5.0. Η προσέγγιση του C5.0 είναι η μελέτη των πεδίων / μεταβλητών του R που διαφοροποιούν καλύτερα το σύνολο δεδομένων S αναφορικά με το εξαρτημένο πεδίο / μεταβλητή C. Δε δίνουν δηλαδή την εστίαση στο C αλλά στο S [(3)-web]. Οι ταξινομητικοί κανόνες μετασχηματίζουν το δέντρο στην ανάποδη φορά του. Στόχος τους είναι δηλαδή να μελετηθούν τα τελικά φύλα του δέντρου με βάση τις ισχύουσες συνθήκες. Τα αποτελέσματα των ταξινομητικών κανόνων έχουν την ακόλουθη μορφή:

EAN	(συνθήκη $A_1$ )
ΚΑΙ	(συνθήκη $A_2$ )
ΚΑΙ	(συνθήκη $A_3$ )
.....	
.....	
ΚΑΙ	(συνθήκη $A_n$ )
ΤΟΤΕ	θα συμβεί το γεγονός $C_j$ .

και δίνεται σε αυτή την περίπτωση το αντίστοιχο πλήθος των περιπτώσεων που καλύπτει ο κανόνας (coverage) καθώς και η ακρίβεια (confidence) του κανόνα. Για παράδειγμα στην εικόνα 3.7 δίνεται η μορφή ενός κανόνα για γεγονός με τιμή "Yes" την οποία θα χρησιμοποιήσουμε στη συνέχεια για την μελέτη των αποτελεσμάτων.

### ΕΙΚΟΝΑ 3.7

Γενική μορφή ταξινομητικού κανόνα

Κανόνας για γεγονός με τιμή "Yes"

if	$A_1 = F$
and	$A_2 > B$
and	$A_3 = 4$
then	$\rightarrow \text{Yes}(\alpha, \beta)$

Το σύνολο  $A = \{ A_1 \cap A_2 \cap A_3 \}$  έχει πλήθος  $\alpha$  εγγραφών. Μέσα στο οποίο οι εγγραφές πλήθους  $X$  έχουν τιμή  $C_j = \text{"Yes"}$ , όπου το  $X$  υπολογίζεται από τον τύπο  $X / \alpha = \beta$ .

Η βασική πληροφορία που φαίνεται από τους κανόνες ταξινόμησης και δε φαίνεται από τα δέντρα αποφάσεων είναι το πλήθος των τελικών κανόνων. Οφέλη προκύπτουν από την γνώση του πλήθους των τελικών κανόνων, για παράδειγμα αν για το γεγονός  $C_j$  με τιμή "Yes" υπάρχουν πέντε κανόνες ενώ για το  $C_j$  γεγονός με τιμή "No" υπάρχουν τριάντα κανόνες αυτό σημαίνει ότι το γεγονός  $C_j$  με τιμή "Yes" μοντελοποιείται πιο λιτά και δε χρειάζεται η λεπτομέρεια που απαιτείται για το γεγονός  $C_j$  με τιμή «No».

Αποτελεί, τέλος, συνέπεια των ιεραρχικών δέντρων να παράγονται κανόνες που επικαλύπτονται οι συνθήκες, σε σχέση με το λογικό συμπέρασμα.

#### 3.3.4. Ο αλγόριθμος Generalized rule induction (G.R.I).

Με βάση την γενική μεθοδολογία σχηματισμού των δέντρων απόφασης, υπάρχουν αρκετοί ακόμα αλγόριθμοι για τη παραγωγή δέντρων οι οποίοι βελτιώνουν τον C5.0. αν και αποτελούν υποπερίπτωσή του. Ο σημαντικότερος από τους οποίους είναι ο G.R.I. Ο παραπάνω αλγόριθμος ανιχνεύει την ύπαρξη όλων των σημαντικών κανόνων στο αρχείο των δεδομένων. Στην πράξη κάνει την ανίχνευση και τον σχηματισμό όλων των δέντρων απόφασης με βάση το θεμελιώδες κριτήριο για τις περιπτώσεις στις οποίες εφαρμόζεται ο κανόνας. Η μεθοδολογία παράγει μικρά σε μήκος δέντρα τα οποία προσδιορίζουν το ζητούμενο αποτέλεσμα ανιχνεύοντας τους

σημαντικούς κανόνες στο αρχείο δεδομένων. Η πιο σπουδαία λοιπόν παράμετρος του αλγορίθμου αυτού είναι το μέγεθος του δέντρου. Και αυτό γιατί, ενώ συνήθως υπάρχουν πολλά δέντρα που αν είναι γνωστό ένα αίτιο προβλέπεται, με ακρίβεια καλύτερης του τυχαίου, το αποτέλεσμα, όταν θέλουμε να συνδυάσουμε περισσότερα από ένα αίτια ελαττώνεται η ακρίβεια πρόγνωσης με αποτέλεσμα να υπολείπεται της άλλης σημαντικής παραμέτρου η οποία είναι η επιθυμητή ακρίβεια.

Εν συντομία ο αλγόριθμος προσπαθεί να εκτιμήσει τις εξής σχέσεις :

EAN Y=y TOTE X=x με πιθανότητα p

Οι τιμές των X και Y είναι δύο πεδία / μεταβλητές του αρχείου δεδομένων τα οποία λαμβάνουν τιμή "x" και "y" αντίστοιχα. Το X περιορίζεται σε ένα μόνο πεδίο / μεταβλητή αλλά το Y μπορεί να είναι συνδυασμός δυνατών τιμών των πεδίων / μεταβλητών του αρχείου δεδομένων. Με βάση αυτά τα στοιχεία ο αλγόριθμος προσπαθεί να εκτιμήσει όλες τις δυνατές δεσμευμένες πιθανότητες [Begy M. et al, 1997] όπως αυτές υπολογίζονται από την ακόλουθη εξίσωση :

$$J(X:Y=y) = p(y) \cdot \left( p(x|y) \cdot \log \frac{p(x|y)}{p(x)} + (1-p(x|y)) \cdot \log \frac{(1-p(x|y))}{(1-p(x))} \right)$$

Στην παραπάνω σχέση  $p(y)$  είναι η πιθανότητα εμφάνισης του «Y=y»,  $p(x)$  είναι αντίστοιχα πιθανότητα εμφάνισης του «X=x» ενώ  $p(x|y)$  είναι η δεσμευμένη πιθανότητα να παρατηρηθεί το x εφόσον είναι γνωστό το y. Έτσι ο αλγόριθμος δεν περιορίζεται στην ανακάλυψη κανόνων ταξινόμησης ενός γεγονότος αλλά στην ανίχνευση όλων των κανόνων που προκύπτουν από τον συνδυασμό των πεδίων / μεταβλητών.

Η παραπάνω σχέση αποτελεί εφαρμογή (ειδίκευση) του GAIN κριτηρίου για  $c=2$ , αφού περιορίζει (στατιστικά δεσμεύει) τη μορφή του επιθυμητού δέντρου σε δύο τιμές.



## ΚΕΦΑΛΑΙΟ 4

### 4.1 Εισαγωγή

Στις επόμενες παραγράφους αναφέρονται οι πηγές από τις οποίες αντλήθηκαν οι πληροφορίες για την δημιουργία του Αρχείου δεδομένων. Περιγράφεται η δημιουργία, η γραμμογράφηση, και η κωδικοποίηση του. Δίνονται τέλος, στοιχεία για τις εγγραφές του αρχείου δεδομένων.

### 4.2 Δημιουργία του αρχείου δεδομένων

Το αρχείο δεδομένων κατασκευάστηκε από στοιχεία που συλλέχθηκαν από διάφορες έγκυρες πηγές:

- η εβδομαδιαία έκδοση "Lloyd's Casualty Week", Lloyd's of London Press,
- οι ετήσιες εκδόσεις "Lloyd's Register of Ships", Lloyd's Register Ltd.

και συμπληρωματικά χρησιμοποιήθηκαν πηγές όπως:

- οι ετήσιες εκδόσεις "World Casualty Statistics", Lloyd's Register Ltd,
- ο Άτλας "Maritime ATLAS", 1993, Lloyd's of London Press Ltd.

Για να καλυφθεί όσο το δυνατόν ευρύτερα και διεπιστημονικά το θέμα έχουν καταγραφεί πλήθος σχετικών πληροφοριών από τις πηγές που προαναφέρθηκαν και αναλυτικότερα:

α) Γενικά στοιχεία (όπως κωδικός εγγραφής ταξιδιού, έτος, μήνας, ημερομηνία, είδος εντύπου).

β) Στοιχεία πλοίων (όπως όνομα, σημαία, τύπος, ονόματα προηγούμενων ιδιοκτητών, πλήθος αλλαγής ιδιοκτητών, όνομα τελευταίου πλοιοκτήτη, όνομα τελευταίου διαχειριστή, έτος που περιήλθε το πλοίο στον τελευταίο ιδιοκτήτη, χωρητικότητα (grt), Νηογνώμονας, έτος κατασκευής, ηλικία, ναυπηγείο, κατασκευαστής, μοντέλο, ισχύς μηχανής, κατασκευαστής μηχανής, ταχύτητα πλοίου.

γ) Στοιχεία ταξιδιού (όπως λιμάνι αναχώρησης και έτος, μήνας, ημερομηνία, λιμάνι άφιξης και έτος, μήνας, ημερομηνία, φορτίο, περιγραφφή συμβάντων και ημερομηνία, μήνας, έτος, χαρακτηρισμός θαλασσιών περιοχών,

ονομασία ευρύτερης γεωγραφική περιοχή, Γεωγραφικό Πλάτος και Μήκος, χαρακτηρισμός ταξιδιού ως επιτυχούς ή μη.

δ) Ανθρώπινες απώλειες (όπως αγνοούμενοι, τραυματίες, νεκροί).

ε) Ζημιές πλοίου (όπως τεχνική περιγραφή, κόστος, χρόνος στο ναυπηγείο, απώλεια χρόνου, ολική απώλεια).

στ) Φορτίο πλοίου (όπως ολική, μερική ή μη αναφερθείσα απώλεια φορτίου, κόστος απολεσθέντος φορτίου).

ζ) Ρύπανση (αν προκλήθηκε ρύπανση ή όχι ή αν δεν αναφέρεται, τύπος χημικού που προκάλεσε την ρύπανση, ποσότητα χημικού).

η) Απορρύπανση (αν έγινε απορρύπανση ή όχι ή αν δεν αναφέρεται, τρόπος που έγινε, κόστος των εργασιών απορρύπανσης).

θ) Καιρικές συνθήκες (όπως ήρεμη θάλασσα, θύελλα, τυφώνας, μικρή ορατότητα, χιονοθύελλα, έλλειψη στοιχείων για τις καιρικές συνθήκες).

Σε μερικές από τις παραπάνω μεταβλητές, οι πληροφορίες που δίνονται από την αρχική πηγή είναι ελλιπείς όπως για παράδειγμα όταν αναφέρεται ρύπανση. Δεν αναφέρεται σε πολλές περιπτώσεις η ποσότητα του ρυπογόνου οπότε, για μελέτη και αυτής της μεταβλητής, πράγμα που ξεφεύγει από τους στόχους τις παρούσας ερευνητικής εργασίας, θα πρέπει να ανατρέξουμε σε άλλη πηγή..

#### 4.3 Γραμμογράφηση και κωδικοποίηση του αρχείου δεδομένων

Οι μεταβλητές που επιλέχθηκαν, σύμφωνα με τους στόχους της παρούσας εργασίας να χρησιμοποιηθούν, αναφέρονται στον πίνακα 4.3.1, επεξηγούνται τα σύμβολά τους και αναφέρονται οι παράγραφοι στις οποίες δίνονται αναλυτικές πληροφορίες για την καλύτερη κατανόηση τους.

Για να εξετασθεί η επίδραση των μεταβλητών του αρχείου στην ολική απώλεια έγινε επιλογή τους από το σύνολο σύμφωνα με το σκοπό της ερευνητικής εργασίας και θα δοθούν αναλυτικές διευκρινήσεις για όσα από τις μεταβλητές κρίθηκε σκόπιμο ότι χρειάζονται. Θα σχολιαστούν επίσης το πλήθος των τιμών των μεταβλητών καθώς και οι τιμές τους.

Το πλέον σημαντικό χαρακτηριστικό κάθε συστήματος κωδικοποίησης είναι το επίπεδο λεπτομέρειας που εμπεριέχουν οι κωδικοί του. Όσο πιο λεπτομερές είναι το σύστημα, τόσο πληρέστερη στατιστική ανάλυση μπορεί να δεχτεί, με την προϋπόθεση βέβαια ότι ο πληθυσμός που πρόκειται να αναλυθεί είναι ικανοποιητικά

μεγάλος. Θα παρουσιασθεί και θα σχολιασθεί ο τρόπος κωδικοποίησης σε μορφή συμβατή, για ευκολότερη επεξεργασία, με το στατιστικό πακέτο S.P.S.S. που θα χρησιμοποιηθεί για ανάλυσή των στοιχείων. Για όλες τις μεταβλητές οι κωδικοί

Πίνακας 4.3.1  
Συμβολισμοί και παράγραφοι των μεταβλητών

Πεδίο/Μεταβλητή	Σύμβολο	Παράγραφος
Πεδίο ταξινόμησης	AA	4.3.24
Ατύχημα	ACCIDE0	4.3.1
Ηλικία	AGE	4.3.3
Είδος περιοχής	AREATYPE	4.3.5
Έτος ναυπήγησης	BUILT1	4.3.6
Νηογνώμονας	CLASS	4.3.4
Χρονολογία ατυχήματος	DATYR	4.3.9
Χωρητικότητα (grt)	GRT1	4.3.2
Σχετικό μέγεθος	GRT1GROU	4.3.19
Πλήθος αλλαγής ιδιοκτητών	OWNER	4.3.10
Μήνας	MONTH	4.3.7
Γεωγραφική περιοχή	NAREA	4.3.8
Πλήθος ατυχημάτων	NOFACC	4.3.21
Πλήθος αλλαγής ιδιοκτητών ομαδοποιημένο	OWNERS	4.3.23
Ακολουθία εγγραφών ανά πλοίο	SEQACC	4.3.20
Τύπος του πλοίου	TYPE	4.3.11
Νηογνώμονας σε ομάδες	GRUCLASS	4.3.16
Σειρά ατυχημάτων	ACCSER	4.3.22
Μη επιτυχές ταξίδι	BADVOYAGE	4.3.12
Χρονική μεταβλητή ατυχήματος	DACC92	4.3.17
Χρονική μεταβλητή καταγραφής ατυχήματος	DAY92	4.3.18
Σημαία	FLAG2	4.3.13
Καιρικές συνθήκες	WEATH	4.3.14
Ολική απώλεια	TOTLOSS	4.3.15



αναλυτικά αναφέρονται στο Παράρτημα Κ.

Κατά την κωδικοποίηση επιλέχθηκε για τις μεταβλητές, που αναφέρονται στη συνέχεια, η ομαδοποίηση να γίνει με βάση τις ομαδοποιήσεις των Lloyd's ώστε να υπάρχει συμβατότητα για το ενδεχόμενο μελλοντικής συγκριτικής μελέτης. Επίσης, η αναφορά στις ομαδοποιήσεις των Lloyd's επιλέχθηκε να γίνει λόγω της γενικής αποδοχής των Lloyd's από την ευρύτερη Ναυτιλιακή Βιομηχανία (Λιμάνια, Ναυπηγεία κ. α.).

Οι μεταβλητές που υπήρχαν πρωτογενώς στη βάση είναι: *Ατύχημα, Χωρητικότητα, Ηλικία, Είδος περιοχής, Έτος ναυπήγησης, Νηογνώμονας, Χρονολογία ατυχήματος, Πλήθος αλλαγής ιδιοκτητών, Μήνας, Γεωγραφική περιοχή, Τύπος πλοίου, Μη επιτυχές ταξίδι, Σημαία, Καιρικές συνθήκες και Ολική απώλεια.*

#### 4.3.1 Ατύχημα

Στη μεταβλητή *Ατύχημα* περιγράφεται το είδος του ατυχήματος σαν συνέπεια του οποίου πιθανά θα οδηγηθεί το πλοίο σε ολική απώλεια. Τα ατυχήματα χωρίστηκαν σε επτά βασικές ομάδες. Ο χωρισμός έγινε με βάση τις διεθνώς αποδεκτές ομαδοποιήσεις των μηνιαίων εκδόσεων Lloyd's Register "World Casualty Statistics" με ενδεικτικές τιμές "FOUNDERED", "MISSING", "FIRE-EXPLOSION", "COLLISION", "CONTACT", "GROUNDED", "HULL /MCHY /EQUIPMENT DAMAGED" όπως φαίνεται και στο Παράρτημα Πι 4.3.1. Ο τρόπος κωδικοποίησης του ατυχήματος είναι ιεραρχικός. Αναλυτικά η ομάδα "HULL /MCHY/ EQUIPMENT DAMAGED" έχει συλλεκτικό κωδικό 8. Σε περίπτωση που υπάρχουν περισσότερες αναλυτικές πληροφορίες για την μεταβλητή *ατύχημα*, για παράδειγμα με τιμή μεταβλητής "BROKE", που έχει κωδικό 81, αυτή περιλαμβάνεται στην ομάδα "HULL /MCHY/ EQUIPMENT DAMAGED" που έχει κωδικό 8. Η αναλυτική σειρά κωδικών φαίνεται στο Παράρτημα Κ 4.3.1. Ανάλογα η υποομάδα "CAPSIZED" περιέχεται στην ομάδα "FOUNDERED". Αξιοπρόσεκτο είναι ότι υπάρχουν καταγεγραμμένα ατυχήματα τα οποία λόγω της φύσης τους, για παράδειγμα το "SEIZURE" καθώς και του απόλυτα συγκεκριμένου στόχου της παρούσας εργασίας δεν μπορούν να περιληφθούν στην τελική ανάλυση.

#### 4.3.2 Χωρητικότητα

Στη μεταβλητή χωρητικότητα καταγράφεται η μεικτή / ολική χωρητικότητα (grt) του πλοίου. Η ποσοτική αυτή μεταβλητή δίνει ένδειξη μεγέθους και διερευνάται η επίδρασή της στο ατύχημα και ειδικότερα στην ολική απώλεια.

#### 4.3.3 Ηλικία

Οι τιμές της μεταβλητής *Ηλικία* φανερώσουν την ηλικία του πλοίου κατά την στιγμή του ατυχήματος. Επιλέχθηκε και για τεχνικούς λόγους ώστε σε συνδυασμό με την μεταβλητή *Έτος ναυπήγησης*, που θα σχολιασθεί στη συνέχεια, να ελέγξει με διασταυρώσεις την αξιοπιστία των εγγραφών.

#### 4.3.4 Νηογνώμονας

Οι Νηογνώμονες ως γνωστόν, παρακολουθούν και ελέγχουν το πλοίο κατά την κατασκευή αλλά και κατά τη διάρκεια της οικονομικής ζωής του και μάλιστα εξετάζουν αν αυτό ανταποκρίνεται στην κλάση στην οποία έχει ταξινομηθεί. Στον βαθμό λοιπόν, που η ποιότητα των υπηρεσιών των φορέων επιθεώρησης επιδρά στη λειτουργία και τεχνική κατάσταση του πλοίου και συνεπώς στο ενδεχόμενο ατυχήματος επιλέχθηκε η μεταβλητή *Νηογνώμονας*. Στη μεταβλητή αυτή καταγράφηκαν πληροφορίες που δίνονται από τις πηγές για τους: "AMERICAN BUREAU OF SHIPPING", "BUREAU VERITAS", "CHINA CLASSIFICATION SOCIETY", "DET NORSKE VERITAS CLASSIFICATION A/S", "GERMANISCHER LLOYD", "KOREAN REGISTER OF SHIPPING", "LLOYD'S REGISTER OF SHIPPING", "NIPPON KAIJI KYOKAI", "POLSKI REJESTR STATKOW", "REGISTRO ITALIANO NAVALE", "RUSSIAN REGISTER OF SHIPPING", "ROMANIAN REGISTER OF SHIPPING", "HELLENIC REGISTER OF SHIPPING" και μερικούς με μικρότερη συχνότητα εμφάνισης Νηογνώμονες όπως φαίνεται και στο Παράρτημα Π1 4.3.4.

Για τους νηογνώμονες χρησιμοποιήθηκε κωδικοποίηση, με βάση τα διεθνώς αποδεκτά, αρχικά χαρακτηριστικά γράμματα, όπως δίνονται στο Παράρτημα Κ 4.3.4.

Για τον “AMERICAN BUREAU OF SHIPPING” για παράδειγμα, χρησιμοποιήθηκε ο κωδικός AB.

#### 4.3.5 *Είδος περιοχής*

Η μεταβλητή *Είδος της περιοχής* του ατυχήματος περιγράφει την περιοχή του ατυχήματος. Επιλέχθηκε για να ελεγχθεί αν έχει σχέση με το ατύχημα και μάλιστα με την ολική απώλεια, η ανοικτή, η ελεγχόμενης κυκλοφορίας θάλασσα ή το λιμάνι, (Παράρτημα Π<sub>1</sub> 4.3.5). Για την κωδικοποίηση έχουμε τους κωδικούς 1, 2, 3 αντίστοιχα με τιμές OPEN SEA, CONTROLLED, PORT, όπως ορίστηκαν στο Παράρτημα Κ 4.3.5.

#### 4.3.6 *Έτος ναυπήγησης*

Οι τιμές της μεταβλητής *Έτος ναυπήγησης* φανεράνουν ουσιαστικά την παλαιότητα του πλοίου την στιγμή του ατυχήματος. Ένα από τα βασικά ερωτήματα που τέθηκαν κατά καιρούς είναι κατά πόσο η παλαιότητα συμμετέχει σαν παράγοντας στην δημιουργία ατυχήματος και κατ’ επέκταση και στην ολική απώλεια και για να δοθεί απάντηση σε αυτό επιλέχθηκε η μεταβλητή *Έτος ναυπήγησης*.

#### 4.3.7 *Μήνας*

Με σκοπό να διερευνηθεί η επίδραση της εποχικότητας και μάλιστα του μήνα επιλέχθηκε αυτή η μεταβλητή (Παράρτημα Π<sub>1</sub> 4.3.7).

#### 4.3.8 *Γεωγραφική περιοχή*

Στη μεταβλητή *Γεωγραφική περιοχή* έχουν καταγραφεί οι περιοχές που συνέβησαν τα ατυχήματα όπως αυτές έχουν ορισθεί με βάση τις διεθνώς αποδεκτές γεωγραφικές περιοχές του άτλαντα των Lloyd’s of London Press Ltd, Maritime Atlas 1993, όπως φαίνονται και στο Παράρτημα Π<sub>1</sub> 4.3.8. Επιλέχθηκε αυτή με σκοπό να αναζητηθούν οι πιο επικίνδυνες για ολική απώλεια περιοχές. Η ομαδοποίηση προέκυψε με κριτήρια σχετικής γειτνίασης των γεωγραφικών περιοχών ενώ η



μεταβλητή παίρνει δώδεκα τιμές οι οποίες παρουσιάστηκαν αναλυτικά στο Παράρτημα Κ 4.3.8.

#### 4.3.9 Χρονολογία ατυχήματος

Η μεταβλητή που δίνει πληροφορία για την χρονολογία ατυχήματος περιλήφθηκε για να διερευνηθεί η επίδραση τις χρονιάς στο γεγονός και να εξαχθούν πιθανά συμπεράσματα (Παράρτημα Π<sub>1</sub> 4.3.9) αλλά και για να ελεγχθούν κατά την προετοιμασία τα δεδομένα. Ειδικότερα, όπως θα αναφερθεί στο επόμενο κεφάλαιο, είναι μια από τις βασικές μεταβλητές για καθορισμό, με χρήση αλγορίθμων, της μοναδικής εγγραφής ατυχήματος.

#### 4.3.10 Πλήθος αλλαγής ιδιοκτητών

Ο αριθμός των διαφορετικών ιδιοκτητών μπορεί να ποικίλει ανάλογα με το πλοίο. Για παράδειγμα σε νεότευκτο πλοίο μπορεί να είναι δύο οι ιδιοκτήτες, η μεταβλητή όμως *Πλήθος αλλαγής ιδιοκτητών* έχει τιμή 1. Σε περίπτωση πρώτης πώλησης η μεταβλητή *Πλήθος αλλαγής ιδιοκτητών* παίρνει τιμή 2 κ.ο.κ. Αυτή ακριβώς την «αλλαγή χεριών» περιγράφει η ποσοτική διακριτή μεταβλητή *Πλήθος αλλαγής ιδιοκτητών*. Επειδή γενικά ο αριθμός που δείχνει πόσες φορές «άλλαξε χέρια» ένα πλοίο είναι μια ένδειξη, σύμφωνα με εμπειρική παρατήρηση, κακοδιαχείρισης και κακής συντήρησης για το πλοίο, για αυτό επιλέχθηκε αυτή η μεταβλητή.

#### 4.3.11 Τύπος πλοίου

Η μεταβλητή αυτή μας πληροφορεί για τον τύπο των πλοίων. Έγιναν ομαδοποιήσεις για την μεταβλητή *Τύπος πλοίου*, με βάση τις ετήσιες εκδόσεις Lloyd's Register "World Casualty Statistics" (Παράρτημα Π<sub>1</sub> 4.3.11). Οι ενδεικτικές τιμές για τις βασικές ομάδες ορίστηκαν οι "BULK", "FERRY", "CONTAINER", και "TANKER". Στην ομάδα "TANKER" περιέχονται και υποομάδες όπως τα "L.N.G" ενώ στην ομάδα "FERRY", περιλήφθηκαν και όλα τα "PASSENGER". Τέλος στην ομάδα "BULK" ενσωματώθηκαν τα "BULK" και τα "GENERAL CARGO". Σημαντική είναι η παρατήρηση, η οποία εύκολα επιβεβαιώθηκε επιστημονικά, με

Πίνακα μονής εισόδου, ότι για τις τιμές “FISHING”, “PONTOON-BARGE”, και “OTHERS” έχουμε ελλιπείς πληροφορίες για τις υπόλοιπες μεταβλητές του αρχείου. Οι τιμές αυτές δεν θα συμπεριληφθούν στην τελική ανάλυση. Εξάλλου για την τιμή “FISHING” δεν έχει έννοια να γίνουν ομαδοποιήσεις για μια σειρά από διαφορετικούς λόγους. Μερικοί από αυτούς είναι ότι κάθε χώρα έχει την δική της πολιτική, όπως δική της χρηματοδότηση και δικό της κανονισμό άρα έχουμε έλλειψη ενιαίας γενικότερης πολιτικής. Πολλά από αυτά δεν είναι επίσημα καταγεγραμμένα για διάφορους λόγους, όπως μικρή χωρητικότητα. Άλλα δεν ακολουθούν τους κανόνες των Lloyd’s όπως τα μικρότερα των 500 grt. Όσον αφορά τους κωδικούς για την μεταβλητή *Τύπος πλοίου*, ο τρόπος κωδικοποίησης είναι ιεραρχικός. Ο τύπος “TANKER” έχει συλλεκτικό κωδικό 3 και σε περίπτωση που υπάρχουν περισσότερο αναλυτικές πληροφορίες παράδειγμα για τύπο “CHEM TANKER” έχουμε κωδικό 32. Οι ομαδοποιήσεις και οι κωδικοί παρουσιάστηκαν σε πλήρη ανάλυση στο Παράρτημα Κ 4.3.11.

#### 4.3.12 *Μη επιτυχές ταξίδι*

Η μεταβλητή χαρακτηρίζει αν το ταξίδι είναι επιτυχές. Μπορεί να δοθεί ο χαρακτηρισμός σαν μη επιτυχούς στην περίπτωση ατυχήματος αλλά και στην περίπτωση που έχουμε καθυστέρηση παράδοσης φορτίου ή μερικής απώλειας αυτού. Χαρακτηρίζεται από υποκειμενικότητα, οριζόμενο κάθε φορά σε σχέση με τον στόχο της ερευνητικής εργασίας. Η μεταβλητή αυτή παίρνει τιμές «ναι» και «όχι» με κωδικούς 1 και 2 αντίστοιχα.

#### 4.3.13 *Σημαία*

Κάθε εμπορικό πλοίο οφείλει να φέρει τη σημαία κάποιου κράτους εφόσον επιθυμεί να προσφέρει μεταφορικό έργο. Το πλοίο που δεν έχει εθνικότητα δεν προστατεύεται από το Διεθνές Δίκαιο και δεν γίνεται δεκτό στα λιμάνια όλου του κόσμου. Η απόκτηση εθνικότητας από ένα πλοίο, γίνεται με την εγγραφή του στα δημόσια μητρώα, νηολόγια ενός κράτους. Υπάρχουν κράτη με μεγάλες απαιτήσεις για να εγγραφούν τα πλοία στα νηολόγια τους, ενώ άλλα παρέχουν ευνοϊκούς όρους και συχνά δίνουν μικρή σημασία στην τήρηση διεθνών κανονισμών. Τέτοια κράτη είναι αυτά των σημαίων ευκαρίας (flags of convenience). Η μεταβλητή *Σημαία* μας

πληροφορεί για την σημαία που έφερε το πλοίο την στιγμή του συγκεκριμένου συμβάντος και οι κωδικοί ορίστηκαν στο Παράρτημα Κ 4.3.13.

#### 4.3.14 Καιρικές συνθήκες

Για να διερευνηθεί η σχέση καιρού και ατυχήματος επιλέχθηκε η μεταβλητή αυτή με τις ενδεικτικές τιμές "CALM WEATHER", "POOR VISIBILITY", "STORM", "SNOW STORM", και "TYPHOON" με κωδικούς από 1 έως 5 αντίστοιχα. Αναλυτικά οι τιμές παρουσιάστηκαν στο Παράρτημα Κ 4.3.14. Επειδή πολλές φορές δεν δίνονται πληροφορίες για τον καιρό, με σκοπό να χρησιμοποιηθούν όσες το δυνατόν περισσότερες πληροφορίες, έγινε αξιοποίηση και των στοιχείων που δίνονται έμμεσα για τον καιρό. Ειδικότερα όπου αναφέρθηκε ότι η ισχύς των ανέμων ήταν από 0 έως 20 Κn έγινε αντιστοίχιση σε κωδικό 1, από 20 έως 40 Κn έγινε αντιστοίχιση σε κωδικό 3 και από 40 Κn και πάνω σε κωδικό 5.

#### 4.3.15 Ολική απώλεια

Η μεταβλητή αυτή δίνει χαρακτηρισμό για το αν ένα ατύχημα κατέληξε η όχι σε ολική απώλεια και αφορά το χρονικό διάστημα από το 1992 έως το 1999 με τιμές «ναι» η «όχι» που αντιστοιχούν στους κωδικούς 1 και 2 (Παράρτημα Κ 4.3.15). Εδώ σημειώνεται ότι υπάρχει και το ενδεχόμενο ένα πλοίο να δηλώνεται σαν ολική απώλεια επειδή το κόστος επιδιόρθωσης των ζημιών ανέρχεται σε μεγαλύτερο ύψος από την συνολική αξία του (Giziakis K, 1996). Πρόβλημα δημιουργείται από την πληρότητα πληροφοριών για τις ολικές απώλειες που αναφέρονται πολλές φορές σε ύστερο χρόνο από τον χρόνο που συνέβη το ατύχημα για μια σειρά από λόγους. Αναλυτικότερα, ένα πλοίο μπορεί να μην έχει δηλωθεί σαν ολική απώλεια σε μία εγγραφή και να έχει δηλωθεί σε κάποια επόμενη εγγραφή. Για να έχουμε όσον το δυνατόν πιο ακριβείς πληροφορίες θα υποδειχθεί λύση, στο μέτρο του δυνατού, για αντιμετώπιση του προβλήματος στο επόμενο κεφάλαιο κατά την προετοιμασία των δεδομένων.

Υπάρχουν μεταβλητές που δημιουργήθηκαν συμπληρωματικά [Maletic J. I. et al, 2000] για τις ανάγκες του καθαρισμού και της ανάλυσης των πληροφοριών του



αρχείου και τα σχετικά αναφέρονται αναλυτικά στο επόμενο κεφάλαιο, ενώ άλλες δημιουργήθηκαν για λόγους χρηστικών.

Οι μεταβλητές είναι: *Πεδίο ταξινόμησης, Σχετικό μέγεθος, Πλήθος αλλαγής ιδιοκτητών ομαδοποιημένο, Πλήθος ατυχημάτων, Ακολουθία εγγραφών ανά πλοίο, Νηογνώμονες ομαδοποιημένοι, Χρονική μεταβλητή ατυχήματος, Χρονική μεταβλητή καταγραφής ατυχήματος και Σειρά ατυχήματος* και στη συνέχεια δίνονται επεξηγηματικές πληροφορίες για αυτές.

#### 4.3.16 Νηογνώμονες ομαδοποιημένοι

Ο κυριότερος εκφραστής της πολιτικής των νηογνώμωνων είναι ο Διεθνής Σύνδεσμος Νηογνώμωνων (IACS), ο οποίος εκπροσωπεί τους μεγαλύτερους νηογνώμονες παγκόσμια και μάλιστα τους "AMERICAN BUREAU OF SHIPPING", "BUREAU VERITAS", "CHINA CLASSIFICATION SOCIETY", "DET NORSKE VERITAS CLASSIFICATION A/S", "GERMANISCHER LLOYD", "KOREAN REGISTER OF SHIPPING", "LLOYD'S REGISTER OF SHIPPING", "NIPPON KAIJI KYOKAI", "POLSKI REJESTR STATKOW", "REGISTRO ITALIANO NAVALE", "RUSSIAN REGISTER OF SHIPPING". Κύριος σκοπός του είναι η βελτιστοποίηση στην κατασκευή των πλοίων ή των μηχανών τους πάντα με στόχο την επίτευξη συνεχώς υψηλότερων προτύπων ασφαλείας όσον αφορά την ασφάλεια του πλοίου, τις θαλάσσιες μεταφορές καθώς και την πρόληψη της θαλάσσιας ρύπανσης. Οι νηογνώμονες είναι ως επί το πλείστον τεχνικοί οργανισμοί αναγνωρισμένοι από όλα τα κράτη, ειδικά μάλιστα οι ανήκοντες στο Σύνδεσμο λόγω της ποιότητας των υπηρεσιών που προσφέρουν. Επιλέχθηκε λοιπόν με υποκειμενικά κριτήρια και σε σχέση με το στόχο της παρούσας διατριβής η ομαδοποίηση των νηογνώμωνων να γίνει όχι στην πιθανή σε δύο ομάδες, τους ανήκοντες στον Διεθνή Σύνδεσμο Νηογνώμωνων και στους υπόλοιπους αλλά σε τέσσερες. Οι τρεις ομάδες αποτελούμενες από τους νηογνώμονες του Συνδέσμου και η τέταρτη ομάδα αποτελούμενη από τους υπόλοιπους, όπως φαίνεται και στο Παράρτημα Π1 4.3.16. Η ομαδοποίηση σε περισσότερες από δύο ομάδες δεν είναι καθόλου δεσμευτική σε σχέση με την ομαδοποίηση σε δύο, αλλά αντίθετα είναι γενικότερη. Μάλιστα αυτού του είδους η ομαδοποίηση δεν αποκλείει το ενδεχόμενο κατά την ταξινόμηση σε κλάσεις που θα παράγει ο αλγόριθμος, ανάλογα με τα δεδομένα, να περιληφθούν

στην ίδια κλάση περισσότερες από μία ομάδες. Αναλυτικά στο Παράρτημα Κ 4.3.16 αναφέρθηκαν οι τιμές της μεταβλητής αυτής.

#### 4.3.17 Χρονική μεταβλητή ατύχηματος

Στις έντυπες πηγές από τις οποίες συλλέχθηκαν τα στοιχεία δίνεται η ημέρα, ο μήνας και το έτος κατά τις οποίες συνέβη το ατύχημα. Στην αρχική καταγραφή χρησιμοποιήθηκαν τρεις ξεχωριστές μεταβλητές. Στην παρούσα εργασία, για την ίδια καταγραφή, χρησιμοποιήθηκε η μεταβλητή *Χρονική μεταβλητή ατύχηματος*. Θεωρήθηκε σαν αρχή των χρόνων η πρώτη του Ιανουαρίου του έτους 1992. Η μεταβλητή αυτή δίνει το χρόνο σε ημέρες που πέρασε από την αρχή των χρόνων μέχρι την χρονική στιγμή που συνέβη το ατύχημα.

Επινοήθηκε δηλαδή τρόπος αντικατάστασης των τριών ξεχωριστών χρονικών μεταβλητών προσδιορισμού του ατύχηματος με μία, την *χρονική μεταβλητή ατύχηματος*. Από αυτόν τον μετασχηματισμό έχουμε σημαντικά οφέλη διότι:

- γίνεται περισσότερο λειτουργικό το αρχείο δεδομένων
- δεν έχουμε απώλεια πληροφορίας
- έχουμε την ημερομηνία εγγραφής προσδιορισμένη ως ένα συνεχές μέγεθος για την τελική επεξεργασία

Επιλέχθηκε η μεταβλητή *Χρονική μεταβλητή ατύχηματος* για τεχνικούς λόγους και θα χρησιμοποιηθεί κυρίως στην διαδικασία προετοιμασίας των δεδομένων και ειδικά στην διαδικασία καθαρισμού του αρχείου (data cleaning).

#### 4.3.18 Χρονική μεταβλητή καταγραφής ατύχηματος

Κατ'αναλογία ορίζουμε την μεταβλητή *Χρονική μεταβλητή καταγραφής ατύχηματος* που δίνει το χρόνο σε ημέρες από την αρχή των χρόνων, την πρώτη Ιανουαρίου του έτους 1992 μέχρι την χρονική στιγμή που καταγράφηκε στις έντυπες πηγές το ατύχημα σε αντικατάσταση των τριών ξεχωριστών χρονικών μεταβλητών. Για τους ίδιους τεχνικούς λόγους και με τα ίδια οφέλη, που προαναφέρθηκαν για την *Χρονική μεταβλητή ατύχηματος*, επιλέξαμε σαν ένα συνεχές μέγεθος και την *Χρονική μεταβλητή καταγραφής ατύχηματος*.

#### 4.3.19 Σχετικό μέγεθος

Από εμπειρική παρατήρηση της αγοράς και με το σκεπτικό ότι σε κάθε τύπου πλοίου διαφορετικό είναι το «μικρό» μέγεθος, η ταξινόμηση των πλοίων σε «μικρά», «μεσαία», «μεγάλα» κ.ο.κ., θα πρέπει να παίρνει υπόψη τους και τα απόλυτα μεγέθη αλλά και την ειδική κατηγορία στην οποία ανήκουν τα συγκεκριμένα πλοία. Η μεταβλητή *Σχετικό μέγεθος* περιγράφει την χωρητικότητα του πλοίου σε σχέση με τον τύπο του πλοίου είναι δηλαδή μεταβλητή που προέκυψε από σύνθεση της χωρητικότητας με τον τύπο του πλοίου. Το σχετικό μέγεθος ομαδοποιείται σε κλάσεις άνισου πλάτους. Αναλυτικά, για κάθε ένα ενδεικτικό τύπο πλοίου υπολογίστηκε το εύρος των παρατηρήσεων και ο αριθμός των κλάσεων επιλέχθηκε να είναι επτά. Στη συνέχεια ορίστηκαν κλάσεις άνισου πλάτους που περιέχουν ίσο αριθμό παρατηρήσεων ώστε να υπάρχει αξιόπιστος αριθμός παρατηρήσεων σε κάθε κλάση. (Παράρτημα Π1 4.3.19).

#### 4.3.20 Ακολουθία εγγραφών ανά πλοίο

Η μεταβλητή *Ακολουθία εγγραφών ανά πλοίο* περιγράφει πόσες εγγραφές για κάθε πλοίο υπάρχουν στο συνολικό αρχείο. Αναλυτικότερα για ένα πλοίο μπορεί να έχουμε μόνο μια εγγραφή ατυχήματος ενώ σε άλλο πλοίο μπορεί να έχουμε τρεις εγγραφές διαφορετική χρονική στιγμή (Παράρτημα Π1 4.3.20). Για να μετρήσουμε λοιπόν πόσες φορές το ίδιο πλοίο έχει ατύχημα, που έχει καταγραφεί, χρησιμοποιήθηκε η μεταβλητή αυτή. Το πλήθος των τιμών της είναι εννιά, όπως φαίνεται και στο Παράρτημα Κ 4.3.20, και για παράδειγμα ο κωδικός 2 αντιστοιχεί σε δύο εγγραφές ατυχημάτων για το ίδιο πλοίο.

#### 4.3.21 Πλήθος ατυχημάτων

Κάθε μία εγγραφή είναι ένα ξεχωριστό συμβάν το οποίο μπορεί να περιλαμβάνει από ένα έως το πολύ πέντε διαφορετικά ατυχήματα. Η μεταβλητή *Πλήθος ατυχημάτων* δίνει τον αριθμό αυτών των διαδοχικών διαφορετικών ατυχημάτων και παίρνει τιμές από 1 έως 5 όπως φαίνεται και στο Παράρτημα Π1 4.3.21.



#### 4.3.22 Σειρά ατυχήματος

Σε μία εγγραφή όπως αναφέρθηκε μπορεί να έχουν συμβεί από ένα έως το πολύ πέντε διαφορετικά ατυχήματα. Η μεταβλητή *Σειρά ατυχήματος* περιγράφει αν το ατύχημα που μελετάμε ήταν στην ακολουθία των ατυχημάτων, για παράδειγμα, πρώτο η τρίτο. Αυτή παίρνει τιμές από 1 έως 5 όπως φαίνεται και στο Παράρτημα Π<sub>1</sub> 4.3.22.

#### 4.3.23 Πλήθος αλλαγής ιδιοκτητών ομαδοποιημένο

Μέσα από διαδικασία βασισμένη σε αυστηρά ποσοτικά κριτήρια, όπως προαναφέρθηκε στη παράγραφο 4.3.10, ορίστηκε η μεταβλητή *Πλήθος αλλαγής ιδιοκτητών*. Στη συνέχεια για λόγους χρηστικούς η ποσοτική αυτή μεταβλητή ταξινομήθηκε σε πέντε διακριτές τάξεις που χαρακτηρίστηκαν με κωδικούς με αριθμούς από 1 έως και 5 δημιουργώντας μια ποιοτική κλίμακα. Η μεταβλητή *Πλήθος αλλαγής ιδιοκτητών ομαδοποιημένο* καταγράφει σε ποιοτική κλίμακα τον αριθμό αλλαγής ιδιοκτητών μέχρι και το συγκεκριμένο συμβάν, όπως φαίνεται και στο Παράρτημα Π<sub>1</sub> 4.3.23.

#### 4.3.24 Πεδίο ταξινόμησης

Με το *Πεδίο ταξινόμησης* έχουμε την δυνατότητα να εκτελούμε μετασχηματισμούς και να διατηρούμε, όταν κρίνεται απαραίτητο, σταθερή την επιθυμητή σειρά εγγραφής. Η επιλογή αυτή της μεταβλητής κρίθηκε απαραίτητη για πραγματοποίηση των διαφόρων μετασχηματισμών στο αρχείο.

Πρέπει να σημειωθεί εδώ, ότι το πλήθος των τιμών της μεταβλητής αυτής, με τιμές σε αύξουσα κλίμακα των φυσικών αριθμών, ήταν ίσο με το πλήθος εγγραφών του αρχικού αρχείου. Μετά την συνολική προετοιμασία των δεδομένων το πλήθος των τιμών της μεταβλητής αυτής περιορίστηκε σημαντικά.

#### 4.4 Εγγραφές και στοιχεία του αρχείου δεδομένων

Το αρχείο δεδομένων που δημιουργήθηκε από τις αρχικές πηγές για τα

ατυχήματα που συνέβησαν από τους τελευταίους μήνες του 1992 έως και το 1999, περιλαμβάνει 38536 εγγραφές.

Στην αρχική μορφή του αρχείου τα ατυχήματα μπορεί να αναφέρονται:

- Σε διαφορετικές εγγραφές και διαφορετικά πλοία
- Σε διαφορετικές εγγραφές και ίδια πλοία και να περιγράφεται το ίδιο η διαφορετικό ατύχημα, οπότε στην περίπτωση αυτή έχουμε τις εξής υποπεριπτώσεις:
  - να υπάρχει αναφορά στο ίδιο πλοίο για το ίδιο ατύχημα οπότε στην περίπτωση αυτή υπάρχει πρόβλημα διπλοεγγραφών και απαιτείται συμπλήρωση η διόρθωση αυτών ώστε να ορισθεί η μοναδική εγγραφή
  - να υπάρχει αναφορά στο ίδιο πλοίο για διαφορετικά ατυχήματα οπότε στην περίπτωση αυτή δεν υπάρχει αντίστοιχο πρόβλημα.

Για το σύνολο όμως των εγγραφών, η επλεγείσα μεθοδολογία δίνει λύσεις. Ακολουθείται διαδικασία καθαρισμού ελέγχου και προετοιμασίας των δεδομένων για την τελική ανάλυσή τους. Αυτή περιγράφεται στο κεφάλαιο που ακολουθεί και στο οποίο εφαρμόζεται η τεχνική Όρυξης δεδομένων στο αρχικό αρχείο.

## ΚΕΦΑΛΑΙΟ 5

### 5.1. Εισαγωγή

Σημαντικό στοιχείο σε αυτό το κεφάλαιο είναι ο προσδιορισμός της μοναδικής εγγραφής του αρχείου που γίνεται στο υποκεφάλαιο 5.2. Ο καθαρισμός του αρχείου δεδομένων (data cleaning) θεωρήθηκε απαραίτητο στάδιο από τον τρόπο δημιουργίας του και σχηματισμού του. Στο υποκεφάλαιο αυτό περιγράφεται η βηματική εργασία καθαρισμού του αρχείου [Vijayshankar R. et al, 2001] και [Maletic J. I. et al, 2000] με σκοπό δημιουργία μοναδικής εγγραφής του ατυχήματος πλοίου. Θα περιγραφεί ο καθαρισμός των δεδομένων του αρχείου, χωρίς αυτό να είναι αυτοσκοπός.

Στο υποκεφάλαιο 5.3 περιγράφεται η επιλογή παραμέτρων για την εφαρμογή του αλγόριθμου C5.0, καθορίζονται οι παράμετροι που επίσης χρησιμοποιήθηκαν για την εφαρμογή των λοιπών τεχνικών καθώς και περιγράφονται οι μεταβλητές που θα χρησιμοποιηθούν για τα μετέπειτα στάδια ανάλυσης του αρχείου.

### 5.2. Διαδικασία καθαρισμού του αρχείου δεδομένων

Είναι απαραίτητος ο καθαρισμός του αρχείου ώστε τα συμπεράσματα που θα προκύψουν από την ανάλυση να είναι έγκυρα και αξιόπιστα. Όπως προαναφέρθηκε, αυτός δεν είναι αυτοσκοπός της παρούσας εργασίας παρόλο που μπορεί να χρησιμοποιηθεί στο μέλλον για αναζήτηση πληθώρας σχέσεων, έτσι ένας δευτερεύον στόχος του καθαρισμού είναι να εξαχθούν ορισμένα χρήσιμα συμπεράσματα για την κατασκευή παρόμοιων βάσεων στο μέλλον.

Ο καθαρισμός αρχείου δεδομένων δεν έχει ακόμα εξελιχθεί σε τυποποιημένη εργασία για να υπάρχει κάποια σχετική μεθοδολογία. Το γεγονός αυτό συμβαίνει για αρκετούς λόγους οι κυριότεροι των οποίων είναι :

- Ο συγκεκριμένος σχεδιασμός του αρχείου.
- Ο τρόπος εισαγωγής των δεδομένων.
- Οι λεπτομέρειες που αφορούν το προς εξέταση συγκεκριμένο πρόβλημα.

Αυτό που είναι κοινά αποδεκτό είναι οι γενικές κατευθύνσεις οι οποίες πρέπει να λαμβάνονται υπόψη σε κάθε διαδικασία καθαρισμού. Η εφαρμογή των



συγκεκριμένων κανόνων, των ιδιαιτεροτήτων καθώς και των προβλημάτων που παρατηρήθηκαν κατά την διάρκειά του, αποτελεί άλλωστε μία από τις σημαντικές συνεισφορές αυτής της ερευνητικής εργασίας.

Πολλοί ειδικοί, όπως ο Aaron Zornes της Meta Group, εκτιμούν ότι επί της συνολικής μεθοδολογίας Όρυξης δεδομένων, το 60% έως 80% αφορά συνήθως την προετοιμασία των δεδομένων.

Η όλη διαδικασία καθαρισμού του παρόντος αρχείου δεδομένων βασίζεται στην ημερολογιακή κατασκευή του αρχείου και τη δυναμική του φύση. Κάθε νεότερη εγγραφή θεωρείται ότι διορθώνει, συμπληρώνει ή ανασκευάζει την προηγούμενη πληροφορία. Όταν μία εγγραφή έχει πολλές ελλείπουσες τιμές τότε πιθανά είναι ανανέωση μίας προηγούμενης. Υπάρχουν και περιπτώσεις που η νεώτερη χρονικά εγγραφή έχει λιγότερες τιμές από την επόμενη και είναι περιπτώσεις οι οποίες αφορούσαν εγγραφή ατυχήματος με λίγα στοιχεία τα οποία συμπληρώθηκαν αργότερα. Τα παραπάνω δεν μπορούν να θεωρηθούν ότι έχουν ισχύ στο σύνολο των εγγραφών του αρχείου δεδομένων αφού για συγκυριακούς ή τυχαίους λόγους υπάρχουν κατά περίπτωση αποκλίσεις από τον παραπάνω κανόνα.

#### 5.2.1. Το Αρχείο δεδομένων

Σαν πρώτο στάδιο εκτελούνται μετατροπές στο αρχείο δεδομένων έτσι ώστε να μπορούν να εκτελεστούν οποιοδήποτε μετασχηματισμοί. Αναλυτικά:

- Ενώνονται όλες οι διαθέσιμες χρονιές σε ένα ενιαίο αρχείο.
- Εξαιρούνται οι εγγραφές οι οποίες αντιστοιχούν σε μη πραγματικά ατυχήματα.
- Δημιουργείται πεδίο ταξινόμησης του αρχείου έτσι ώστε σε ύστερη φάση να μπορούν να ανασκευαστούν όσες χρήσιμες πληροφορίες είναι απαραίτητες.
- Εισάγονται ετικέτες στο αρχείο έτσι ώστε να ανιχνεύεται εύκολα η φύση των τιμών των μεταβλητών.

#### 5.2.2. Καθορισμός ομάδων μεταβλητών

Στο στάδιο αυτό καθορίζονται τα σύνολα / ομάδες των μεταβλητών και η σχέση τους με τα ερευνητικά ζητήματα της εργασίας.

Χωρίζονται οι μεταβλητές στις ακόλουθες ομάδες:

- Ομάδα Α : Οι μεταβλητές προσδιορισμού πλοίου,
- Ομάδα Β : Οι χρονικές μεταβλητές,
- Ομάδα Γ : Οι μεταβλητές μοναδικότητας ατυχήματος,
- Ομάδα Δ : Οι προς διόρθωση μεταβλητές,
- Ομάδα Ε : Οι υπόλοιπες μεταβλητές του αρχείου.

### 5.2.3. Ορισμός νέων μεταβλητών για τον καθαρισμό του αρχείου δεδομένων

Μία σειρά νέων σύνθετων μεταβλητών κρίθηκαν ως απαραίτητες για το καθαρισμό [Hernandez M. et al, 1997] του αρχείου δεδομένων πιο συγκεκριμένα:

Μεταβλητές που προκύπτουν ως μετασχηματισμοί ενός ή περισσότερων μεταβλητών του αρχείου και αυτές είναι:

- Χρονική μεταβλητή καταγραφής ατυχήματος προσδιορισμένη ως ένα συνεχές μέγεθος.
- Χρονική μεταβλητή ατυχήματος προσδιορισμένη ως ένα συνεχές μέγεθος.
- Μοναδική εγγραφή πλοίου.
- Μοναδική εγγραφή ατυχήματος.

Καθώς, όπως προαναφέρθηκε, και το Πεδίο ταξινόμησης του αρχείου δεδομένων.

Μεταβλητές των οποίων οι τιμές τους εξαρτώνται με κάποιες συνθήκες από τιμές άλλων εγγραφών και αυτές είναι:

- Φίλτρο μοναδικότητας της εγγραφής,
- Πεδίο ανίχνευσης ποιότητας της εγγραφής,
- Πεδίο αλλαγής κάποιας μεταβλητής της εγγραφής.

### 5.2.4. Έλεγχος των μεταβλητών μονοδιάστατα

Στο στάδιο αυτό εφαρμόζεται ο έλεγχος κάθε μιας μεταβλητής χωριστά. Κατασκευάζεται οι πίνακες τιμών της μεταβλητής και στο παράρτημα παρουσιάζονται οι Πίνακες συχνότητας για κάθε μία από τις εξεταζόμενες μεταβλητές. Με βάση αυτούς τους πίνακες γίνονται οι παρακάτω έλεγχοι:

- Ελέγχονται οι τιμές των μεταβλητών ώστε να βρίσκονται στο σωστό εύρος του πεδίου τιμών τους.

- Διορθώνονται τιμές μετά από έλεγχο για λογικά σφάλματα και συμπληρώνονται ελλείπουσες ετικέτες. Για παράδειγμα, το 19997 μπορεί να θεωρηθεί κατ' αρχήν έγκυρη τιμή αλλά όταν θεωρηθεί τιμή της μεταβλητής «Χρονολογία ατυχήματος» με όρια από 1992 έως 1999 τότε θα πρέπει να έχει δείκτη διόρθωσης.

- Ελέγχονται προσεκτικά οι ακραίες τιμές της κατανομής
- Ελέγχεται συνολικά, με βάση κάποια πρότερα στοιχεία που είναι σε γνώση του ερευνητή, ο Πίνακας σχετικών συχνοτήτων της κατανομής. Αυτό αποτελεί σημείο εξαιρετικής σημασίας αφού πρόκειται ουσιαστικά για τις ιδιότητες του πληθυσμού πάνω στον οποίο θα παραχθούν τα όποια στατιστικά συμπεράσματα.

#### 5.2.5. Έλεγχος των μεταβλητών με διασταυρώσεις

Άλλο στάδιο στην διαδικασία καθαρισμού είναι να εξεταστεί κατά πόσο έχουν παρατηρηθεί μη λογικά αποτελέσματα όταν γίνονται συσχετίσεις μεταξύ δύο διαφορετικών μεταβλητών.

Έγιναν έλεγχοι που αντιστοιχούν στις ακόλουθες διασταυρώσεις τιμών μεταβλητών:

- Ηλικίας και έτους ναυπήγησης πλοίου.
- Του είδους πλοίου και της αντίστοιχης χωρητικότητας.
- Της ημερομηνίας ατυχήματος και της ημερομηνίας εγγραφής. Πρέπει η ημερομηνία ατυχήματος να είναι μικρότερη της ημερομηνίας της εγγραφής.
- Οι τιμές των ατυχημάτων, ώστε να μην εμφανίζονται σαν αντιφατικές μεταξύ τους.

#### 5.2.6. Επιπλέον έλεγχοι εγγραφών

Στο στάδιο αυτό γίνονται αρκετοί έλεγχοι με την παρέμβαση του ερευνητή με σκοπό την ανίχνευση προβληματικών καταστάσεων οι οποίοι είναι αδύνατο να ελεγχθούν αλγοριθμικά.

Οι έλεγχοι αυτοί είναι των εξής μορφών:

- Ο λογικός έλεγχος των εγγραφών με εποπτεία του αρχείου δεδομένων,



- Ο δειγματοληπτικός έλεγχος της ορθότητας των στοιχείων συγκρίνοντας τα ηλεκτρονικά στοιχεία με τα πρωτότυπα.

### 5.2.7. Γενικές διαδικασίες για ορισμό της μοναδικής εγγραφής του πλοίου

Κεντρικό σημείο του καθαρισμού του αρχείου δεδομένων είναι ο καθορισμός της μοναδικής εγγραφής του πλοίου. Με βάση τον μοναδικό καθορισμό του πλοίου θα συνεχίζεται η εξέταση για κάθε μία από τις μεταβλητές των υπόλοιπων ομάδων. Για το σχηματισμό της μοναδικής τιμής τους πλοίου λήφθηκαν υπόψη οι βασικές αρχές κατασκευής και δόμησης του αρχείου δεδομένων.

Η διαδικασία εύρεσης της μοναδικής τιμής προσδιορισμού του πλοίου είναι μία αναδρομική ακολουθία αφού ουσιαστικά η αλλαγή μίας τιμής δημιουργεί νέα δεδομένα για τα οποία θα πρέπει να επαναληφθεί η ίδια επεξεργασία.

Οι γενικοί κανόνες που εφαρμόστηκαν για τον προσδιορισμό της μοναδικής εγγραφής αναφέρονται συνοπτικά παρακάτω:

- Σε περίπτωση που ορισμένες μεταβλητές έχουν ελλείπουσες τιμές και οι υπόλοιπες μεταβλητές έχουν ίδιες τιμές, τότε αντικαθίστανται οι ελλείπουσες τιμές ώστε να αναφερόμαστε στο ίδιο πλοίο.
- Σε περίπτωση που 3 από τις 4 μεταβλητές έχουν ίδιες τιμές και οι άλλες διαφέρουν θεωρείται ότι έχουμε το ίδιο πλοίο και διορθώνεται η «μη σύμφωνη τιμή».
- Δημιουργείται ξανά η μοναδική τιμή που προσδιορίζει το πλοίο.
- Η διαδικασία εξελίσσεται αναδρομικά έως ότου αποκατασταθούν στο ορθόν όλες οι προς διόρθωση ή οι προς συμπλήρωση τιμές.

Σε κάθε περίπτωση λαμβάνεται υπόψη ο χρόνος στον οποίο γίνεται η εγγραφή καθώς και ο χρόνος στον οποίο συμβαίνει το ατύχημα.

Προφανώς, σε κάθε στάδιο στο οποίο γίνεται αλλαγή των τιμών του αρχείου δεδομένων έχει προηγηθεί και έλεγχος των τιμών που θα αντικατασταθούν από τον ερευνητή, αν δεν είναι προφανής η ορθότητα της αντικατάστασης.

### 5.2.8. Αντικατάσταση των ελλειπουσών ( missing) τιμών

Από το όνομα πλοίου αφαιρέθηκαν όλοι οι μη λεκτικοί χαρακτήρες όπως παύλες και κενά γιατί, χωρίς να αλλάζουν το όνομα, αυξάνουν τη πιθανότητα να έχει

γίνει λάθος στη εισαγωγή των δεδομένων. Για τον ίδιο λόγο όλα τα γράμματα μετατράπηκαν σε κεφαλαία. Στη συνέχεια το όνομα του πλοίου επανακωδικοποιήθηκε σε αριθμητική μεταβλητή με ένα προς ένα αντιστοιχία. Τούτο έγινε για να είναι εύκολη η επεξεργασία και να λειτουργεί ο αλγόριθμος με μεγαλύτερη ταχύτητα και αξιοπιστία.

Για την ακριβή ανίχνευση της μοναδικής ταυτότητας του πλοίου μετασηματίστηκαν οι ελλείπουσες τιμές των βασικών προσδιοριστικών μεταβλητών, όνομα (name), έτος ναυπήγησης (built1), η χωρητικότητα (grt1) και είδος πλοίου (type), σε τιμή «-1».

Η διαδικασία αντικατάστασης λειτουργεί με το κανόνα  $\frac{3}{4}$ . Αφού ταξινομούνται τα δεδομένα με βάση και τις τέσσερις βασικές μεταβλητές εξετάζεται κατά πόσο δύο εγγραφές έχουν στην τέταρτη μεταβλητή τιμή «-1» για την εγγραφή j και στην εγγραφή j+1 τιμή μεγαλύτερη του μηδενός. Σε αυτή την περίπτωση η τιμή «-1» αντικαθίσταται από την αντίστοιχη αριθμητική της. Στον Πίνακα 5.1 που ακολουθεί παρουσιάζεται μία τέτοια περίπτωση.

ΠΙΝΑΚΑΣ 5.1

Ενδεικτικός πίνακας εύρεσης και αντικατάστασης ελλειπουσών τιμών

Αα	name	built1	grt1	type
12134	1267	1980	4678	-1
12367	1267	1980	4678	12

Αφού έχει γίνει η ταξινόμηση κατά αύξουσα σειρά, η «-1» θα είναι πρότερη της πραγματικής τιμής και η «-1» θα γίνει 12. Η διαδικασία συνεχίζεται για όσες τιμές «-1» υπάρχουν.

Το επόμενο βήμα είναι να παραχθεί μία νέα ταξινόμηση και να ακολουθηθεί ξανά η ίδια διαδικασία.

Στη συνέχεια η ίδια διαδικασία ακολουθείται και όταν υπάρχουν δύο ελλείπουσες τιμές από τις τέσσερες διαφορετικές μεταβλητές αρκεί να επιλέγουμε τρεις κατάλληλες μεταβλητές. Η μεταβλητή του ονόματος όμως είναι απαραίτητο να είναι συμπληρωμένη με έλεγχο των λοιπών τιμών σε όσο μεγαλύτερο αριθμό εγγραφών από τον ερευνητή.

Τέλος σε περίπτωση που έχουν μείνει ακόμα ελλείπουσες τιμές σε κάποια μεταβλητή για κάποια εγγραφή δημιουργείται μια νέα μεταβλητή στο αρχείο δεδομένων το οποίο δηλώνει ότι η συγκεκριμένη εγγραφή είναι ύποπτη για σφάλματα και αρχικά να μην χρησιμοποιηθεί στο στάδιο της μοντελοποίησης.

### 5.2.9. Αντικατάσταση των διαφορετικών τιμών

Στο στάδιο αυτό ακολουθείται η ίδια διαδικασία με το προηγούμενο υποκεφάλαιο με την προσθήκη όμως της εποπτείας του ερευνητή στην αντικατάσταση των διαφορετικών τιμών, αφού ταξινομηθούν κατά αύξουσα σειρά ως προς την προς έλεγχο μεταβλητή.

Βήμα 1. Το αρχείο δεδομένων ταξινομείται με βάση τις τέσσερις βασικές μεταβλητές.

Βήμα 2. Εφόσον υπάρχουν περισσότερες από δύο εγγραφές με τις ίδιες τιμές στις τρεις πρώτες μεταβλητές ακολουθείται το βήμα 3 αλλιώς ακολουθείται το βήμα 6.

Βήμα 3. Ελέγχεται η εγγραφή  $j$  με την εγγραφή  $j+1$  με βάση τον ακόλουθα κανόνα. Η τιμή στη μεταβλητή  $x_{kj}$  για την εγγραφή  $j$  γίνεται ίση με την  $x_{k,j+1}$  εφόσον η  $x_{k,j+1}$  δεν έχει ελλείπουσα τιμή.

Βήμα 4. Το βήμα 3 συνεχίζεται για το σύνολο των εγγραφών που έχουν ίδιες τιμές στις τρεις πρώτες μεταβλητές.

Βήμα 5. Τα δεδομένα ταξινομούνται με το ίδιο τρόπο και ξαναγίνεται η ίδια διαδικασία.

Βήμα 6. Όταν δεν παράγονται πια αλλαγές στα δεδομένα του αρχείου δημιουργείται μία νέα ταξινόμηση.

Ο αλγόριθμος σταματάει όταν έχουν τελειώσει όλες οι ταξινομήσεις. Στον Πίνακα 5.2 που ακολουθεί παρουσιάζεται μία τέτοια περίπτωση όπου ο κωδικός 3 θα αντικατασταθεί με τον πληρέστερη πληροφορία κωδικό 34.



## ΠΙΝΑΚΑΣ 5.2

Ενδεικτικός πίνακας εύρεσης και αποκατάστασης στη σωστή τιμή μιας μεταβλητής

Aa	name	built1	grt1	type
22134	12451	1990	64678	3
22367	12451	1990	64678	34
22380	12451	1990	64678	34

Σε περίπτωση που δύο στις τέσσερες μεταβλητές είναι ίδιες εξετάζονται με επιτόπιο έλεγχο οι τιμές και με βάση τους κωδικούς λαμβάνεται απόφαση για την ορθή κατηγοριοποίηση του πλοίου.

Αν τρεις από τις τέσσερες μεταβλητές διαφέρουν αλλά η τέταρτη είναι το όνομα του πλοίου πάλι ελέγχεται η ταυτότητα του πλοίου αφού είναι δυνατόν οι διαφορές να οφείλονται σε συγκεντρωτικούς κωδικούς, σε λάθη εισαγωγής στοιχείων ή και σε ελλιπή πληροφορία από την αρχική πηγή από την οποία δημιουργήθηκε το αρχείο δεδομένων.

Σε περίπτωση που διαπιστωθεί συνωνυμία που αντιστοιχεί σε διαφορετικά πλοία, μετά από έλεγχο, ο κωδικός προσδιορισμού του ονόματος του πλοίου κωδικοποιείται ξανά έτσι ώστε να αποδίδεται μοναδικά.

Τελικά το αποτέλεσμα όλης αυτής της διαδικασίας είναι να κατασκευαστεί μια νέα μεταβλητή μοναδικού προσδιορισμού του πλοίου (ship) από την μεταβλητή όνομα (name), έτος ναυπήγησης (built1), χωρητικότητα (grt1) και είδος πλοίου (type) με βάση το οποίο μπορεί πλέον να ανιχνευτούν οι προς εξέταση υποθέσεις.

### 5.2.10. Μοναδική εγγραφή του Ατυχήματος

Στο στάδιο αυτό καθορίζονται πλήρως τα ατυχήματα για κάθε πλοίο αφού βασική εγγραφή του αρχείου δεδομένων είναι το ατύχημα.

Το αρχείο δεδομένων ταξινομείται με βάση το πλοίο (ship) την ημερομηνία εγγραφής (day92), την ημερομηνία ατυχήματος (dacc92), και την περιοχή (narea) όπου έχει συμβεί το συμβάν.

Αρχική συνθήκη που εφαρμόστηκε είναι ότι για κάθε πλοίο με ίδιο κωδικό ονόματος το οποίο έχει πάθει ατύχημα μέσα στην ίδια γεωγραφική περιοχή σε μικρό χρονικό διάστημα θα αναφερόμαστε στο ίδιο ατύχημα.

Σε περιπτώσεις ελλειπουσών τιμών καθώς και περιπτώσεις διαφορετικών τιμών ακολουθήθηκαν διαδικασίες συμπλήρωσης και αντικατάστασης ανάλογες με αυτές που περιγράφηκαν με λεπτομέρειες στα υποκεφάλαια 5.2.8 και 5.2.9.

Αρχικά τα πλοία με ίδιες τις τέσσερις βασικές μεταβλητές μπορεί να είχαν:

- Το ίδιο ατύχημα την ίδια χρονική στιγμή στην ίδια γεωγραφική περιοχή.
- Το ίδιο ατύχημα διαφορετική χρονική στιγμή στην ίδια γεωγραφική περιοχή.
- Διαφορετικό ατύχημα διαφορετική χρονική στιγμή στην ίδια γεωγραφική περιοχή.
- Διαφορετικό ατύχημα διαφορετική χρονική στιγμή στην διαφορετική γεωγραφική περιοχή κ.ο.κ.

Για να εξεταστούν διπλοεγγραφές για ίδια ατυχήματα αναγκαίο είναι να κατασκευαστεί μια νέα μεταβλητή (samea) η οποία λαμβάνει τις εξής τιμές :

Av samea = 1 πρόκειται για την πρώτη ημερολογιακά εγγραφή.

Av samea = 2 υπάρχει δεύτερη εγγραφή με νέα στοιχεία τα οποία θα αντικαταστήσουν ή θα διορθώσουν αυτά της προηγούμενης εγγραφής.

Av samea = 3 υπάρχει και τρίτη εγγραφή για το ίδιο ατύχημα. κ.ο.κ.

Η απαλοιφή των διπλών εγγραφών ακολουθεί την παρακάτω περιγραφόμενη λογική σειρά με χρήση της μεταβλητής samea:

Εφόσον έχει διορθωθεί η εγγραφή τάξης  $j$  για την μεταβλητή samea = 1 με την τάξης  $j+1$  εγγραφή για την μεταβλητή samea = 2, απαλείφεται η εγγραφή τάξης  $j+1$  και ο κωδικός της μεταβλητής samea = 3 γίνεται samea = 2 ενώ ο κωδικός της μεταβλητής samea = 2 γίνεται samea = 1 έτσι ώστε να συνεχιστεί η διαδικασία καθαρισμού.

Αποτέλεσμα των ανωτέρω :

- Σε περίπτωση που ένα ατύχημα έχει καταχωρηθεί σε δύο διαφορετικές εγγραφές συμπληρώνονται οι εγγραφές του ατυχήματος στη μοναδική εγγραφή.
- Σε περίπτωση που, το με νεώτερη χρονικά καταγραφή ατύχημα είναι πιο λεπτομερής κωδικοποίηση, του με παλαιότερη καταγραφή ατυχήματος, τότε κρατάμε το με πληρέστερη κωδικοποίηση ατύχημα.

- Για κάθε μοναδική εγγραφή του ατυχήματος, και άρα του αρχείου δημιουργείται το πλήρες σύνολο των ατυχημάτων που έχουν συμβεί.
- Σε περίπτωση που το νέο ατύχημα δεν μπορεί να συμβεί ταυτόχρονα με το παλαιότερο τότε εξετάζεται αν έχουν συμβεί δύο διαφορετικά ατυχήματα στο ίδιο πλοίο, δηλαδή ισχύει ορθά η διαφοροποίηση των εγγραφών.

Ο ανωτέρω γενικός κανόνας υπάρχει περίπτωση να μην ισχύει σε ορισμένες μεμονωμένες περιπτώσεις για αυτό και γίνεται έλεγχος από τον ερευνητή σε περίπτωση που σε κάποιο πλοίο έχουν συμβεί περισσότερα από ένα ατυχήματα.

Το αποτέλεσμα όλης αυτής της διαδικασίας που περιγράφηκε είναι να κατασκευαστούν εγγραφές που θα αφορούν διαφορετικά και ξεχωριστά ατυχήματα.

Τελικά από τον συνδυασμό των παραπάνω κλειδιών και διαδικασιών που περιγράφονται στα υποκεφάλαια 5.2.7 έως και 5.2.10 δημιουργείται η μοναδική εγγραφή του ατυχήματος η οποία ενώνει το πλοίο και το ατύχημα.

#### 5.2.11. Διόρθωση των υπολοίπων μεταβλητών

Εφόσον έχει προσδιοριστεί μοναδικά το πλοίο και το ατύχημα η διόρθωση των υπολοίπων στοιχείων των μεταβλητών ακολουθεί την πεπατημένη διαδικασία.

Βήμα 1. Το αρχείο δεδομένων ταξινομείται πρώτα κατά πλοίο και ύστερα χρονικά.

Βήμα 2. Εφόσον υπάρχουν περισσότερες από δύο εγγραφές με το ίδιο ατύχημα ακολουθεί το βήμα 3 αλλιώς ακολουθεί το βήμα 7.

Βήμα 3. Ελέγχεται η εγγραφή τάξης  $j$  με την εγγραφή τάξης  $j+1$  με βάση τον ακόλουθο κανόνα: Η τιμή της μεταβλητής  $x_{kj}$  για την εγγραφή τάξης  $j$  γίνεται ίση με την  $x_{k,j+1}$  εφόσον η  $x_{k,j+1}$  δεν έχει ελλείπουσα τιμή.

Βήμα 4. Το βήμα 3 συνεχίζεται για όλο το σύνολο των προς διόρθωση μεταβλητών της ομάδας  $\Delta$ .

Βήμα 5. Διαμορφώνεται το πεδίο ποιότητας για κάθε μία εγγραφή.



Βήμα 6. Αφού έχει χρησιμοποιηθεί όλη η πληροφορία της εγγραφής, η εγγραφή τάξης  $j+1$  αφαιρείται από το αρχείο δεδομένων και στη θέση της εισάγεται η επόμενη εγγραφή τάξης  $j+2$  η οποία εξαιτίας της διαγραφής της εγγραφής τάξης  $j+2$  γίνεται εγγραφή τάξης  $j+1$ .

Βήμα 7. Ελέγχεται εκ νέου η εγγραφή τάξης  $j$  με την εγγραφή τάξης  $j+1$ .

### 5.2.12. Βαθμός αξιοπιστίας εγγραφής

Κάθε μία εγγραφή του τελικού αρχείου μπορεί να χαρακτηριστεί με βάση το βαθμό πληρότητας των εγγραφών της.

Μία εγγραφή μπορεί να έχει ελλείπουσες τιμές ή να είναι πλήρης. Εφόσον είναι πλήρης, μπορεί να αναφέρεται σε ένα ή σε περισσότερα από ένα ατυχήματα. Επίσης ένα ακόμα διακριτικό στοιχείο για μία εγγραφή είναι αν θα έχουν μεταβληθεί κάποια από τα στοιχεία της ή αν δεν έχει μεταβληθεί καθόλου. Αυτές που δεν έχουμε αλλάξει με τους «λογικούς ελέγχους» μπορούμε να ελέγξουμε δειγματοληπτικά.

Εξαιτίας της σημαντικότητας της μεταβλητής της ολικής απώλειας η μεταβλητή αυτή έχει ελεγχθεί χρησιμοποιώντας και τα δεδομένα μίας δεύτερης πηγής πληροφοριών της έκδοσης των Lloyd's Register "World Casualty Statistics" εκτός από την "Lloyd's Casualty week" των Lloyd's of London Press. Τα αποτελέσματα της σύγκρισης επιβεβαιώνουν την αξιοπιστία των πληροφοριών του αρχείου δεδομένων [Kohavi R., 1995] και παρουσιάζονται στον Πίνακα 5.3.

### ΠΙΝΑΚΑΣ 5.3

Πίνακας επαλήθευσης με διασταύρωση (cross validation)  
της ολικής απώλειας από δύο πηγές

Πηγή 1	Πηγή 2	Ολική απώλεια Ναι	Ολική απώλεια Όχι	Σύνολο
Ολική απώλεια : Ναι		10,2%	0,5%	10,7%
Ολική απώλεια : Όχι		3,0%	86,3%	89,3%
Σύνολο		13,2%	86,8%	100%

Το 3% των πλοίων, τα οποία με βάση τα δεδομένα από τη δεύτερη πηγή, έχουν ολική απώλεια αλλά στο αρχικό αρχείο δεδομένων δεν έχουν, οφείλεται στους εξής λόγους:

- Λάθη κατά την εισαγωγή δεδομένων,
- Η δεύτερη πηγή έχει περισσότερο ενημερωμένα στοιχεία,
- Η πρώτη πηγή περιλαμβάνει τα στοιχεία με διαφορετική δομή έχοντας συχνά πολλαπλές αναφορές άρα και εγγραφές ατυχημάτων για το κάθε πλοίο. Έτσι είναι πιθανό ένα πλοίο να μην έχει δηλωθεί ως ολική απώλεια σε μία εγγραφή και να έχει δηλωθεί σε κάποια επόμενη εγγραφή.

### 5.3. Επιλογή παραμέτρων για την εφαρμογή του αλγόριθμου C5.0 και των λοιπών τεχνικών

Αφού σχολιάστηκαν στα προηγούμενα υποκεφάλαια οι διαδικασίες και έγινε η εφαρμογή τους για τον καθαρισμό του αρχείου, στη συνέχεια παρουσιάζονται τα βήματα για την εφαρμογή του αλγορίθμου C5.0 για την δημιουργία δένδρων καθώς και την εφαρμογή των λοιπών μεθοδολογιών στο αρχείο δεδομένων με σκοπό την εξαγωγή αποτελεσμάτων.

Περισσότερο αναλυτικά :

1. Διατύπωση του προβλήματος και των περιορισμών.

Το ουσιαστικό προς διερεύνηση πρόβλημα είναι η κατασκευή κανόνων που να προσδιορίζουν τα ναυτικά ατυχήματα και μάλιστα την ολική απώλεια των πλοίων από 1992 έως 1999.

2. Σχηματισμός του αναγκαίου αρχείου δεδομένων με τις απαραίτητες μεταβλητές και τις τιμές αυτών των μεταβλητών.

Στα προηγούμενα κεφάλαια έχει παρουσιαστεί ολόκληρη η διαδικασία σχηματισμού του αρχείου και προσδιορισμού της μοναδικής εγγραφής του αρχείου. Από όλη την πληροφορία του αρχείου χρησιμοποιήθηκε ένα υποσύνολο από τις μεταβλητές οι οποίες είναι επιθυμητό να μελετηθούν.

3. Προσδιορισμός των απαραίτητων μετασχηματισμών καθώς και κατασκευή νέων συνθετικών μεταβλητών είτε με ποσοτικές είτε με ποιοτικές πληροφορίες.

Στο στάδιο αυτό σχηματίζεται μία νέα σειρά μεταβλητών και κωδικοποιήσεων χρήσιμων για την μετέπειτα επεξεργασία. Σκοπός του σταδίου,

μετά την πρωτογενή επεξεργασία και επισκόπηση των κατανομών, είναι να γίνει ανακωδικοποίηση και μετασηματισμός των βασικών μεταβλητών έτσι ώστε να μετατραπεί η ποσοτική πληροφορία και σε ποιοτική. Ο σκοπός που γίνεται αυτό είναι για να απεγκλωβιστεί ο «θόρυβος» [Quinlan J. R., 1986a], που είναι έμφυτος σε συνεχείς παρατηρήσεις, από την πληροφορία των δεδομένων. Τέτοιοι μετασηματισμοί έγιναν, θα γίνουν και στη συνέχεια όπου χρειαστεί, και είναι ιδιαίτερα χρήσιμοι για την ανάλυση αφού επιτρέπουν την εξάλειψη του θορύβου από την πληροφορία των δεδομένων.

#### 4. Επιλογή του αλγορίθμου ανάλυσης (ή κατασκευή νέου) - «μοντέλο».

Το σημαντικότερο σημείο στην μεθοδολογία είναι η επιλογή του κατάλληλου αλγορίθμου για την ανάλυση των στοιχείων. Για αυτήν έχουν επιλεχθεί τα δέντρα αποφάσεων για τους λόγους που αναφέρθηκαν στο τρίτο κεφάλαιο.

Κάθε ξεχωριστή επιλογή των παραμέτρων παράγει και διαφορετικά δέντρα άρα και διαφορετικά αποτελέσματα. Σαν εξαρτημένη ή προς διερεύνηση παράμετρος επιλέχθηκε η ολική απώλεια του πλοίου και ως ανεξάρτητες ένα σύνολο λοιπών παραγόντων για την δημιουργία δένδρων. Το αρχικό λοιπόν σενάριο θα υλοποιηθεί με την ακόλουθη επιλογή παραμέτρων :

• Ολική απώλεια	εξαρτημένη- -(προς διερεύνηση)
• Νηογώνονας	ανεξάρτητη
• Σχετικό μέγεθος	ανεξάρτητη
• Μήνας	ανεξάρτητη
• Γεωγραφική περιοχή	ανεξάρτητη
• Ατύχημα	ανεξάρτητη
• Καιρικές συνθήκες	ανεξάρτητη
• Είδος περιοχής	ανεξάρτητη
• Πλήθος αλλαγής ιδιοκτητών ομάδ.	ανεξάρτητη
• Τύπος πλοίου	ανεξάρτητη
• Πλήθος ατυχημάτων	ανεξάρτητη
• Χρονολογία ατυχήματος	ανεξάρτητη
• Έτος ναυπήγησης	ανεξάρτητη



• Χωρητικότητα	ανεξάρτητη
• Νηογνώμονες ομαδοποιημένοι	ανεξάρτητη
• Ηλικία	ανεξάρτητη
• Ακολουθία εγγραφών ανά πλοίο	ανεξάρτητη
• Πεδίο ταξινόμησης	ανεξάρτητη
• Πλήθος αλλαγής ιδιοκτητών	ανεξάρτητη
• Χρονική μεταβλητή ατυχήματος	ανεξάρτητη
• Χρονική μεταβλητή καταγραφής ατυχ.	ανεξάρτητη
• Σημαία	ανεξάρτητη
• Σειρά ατυχημάτων	ανεξάρτητη
• Μη επιτυχές ταξίδι	ανεξάρτητη

Επίσης επιλέχθηκε τα τελικά φύλλα του δέντρου να έχουν τουλάχιστον δέκα εγγραφές και να μην υπάρχει διαφορά στα κόστη σφάλματος ταξινόμησης της ολικής και της μη ολικής απώλειας.

Με τις ίδιες ακριβώς παραμέτρους και τον ίδιο διαχωρισμό εξαρτημένων και ανεξάρτητων που αναφέρθηκαν προηγουμένως παρήχθησαν και οι κανόνες ταξινόμησης.

Για την ανάλυση αξιοπιστίας η οποία δίνει τη σχετική βαρύτητα στην μη γραμμική συνάρτηση εκτίμησης της ολικής απώλειας, με τη χρήση των νευρωνικών δικτύων, συνοπτικές πληροφορίες για τα οποία δίνονται στο Παράρτημα Α, η τοπολογία που επιλέχθηκε είναι :

Input Layer : Όλα τα προαναφερθέντα ανεξάρτητα πεδία  
/ μεταβλητές

Hidden Layer: Ένα

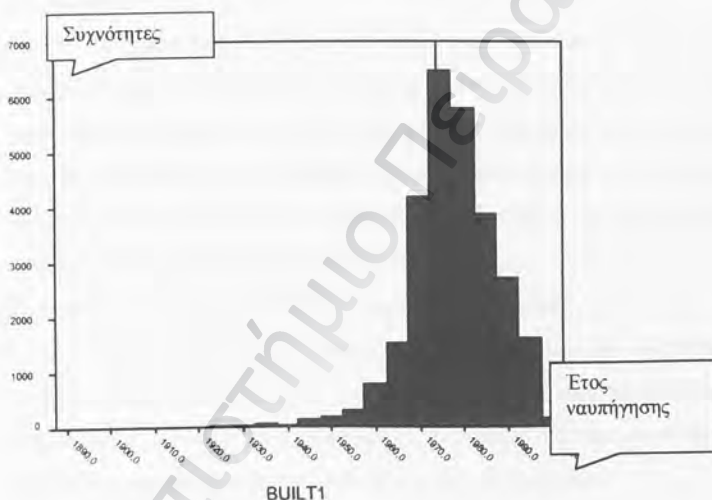
Output Layer : Ολική απώλεια

Με χρήση της μεθοδολογίας Boosting [Drucker H., 1999], το αρχείο «εξισορροπείται» (“balanced”), με χρήση συντελεστή, σε σχέση με την ολική απώλεια και μπορούμε να δοκιμάσουμε να δημιουργήσουμε δένδρα ώστε να μελετήσουμε την επίδραση ενός υποσυνόλου παραγόντων [Hand D. et al, 2001] στην ολική απώλεια. Θα υλοποιηθεί αυτή τη φορά σενάριο με την επιλογή παραμέτρων όπως Τύπος πλοίου, Γεωγραφική περιοχή, Νηογνώμονες ομαδοποιημένοι, Πλήθος αλλαγής ιδιοκτητών ομαδοποιημένο, Έτος ναυπήγησης

σαν ανεξάρτητες και Ολική απώλεια σαν εξαρτημένη και άρα προς διερεύνηση μεταβλητή.

Για την εφαρμογή του αλγορίθμου G.R.I μια από τις παραμέτρους που επιλέχθηκαν είναι το Έτος ναυπήγησης. Η παράμετρος αυτή μετασχηματίστηκε από την ποσοτική της μέτρηση σε ποιοτική κατηγοριοποίηση με το όνομα Έτος ναυπήγησης -74- ( BUILT 74 ) με δύο τιμές πλοία παλιά ή νεότερα. Παλιά που έχουν ναυπηγηθεί πριν το έτος 1974 ή νεότερα μετά το έτος 1974, όπως φαίνεται και από την κατανομή στο διάγραμμα 5.4.

ΔΙΑΓΡΑΜΜΑ 5.4  
Κατανομή έτους ναυπήγησης



Συνολικά από τις παραμέτρους για την εφαρμογή του αλγορίθμου G.R.I ως εξαρτημένο μέγεθος επιλέχθηκε η ολική απώλεια και ως ανεξάρτητα τα παρακάτω κατηγορικά:

- |                              |                              |
|------------------------------|------------------------------|
| • Ολική απώλεια              | προς διερεύνηση (εξαρτημένο) |
| • Έτος ναυπήγησης -74-       | ανεξάρτητο                   |
| • Τύπος πλοίου               | ανεξάρτητο                   |
| • Γεωγραφική περιοχή         | ανεξάρτητο                   |
| • Νηογνώμονες ομαδοποιημένοι | ανεξάρτητο                   |

5. Ερμηνεία και αξιολόγηση του μοντέλου για μελλοντική χρήση.

Οι κανόνες που παράγονται με βάση τα δέντρα αποφάσεων χρησιμοποιούνται για τη παραγωγή συμπερασμάτων για τους λόγους και τα αίτια που προκαλούν το παρατηρούμενο γεγονός. Ακόμα σπουδαιότερο είναι το γεγονός ότι ταξινομούν τους παραπάνω λόγους σε σειρά σημαντικότητας εμπλουτίζοντας την αντίληψη του ερευνητή για το φαινόμενο.

Σε αρκετές περιπτώσεις τα παραχθέντα συμπεράσματα του μοντέλου δίνουν το έναυσμα για ακόμα βαθύτερη διερεύνηση των ερευνητικών δεδομένων με την αντίληψη που προσφέρουν μέσω των αποτελεσμάτων. Το γεγονός αυτό μπορεί να οδηγήσει στη επιλογή νέων παραμέτρων ως μεταβλητών ορισμού του μοντέλου.

Ο αλγόριθμος GRI έχει μία απλή αλλά σημαντική χρήση. Σε περίπτωση που βάση των επιλεχθέντων παραμέτρων δεν παραχθεί κανένας κανόνας, τότε αυτόματα εξάγεται το συμπέρασμα ότι δεν υπάρχουν «γενικού» κανόνες στο αρχείο δεδομένων και τα όποια συμπεράσματα πρέπει να ανιχνευτούν πια με ποιοτικό τρόπο, εξετάζοντας τη συμπεριφορά κάθε μιας εγγραφής δηλαδή στην περίπτωση μας του ατυχήματος ενός πλοίου.

6. Συμπεράσματα και τροποποιήσεις για μελλοντική χρήση.

Τέλος, θα δοθούν συμπεράσματα από τις επί μέρους διαδικασίες επεξεργασίας καθώς και γενικά συμπεράσματα από την συνολική ανάλυση των αποτελεσμάτων και θα διατυπωθούν προτάσεις για μελλοντικές ερευνητικές εργασίες, που θα προκύψουν από το σύνολο της διατριβής.



## ΚΕΦΑΛΑΙΟ 6

### 6.1 Εισαγωγή

Στο κεφάλαιο αυτό παρουσιάζονται τα αποτελέσματα από το σύνολο των αναλύσεων και διαδικασιών που εφαρμόστηκαν. Καταρχήν παρουσιάζονται βασικά ποσοστιαία αποτελέσματα στο σύνολο των εγγραφών μετά τις διαδικασίες προετοιμασίας του αρχείου.

Το ποσοστό των περιπτώσεων στο σύνολο των εγγραφών που μελετήθηκαν ανέρχεται σε 82 % αν εξαιρέσουμε, για λόγους που εξηγήθηκαν στο κεφάλαιο παρουσίασης των δεδομένων στην παράγραφο 4.3.11, εγγραφές που αφορούν τύπους πλοίων όπως “Fishing”, “Pontoon-Barge”, “Miscellaneous activities”, και “Others” σε ποσοστό 18 %.

Στο σύνολο των εγγραφών του αρχείου, δεν περιελήφθησαν, σύμφωνα με τον ορισμό του ναυτικού ατυχήματος που δόθηκε στο 2.2.1 υποκεφάλαιο, σε ποσοστό 20 % τα μη “πραγματικά ατυχήματα” όπως “Seizure” κ.τ.λ., οπότε το ποσοστό των περιπτώσεων ατυχημάτων που μελετήθηκαν ανέρχεται σε 80 %.

Το ποσοστό των τελικά, μετά τις διαδικασίες καθαρισμού των δεδομένων, διαφορετικών συμβάντων που μελετήθηκαν ανέρχεται σε 53 % στο σύνολο των εγγραφών.

Σε σύνολο 20337 διαφορετικών εγγραφών εμφανίζονται 17081 διαφορετικά πλοία, δηλαδή ποσοστό 16 % των πλοίων επανεμφανίζει σε άλλη χρονική στιγμή ατυχήματα.

Στα υποκεφάλαια που ακολουθούν παρουσιάζονται τα αποτελέσματα από την ανάλυση των μεταβλητών του αρχείου δεδομένων για το σύνολο των εγγραφών του, που προέκυψε τελικά μετά τις διαδικασίες καθαρισμού των δεδομένων. Στη συνέχεια δίνονται τα αποτελέσματα που παρήχθησαν με χρήση του αλγορίθμου C5.0 που παράγει τα δέντρα αποφάσεων και τους ταξινομητικούς κανόνες. Ακολουθούν αναφέρονται οι παρατηρήσεις από την ανάλυση της σχετικής βαρύτητας όλων των ανεξάρτητων μεταβλητών που χρησιμοποιήθηκαν για την πρόβλεψη της εξαρτημένης μεταβλητής με χρήση νευρωνικών δικτύων. Τέλος, δίνονται τα αποτελέσματα που παρήχθησαν με εφαρμογή μεθοδολογίας Boosting και εφαρμογή του G.R.I αλγορίθμου.

## 6.2 Αποτελέσματα από την ανάλυση των μεταβλητών

Η ανάλυση των δεδομένων είναι ένα βασικό στοιχείο της παρούσας εργασίας αφού δίνει για πρώτη φορά με αυτή την μεθοδολογία μια μεγάλης κλίμακας, όπως είχε γίνει παλαιότερα (Giziakis K. P.h.d, 1987), ανάλυση των ατυχημάτων.

Από την πληθώρα σχέσεων για τα ατυχήματα που καταγράφηκαν στο σύνολο των 20337 εγγραφών του αρχείου αναφέρονται στη συνέχεια οι σημαντικότερες και αναλυτικά παρουσιάζονται στο Παράρτημα Π1.

Η συχνότερα παρατηρούμενη αιτία ναυτικών ατυχημάτων είναι η μηχανολογική βλάβη (HULL/MCHY/EQUIP.DAMAGED) που παρατηρείται στο ένα τρίτο των περιπτώσεων. Αρκετά σημαντικές αιτίες ατυχημάτων είναι η σύγκρουση με άλλο πλοίο (COLLISION WITH SHIP) και η προσάραξη (GROUNDED). Οι υπόλοιπες ομάδες ατυχημάτων αναφέρθηκαν με μικρότερα ποσοστά

Περισσότερα ατυχήματα έχουν καταγραφεί τον Ιανουάριο με ποσοστό 9,9 % ενώ λιγότερα τον Απρίλιο με ποσοστό 7,5 % και τον Σεπτέμβριο με ποσοστό 7,1 %.

Οι γεωγραφικές περιοχές: N. SEA, BALTIC, KIEL και JAPAN, S. CHINA & E. INDIES καθώς και W. INDIES, U.S. EASTERN, GREAT LAKES, NEWFOUNDLAND, GULF OF MEXICO και συγκεντρώνουν τα υψηλότερα ποσοστά 24,8 %, 20 % και 14,7 % αντίστοιχα.

Οι καιρικές συνθήκες συνήθως δεν αναφέρονται. Πάντως στις περιπτώσεις που αναφέρονται η καταιγίδα STORM εμφανίζεται με το υψηλότερο ποσοστό 11,5 %.

Μόνο το 30% των πλοίων που είχε ατύχημα είχε μόνο ένα ιδιοκτήτη.

Τέλος το ποσοστό των περιπτώσεων ολικής απώλειας στο σύνολο των εγγραφών για το χρονικό διάστημα από 1992 έως 1999 ανέρχεται 7,3 %.

Ιδιαίτερα ενδιαφέρουσα είναι η διερεύνηση του μήνα, σαν παράγοντα χρόνου, με την ευρύτερη Γεωγραφική περιοχή που κατεγράφη το ατύχημα. Στον πίνακα 6.1 δίνονται τα ποσοστά εμφάνισης ατυχήματος ανά ευρύτερη γεωγραφική περιοχή στην διάρκεια των μηνών για τα έτη από 1992 έως και 1999. Στον πίνακα 6.1, για παράδειγμα, φαίνεται ότι στον Ινδικό ατυχήματα έχουν παρατηρηθεί με μεγαλύτερη συχνότητα τον Φεβρουάριο, τον Μάρτιο και τον Απρίλιο με ποσοστά 13 %, 13 %, και 18,5 % αντίστοιχα, έναντι του 8,3 % του μέσου όρου.

## ΠΙΝΑΚΑΣ 6.1

Ποσοστό εμφάνισης ατυχήματος στις Γεωγραφικές περιοχές ανά Μήνα

	Jan	Feb	Mar	Apr	May	Jun	Jul	Aug	Sep	Oct	Nov	Dec
N. SEA,												
BALTIC, KIEL	12,2	11,3	9,5	7,5	7,6	6,3	5,8	6,4	6,7	8,3	8,6	9,8
MEDITERRAN												
EAN, SUEZ	10,6	9,9	10,4	8,3	6,8	5,9	9,1	8,5	5,2	7,9	8,3	9,1
RED SEA E.&												
W. AFRICAN												
COAST	9,7	8,3	7,7	7,4	10,2	10,9	9,6	7,3	6,7	8,2	7,5	6,5
W. INDIES	9,6	8,6	8,4	8,0	8,1	7,7	6,8	9,8	6,2	8,4	7,9	10,4
INDIAN												
OCEAN &												
ANTARTIC	1,9	13,0	13,0	18,5	1,9	7,4	9,3	7,4	18,5	7,4	1,9	0,0
GULF, BAY OF												
BENGAL	8,8	6,4	4,7	5,2	6,4	15,3	12,6	12,4	8,1	7,7	6,9	5,7
JAPAN, S.												
CHINA & E.												
INDIES	9,0	9,2	8,3	7,1	9,1	7,4	9,5	8,8	7,4	8,2	7,9	8,0
PACIFIC	13,9	7,9	4,2	6,0	6,3	11,5	9,7	9,7	6,0	6,3	7,9	10,6
AUSTRALIA	7,3	9,6	10,2	6,3	8,9	4,1	5,8	10,2	10,9	4,5	9,9	12,2
ATLANTIC	12,3	8,7	8,4	6,0	8,7	9,0	7,9	8,4	7,7	5,6	7,9	9,3
W. COAST S-N.												
AMERICA	9,5	9,6	10,2	7,0	6,5	8,2	9,8	8,3	9,2	9,5	6,5	5,6
ALASKA,												
U.S.S.R. ARTIC	14,2	10,0	8,6	8,3	6,0	6,8	8,6	11,3	6,5	8,0	6,0	5,5
Σύνολο περιοχών	10,4	9,5	8,7	7,3	7,9	7,8	8,2	8,6	6,9	8,0	8,0	8,7

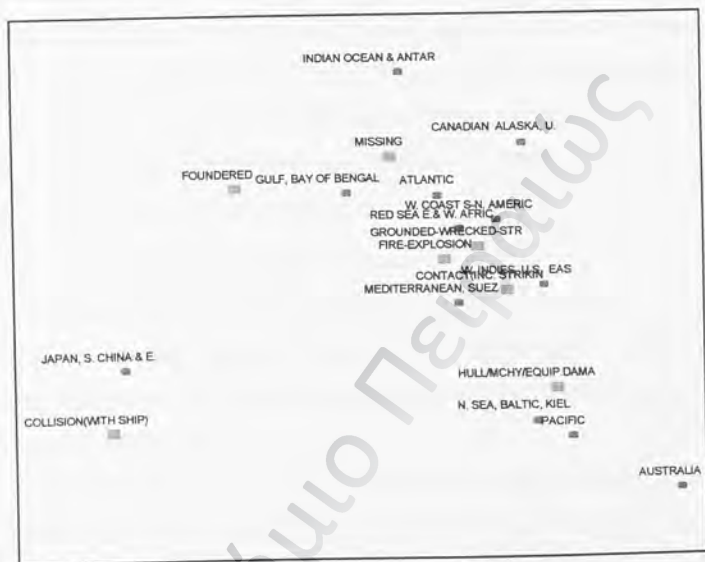
Τέλος για να αναζητηθούν σχέσεις στα δεδομένα αρχικά απαιτείται και η οπτική παρατήρηση (data understanding / data visualization), έτσι με το διάγραμμα (Row and Column Points / Symetrical Normalization) 6.2 δίνεται η δυνατότητα να



οπτικοποιήσουμε και να ελέγξουμε την σχέση των μεταβλητών ανά δύο. Στο διάγραμμα αυτό παρουσιάζεται η σχέση του Ατυχήματος και της Γεωγραφικής περιοχής που συνέβη αυτό.

## ΔΙΑΓΡΑΜΜΑ 6.2

Σχετική σύγκριση Γεωγραφικής περιοχής και Ατυχήματος



Με τα μεγαλύτερα και σε σχήμα ορθογωνίου παραλληλογράμμου σχήματα απεικονίζονται οι τιμές των ατυχημάτων, ενώ με τα μικρότερα σχήματος πολύγωνου οι γεωγραφικές περιοχές. Αυτό που έχει σημασία σε αυτό το διάγραμμα είναι η σχετική απόσταση των σημείων μεταξύ τους, δεν έχουν σημασία τα απόλυτα νούμερα γιατί αντιστοιχούν σε «μη πραγματικούς άξονες». Μια παρατήρηση του διαγράμματος δείχνει ότι το συχνότερο παρατηρούμενο ατύχημα, για παράδειγμα, στην γεωγραφική περιοχή JAPAN, S. CHINA & E. INDIES είναι το ατύχημα COLLISION και όχι το HULL/MCHY/EQUIP.DAMAGED. Αυτό προκύπτει από παρατήρηση του διαγράμματος στο οποίο η απόσταση του σημείου που αντιστοιχεί στην γεωγραφική περιοχή JAPAN, S. CHINA & E. INDIES από το σημείο που αντιστοιχεί στο ατύχημα COLLISION είναι κατά πολύ μικρότερη από την απόσταση ανάμεσα στο σημείο που αντιστοιχεί στην γεωγραφική περιοχή JAPAN, S. CHINA &

Ε. INDIES και το σημείο που αντιστοιχεί στο ατύχημα HULL/MCHY/EQUIP.DAMAGED.

Μελετώντας την ολική απώλεια σαν εξαρτημένη μεταβλητή σε σχέση με τους υπολοίπους άλλους ανεξάρτητους παράγοντες, μια και αυτός είναι ένας στόχος της παρούσας εργασίας, προκύπτουν ενδιαφέροντα συμπεράσματα από τους Πίνακες που παρατίθενται στη συνέχεια. Οι επιλεγμένοι πίνακες που ακολουθούν παρουσιάζουν τις βασικές πληροφορίες από την επεξεργασία του αρχείου. Σε κάθε ένα πίνακα παρουσιάζονται τρεις στήλες. Η πρώτη στήλη περιέχει την προς εξέταση μεταβλητή και τις τιμές που έχει αυτή στο αρχείο. Η δεύτερη στήλη παρουσιάζει το ποσοστό ολικής απώλειας για τη συγκεκριμένη κατηγορία μεταβλητών. Η τρίτη και σημαντικότερη στήλη καταγράφει τις εγγραφές στις οποίες παρατηρείται η τιμή των μεταβλητών αυτών στο σύνολο των εγγραφών του αρχείου.

Οι πίνακες αυτοί διαβάζονται δίνοντας σημασία:

- και στο πλήθος εμφανίσεων (ή όχι) του καθενός γεγονότος
- και στο ποσοστό ολικής απώλειας για το συγκεκριμένο γεγονός.

Σε κάθε πίνακα στην δεύτερη στήλη το ποσοστό ολικής απώλειας είναι διατεταγμένο κατά φθίνουσα τάξη.

Στον Πίνακα 6.3 κατανομής Ατυχημάτων, για παράδειγμα, η φωτιά / έκρηξη καταγράφηκε σε 1823 εγγραφές και από αυτές τις εγγραφές στο 10,3 % έχει παρατηρηθεί ολική απώλεια. Μεγαλύτερο ποσοστό ολικής απώλειας έχουμε σε πλοία που έχει συμβεί ατύχημα με ενδεικτική τιμή FOUNDERED.

ΠΙΝΑΚΑΣ 6.3

Κατανομή Ατυχημάτων

Ατύχημα	% Ολ.Απ.	εγγραφές
1 FOUNDERED	54,7	1022
2 MISSING	22,1	407
3 FIRE-EXPLOSION	10,3	1823
6 GROUNDED	5,9	3819
4 COLLISION(WITH SHIP)	4,6	4073
8 HULL/MCHY/EQUIP.DAMAGED	2,5	7069
5 CONTACT(INC. STRIKING ICE)	2,3	2097
	7,3	20337

Από τον Πίνακα 6.4 κατανομής Γεωγραφικών περιοχών, συμπεραίνουμε ότι οι λιγότερες εγγραφές παρατηρούνται στην Αυστραλία και στον Ινδικό. Όμως τα ποσοστά ολικής απώλειας σε αυτές διαφέρουν και μάλιστα για την περιοχή του Ινδικού έχουμε το υψηλότερο ποσοστό. Αντίθετα οι περισσότερες εγγραφές εμφανίζονται στις θάλασσες της Βόρειας Ευρώπης και της Ιαπωνίας-Κίνας. Πάλι όμως υπάρχει μία σημαντική διαφορά στα ποσοστά ολικής απώλειας. Παρατηρούμε επίσης, ότι εμφανίζεται το μικρότερο ποσοστό ολικής απώλειας στις περιπτώσεις που το ατύχημα συνέβη στην γεωγραφική περιοχή της Αυστραλίας.

ΠΙΝΑΚΑΣ 6.4  
Κατανομή Γεωγραφικών περιοχών

Γεωγραφική περιοχή	% Ολ.Απ	εγγραφές
5 INDIAN OCEAN & ANTARTIC	20,5	39
6 GULF, BAY OF BENGAL	12,8	1356
7 JAPAN, S. CHINA & E. INDIES	12,0	4063
3 RED SEA E.& W. AFRICAN COAST	11,8	1157
12 CANADIAN AL, U.S.S.R. ARTIC & BER. SEA, ICELAND	8,5	495
2 MEDITERRANEAN, SUEZ	6,6	2828
11 W. COAST S-N. AMERICA, PANAMA CAPE HORN	6,4	708
10 ATLANTIC	5,9	919
4 W. INDIES, U.S.EAST, GULF. MEX, GR.LAKES, NEWFOU	4,3	2984
8 PACIFIC	4,1	241
1 N. SEA, BALTIC, KIEL	4,0	5049
9 AUSTRALIA	1,4	498
	7.3	20337

Ο πίνακας 6.5 θα διαβασθεί με την υπενθύμιση ότι όταν λέμε πλήθος αλλαγής ιδιοκτητών ομαδοποιημένο εννοούμε το πόσες φορές το πλοίο «άλλαξε χέρια». Στον συγκεκριμένο πίνακα μπορούμε να παρατηρήσουμε ότι όταν το πλοίο έχει «αλλάξει χέρια» πάνω από πέντε φορές τότε εμφανίζεται και το μεγαλύτερο ποσοστό 8.8 % ολικής απώλειας.



### ΠΙΝΑΚΑΣ 6.5

Κατανομή Πλήθους αλλαγής ιδιοκτητών ομαδοποιημένο

Πλήθος αλλαγής ιδιοκτητών ομαδοποιημένο	% Ολ.Απ.	εγγραφές
4 Αλλαγές ιδιοκτητών άνω των πέντε	8,8	1959
2 Αλλαγές ιδιοκτητών δύο	6,5	3601
3 Αλλαγές ιδιοκτητών τρεις με τέσσερες	5,8	4181
1 Ένας ιδιοκτήτης	4,1	6195
5 No info	13,0	4401
	7,3	20337

Από τον πίνακα 6.6 φαίνεται ότι υψηλότερα ποσοστά ολικής απώλειας παρατηρείται στις περιπτώσεις Τυφώνα 15,4 % (TYPHOON) και Θύελλας 15,2 % (STORM).

### ΠΙΝΑΚΑΣ 6.6

Κατανομή Καιρικών συνθηκών

Καιρικές συνθήκες	% Ολ.Απ.	εγγραφές
5 TYPHOON	15,4	622
3 STORM	15,2	2336
2 POOR VISIBILITY	8,0	440
1 CALM WEATHER	6,7	608
4 SNOWSTORM	6,1	82
6 No info	5,8	16249
	7,3	20337

### ΠΙΝΑΚΑΣ 6.7

Κατανομή Είδους περιοχής

Είδος περιοχής	% Ολ.Απ.	εγγραφές
1 OPEN SEA	9,9	7361
3 PORT	5,8	8920
2 CONTROLLED	5,7	4056
	7,3	20337

Στον πίνακα 6.7 φαίνεται ότι υψηλότερο ποσοστό ολικής απώλειας παρατηρείται στις περιπτώσεις που το ατύχημα έγινε στις ανοικτές θάλασσες 9.9 % και για πλήθος 7361 εγγραφών.

ΠΙΝΑΚΑΣ 6.8  
Κατανομή Νηογνώμονα

Νηογνώμονας	% Ολ.Απ.	εγγραφές
KI Biro Klasifikasi Indonesia	29,0	93
HR Hellenic Register of Shipping	11,9	286
CC China Classification	9,2	120
KR Korean Register of Shipping	7,3	384
BV Bureau Veritas	7,2	1898
RI Registro Italiano Navale	6,9	288
NK Nippon Kaiji Kyokai	6,5	1777
RS Russian Register of Shipping	5,9	581
GL Germanischer Lloyd	5,0	1645
IR Indian Register of Shipping	4,7	86
PR Polski Rejestr Statkow	4,7	150
LR Lloyd's Register of Shipping	4,5	4446
NV Det Norske Veritas Classification A/S	3,7	1545
RN Romanian Register of Shipping	2,4	166
AB American Bureau of Shipping	2,3	1663
Oth Other Classifications	15,3	5209
	7.3	20337

Στην πρώτη θέση από πλευράς ολικών απωλειών, σύμφωνα με τον πίνακα 6.8 είναι ο νηογνώμονας της Ινδονησίας με ποσοστό 29 % και 93 εγγραφές των πλοίων. Αξιοσημείωτο από τον πίνακα 6.8 το πολύ χαμηλό ποσοστό 2,3 %, στις 1663 εγγραφές, των πλοίων που ήταν εγγεγραμμένα στον Αμερικανικό Νηογνώμονα όταν καταγράφηκε το ατύχημα.

Στον πίνακα 6.9 παρουσιάζεται το ποσοστό και το πλήθος των εγγραφών ανά μήνα για το σύνολο των καταγεγραμμένων ατυχημάτων που οδήγησαν σε ολική απώλεια και μάλιστα τον Σεπτέμβριο και Μάρτιο εμφανίζονται τα μικρότερα ποσοστά, 5,9 % και 5,2 % αντίστοιχα, ολικής απώλειας.

ΠΙΝΑΚΑΣ 6.9  
Κατανομή Μηνών

Μήνας	% Ολ.Απ.	εγγραφές
November	8,6	1624
July	8,4	1716
August	8,2	1737
February	8,0	1834
October	7,9	1749
December	7,8	1748
May	7,7	1562
June	6,7	1611
Janauary	6,4	2008
April	6,1	1523
September	5,9	1448
March	5,2	1777
	7.3	20337

ΠΙΝΑΚΑΣ 6.10  
Κατανομή Τύπου πλοίου

Τύπος πλοίου	% Ολ.Απ	εγγραφές
3 FERRY	8,9	2722
1 BULK	7,5	12616
2 TANKER	6,5	3716
4 CONTAINER	3,8	1283
	7.3	20337



Για τον πίνακα 6.10 θα πρέπει να σημειωθεί ότι στον συλλεκτικό κωδικό FERRY με ποσοστό 8,9 % περιέχονται και οι επί μέρους κατηγορίες όπως PASSENGER, PASSENGER/RO-RO CARGO κ.α.. Επίσης έχουν αναλυθεί στον συγκεκριμένο πίνακα, οι επιλεγείσες για ανάλυση τέσσερες συλλεκτικές τιμές των μεταβλητών, όπως αναλυτικά αναφέρονται και σχολιάζονται στο υποκεφάλαιο 4.4.

Με την Όρυξη δεδομένων διερευνώνται και σχέσεις που μπορεί να μην είναι η να μην θεωρηθούν καταρχήν σημαντικές. Εντύπωση έκανε και καταγράφεται η σχετική θέση της εβδομάδας στο μήνα, όπως αυτή απεικονίζεται στον Πίνακα 6.11.

ΠΙΝΑΚΑΣ 6.11

Σχετική θέση της εβδομάδας στο μήνα

Σχετική θέση της εβδομάδας στο μήνα	% Ολ.Απ.	εγγραφές
4 Τέταρτη εβδομάδα	8,3	4132
1 Πρώτη εβδομάδα	7,9	3636
3 Τρίτη εβδομάδα	7,2	5575
2 Δεύτερη εβδομάδα	6,3	6994
	7.3	20337

Είναι στα όρια της στατιστικής σημαντικότητας ( $p\text{-value} = 0,063$ ), και απλά αναφέρει ότι φαίνεται να υπάρχουν αναλογικά περισσότερες περιπτώσεις ολικής απώλειας στις αρχές ή στα τέλη ενός μήνα, ενώ όπως διαπιστώνεται από τη τελευταία στήλη των εγγραφών τα σχετικά περισσότερα συμβάντα καταγράφονται στα μέσα του μήνα, δεύτερη και τρίτη εβδομάδα.

ΠΙΝΑΚΑΣ 6.12

Μέσοι όροι για την ολική και όχι ολική απώλεια για Έτος ναυπήγησης, Ηλικία Χωρητικότητα, Πλήθος αλλαγής ιδιοκτητών

	Ολική απώλεια - Ναι	Ολική απώλεια - Όχι
Έτος ναυπήγησης	1974	1978
Ηλικία	23 έτη	17 έτη
Χωρητικότητα	8310	15682
Πλήθος αλλαγής ιδιοκτητών	2,85	2,46

Χαρακτηριστικές είναι και οι αντιπαραβολές των μέσων όρων των μεταβλητών (Παράρτημα Π<sub>3</sub>) για την ολική και όχι ολική απώλεια. Τα πλοία με ολική απώλεια έχουν ναυπηγηθεί κατά μέσο όρο το 1974, ενώ αυτά που δεν έχουν ολική απώλεια το 1978. Το έτος ναυπήγησης και άλλα χαρακτηριστικά όπως ηλικία, χωρητικότητα, πλήθος αλλαγής ιδιοκτητών παρουσιάζονται στον Πίνακα 6.12.

Εξάλλου μπορούμε για κάθε ένα παράγοντα και την Ολική απώλεια να παρουσιάσουμε αναλύσεις όπως:

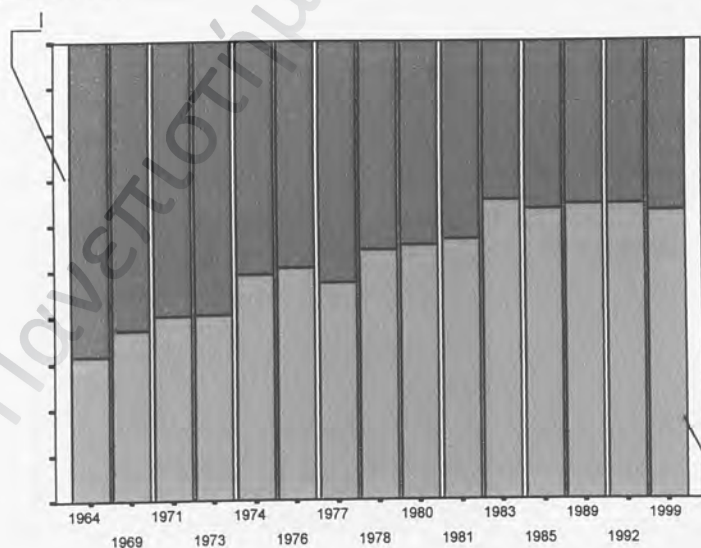
- Ανάλυση Έτους ναυπήγησης και Ολικής απώλειας

Στο διάγραμμα (stacked bar) 6.13, χρονιάς ναυπήγησης και ολικής απώλειας, η ολική απώλεια του πλοίου αναλύεται και με βάση τη χρονιά κατασκευής.

Διάγραμμα 6.13

Έτος ναυπήγησης σε σχέση με ολική και μη ολική απώλεια

Ολική απώλεια %



Έτος ναυπήγησης

Μη ολική απώλεια

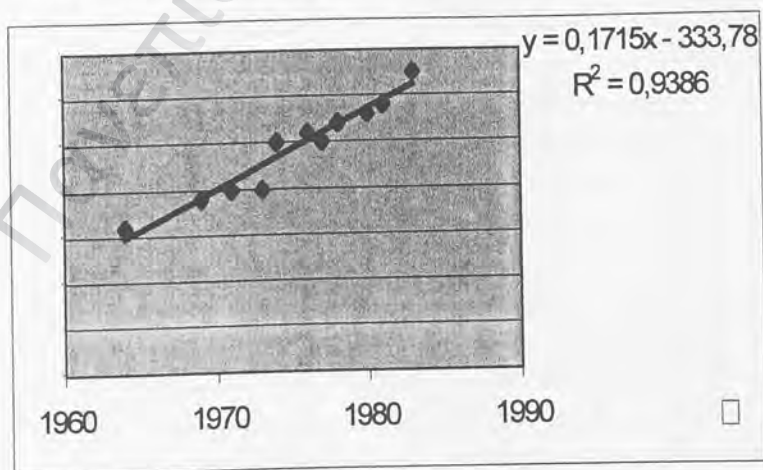
Στο διάγραμμα ο κατακόρυφος άξονας δείχνει τα ποσοστά ολικής απώλειας και μάλιστα κάθε υποδιαίρεση στο κατακόρυφο άξονα είναι και ένα δεκατημόριο (decile) ενώ ο οριζόντιος άξονας δείχνει τη χρονιά κατασκευής. Οι ράβδοι (bar) που ξεκινούν από τον οριζόντιο άξονα δίνουν τα ποσοστά των πλοίων στα οποία δεν συνέβη ολική απώλεια και οι συνεχόμενοι στις προαναφερθείσες ράβδους, αλλά διαφορετικού χρώματος, δίνουν τα ποσοστά των πλοίων στα οποία συνέβη ολική απώλεια.

Τέλος το διάγραμμα είναι κανονικοποιημένο στο 100% και ο οριζόντιος άξονας υπολογίζεται με βάση τον αριθμό των πλοίων και όχι με χρονικό εύρος (έτσι δεν έχουμε ίσους πλάτους χρονικά διαστήματα), ώστε να υπάρχει αξιόπιστος αριθμός παρατηρήσεων σε κάθε ράβδο του διαγράμματος.

Η σχέση χρονιάς ναυπήγησης και ολικής απώλειας μπορεί να αποδοθεί γραμμικά με μορφή  $y = ax + b$  έως το 1983, όπως αποδεικνύεται με την μέθοδο ελαχίστων τετραγώνων για ικανοποιητική τιμή του  $R^2 = 0,938$  και τιμές των εκτιμητριών ελαχίστων τετραγώνων (least square estimators) τις  $a = 0,1715$  και  $b = -333,78$  και παρουσιάζεται στο γράφημα 6.14. Άρα συμπεραίνουμε ότι όσο παλαιότερο από το 1983 έτος ναυπήγησης, έχει το πλοίο τόσο μεγαλύτερα ποσοστά ολικής απώλειας παρατηρούνται.

Γράφημα 6.14

Προσαρμογή ευθείας ελαχίστων τετραγώνων στο διάγραμμα διασποράς





Με ανάλογο τρόπο αποδεικνύουμε ότι η σχέση χρονιάς ναυπήγησης και ολικής απώλειας είναι γραμμική της μορφής  $y = c$ . Άρα για τα πλοία που ναυπηγήθηκαν μετά το 1984 φαίνεται η τάση να είναι σταθερή και δεν παρατηρείται για αυτά τα πλοία αξιοσημείωτη αύξηση ή ελάττωση του ποσοστού ολικής απώλειας.

- Ανάλυση Χωρητικότητας και Νηογνώμονα σε σχέση με την ολική απώλεια

Οι πληροφορίες από τον Πίνακα 6.15, Χωρητικότητα και Νηογνώμονα, που δίνει τις μέσους όρους χωρητικότητας για τα πλοία που είναι εγγεγραμμένα σε κάθε

ΠΙΝΑΚΑΣ 6.15

Χωρητικότητα, μέση τιμή σε grt, και Νηογνώμονας σε σχέση με την ύπαρξη ή όχι ολική απώλεια

	total loss	
	No	Yes
AB	25242	23926
BV	11312	6844
CC	11638	13661
GL	7087	3177
HR	8265	3001
IR	16794	9550
KI	4349	1209
KR	13417	2963
LR	16195	8474
NK	19356	15154
NV	27519	19672
oth	14409	6994
PR	10594	8519
RI	12218	16120
RN	9029	2963
RS	7825	7448
ZC	12783	9070

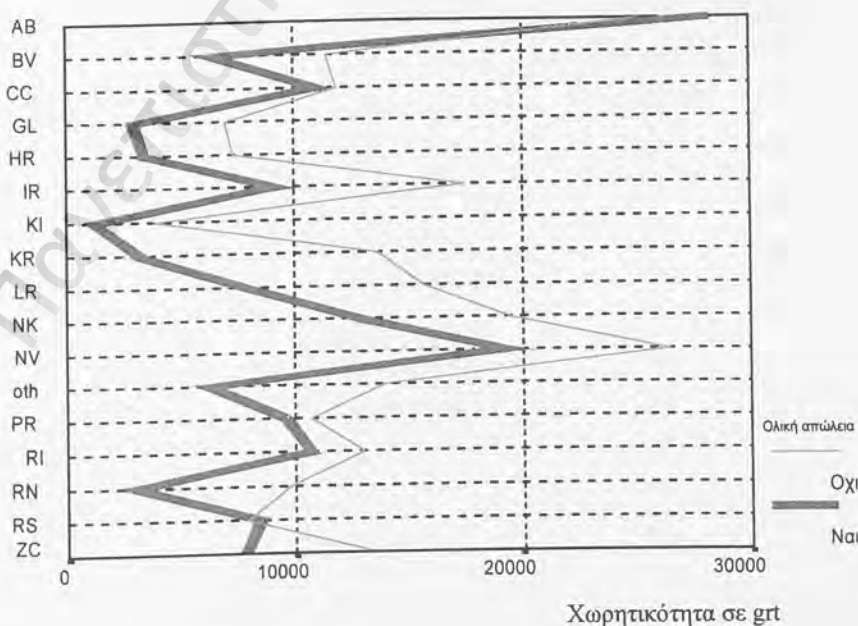
νηογνώμονα και είχαν ατύχημα, όπως αυτά είναι χωρισμένα σε δύο σύνολα για την ολική απώλεια «Yes» και την μη ολική απώλεια «No» αναπαριστώνται στο διάγραμμα 6.16.

Για την υπολογισθείσα τιμή του  $P_{value}$  για την αλληλεπίδραση (interaction) μέσου όρου χωρητικότητας και νηογνώμονα η σχέση δεν είναι στατιστικά σημαντική ( $P_{value} > 0.05$ ), άρα οι γραμμές πρέπει να είναι σχεδόν παράλληλες το οποίο φαίνεται και στο 6.16 διάγραμμα (graph line multiple). Πρακτικά αυτό σημαίνει ότι για κάθε νηογνώμονα τα πλοία μικρότερης σε μέσο όρο χωρητικότητας έχουν μεγαλύτερη συχνότητα εμφάνισης ολικής από ότι μη ολικής απώλειας όπως φαίνεται και από τον πίνακα 6.15. Άρα και για τους θεωρούμενους «αυστηρούς» νηογνώμονες, επομένως ανεξάρτητα νηογνώμονα, παρατηρείται ολική απώλεια κυρίως στις μικρές χωρητικότητες παρά στις μεγάλες.

Παρατηρούμε επίσης ότι ορισμένοι νηογνώμονες, όπως για παράδειγμα ο NV, έχουν μεγαλύτερο μέσο όρο χωρητικότητας των πλοίων που είναι εγγεγραμμένα σε αυτούς σε σχέση με άλλους, όπως ο IR, στην περίπτωση ολικής απώλειας.

ΔΙΑΓΡΑΜΜΑ 6.16

Χωρητικότητα, μέση τιμή σε grt, και Νηογνώμονας σε σχέση με την ολική και όχι ολική απώλεια



### 6.3 Αποτελέσματα από την ανάλυση του αρχείου δεδομένων με δέντρα αποφάσεων

Η χρήση της μεθοδολογίας δέντρων αποφάσεων βοηθάει την κατασκευή «ιεραρχικών διαδρομών», όπως αναφέρθηκε στο υποκεφάλαιο 3.3.1, για την ανάλυση της σημαντικότητας των διαφόρων παραγόντων. Όπως είναι προφανές και από το υποκεφάλαιο 6.2 μπορούν να κατασκευαστούν αναρίθμητα πολλές αναλύσεις.

Η ανάγκη όμως είναι να μελετηθεί η ολική απώλεια ως ένα μοναδικό φαινόμενο το οποίο θα εκδηλωθεί (ή όχι) μετά από μία σειρά ανεξάρτητων μεταξύ τους παραγόντων. Η προσέγγιση αυτή οδηγεί στην εφαρμογή ιεραρχικών δέντρων για την διερεύνηση του φαινομένου. Επειδή αποτελεί αντικείμενο της εργασίας και η διερεύνηση των ατυχημάτων η ανάλυση έχει πραγματοποιηθεί στο σύνολο του αρχείου δεδομένων που προέκυψε από τον καθαρισμό όπου η ολική απώλεια αντιστοιχεί στο 7.3 % των εγγραφών.

Κρίθηκε σκόπιμο καταρχήν να αναφερθούν οι μεταβλητές οι οποίες χρησιμοποιήθηκαν για την δημιουργία δένδρων με χρήση του αλγορίθμου και να δοθούν βασικά στοιχεία σε σχέση με τον τύπο και το είδος τους στον πίνακα 6.17.

Πίνακας 6.17

Συνοπτική παρουσίαση τύπου και είδους των μεταβλητών

Πεδία / Μεταβλητές (Field)	Τύπος (Type)	Είδος μεταβλητής (Direction)
		OUT →Εξαρτημένη μεταβλητή
TOTLOSS	Flag	IN
ACCIDE0	Set	IN
AGE	Integer Range	IN
AREATYPE	Set	IN
BUILT1	Integer Range	IN
CLASS	Set	IN
DATYR	Integer Range	IN
GRT1	Integer Range	IN
GRT1GROU	Set	IN
GRUCLASS	Set	IN
MONTH	Set	IN



Πεδία / Μεταβλητές (Field)	Τύπος (Type)	Είδος μεταβλητής (Direction)
NAREA	Set	IN
NOFACC	Real Range	IN
OWNERS	Set	IN
SEQACC	Integer Range	IN
TYPE	Set	IN
WEATH	Set	IN
AA	Integer Range	NONE (*1)
ACCSER	Real Range	NONE (*6)
BADVOYAG	Flag	NONE (*5)
DACC92	Real Range	NONE (*2)
DAY92	Real Range	NONE (*2)
FLAG	Set	NONE (*4)
OWNER	Real Range	NONE (*3)

Ο λόγος που δε χρησιμοποιήθηκαν τα παραπάνω πεδία/ μεταβλητές είναι:

\*1→ Είναι το κλειδί του αρχείου δεδομένων.

\*2→ Δεν προσφέρει πληροφορία στην ανάλυση του δέντρου.

\*3→ Η OWNERS μεταφέρει σχεδόν στον επιθυμητό βαθμό την επιθυμητή πληροφορία άρα δεν χρειάζεται η OWNER για την ανάλυση.

\*6→ Η ACCSER είναι ενδογενής πληροφορία του αρχείου δεδομένων και είναι συνάρτηση άλλων μεταβλητών.

\*4 → Η ύπαρξη της μαζί με τις άλλες μεταβλητές συσχετιζόμενη πλήρως με αυτές δεν επιτρέπει στον αλγόριθμο να παράγει αποτέλεσμα (error).

\*5 → Είναι άμεση συνάρτηση της ολικής απώλειας και εκτιμάται με κριτήρια ανάλογα με το ζητούμενο κάθε φορά.

Παρουσιάζονται στη συνέχεια τα αποτελέσματα από την εκτέλεση του αλγόριθμου και μάλιστα από τους διαχωρισμούς που εμφανίζονται στο δένδρο απόφασης μπορούμε να κάνουμε παρατηρήσεις, ενώ το συνολικό δέντρο Δ<sub>1</sub> παρουσιάζεται στο Παράρτημα Δ.

Πρώτος διαχωριστικός παράγοντας είναι το είδος του ατυχήματος. Έτσι διαφοροποιούνται τα ατυχήματα με ενδεικτική τιμή FOUNDERED, με την παρατήρηση που προαναφέρθηκε και στο τέταρτο κεφάλαιο ότι στην ομάδα

FOUNDNERED εμπεριέχονται υποομάδες όπως CAPSIZED κ.α., από τα υπόλοιπα ατυχήματα.

ACCIDE0 ['FOUNDERED'] [Mode: Yes] (2958)

η

ACCIDE0 ['COLLISION (WITH SHIP)' 'CONTACT (INC. STRIKING ICE)'  
'FIRE/EXPLOSION' 'GROUNDED' 'HULL/MCHY/EQUIP.DAMAGED'  
'MISSING'] [Mode: No] (24706,0.949) -> No

Ακολούθως ενώ τα λοιπά ατυχήματα δεν διαφοροποιούνται για τα πλοία με ατύχημα FOUNDERED γίνεται ακόμα ένας διαχωρισμός ο οποίος τα διαφοροποιεί ως προς την ολική απώλεια. Αυτός ο διαχωρισμός γίνεται με βάση την χωρητικότητα του πλοίου. Το μέγεθος αυτό όμως είναι συνεχές οπότε ο αλγόριθμος διαπιστώνει ότι πλοία με χωρητικότητα (grt) πάνω από 16518 έχουν αναλογικά μικρότερη πιθανότητα να συμβεί σε αυτά ολική απώλεια από ότι στα μικρότερα.

ACCIDE0 ['FOUNDERED'] [Mode: Yes] (2958)

GRT1 =< 16518 [Mode: Yes] (2356)

η

GRT1 > 16518 [Mode: No] (602)

Με ακόμα μεγαλύτερη λεπτομέρεια για τα μικρά πλοία, παρατηρείται ότι σημαντικό ρόλο έχει και η ευρύτερη γεωγραφική περιοχή που έχει παρατηρηθεί το ατύχημα.

ACCIDE0 ['FOUNDERED']

GRT1 =< 16518 [Mode: Yes] (2356.14)

NAREA [A:'ATLANTIC' 'AUSTRALIA' 'CANADIAN ALASKA,  
U.S.S.R. ARTIC & BERNING SEA, ICELAND' 'PACIFIC'  
'MEDITERRANEAN, SUEZ' 'N. SEA, BALTIC, KIEL'  
'W. COAST S - N. AMERICA, PANAMA, CAPE HORN'  
'W.INDIES U. S. EASTERN, GULF OF MEXICO,  
GREAT LAKES, NEWFOUNDLAND'] [Mode: No] (1128)

η

NAREA [B: 'GULF, BAY OF BENGAL' 'INDIAN OCEAN &  
ANTARCTIC' 'JAPAN S CHINA & E. INDIES' 'RED  
SEA E.& W. AFRICAN COAST']  
[Mode: Yes] (1228)

Δηλαδή, ο αλγόριθμος χώρισε τώρα σε δύο ομάδες τις γεωγραφικές περιοχές του ατυχήματος. Οι γεωγραφικές περιοχές «υψηλού κινδύνου» Β έχουν σαν επικρατούν γεγονός την ολική απώλεια.

Θα σχολιασθεί ακόμη ότι ο τέταρτος κατά σειρά διαχωρισμός γίνεται σε σχέση με τον μήνα και μάλιστα για τους μήνες AUG, DEC, FEB, JUL, MAY και

ACCIDE0 ['FOUNDERED']

GRT1 =< 16518 [Mode: Yes] (2356.14)

NAREA [B: 'GULF, BAY OF BENGAL' 'INDIAN OCEAN &  
ANTARCTIC' 'JAPAN S CHINA & E. INDIES' 'RED SEA  
E.& W. AFRICAN COAST'] [Mode: Yes] (1228)

MONTH ['AUG' 'DEC' 'FEB' 'JUL' 'MAY' 'NOV' 'OCT']

[Mode: Yes](730.393, 0.703) -> Yes

η MONTH ['APR' 'JAN' 'JUN' 'MAR' 'SEP'] [Mode: Yes] (497.432)

NOV, OCT το πλήθος περιπτώσεων που καλύπτει ο κανόνας (coverage) είναι κατά προσέγγιση 730 και 0,703 η ακρίβεια (confidence) του κανόνα.

Με βάση τους ομαδοποιημένους νηογνώμονες γίνεται ο πέμπτος κατά

ACCIDE0 ['FOUNDERED']

GRT1 =< 16518 [Mode: Yes] (2356.14)

NAREA [B: 'GULF, BAY OF BENGAL' 'INDIAN OCEAN &  
ANTARCTIC' 'JAPAN S CHINA & E. INDIES' 'RED SEA  
E.& W. AFRICAN COAST'] [Mode: Yes] (1228)

MONTH ['APR' 'JAN' 'JUN' 'MAR' 'SEP'] [Mode: Yes] (497.432)

GRUCLASS ['AB-GL-NK' 'BV-CC-KR-NV-RI-PR-RS' 'LR']

[Mode: No] (302.39)

η GRUCLASS ['OTHERS'] [Mode: Yes] (195.042)



σειρά διαχωρισμός σε δύο ομάδες, στην ομάδα που περιλαμβάνει τους AB-GL-NK-BV-CC-KR-NV-RI-PR-RS-LR και στην ομάδα που περιλαμβάνει όλους τους υπόλοιπους νηογνώμονες (OTHERS) για τους οποίους έχουμε σαν επικρατούν γεγονός την ολική απώλεια.

Οι επόμενες ομαδοποιήσεις, για τα πλοία τα ανήκοντα στους λουπούς νηογνώμονες, γίνονται με βάση τις μεταβλητές πλήθους της αλλαγής ιδιοκτητών ομαδοποιημένο, της ακολουθίας εγγραφών ανά πλοίο και της χρονολογίας ατυχήματος, όπως παρουσιάζεται στο συνολικό δέντρο Δ<sub>1</sub> (Παράρτημα Δ) που προσφέρει ακόμα μεγαλύτερη και λεπτομερή ανάλυση.

Το παραπάνω δέντρο παράγει τον καλύτερο δυνατό ταξινομητικό κανόνα για το σύνολο του αρχείου δεδομένων. Ο Πίνακας 6.18 συγκρίνει την εκτιμώμενη τιμή της ολικής απώλειας, με βάση τη συναρτησιακή μορφή του δέντρου, με την πραγματική [Quinlan J. R., 1986]. Όπως φαίνεται ο κανόνας δουλεύει σωστά στο 92,44% των εγγραφών. Η αποτελεσματικότητα του κανόνα δεν είναι ίδια όμως στις τιμές της ολικής απώλειας. Ενώ ο κανόνας δουλεύει σχεδόν τέλεια με ποσοστό 98 % στην πρόγνωση του γεγονότος ότι για ένα πλοίο δε θα συμβεί ολική απώλεια δεν ισχύει το ίδιο για την πρόβλεψη του αντίθετου γεγονότος, κάτι που συμβαίνει ούτε για το 50% των περιπτώσεων.

Αξιίζει εδώ να σημειωθεί ότι το 42,5% παραμένει παρόλα αυτά ένα πολύ καλό ποσοστό συγκρινόμενο με άλλες μεθόδους όπως αυτό αποδεικνύεται στη συνέχεια.

ΠΙΝΑΚΑΣ 6.18

Αξιοπιστίας ταξινόμησης του δένδρου Δ<sub>1</sub>

	Σύνολο	Ολική απώλεια		Ολική απώλεια	
		Οχι	Ναι	Ναι	Ναι
Σωστό	25573 ( 92.44%)	24409 ( 97.94%)	1164 ( 42.45%)		
Λάθος	2091 ( 7.56%)	513 ( 2.06%)	1578 ( 57.55%)		

Αν το σχετικό κόστος σφάλματος ταξινόμησης των περιπτώσεων ολικής απώλειας σε μη ολική απώλεια οριστεί να είναι τριπλάσιο από όλα τα άλλα [Fawcett T., 2001] και [Quinlan, J. R., 1986], παράγεται ένα διαφορετικό δέντρο με μικρότερη

γενική προγνωστική ικανότητα, όπως φαίνεται στον Πίνακα 6.19 το οποίο προβλέπει όμως καλύτερα τις περιπτώσεις ολικής απώλειας.

ΠΙΝΑΚΑΣ 6.19  
Αξιοπιστίας ταξινόμησης του δένδρου Δ<sub>2</sub>

	Σύνολο	Ολική		Ναι ολ.	
		απώλεια	Όχι	απώλεια	Ναι
Σωστό	25395( 91.80%)	23190( 93.05%)		2205( 80.42%)	
Λάθος	2269( 8.20%)	1732( 6.95%)		537( 19.58%)	

Τώρα παρατηρείται μία σχεδόν 100% αύξηση της «απόδοσης» του μοντέλου για τις περιπτώσεις της ολικής απώλειας «Ναι» σε βάρος βέβαια της απόδοσής του για τις περιπτώσεις της όχι ολικής απώλειας «Όχι». Η συνολική απόδοση του μοντέλου είναι συνολικά ελάχιστα μικρότερη.

Το δέντρο Δ<sub>2</sub> που παράγεται παρουσιάζεται συνολικά στο Παράρτημα Δ και μοιάζει αρκετά με το βασικό Δ<sub>1</sub> αφού χωρίζει ξανά τις περιπτώσεις με βάση το ατύχημα ACCIDE0. Πρώτος διαχωριστικός παράγοντας είναι πάλι το είδος του ατυχήματος. Έτσι διαφοροποιούνται τα ατυχήματα με ενδεικτική τιμή FOUNDERED, από τα υπόλοιπα ατυχήματα.

ACCIDE0 ['COLLISION (WITH SHIP)' 'CONTACT (INC. STRIKING ICE)'  
'FIRE/EXPLOSION' 'GROUNDED' 'MISSING'  
'HULL/MCHY/EQUIP.DAMAGED'] [Mode: No] (24707.29)

η  
ACCIDE0 ['FOUNDERED'] [Mode: Yes] (2957.98)

Αλλά τώρα πια διαχωρίζονται και τα λοιπά ατυχήματα COLLISION (WITH SHIP), CONTACT (INC. STRIKING ICE), FIRE-EXPLOSION, GROUNDED, MISSING, HULL/MCHY/EQUIP. DAMAGED έτσι ώστε να ανιχνευτούν καλύτερα όλες οι εναλλακτικές περιπτώσεις.

Τα λοιπά ατυχήματα διαχωρίζονται σε FIRE - EXPLOSION, MISSING, όπως φαίνεται αναλυτικά στη συνέχεια, και COLLISION (WITH SHIP), CONTACT (INC. STRIKING ICE), GROUNDED, HULL/MCHY/EQUIP. DAMAGED.

ACCIDE0 ['COLLISION(WITH SHIP)' 'CONTACT(INC. STRIKING ICE)' 'FIRE-  
EXPLOSION' 'GROUNDED' 'MISSING'

'HULL/MCHY/EQUIP.DAMAGED'] [Mode: No] (24707.29)

ACCIDE0 ['COLLISION(WITH SHIP)' 'CONTACT(INC. STRIKING ICE)',  
'GROUNDED' 'HULL/MCHY/EQUIP.DAMAGED']

[Mode: No] (21589.196)

η

ACCIDE0 ['FIRE-EXPLOSION' 'MISSING'] [Mode: No] (3120.009)

Ενδιαφέρον στο δένδρο Δ<sub>2</sub> είναι ότι και τώρα η χωρητικότητα (grt) διαχωρίζεται στα 16518 το οποίο θεωρείται σαν φυσικό, αφού τα δύο δέντρα είναι συγγενή.

ACCIDE0 ['FOUNDERED'] [Mode: Yes] (2957.98)

GRT1 =< 16518 [Mode: Yes] (2356.12)

η

GRT1 > 16518 [Mode: Yes] (601.861)

#### 6.4 Αποτελέσματα από την ανάλυση των δεδομένων του αρχείου με ταξινομητικούς κανόνες

Μελετώντας το βασικό δέντρο με κατεύθυνση από τα φύλλα έως τη ρίζα του, ενώ αποκτάμε επίγνωση για τη σημαντικότητα των διαφόρων μεταβλητών στην ταξινόμηση ενός γεγονότος σε "Ναι" ή "Όχι", δεν υπάρχει τελικά εύκολα απευθείας γνώση του ίδιου του τελικού γεγονότος. Ο ρόλος των κανόνων ταξινόμησης είναι να παρουσιάσουν για κάθε ένα φύλλο, τον κανόνα ο οποίος οδηγεί σε αυτό, με γλώσσα λογικής συνόλων της μορφής:

«EAN έχει συμβεί το X»

ΚΑΙ

«EAN έχει συμβεί το Y»

TOTE

με πιθανότητα ζ%

θα συμβεί το Z.



Από το βασικό δέντρο, σχηματίζονται 23 κανόνες για την ολική απώλεια και 26 κανόνες για την μη ολική απώλεια. Οι κανόνες αυτοί παρουσιάζονται στο Παράρτημα Τ, που αποτελείται από το Παράρτημα Τ<sub>Yes</sub> και το Παράρτημα Τ<sub>No</sub> για την ολική και μη ολική απώλεια αντίστοιχα. Οι κανόνες αυτοί καταφέρνουν να ενώσουν πολλές φαινομενικά ανεξάρτητες μεταξύ τους περιπτώσεις ώστε να παραχθούν συμπεράσματα σε ευρύτερο επίπεδο. Οι κανόνες για την ολική απώλεια είναι 23 διαφορετικά σενάρια, τα οποία οδηγούν με σημαντικά μεγάλες πιθανότητες στην ολική απώλεια του πλοίου. Ο Πίνακας 6.20 παρουσιάζει αυτές τις περιπτώσεις κατά φθίνουσα πιθανότητα.

ΠΙΝΑΚΑΣ 6.20

Πιθανότητα ολικής απώλειας ανά κανόνα

Κανόνας	Πλήθος	Πιθανότητα ολικής απώλειας
17	11	1,00
14	11	0,93
16	12	0,83
22	11	0,82
2	32	0,81
12	14	0,78
11	27	0,78
1	21	0,76
20	12	0,75
5	48	0,73
18	731	0,70
7	20	0,70
15	103	0,70
13	25	0,69
21	13	0,69
4	14	0,64
19	25	0,64
8	47	0,64
6	290	0,64
23	131	0,63
10	87	0,62

Αυτό που είναι εξίσου σημαντικό είναι η διερεύνηση των συστατικών των κανόνων που οδηγούν στα συμπεράσματα. Αυτή περιλαμβάνει τα εξής διακριτά συστατικά:

- Τις μεταβλητές που συμμετέχουν στον κανόνα.
- Τις συγκεκριμένες τιμές με τις οποίες συμμετέχουν οι μεταβλητές στον κανόνα.

Ένας σχολιασμός που μπορεί να γίνει για τον κανόνα «17», που ακολουθεί είναι ότι σε όλες τις περιπτώσεις των συνθηκών που συνθέτουν τον κανόνα αυτόν που παρουσιάζεται μαζί με όλους τους υπόλοιπους στο Παράρτημα Yes έχει καταγραφεί ολική απώλεια.

Rule #17 for Yes:

```
if ACCIDE0 = ['FOUNDERED']
and GRT1 <= 16518
and NAREA = ['GULF, BAY OF BENGAL' 'INDIAN OCEAN &
ANTARTIC' 'JAPAN, S. CHINA & E. INDIES' 'RED SEA
E.& W. AFRICAN COAST']
and MONTH = ['APR' 'JAN' 'JUN' 'MAR' 'SEP']
and GRUCLASS = ['OTHERS']
and OWNERS = ['2', '5***']
then -> Yes (10.593, 1.0)
```

Ο κανόνας «14», που ακολουθεί επιβεβαιώνει τα παρατηρηθέντα και στην αρχή του παρόντος υποκεφαλαίου, σχετικά με την ιεράρχηση των μεταβλητών ατύχημα, χωρητικότητα, γεωγραφική περιοχή, μήνας, ομαδοποιημένοι νηογνώμονες, πλήθος αλλαγής ιδιοκτητών ομαδοποιημένο, ακολουθία εγγραφών ανά πλοίο, χρονολογία του ατυχήματος καθώς και τον διαχωρισμό των τιμών τους.

Rule #14 for Yes:

```
if ACCIDE0 = ['FOUNDERED']
and GRT1 <= 16518
and NAREA = ['GULF, BAY OF BENGAL' 'INDIAN OCEAN &
ANTARTIC' 'JAPAN, S. CHINA & E. INDIES' 'RED SEA
E.&W. AFRICAN COAST']
and MONTH = ['APR' 'JAN' 'JUN' 'MAR' 'SEP']
```

```

and GRUCLASS = ['OTHERS']
and OWNERS = ['1', 'No info']
and SEQACC =< 1
and DATYR > 93
and DATYR =< 95
and MONTH = ['JAN']
then -> Yes (10.779, 0.926)

```

Ένα ακόμη πλεονέκτημα από την εφαρμογή της μεθόδου είναι η δυνατότητα που προσφέρεται από τις συγκρίσεις που γίνονται μεταξύ «συμπληρωματικών κανόνων». Για παράδειγμα ο κανόνας «2» του Παραρτήματος T<sub>No</sub> για μη ολική απώλεια απλοποιημένος έχει τη μορφή :

Rule #2 for No:

```

if ACCIDE0 = ['FOUNDERED']
and GRT1 =< 16518
and NAREA = ['T: ίδιες με τον Rule #1 for Yes']
and AREATYPE = ['CONTROLLED' 'PORT']
and NOFACC =< 1
then -> No (...)

```

Ενώ ο ταξινομητικός κανόνας «1» του Παραρτήματος T<sub>Yes</sub> για την ολική απώλεια έχει την ακόλουθη μορφή :

Rule #1 for Yes:

```

if ACCIDE0 = ['FOUNDERED']
and GRT1 =< 16518
and NAREA = ['T: ίδιες με τον Rule #2 for No']
and AREATYPE = ['CONTROLLED' 'PORT'] and ...
and NOFACC > 1 and ...
then -> Yes (...)

```

Από την αντιπαράθεση των δύο κανόνων μπορούμε να βγάλουμε συμπεράσματα. Από τους ανωτέρω κανόνες λοιπόν συμπεραίνουμε ότι παρόλο που



και στις δύο περιπτώσεις έχουμε ίδιες τιμές για το ατύχημα, την χωρητικότητα, την Γεωγραφική περιοχή και το είδος της, η πρώτη καταγραφή ενός ατυχήματος πλοίου NOFACC  $\leq 1$  στο αρχείο δεδομένων δεν συνοδεύεται αυτόματα με ολική απώλεια. Στην περίπτωση όμως ύπαρξης και δεύτερης καταγραφής ατυχήματος NOFACC  $> 1$  αυξάνει απότομα η πιθανότητα για ολική απώλεια.

#### 6.5 Αποτελέσματα από την ανάλυση του αρχείου δεδομένων με άλλες μεθόδους

Η ικανότητα ενός έμπειρου συστήματος να προβλέπει την ολική απώλεια ενός πλοίου μπορεί να υπολογιστεί και με τη βοήθεια των νευρωνικών δικτύων. Τα νευρωνικά δίκτυα [Barlow H., 1989] αποτελούν μη γραμμικά συστήματα που έχουν τη δυνατότητα να εκτιμούν το προς διερεύνηση γεγονός με καλό ποσοστό. Ο λόγος που δε χρησιμοποιείται ευρύτατα ως τεχνική είναι γιατί τα αποτελέσματα εξαιτίας αυτής της μη γραμμικότητας δεν μπορούν εύκολα να ερμηνευτούν. Αυτό που είναι εύκολο να ερμηνευτεί και αποτελεί ένα ενδιαφέρον στοιχείο για την παρούσα διατριβή είναι η ανάλυση αξιοπιστίας.

Η ανάλυση αξιοπιστίας δίνει τη σχετική βαρύτητα στην μη γραμμική συνάρτηση εκτίμησης του προς διερεύνηση μεγέθους, στην περίπτωση μας της ολικής απώλειας [Barlow H., 1989] και [Hand D. et al., 2001]. Βάζει σε σχέση διάταξης τα μέσα βάρη από τις δοθείσες στην είσοδο του δέντρου (INPUT) μεταβλητές (παράγοντες) παίρνοντας υπόψη το σύνολο των βαρών που υπάρχουν στον πληθυσμό όλων των εξωτερικών ανεξάρτητων μεταβλητών.

Ο πίνακας 6.21 παρουσιάζει σε φθίνουσα σειρά βαρύτητας τις κυριότερες μεταβλητές που χρησιμοποιήθηκαν για την πρόβλεψη της ολικής απώλειας, πρόβλεψη που επιτεύχθηκε με ακρίβεια 91.6%. Στη πρώτη στήλη αναφέρονται οι μεταβλητές (παράγοντες) ενώ στη δεύτερη στήλη για κάθε μια μεταβλητή παρουσιάζεται η σχετική βαρύτητα της.

Ενδιαφέρον αποτελεί από την παρατήρηση του πίνακα ότι το είδος ατυχήματος είναι μόλις στην πέμπτη θέση ενώ το πόσα διαφορετικά ατυχήματα έχουν συμβεί δε φαίνεται να παίζει κανένα σημαντικό ρόλο. Αντίθετα ο νηογώμονας και το σχετικό μέγεθος του πλοίου φαίνεται να παίζουν σημαντικό ρόλο. Σε αντιπαραβολή η χωρητικότητα από μόνη της δε φαίνεται να ασκεί κάποιο σημαντικό ρόλο.

## ΠΙΝΑΚΑΣ 6.21

Σχετική βαρύτητα παραγόντων για πρόβλεψη της ολικής απώλειας

Παράγοντας / μεταβλητή	Σχετική βαρύτητα
Νηογώμονας	0,15
Σχετικό μέγεθος	0,11
Μήνας	0,10
Γεωγραφική περιοχή	0,09
Ατύχημα	0,09
Καιρικές συνθήκες	0,08
Είδος περιοχής	0,07
Πλήθος αλλαγής ιδιοκτητών ομαδοποιημένο	0,06
Τύπος πλοίου	0,05
Πλήθος ατυχημάτων	0,03
Χρονολογία ατυχήματος	0,03
Έτος ναυπήγησης	0,02
Χωρητικότητα	0,02

Το αποτελέσματα αυτά δεν έρχονται σε αντίθεση με τα παραχθέντα από δέντρα αποφάσεων, πρόκειται για μία τελείως διαφορετική προσέγγιση και τεχνική, γιατί τα νευρωνικά δίκτυα μελετάνε τα δεδομένα συνολικά λαμβάνοντας υπόψη οποιοσδήποτε συντελεστής στάθμισης, ενώ τα δέντρα αποφάσεων παράγουν πιο κοντά στην πραγματικότητα κατανοητά συμπεράσματα για την πρόβλεψη.

Ένα σημείο αξιο προσοχής είναι ότι όλη η έρευνα είναι μεροληπτική υπέρ της μη εμφάνισης ολικής απώλειας του πλοίου λόγω της σχετικά μικρής εμφάνισης της. Αυτό από τη μία δίνει ακριβή συμπεράσματα από την άλλη όμως δεν παρουσιάζει τη σχετική εικόνα της βαρύτητας των παραγόντων στην ολική απώλεια. Με χρήση της μεθόδου Boosting [Drucker H., 1999] πολλαπλασιάζοντας τις περιπτώσεις ολικής απώλειας με ένα συντελεστή της τάξης του 12 (12,778455) το αρχείο «εξισορροπείται» (“balanced”) αναφορικά με την ολική απώλεια αναιρώντας έτσι την επίδραση του μεγέθους. Με δεδομένο ότι έχουμε 18861 εγγραφές για την μη ολική απώλεια και 1476 εγγραφές για την ολική απώλεια προτάθηκε από τον αλγόριθμο σαν συντελεστής για το βέλτιστο “boosting” των τιμών των μεταβλητών το

12,778455 διότι  $12,78 \cdot 1476 = 18863$ , αριθμός ο οποίος προσεγγίζει ικανοποιητικά τον αριθμό των εγγραφών για την μη ολική απώλεια.

Με βάση αυτή την μέθοδο μπορούμε να μελετήσουμε την επίδραση ενός υποσυνόλου παραγόντων [Hand D. et al, 2001] των NAREA - BUILT1 - OWNERS - TYPE - GRUCLASS, (Πίνακας 6.22), στην ολική απώλεια TOTLOSS.

Πίνακας 6.22

Συνοπτική παρουσίαση τύπου και είδους μεταβλητών που επελέγησαν

Πεδίο/μεταβλητή (Field)	Τύπος (Type)	Είδος μεταβλητής (Direction)
TOTLOSS	Flag	OUT
NAREA	Set	IN
BUILT1	Integer Range	IN
OWNERS	Set	IN
TYPE	Set	IN
GRUCLASS	Set	IN

Η εκτέλεση του αλγόριθμου παράγει το δέντρο  $\Delta_3$  που παρουσιάζεται συνολικά στο Παράρτημα Δ. Για τους διαχωρισμούς που εμφανίζονται στο δένδρο απόφασης Δένδρο  $\Delta_3$  μπορούμε να κάνουμε παρατηρήσεις.

Στον δένδρο  $\Delta_3$ , ανεξαρτήτως ατυχήματος, η μεταβλητή η οποία τα διαφοροποιεί ως προς την ολική απώλεια είναι οι ομαδοποιημένοι νηογνώμονες GRUCLASS και μάλιστα για τιμές A: AB-GL-NK BV-CC-KR-NV-RI-PR-RS LR και για τιμές B: OTHERS.

GRUCLASS A: ['AB-GL-NK' 'BV-CC-KR-NV-RI-PR-RS' 'LR'] [Mode: No]

η

GRUCLASS B: ['OTHERS'] [Mode: Yes]

Αξιοσημείωτο είναι ότι η ομαδοποίηση των τιμών της μεταβλητής ομαδοποιημένοι νηογνώμονες δεν έγινε με κριτήριο τους ανήκοντες στον Διεθνή Σύνδεσμο Νηογνώμωνων (International Assosiation of Classification Societies) και



στους υπόλοιπους νηογνώμονες μη ανήκοντες στον Διεθνή Σύνδεσμο Νηογνώμωνων αλλά αντίθετα χωρίστηκαν σε τέσσερες κλάσεις όπως αυτές ορίστηκαν στην παράγραφο 4.3.16. Ο αλγόριθμος εν τούτοις διαχώρισε τους ομαδοποιημένους νηογνώμονες σε δύο ομάδες. Η ομάδα Α έχει τρεις τιμές για την μεταβλητή αυτή, τις AB-GL-NK και BV-CC-KR-NV-RI-PR-RS και LR, που είναι οι νηογνώμονες που ανήκουν στον Διεθνή Σύνδεσμο Νηογνώμωνων, και δεν οδηγεί σε ολική απώλεια. Αντιθέτως η ομάδα Β έχει τιμή για την μεταβλητή αυτή την OTHERS, δηλαδή τους υπόλοιπους νηογνώμονες και για επικρατούσα τιμή την ολική απώλεια.

Στην συνέχεια εισέρχεται μια ακόμα διαχωριστική μεταβλητή, για τους νηογνώμονες GRUCLASS A και μάλιστα ο σημαντικότερος παράγοντας που τους διαχωρίζει είναι η χρονιά ναυπήγησης (BUILT1). Αν το έτος ναυπήγησης είναι μικρότερο ή ίσο του 1974 τότε θα έχουμε αναλογικά μεγαλύτερη πιθανότητα να συμβεί ολική απώλεια παρά αν το έτος ναυπήγησης είναι μεγαλύτερο του 1974.

GRUCLASS A: ['AB-GL-NK' 'BV-CC-KR-NV-RI-PR-RS' 'LR'] [Mode: No]

BUILT1 =< 1974 [Mode: Yes]

η

BUILT1 > 1974 [Mode: No]

Για τους υπόλοιπους νηογνώμονες πέραν από τους ανήκοντες στον Διεθνή

GRUCLASS B: ['OTHERS'] [Mode: Yes]

NAREA ['ATLANTIC' 'AUSTRALIA' 'MEDITERRANEAN, SUEZ' 'N.SEA,  
BALTIC, KIEL' 'PACIFIC' 'W. COAST S - N. AMERICA,  
PANAMA, CAPE HORN' 'W.INDIES U. S. EASTERN, GULF  
OF MEXICO, GREAT LAKES, NEWFOUNDLAND'] [Mode: Yes]

η

NAREA ['GULF. BAY OF BENGAL' 'INDIAN OCEAN & ANTARCTIC'  
'JAPAN S CHINA & E. INDIES' 'SEA E.& W. AFRICAN COAST'  
'CANADIAN ALASKA,U.S.S.R.ARTIC & BERNING SEA,  
ICELAND'] [Mode: Yes]

Σύνδεσμο Νηογνώμωνων ο σημαντικότερος παράγοντας είναι η γεωγραφική περιοχή πλεύσης και ανεξάρτητα περιοχής όμως έχουμε μεγάλη πιθανότητα να συμβεί ολική

απώλεια εξαρτώμενη από τις τιμές των υπόλοιπων μεταβλητών όπως αυτές φαίνονται στο Παράρτημα Δ στο δένδρο Δ<sub>3</sub>.

Το παραπάνω δέντρο μπορεί να παρουσιαστεί περισσότερο ανοιγμένο [Quinlan J. R., 1986] από τη ρίζα μέχρι τα φύλλα και είναι το καλύτερο δέντρο με την έννοια ότι οποιοδήποτε άλλο δέντρο δε διαχωρίζει τόσο καλά την ολική απώλεια.

Σε περίπτωση που θέλουμε να παρατηρήσουμε όλα τα σχετικά δέντρα αρκεί να εκτελέσουμε εναλλακτικά τον αλγόριθμο G.R.I. που όπως αναφέρθηκε αποτελεί μία σειρά επαναλαμβανόμενων δέντρων αποφάσεων για όλα τα δυνατά σενάρια.

Για την εκτέλεση του αλγόριθμου επιλέχθηκαν οι παράμετροι TOTLOSS-NAREA-OWNERS-TYPE-GRUCLASS και η παράμετρος BUILT 74 (Πίνακας 6.23). Αυτή όπως προαναφέρθηκε στο υποκεφάλαιο 5.4 διαχωρίζει τα πλοία που έχουν ναυπηγηθεί μέχρι το έτος 1974 και μετά το 1974. Με σκοπό να παρατηρηθούν τα «παλιά» πλοία επιλέχθηκε η μεταβλητή BUILT 74, χαρακτηριζόμενη ως προς τον τύπο αυτής σαν «σημαία» (flag) και καταγράφει εάν έχει ναυπηγηθεί πριν ή μετά το 1974.

Πίνακας 6.23

Συνοπτική παρουσίαση τύπου και είδους μεταβλητών που επελέγησαν

Πεδίο / μεταβλητή (Field)	Τύπος (Type)	Είδος μεταβλητής (Direction)
TOTLOSS	Flag	OUT
NAREA	Set	IN
OWNERS	Set	IN
TYPE	Set	IN
GRUCLASS	Set	IN
BUILT 74	Flag	IN

Ο αλγόριθμος έχει εκτελεστεί για δέντρο βάθους το πολύ τρία για λόγους οικονομίας και απλότητας και με το 50% να είναι το μέτρο σύγκρισης για την ολική απώλεια [Drucker H., 1999] σύμφωνα με την Boosting μεθοδολογία.

Τα αποτελέσματα του αλγόριθμου και μάλιστα ο πλήρης συνδυασμός των παραγόντων παρουσιάζονται σε πίνακες στο Παράρτημα Π, ενώ εδώ αναφέρεται ένα μέρος των συνδυασμών.

Παράγοντες που ο συνδυασμός τους οδηγεί στην ολική απώλεια	percent
GRUCLASS = OTHERS & NAREA = JAPAN, S. CHINA & E. INDIES & TYPE = FERRY	0,85
GRUCLASS = OTHERS & OWNERS = No info & NAREA = JAPAN, S. CHINA & E. INDIES & TYPE = FERRY	0,84
TYPE = BULK & BUILT 74 < 1975 & OWNERS = No info & NAREA = GULF, BAY OF BENGAL	0,84
OWNERS = No info & NAREA = JAPAN, S. CHINA & E. INDIES & TYPE = FERRY	0,83
BUILT 74 < 1975 & GRUCLASS = OTHERS & OWNERS = No info & NAREA = GULF, BAY OF BENGAL	0,83
BUILT 74 < 1975 & OWNERS = No info & NAREA = GULF, BAY OF BENGAL	0,82
BUILT 74 < 1975 & NAREA = JAPAN, S. CHINA & E. INDIES & TYPE = FERRY	0,82
TYPE = BULK & GRUCLASS = OTHERS & OWNERS = No info & NAREA = GULF, BAY OF BENGAL	0,81
GRUCLASS = OTHERS & OWNERS = No info & NAREA = GULF, BAY OF BENGAL	0,81
TYPE = BULK & OWNERS = No info & NAREA = GULF, BAY OF BENGAL	0,81
OWNERS = No info & NAREA = GULF, BAY OF BENGAL	0,80
TYPE = BULK & BUILT 74 < 1975 & GRUCLASS = OTHERS & NAREA = GULF, BAY OF BENGAL	0,80
TYPE = BULK & BUILT 74 < 1975 & NAREA = JAPAN, S. CHINA & E. INDIES & OWNERS = 2	0,80



Στη συνέχεια από τα αποτελέσματα που προέκυψαν σχολιάζεται, για παράδειγμα, ο συνδυασμός:

TYPE	=	['BULK']	&
BUILT 74	<	1975	&
OWNERS	=	['No info']	&
NAREA	=	['GULF, BAY OF BENGAL']	

που οδηγεί στην ολική απώλεια με ποσοστό 0.84 %. Οι τιμές για το έτος ναυπήγησης, BUILT 74, μικρότερες του 1975, η τιμή για την μεταβλητή τύπος πλοίου, TYPE, είναι BULK, για το πλήθος αλλαγής ιδιοκτητών ομαδοποιημένο, OWNERS, να μην έχουμε πληροφορίες και οι τιμές για την μεταβλητή Γεωγραφική περιοχή, NAREA, είναι ο Περσικός κόλπος και κόλπος της Βεγγάλης.

Ο καλύτερος κανόνας με ποσοστό 0.85% για την ολική απώλεια συνδυάζει τα δεδομένα :

GRUCLASS	=	['OTHERS']	&
NAREA	=	['JAPAN, S. CHINA & E. INDIES']	&
TYPE	=	['FERRY']	

Το πλήθος των αποτελεσμάτων που προαναφέρθηκαν μετά την εφαρμογή των μεθοδολογιών στα δεδομένα μας, δίνει την δυνατότητα να εξάγουμε συμπεράσματα που παρουσιάζονται στο κεφάλαιο που ακολουθεί.

## ΚΕΦΑΛΑΙΟ 7

### 7.1 Εισαγωγή

Κατά την διαδικασία εκπόνησης της διατριβής ερευνήθηκε το κύριο ερώτημα που τέθηκε στο πρώτο κεφάλαιο και προέκυψαν παράπλευρα συμπεράσματα και παρατηρήσεις σε σχέση με την προετοιμασία του αρχείου δεδομένων καθώς και σχετικά με τον καθαρισμό του.

Για το αρχείο παρατηρήθηκε ότι λόγω της ημερολογιακής κατασκευής του ένα γεγονός μπορεί να αναφέρεται σε περισσότερες από μία εγγραφές. Κάθε νεότερη εγγραφή του ίδιου ατυχήματος διορθώνει, συμπληρώνει ή ανασκευάζει την προηγούμενη πληροφορία. Άρα μία εγγραφή πολλές φορές απαιτεί διορθώσεις, μεταβολές, για να είναι πλήρης. Επομένως, είναι απαραίτητη η προετοιμασία των δεδομένων, λόγω της φύσης τους, γεγονός που απαιτεί την δημιουργία και χρήση αλγορίθμων ώστε να ορισθεί η μοναδικότητα της κάθε εγγραφής. Επίσης για την διαδικασία καθαρισμού του αρχείου δεδομένων, όπως αυτός περιγράφηκε αναλυτικά στο πέμπτο κεφάλαιο, επινοήθηκαν αλγόριθμοι, προσαρμοσμένοι στις ιδιαιτερότητες του αρχείου, οι οποίοι έκαναν αντικατάσταση των ελλειπουσών τιμών όπου υπήρχε η δυνατότητα καθώς και αποκατάσταση επί το ορθόν των διαφορετικών τιμών.

Συνολικά εξήχθησαν χρήσιμα συμπεράσματα, από το πέμπτο κεφάλαιο, σχετικά με την μεθοδολογία κατασκευής και καθαρισμού παρόμοιων αρχείων δεδομένων στο μέλλον, μια και ο καθαρισμός και κατασκευή αρχείου δεδομένων δεν έχει ακόμα εξελιχθεί σε τυποποιημένη διαδικασία.

Πρόβλημα δημιουργείται από τις ολικές απώλειες που αναφέρονται πολλές φορές σε ύστερο χρόνο από τον χρόνο που συνέβη το ατύχημα για μια σειρά από λόγους. Ένα πλοίο μπορεί να μην έχει δηλωθεί σαν ολική απώλεια σε μία εγγραφή και να έχει δηλωθεί σε κάποια επόμενη εγγραφή. Απαραίτητες είναι λοιπόν οι πληροφορίες και η διασταύρωση με μια δεύτερη πηγή που να έχει περισσότερο ενημερωμένα στοιχεία καθώς και ο έλεγχος αξιοπιστίας της διασταύρωσης αυτής, όπως έγινε στην παρούσα διατριβή, ώστε να έχουμε ακριβείς πληροφορίες.

Σημειώνεται τέλος, ότι σε σύνολο 20337 διαφορετικών εγγραφών εμφανίζονται 17081 διαφορετικά πλοία, δηλαδή ποσοστό 16 % των πλοίων επανεισφέρει σε άλλη χρονική στιγμή ατυχήματα.

## 7.2 Συμπεράσματα και σχόλια από την ανάλυση των μεταβλητών

Από την πληθώρα σχέσεων που καταγράφηκαν και αναδείχθηκαν στο υποκεφάλαιο 6.2, στο σύνολο των εγγραφών του αρχείου, αναφέρονται στη συνέχεια οι σημαντικότερες από την ανάλυση των πρωτογενών στοιχείων όπως:

- Η μηχανολογική βλάβη HULL/MCHY/EQUIP.DAMAGED είναι η συχνότερα παρατηρούμενη αιτία ναυτικών ατυχημάτων, που παρατηρείται στο ένα τρίτο των περιπτώσεων. Τούτο δείχνει την διαχειριστική πολυπλοκότητα όσον αφορά την πρόληψη των αστοχιών στην κατασκευή και τον εξοπλισμό του πλοίου. Επίσης, αναδεικνύει τις συνέπειες τις ανεπαρκούς συντήρησης.

- Περισσότερα ατυχήματα έχουν καταγραφεί τον Ιανουάριο που συνδυάζεται με την υψηλή κυκλοφορία και τις καιρικές συνθήκες στο Βόρειο Ημισφαίριο. Λιγότερα τον Απρίλιο για το Βόρειο και τον Σεπτέμβριο για το Νότιο Ημισφαίριο.

- Οι γεωγραφικές περιοχές: N. SEA, BALTIC, KIEL και JAPAN, S. CHINA & E. INDIES καθώς και W. INDIES, U.S. EASTERN, GREAT LAKES, NEWFOUNDLAND, GULF OF MEXICO, και συγκεντρώνουν, αναμενόμενο λόγω υψηλής κυκλοφορίας, περισσότερα από τα μισά ατυχήματα.

- Από την συσχέτιση του χρονικού παράγοντα με την Γεωγραφική περιοχή φαίνεται ότι στον Ινδικό τα ατυχήματα έχουν παρατηρηθεί από το Φεβρουάριο έως και τον Απρίλιο εποχή κατά την οποία επικρατούν στην περιοχή ιδιόμορφες καιρικές συνθήκες.

- Τέλος, το ποσοστό των περιπτώσεων ολικής απώλειας στο σύνολο των εγγραφών για το χρονικό διάστημα από 1992 έως 1999 ανέρχεται σε 7,3%.

Η διερεύνηση της ολικής απώλειας έχει εκτιμηθεί ως ο βασικός στόχος της παρούσας εργασίας. Μελετώντας την ολική απώλεια με βάση ορισμένους άλλους ανεξάρτητους παράγοντες από πίνακες του υποκεφαλαίου 6.2 και τους πίνακες που παρατίθενται στο Παράρτημα Π2 προκύπτουν τα εξής ενδιαφέροντα συμπεράσματα :

Εμφανίζονται μεγαλύτερα ποσοστά ολικής απώλειας στις περιπτώσεις που :

- Έγινε στον Ινδικό Ωκεανό στον οποίο έχουμε ως γνωστό τον υψηλότερο μέσο κυματισμό άρα και τουλάχιστον την μεγαλύτερη καταπόνηση των σκαφών.

- Έχει συμβεί στα πλοία που έχουν αλλάξει πολλούς ιδιοκτήτες. Στον βαθμό που η «αλλαγή χεριών» συμβαδίζει με την διαχειριστική και λειτουργική ασυνέχεια



αναδεικνύεται η έλλειψη μεταφοράς εμπειριών, κατά την αλλαγή των ιδιοκτητών, που αφορούν την διαχείριση (γραφεία) και λειτουργία (πλήρωμα) του πλοίου.

- Είναι πλοία με νηογνώμονα BIRO KLASIFICASI INDONESIA.
- Ο καιρός που επικρατούσε ήταν STORM και TYPHOON.
- Έχει συμβεί στις ανοιχτές θάλασσες.
- Έχει συμβεί σε πλοία με ατύχημα ενδεικτική τιμή FOUNDERED.

Εμφανίζονται μικρότερα ποσοστά ολικής απώλειας στις περιπτώσεις που το ατύχημα συνέβη :

- Τον Σεπτέμβριο και Μάρτιο που είναι Άνοιξη στα δύο Ημισφαίρια.
- Στην γεωγραφική περιοχή της Αυστραλίας.
- Στα πλοία τα ανήκοντα στον Αμερικανικό Νηογνώμονα (AB) και μάλιστα έχουμε πολύ χαμηλά ποσοστά ολικής απώλειας για αυτόν. Παρατηρούμε γενικά ότι οι νηογνώμονες διαφοροποιούνται μεταξύ τους ως προς αυτήν.

Άλλα αξιοσημείωτα συμπεράσματα συσχέτισης παραγόντων με την ολική απώλεια που προέκυψαν από την ανάλυση στο υποκεφάλαιο 6.2 είναι:

- Ο μέσος όρος ηλικίας των πλοίων στα οποία είχαμε ολική απώλεια είναι 23 έτη σε σχέση με 17 έτη για τα πλοία που δεν είχαμε. Σε μεγαλύτερη ηλικία έχουμε μειωμένη ανθεκτικότητα και αντοχή του πλοίου σε πιθανό ατύχημα και στην πρόληψη παραγώνων συνεπειών με την μορφή επακόλουθων ατυχημάτων που οδηγούν σε ολική απώλεια. Στην περίπτωση αυτή βέβαια πρέπει να ληφθεί υπόψη και η παρατήρηση της παραγράφου 4.3.15 στην οποία σημειώνεται ότι υπάρχει ενδεχόμενο ένα πλοίο να δηλώνεται σαν ολική απώλεια για άλλους λόγους (Giziakis K, 1996)

• Από την συσχέτιση έτους ναυπήγησης με την ολική απώλεια αναδείξαμε ότι όσο παλαιότερο, από το 1983, έτος ναυπήγησης έχει το πλοίο τόσο μεγαλύτερα ποσοστά ολικής απώλειας παρατηρούνται, ενώ δεν παρατηρείται για τα πλοία αξιοσημείωτη μεταβολή του ποσοστού ολικής απώλειας μετά το 1984.

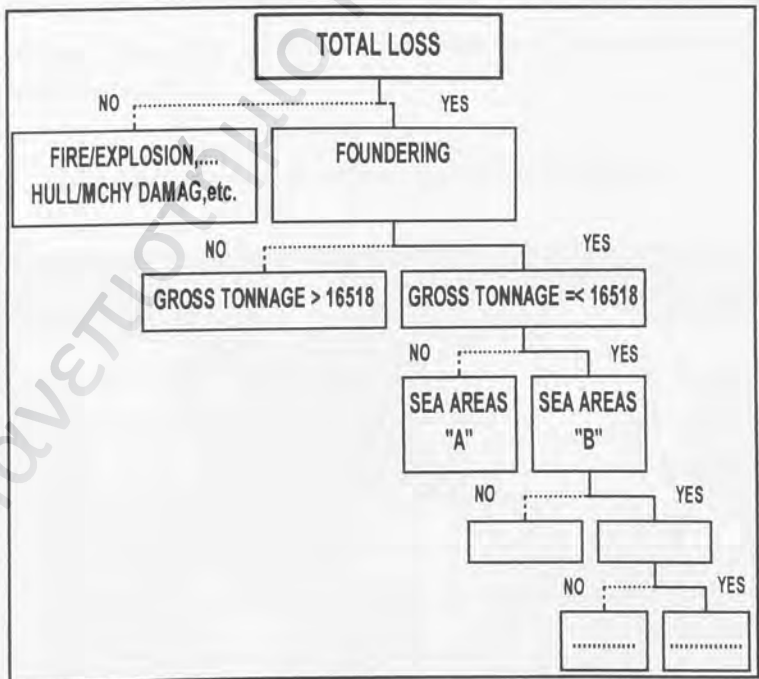
• Άξιο λόγου είναι το συμπέρασμα ότι τα πλοία μικρότερης κατά μέσο όρο χωρητικότητας, ανεξαρτήτως νηογνώμονα, έχουν μεγαλύτερη συχνότητα εμφάνισης ολικής απώλειας. Παρατηρούμε επίσης ότι για ορισμένους νηογνώμονες τα πλοία τους που υπέστησαν ολική απώλεια έχουν μεγαλύτερη κατά μέσο όρο χωρητικότητα από άλλα, όπως για παράδειγμα ο Νορβηγικός ο οποίος έχει εγγεγραμμένα πλοία μεγάλης κατά μέσο όρο χωρητικότητας.

### 7.3 Συμπεράσματα από την ανάλυση του αρχείου με τις επιλεγείσες μεθοδολογίες

Με χρήση μεθοδολογίας δέντρων αποφάσεων κατά την Όρυξη δεδομένων, εντοπίστηκαν τα χαρακτηριστικά που έχουν την μεγαλύτερη επιρροή στο γεγονός της ολικής απώλειας.

Ο αλγόριθμος των δένδρων ταξινομήσε, όπως δείχθηκε στο υποκεφάλαιο 6.3 και διαγραμματικά απεικονίζεται στο Δένδρο 7.1, μεταξύ των άλλων τις μεταβλητές, στα τρία πρώτα επίπεδα, και μάλιστα αναλυτικότερα:

- ομαδοποίησε τις κατηγορίες ατυχήματος άλλες σε ιδιαίτερα επικίνδυνες (CAPSIZED/ FOUNDERED) και άλλες σε λιγότερο (FIRE/COLLISSION κ.α.).
- διαχώρισε τα πλοία ανάλογα με την χωρητικότητά τους και υπέδειξε σαν περισσότερα επικίνδυνα για ολική απώλεια αυτά με χωρητικότητα μικρότερη από 16518 grt, τα οποία είναι και περισσότερα ευάλωτα.



ΔΕΝΔΡΟ 7.1

Διαγραμματική απεικόνιση των διαχωρισμών

## ΥΠΟΜΝΗΜΑ ΔΕΝΔΡΟΥ 7.1

### SEA AREAS "A":

8: ATLANTIC, 7: AUSTRALIA, 12: CANADIAN ALASKA / U.S.S.R. ARTIC / BERING SEA / ICELAND, 2: MEDITERRANEAN/ SUEZ, 1: N. SEA / BALTIC / KIEL, 11: PACIFIC, 10: W. COAST S-N. AMERICA / PANAMA / CAPE HORN, 9: W. INDIES / U.S. EASTERN / GULF OF MEXICO / GREAT LAKES / NEW FOUNDLAND.

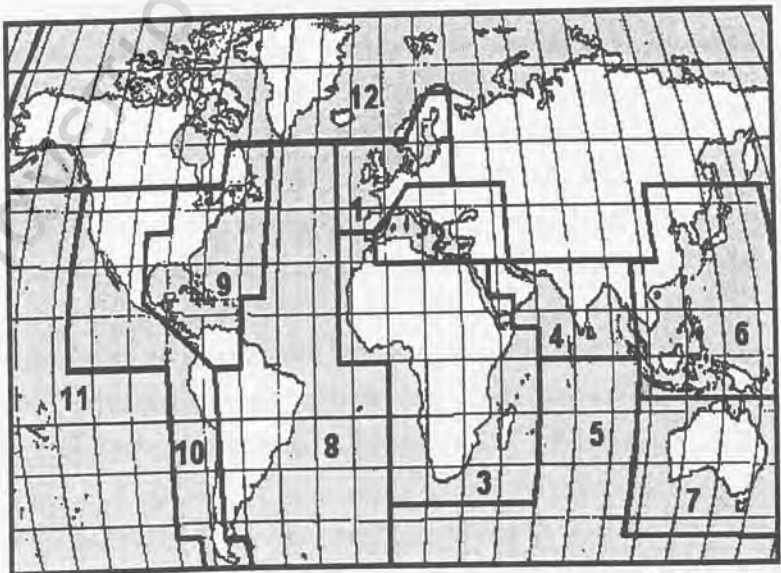
### SEA AREAS "B":

4: GULF / BAY OF BENGAL, 5: INDIAN OCEAN / ANTARTIC, 6: JAPAN / S. CHINA / E. INDIES, 3: RED SEA / E. & W. AFRICAN COAST.

• χώρισε τις γεωγραφικές περιοχές του ατυχήματος σε περιοχές «υψηλού κινδύνου» SEA AREAS "B" (ΥΠΟΜΝΗΜΑ ΔΕΝΔΡΟΥ 7.1) που έχουν σαν επικρατούν γεγονός την ολική απόλεια και σε περιοχές που δεν έχουν την ολική απόλεια SEA AREAS "A" (ΥΠΟΜΝΗΜΑ ΔΕΝΔΡΟΥ 7.1). Αυτές απεικονίζονται στον παγκόσμιο Χάρτη 7.2 με την κωδικοποίηση που παρουσιάζεται στο ΥΠΟΜΝΗΜΑ ΔΕΝΔΡΟΥ 7.1

### ΧΑΡΤΗΣ 7.2

Γεωγραφικές περιοχές (με κωδικοποίηση από το Υπόμνημα 7.1)





Σχηματίζοντας μια ευρύτερη ζώνη, οι περιοχές Ερυθρά Θάλασσα, Περσικός Κόλπος, Αφρικανικές ακτές, Κόλπος της Βεγγάλης, Ινδικός, Θάλασσα Ιαπωνίας και Νότιας Κίνας, λόγω μεγάλης κυκλοφορίας και ιδιόμορφων καιρικών φαινομένων επιβεβαιώνουν, όπως ήταν αναμενόμενο, τον χαρακτηρισμό τους σαν περιοχές «υψηλού κινδύνου».

Η ιεράρχηση των παραγόντων που προκύπτει από την διαδικασία ομαδοποίησης με τα δένδρα, με ποσοστό αξιοπιστίας της ταξινόμησης 92,44 %, επέτρεψε καταρχήν να διαφανεί κατά προτεραιότητα η σημαντικότητα του ατυχήματος, της χωρητικότητας και της γεωγραφικής περιοχής. Παρατηρούμε, στο βασικό δένδρο Δ<sub>1</sub> που σχολιάστηκε στο υποκεφάλαιο 6.3, ότι η εποχικότητα / μήνας της δεν παίζει σημαντικό ρόλο. Εν τούτοις για την περίοδο των Μουσώνων για τις σχολιασθείσες γεωγραφικές περιοχές και χωρητικότητες καθώς και για τα προαναφερθέντα ατυχήματα ο νηογνώμονας στον οποίο είναι εγγεγραμμένο το πλοίο είναι σημαντικός παράγοντας. Αναλυτικά τα πλοία τα ανήκοντα στους λουπούς νηογνώμονες, πλην αυτών που ανήκουν στον Διεθνή Σύνδεσμο Νηογνομώνων (I.A.C.S), έχουν σαν επικρατούσα τιμή την ολική απώλεια.

Κατά κύριο λόγο δεν διαχώρισε τα πλοία ανάλογα με την ηλικία, τον τύπο τους και τις καιρικές συνθήκες που επικρατούσαν την χρονική στιγμή του ατυχήματος.

Έγινε δυνατή, πράγμα απαραίτητο για τις αναλύσεις άρα και για ένα σύστημα αποφάσεων, η δημιουργία όπως δείχθηκε στο υποκεφάλαιο 6.4 κανόνων ταξινόμησης. Σε αυτούς φαίνεται για κάθε ένα φύλλο του δένδρου ο κανόνας που οδηγεί σε αυτό καθώς και το πλήθος των τελικών κανόνων, πράγμα που δε φαίνεται στα δέντρα αποφάσεων.

Σχολιάζεται ιδιαίτερα το γεγονός της εύκολης χρήσης τους. Ειδικότερα από την σύγκριση που γίνεται μεταξύ των συμπληρωματικών κανόνων, όπως αναλυτικά σχολιάστηκε, από τον Rule #2 for No και από τον Rule #1 for Yes, συμπεραίνουμε ότι η πρώτη καταγραφή ενός ατυχήματος πλοίου στο αρχείο δεδομένων δε συνοδεύεται αυτόματα με ολική απώλεια. Στην περίπτωση όμως ύπαρξης και δεύτερης καταγραφής ατυχήματος αυξάνει απότομα η πιθανότητα για ολική απώλεια.

Με ακρίβεια 91.6% έγινε δυνατό να επιτευχθεί ανάλυση αξιοπιστίας, χρησιμοποιώντας τα νευρωνικά δίκτυα, που έδωσε την σχετική βαρύτητα των

παραγόντων που χρησιμοποιήθηκαν για την πρόβλεψη της ολικής απώλειας. Ειδικότερα, από τα αποτελέσματα στο υποκεφάλαιο 6.5, παρατηρούμε ότι ο καιρός δεν είναι από τους πρώτους σε σπουδαιότητα παράγοντες πράγμα που συμφωνεί και με παλαιότερη επιστημονική ανακοίνωση των ερευνητών του Εργαστηρίου Ναυτικών Ατυχημάτων του Πανεπιστημίου Πειραιά [Giziakis et al, 1998].

Στο υποκεφάλαιο 6.5 δείχθηκε επίσης ότι μπορούμε να εξειδικεύσουμε την ανάλυση επίδρασης των παραγόντων, με τον αποκλεισμό μερικών από αυτούς με βάση τις γενικότερες παρατηρήσεις στο αρχικό δένδρο, και με χρήση μεθοδολογίας Boosting να μελετήσουμε την επίδραση ενός υποσυνόλου παραγόντων στο γεγονός της ολικής απώλειας.

Ανεξαρτήτως ατυχήματος, ο αλγόριθμος που παρήγαγε το δένδρο  $\Delta_3$  διαχώρισε τους ομαδοποιημένους νηογνώμονες σε δύο ομάδες. Η μια ομάδα, στην οποία σημειωτέον ομαδοποιούνται τρεις κλάσεις, περιλαμβάνει τους νηογνώμονες AB-GL-NK-BV-CC-KR-NV-RI-PR-RS και LR, που ανήκουν στον Διεθνή Σύνδεσμο Νηογνώμωνων (I.A.C.S). Για αυτούς δεν έχουμε σαν επικρατούν γεγονός την ολική απώλεια ενώ στη δεύτερη ομάδα των υπολοίπων νηογνώμωνων παρατηρείται ολική απώλεια. Ως γνωστόν ο Διεθνής Σύνδεσμος Νηογνώμωνων εκπροσωπεί τους μεγαλύτερους νηογνώμονες παγκόσμια και ένας από τους κύριους σκοπούς του είναι η προώθηση συνεχώς υψηλότερων πρότυπων όσον αφορά την ασφάλεια του πλοίου.

Ο αλγόριθμος διαχώρισε τους νηογνώμονες σε αναμενόμενες ομάδες ανεξαρτήτως ατυχήματος. Η μη διάκριση των νηογνώμωνων που ανήκουν στον Διεθνή Σύνδεσμο Νηογνώμωνων, στην συγκεκριμένη ομαδοποίηση νηογνώμωνων, μεταξύ τους αναδεικνύει την γενικότερη αξιοπιστία των προσφερομένων υπηρεσιών του συγκεκριμένου Συνδέσμου στο σύνολό του.

Μετά τον πρώτο διαχωρισμό όμως, η αποτελεσματικότητα των υπηρεσιών που παρέχει ο Διεθνής Σύνδεσμος Νηογνώμωνων ελέγχεται για τα παλιά πλοία που το έτος ναυπήγησης είναι παλιότερο του 1974. Στα παλιά αυτά πλοία το έργο επιθεώρησης και εντοπισμού βλαβών καθίσταται επίπονο και ιδιαίτερα δύσκολο.

Δευτερευόντως, ο διαχωρισμός που έκανε ο αλγόριθμος σε δύο ομάδες στην μία των οποίων ομαδοποιούνται τρεις κλάσεις που περιλαμβάνουν τους ανήκοντες στον Διεθνή Σύνδεσμο Νηογνώμωνων, δείχνει για μια φορά ακόμη την ευαισθησία, τις δυνατότητες, την αξιοπιστία του αλγόριθμου άρα και την ορθότητα επιλογής του.

Σε συνέχεια των ανωτέρω και στα πλαίσια της εξειδίκευσης της ανάλυσης της επίδρασης των παραγόντων παρουσιάστηκε μια εύχρηστη εφαρμογή, που με



χρήση του αλγόριθμου G.R.I. μπορέσαμε να αναπτύξουμε όλα τα σχετικά δέντρα με περιορισμό του βάθους του δένδρου, με ευδιάκριτα αποτελέσματα. Στην περίπτωση αυτή με σκοπό να παρατηρήσουμε τα «παλιά», που έχουν ναυπηγηθεί μέχρι το 1974, πλοία προέκυψαν οι συνδυασμοί των παραγόντων που επιλέχθηκαν και τα αντίστοιχα ποσοστά τους, όπως δείχθηκε στο υποκεφάλαιο 6.5, που οδηγούν στην ολική απώλεια. Μάλιστα ο κανόνας με το μεγαλύτερο ποσοστό 85% για την ολική απώλεια βασικά συνδυάζει δεδομένα όπως γεωγραφική περιοχή αυτή της Ιαπωνίας-N. Κίνας, τύπο πλοίου FERRY και τους υπόλοιπους νηογνώμονες, πλην του Διεθνούς Συνδέσμου Νηογνώμωνων.

Στα στοιχεία που έχουν συλλεχθεί, με τις μεθοδολογίες που αναφέρθηκαν, έγινε επεξεργασία των πληροφοριών και εξήχθησαν πολύτιμα συμπεράσματα ώστε να γίνονται με έμφαση σε ορισμένους παράγοντες οι απαραίτητοι έλεγχοι, να παρθούν μέτρα και να γίνονται προβλέψεις με σκοπό την αποφυγή ατυχημάτων και ειδικότερα της ολικής απώλειας. Τα αποτελέσματα πιστεύουμε ότι θα βοηθήσουν τους οργανισμούς της Ευρωπαϊκής Ένωσης και τους Παγκόσμιους Οργανισμούς ως προς την αύξηση της ασφάλειας στη Ναυτιλία στα ύδατα της Ευρωπαϊκής Ένωσης και σε παγκόσμιο επίπεδο.

Ένα από τα πλεονέκτημα της συγκεκριμένης μεθοδολογία ανάλυσης δεδομένων είναι το ότι μας παρέχει την ευχέρεια να παρατηρούμε το γεγονός της ολικής απώλειας στην τομή των ανεξαρτήτων μεταβλητών / παραγόντων που οδηγούν σε αυτήν. Αξιοσημείωτο είναι ότι μπορούμε να συμπεράνουμε ότι κατά προτεραιότητα, όπως αναδείχθηκε από την ανάλυση στο υποκεφάλαιο 6.3, στις συγκεκριμένες γεωγραφικές περιοχές στα μικρού μεγέθους πλοία πρέπει να δοθεί προσοχή στα ατυχήματα CAPSIZED, διότι αυτός ο συνδυασμός παραγόντων οδηγεί σε ολική απώλεια. Προκύπτει λοιπόν, μια άμεση ωφέλεια:

- Για φορείς, όπως οι Λιμενικές Αρχές οι οποίες με βάση την αξιοποίηση των συμπερασμάτων από την ανάλυση του αρχείου δεδομένων μπορούν να κατευθύνουν τις Υπηρεσίες Ελέγχου που διαθέτουν προς τα πλοία με το μεγαλύτερο κίνδυνο.

- Για τους Πλοιοκτήτες οι οποίοι μπορούν να αποφεύγουν ορισμένους παράγοντες και λειτουργίες με υψηλή επικινδυνότητα που υποδεικνύονται από τους παραγόμενους κανόνες στους οποίους πρέπει να δίνεται ιδιαίτερη προσοχή. Χωρίς βέβαια να θεωρηθεί αμελητέο για αυτούς το όφελος, στην περίπτωση αγοράς μεταχειρισμένου πλοίου, από την γνώση των ατυχημάτων που είχε το πλοίο κατά τη



διάρκεια της ζωής του, γνώση δηλαδή του ιστορικού του πλοίου, που προκύπτει από την επεξεργασία δεδομένων πρωτογενώς.

Τα συμπεράσματα θα βοηθήσουν στην μείωση του αριθμού των ατυχημάτων και μάλιστα αυτών που οδηγούν στην ολική απώλεια, οδηγώντας τους αρμόδιους φορείς να θεσμοθετήσουν κατάλληλα μέτρα. Αναλυτικότερα και ενδεικτικά αναφέρεται να ληφθούν μέτρα κατά προτεραιότητα παρέμβασης:

α) για αποτροπή CAPSIZED (FOUNDERED) λόγω απώλειας της ευστάθειας των πλοίων, που μπορεί να προκύψει από μετατόπιση φορτίου, εισροή υδάτων κ.λ.π.

β) με απόδοση έμφασης στα μικρού μεγέθους πλοία

γ) για συγκεκριμένες γεωγραφικές περιοχές κυκλοφορίας αυτών των πλοίων.

Οι νηογνώμονες μπορούν είτε να χρησιμοποιήσουν τα συμπεράσματα από την ανάλυση, είτε μπορούν να πάρουν πρωτογενές υλικό όσον αφορά τα ατυχήματα, προκειμένου να βελτιώσουν τους κανόνες τους. Τα αποτελέσματα των αναλύσεων μπορούν να βοηθήσουν, έτσι ώστε οι Επιτροπές Ασφαλείας του I.M.O, του N.E.E, της E.E.E να μπορούν να λάβουν τα αποτελεσματικότερα μέτρα, όσον αφορά την μείωση του αριθμού, το κόστους των ατυχημάτων και μάλιστα της ολικής απώλειας και γενικότερα των συνεπειών που προκύπτουν από αυτά.

Τα δένδρα προσφέρονται για υψηλού βαθμού πρόβλεψη [(15)-web] [(16)-web] του γεγονότος που εξετάζουμε, χωρίς να έχει καν οριστεί η χρήση των συντελεστών βαρύτητας για τα σφάλματα ταξινόμησης (misclassification costs).

Οι ενδιαφερόμενοι φορείς μπορούν να είναι σε ετοιμότητα ώστε να δράσουν γρήγορα και προς την σωστή κατεύθυνση σε κάθε προβλεπόμενο ατύχημα διότι χρησιμοποιώντας την ανάλυση με τα δένδρα έχουμε πολύ καλύτερη προβλεπτική ικανότητα για το γεγονός, από ότι αν χρησιμοποιήσουμε τις υπόλοιπες μεθόδους ανάλυσης.

Το ότι έχουμε με τα δένδρα πολύ καλύτερη προβλεπτική ικανότητα θα μπορούσε να δειχθεί αν δοκιμασθεί εναλλακτικά μοντέλο παλινδρόμησης [Dillon R.W. et al. 1984] για την πρόβλεψη της ολικής απώλειας. Ενδεικτικά η εκτέλεση αυτής της μεθόδου με το βασικό σύνολο μεταβλητών (BUILT1, NAREA, MONTH, OWNERS, WEATH, NOFACC, GRUCLASS, TYPE) δίνει αποθαρρυντικά αποτελέσματα.

Το  $R^2$  είναι εξαιρετικά χαμηλό όπως φαίνεται και από τον πίνακα 7.3 της συνοπτικής αναφοράς και ακόμα χειρότερα η παλινδρόμηση, όπως φαίνεται από τον πίνακα ταξινόμησης 7.4 ολικής και μη ολικής απώλειας, λειτουργεί εξαιρετικά καλά

Πίνακας 7.3  
Συνοπτικής αναφοράς παραμέτρων

Model Summary

Step	-2 Log likelihood	Cox & Snell R Square	Nagelkerke R Square
1	8966,730	,077	,189

για ένα από τα δύο σύνολα περιπτώσεων, την μη ολικής απώλειας No και πολύ άσχημα για το δεύτερο της ολικής απώλειας Yes με ποσοστό μόλις 3.7%.

Πίνακας 7.4  
Ταξινόμησης ολικής και μη ολικής απώλειας

Classification Table<sup>a</sup>

Observed		Predicted			
		total loss		Percentage Correct	
		No	Yes		
Step 1	total loss	No	18794	67	99,6
		Yes	1421	55	3,7
Overall Percentage					92,7

a. The cut value is ,500

Είναι σημαντική η δυνατότητα των δένδρων να παρέχουν, με μεταβλητές εισόδου (INPUT) για παράδειγμα, το ατύχημα, την χωρητικότητα, τις γεωγραφικές περιοχές που συνέβησαν τα ατυχήματα κ.α., σημαντικά στοιχεία πρόβλεψης στην έξοδο (OUTPUT) για την ολική απώλεια καθώς και διερεύνηση των βασικών παραγόντων που οδηγούν σε αυτήν. Αυτή η ικανότητα πρόβλεψης είναι ιδιαίτερα χρήσιμη για αποφυγή αυτής καθαυτής της ολικής απώλειας καθώς και τον περιορισμό των επακόλουθών της για:

- Τις Κρατικές Υπηρεσίες και Διεθνείς Οργανισμούς που ασχολούνται με την Ναυτιλία και οι οποίοι μπορούν να οργανωθούν ως προς τον σχεδιασμό, την προετοιμασία και αντιμετώπιση εκτάκτων αναγκών και μάλιστα μπορούν να επέμβουν γρήγορα και προς την σωστή κατεύθυνση λόγω της πρόβλεψης που θα τους παρέχεται. Η αντιμετώπιση εκτάκτων αναγκών θα γίνεται σύμφωνα με τα οριζόμενα από τις συμβάσεις για την ανάπτυξη και την εφαρμογή τους σε τοπικό περιφερειακό

και εθνικό επίπεδο καθώς και στην περίπτωση προβλέψεων, σχετικών με τις γεωγραφικές περιοχές, με ενεργοποίηση των διακρατικών η διεθνών φορέων.

- Τους Οργανισμούς τους σχετικούς με την προστασία του Περιβάλλοντος. Οι αρμόδιοι φορείς και οι εθελοντικές οργανώσεις, μπορούν να προτείνουν μέτρα και να επέμβουν κυρίως για μείωση των συνεπειών που προέρχεται από ατυχηματική μόλυνση του περιβάλλοντος.

- Τις Ασφαλιστικές Εταιρίες. Όταν έχουμε εντοπίσει τους παράγοντες που χαρακτηρίζονται «υψηλού κινδύνου» είναι σε θέση οι Ασφαλιστικές εταιρίες να διαμορφώσουν σωστή Τιμολογιακή Πολιτική. Επίσης, δύνανται να αξιολογήσουν συγκριτικά τους παράγοντες «υψηλού κινδύνου», έτσι ώστε να γνωρίζουν τα απαιτούμενα κεφάλαια προς αποζημίωση.

Η οποιαδήποτε εξαγωγή συμπερασμάτων για το μέλλον απαιτεί την συμμετοχή του ερευνητή, ο οποίος οριοθετεί συγκεκριμένες παραμέτρους για να παράγει εναλλακτικά συμπεράσματα. Υπάρχουν ορισμένα χαρακτηριστικά τα οποία είναι πολύ καλά για την πρόβλεψη, όπως οι μεταβλητές που χρησιμοποιήθηκαν στην παρούσα ανάλυση. Άλλα μπορούν να μην επηρεάζουν καθόλου τα τελικά αποτελέσματα, όπως η μεταβλητή "BADVOYAGE" και κατά συνέπεια να αποβαίνουν περιττά. Τα παρόμοια και συσχετιζόμενα δεδομένα αυξάνουν αρκετά το μέγεθος των αρχείων χωρίς να προσθέτουν σημαντική πληροφορία. Η μεταβλητή "OWNERS", για παράδειγμα, μεταφέρει σχεδόν στον επιθυμητό βαθμό την επιθυμητή πληροφορία άρα δεν χρειάζεται για την ανάλυση η μεταβλητή "OWNER". Επιπλέον, αυξάνουν και το μέγεθος των προβλημάτων με αποτέλεσμα να καθυστερεί σημαντικά η εξαγωγή αποτελεσμάτων αλλά και να καθίστανται τα μοντέλα Όρυξης μη εφαρμόσιμα.

Επίσης μία συμβολή της παρούσας διατριβής στην απάντηση των γενικότερων ερευνητικών ερωτημάτων που τίθενται, είναι να δώσει τις πρώτες βάσεις για ένα σύστημα υποστήριξης αποφάσεων (Decision Making Systems):

- είτε μέσω των πρωτογενώς επεξεργασμένων δεδομένων
- είτε με τη μεθοδολογία που ακολουθήθηκε
- είτε με τους παραγόμενους κανόνες και συμπεράσματα.

Τέλος, επειδή η διαδικασία παροχής πληροφοριών είναι γενικότερη έννοια από αυτήν της όρυξης δεδομένων και αφορά συστήματα υποστήριξης αποφάσεων είναι ωφέλιμη σε όλους τους φορείς Νηογνώμονες, Λιμενικές Αρχές, Ασφαλιστικές Εταιρίες, Ι.Μ.Ο.



NEE, E.E.E, Πλοιοκτήτες, Περιβαλλοντικούς Οργανισμούς που ασχολούνται με τη Ναυτιλία.

#### 7.4 Προτάσεις για μελλοντικές ερευνητικές εργασίες

Σαν απόρροια των συμπερασμάτων του υποκεφαλαίου 7.3 αλλά και των συνολικών διαδικασιών και αναλύσεων που πραγματοποιήθηκαν σε όλη την διατριβή μπορούν να διατυπωθούν παρατηρήσεις και προτάσεις για μελλοντικές ερευνητικές εργασίες.

Όσον αφορά την κωδικοποίηση σημειώνεται ότι κάποιοι από τους κωδικούς ήταν πιο γενικοί από άλλους, με αποτέλεσμα την συχνότερη χρησιμοποίησή τους, ακόμη και αν μπορούσε να επιλεγεί ένας εξειδικευμένος κωδικός. Σε τέτοιες περιπτώσεις προτείνεται να χρησιμοποιείται συνδυασμός κωδικών για να περιγράψουν την αιτία ενός ατυχήματος.

Σε ορισμένα σενάρια δεν παράγονται δένδρα, μια προτεινόμενη λύση είναι με εμπλουτισμό των εγγραφών του αρχείου δεδομένων θα μπορούσε να γίνει δυνατή η ιεράρχηση των μεταβλητών και ο περαιτέρω διαχωρισμός στις επιμέρους τιμές τους.

Λόγω της διαφοράς του σχετικού μεγέθους της ολικής απώλειας οι κανόνες προβλέπουν με καλύτερη αναλογία, «ευστοχία», τότε για ένα πλοίο δε θα συμβεί ολική απώλεια από το να συμβεί. Σε περίπτωση που αντικείμενο κάποιας εργασίας είναι να ελεγχθεί η σχετική επίδραση ενός συνόλου παραγόντων στην ολική απώλεια θα πρέπει να υλοποιηθεί συγκεκριμένο πειραματικό σχέδιο που να εξομαλύνει τις διαφορές στην κατανομή της ολικής απώλειας, όπως αυτό δείχθηκε στο προηγούμενο κεφάλαιο.

Θα μπορούσαν να γίνουν προτάσεις:

- για συνέχιση συγκέντρωσης πληροφοριών για τα ατυχήματα, ώστε να αυξηθεί το πλήθος των παρατηρήσεων και ο αλγόριθμος να γίνει αποτελεσματικότερος μιας και χρειάζεται το μεγαλύτερο δυνατό πλήθος των παρατηρήσεων,
- για αξιοποίηση της ήδη αποκτηθείσας εμπειρίας στην συλλογή νέων στοιχείων, για παράδειγμα, μη συλλογή πληροφοριών για μεταβλητές απόλυτα συσχετιζόμενες, ώστε να γίνει πιο εύχρηστο το αρχείο δεδομένων,
- για μετατροπή του αρχείου δεδομένων σε βάση σχεσιακή.

- για αφαίρεση, χωρίς να αλλοιωθεί το τελικό αποτέλεσμα νέας ανάλυσης, των χαρακτηριστικών που δεν χρησιμοποιήθηκαν και άρα δεν χρειάζονται.

Η αποκτηθείσα ήδη εμπειρία, τα νέα προγράμματα και οι αλγόριθμοι μπορούν να αξιοποιηθούν στην ανάλυση των νέων στοιχείων που θα συλλεχθούν.

Η συσχέτιση του πίνακα ποσοστού εμφάνισης ολικής απώλειας στις γεωγραφικές περιοχές ανά μήνα με τις εκάστοτε καιρικές συνθήκες μπορεί να οδηγήσει σε εξέταση του ρόλου και της σημασίας των καιρικών συνθηκών στα ναυτικά ατυχήματα με βάση ένα σύνολο άλλων παραγόντων.

Το Νευρωνικό προβλέπει με ακρίβεια 92% σε σχέση με την υπάρχουσα πληροφορία, θα μπορούσε για το υπόλοιπο 8% να διερευνηθεί, πράγμα που ξεφεύγει από τους στόχους της παρούσας εργασίας, για ποιους παράγοντες δεν μπορεί να γίνει πρόβλεψη στο 8%.

Θα μπορούσε επίσης, να μελετηθεί το κόστος των ατυχημάτων και μάλιστα της ολικής απώλειας γενικότερα δε το κόστος των συνεπειών που προκύπτουν από αυτά.

Γενικά, επειδή η όλη διαδικασία Όρυξης δεδομένων είναι κυκλική, όπως και όλες οι διαδικασίες «γνώσης», από τα αποτελέσματα της ανάλυσης μπορεί να γεννηθεί μία νέα υπόθεση ή ακόμα περισσότερο να αναδιατυπωθεί σε νέα βάση η αρχική υπόθεση για περαιτέρω διερεύνηση.

Κατά την προετοιμασία του αρχείου δεδομένων συμπεράναμε ότι μια εγγραφή συνεπάγεται πολλές διορθώσεις και μεταβολές, πράγμα ιδιαίτερα σημαντικό για μελλοντική διαδικασία συμπλήρωσης του αρχείου για τα επόμενα έτη 2000 έως 2002. Οι διορθώσεις και οι μεταβολές θα γίνονται συνέχεια λόγω της δυναμικής φύσης του αρχείου δεδομένων και ο μελλοντικός ερευνητής θα αντιμετωπίσει και αυτός το πρόβλημα. Αισιοδοξούμε ότι θα το ξεπεράσει με την βοήθεια της προσφερόμενης εμπειρίας από την παρούσα εργασία. Άλλωστε αυτή πιστεύουμε ότι θα είναι μια από τις συνεισφορές της παρούσας διατριβής στην αντιμετώπιση συναφών προβλημάτων, που τίθενται προς διερεύνηση και επίλυση από την ευρύτερη επιστημονική κοινότητα.

## ΒΙΒΛΙΟΓΡΑΦΙΑ

Andrews PJD, Sleeman DH, Statham PFX, et al., (2002), Predicting recovery in patients suffering from traumatic brain injury by using admission variables and physiological data: a comparison between decision tree analysis and logistic regression. *J Neurosurgery* 97 (2): 326-336 Aug 2002.

Arning A., Agrawal R., and Raghavan P., 2000, "A Linear Method for Deviation Detection in Large Databases", *VLDB Journal* 8(3-4), p.p. 237-253, <http://www.informatik.uni-trier.de/~ley/db/journals/vldb/KNorrNT00.html>

Bardi E., 1996, "Ανάλυση Κινδύνου Ατυχημάτων στις θαλάσσιες παράκτιες οδούς". Πρακτικά Συνεδρίου: Ελληνικές Ακτές και Θάλασσες στο 2000. Πανεπιστήμιο Πειραιά, Πειραιάς, σελ 345 - 353

Barlow H., 1989, "Unsupervised learning", *Neural Computation*, 1, pp295-311.

Berry M.J.A., Linoff G., 1997, "Data Mining Techniques for Marketing Sales, and Customer Support", Chapter 8, New York: John Wiley and Sons.

Bonchi F., Giannotti F., Mainetto G., Pedreschi D., 1999, Dipartimento di Informatica, Università di Pisa, C.so Italia 40, 56125 Pisa, "Using data mining techniques in fiscal fraud detection", In Proc. DaWak'99, First Int. Conf., Francesco Bonchi's Publications, Italy <http://www.di.unipi.it/~bonchi/publication.html>

Breiman L., Friedman J., Olshen H., Stone CJ., 1996, "A method of choosing multiway partitions for classification and decision trees", *Journal of Applied Statistics* 18(1):49- 62. <http://www.sys.uea.ac.uk/~bli/KDD96.ps>

Brian J. R., 1999, "Data Mining and Science" CLRC Rutherford Appleton Laboratory, Chilton, Didcot, Oxon OX110Qx, UK.  
[http://www.ercim.org/publication/ws-proceedings/12th-EDRG/EDRG12\\_Re.pdf](http://www.ercim.org/publication/ws-proceedings/12th-EDRG/EDRG12_Re.pdf)

Bureau of Transport and Communications Economics, 1994, "Structural Failure of large Bulk Ships", Rep. 85, Commonwealth of Australia, ISBN 1034-4152

Burrows P., 1979, "The economic theory of pollution control" University of York, M. Robertson.

Casualty Information System, LR-LLP Lloyd's of London Press Ltd.

Cover T. and Thomas J., 1991. "Elements of Information Theory", volume-1. John Wiley and Sons, New York.



Dillon R. W. & Goldstein M., 1984, "Multivariate Analysis Methods and Applications" John Wiley , ISBN 0-471-08317-8.

Drucker H., 1999, "Boosting using neural networks," in Sharkey (ed)., pp.51-78.

Durvasula S, Lysonski S, Mehta SC (2002), Understanding the interfaces: How ocean freight shipping lines can maximize satisfaction, Industrial Marketing Management 31 (6): 491-504 Sep 2002.

Fawcett Tom, 2001, HP Labs, MS 1U-4B. 1501 Page Mill Rd. Palo Alto California, CA 94304-1126. Using Rule Sets to Maximize ROC Performance. tfawcett@acm.org. <http://www.purl.org/NET/tfawcett/papers/ICDM-submission.ps.gz>

Fayyad U., Piatetsky-Shapiro, Smyth, and Uthurusamy, 1996, "Advances in Knowledge Discovery and Data Mining", MIT Press.

Giziakis K. Ph.d., 1987, "The cost of Maritime Traffic Accidents Worldwide with particular reference to NW European Waters". Guildhall University.

Giziakis K. ,1996, " Κριτική του περιεχομένου των μεταβλητών που χρησιμοποιούνται στην ανάλυση των ατυχημάτων στην παγκόσμιο ποντοπόρο ναυτιλία ", Τόμος Πανεπιστημίου Πειραιά προς τιμή του Καθηγητού Σταυρόπουλου Κ., Πειραιάς.

Giziakis K. and Giziaki E. ,1996, "Risk Assessment Methods in shipping safety". Journal of Science and Technology, Cyprus , No 1.

Giziakis K. and Kokotos D., 1996, "Ανάγκες και Οφέλη από τη Δημιουργία Βάσης Δεδομένων για τα Ναυτιλιακά Ατυχήματα", Πρακτικά Συνεδρίου: "Ελληνικές Ακτές και Θάλασσες στο 2000", Πανεπιστήμιο Πειραιά, Πειραιάς, σελ. 355-360.

Giziakis K and Kokotos D., 1996, "Τα ατυχήματα της Ελληνικής και Παγκόσμιας Ακτοπλοίας. Η επίδρασή τους στην οικονομική αποδοτικότητα". Πρακτικά 2η Ημερίδας Οικονομικής των Μεταφορών: "Ακτοπλοία και Αερομεταφορές - Ανταγωνισμός και Συμπληρωματικότητα", Πανεπιστήμιο Πειραιά, σελ. 121-132.

Giziakis K and Kokotos D., 1997, "Ανάγκη εφαρμογής Ποιοτικού Ελέγχου για τη μείωση των ατυχημάτων στην Ναυτιλία". Πρακτικά Συνεδρίου: 2ο Συνέδριο "Συστήματα Ποιότητας στις Επιχειρήσεις. Ευρωπαϊκή πρόκληση - προοπτική", Αθήνα, σελ. 101-104.

Giziakis K and Kokotos D., 1998, "Ανάλυση παραγόντων που επιδρούν στα Ναυτικά Ατυχήματα σε σχέση με τις καιρικές συνθήκες", Πρακτικά Ενδέκατου

Πανελληνίου Συνεδρίου Στατιστικής “Εφαρμοσμένα Στοχαστικά Μοντέλα και Ανάλυση Δεδομένων”, Μ.Α.Ι.Χ Χανιά, σελ. 13-22.

Giziakis K and Kokotos D., 1998, “Η Ασφάλεια στην Ελληνική Επιβατηγό Ναυτιλία”, Πρακτικά Συνεδρίου: Thalassa Forum 98, Αθήνα, σελ. 155-158.

Giziakis K and Kokotos D., 1998, “Μόλυνση Περιβάλλοντος από ατυχηματική ρύπανση πλοίων”. Πρακτικά Δωδέκατου Εθνικού Συνεδρίου Ελληνικής Εταιρίας Επιχειρησιακών Ερευνών, “Επιχείρηση και Νέες Τεχνολογίες, Σύγχρονα εργαλεία Υποστήριξης, Διαχείρισης και Λειτουργίας”. Πανεπιστήμιο Αιγαίου, Σάμος, σελ. 139-146.

Giziakis K and Kokotos D., 2000, “Στατιστικά στοιχεία για ατυχήματα Δεξαμενόπλοιων” Πρακτικά Συνεδρίου: Thalassa Forum 2000, Αθήνα.

Giziakis K., Polydoropoulou A., Kantzos K., 2001, “Risk analysis of accidents in main Greek Ports”, 2nd International Conference: Safety of Maritime Transport, University of the Aegean, Chios-Greece.

Giziakis K., 1996, “Οικονομική Θεώρηση για τον έλεγχο της μόλυνσης θαλασσίου περιβάλλοντος” Πρακτικά Συνεδρίου: “Ελληνικές Ακτές και Θάλασσες το 2000”, Πανεπιστήμιο Πειραιά, Πειραιάς, σελ. 299-308.

Giziakis, K., Giziaki, E., Pardali-Lainou, A., Michalopoulos, V., Kokotos, D., 1996, “Minimising the Risk of Failure for an Effective and Reliable European Shipping Network”, Bergen Norway, Third European Research Roundtable Conference on Shortsea Shipping: Building European ShortSea Networks, Delft University Press, pp 456-470.

Goulielmos A.M. & Giziakis K., 1995, “Treatment of uncompensated cost of marine accidents in a model of Welfare Economics”, Proceedings of JAME Conference, Boston M.I.T. U.S.A., Session 8 - 8.3.

Goulielmos A.M. & Giziakis K., 1997, “Έλεγχος ποιότητας στη Ναυτιλιακή Επιχείρηση και στο Πλοίο”, Εκδόσεις Α. Σταμούλης .

Graham J. Williams and Zhexue Huang, 1997. Cooperative Research Centre for Advanced Computational Systems, CSIRO Mathematical and Information Sciences, GPO Box 664 Canberra 2601 Australia, Graham.Williams@cmis.Csiro.au, “Mining the Knowledge Mine: The Hot Spots Methodology for Mining Large Real World Databases”, Volume 1342, Advanced Topics in Artificial Intelligence, Abdul Sattar (Ed.), Springer-Verlag.

Hand D., Mannila H., Smyth P., 2001, "Principles of Data Mining" M.I.T Press, Massachusetts.

Hashemi R. R., Pearce B. A., Arani R. B., Hinson W. G., and Paule M. G., 1997, "A Fusion of Rough Sets, Modified Rough sets, and Genetic Algorithms for Hybrid Diagnostic Systems", Rough Sets and Data Mining: Analysis of Imprecise Data, Kluwer Academic Publishers, Boston, pp 149-176.

Hashemi R. Ray, Louis A. Le Blanc, Conway T. Rucks, and Angela Shearry, 1995, University of Arkansas at Little Rock, 2801 South University Avenue, Little Rock, AR 72204-1099, U.S.A. "A Neural Network for Transportation Safety Modeling", Expert Systems with Applications Vol 9. N 3, pp. 247-256.

Hernandez M. and S. Stolfo S., 1997, "Real-world data is dirty: Data cleaning and the merge/purge problem". Data Mining and Knowledge Discovery, 2(1)..

Holsheimer M., Siebes A., 1994, "Data Mining - The Search for Knowledge in Databases", Technical Report CS-R9406, CWI, Amsterdam, The Netherlands.

I.M.O. A.849(20) (Resolution) 1997 "Κώδικας για τη Διερεύνηση Ναυτικών Ατυχημάτων και Συμβάντων".

Iba W., Wogulis J. & Langley P., 1988, "Trading off Simplicity and Coverage in Incremental Concept Learning" Proc. 5th Conf. on Machine Learning.

Ismail M.A., and Kamel M.S, 1989, "Multidimensional data clustering utilizing hybrid search strategies", Pattern Recognition, 22, 75-89.

Karydis P and Vasilakos E., 1998, "Event Tree Method: Η διερεύνηση των ναυτικών ατυχημάτων με τη μέθοδο των δικτύων γεγονότων", Πρακτικά 1<sup>ου</sup> Πανελληνίου Συνεδρίου με διεθνή συμμετοχή: Ασφάλεια θαλάσσιων Μεταφορών, Πανεπιστήμιο Πειραιά, Πειραιάς σελ 97-110.

Kohavi R., 1995, "A study of crossvalidation and bootstrap for accuracy estimation and model selection", International Joint Conference on Artificial Intelligence (IJCAI).

Kokotos D. and Ginis D., 2000, "Ρύπανση από ατυχήματα δεξαμενόπλοιων. Στατιστικά στοιχεία. ", Τόμος Πανεπιστημίου Πειραιά προς τιμή της Καθηγήτριας Γ. Νικολάου-Σμοκοβίτη, Πειραιάς, σελ. 631-652.

Kokotos D. and Ginis D., 2001, "Στατιστική παρουσίαση παραμέτρων που σχετίζονται με τα Ναυτιλιακά ατυχήματα", Επιστημονική Επετηρίδα Πανεπιστημίου Πειραιά προς τιμή του Καθηγητή Δ. Κοδοσάκη, Πειραιάς, σελ. 541-564.



Kokotos D., 1996, "Accident Database for Economic Appraisal and Marine Safety", International Shipping Conference: Thalassa 96 Forum, Evgenides Foundation, Athens.

Kokotos D., 1997, "Περιγραφική Στατιστική παρουσίαση των ατυχημάτων Πυρκαϊάς, βασισμένη σε αξιοποίηση των πληροφοριών της Βάσης Δεδομένων για τα Ναυτικά Ατυχήματα". Επιστημονική Επετηρίδα Πανεπιστημίου Πειραιά προς τιμή του Καθηγητή Β.Ν. Μεταξά, Πειραιάς, σελ.109-115.

Kokotos D., 1998, "Μια Στατιστική ανάλυση δεδομένων ατυχηματικής ρύπανσης", Πρακτικά 1<sup>ου</sup> Πανελληνίου Συνεδρίου με διεθνή συμμετοχή: Ασφάλεια θαλάσσιων Μεταφορών, Πανεπιστήμιο Πειραιά, Πειραιάς, σελ 302-311.

Kokotos D., 2000, "Ανάλυση δεδομένων για ατυχήματα δεξαμενόπλοιων " Τόμος Πανεπιστημίου Πειραιά προς τιμή της Καθηγήτριας Γ. Νικολάου-Σμοκοβίτη, Πειραιάς σελ. 619-629.

Kokotos D., 2001, "Νηογώμονες και Ναυτιλιακά ατυχήματα. Μια στατιστική ανάλυση ". Επιστημονική Επετηρίδα Πανεπιστημίου Πειραιά προς τιμή του Καθηγητή Δ. Κοδοσάκη, Πειραιάς, σελ. 565-573,.

Kokotos D., 2002, "Εφαρμογή μεθοδολογίας data mining για παραγωγή κανόνων με σκοπό την πρόγνωση ναυτικών ατυχημάτων", Τόμος Πανεπιστημίου Πειραιά προς τιμή του Καθηγητή Α. Λάζαρη, Πειραιάς.

Kokotos D., 2002, "Εφαρμογή των δέντρων αποφάσεων στη λήψη των αποφάσεων με τη χρήση μεθοδολογίας Όρυξης δεδομένων στα ναυτικά ατυχήματα.", 15<sup>ο</sup> Πανελλήνιο Συνέδριο Στατιστικής, Ε.Σ.Ι, Ιωάννινα.

Kosuda S, Ichihara K, Watanabe M, Kobayashi H, Kusano S, (2002). Decision-tree sensitivity analysis for cost-effectiveness of whole-body FDG PET in the management of patients with non-small-cell lung carcinoma in Japan *Annals of Nuclear Medicine* 16 (4): 263-271 Jun 2002.

Kristiansen S., 1995, "An approach for systematic learning from accidents", IMAS 95, Conference on Management and Operation of Ships.

Kula E., 1992, "Economics of Natural Resources and the Environment" Chapman and Hall.

Le Blanc<sup>1</sup> A. L., Hashemi<sup>2</sup> R.R., Rucks<sup>3</sup> T. C., 2001, "Pattern development for vessel accidents: a comparison of statistical and neural computing techniques", <sup>1</sup>Berry College, Box 5024, Mount Berry, GA30149-5024, U.S.A., <sup>2</sup>Department of Computer Science, University of Arkansas at Little Rock, 2801 South University Avenue, Little

Rock, AR 72204-1099, U.S.A., <sup>3</sup>Department of Marketing, University of Arkansas at Little Rock, 2801 South University Avenue, Little Rock, AR 72204-1099, U.S.A., Expert Systems with Applications Vol 20. N 3, pp. 163-171.

Lian J, Lai XM, Lin ZQ, Yao FS, (2002), Application of data mining and process knowledge discovery in sheet metal assembly dimensional variation diagnosis Journal of Materials Processing Technology 129 (1-3): 315-320 Oct 11 2002.

Lloyd's Casualty week, Lloyd's of London Press Ltd.

Lloyd's Register of Ships, Lloyd's Register Ltd.

M.I.T Technology Review, 2001 Jan/Feb.

MacKay, D., 1998, "Information Theory Inference and Learning Algorithm". <http://wol.ra.phy.cam.ac.uk/mackay/itprn/#book>

Maletic I. Jonathan, Marcus Andrian, 2000, "Data Cleaning: Beyond Integrity Analysis" <http://citeseer.nj.nec.com/maletic00data.html>

Mallen, J., 1995, "Utilizing Incomplete Domain Knowledge in an Information Theoretic guided Inductive Knowledge Discovery Algorithm". PhD Thesis, University of Portsmouth, UK.

Maritime ATLAS, 1993, Lloyd's of London Press Ltd, 10<sup>th</sup> Edition.

Markham J. M., Rhodes C. P., 1999, "Maximising entropy to deduce an initial probability distribution for a causal network", International journal of Uncertainty, Fuzziness and Knowledge based systems, Vol7, No1.

Michalski, R. & Tecuci, G., 1994, "Machine Learning: A Multistrategy Approach", Volume Four, Morgan Kaufmann, California.

Michalski, R., Carbonnel, J. & Mitchell, T., 1986, "Machine Learning: An Artificial Intelligence Approach", Volume Two, Morgan Kaufmann, California.

Nakaya Akihiro, Furukawa Hideharu, and Morishita Shinichi, 1999, University of Tokyo, 4-6-1 Shirokanedai, Minato Ward, Tokyo 108-8639, Japan {nakaya, moris} @ims.u-tokyo.ac.jp, furukawa@ims.u-tokyo.ac.jp., "Weighted Majority Decision among Several Region Rules for Scientific Discovery", <http://www.is.s.u-tokyo.ac.jp/annual-reviews/1999/review99.ps.gz>

Nakaya Akihiro, Furukawa Hideharu, and Morishita Shinichi, 2000, University of Tokyo, "Tests for Classification and Regression", Transaction of IEICE, Vol.E83-D, No.1, pp.52-60, <http://www.gi.k.u-tokyo.ac.jp/~moris/data-mining-papers.htm>

NASA Workshop, 1999, "NASA Workshop on Issues in the Application of Data Mining to Scientific Data".

Norusis M, 1993, "Spss for Windows User' Guide, pp.267-280.

Ogan K, Lotan Y, Koeneman K, Pearle MS, Cadeddu, (2002), Laparoscopic versus open retroperitoneal lymph node dissection: A cost analysis Journal of Urology 168 (5): 1945-1949 Nov 2002.

Piatetsky-Shapiro, G. & Frawley, W., 1991, "Knowledge Discovery in Databases", AAAI Press.

Psarafitis Ch, Karydis P., Desipris N, Panagakos G., Ventikos N., 1998, "Ο ανθρώπινος παράγοντας ως παράμετρος των ναυτικών ατυχημάτων", Πρακτικά 1<sup>ου</sup> Πανελληνίου Συνεδρίου με διεθνή συμμετοχή: Ασφάλεια θαλάσσιων Μεταφορών, Πανεπιστήμιο Πειραιά, Πειραιάς, σελ 41-53.

Psarafitis Ch, Panagakos G., Desipris N, Ventikos N., 1998, "An Analysis of Maritime Transportation Risk Factors ", Proceedings of the 8<sup>th</sup> Conference ISOPE, Montreal, Canada Vol IV, pp 484-492.

Quinlan J. R., 1986, "Induction of decision trees", Machine Learning,1, pp 81-106

Quinlan J. R., 1986a, "The Effect of Noise on Concept Learning", in Michalski et al, Chapter 6, Machine Learning: An Artificial Intelligence Approach, volume 2. Los Altos, CA: Morgan Kaufmann

Quinlan J. R., 1987, "Simplifying decision trees", International Journal of Man-Machine Studies, 27, pp221-234.

Quinlan J. R, 1993, "C4.5: Programs for Machine Learning" Morgan Kaufmann, San Mateo, California, ISBN 1-55860

SAFECO I, 1998, & SAFECO II, 1999, Safety of Shipping Coastal Waters ,DGVII, Γενική Διεύθυνση Μεταφορών Ευρωπαϊκής Ένωσης.

Schlimmer J., 1987, "Concept Acquisition through Representational Adjustment", Dept. of Information and Computer Science, University of California, Irvine, IBM Journal of Research and Development, 11, 601-617, Tech. Rep. No.87-19. <http://www.vuse.vanderbilt.edu/~dfisher/tech-reports/tr-88-05/node16.html>

Smyth, P. & Goodman, R., 1992, "An Information Theoretic Approach to Rule Induction from Databases", IEEE Transactions on Knowledge and Data Engineering, Vol. 4, No. 4.



The International Tanker Owners Pollution Federation, 1991, "Response to marine oil spills", Witherby Co. Ltd, London,.

Thoms MC, Sheldon F, (2002), An ecosystem approach for determining environmental water allocations in Australian dryland river systems: the role of geomorphology. *Geomorphology* 47 (2-4): 153-168 Oct 1 2002.

Tong WD, Hong HX, Fang H, Xie Q, Perkins R, (2002), From decision tree to decision forest - A novel chemometrics approach for QSAR study. Abstracts of Papers of the American Chemical Society 224: 280-COMP Part 1 Aug 18 2002.

Two Crows, 1999, "Induction to Data Mining and Knowledge Discovery", Third Edition, ISBN:1-892095-02-5 by Two Crows Corporation, 10500 Falls Road, Potomac, MD 20854 U.S.A.

Tzannatos E. "Ναυτική Τεχνολογία" Σημειώσεις 2002.

Vijayshankar Raman and Joseph M. Hellerstein, 2001, University of California at Berkeley, {rshankar,jmh}@cs.berkeley.edu, "Potter'Wheel: An Interactive Data Cleaning System", Proceedings of the 27<sup>th</sup> VLDB Conference, Roma, Italy.

Vlachos G., Alexopoulos A., 1995, Τεχνικό-οικονομικές απόψεις της θαλάσσιας διακίνησης των αγαθών και της προστασίας του θαλάσσιου περιβάλλοντος, Εκδόσεις Α. Σταμούλης.

Vlachos G., Alexopoulos A., 1996, "Διεθνείς Οργανισμοί και Ναυτιλιακή Πολιτική", Εκδόσεις Α. Σταμούλης.

Wei CP, Chiu IT, (2002), Turning telecommunications call details to churn prediction: a data mining approach *Expert Systems with Applications* 23 (2): 103-112 Aug 2002.

Westphal C., Blaxton T., 1998, "Data Mining Solutions", John Wiley.

Winston, P.H., 1992, "Artificial Intelligence", Third Edition Addison-Wesley.  
World Casualty Statistics, Lloyd's Register Ltd.

Yang, H. and Amari, S., 1997, "Adaptive on line learning algorithms for blind separation maximum entropy and minimum mutual information", *Neural Computation*, 9:1457-1482.

## ΑΝΑΦΟΡΕΣ στο ΔΙΑΔΙΚΤΥΟ (web)

- (1) <http://cgm.cs.mcgill.ca/~soss/cs644/projects/simon/Entropy.html>
- (2) <http://wol.ra.phy.cam.ac.uk/mackay/itprnn/#book>
- (3) <http://www.cs.cornell.edu/home/cardie/tutorial/dtrees>
- (4) <http://www.cse.unsw.edu.au/~quinlan/>
- (5) <http://www.dbmsmag.com>
- (6) <http://www.ics.uci.edu/~kibler/pubs/mlw92.ps>
- (7) <http://www.ltg.ed.ac.uk/~chrisbr/dilbook>
- (8) <http://www.rulequest.com/index.Html>
- (9) <http://www.rulequest.com/r113.html>
- (10) <http://www.rulequest.com/see5-previous.html>
- (11) <http://www.spss.com>
- (12) <http://www.datamining.com>
- (13) <http://yoda.cis.temple.edu:8080/UGAIWWW/lectures/C45/>
- (14) <ftp://ftp.ics.uci.edu/pub/machine-learning-databases>
- (15) [http://dms.irb.hr/tutorial/tut\\_dtrees.php](http://dms.irb.hr/tutorial/tut_dtrees.php)
- (16) [http://www.cs.uregina.ca/~dbd/cs831/notes/ml/dtrees/4\\_dtrees1.html](http://www.cs.uregina.ca/~dbd/cs831/notes/ml/dtrees/4_dtrees1.html)
- (17) <http://www.cs.uregina.ca/~dbd/cs831/notes/ml/dtrees/c4.5/tutorial.html>
- (18) <ftp://ftp.ics.uci.edu/pub/machine-learning-databases>
- (19) <http://www.erg.sri.com/people/marie/papers/cInl.ps>

## ΠΑΡΑΡΤΗΜΑΤΑ

### ΠΑΡΑΡΤΗΜΑ

- Σ. Συμβολισμοί μεταβλητών του αρχείου δεδομένων.
- Κ. Κωδικοί που χρησιμοποιήθηκαν.

Αρχείο αποτελεσμάτων S.P.S.S στα οποία ο χρήστης μπορεί να βλέπει την επίδραση διαφόρων παραγόντων στην ολική απώλεια του πλοίου.

- Π<sub>1</sub> Πίνακες μονοδιάστατης ανάλυσης.
- Π<sub>2</sub> Πίνακες δυσδιάστατης ανάλυσης.
- Π<sub>3</sub> Περιγραφικά στοιχεία.
- Δ. Δέντρα αποφάσεων για την ολική απώλεια.
- Τ. Ταξινομητικοί κανόνες.
- Π. Παράγοντες που ο συνδυασμός τους οδηγεί σε ολική απώλεια.

### ΠΑΡΑΡΤΗΜΑ Α

- Α. Η μεθοδολογία των νευρωνικών δικτύων.

### ΠΑΡΑΡΤΗΜΑ Β

- Α. Νέοι αλγόριθμοι και προγράμματα.
- Β. Ο αλγόριθμος C5.0. στα δεδομένα του αρχείου σε γλώσσα C.



## ΠΑΡΑΡΤΗΜΑ

### Σ. Συμβολισμοί μεταβλητών του αρχείου δεδομένων

Σύμβολο	Ερμηνεία
AA	Πεδίο ταξινόμησης
ACCIDE0	Ατύχημα
AGE	Ηλικία
AREATYPE	Είδος περιοχής
BUILT1	Έτος ναυπήγησης
CLASS	Νηογώμονας
DATYR	Χρονολογία ατυχήματος
GRT1	Χωρητικότητα (gri)
GRT1GROU	Σχετικό μέγεθος
OWNER	Πλήθος αλλαγής ιδιοκτητών
MONTH	Μήνας
NAREA	Γεωγραφική περιοχή
NOFACC	Πλήθος ατυχημάτων
OWNERS	Πλήθος αλλαγής ιδιοκτητών ομαδοποιημένο
SEQACC	Ακολουθία εγγραφών ανά πλοίο
TYPE	Τύπος του πλοίου
GRUCLASS	Νηογώμονες ομαδοποιημένοι
ACCSER	Σειρά ατυχήματος
BADVOYAGE	Μη επιτυχές ταξίδι
DACC92	Χρονική μεταβλητή ατυχήματος
DAY92	Χρονική μεταβλητή καταγραφής ατυχήματος
FLAG2	Σημεία
WEATH	Καιρικές συνθήκες
TOTLOSS	Ολική απώλεια

Κ Κωδικοί που χρησιμοποιήθηκαν.

4.3.1 Κωδικοί της μεταβλητής ACCIDEO

- 1 FOUNDERED
- 2 MISSING
- 3 FIRE-EXPLOSION
- 4 COLLISION (WITH ANOTHER SHIP)
- 5 CONTACT (INC. STRIKING ICE-ICEBERG SEE0
- 6 GROUNDED
- 8 HULL/MCHY/EQUIPMENT DAMAGED/

Αναλυτικά:

- 1 FOUNDERED
- 11 CAPSIZED
- 12 SANK-FOUNDERED
- 13 PRESUMED TO HAVE FOUNDERED
- 14 REPORTED SINKING
- 15 PART SUBMERGED
  
- 2 MISSING
- 20 MISSING
- 21 OVERDUE-REPORTED MISSING
  
- 3 FIRE-EXPLOSION
- 30 EXPLOSION
- 31 FIRE
- 32 SPARK (ELECTRIC)
- 33 SPARK (GAS CUTTING OR WELDING)
- 34 SPONTANEOUS COMBUSTION
- 35 STRUCK BY LIGHTNING
  
- 4 COLLISION (WITH ANOTHER SHIP)

- 40 COLLISION  
41 COLLISION (STRUCK BY)  
42 COLLISION AND CUT IN TWO-THREE
- 5 CONTACT (INC. STRIKING ICE-ICEBERG SEE  
50 CONTACT-STRUCK  
51 CONTACT-STRUCK BY
- 6 GROUNDED  
60 BEACHED (DELIBERATELY)  
61 GROUNDED-WRECKED-STRANDED  
62 WEDGED IN  
63 WEDGED UNDER  
64 TOUCHED BOTTOM (NOT HARD AND FAST)  
65 HOLED (WHILE AGROUND-NOT EVENT 1)
- 7 WAR LOSS-HOSTILITIES  
70 STRUCK BY PROJECTILE  
71 SABOTAGE
- 8 HULL/MCHY/EQUIPMENT DAMAGED/FAILURE  
(INC. SPRINGING LEAKS, DRAGGING ANCHORS, LOST RUDDER  
ETC.)  
80 BREAKDOWN  
81 BROKE  
82 BROKE IN TWO-THREE  
83 LOST (ITEM)  
84 CORRODED  
85 CRACKED  
86 DAMAGED  
87 OVERHEATED  
88 DAMAGE-DUE EXTINGUISHING OPERATIONS  
89 DEFECT  
801 DRIFTED



- 802 FAILURE
- 803 FRACTURED
- 804 FLOODED
- 805 LEAKED/SPILLED (NOT SPRANG LEAK)
- 806 DRAGGED
- 807 TROUBLE/DIFFICULTY
- 808 TOOK WATER
  
- 9 MISCELLANEOUS (INC. CARGO SHIFTING, TROUBLE)
- 90 CONTAMINATION
- 91 LISTED
- 92 ICED UP
- 93 WASHED-OVERFLOWED
- 94 SHIFTED
- 96 GAVE AWAY
- 97 GAS ACCUMULATION
- 98 NAVIGATIONAL ERROR
- 99 UNKNOWN
- 901 SEIZURE
- 902 SHORT CIRCUIT
- 903 SPRANG LEAKS
- 905 VIBRATION
- 906 WIPED

#### 4.3.4 Κωδικοί της μεταβλητής CLASS

- KI BIRO KLASIFICASI INDONESIA
- HR HELLENIC REGISTER OF SHIPPING
- CC CHINA CLASSIFICATION
- AB AMERICAN BUREAU OF SHIPPING
- BV BUREAU VERITAS
- GL GERMANISCHER LLOYD
- IR INDIAN REGISTER OF SHIPPING
- KR KOREAN REGISTER OF SHIPPING

LR LLOYD'S REGISTER OF SHIPPING  
NK NIPPON KAIJI KYOKAI  
NV DET NORSKE VERITAS CLASSIFICATION A/S  
PR POLSKI REJESTR STATKOW  
RI REGISTRO ITALIANO NAVALE  
RN ROMANIAN REGISTER OF SHIPPING  
RS RUSSIAN REGISTER OF SHIPPING  
Oth OTHER CLASSIFICATIONS

#### 4.3.5 Κωδικοί της μεταβλητής AREATYPE

1. OPEN SEA
2. CONTROLLED
3. PORT

#### 4.3.8 Κωδικοί της μεταβλητής NAREA

- 1 N. SEA, BALTIC, KIEL
- 2 MEDITERRANEAN, SUEZ
- 3 RED SEA E.& W. AFRICAN COAST
- 4 W. INDIES, U.S.EAST,GULF. MEX, GR.LAKES, NEWFOU
- 5 INDIAN OCEAN & ANTARTIC
- 6 GULF, BAY OF BENGAL
- 7 JAPAN, S. CHINA & E. INDIES
- 8 PACIFIC
- 9 AUSTRALIA
- 10 ATLANTIC
- 11 W. COAST S-N. AMERICA,PANAMA CAPE HORN
- 12 CANADIAN AL, U.S.S.R. ARTIC & BER. SEA,ICELAND

Αναλυτικά:

1. BR. ISLES, N. SEA, E. CHNL, BISCAY BAY
2. BALTIC
3. KIEL CANAL

4. E. MEDITERRANEAN & BLACK SEA
5. W. MEDITERRANEAN
6. SUEZ CANAL
7. RED SEA
8. GULF (PERSIAN)
9. E. AFRICAN COAST
10. INDIAN OCEAN & ANTARTIC
11. BAY OF BENGAL
12. S. CHINA & E. INDIES
13. JAPAN
14. N. PACIFIC
15. S. PACIFIC
16. AUSTRALIA
17. CAPE HORN
18. S. ATLANTIC, E. COAST, S. AMERICA
19. W. AFRICA COAST
20. W. COAST, S. AMERICA
21. N. AMERICA PACIFIC COAST
22. PANAMA CANAL
23. W. INDIES
24. U.S. EASTERN SEA BOARD
25. GULF OF MEXICO
26. GREAT LAKES
27. CANADIAN ARTIC & ALASKA
28. NEWFOUND LAND
29. ICELAND
30. U.S.S.R. ARTIC & BERNING SEA
31. N. ATLANTIC

#### 4.3.11 Κωδικοί της μεταβλητής TYPE

- 1-2. BULK
3. TANKER
- 4-9. FERRY



## 6. CONTAINER

Αναλυτικά:

- 2 BULK CARRIER
- 21 ORE CARRIER
- 22 BULK CARRIER
- 23 BULK/OIL CARRIER
- 24 ORE/OIL CARRIER
- 25 CEMENT CARRIER
- 26 WOODCHIPS CARRIER
- 27 UREA CARRIER
  
- 1 OTHER DRY CARGO
- 11 REFRIGERATED
- 12 GENERAL CARGO
- 13 PALLETIZED CARGO
- 14 DECK CARGO
- 15 LIVESTOCK CARRIER
- 16 BARGE CARRIER
  
- 3 TANKER
- 31 CHEM / OIL T.
- 32 CHEM TANKER
- 33 OIL TANKER
- 34 LIG-PET GAS T.
- 35 LIG-NOT PET GAS T. (MOLASSES,FRUIT)
- 36 LIG. GAS /CHEM TANKER
- 37 LIG. NAT. GAS TANKER
  
- 4 FERRY
- 41 RO-RO CARGO
- 42 CONTAINER/RO-RO CARGO
- 43 VEHICLES CARRIER
- 44 PASSENGER/RO-RO CARGO

- 45 PASSENGER/ GENERAL CARGO
  
- 5 FISHING
- 51 FISH CARRIER
- 52 FISH FACTORY
- 53 TRAWLER
  
- 6 CONTAINER
  
- 7 PONTOON
  
- 8 MISCELLANEOUS ACTIVITIES
- 81 OFFSHORE SUPPLY
- 82 ICE BREAKER
- 83 DREDGER
- 84 TUG/PUSHER TUG
- 85 DRILL RIG
- 86 FLOAT GRANE
- 87 FLOAT PR DOCK
  
- 9 PASSENGER
- 91 CATAMARAN
- 92 HOVERCRAFT
  
- 0 OTHER (INC. RESEARCH)

#### 4.3.13 Κωδικοί της μεταβλητής FLAG2

- 1 GREECE (GRC)
- 2 INDIA (IND)
- 3 SINGAPORE (SNG)
- 4 GREAT BRITAIN (GBI) (ENGLAND, U.K.)
- 5 KOREA (SOUTH) (KRS)
- 6 SPAIN (SPN)

- 7 LEBANON (LEB)
- 8 CHILE (CHL)
- 9 CYPRUS (CYP)
- 10 PANAMA (PAN)
- 11 LIBERIA (LIB)
- 12 CHINA, REPUBLIC OF (TAIWAN) (CHT)
- 13 BRAZIL (BRZ)
- 14 ITALY (ITL)
- 15 JAPAN (JPN)
- 16 SWEDEN (SWD)
- 17 NORWAY (NOR)
- 18 FRANCE (FRA)
- 19 UNITED STATES OF AMERICA (USA)
- 20 DOMINICAN REPUBLIC (DOM)
- 21 INDONESIA (IDA)
- 22 SRI LANKA (SRI)
- 23 DENMARK (DEN)
- 24 FIJI
- 25 NETHERLANDS (NTH)
- 26 GERMANY, FEDERAL REPUBLIC OF (GFR)
- 27 U.S.S.R. (USR)
- 28 EGYPT (EGY)
- 29 SWITZERLAND (SWZ)
- 30 POLAND (POL)
- 31 SOUTH AFRICA (SAF)
- 32 UNITED ARAB EMIRATES (UAE)
- 33 PHILIPPINES (PHI)
- 34 AUSTRALIA (AST)
- 35 FINLAND (FIN)
- 36 NETHERLANDS ANTILLES & ARUBA (NEA)
- 37 HONG KONG (HKG)
- 38 ECUADOR (ECU)
- 39 ARGENTINA (ARG)
- 40 ALGERIA (ALG)



- 42 MONACO (MON)
- 43 TURKEY (TRK)
- 44 VENEZUELA (VEN)
- 45 CAYMAN ISLANDS (CAY)
- 46 BERMUDA (BER)
- 47 CANADA (CAN)
- 48 PERU (PER)
- 49 OMAN (OMN)
- 50 YUGOSLAVIA (YUG)
- 51 HONDURAS (HON)
- 52 BULGARIA (BUL)
- 53 ROMANIA (ROM)
- 54 ISRAEL (ISR)
- 55 SAUDI ARABIA (SAU)
- 56 IRAN (IRN)
- 57 BURMA (BMA)
- 58 SOMALI REPUBLIC (SOM)
- 59 TUNISIA (TUN)
- 60 MEXICO (MEX)
- 61 BAHRAIN (BRN)
- 62 GUATEMALA (GUA)
- 63 KUWAIT (KUW)
- 64 MADAGASCAR (MGY)
- 65 GERMAN DEMOCRATIC REPUBLIC (GDR)
- 66 KENYA (KEN)
- 67 LIBYA (LBY)
- 68 NICARAGUA (NIC)
- 69 COLOMBIA (COL)
- 70 MALDIVE ISLANDS (MDV)
- 71 ALBANIA (ALB)
- 72 CUBA (CUB)
- 73 GHANA (GHA)
- 75 THAILAND (THA)
- 76 PAKISTAN (PAK)

- 77 MALAYSIA (MAL)  
78 PORTUGAL (PTG)  
80 ETHIOPIA (ETH)  
82 HUNGARY (HUN)  
83 DUBAI (DUB)  
84 TONGA (TON)  
85 MOROCCO (MOR)  
86 IRAQ (IRQ)  
87 URUGUAY (URG)  
88 BELGIUM (BLG)  
89 BANGLADESH (BNG)  
90 AUSTRIA (AUS)  
93 CHINA, PEOPLE'S REP. OF (CHR)  
94 VIETNAM (VNM)  
95 PAPUA NEW GUINEA (PNG)  
96 FAEROES (FAR)  
97 SYRIA (SYR)  
98 TANZANIA (TAN)  
101 NIGERIA (NIG)  
102 ICELAND (ICL)  
104 WESTERN SAMOA  
105 IRISH REPUBLIC (IRP)  
106 ANGOLA (ALA)  
107 NEW ZELAND (NZL)  
108 SURINAM (SUR)  
110 IVORY COAST  
111 TRINIDAD & TOBAGO (TRI)  
112 QUATAR  
113 MAURITIUS  
114 YEMENI  
115 SHARJAH (SHA)  
116 MALTA (MTA)  
117 BRITISH VIRGIN ISLANDS (BVI)  
118 UNITED STATES OF7 AMERICA (LAKES) (USL)

- 119 GUYANA (GUY)  
121 GREENLAND (GRN)  
122 VANUATU (VAN)  
124 SAINT VINCENT (SVC)  
125 MAURITANIA (MAU)  
127 SENEGAL (SEN)  
130 GUINEA BISSAU (GNB)  
132 GAMBIA (GAM)  
133 SAINT LUSIA (SLC)  
134 CANADA (LAKES) (CNL)  
136 DOMINICA (DMA)  
137 ANGUILA (AGL)  
139 SOLOMON ISLANDS (SLI)  
140 GIBRALTAR (GIB)  
142 COSTA RICA  
146 SAINT KITTS  
147 TURKS & CAICOS IS. (TKC)  
149 MOZAMBIQUE (MOZ)  
154 CAPE VERDE  
200 AJMAN (AJM)  
201 ANTIGUA 7 BARBUDA (ABB)  
202 BAHAMAS (BAH)  
203 MARSHALL ISLANDS (MAI)  
208 FALKLAND ISL  
209 CROATIA  
210 BELIZE (BZE)  
211 PUERTO RICO  
212 RUSSIA  
213 COMORO (CMR)  
214 COOK ISLANDS (CKI)  
215 AZORES  
216 BARBADOS  
217 BURUNDI  
218 BOLIVIA



- 219 TUVALU  
220 CAMEROON (CMN)  
221 CONGO  
222 CZECHOSLOVAKIA  
223 DENMARK (INTERNATIONAL) (DIS)  
224 EL SALVADOR  
225 FRENCH ANTARCTIC TERRITORY (FAT)  
226 GABON  
227 GRENADA  
228 DANAM  
229 HAITI  
230 ESTHONIA  
231 JAMAICA  
232 JORDAN  
233 KOREA (NORTH) (KRN)  
234 LUXEMBOURG  
235 UKRAINE  
236 NORWAY (INTERNATIONAL) (NIS)  
237 PARAGUAY (PAR)  
238 NAMIBIA  
239 SEYCHELLES  
240 SIERRA LEONE (SIE)  
241 LITHUANIA  
242 SUDAN  
243 TOGO (TOG)  
244 SAO TOME AND PRINCIPE  
245 MYANMAR (UNION OF MYANMAR)  
246 CAMBODIAN  
247 MADEIRA  
248 DGIBOUTI  
249 AZERBAIJAN  
250 GEORGIAN  
251 KERGUELEN ISLANDS  
252 MALAGASY

## 253 ISLE OF MAN

### 4.3.14 Κωδικοί της μεταβλητής WEATH

1. CALM WEATHER (CALM SEAS/GOOD WEATHER/GOOD VISIBILITY)
2. POOR VISIBILITY (FOG/MIST)
3. STORM (HEAVY WEATHER / BAD WEATHER / HEAVY SEAS / ROYGH SEAS / SOYALL / HEAVY SWELL)
4. SNOW STORM (SNOW / FREEZING CONDITIONS)
5. TYPHOON (HURRICANE / CYCLONE / TORNADO / FREAKWEATHER CONDITIONS / FREAK SEAS)
6. NO INFORMATION

### 4.3.16 Κωδικοί της μεταβλητής GRUCLASS

- 1 AB-GL-NK
- 2 BV-CC-KR-NV-RI-PR-RS
- 3 LR
- 4 OTHERS

### 4.3.19 Κωδικοί της μεταβλητής GRT1GROU

- 1 ΠΡΩΤΗ ΚΛΑΣΗ ΤΗΣ GRT1GROU
- 2 ΔΕΥΤΕΡΗ ΚΛΑΣΗ ΤΗΣ GRT1GROU
- 3 ΤΡΙΤΗ ΚΛΑΣΗ ΤΗΣ GRT1GROU
- 4 ΤΕΤΑΡΤΗ ΚΛΑΣΗ ΤΗΣ GRT1GROU
- 5 ΠΕΜΠΤΗ ΚΛΑΣΗ ΤΗΣ GRT1GROU
- 6 ΕΚΤΗ ΚΛΑΣΗ ΤΗΣ GRT1GROU
- 7 ΕΒΔΟΜΗ ΚΛΑΣΗ ΤΗΣ GRT1GROU
- 8 ΔΕΝ ΕΧΟΥΜΕ ΠΛΗΡΟΦΟΡΙΑ

### 4.3.20 Κωδικοί της μεταβλητής SEQACC

- 1 ΕΝΑ ΑΤΥΧΗΜΑ/ΠΛΟΙΟ
- 2 ΔΥΟ ΑΤΥΧΗΜΑΤΑ/ΠΛΟΙΟ

- 3 ΤΡΙΑ ΑΤΥΧΗΜΑΤΑ/ΠΛΟΙΟ
- 4 ΤΕΣΣΑΡΑ ΑΤΥΧΗΜΑΤΑ/ΠΛΟΙΟ
- 5 ΠΕΝΤΕ ΑΤΥΧΗΜΑΤΑ/ΠΛΟΙΟ
- 6 ΕΞΗ ΑΤΥΧΗΜΑΤΑ/ΠΛΟΙΟ
- 7 ΕΠΤΑ ΑΤΥΧΗΜΑΤΑ/ΠΛΟΙΟ
- 7 ΟΚΤΩ ΑΤΥΧΗΜΑΤΑ/ΠΛΟΙΟ
- 9 ΕΝΝΙΑ ΑΤΥΧΗΜΑΤΑ/ΠΛΟΙΟ

#### 4.3.23 Κωδικοί της μεταβλητής OWNERS

- 1 ΠΡΩΤΟ ΧΕΡΙ
- 2 ΔΕΥΤΕΡΟ ΧΕΡΙ
- 3 ΤΡΙΑ Η ΤΕΣΣΑΡΑ ΧΕΡΙΑ
- 4 ΑΠΟ ΠΕΝΤΕ ΚΑΙ ΠΑΝΩ ΧΕΡΙΑ
- 5 ΔΕΝ ΕΧΟΥΜΕ ΠΛΗΡΟΦΟΡΙΑ

Πανεπιστήμιο Πειραιώς



ΑΡΧΕΙΟ ΑΠΟΤΕΛΕΣΜΑΤΩΝ S.P.S.S ΣΤΑ ΟΠΟΙΑ Ο ΧΡΗΣΤΗΣ ΜΠΟΡΕΙ ΝΑ ΒΛΕΠΕΙ ΤΗΝ ΕΠΙΔΡΑΣΗ ΔΙΑΦΟΡΩΝ ΠΑΡΑΓΟΝΤΩΝ ΣΤΗΝ ΟΛΙΚΗ ΑΠΩΛΕΙΑ ΤΟΥ ΠΛΟΙΟΥ.

Π<sub>1</sub> Πίνακες μονοδιάστατης ανάλυσης

4.3.1 ACCIDE0

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	1 FOUNDERED	1022	5,0	5,0	5,0
	2 MISSING	407	2,0	2,0	7,0
	3 FIRE EXPLOS.	1823	9,0	9,0	16,0
	4 COLLISION	4073	20,0	20,0	36,0
	5 CONTACT	2097	10,3	10,3	46,3
	6 GROUNDED	3819	18,8	18,8	65,1
	8 HULL/M/E.DAM.	7096	34,9	34,9	100,0
	Total	20337	100,0	100,0	

4.3.4 CLASS

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	AB	1663	8,2	8,2	8,2
	BV	1898	9,3	9,3	17,5
	CC	120	,6	,6	18,1
	GL	1645	8,1	8,1	26,2
	HR	286	1,4	1,4	27,6
	IR	86	,4	,4	28,0
	KI	93	,5	,5	28,5
	KR	384	1,9	1,9	30,4
	LR	4446	21,9	21,9	52,2
	NK	1777	8,7	8,7	61,0
	NV	1545	7,6	7,6	68,6

	PR	150	,7	,7	69,3
	RI	288	1,4	1,4	70,7
	RN	166	,8	,8	71,5
	RS	581	2,9	2,9	74,4
	oth	5209	25,6	25,6	100,0
	Total	20337	100,0	100,0	

#### 4.3.5 AREATYPE

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	1 OPEN SEA	7361	36,2	36,2	36,2
	2 CONTROLLED	4056	19,9	19,9	56,1
	3 PORT	8920	43,9	43,9	100,0
	Total	20337	100,0	100,0	

#### 4.3.7 MONTH

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	Apr	1523	7,5	7,5	7,5
	Aug	1737	8,5	8,5	16,0
	Dec	1748	8,6	8,6	24,6
	Feb	1834	9,0	9,0	33,6
	Jan	2008	9,9	9,9	43,5
	Jul	1716	8,4	8,4	52,0
	Jun	1611	7,9	7,9	59,9
	Mar	1777	8,7	8,7	68,6
	May	1562	7,7	7,7	76,3
	Nov	1624	8,0	8,0	84,3
	Oct	1749	8,6	8,6	92,9
	Sep	1448	7,1	7,1	100,0
	Total	20337	100,0	100,0	

## 4.3.8 NAREA

		Frequency	Percent	Valid Percent	Cumulative Percent
VI	1 N. SEA,BALTIC, KIEL	5049	24,8	24,8	24,8
	2 SUEZ,MEDITERRANEAN	2828	13,9	13,9	38,7
	3 RED S.E.&W.AFR.COAST	1157	5,7	5,7	44,4
	4 W. INDIES,U.S. EAST, G. MEXICO, GR.LAKES, NFOU	2984	14,7	14,7	59,1
	5 INDIAN.OCEAN& ANT.	39	,2	,2	59,3
	6 GULF, BAY OF BENGAL	1356	6,7	6,7	66,0
	7 JAPAN, S.CHINA &E.IND	4063	20,0	20,0	85,9
	8 PACIFIC OCEAN	241	1,2	1,2	87,1
	9 AUSTRALIA	498	2,4	2,4	89,6
	10 ATLANTIC	919	4,5	4,5	94,1
	11 W.COAST, S-N. AMERICA,PANAMA,HORN	708	3,5	3,5	97,6
	12 CAN. ALASKA,U.S.S.R. ARTIC &BER. SEA. ICELAND	495	2,4	2,4	100,0
	Total	20337	100,0	100,0	

## 4.3.9 DATYR

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	92	217	1,1	1,1	1,1
	93	3419	16,8	16,8	17,9
	94	3538	17,4	17,4	35,3
	95	3456	17,0	17,0	52,3
	96	2795	13,7	13,7	66,0
	97	2780	13,7	13,7	79,7
	98	2191	10,8	10,8	90,5
	99	1941	9,5	9,5	100,0
	Total	20337	100,0	100,0	



## 4.3.11 TYPE

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	1-2 BULK	12616	62,0	62,0	62,0
	3 TANKER	3716	18,3	18,3	80,3
	4-9 FERRY	2722	13,4	13,4	93,7
	6 CONTAINER	1283	6,3	6,3	100,0
	Total	20337	100,0	100,0	

## 4.3.14 WEATH

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	1 CALM WEATHER	608	3,0	3,0	3,0
	2 POOR VISIBILITY	440	2,2	2,2	5,2
	3 STORM	2336	11,5	11,5	16,7
	4 SNOWSTORM	82	,4	,4	17,1
	5 TYPHOON	622	3,1	3,1	20,2
	No info	16249	79,8	79,8	100,0
	Total	20337	100,0	100,0	

## 4.3.15 TOTLOSS

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	2 No	18861	92,7	92,7	92,7
	1 Yes	1476	7,3	7,3	100,0
	Total	20337	100,0	100,0	

## 4.3.16 GRUCLASS

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	1 AB-GL-NK	5085	25,0	25,0	25,0
	2 BV-CC-KR-NV-RI-PR-RS	4966	24,4	24,4	49,4
	3 LR	4446	21,9	21,9	71,3
	4 OTHERS	5840	28,7	28,7	100,0
	Total	20337	100,0	100,0	

## 4.3.19 GRT1GROU

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	1 ΠΡΩΤΗ ΚΛΑΣΗ	12674	62,3	62,3	62,3
	2 ΔΕΥΤ. ΚΛΑΣΗ	2979	14,6	14,6	76,9
	3 ΤΡΙΤΗ ΚΛΑΣΗ	1259	6,2	6,2	83,1
	4 ΤΕΤΑΡ. ΚΛΑΣΗ	715	3,5	3,5	86,6
	5 ΠΕΜΠ. ΚΛΑΣΗ	291	1,4	1,4	88
	6 ΕΚΤΗ ΚΛΑΣΗ	197	1,0	1,0	89
	7 ΕΒΔΟΜ. ΚΛΑΣΗ	100	,5	,5	89,5
	No info	2122	10,5	10,5	100,0
	Total	20337	100,0	100,0	

## 4.3.20 SEQACC

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	1 ΕΝΑ ΑΤΥΧΗΜΑ/ΠΛΟΙΟ	17081	84,0	84,0	84,0
	2 ΔΥΟ ΑΤΥΧΗΜΑΤΑ/ΠΛΟΙΟ	2361	11,6	11,6	95,6
	3 ΤΡΙΑ ΑΤΥΧΗΜΑΤΑ/ΠΛΟΙΟ	642	3,2	3,2	98,8
	4 ΤΕΣΣΕΡΑ ΑΤΥΧΗΜΑΤΑ/ΠΛΟΙΟ	182	,9	,9	99,7
	5 ΠΕΝΤΕ ΑΤΥΧΗΜΑΤΑ/ΠΛΟΙΟ	39	,2	,2	99,8

	6 ΕΞΗ ΑΤΥΧΗΜΑΤΑ/ΠΛΟΙΟ	21	,1	,1	99,9
	7 ΕΠΤΑ ΑΤΥΧΗΜΑΤΑ/ΠΛΟΙΟ	7	,0	,0	100,0
	8 ΟΚΤΩ ΑΤΥΧΗΜΑΤΑ/ΠΛΟΙΟ	2	,0	,0	100,0
	9 ΕΝΝΙΑ ΑΤΥΧΗΜΑΤΑ/ΠΛΟΙΟ	2	,0	,0	100,0
	Total	20337	100,0	100,0	

#### 4.3.21 NOFACC

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	1 ΕΝΑ ΑΤΥΧΗΜΑ	15436	15436	75,9	75,9
	2 ΔΥΟ ΑΤΥΧΗΜΑΤΑ	3989	3989	19,6	95,5
	3 ΤΡΙΑ ΑΤΥΧΗΜΑΤΑ	691	691	3,4	98,9
	4 ΤΕΣΣΑΡΑ ΑΤΥΧΗΜΑΤΑ	173	173	,9	99,8
	5 ΠΕΝΤΕ ΑΤΥΧΗΜΑΤΑ	48	48	,2	100,0
	Total	20337	20337	100,0	

#### 4.3.23 OWNERS

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	1 ΠΡΩΤΟ ΧΕΡΙ	6195	30,5	30,5	30,5
	2 ΔΕΥΤΕΡΟ ΧΕΡΙ	3601	17,7	17,7	48,2
	3 ΤΡΙΑ-ΤΕΣΣΕΡΑ ΧΕΡΙΑ	4181	20,6	20,6	68,7
	4 ΠΕΝΤΕ ΚΑΙ ΠΑΝΩ ΧΕΡΙΑ	1959	9,6	9,6	78,4
	5 No info	4401	21,6	21,6	100,0
	Total	20337	100,0	100,0	



Π<sub>2</sub> Πίνακες δυσδιάστατης ανάλυσης

NAREA

		Total	Totlos					
			No			Yes		
			Count	Row %	Col %	Cou nt	Row %	Col %
NAR EA	1 N. SEA, BALTIC, KIEL	5049	4849	96,0 %	25,7 %	200	4,0 %	13,6 %
	2 SUEZ, MEDITERRANEAN	2828	2640	93,4 %	14,0 %	188	6,6 %	12,7 %
	3 RED SEA E.& W. AFRICAN COAST	1157	1021	88,2 %	5,4 %	136	11,8 %	9,2 %
	4W. INDIES, U.S. EAST,G. MEXICO, GR LAK., NEWFOU	2984	2857	95,7 %	15,1 %	127	4,3 %	8,6% %
	5 INDIAN OC. & ANTARTIC	39	31	79,5 %	,2% %	8	20,5 %	,5% %
	6 GULF, BAY OF BENGAL	1356	1183	87,2 %	6,3 %	173	12,8 %	11,7 %
	7 JAPAN, S. CHINA & E. INDIES	4063	3577	88,0 %	19,0 %	486	12,0 %	32,9 %
	8 PACIFIC	241	231	95,9 %	1,2 %	10	4,1 %	,7% %
	9 AUSTRALIA	498	491	98,6 %	2,6 %	7	1,4 %	,5% %
	10,00 ATLANTIC	919	865	94,1 %	4,6 %	54	5,9 %	3,7% %
	11 W.COAST,SN.AM ERICA,PAN.HORN	708	663	93,6 %	3,5 %	45	6,4 %	3,0% %
	12ALASKA,U.S.S.R. AR.&BER.,ICELAN	495	453	91,5 %	2,4 %	42	8,5 %	2,8% %
Total		20337	18861	92,7 %	100, 0%	1476	7,3 %	100,0 %

AREA TYPE

		Total	total loss					
			No			Yes		
			Count	Row %	Col %	Count	Row %	Col %
AREA TYPE	1 OPEN SEA	7361	6631	90,1%	35,2%	730	9,9%	49,5%
	2CONTROLLED	4056	3825	94,3%	20,3%	231	5,7%	15,7%
	3 PORT	8920	8405	94,2%	44,6%	515	5,8%	34,9%
Total		20337	18861	92,7%	100,0%	1476	7,3%	100,0%

WEATH

		Total	totloss					
			No			Yes		
			Count	Row %	Col %	Count	Row %	Col %
WEATH	1 CALM WEATH	608	567	93,3%	3,0%	41	6,7%	2,8%
	2 POOR VISIBILITY	440	405	92,0%	2,1%	35	8,0%	2,4%
	3 STORM	2336	1981	84,8%	10,5%	355	15,2%	24,1%
	4SNOWSTOR	82	77	93,9%	,4%	5	6,1%	,3%
	5 TYPHOON	622	526	84,6%	2,8%	96	15,4%	6,5%
	6 No info	16249	15305	94,2%	81,1%	944	5,8%	64,0%
Total		20337	18861	92,7%	100,%	1476	7,3%	100,0%

ACCIDEO

	Total	total loss						
		No			Yes			
ACCIDEO		Count	Row %	Col %	Count	Row %	Col %	
1 FOUNDERED	1022	463	45,3%	2,5%	559	54,7%	37,9%	
2 MISSING	407	317	77,9%	1,7%	90	22,1%	6,1%	
3 FIRE-EXP.	1823	1635	89,7%	8,7%	188	10,3%	12,7%	
4 COLLISION	4073	3887	95,5%	20,6%	186	4,6%	12,6%	
5 CONTACT	2097	2049	97,6%	10,9%	48	2,3%	3,3%	
6 GROUNDED	3819	3592	94,1%	19,0%	227	5,9%	15,4%	
8 H.M.E.DAM.	7096	6918	97,5%	36,7%	178	2,5%	12,1%	
Total	20337	18861	92,7%	100,0%	1476	7,3%	100,0%	

DATYR

		Total	total loss					
			No			Yes		
DATYR		Count	Row %	Col %	Count	Row %	Col %	
DATYR	92	217	195	89,9%	1,0%	22	10,1%	1,5%
	93	3419	3153	92,2%	16,7%	266	7,8%	18,0%
	94	3538	3293	93,1%	17,5%	245	6,9%	16,6%
	95	3456	3257	94,2%	17,3%	199	5,8%	13,5%
	96	2795	2559	91,6%	13,6%	236	8,4%	16,0%
	97	2780	2613	94,0%	13,9%	167	6,0%	11,3%
	98	2191	2029	92,6%	10,8%	162	7,4%	11,0%
	99	1941	1762	90,8%	9,3%	179	9,2%	12,1%
Total		20337	18861	92,7%	100,0%	1476	7,3%	100,0%



## GRT1GROU

		Total	total loss					
			No			Yes		
			Count	Row %	Col %	Count	Row %	Col %
GRT1GROU	1	12674	11760	92,8%	62,4%	914	7,2%	61,9%
	2	2979	2894	97,1%	15,3%	85	2,9%	5,8%
	3	1259	1206	95,8%	6,4%	53	4,2%	3,6%
	4	715	693	96,9%	3,7%	22	3,1%	1,5%
	5	291	269	92,4%	1,4%	22	7,6%	1,5%
	6	197	190	96,4%	1,0%	7	3,6%	,5%
	7	100	99	99,0%	,5%	1	1,0%	,1%
	No info	2122	1750	82,5%	9,3%	372	17,5%	25,2%
Total		20337	18861	92,7%	100,0%	1476	7,3%	100,0%

## OWNERS

	Total	total loss					
		No			Yes		
OWNERS		Count	Row %	Col %	Count	Row %	Col %
1	6195	5939	95,9%	31,5%	256	4,1%	17,3%
2	3601	3367	93,5%	17,9%	234	6,5%	15,9%
3	4181	3939	94,2%	20,9%	242	5,8%	16,4%
4	1959	1787	91,2%	9,5%	172	8,8%	11,7%
5 No info	4401	3829	87,0%	20,3%	572	13,0%	38,8%
Total	20337	18861	92,7%	100,0%	1476	7,3%	100,0%

## MONTH

	Total	total loss					
MONTH		No			Yes		
		Count	Row %	Col %	Count	Row %	Col %
Apr	1523	1430	93,9%	7,6%	93	6,1%	6,3%
Aug	1737	1595	91,8%	8,5%	142	8,2%	9,6%
Dec	1748	1611	92,2%	8,5%	137	7,8%	9,3%
Feb	1834	1687	92,0%	8,9%	147	8,0%	10,0%
Jan	2008	1879	93,6%	10,0%	129	6,4%	8,7%
Jul	1716	1572	91,6%	8,3%	144	8,4%	9,8%
Jun	1611	1503	93,3%	8,0%	108	6,7%	7,3%
Mar	1777	1685	94,8%	8,9%	92	5,2%	6,2%
May	1562	1442	92,3%	7,6%	120	7,7%	8,1%
Nov	1624	1485	91,4%	7,9%	139	8,6%	9,4%
Oct	1749	1610	92,1%	8,5%	139	7,9%	9,4%
Sep	1448	1362	94,1%	7,2%	86	5,9%	5,8%
Total	20337	18861	92,7%	100,0%	1476	7,3%	100,0%

## CLASS

	Total	total loss					
		No			Yes		
CLAS S		Count	Row %	Col %	Count	Row %	Col %
AB	1663	1624	97,7%	8,6%	39	2,3%	2,6%
BV	1898	1762	92,8%	9,3%	136	7,2%	9,2%
CC	120	109	90,8%	,6%	11	9,2%	,7%
GL	1645	1563	95,0%	8,3%	82	5,0%	5,6%
HR	286	252	88,1%	1,3%	34	11,9%	2,3%
IR	86	82	95,3%	,4%	4	4,7%	,3%

KI	93	66	71,0%	,3%	27	29,0%	1,8%
KR	384	356	92,7%	1,9%	28	7,3%	1,9%
LR	4446	4245	95,5%	22,5%	201	4,5%	13,6%
NK	1777	1662	93,5%	8,8%	115	6,5%	7,8%
NV	1545	1488	96,3%	7,9%	57	3,7%	3,9%
PR	150	143	95,3%	,8%	7	4,7%	,5%
RI	288	268	93,1%	1,4%	20	6,9%	1,4%
RN	166	162	97,6%	,9%	4	2,4%	,3%
RS	581	547	94,1%	2,9%	34	5,9%	2,3%
oth	6754	6020	89,1%	31,9%	734	10,9%	49,7%
Total	20337	18861	92,7%	100,0%	1476	7,3%	100,0%

#### NOFACC

	Total	total loss					
		No			Yes		
NOFACC		Count	Row %	Col %	Count	Row %	Col %
1	15436	14621	94,7%	77,5%	815	5,3%	55,2%
2	3989	3527	88,4%	18,7%	462	11,6%	31,3%
3	691	555	80,3%	2,9%	136	19,7%	9,2%
4	173	125	72,3%	0,7%	48	27,7%	3,3%
5	48	33	68,8%	0,2%	15	31,3%	1,0%
Total	20337	18861	92,7%	100,0%	1476	7,3%	100,0%



Π<sub>3</sub> Περιγραφικά στοιχεία (Descriptives)

OWNER

	N	Mean	Std. Deviation	Std. Error	95%Conf. Int. for M	Lower Bound	Upper Bound	Minimum	Maximum
TOTAL LOSS No	15032	2,4665	1,76775	,01442	2,4383	2,4948		1,	15,00
TOTAL LOSS Yes	904	2,8518	1,85765	,06178	2,7305	2,9730		1,	11,00
Total	15936	2,4884	1,77515	,01406	2,4608	2,5160		1,	15,00

GRT1

	N	Mean	Std. Dev.	Std. Error	95%Conf. Int. for M	Lower Bound	Upper Bound	Minimum	Maximum
TOTAL LOSS No	17111	15682,71	23184,546	177,240	15335,30	16030,11		1	598002
TOTAL LOSS Yes	1104	8310,48	15504,996	466,646	7394,86	9226,09		4	159222
Total	18215	15235,88	22860,347	169,382	14903,87	15567,88		1	598002

## AGE

	N	Mean	Std. Dev.	Std. Error	95%Conf. Int. for M		Minimum	Maximum
						Lower Bound	Upper Bound	
TOTAL LOSS No	16561	17,17	9,394	,073	17,03	17,32	0	99
TOTAL LOSS Yes	974	22,70	10,501	,336	22,04	23,36	0	106
Total	17535	17,48	9,543	,072	17,34	17,62	0	106

## BUILT1

	N	Mean	Std. Dev.	Std. Error	95%Conf. Int. for M		Minimum	Maximum
						Lower Bound	Upper Bound	
TOTAL LOSS No	18861	1978,40	9,511	,069	1978,27	1978,54	1898	1999
TOTAL LOSS Yes	1476	1974,55	10,14	,264	1974,04	1975,07	1891	1999
Total	20337	1978,12	9,609	,067	1977,99	1978,26	1891	1999

## DAY92

	N	Mean	Std. Dev	Std. Error	95%Conf. Int. for M		Min	Maximum
						Lower Bound	Upper Bound	
TOTAL LOSS No	18860	1482,3244	714,07168	5,19961	1472,1327	1492,5161	342,	2904,
TOTAL LOSS Yes	1477	1512,3645	751,49073	19,56052	1473,9951	1550,7339	342,	2904,
Total	20337	1484,5047	716,87646	5,02703	1474,6513	1494,3581	342,00	2904,00

## DACC92

	N	Mean	Std. Dev.	Std. Error	95%Conf. Int. for M		Minimum	Maximum
						Lower Bound	Upper Bound	
TOTAL LOSS No	18861	1454,3982	723,10192	5,26523	1444,0779	1464,7185	-8708,0	2904,00
TOTAL LOSS Yes	1476	1423,5617	804,85688	20,94959	1382,4675	1464,6558	-4130,0	2904,00
Total	20337	1452,1602	729,36629	5,11449	1442,1354	1462,1850	-8708,0	2904,00



## SEQACC

	N	Mean	Std. Deviation	Std. Error	95%Conf Int. for M		Minimum	Maximum
					Lower Bound	Upper Bound		
TOTAL LOSS No	18861	1,23	,607	,004	1,22	1,24	1	9
TOTAL LOSS Yes	1476	1,12	,391	,010	1,10	1,14	1	4
Total	20337	1,22	,595	,004	1,21	1,23	1	9

## NOFACC

	N	Mean	Std. Dev.	Std. Error	95%Conf Int. for M		Minimum	Maximum
					Lower Bound	Upper Bound		
TOTAL LOSS No	18861	1,2727	,56390	,00411	1,2647	1,2808	1,00	5,00
TOTAL LOSS Yes	1476	1,6355	,85644	,02229	1,5918	1,6792	1,00	5,00
Total	20337	1,2991	,59746	,00419	1,2908	1,3073	1,00	5,00

Δ. Δέντρα αποφάσεων για την ολική απώλεια

ΔΕΝΔΡΟ ΑΠΟΦΑΣΗΣ Δ<sub>1</sub>

ACCIDE0 ['COLLISION(WITH SHIP)' 'CONTACT(INC. STRIKING ICE)' 'FIRE-  
EXPLOSION' 'GROUNDED' 'HULL/MCHY/EQUIP.DAMAGED' 'MISSING']  
[Mode: No] (24706, 0.949) -> No

ACCIDE0 ['FOUNDERED'] [Mode: Yes] (2958)

GRT1 =< 16518 [Mode: Yes] (2356.14)

NAREA ['ATLANTIC' 'AUSTRALIA' 'CANADIAN' 'ALASKA, U.S.S.R.  
ARTIC & BERNING SEA, ICELAND' 'MEDITERRANEAN, SUEZ' 'N. SEA,  
BALTIC, KIEL' 'PACIFIC' 'W. COAST S-N. AMERICA, PANAMA CAPE HORN'  
'W. INDIES, U.S. EASTERN, GULF OF MEXICO, GREAT LAKES, NEWFOU']  
[Mode: No] (1128.314)

AREATYPE ['controlled' 'port'] [Mode: No] (608.631)

NOFACC =< 1 [Mode: No] (314.129, 0.679) -> No

NOFACC > 1 [Mode: No] (294.502)

GRT1GROU ['1'] [Mode: No] (210)

WEATH ['CALM WEATHER' 'TYPHOON'] [Mode: Yes] (21, 0.762)

-> Yes

WEATH ['No info'] [Mode: No] (136)

BUILT1 =< 1977 [Mode: No] (99)

OWNERS ['1' '3-4' '5++'] [Mode: No] (52, 0.75) -> No

OWNERS ['2' 'No info'] [Mode: Yes] (47)

BUILT1 =< 1968 [Mode: No] (15, 0.667) -> No

BUILT1 > 1968 [Mode: Yes] (32, 0.813) -> Yes

BUILT1 > 1977 [Mode: No] (37, 0.892) -> No

WEATH ['POOR VISIBILITY' 'SNOWSTORM'] [Mode: No] (14,

0.643) -> No

WEATH ['STORM'] [Mode: Yes] (39)

GRUCLASS ['AB-GL-NK' 'BV-CC-KR-NV-RI-PR-RS' 'LR']

[Mode: Yes] (31, 0.613) -> Yes

GRUCLASS ['OTHERS'] [Mode: No] (8, 0.875) -> No  
 GRT1GROU ['2' '4'] [Mode: Yes] (14, 0.643) -> Yes  
 GRT1GROU ['3' '5' '6' '7'] [Mode: No] (2, 1.0) -> No  
 GRT1GROU ['No info'] [Mode: Yes] (68.502)  
 WEATH ['CALM WEATHER' 'No info' 'POOR VISIBILITY'  
 'SNOWSTORM'] [Mode: Yes] (47.792, 0.733) -> Yes  
 WEATH ['STORM' 'TYPHOON'] [Mode: No] (20.71, 0.692) -> No  
 AREATYPE ['open sea'] [Mode: Yes] (519.683)  
 TYPE ['BULK' 'FERRY'] [Mode: Yes] (451.229)  
 CLASS ['AB' 'HR' 'KI' 'KR' 'NV' 'RI'] [Mode: No] (53, 0.623) -> No  
 CLASS ['BV' 'CC' 'IR' 'NK' 'PR' 'RN' 'RS' 'oth'] [Mode: Yes] (290.229,  
 0.637) -> Yes  
 CLASS ['GL'] [Mode: No] (46)  
 OWNERS ['1' '2' '3-4' 'No info'] [Mode: No] (26, 0.731) -> No  
 OWNERS ['5\*\*\*'] [Mode: Yes] (20, 0.7) -> Yes  
 CLASS ['LR'] [Mode: Yes] (62)  
 NAREA ['ATLANTIC' 'CANADIAN ALASKA, U.S.S.R. ARTIC &  
 BERNING SEA, ICELAND' 'PACIFIC' 'W. COAST S-N. AMERICA, PANAMA  
 CAPE HORN'] [Mode: No] (15, 0.8) -> No  
 NAREA ['AUSTRALIA' 'MEDITERRANEAN, SUEZ' 'N. SEA,  
 BALTIC, KIEL' 'W. INDIES, U.S. EASTERN, GULF OF MEXICO, GREAT  
 LAKES, NEWFOU'] [Mode: Yes] (47, 0.638) -> Yes  
 TYPE ['CONTAINER' 'TANKER'] [Mode: No] (68.454, 0.656) -> No  
 AREATYPE ['No info'] [Mode: No] (0.0) -> No  
 NAREA ['GULF, BAY OF BENGAL' 'INDIAN OCEAN & ANTARTIC'  
 'JAPAN, S. CHINA & E. INDIES' 'RED SEA E.& W. AFRICAN COAST'] [Mode:  
 Yes] (1227.825)  
 MONTH ['Apr' 'Jan' 'Jun' 'Mar' 'Sep'] [Mode: Yes] (497.432)  
 GRUCLASS ['AB-GL-NK' 'BV-CC-KR-NV-RI-PR-RS' 'LR'] [Mode: No]  
 (302.39)  
 NAREA ['GULF, BAY OF BENGAL'] [Mode: No] (63.797, 0.734) ->  
 No  
 NAREA ['INDIAN OCEAN & ANTARTIC'] [Mode: Yes] (4, 0.5) ->  
 Yes



NAREA ['JAPAN, S. CHINA & E. INDIES'] [Mode: Yes] (190.796)  
 OWNERS ['1'] [Mode: No] (47, 0.66) -> No  
 OWNERS ['2' '5\*\*\*' 'No info'] [Mode: Yes] (83.797, 0.618) -> Yes  
 OWNERS ['3-4'] [Mode: Yes] (60)  
 MONTH ['Apr' 'Jan' 'Sep'] [Mode: No] (33, 0.727) -> No  
 MONTH ['Jun' 'Mar'] [Mode: Yes] (27, 0.778) -> Yes  
 NAREA ['RED SEA E.& W. AFRICAN COAST'] [Mode: No] (43.797)  
 GRUCLASS ['AB-GL-NK' 'BV-CC-KR-NV-RI-PR-RS'] [Mode: No]  
 (30, 0.633) -> No  
 GRUCLASS ['LR'] [Mode: Yes] (13.796, 0.783) -> Yes  
 GRUCLASS ['OTHERS'] [Mode: Yes] (195.042)  
 OWNERS ['1' 'No info'] [Mode: Yes] (183.449)  
 SEQACC =< 1 [Mode: Yes] (171.449)  
 DATYR =< 95 [Mode: No] (68.623)  
 DATYR =< 93 [Mode: Yes] (21.524, 0.694) -> Yes  
 DATYR > 93 [Mode: No] (47.099)  
 MONTH ['Apr' 'Jun' 'Mar' 'Sep'] [Mode: No] (36.32, 0.83) ->  
 No  
 MONTH ['Jan'] [Mode: Yes] (10.779, 0.926) -> Yes  
 DATYR > 95 [Mode: Yes] (102.827, 0.699) -> Yes  
 SEQACC > 1 [Mode: Yes] (12, 0.833) -> Yes  
 OWNERS ['2' '5\*\*\*'] [Mode: Yes] (10.593, 1.0) -> Yes  
 OWNERS ['3-4'] [Mode: No] (1, 1.0) -> No  
 MONTH ['Aug' 'Dec' 'Feb' 'Jul' 'May' 'Nov' 'Oct'] [Mode: Yes] (730.393, 0.703)  
 -> Yes  
 GRT1 > 16518 [Mode: No] (601.86)  
 GRT1GROU ['1' '2' '3' '4' '5' '7'] [Mode: No] (471)  
 BUILT1 =< 1973 [Mode: No] (137)  
 CLASS ['AB' 'BV' 'GL' 'CC' 'IR' 'KI' 'KR' 'NV' 'PR' 'RN'] [Mode: No] (39,  
 0.769) -> No  
 CLASS ['HR' 'RI' 'RS' 'oth'] [Mode: Yes] (25, 0.64) -> Yes  
 CLASS ['LR'] [Mode: No] (22)  
 GRT1 =< 22841 [Mode: Yes] (12, 0.75) -> Yes  
 GRT1 > 22841 [Mode: No] (10, 0.9) -> No

CLASS ['NK'] [Mode: Yes] (51)

NOFACC =< 1 [Mode: No] (38, 0.553) -> No

NOFACC > 1 [Mode: Yes] (13, 0.692) -> Yes

BUILT1 > 1973 [Mode: No] (334)

NAREA ['ATLANTIC' 'AUSTRALIA' 'CANADIAN ALASKA, U.S.S.R.  
ARTIC & BERNING SEA, ICELAND' 'INDIAN OCEAN & ANTARTIC' 'N. SEA,  
BALTIC, KIEL' 'W. INDIES, U.S. EASTERN, GULF OF MEXICO, GREAT  
LAKES, NEWFOU'] [Mode: No] (154, 0.981) -> No

NAREA ['GULF, BAY OF BENGAL' 'JAPAN, S. CHINA & E. INDIES'  
'MEDITERRANEAN, SUEZ' 'PACIFIC' 'RED SEA E.& W. AFRICAN COAST' 'W.  
COAST S-N. AMERICA, PANAMA CAPE HORN'] [Mode: No] (180)

AREATYPE ['controlled' 'port'] [Mode: No] (136, 0.875) -> No

AREATYPE ['open sea'] [Mode: No] (44)

GRT1 =< 21716 [Mode: Yes] (11, 0.818) -> Yes

GRT1 > 21716 [Mode: No] (33, 0.727) -> No

GRT1GROU ['6' 'No info'] [Mode: Yes] (130.86, 0.634) -> Yes

ΔΕΝΔΡΟ ΑΠΟΦΑΣΗΣ Δ<sub>2</sub>

- ACCIDE0 ['COLLISION(WITH SHIP)' 'CONTACT(INC. STRIKING ICE)' 'FIRE-  
EXPLOSION' 'GROUNDED'  
'HULL/MCHY/EQUIP.DAMAGED' 'MISSING']  
[Mode: No] (24707.29)
- ACCIDE0 ['COLLISION(WITH SHIP)' 'CONTACT(INC. STRIKING ICE)'  
'GROUNDED' 'HULL/MCHY/EQUIP.DAMAGED']  
[Mode: No] (21589.196)
- WEATH ['CALM WEATHER' 'No info' 'POOR VISIBILITY' 'SNOWSTORM']  
[Mode: No] (18143.998)
- NOFACC =< 1 [Mode: No] (12138.3, 0.986) -> No
- NOFACC > 1 [Mode: No] (6002.699)
- WEATH ['STORM' 'TYPHOON'] [Mode: No] (3444.96)
- ACCIDE0 ['FIRE-EXPLOSION' 'MISSING'] [Mode: No] (3120.009)
- NOFACC =< 1 [Mode: No] (1813)
- NOFACC > 1 [Mode: No] (1306.99)
- ACCIDE0 ['FOUNDERED'] [Mode: Yes] (2957.98)
- GRT1 =< 16518 [Mode: Yes] (2356.12)
- NAREA ['ATLANTIC' 'AUSTRALIA' 'CANADIAN ALASKA, U.S.S.R.  
ARTIC & BERNING SEA, ICELAND'  
'MEDITERRANEAN, SUEZ' 'N. SEA, BALTIC, KIEL'  
'PACIFIC' 'W. COAST S-N. AMERICA, PANAMA  
CAPE HORN' 'W. INDIES, U.S. EASTERN, GULF OF  
MEXICO, GREAT LAKES, NEWFOU'] [Mode: Yes]  
(1128.315)
- NAREA ['GULF, BAY OF BENGAL' 'INDIAN OCEAN & ANTARTIC'  
'JAPAN, S. CHINA & E. INDIES' 'RED SEA E.& W.  
AFRICAN COAST'] [Mode: Yes] (1227.824, 0.632) ->  
Yes
- GRT1 > 16518 [Mode: Yes] (601.861)



ΔΕΝΔΡΟ ΑΠΟΦΑΣΗΣ Δ<sub>3</sub>

- GRUCLASS ['AB-GL-NK' 'BV-CC-KR-NV-RI-PR-RS' 'LR'] [Mode: No]  
 BUILT1 =< 1974 [Mode: Yes]  
 NAREA ['ATL' 'AUS' 'ARTIC' 'INDIAN' 'MED' 'N.SEA' 'PACIFIC' 'GULF OF MEXICO'] [Mode: No]  
 BUILT1 =< 1968 [Mode: Yes]  
 BUILT1 > 1968 [Mode: No]  
 NAREA ['BENGAL' 'JAPAN' 'AFRICA' 'W.COAST'] [Mode: Yes]  
 BUILT1 > 1974 [Mode: No]  
 NAREA ['ATL' 'AUS' 'ARTIC' 'MED' 'N.SEA' 'PACIFIC' 'W.COAST' 'GULF OF MEXICO'] [Mode: No]  
 NAREA ['ATL' 'ARTIC' 'MED' 'N.SEA' 'PACIFIC' 'W.COAST' 'GULF OF MEXICO'] [Mode: No]  
 BUILT1 =< 1981 [Mode: No]  
 BUILT1 > 1981 [Mode: No]  
 NAREA ['AUS'] [Mode: No] -> No  
 NAREA ['BENGAL' 'INDIAN' 'JAPAN' 'AFRICA'] [Mode: No]  
 GRUCLASS ['OTHERS'] [Mode: Yes]  
 NAREA ['ATL' 'AUS' 'MED' 'N.SEA' 'PACIFIC' 'W.COAST' 'GULF OF MEXICO'] [Mode: Yes]  
 TYPE ['BULK' 'FERRY' 'TANKER'] [Mode: Yes]  
 OWNERS ['1'] [Mode: No]  
 NAREA ['ATL'] [Mode: No]  
 NAREA ['AUS'] [Mode: No]  
 NAREA ['MED'] [Mode: No]  
 NAREA ['N.SEA'] [Mode: No]  
 NAREA ['PACIFIC'] [Mode: No] -> No  
 NAREA ['W.COAST'] [Mode: No]  
 NAREA ['GULF OF MEXICO'] [Mode: Yes]  
 OWNERS ['2' '3-4' '5+' 'No info'] [Mode: Yes]  
 BUILT1 =< 1980 [Mode: Yes]  
 BUILT1 > 1980 [Mode: No]

TYPE ['CONTAINER'] [Mode: No]  
NAREA ['ARTIC' 'BENGAL' 'INDIAN' 'JAPAN' 'AFRICA'] [Mode: Yes]  
BUILT1 =< 1982 [Mode: Yes]  
OWNERS ['1' '2' 'No info'] [Mode: Yes]  
OWNERS ['3-4' '5++'] [Mode: Yes]  
BUILT1 > 1982 [Mode: Yes]  
TYPE ['BULK' 'CONTAINER' 'TANKER'] [Mode: Yes]  
TYPE ['FERRY'] [Mode: Yes]

Πανεπιστήμιο Πειραιώς

## Τ. Ταξινομητικοί κανόνες

T<sub>Yes</sub> Ταξινομητικοί κανόνες για την ολική απώλεια

Rules for Yes:

Rule #1 for Yes:

if ACCIDE0 = FOUNDERED  
and GRT1 <= 16518  
and NAREA = [ATLANTIC AUSTRALIA CANADIAN ALASKA, U.S.S.R.  
ARTIC & BERNING SEA, ICELAND MEDITERRANEAN, SUEZ N. SEA,  
BALTIC, KIEL PACIFIC W. COAST S-N. AMERICA, PANAMA CAPE HORN W.  
INDIES, U.S. EASTERN, GULF OF MEXICO, GREAT LAKES, NEWFOU]  
and AREATYPE = [controlled port]  
and NOFACC > 1  
and GRT1GROU = 1  
and WEATH = [CALM WEATHER TYPHOON]  
then -> Yes (21, 0.762)

Rule #2 for Yes:

if ACCIDE0 = FOUNDERED  
and GRT1 <= 16518  
and NAREA = [ATLANTIC AUSTRALIA CANADIAN ALASKA, U.S.S.R.  
ARTIC & BERNING SEA, ICELAND MEDITERRANEAN, SUEZ N. SEA,  
BALTIC, KIEL PACIFIC W. COAST S-N. AMERICA, PANAMA CAPE HORN W.  
INDIES, U.S. EASTERN, GULF OF MEXICO, GREAT LAKES, NEWFOU]  
and AREATYPE = [controlled port]  
and NOFACC > 1  
and GRT1GROU = 1  
and WEATH = No info  
and BUILT1 > 1968  
and BUILT1 <= 1977  
and OWNERS = [2 No info]  
then -> Yes (32, 0.813)



Rule #3 for Yes:

if ACCIDE0 = FOUNDERED

and GRT1 =< 16518

and NAREA = [ATLANTIC AUSTRALIA CANADIAN ALASKA, U.S.S.R.  
ARTIC & BERNING SEA, ICELAND MEDITERRANEAN, SUEZ N. SEA,  
BALTIC, KIEL PACIFIC W. COAST S-N. AMERICA,PANAMA CAPE HORN W.  
INDIES, U.S. EASTERN,GULF OF MEXICO, GREAT LAKES, NEWFOU]

and AREATYPE = [controlled port]

and NOFACC > 1

and GRT1GROU = 1

and WEATH = STORM

and GRUCLASS = [AB-GL-NK BV-CC-KR-NV-RI-PR-RS LR]

then -> Yes (31, 0.613)

Rule #4 for Yes:

if ACCIDE0 = FOUNDERED

and GRT1 =< 16518

and NAREA = [ATLANTIC AUSTRALIA CANADIAN ALASKA, U.S.S.R.  
ARTIC & BERNING SEA, ICELAND MEDITERRANEAN, SUEZ N. SEA,  
BALTIC, KIEL PACIFIC W. COAST S-N. AMERICA,PANAMA CAPE HORN W.  
INDIES, U.S. EASTERN,GULF OF MEXICO, GREAT LAKES, NEWFOU]

and AREATYPE = [controlled port]

and NOFACC > 1

and GRT1GROU = [ 2 4]

then -> Yes (14, 0.643)

Rule #5 for Yes:

if ACCIDE0 = FOUNDERED

and GRT1 =< 16518

and NAREA = [ATLANTIC AUSTRALIA CANADIAN ALASKA, U.S.S.R.  
ARTIC & BERNING SEA, ICELAND MEDITERRANEAN, SUEZ N. SEA,  
BALTIC, KIEL PACIFIC W. COAST S-N. AMERICA,PANAMA CAPE HORN W.  
INDIES, U.S. EASTERN,GULF OF MEXICO, GREAT LAKES, NEWFOU]

and AREATYPE = [controlled port]  
and NOFACC > 1  
and GRT1GROU = No info  
and WEATH = [CALM WEATHER No info POOR VISIBILITY  
SNOWSTORM]  
then -> Yes (47.792, 0.733)

Rule #6 for Yes:

if ACCIDE0 = FOUNDERED  
and GRT1 <= 16518  
and NAREA = [ATLANTIC AUSTRALIA CANADIAN ALASKA, U.S.S.R.  
ARTIC & BERNING SEA, ICELAND MEDITERRANEAN, SUEZ N. SEA,  
BALTIC, KIEL PACIFIC W. COAST S-N. AMERICA, PANAMA CAPE HORN W.  
INDIES, U.S. EASTERN, GULF OF MEXICO, GREAT LAKES, NEWFOU]  
and AREATYPE = open sea  
and TYPE = [BULK FERRY]  
and CLASS = [BV CC IR NK PR RN RS oth]  
then -> Yes (290.229, 0.637)

Rule #7 for Yes:

if ACCIDE0 = FOUNDERED  
and GRT1 <= 16518  
and NAREA = [ATLANTIC AUSTRALIA CANADIAN ALASKA, U.S.S.R.  
ARTIC & BERNING SEA, ICELAND MEDITERRANEAN, SUEZ N. SEA,  
BALTIC, KIEL PACIFIC W. COAST S-N. AMERICA, PANAMA CAPE HORN W.  
INDIES, U.S. EASTERN, GULF OF MEXICO, GREAT LAKES, NEWFOU]  
and AREATYPE = open sea  
and TYPE = [BULK FERRY]  
and CLASS = GL  
and OWNERS = 5++  
then -> Yes (20, 0.7)

Rule #8 for Yes:

if ACCIDE0 = FOUNDERED

and GRT1 =< 16518

and NAREA = [ATLANTIC AUSTRALIA CANADIAN ALASKA, U.S.S.R.  
ARTIC & BERNING SEA, ICELAND MEDITERRANEAN, SUEZ N. SEA,  
BALTIC, KIEL PACIFIC W. COAST S-N. AMERICA, PANAMA CAPE HORN W.  
INDIES, U.S. EASTERN, GULF OF MEXICO, GREAT LAKES, NEWFOU]

and AREATYPE = open sea

and TYPE = [BULK FERRY]

and CLASS = LR

and NAREA = [AUSTRALIA MEDITERRANEAN, SUEZ N. SEA, BALTIC,  
KIEL W. INDIES, U.S. EASTERN, GULF OF MEXICO, GREAT LAKES,  
NEWFOU]

then -> Yes (47, 0.638)

Rule #9 for Yes:

if ACCIDE0 = FOUNDERED

and GRT1 =< 16518

and NAREA = [GULF, BAY OF BENGAL INDIAN OCEAN & ANTARTIC  
JAPAN, S. CHINA & E. INDIES RED SEA E.& W. AFRICAN COAST]

and MONTH = [Apr Jan Jun Mar Sep]

and GRUCLASS = [AB-GL-NK BV-CC-KR-NV-RI-PR-RS LR]

and NAREA = INDIAN OCEAN & ANTARTIC

then -> Yes (4, 0.5)

Rule #10 for Yes:

if ACCIDE0 = FOUNDERED

and GRT1 =< 16518

and NAREA = [GULF, BAY OF BENGAL INDIAN OCEAN & ANTARTIC  
JAPAN, S. CHINA & E. INDIES RED SEA E.& W. AFRICAN COAST]

and MONTH = [Apr Jan Jun Mar Sep]

and GRUCLASS = [AB-GL-NK BV-CC-KR-NV-RI-PR-RS LR]

and NAREA = JAPAN, S. CHINA & E. INDIES

and OWNERS = [2 5++ No info]

then -> Yes (83.797, 0.618)

Rule #11 for Yes:

if ACCIDE0 = FOUNDERED

and GRT1 =< 16518

and NAREA = [GULF, BAY OF BENGAL INDIAN OCEAN & ANTARTIC  
JAPAN, S. CHINA & E. INDIES RED SEA E.& W. AFRICAN COAST]

and MONTH = [Apr Jan Jun Mar Sep]

and GRUCLASS = [AB-GL-NK BV-CC-KR-NV-RI-PR-RS LR]

and NAREA = JAPAN, S. CHINA & E. INDIES

and OWNERS = 3-4

and MONTH = [Jun Mar]

then -> Yes (27, 0.778)

Rule #12 for Yes:

if ACCIDE0 = FOUNDERED

and GRT1 =< 16518

and NAREA = [GULF, BAY OF BENGAL INDIAN OCEAN & ANTARTIC  
JAPAN, S. CHINA & E. INDIES RED SEA E.& W. AFRICAN COAST]

and MONTH = [Apr Jan Jun Mar Sep]

and GRUCLASS = [AB-GL-NK BV-CC-KR-NV-RI-PR-RS LR]

and NAREA = RED SEA E.& W. AFRICAN COAST

and GRUCLASS = LR

then -> Yes (13,796, 0.783)

Rule #13 for Yes:

if ACCIDE0 = FOUNDERED

and GRT1 =< 16518

and NAREA = [GULF, BAY OF BENGAL INDIAN OCEAN & ANTARTIC  
JAPAN, S. CHINA & E. INDIES RED SEA E.& W. AFRICAN COAST]

and MONTH = [Apr Jan Jun Mar Sep]

and GRUCLASS = OTHERS

and OWNERS = [1 No info]

and SEQACC =< 1

and DATYR =< 93

then -> Yes (21,524, 0.694)



Rule #14 for Yes:

if ACCIDE0 = FOUNDERED  
and GRT1 =< 16518  
and NAREA = [GULF, BAY OF BENGAL INDIAN OCEAN & ANTARTIC  
JAPAN, S. CHINA & E. INDIES RED SEA E.& W. AFRICAN COAST]  
and MONTH = [Apr Jan Jun Mar Sep]  
and GRUCLASS = OTHERS  
and OWNERS = [1 No info]  
and SEQACC =< 1  
and DATYR > 93  
and DATYR =< 95  
and MONTH = Jan  
then -> Yes (10.779, 0.926)

Rule #15 for Yes:

if ACCIDE0 = FOUNDERED  
and GRT1 =< 16518  
and NAREA = [GULF, BAY OF BENGAL INDIAN OCEAN & ANTARTIC  
JAPAN, S. CHINA & E. INDIES RED SEA E.& W. AFRICAN COAST]  
and MONTH = [Apr Jan Jun Mar Sep]  
and GRUCLASS = OTHERS  
and OWNERS = [1 No info]  
and SEQACC =< 1  
and DATYR > 95  
then -> Yes (102.827, 0.699)

Rule #16 for Yes:

if ACCIDE0 = FOUNDERED  
and GRT1 =< 16518  
and NAREA = [GULF, BAY OF BENGAL INDIAN OCEAN & ANTARTIC  
JAPAN, S. CHINA & E. INDIES RED SEA E.& W. AFRICAN COAST]  
and MONTH = [Apr Jan Jun Mar Sep]  
and GRUCLASS = OTHERS

and OWNERS == [1 No info]  
and SEQACC > 1  
then -> Yes (12, 0.833)

Rule #17 for Yes:

if ACCIDE0 == FOUNDERED  
and GRT1 =< 16518  
and NAREA == [GULF, BAY OF BENGAL INDIAN OCEAN & ANTARTIC  
JAPAN, S. CHINA & E. INDIES RED SEA E.& W. AFRICAN COAST]  
and MONTH == [Apr Jan Jun Mar Sep]  
and GRUCLASS == OTHERS  
and OWNERS == [2 5++]  
then -> Yes (10.593, 1.0)

Rule #18 for Yes:

if ACCIDE0 == FOUNDERED  
and GRT1 =< 16518  
and NAREA == [GULF, BAY OF BENGAL INDIAN OCEAN & ANTARTIC  
JAPAN, S. CHINA & E. INDIES RED SEA E.& W. AFRICAN COAST]  
and MONTH == [Aug Dec Feb Jul May Nov Oct]  
then -> Yes (730.393, 0.703)

Rule #19 for Yes:

if ACCIDE0 == FOUNDERED  
and GRT1 > 16518  
and GRT1GROU == [ 1 2 3 4 5 7]  
and BUILT1 =< 1973  
and CLASS == [CC RI RS oth]  
then -> Yes (25, 0.64)

Rule #20 for Yes:

if ACCIDE0 == FOUNDERED  
and GRT1 > 16518  
and GRT1 =< 22841

and GRT1GROU == [ 1 2 3 4 5 7]  
and BUILT1 <= 1973  
and CLASS == LR  
then -> Yes (12, 0.75)

Rule #21 for Yes:

if ACCIDE0 == FOUNDERED  
and GRT1 > 16518  
and GRT1GROU == [ 1 2 3 4 5 7]  
and BUILT1 <= 1973  
and CLASS == NK  
and NOFACC > 1  
then -> Yes (13, 0.692)

Rule #22 for Yes:

if ACCIDE0 == FOUNDERED  
and GRT1 > 16518  
and GRT1 <= 21716  
and GRT1GROU == [ 1 2 3 4 5 7]  
and BUILT1 > 1973  
and NAREA == [GULF, BAY OF BENGAL JAPAN, S. CHINA & E. INDIES  
MEDITERRANEAN, SUEZ PACIFIC RED SEA E.& W. AFRICAN COAST W.  
COAST S-N. AMERICA, PANAMA CAPE HORN]  
and AREATYPE == open sea  
then -> Yes (11, 0.818)

Rule #23 for Yes:

if ACCIDE0 == FOUNDERED  
and GRT1 > 16518  
and GRT1GROU == [ 6 No info]  
then -> Yes (130.86, 0.634)

Default : -> No

T<sub>No</sub> Ταξινομητικοί κανόνες για ατυχήματα που δεν οδηγούν στην ολική απώλεια

Rules for No:

Rule #1 for No:

if ACCIDE0 = [COLLISION(WITH SHIP) CONTACT(INC. STRIKING  
ICE) FIRE-EXPLOSION GROUNDED HULL/MCHY/EQUIP.DAMAGED  
MISSING]

then -> No (24706, 0.949)

Rule #2 for No:

if ACCIDE0 = FOUNDERED

and GRT1 <= 16518

and NAREA = [ATLANTIC AUSTRALIA CANADIAN ALASKA, U.S.S.R.  
ARTIC & BERNING SEA, ICELAND MEDITERRANEAN, SUEZ N. SEA,  
BALTIC, KIEL PACIFIC W. COAST S-N. AMERICA,PANAMA CAPE HORN W.  
INDIES, U.S. EASTERN,GULF OF MEXICO, GREAT LAKES, NEWFOU]

and AREATYPE = [controlled port]

and NOFACC <= 1

then -> No (314.129, 0.679)

Rule #3 for No:

if ACCIDE0 = FOUNDERED

and GRT1 <= 16518

and NAREA = [ATLANTIC AUSTRALIA CANADIAN ALASKA, U.S.S.R.  
ARTIC & BERNING SEA, ICELAND MEDITERRANEAN, SUEZ N. SEA,  
BALTIC, KIEL PACIFIC W. COAST S-N. AMERICA,PANAMA CAPE HORN W.  
INDIES, U.S. EASTERN,GULF OF MEXICO, GREAT LAKES, NEWFOU]

and AREATYPE = [controlled port]

and NOFACC > 1

and GRT1GROU = 1

and WEATH = No info

and BUILT1 <= 1977

and OWNERS = [1 3-4 5++]



then -> No (52, 0.75)

Rule #4 for No:

if ACCIDE0 = FOUNDERED

and GRT1 <= 16518

and NAREA = [ATLANTIC AUSTRALIA CANADIAN ALASKA, U.S.S.R.  
ARTIC & BERNING SEA, ICELAND MEDITERRANEAN, SUEZ N. SEA,  
BALTIC, KIEL PACIFIC W. COAST S-N. AMERICA, PANAMA CAPE HORN W.  
INDIES, U.S. EASTERN, GULF OF MEXICO, GREAT LAKES, NEWFOU]

and AREATYPE = [controlled port]

and NOFACC > 1

and GRT1GROU = 1

and WEATH = No info

and BUILT1 <= 1968

and OWNERS = [2 No info]

then -> No (15, 0.667)

Rule #5 for No:

if ACCIDE0 = FOUNDERED

and GRT1 <= 16518

and NAREA = [ATLANTIC AUSTRALIA CANADIAN ALASKA, U.S.S.R.  
ARTIC & BERNING SEA, ICELAND MEDITERRANEAN, SUEZ N. SEA,  
BALTIC, KIEL PACIFIC W. COAST S-N. AMERICA, PANAMA CAPE HORN W.  
INDIES, U.S. EASTERN, GULF OF MEXICO, GREAT LAKES, NEWFOU]

and AREATYPE = [controlled port]

and NOFACC > 1

and GRT1GROU = 1

and WEATH = No info

and BUILT1 > 1977

then -> No (37, 0.892)

Rule #6 for No:

if ACCIDE0 = FOUNDERED

and GRT1 <= 16518

and NAREA = [ATLANTIC AUSTRALIA CANADIAN ALASKA, U.S.S.R.  
ARTIC & BERNING SEA, ICELAND MEDITERRANEAN, SUEZ N. SEA,  
BALTIC, KIEL PACIFIC W. COAST S-N. AMERICA, PANAMA CAPE HORN W.  
INDIES, U.S. EASTERN, GULF OF MEXICO, GREAT LAKES, NEWFOU]

and AREATYPE = [controlled port]

and NOFACC > 1

and GRT1GROU = 1

and WEATH = [POOR VISIBILITY SNOWSTORM]

then -> No (14, 0.643)

Rule #7 for No:

if ACCIDE0 = FOUNDERED

and GRT1 =< 16518

and NAREA = [ATLANTIC AUSTRALIA CANADIAN ALASKA, U.S.S.R.  
ARTIC & BERNING SEA, ICELAND MEDITERRANEAN, SUEZ N. SEA,  
BALTIC, KIEL PACIFIC W. COAST S-N. AMERICA, PANAMA CAPE HORN W.  
INDIES, U.S. EASTERN, GULF OF MEXICO, GREAT LAKES, NEWFOU]

and AREATYPE = [controlled port]

and NOFACC > 1

and GRT1GROU = 1

and WEATH = STORM

and GRUCLASS = OTHERS

then -> No (8, 0.875)

Rule #8 for No:

if ACCIDE0 = FOUNDERED

and GRT1 =< 16518

and NAREA = [ATLANTIC AUSTRALIA CANADIAN ALASKA, U.S.S.R.  
ARTIC & BERNING SEA, ICELAND MEDITERRANEAN, SUEZ N. SEA,  
BALTIC, KIEL PACIFIC W. COAST S-N. AMERICA, PANAMA CAPE HORN W.  
INDIES, U.S. EASTERN, GULF OF MEXICO, GREAT LAKES, NEWFOU]

and AREATYPE = [controlled port]

and NOFACC > 1

and GRT1GROU = [3 5 6 7]

then -> No (2, 1.0)

Rule #9 for No:

if ACCIDE0 = FOUNDERED

and GRT1 <= 16518

and NAREA = [ATLANTIC AUSTRALIA CANADIAN ALASKA, U.S.S.R.  
ARTIC & BERNING SEA, ICELAND MEDITERRANEAN, SUEZ N. SEA,  
BALTIC, KIEL PACIFIC W. COAST S-N. AMERICA, PANAMA CAPE HORN W.  
INDIES, U.S. EASTERN, GULF OF MEXICO, GREAT LAKES, NEWFOU]

and AREATYPE = [controlled port]

and NOFACC > 1

and GRT1GROU = No info

and WEATH = [STORM TYPHOON]

then -> No (20.71, 0.692)

Rule #10 for No:

if ACCIDE0 = FOUNDERED

and GRT1 <= 16518

and NAREA = [ATLANTIC AUSTRALIA CANADIAN ALASKA, U.S.S.R.  
ARTIC & BERNING SEA, ICELAND MEDITERRANEAN, SUEZ N. SEA,  
BALTIC, KIEL PACIFIC W. COAST S-N. AMERICA, PANAMA CAPE HORN W.  
INDIES, U.S. EASTERN, GULF OF MEXICO, GREAT LAKES, NEWFOU]

and AREATYPE = open sea

and TYPE = [BULK FERRY]

and CLASS = [AB CC KI KR NV RI]

then -> No (53, 0.623)

Rule #11 for No:

if ACCIDE0 = FOUNDERED

and GRT1 <= 16518

and NAREA = [ATLANTIC AUSTRALIA CANADIAN ALASKA, U.S.S.R.  
ARTIC & BERNING SEA, ICELAND MEDITERRANEAN, SUEZ N. SEA,  
BALTIC, KIEL PACIFIC W. COAST S-N. AMERICA, PANAMA CAPE HORN W.  
INDIES, U.S. EASTERN, GULF OF MEXICO, GREAT LAKES, NEWFOU]

and AREATYPE = open sea  
and TYPE = [BULK FERRY]  
and CLASS = GL  
and OWNERS = [1 2 3-4 No info]  
then -> No (26, 0.731)

Rule #12 for No:

if ACCIDE0 = FOUNDERED  
and GRT1 <= 16518  
and NAREA = [ATLANTIC AUSTRALIA CANADIAN ALASKA, U.S.S.R.  
ARTIC & BERNING SEA, ICELAND MEDITERRANEAN, SUEZ N. SEA,  
BALTIC, KIEL PACIFIC W. COAST S-N. AMERICA,PANAMA CAPE HORN W.  
INDIES, U.S. EASTERN,GULF OF MEXICO, GREAT LAKES, NEWFOU]  
and AREATYPE = open sea  
and TYPE = [BULK FERRY]  
and CLASS = LR  
and NAREA = [ATLANTIC CANADIAN ALASKA, U.S.S.R. ARTIC &  
BERNING SEA, ICELAND PACIFIC W. COAST S-N. AMERICA,PANAMA  
CAPE HORN]  
then -> No (15, 0.8)

Rule #13 for No:

if ACCIDE0 = FOUNDERED  
and GRT1 <= 16518  
and NAREA = [ATLANTIC AUSTRALIA CANADIAN ALASKA, U.S.S.R.  
ARTIC & BERNING SEA, ICELAND MEDITERRANEAN, SUEZ N. SEA,  
BALTIC, KIEL PACIFIC W. COAST S-N. AMERICA,PANAMA CAPE HORN W.  
INDIES, U.S. EASTERN,GULF OF MEXICO, GREAT LAKES, NEWFOU]  
and AREATYPE = open sea  
and TYPE = [CONTAINER TANKER]  
then -> No (68.454, 0.656)

Rule #14 for No:

if ACCIDE0 = FOUNDERED



and GRT1 =< 16518

and NAREA = [ATLANTIC AUSTRALIA CANADIAN ALASKA, U.S.S.R.  
ARTIC & BERNING SEA, ICELAND MEDITERRANEAN, SUEZ N. SEA,  
BALTIC, KIEL PACIFIC W. COAST S-N. AMERICA, PANAMA CAPE HORN W.  
INDIES, U.S. EASTERN, GULF OF MEXICO, GREAT LAKES, NEWFOU]

then -> No (0.0)

Rule #15 for No:

if ACCIDE0 = FOUNDERED

and GRT1 =< 16518

and NAREA = [GULF, BAY OF BENGAL INDIAN OCEAN & ANTARTIC  
JAPAN, S. CHINA & E. INDIES RED SEA E.& W. AFRICAN COAST]

and MONTH = [Apr Jan Jun Mar Sep]

and GRUCLASS = [AB-GL-NK BV-CC-KR-NV-RI-PR-RS LR]

and NAREA = GULF, BAY OF BENGAL

then -> No (63.797, 0.734)

Rule #16 for No:

if ACCIDE0 = FOUNDERED

and GRT1 =< 16518

and NAREA = [GULF, BAY OF BENGAL INDIAN OCEAN & ANTARTIC  
JAPAN, S. CHINA & E. INDIES RED SEA E.& W. AFRICAN COAST]

and MONTH = [Apr Jan Jun Mar Sep]

and GRUCLASS = [AB-GL-NK BV-CC-KR-NV-RI-RN-RS LR]

and NAREA = JAPAN, S. CHINA & E. INDIES

and OWNERS = 1

then -> No (47, 0.66)

Rule #17 for No:

if ACCIDE0 = FOUNDERED

and GRT1 =< 16518

and NAREA = [GULF, BAY OF BENGAL INDIAN OCEAN & ANTARTIC  
JAPAN, S. CHINA & E. INDIES RED SEA E.& W. AFRICAN COAST]

and MONTH = [Apr Jan Jun Mar Sep]

and GRUCLASS = [AB-GL-NK BV-CC-KR-NV-RI-RN-RS LR]

and NAREA = JAPAN, S. CHINA & E. INDIES

and OWNERS = 3-4

and MONTH = [Apr Jan Sep]

then -> No (33, 0.727)

Rule #18 for No:

if ACCIDE0 = FOUNDERED

and GRT1 <= 16518

and NAREA = [GULF, BAY OF BENGAL INDIAN OCEAN & ANTARTIC  
JAPAN, S. CHINA & E. INDIES RED SEA E.& W. AFRICAN COAST]

and MONTH = [Apr Jan Jun Mar Sep]

and GRUCLASS = [AB-GL-NK BV-CC-KR-NV-RI-RN-RS LR]

and NAREA = RED SEA E.& W. AFRICAN COAST

and GRUCLASS = [AB-GL-NK BV-CC-KR-NV-RI-RN-RS]

then -> No (30, 0.633)

Rule #19 for No:

if ACCIDE0 = FOUNDERED

and GRT1 <= 16518

and NAREA = [GULF, BAY OF BENGAL INDIAN OCEAN & ANTARTIC  
JAPAN, S. CHINA & E. INDIES RED SEA E.& W. AFRICAN COAST]

and MONTH = [Apr Jan Jun Mar Sep]

and GRUCLASS = OTHERS

and OWNERS = [1 No info]

and SEQACC <= 1

and DATYR > 93

and DATYR <= 95

and MONTH = [Apr Jun Mar Sep]

then -> No (36.32, 0.83)

Rule #20 for No:

if ACCIDE0 = FOUNDERED

and GRT1 <= 16518

and NAREA = [GULF, BAY OF BENGAL INDIAN OCEAN & ANTARTIC  
JAPAN, S. CHINA & E. INDIES RED SEA E.& W. AFRICAN COAST]

and MONTH = [Apr Jan Jun Mar Sep]

and GRUCLASS = OTHERS

and OWNERS = 3-4

then -> No (1, 1.0)

Rule #21 for No:

if ACCIDE0 = FOUNDERED

and GRT1 > 16518

and GRT1GROU = [ 1 2 3 4 5 7]

and BUILT1 =< 1973

and CLASS = [AB BV GL CC IR KI KR NV PR RN ]

then -> No (39, 0.769)

Rule #22 for No:

if ACCIDE0 = FOUNDERED

and GRT1 > 22841

and GRT1GROU = [ 1 2 3 4 5 7]

and BUILT1 =< 1973

and CLASS = LR

then -> No (10, 0.9)

Rule #23 for No:

if ACCIDE0 = FOUNDERED

and GRT1 > 16518

and GRT1GROU = [ 1 2 3 4 5 7]

and BUILT1 =< 1973

and CLASS = NK

and NOFACC =< 1

then -> No (38, 0.553)

Rule #24 for No:

if ACCIDE0 = FOUNDERED

and GRT1 > 16518

and GRT1GROU = [ 1 2 3 4 5 7]

and BUILT1 > 1973

and NAREA = [ATLANTIC AUSTRALIA CANADIAN ALASKA, U.S.S.R.  
ARTIC & BERNING SEA, ICELAND INDIAN OCEAN & ANTARTIC N. SEA,  
BALTIC, KIEL W. INDIES, U.S. EASTERN, GULF OF MEXICO, GREAT  
LAKES, NEWFOU]

then -> No (154, 0.981)

Rule #25 for No:

if ACCIDE0 = FOUNDERED

and GRT1 > 16518

and GRT1GROU = [ 1 2 3 4 5 7]

and BUILT1 > 1973

and NAREA = [GULF, BAY OF BENGAL JAPAN, S. CHINA & E. INDIES  
MEDITERRANEAN, SUEZ PACIFIC RED SEA E.& W. AFRICAN COAST W.  
COAST S-N. AMERICA, PANAMA CAPE HORN]

and AREATYPE = [controlled port]

then -> No (136, 0.875)

Rule #26 for No:

if ACCIDE0 = FOUNDERED

and GRT1 > 21716

and GRT1GROU = [ 1 2 3 4 5 7]

and BUILT1 > 1973

and NAREA = [GULF, BAY OF BENGAL JAPAN, S. CHINA & E. INDIES  
MEDITERRANEAN, SUEZ PACIFIC RED SEA E.& W. AFRICAN COAST W.  
COAST S-N. AMERICA, PANAMA CAPE HORN]

and AREATYPE = open sea

then -> No (33, 0.727)



Π. Παράγοντες που ο συνδυασμός τους οδηγεί σε ολική απώλεια

Συνδυασμός παραγόντων	percent
GRUCLASS = OTHERS & NAREA = JAPAN, S. CHINA & E. INDIES & TYPE = FERRY	0,85
GRUCLASS = OTHERS & OWNERS = No info & NAREA = JAPAN, S. CHINA & E. INDIES & TYPE = FERRY	0,84
TYPE = BULK & BUILT74 < 1975 & OWNERS = No info & NAREA = GULF, BAY OF BENGAL	0,84
OWNERS = No info & NAREA = JAPAN, S. CHINA & E. INDIES & TYPE = FERRY	0,83
BUILT74 < 1975 & GRUCLASS = OTHERS & OWNERS = No info & NAREA = GULF, BAY OF BENGAL	0,83
BUILT74 < 1975 & OWNERS = No info & NAREA = GULF, BAY OF BENGAL	0,82
BUILT74 < 1975 & NAREA = JAPAN, S. CHINA & E. INDIES & TYPE = FERRY	0,82
TYPE = BULK & GRUCLASS = OTHERS & OWNERS = No info & NAREA = GULF, BAY OF BENGAL	0,81
GRUCLASS = OTHERS & OWNERS = No info & NAREA = GULF, BAY OF BENGAL	0,81
TYPE = BULK & OWNERS = No info & NAREA = GULF, BAY OF BENGAL	0,81
OWNERS = No info & NAREA = GULF, BAY OF BENGAL	0,80
TYPE = BULK & BUILT74 < 1975 & GRUCLASS = OTHERS & NAREA = GULF, BAY OF BENGAL	0,80
TYPE = BULK & BUILT74 < 1975 & NAREA = JAPAN, S. CHINA & E. INDIES & OWNERS = 2	0,80
BUILT74 < 1975 & GRUCLASS = OTHERS & NAREA = JAPAN, S. CHINA & E. INDIES	0,80
BUILT74 < 1975 & GRUCLASS = OTHERS & NAREA = GULF, BAY OF BENGAL	0,80

GRUCLASS = OTHERS & NAREA = GULF, BAY OF BENGAL	0,79
BUILT74 < 1975 & GRUCLASS = OTHERS & OWNERS = No info & NAREA = JAPAN, S. CHINA & E. INDIES	0,79
BUILT74 < 1975 & OWNERS = No info & NAREA = JAPAN, S. CHINA & E. INDIES	0,79
TYPE = BULK & GRUCLASS = OTHERS & NAREA = GULF, BAY OF BENGAL	0,78
TYPE = BULK & BUILT74 < 1975 & GRUCLASS = OTHERS & NAREA = JAPAN, S. CHINA & E. INDIES	0,78
BUILT74 < 1975 & NAREA = JAPAN, S. CHINA & E. INDIES & OWNERS = 2	0,77
GRUCLASS = OTHERS & NAREA = JAPAN, S. CHINA & E. INDIES	0,77
GRUCLASS = OTHERS & NAREA = JAPAN, S. CHINA & E. INDIES & OWNERS = 1	0,77
GRUCLASS = OTHERS & NAREA = RED SEA E.& W. AFRICAN COAST	0,77
BUILT74 < 1975 & GRUCLASS = OTHERS & NAREA = RED SEA E.& W. AFRICAN COAST	0,77
TYPE = BULK & GRUCLASS = OTHERS & NAREA = JAPAN, S. CHINA & E. INDIES	0,76
GRUCLASS = OTHERS & OWNERS = No info & NAREA = JAPAN, S. CHINA & E. INDIES	0,76
BUILT74 < 1975 & NAREA = JAPAN, S. CHINA & E. INDIES & TYPE = TANKER	0,76
TYPE = BULK & BUILT74 < 1975 & OWNERS = No info & NAREA = JAPAN, S. CHINA & E. INDIES	0,76
TYPE = BULK & BUILT74 < 1975 & NAREA = JAPAN, S. CHINA & E. INDIES & OWNERS = 5++	0,75
NAREA = JAPAN, S. CHINA & E. INDIES & TYPE = FERRY	0,75
GRUCLASS = OTHERS & OWNERS = No info & NAREA = RED SEA E.& W. AFRICAN COAST	0,75
TYPE = BULK & GRUCLASS = OTHERS & OWNERS = No info & NAREA = JAPAN, S. CHINA & E. INDIES	0,75

GRUCLASS = BV-CC-KR-NV-RI-PR-RS & NAREA = JAPAN, S. CHINA & E. INDIES & OWNERS = 5++	0,75
TYPE = BULK & BUILT74 < 1975 & GRUCLASS = BV-CC-KR-NV- RI-PR-RS & NAREA = GULF, BAY OF BENGAL	0,75
OWNERS = No info & NAREA = RED SEA E.& W. AFRICAN COAST	0,74
GRUCLASS = OTHERS & OWNERS = No info & TYPE = FERRY	0,74
BUILT74 < 1975 & NAREA = JAPAN, S. CHINA & E. INDIES	0,74
OWNERS = No info & NAREA = JAPAN, S. CHINA & E. INDIES	0,74
BUILT74 < 1975 & GRUCLASS = OTHERS & OWNERS = No info & TYPE = FERRY	0,74
TYPE = BULK & GRUCLASS = OTHERS & OWNERS = 2	0,74
GRUCLASS = OTHERS & TYPE = FERRY	0,73
GRUCLASS = OTHERS & OWNERS = No info & NAREA = CANADIAN ALASKA, U.S.S.R. ARTIC & BERNING SEA, ICELAND	0,73
TYPE = BULK & BUILT74 < 1975 & NAREA = JAPAN, S. CHINA & E. INDIES	0,73
GRUCLASS = OTHERS & NAREA = CANADIAN ALASKA, U.S.S.R. ARTIC & BERNING SEA, ICELAND	0,73
TYPE = BULK & GRUCLASS = OTHERS & NAREA = RED SEA E.& W. AFRICAN COAST	0,73
TYPE = BULK & OWNERS = No info & NAREA = JAPAN, S. CHINA & E. INDIES	0,73
BUILT74 < 1975 & GRUCLASS = BV-CC-KR-NV-RI-PR-RS & NAREA = JAPAN, S. CHINA & E. INDIES	0,73
BUILT74 < 1975 & GRUCLASS = BV-CC-KR-NV-RI-PR-RS & NAREA = GULF, BAY OF BENGAL	0,73
GRUCLASS = BV-CC-KR-NV-RI-PR-RS & NAREA = JAPAN, S. CHINA & E. INDIES & OWNERS = 2	0,73
TYPE = BULK & GRUCLASS = BV-CC-KR-NV-RI-PR-RS & NAREA = JAPAN, S. CHINA & E. INDIES & OWNERS = 2	0,73
OWNERS = No info & NAREA = CANADIAN ALASKA, U.S.S.R. ARTIC & BERNING SEA, ICELAND	0,73



BUILT74 < 1975 & OWNERS = No info & TYPE = FERRY	0,73
OWNERS = No info & TYPE = FERRY	0,73
BUILT74 < 1975 & GRUCLASS = OTHERS & TYPE = FERRY	0,73
GRUCLASS = OTHERS & OWNERS = No info & NAREA = JAPAN, S. CHINA & E. INDIES & TYPE = TANKER	0,73
TYPE = BULK & BUILT74 < 1975 & NAREA = JAPAN, S. CHINA & E. INDIES & GRUCLASS = AB-GL-NK	0,72
OWNERS = 5++ & NAREA = RED SEA E.& W. AFRICAN COAST	0,72
TYPE = BULK & OWNERS = No info & NAREA = RED SEA E.& W. AFRICAN COAST	0,72
TYPE = BULK & BUILT74 < 1975 & NAREA = JAPAN, S. CHINA & E. INDIES & OWNERS = 3-4	0,72
TYPE = FERRY & NAREA = RED SEA E.& W. AFRICAN COAST	0,72
BUILT74 < 1975 & OWNERS = 5++ & NAREA = GULF, BAY OF BENGAL	0,72
TYPE = BULK & BUILT74 < 1975 & GRUCLASS = OTHERS & OWNERS = No info	0,72
BUILT74 < 1975 & NAREA = JAPAN, S. CHINA & E. INDIES & OWNERS = 5++	0,72
TYPE = BULK & GRUCLASS = OTHERS & OWNERS = No info & NAREA = RED SEA E.& W. AFRICAN COAST	0,72
TYPE = BULK & BUILT74 < 1975 & OWNERS = No info	0,72
BUILT74 < 1975 & NAREA = JAPAN, S. CHINA & E. INDIES & OWNERS = 3-4	0,72
GRUCLASS = OTHERS & NAREA = JAPAN, S. CHINA & E. INDIES & TYPE = TANKER	0,72
GRUCLASS = OTHERS & OWNERS = 2	0,71
GRUCLASS = OTHERS & OWNERS = 1 & TYPE = FERRY	0,71
BUILT74 < 1975 & GRUCLASS = OTHERS & OWNERS = No info	0,71
TYPE = BULK & BUILT74 < 1975 & GRUCLASS = BV-CC-KR-NV- RI-PR-RS & NAREA = JAPAN, S. CHINA & E. INDIES	0,71
BUILT74 < 1975 & OWNERS = No info	0,71
TYPE = BULK & BUILT74 < 1975 & NAREA = GULF, BAY OF	0,71



BENGAL	
TYPE = BULK & BUILT74 < 1975 & GRUCLASS = OTHERS	0,71
TYPE = BULK & NAREA = JAPAN, S. CHINA & E. INDIES & OWNERS = 5++	0,70
NAREA = JAPAN, S. CHINA & E. INDIES & OWNERS = 3-4 & GRUCLASS = LR	0,70
TYPE = BULK & BUILT74 < 1975 & NAREA = JAPAN, S. CHINA & E. INDIES & GRUCLASS = LR	0,70
BUILT74 < 1975 & NAREA = GULF, BAY OF BENGAL	0,70
BUILT74 < 1975 & GRUCLASS = OTHERS	0,70
TYPE = BULK & BUILT74 < 1975 & NAREA = RED SEA E.& W. AFRICAN COAST	0,70
BUILT74 < 1975 & NAREA = RED SEA E.& W. AFRICAN COAST	0,70
BUILT74 < 1975 & NAREA = JAPAN, S. CHINA & E. INDIES & GRUCLASS = AB-GL-NK	0,70
TYPE = BULK & NAREA = JAPAN, S. CHINA & E. INDIES & OWNERS = 2	0,69
OWNERS = No info & NAREA = JAPAN, S. CHINA & E. INDIES & TYPE = TANKER	0,69
BUILT74 < 1975 & GRUCLASS = OTHERS & OWNERS = No info & TYPE = TANKER	0,69
TYPE = BULK & OWNERS = 5++ & NAREA = GULF, BAY OF BENGAL	0,69
BUILT74 < 1975 & NAREA = JAPAN, S. CHINA & E. INDIES & GRUCLASS = LR	0,69
BUILT74 < 1975 & OWNERS = No info & TYPE = TANKER	0,69
BUILT74 < 1975 & GRUCLASS = BV-CC-KR-NV-RI-PR-RS & NAREA = RED SEA E.& W. AFRICAN COAST	0,69
OWNERS = 5++ & NAREA = GULF, BAY OF BENGAL	0,68
NAREA = JAPAN, S. CHINA & E. INDIES & OWNERS = 2	0,68
TYPE = BULK & BUILT74 < 1975 & GRUCLASS = BV-CC-KR-NV- RI-PR-RS & OWNERS = 5++	0,68
BUILT74 < 1975 & NAREA = JAPAN, S. CHINA & E. INDIES &	0,67

OWNERS = 1	
BUILT74 < 1975 & GRUCLASS = OTHERS & OWNERS = No info & NAREA = MEDITERRANEAN, SUEZ	0,67
GRUCLASS = OTHERS & OWNERS = No info	0,67
BUILT74 < 1975 & GRUCLASS = OTHERS & TYPE = TANKER	0,67
TYPE = BULK & BUILT74 < 1975 & OWNERS = No info & NAREA = N. SEA, BALTIC, KIEL	0,67
TYPE = BULK & GRUCLASS = OTHERS & OWNERS = No info	0,67
TYPE = BULK & GRUCLASS = BV-CC-KR-NV-RI-PR-RS & NAREA = GULF, BAY OF BENGAL	0,67
BUILT74 < 1975 & OWNERS = No info & NAREA = MEDITERRANEAN, SUEZ	0,67
GRUCLASS = OTHERS	0,66
TYPE = BULK & BUILT74 < 1975 & OWNERS = No info & NAREA = MEDITERRANEAN, SUEZ	0,66
GRUCLASS = OTHERS & OWNERS = No info & TYPE = TANKER	0,66
NAREA = JAPAN, S. CHINA & E. INDIES & OWNERS = 5++	0,66
TYPE = BULK & BUILT74 < 1975 & GRUCLASS = OTHERS & NAREA = MEDITERRANEAN, SUEZ	0,66
OWNERS = No info	0,66
TYPE = BULK & NAREA = GULF, BAY OF BENGAL	0,66
GRUCLASS = OTHERS & TYPE = TANKER	0,66
TYPE = BULK & GRUCLASS = OTHERS	0,65
BUILT74 < 1975 & GRUCLASS = OTHERS & NAREA = MEDITERRANEAN, SUEZ	0,65
TYPE = BULK & OWNERS = No info	0,65
NAREA = GULF, BAY OF BENGAL	0,65
TYPE = BULK & NAREA = JAPAN, S. CHINA & E. INDIES	0,65
TYPE = BULK & GRUCLASS = BV-CC-KR-NV-RI-PR-RS & OWNERS = 5++	0,65
NAREA = JAPAN, S. CHINA & E. INDIES & GRUCLASS = AB-GL- NK & OWNERS = 3-4	0,65
TYPE = BULK & GRUCLASS = OTHERS & OWNERS = No info &	0,65

NAREA = MEDITERRANEAN, SUEZ	
GRUCLASS = OTHERS & OWNERS = No info & NAREA = N. SEA, BALTIC, KIEL & TYPE = FERRY	0,65
NAREA = JAPAN, S. CHINA & E. INDIES & OWNERS = 3-4	0,64
TYPE = BULK & BUILT74 < 1975 & OWNERS = 3-4 & NAREA = GULF, BAY OF BENGAL	0,64
BUILT74 < 1975 & NAREA = W. COAST S-N. AMERICA, PANAMA CAPE HORN	0,64
BUILT74 < 1975 & GRUCLASS = BV-CC-KR-NV-RI-PR-RS & OWNERS = 5++	0,64
GRUCLASS = OTHERS & TYPE = FERRY & NAREA = MEDITERRANEAN, SUEZ	0,64
TYPE = BULK & BUILT74 < 1975 & GRUCLASS = AB-GL-NK & OWNERS = 5++	0,64
OWNERS = No info & TYPE = TANKER	0,64
TYPE = BULK & BUILT74 < 1975 & OWNERS = 5++	0,64
TYPE = BULK & GRUCLASS = BV-CC-KR-NV-RI-PR-RS & NAREA = JAPAN, S. CHINA & E. INDIES	0,64
NAREA = JAPAN, S. CHINA & E. INDIES	0,64
TYPE = BULK & OWNERS = No info & NAREA = MEDITERRANEAN, SUEZ	0,63
GRUCLASS = OTHERS & OWNERS = No info & NAREA = ATLANTIC	0,63
GRUCLASS = OTHERS & OWNERS = No info & NAREA = MEDITERRANEAN, SUEZ	0,63
NAREA = RED SEA E. & W. AFRICAN COAST	0,63
TYPE = BULK & BUILT74 < 1975 & GRUCLASS = OTHERS & NAREA = N. SEA, BALTIC, KIEL	0,63
GRUCLASS = BV-CC-KR-NV-RI-PR-RS & NAREA = GULF, BAY OF BENGAL	0,63
TYPE = BULK & NAREA = JAPAN, S. CHINA & E. INDIES & OWNERS = 3-4	0,63
BUILT74 < 1975 & OWNERS = No info & NAREA = N. SEA, BALTIC,	0,63



KIEL

GRUCLASS = OTHERS & NAREA = N. SEA, BALTIC, KIEL & TYPE = FERRY	0,63
GRUCLASS = BV-CC-KR-NV-RI-PR-RS & NAREA = JAPAN, S. CHINA & E. INDIES	0,62
OWNERS = No info & NAREA = N. SEA, BALTIC, KIEL & TYPE = FERRY	0,62
OWNERS = No info & NAREA = MEDITERRANEAN, SUEZ	0,62
TYPE = BULK & NAREA = RED SEA E.& W. AFRICAN COAST	0,62
TYPE = TANKER & NAREA = RED SEA E.& W. AFRICAN COAST	0,62
TYPE = BULK & BUILT74 < 1975 & GRUCLASS = OTHERS & NAREA = W. INDIES, U.S. EASTERN, GULF OF MEXICO, GREAT LAKES, NEWFOU	0,62
BUILT74 < 1975 & GRUCLASS = OTHERS & OWNERS = No info & NAREA = N. SEA, BALTIC, KIEL	0,62
TYPE = BULK & GRUCLASS = OTHERS & NAREA = MEDITERRANEAN, SUEZ	0,62
BUILT74 < 1975 & OWNERS = 3-4 & NAREA = GULF, BAY OF BENGAL	0,62
OWNERS = No info & NAREA = ATLANTIC	0,61
BUILT74 < 1975 & GRUCLASS = BV-CC-KR-NV-RI-PR-RS & OWNERS = 2	0,61
GRUCLASS = OTHERS & NAREA = MEDITERRANEAN, SUEZ	0,61
BUILT74 < 1975 & GRUCLASS = OTHERS & OWNERS = No info & NAREA = W. INDIES, U.S. EASTERN, GULF OF MEXICO, GREAT LAKES, NEWFOU	0,61
BUILT74 < 1975 & GRUCLASS = BV-CC-KR-NV-RI-PR-RS & TYPE = TANKER	0,61
GRUCLASS = BV-CC-KR-NV-RI-PR-RS & NAREA = JAPAN, S. CHINA & E. INDIES & TYPE = TANKER	0,61
TYPE = BULK & BUILT74 < 1975	0,61
GRUCLASS = BV-CC-KR-NV-RI-PR-RS & NAREA = RED SEA E.& W. AFRICAN COAST	0,61



BUILT74 < 1975 & GRUCLASS = OTHERS & NAREA = W. INDIES, U.S. EASTERN,GULF OF MEXICO, GREAT LAKES, NEWFOU	0,61
TYPE = TANKER & NAREA = GULF, BAY OF BENGAL	0,60
GRUCLASS = BV-CC-KR-NV-RI-PR-RS & OWNERS = 5++	0,60
TYPE = BULK & BUILT74 < 1975 & OWNERS = 2	0,60
TYPE = BULK & NAREA = JAPAN, S. CHINA & E. INDIES & GRUCLASS = AB-GL-NK & OWNERS = 3-4	0,60
BUILT74 < 1975 & OWNERS = No info & NAREA = W. INDIES, U.S. EASTERN,GULF OF MEXICO, GREAT LAKES, NEWFOU	0,60
BUILT74 < 1975 & OWNERS = 5++	0,60
GRUCLASS = OTHERS & NAREA = ATLANTIC	0,60
TYPE = BULK & OWNERS = 5++	0,60
GRUCLASS = OTHERS & OWNERS = 1	0,59
BUILT74 < 1975 & OWNERS = 3-4 & NAREA = RED SEA E.& W. AFRICAN COAST	0,59
TYPE = BULK & BUILT74 < 1975 & GRUCLASS = BV-CC-KR-NV- RI-PR-RS & OWNERS = 2	0,59
BUILT74 < 1975	0,59
TYPE = BULK & GRUCLASS = OTHERS & OWNERS = No info & NAREA = N. SEA, BALTIC, KIEL	0,59
BUILT74 < 1975 & GRUCLASS = OTHERS & OWNERS = 1	0,59
BUILT74 < 1975 & GRUCLASS = AB-GL-NK & OWNERS = 5++	0,59
TYPE = BULK & BUILT74 < 1975 & GRUCLASS = BV-CC-KR-NV- RI-PR-RS	0,59
TYPE = BULK & NAREA = W. INDIES, U.S. EASTERN,GULF OF MEXICO, GREAT LAKES, NEWFOU & OWNERS = 5++	0,59
BUILT74 < 1975 & OWNERS = 1 & TYPE = TANKER	0,59
GRUCLASS = BV-CC-KR-NV-RI-PR-RS & NAREA = JAPAN, S. CHINA & E. INDIES & OWNERS = 3-4	0,59
TYPE = BULK & BUILT74 < 1975 & GRUCLASS = BV-CC-KR-NV- RI-PR-RS & NAREA = MEDITERRANEAN, SUEZ	0,59
BUILT74 < 1975 & GRUCLASS = OTHERS & NAREA = N. SEA, BALTIC, KIEL	0,59

NAREA = JAPAN, S. CHINA & E. INDIES & GRUCLASS = AB-GL- NK & OWNERS = 2	0,59
OWNERS = 1 & NAREA = RED SEA E.& W. AFRICAN COAST	0,58
BUILT74 < 1975 & TYPE = TANKER	0,58
BUILT74 < 1975 & OWNERS = 2 & TYPE = FERRY	0,58
BUILT74 < 1975 & OWNERS = 2	0,58
TYPE = BULK & BUILT74 < 1975 & GRUCLASS = OTHERS & OWNERS = 1	0,58
BUILT74 < 1975 & GRUCLASS = BV-CC-KR-NV-RI-PR-RS	0,58
BUILT74 < 1975 & NAREA = CANADIAN ALASKA, U.S.S.R. ARTIC & BERNING SEA, ICELAND	0,58
GRUCLASS = BV-CC-KR-NV-RI-PR-RS & OWNERS = 2 & NAREA = MEDITERRANEAN, SUEZ	0,58
OWNERS = 2 & NAREA = GULF, BAY OF BENGAL	0,58
TYPE = BULK & OWNERS = No info & NAREA = N. SEA, BALTIC, KIEL	0,57
BUILT74 < 1975 & TYPE = FERRY	0,57
TYPE = BULK & GRUCLASS = OTHERS & OWNERS = 1	0,57
TYPE = BULK & BUILT74 < 1975 & NAREA = MEDITERRANEAN, SUEZ & OWNERS = 5++	0,57
TYPE = BULK & BUILT74 < 1975 & GRUCLASS = AB-GL-NK	0,57
BUILT74 < 1975 & NAREA = ATLANTIC	0,57
TYPE = BULK & GRUCLASS = BV-CC-KR-NV-RI-PR-RS & NAREA = RED SEA E.& W. AFRICAN COAST	0,57
NAREA = JAPAN, S. CHINA & E. INDIES & TYPE = TANKER	0,57
TYPE = BULK & BUILT74 < 1975 & NAREA = MEDITERRANEAN, SUEZ	0,57
TYPE = BULK & NAREA = JAPAN, S. CHINA & E. INDIES & GRUCLASS = LR	0,57
TYPE = BULK & BUILT74 < 1975 & GRUCLASS = AB-GL-NK & OWNERS = 2	0,56
GRUCLASS = OTHERS & OWNERS = No info & NAREA = N. SEA, BALTIC, KIEL	0,56

TYPE = BULK & OWNERS = 3-4 & NAREA = GULF, BAY OF BENGAL	0,56
BUILT74 < 1975 & GRUCLASS = BV-CC-KR-NV-RI-PR-RS & OWNERS = 3-4 & TYPE = TANKER	0,56
TYPE = BULK & NAREA = JAPAN, S. CHINA & E. INDIES & GRUCLASS = AB-GL-NK	0,56
TYPE = BULK & BUILT74 < 1975 & OWNERS = 3-4 & GRUCLASS = LR	0,56
TYPE = FERRY	0,55
TYPE = BULK & GRUCLASS = AB-GL-NK & NAREA = GULF, BAY OF BENGAL	0,55
TYPE = BULK & GRUCLASS = AB-GL-NK & OWNERS = 5++	0,55
OWNERS = 5++	0,55
TYPE = BULK & BUILT74 < 1975 & GRUCLASS = BV-CC-KR-NV-RI-PR-RS & OWNERS = 3-4	0,55
TYPE = BULK & GRUCLASS = OTHERS & NAREA = N. SEA, BALTIC, KIEL	0,55
TYPE = BULK & BUILT74 < 1975 & OWNERS = 3-4	0,55
TYPE = BULK & GRUCLASS = LR & OWNERS = 5++	0,55
OWNERS = 3-4 & NAREA = RED SEA E.& W. AFRICAN COAST	0,55
OWNERS = No info & NAREA = N. SEA, BALTIC, KIEL	0,55
TYPE = BULK & BUILT74 < 1975 & NAREA = ATLANTIC	0,55
NAREA = CANADIAN ALASKA, U.S.S.R. ARTIC & BERNING SEA, ICELAND	0,54
GRUCLASS = LR & NAREA = RED SEA E.& W. AFRICAN COAST	0,54
TYPE = BULK & BUILT74 < 1975 & GRUCLASS = LR & OWNERS = 5++	0,54
TYPE = BULK & BUILT74 < 1975 & NAREA = N. SEA, BALTIC, KIEL & OWNERS = 5++	0,54
TYPE = BULK & GRUCLASS = OTHERS & OWNERS = No info & NAREA = W. INDIES, U.S. EASTERN, GULF OF MEXICO, GREAT LAKES, NEWFOU	0,54
OWNERS = 3-4 & NAREA = GULF, BAY OF BENGAL	0,54



NAREA = W. INDIES, U.S. EASTERN, GULF OF MEXICO, GREAT LAKES, NEWFOU & OWNERS = 5++	0,54
BUILT74 < 1975 & OWNERS = 3-4 & GRUCLASS = LR	0,53
TYPE = BULK & BUILT74 < 1975 & GRUCLASS = AB-GL-NK & OWNERS = 3-4	0,53
BUILT74 < 1975 & TYPE = TANKER & OWNERS = 2	0,53
BUILT74 < 1975 & OWNERS = 2 & NAREA = MEDITERRANEAN, SUEZ	0,53
TYPE = BULK & OWNERS = 3-4 & NAREA = RED SEA E. & W. AFRICAN COAST	0,53
BUILT74 < 1975 & OWNERS = 3-4	0,53
BUILT74 < 1975 & NAREA = MEDITERRANEAN, SUEZ	0,53
TYPE = BULK & NAREA = MEDITERRANEAN, SUEZ & OWNERS = 5++	0,53
BUILT74 < 1975 & OWNERS = 3-4 & TYPE = TANKER	0,53
TYPE = BULK & OWNERS = No info & NAREA = W. INDIES, U.S. EASTERN, GULF OF MEXICO, GREAT LAKES, NEWFOU	0,53
GRUCLASS = OTHERS & NAREA = N. SEA, BALTIC, KIEL	0,53
BUILT74 < 1975 & GRUCLASS = LR & OWNERS = 2	0,53
GRUCLASS = OTHERS & OWNERS = No info & NAREA = W. INDIES, U.S. EASTERN, GULF OF MEXICO, GREAT LAKES, NEWFOU	0,53
TYPE = BULK & NAREA = CANADIAN ALASKA, U.S.S.R. ARTIC & BERNING SEA, ICELAND	0,52
BUILT74 < 1975 & GRUCLASS = AB-GL-NK	0,52
BUILT74 < 1975 & GRUCLASS = BV-CC-KR-NV-RI-PR-RS & OWNERS = 3-4	0,52
NAREA = JAPAN, S. CHINA & E. INDIES & GRUCLASS = LR	0,52
BUILT74 < 1975 & GRUCLASS = LR & TYPE = TANKER	0,52
TYPE = BULK & BUILT74 < 1975 & GRUCLASS = LR & OWNERS = 2	0,52
BUILT74 < 1975 & GRUCLASS = AB-GL-NK & OWNERS = 3-4	0,52
OWNERS = No info & NAREA = W. INDIES, U.S. EASTERN, GULF	0,52



OF MEXICO, GREAT LAKES, NEWFOU	
BUILT74 < 1975 & GRUCLASS = BV-CC-KR-NV-RI-PR-RS & OWNERS = 1	0,52
BUILT74 < 1975 & OWNERS = 1 & TYPE = FERRY	0,52
TYPE = BULK & OWNERS = 2 & NAREA = MEDITERRANEAN, SUEZ	0,52
TYPE = BULK & NAREA = JAPAN, S. CHINA & E. INDIES & OWNERS = 1	0,52
BUILT74 < 1975 & GRUCLASS = BV-CC-KR-NV-RI-PR-RS & NAREA = MEDITERRANEAN, SUEZ	0,52
GRUCLASS = OTHERS & NAREA = W. INDIES, U.S. EASTERN, GULF OF MEXICO, GREAT LAKES, NEWFOU	0,51
GRUCLASS = AB-GL-NK & NAREA = W. COAST S-N. AMERICA, PANAMA CAPE HORN	0,51
OWNERS = 3-4 & NAREA = ATLANTIC	0,51
TYPE = BULK & GRUCLASS = OTHERS & NAREA = W. INDIES, U.S. EASTERN, GULF OF MEXICO, GREAT LAKES, NEWFOU	0,51
TYPE = BULK & BUILT74 < 1975 & OWNERS = 1 & GRUCLASS = AB-GL-NK	0,51
TYPE = BULK & BUILT74 < 1975 & GRUCLASS = LR	0,51
BUILT74 < 1975 & TYPE = FERRY & NAREA = MEDITERRANEAN, SUEZ	0,51
NAREA = JAPAN, S. CHINA & E. INDIES & GRUCLASS = AB-GL- NK	0,51
TYPE = BULK	0,51
OWNERS = 2 & NAREA = RED SEA E. & W. AFRICAN COAST	0,51
TYPE = BULK & GRUCLASS = BV-CC-KR-NV-RI-PR-RS & OWNERS = 2	0,51
BUILT74 < 1975 & NAREA = MEDITERRANEAN, SUEZ & OWNERS = 5++	0,51
GRUCLASS = OTHERS & OWNERS = No info & TYPE = CONTAINER	0,51
OWNERS = 2 & TYPE = FERRY	0,51

GRUCLASS = AB-GL-NK & NAREA = GULF, BAY OF BENGAL	0,50
TYPE = BULK & GRUCLASS = BV-CC-KR-NV-RI-PR-RS & NAREA = MEDITERRANEAN, SUEZ	0,50
TYPE = BULK & NAREA = MEDITERRANEAN, SUEZ	0,50
BUILT74 < 1975 & OWNERS = 1	0,50
GRUCLASS = BV-CC-KR-NV-RI-PR-RS & OWNERS = 2	0,50
GRUCLASS = AB-GL-NK & OWNERS = 5++	0,50
OWNERS = No info & TYPE = CONTAINER	0,50
BUILT74 < 1975 & GRUCLASS = LR	0,50
TYPE = BULK & BUILT74 < 1975 & NAREA = W. INDIES, U.S. EASTERN, GULF OF MEXICO, GREAT LAKES, NEWFOU	0,50
GRUCLASS = LR & OWNERS = 5++	0,50
BUILT74 < 1975 & GRUCLASS = AB-GL-NK & OWNERS = 2	0,50
BUILT74 < 1975 & GRUCLASS = LR & OWNERS = 5++	0,49
TYPE = BULK & BUILT74 < 1975 & OWNERS = 3-4 & NAREA = MEDITERRANEAN, SUEZ	0,49
GRUCLASS = BV-CC-KR-NV-RI-PR-RS & OWNERS = 3-4 & TYPE = TANKER	0,49
GRUCLASS = OTHERS & TYPE = CONTAINER	0,49
TYPE = BULK & BUILT74 < 1975 & GRUCLASS = LR & NAREA = W. INDIES, U.S. EASTERN, GULF OF MEXICO, GREAT LAKES, NEWFOU	0,49
TYPE = BULK & OWNERS = 2	0,49
TYPE = BULK & NAREA = W. COAST S-N. AMERICA, PANAMA CAPE HORN	0,49
TYPE = FERRY & NAREA = MEDITERRANEAN, SUEZ	0,49
TYPE = BULK & BUILT74 < 1975 & NAREA = N. SEA, BALTIC, KIEL	0,49
TYPE = BULK & GRUCLASS = BV-CC-KR-NV-RI-PR-RS	0,48
BUILT74 < 1975 & NAREA = W. INDIES, U.S. EASTERN, GULF OF MEXICO, GREAT LAKES, NEWFOU	0,48
TYPE = BULK & BUILT74 < 1975 & OWNERS = 1	0,48
NAREA = MEDITERRANEAN, SUEZ	0,48

NAREA = MEDITERRANEAN, SUEZ & OWNERS = 5++	0,48
BUILT74 < 1975 & GRUCLASS = BV-CC-KR-NV-RI-PR-RS & NAREA = W. INDIES, U.S. EASTERN, GULF OF MEXICO, GREAT LAKES, NEWFOU	0,48
TYPE = BULK & BUILT74 < 1975 & GRUCLASS = AB-GL-NK & NAREA = N. SEA, BALTIC, KIEL	0,48
TYPE = BULK & NAREA = N. SEA, BALTIC, KIEL & OWNERS = 5++	0,48
OWNERS = 2 & NAREA = MEDITERRANEAN, SUEZ	0,47
OWNERS = 1 & NAREA = W. COAST S-N. AMERICA, PANAMA CAPE HORN	0,47
BUILT74 < 1975 & OWNERS = 2 & NAREA = W. INDIES, U.S. EASTERN, GULF OF MEXICO, GREAT LAKES, NEWFOU	0,47
NAREA = JAPAN, S. CHINA & E. INDIES & OWNERS = 1	0,47
BUILT74 < 1975 & NAREA = N. SEA, BALTIC, KIEL & OWNERS = 5++	0,47
TYPE = TANKER	0,47
OWNERS = 2	0,47
OWNERS = 1 & NAREA = GULF, BAY OF BENGAL	0,47
GRUCLASS = AB-GL-NK & NAREA = RED SEA E. & W. AFRICAN COAST	0,47
TYPE = BULK & GRUCLASS = BV-CC-KR-NV-RI-PR-RS & OWNERS = 3-4	0,47
TYPE = BULK & BUILT74 < 1975 & GRUCLASS = BV-CC-KR-NV-RI-PR-RS & OWNERS = 1	0,47
BUILT74 < 1975 & OWNERS = 1 & GRUCLASS = AB-GL-NK	0,47
GRUCLASS = BV-CC-KR-NV-RI-PR-RS	0,47
NAREA = W. COAST S-N. AMERICA, PANAMA CAPE HORN	0,46
BUILT74 < 1975 & GRUCLASS = LR & NAREA = W. INDIES, U.S. EASTERN, GULF OF MEXICO, GREAT LAKES, NEWFOU	0,46
GRUCLASS = BV-CC-KR-NV-RI-PR-RS & NAREA = JAPAN, S. CHINA & E. INDIES & OWNERS = 1	0,46
TYPE = BULK & OWNERS = 3-4 & GRUCLASS = LR	0,46

TYPE = BULK & NAREA = JAPAN, S. CHINA & E. INDIES & OWNERS = 1 & GRUCLASS = LR	0,46
GRUCLASS = BV-CC-KR-NV-RI-PR-RS & NAREA = MEDITERRANEAN, SUEZ	0,46
GRUCLASS = BV-CC-KR-NV-RI-PR-RS & OWNERS = 3-4	0,45
TYPE = BULK & OWNERS = 3-4	0,45
BUILT74 < 1975 & OWNERS = 3-4 & NAREA = MEDITERRANEAN, SUEZ	0,45
BUILT74 < 1975 & GRUCLASS = LR & TYPE = FERRY	0,45
TYPE = FERRY & NAREA = W. INDIES, U.S. EASTERN, GULF OF MEXICO, GREAT LAKES, NEWFOU	0,45
TYPE = BULK & GRUCLASS = AB-GL-NK & NAREA = MEDITERRANEAN, SUEZ	0,45
BUILT74 < 1975 & NAREA = N. SEA, BALTIC, KIEL	0,45
NAREA = ATLANTIC	0,44
GRUCLASS = BV-CC-KR-NV-RI-PR-RS & TYPE = TANKER	0,44
OWNERS = 3-4	0,44
NAREA = JAPAN, S. CHINA & E. INDIES & TYPE = CONTAINER	0,44
OWNERS = 3-4 & TYPE = TANKER	0,44
GRUCLASS = LR & NAREA = GULF, BAY OF BENGAL	0,44
OWNERS = 3-4 & GRUCLASS = LR	0,44
BUILT74 < 1975 & GRUCLASS = BV-CC-KR-NV-RI-PR-RS & TYPE = FERRY	0,44
BUILT74 < 1975 & TYPE = CONTAINER	0,44
BUILT74 < 1975 & GRUCLASS = AB-GL-NK & NAREA = N. SEA, BALTIC, KIEL	0,43
TYPE = BULK & OWNERS = 3-4 & NAREA = MEDITERRANEAN, SUEZ	0,43
TYPE = TANKER & OWNERS = 2	0,43
TYPE = BULK & GRUCLASS = AB-GL-NK & OWNERS = 2	0,43
TYPE = BULK & GRUCLASS = AB-GL-NK & OWNERS = 3-4	0,43
TYPE = BULK & GRUCLASS = AB-GL-NK	0,43
OWNERS = 1 & TYPE = FERRY	0,43



BUILT74 < 1975 & GRUCLASS = AB-GL-NK & TYPE = TANKER	0,43
TYPE = BULK & BUILT74 < 1975 & NAREA = N. SEA, BALTIC, KIEL & OWNERS = 3-4	0,43
TYPE = BULK & BUILT74 < 1975 & NAREA = N. SEA, BALTIC, KIEL & GRUCLASS = LR	0,42
GRUCLASS = BV-CC-KR-NV-RI-PR-RS & TYPE = FERRY & NAREA = MEDITERRANEAN, SUEZ	0,42
GRUCLASS = AB-GL-NK & OWNERS = 3-4	0,42
TYPE = BULK & BUILT74 < 1975 & GRUCLASS = BV-CC-KR-NV- RI-PR-RS & NAREA = N. SEA, BALTIC, KIEL	0,42
GRUCLASS = BV-CC-KR-NV-RI-PR-RS & TYPE = FERRY	0,42
OWNERS = 3-4 & TYPE = FERRY	0,42
TYPE = BULK & NAREA = ATLANTIC	0,42
TYPE = BULK & GRUCLASS = LR & OWNERS = 2	0,41
GRUCLASS = BV-CC-KR-NV-RI-PR-RS & OWNERS = 3-4 & NAREA = MEDITERRANEAN, SUEZ	0,41
GRUCLASS = AB-GL-NK & OWNERS = 2	0,41
BUILT74 < 1975 & NAREA = N. SEA, BALTIC, KIEL & TYPE = TANKER	0,41
BUILT74 < 1975 & OWNERS = 1 & NAREA = MEDITERRANEAN, SUEZ	0,41
TYPE = BULK & GRUCLASS = LR	0,41
GRUCLASS = LR & OWNERS = 2	0,40
BUILT74 < 1975 & OWNERS = 1 & GRUCLASS = LR	0,40
TYPE = BULK & BUILT74 < 1975 & NAREA = N. SEA, BALTIC, KIEL & OWNERS = 2	0,40
GRUCLASS = AB-GL-NK & TYPE = FERRY	0,40
OWNERS = 3-4 & NAREA = MEDITERRANEAN, SUEZ	0,40
BUILT74 < 1975 & GRUCLASS = LR & NAREA = MEDITERRANEAN, SUEZ	0,40
NAREA = N. SEA, BALTIC, KIEL & OWNERS = 5++	0,40
NAREA = N. SEA, BALTIC, KIEL & TYPE = FERRY	0,40
TYPE = TANKER & NAREA = MEDITERRANEAN, SUEZ	0,40

NAREA = JAPAN, S. CHINA & E. INDIES & OWNERS = 1 & GRUCLASS = LR	0,39
NAREA = JAPAN, S. CHINA & E. INDIES & GRUCLASS = AB-GL- NK & TYPE = TANKER	0,39
GRUCLASS = AB-GL-NK & TYPE = TANKER & OWNERS = 2	0,39
BUILT74 < 1975 & GRUCLASS = BV-CC-KR-NV-RI-PR-RS & NAREA = N. SEA, BALTIC, KIEL	0,39
BUILT74 < 1975 & OWNERS = 3-4 & TYPE = FERRY	0,39
BUILT74 < 1975 & OWNERS = 3-4 & NAREA = W. INDIES, U.S. EASTERN, GULF OF MEXICO, GREAT LAKES, NEWFOU	0,39
GRUCLASS = BV-CC-KR-NV-RI-PR-RS & OWNERS = 1 & NAREA = MEDITERRANEAN, SUEZ	0,39
GRUCLASS = BV-CC-KR-NV-RI-PR-RS & OWNERS = 1 & TYPE = FERRY	0,38
GRUCLASS = AB-GL-NK	0,38
GRUCLASS = BV-CC-KR-NV-RI-PR-RS & OWNERS = 1 & TYPE = TANKER	0,38
GRUCLASS = LR	0,38
OWNERS = 1 & NAREA = N. SEA, BALTIC, KIEL & TYPE = FERRY	0,38
GRUCLASS = AB-GL-NK & NAREA = MEDITERRANEAN, SUEZ	0,37
BUILT74 < 1975 & TYPE = TANKER & NAREA = MEDITERRANEAN, SUEZ	0,37
OWNERS = 1 & GRUCLASS = LR & TYPE = FERRY	0,37
BUILT74 < 1975 & NAREA = N. SEA, BALTIC, KIEL & OWNERS = 3- 4	0,37
TYPE = BULK & NAREA = W. INDIES, U.S. EASTERN, GULF OF MEXICO, GREAT LAKES, NEWFOU	0,37
TYPE = BULK & OWNERS = 1 & NAREA = MEDITERRANEAN, SUEZ	0,37
TYPE = BULK & OWNERS = 1	0,37
TYPE = BULK & NAREA = N. SEA, BALTIC, KIEL	0,37
BUILT74 < 1975 & OWNERS = 1 & NAREA = W. INDIES, U.S. EASTERN, GULF OF MEXICO, GREAT LAKES, NEWFOU	0,36

NAREA = W. INDIES, U.S. EASTERN, GULF OF MEXICO, GREAT LAKES, NEWFOU	0,36
BUILT74 < 1975 & OWNERS = 1 & NAREA = N. SEA, BALTIC, KIEL	0,36
GRUCLASS = BV-CC-KR-NV-RI-PR-RS & NAREA = W. INDIES, U.S. EASTERN, GULF OF MEXICO, GREAT LAKES, NEWFOU	0,36
BUILT74 < 1975 & NAREA = N. SEA, BALTIC, KIEL & GRUCLASS = LR	0,36
OWNERS = 1	0,36
GRUCLASS = BV-CC-KR-NV-RI-PR-RS & NAREA = ATLANTIC	0,35
TYPE = BULK & NAREA = N. SEA, BALTIC, KIEL & OWNERS = 3-4	0,35
NAREA = JAPAN, S. CHINA & E. INDIES & OWNERS = 1 & TYPE = TANKER	0,35
TYPE = BULK & GRUCLASS = LR & NAREA = W. INDIES, U.S. EASTERN, GULF OF MEXICO, GREAT LAKES, NEWFOU	0,35
BUILT74 < 1975 & GRUCLASS = AB-GL-NK & NAREA = W. INDIES, U.S. EASTERN, GULF OF MEXICO, GREAT LAKES, NEWFOU	0,35
GRUCLASS = BV-CC-KR-NV-RI-PR-RS & OWNERS = 1	0,35
GRUCLASS = BV-CC-KR-NV-RI-PR-RS & NAREA = N. SEA, BALTIC, KIEL & OWNERS = 3-4	0,35
OWNERS = 1 & NAREA = MEDITERRANEAN, SUEZ	0,35
TYPE = BULK & GRUCLASS = AB-GL-NK & NAREA = N. SEA, BALTIC, KIEL & OWNERS = 3-4	0,35
NAREA = N. SEA, BALTIC, KIEL	0,35
GRUCLASS = LR & TYPE = TANKER	0,34
TYPE = BULK & GRUCLASS = LR & NAREA = MEDITERRANEAN, SUEZ	0,34
BUILT74 < 1975 & NAREA = N. SEA, BALTIC, KIEL & OWNERS = 2	0,34
TYPE = CONTAINER	0,34
GRUCLASS = LR & TYPE = FERRY	0,34
OWNERS = 2 & NAREA = W. INDIES, U.S. EASTERN, GULF OF MEXICO, GREAT LAKES, NEWFOU	0,34
TYPE = BULK & OWNERS = 1 & GRUCLASS = AB-GL-NK	0,33



TYPE = TANKER & NAREA = W. INDIES, U.S. EASTERN, GULF OF MEXICO, GREAT LAKES, NEWFOU	0,33
TYPE = BULK & GRUCLASS = BV-CC-KR-NV-RI-PR-RS & NAREA = W. INDIES, U.S. EASTERN, GULF OF MEXICO, GREAT LAKES, NEWFOU	0,33
GRUCLASS = LR & NAREA = MEDITERRANEAN, SUEZ	0,33
TYPE = BULK & GRUCLASS = BV-CC-KR-NV-RI-PR-RS & OWNERS = 1	0,33
TYPE = BULK & GRUCLASS = BV-CC-KR-NV-RI-PR-RS & NAREA = N. SEA, BALTIC, KIEL	0,33
GRUCLASS = LR & NAREA = W. INDIES, U.S. EASTERN, GULF OF MEXICO, GREAT LAKES, NEWFOU	0,32
TYPE = BULK & OWNERS = 2 & NAREA = W. INDIES, U.S. EASTERN, GULF OF MEXICO, GREAT LAKES, NEWFOU	0,32
GRUCLASS = AB-GL-NK & NAREA = N. SEA, BALTIC, KIEL & OWNERS = 2	0,31
NAREA = N. SEA, BALTIC, KIEL & OWNERS = 3-4	0,31
OWNERS = 1 & TYPE = TANKER	0,31
BUILT74 < 1975 & NAREA = N. SEA, BALTIC, KIEL & TYPE = FERRY	0,31
TYPE = BULK & GRUCLASS = AB-GL-NK & NAREA = N. SEA, BALTIC, KIEL	0,30
GRUCLASS = AB-GL-NK & NAREA = N. SEA, BALTIC, KIEL & OWNERS = 3-4	0,30
GRUCLASS = BV-CC-KR-NV-RI-PR-RS & NAREA = N. SEA, BALTIC, KIEL	0,29
TYPE = BULK & OWNERS = 1 & GRUCLASS = LR	0,29
GRUCLASS = AB-GL-NK & TYPE = TANKER	0,28
OWNERS = 1 & GRUCLASS = LR	0,28
NAREA = JAPAN, S. CHINA & E. INDIES & OWNERS = 1 & GRUCLASS = AB-GL-NK	0,28
TYPE = BULK & NAREA = N. SEA, BALTIC, KIEL & OWNERS = 2	0,27
GRUCLASS = BV-CC-KR-NV-RI-PR-RS & NAREA = N. SEA,	0,27



BALTIC, KIEL & TYPE = FERRY	
OWNERS = 3-4 & NAREA = W. INDIES, U.S. EASTERN, GULF OF MEXICO, GREAT LAKES, NEWFOU	0,27
TYPE = BULK & NAREA = N. SEA, BALTIC, KIEL & GRUCLASS = LR	0,27
GRUCLASS = BV-CC-KR-NV-RI-PR-RS & OWNERS = 1 & NAREA = N. SEA, BALTIC, KIEL	0,27
TYPE = BULK & OWNERS = 3-4 & NAREA = W. INDIES, U.S. EASTERN, GULF OF MEXICO, GREAT LAKES, NEWFOU	0,26
OWNERS = 1 & GRUCLASS = AB-GL-NK	0,26
GRUCLASS = AB-GL-NK & NAREA = N. SEA, BALTIC, KIEL	0,26
NAREA = N. SEA, BALTIC, KIEL & TYPE = TANKER	0,26
TYPE = BULK & GRUCLASS = BV-CC-KR-NV-RI-PR-RS & OWNERS = 1 & NAREA = N. SEA, BALTIC, KIEL	0,25
NAREA = N. SEA, BALTIC, KIEL & GRUCLASS = LR	0,24
TYPE = BULK & GRUCLASS = AB-GL-NK & NAREA = W. INDIES, U.S. EASTERN, GULF OF MEXICO, GREAT LAKES, NEWFOU	0,23
NAREA = N. SEA, BALTIC, KIEL & OWNERS = 2	0,23
OWNERS = 1 & NAREA = N. SEA, BALTIC, KIEL	0,22
OWNERS = 1 & NAREA = W. INDIES, U.S. EASTERN, GULF OF MEXICO, GREAT LAKES, NEWFOU	0,21
TYPE = BULK & NAREA = AUSTRALIA	0,21
OWNERS = 1 & TYPE = CONTAINER	0,20
GRUCLASS = AB-GL-NK & NAREA = W. INDIES, U.S. EASTERN, GULF OF MEXICO, GREAT LAKES, NEWFOU	0,20
TYPE = BULK & OWNERS = 1 & NAREA = W. INDIES, U.S. EASTERN, GULF OF MEXICO, GREAT LAKES, NEWFOU	0,20
TYPE = BULK & OWNERS = 1 & NAREA = N. SEA, BALTIC, KIEL	0,19
GRUCLASS = AB-GL-NK & TYPE = CONTAINER	0,18
OWNERS = 1 & GRUCLASS = AB-GL-NK & TYPE = TANKER	0,18
OWNERS = 1 & NAREA = N. SEA, BALTIC, KIEL & GRUCLASS = LR	0,17
OWNERS = 1 & GRUCLASS = LR & NAREA = W. INDIES, U.S.	0,17

EASTERN, GULF OF MEXICO, GREAT LAKES, NEWFOU  
OWNERS = 1 & GRUCLASS = AB-GL-NK & NAREA = N. SEA,  
BALTIC, KIEL  
NAREA = AUSTRALIA

0,16

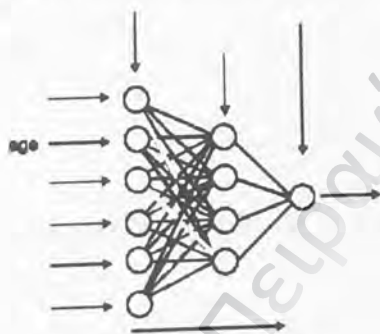
0,15

Πανεπιστήμιο Πειραιώς

## ΠΑΡΑΡΤΗΜΑ Α

### N. Η μεθοδολογία των νευρωνικών δικτύων.

Δίνονται στη συνέχεια κάποια βασικά σχόλια για τα νευρωνικά δίκτυα. Αυτά είναι μοντέλα που μιμούνται τη λειτουργία του νευρικού συστήματος. Η βασική τους



ΣΧΗΜΑ

μονάδα είναι οι νευρώνες – οι κύκλοι στο σχήμα. Αυτοί ακριβώς οι νευρώνες οργανώνονται σε επίπεδα για να επεξεργαστούν τα δεδομένα (αριστερή είσοδος) ώστε με τη βοήθεια ενός ή περισσότερων ενδιάμεσων σταδίων (layers) να εκτιμήσουν το αποτέλεσμα (δεξιά έξοδος του γραφήματος).

Κάθε μετασχηματισμός των δεδομένων από τα αριστερά προς τα δεξιά γίνεται με τη βοήθεια συντελεστήν στάθμισης. Αυτοί οι συντελεστές αρχικά ορίζονται στη τύχη και με διαδοχικά βήματα εκτιμούνται έτσι ώστε το τελικό αποτέλεσμα να εκτιμάται με το καλύτερο δυνατό τρόπο. Τα νευρωνικά χρησιμοποιούν σύνολο εκπαίδευσης και επαλήθευσης για να παράγουν έγκυρα αποτελέσματα και αυτή ακριβώς η διαδικασία, που περιγράφηκε παραπάνω, ονομάζεται «εκπαίδευση». Στο τέλος της παράγεται ένα δίκτυο το οποίο είναι ένα μη γραμμικό μοντέλο που εκτιμά με τον καλύτερο δυνατό τρόπο το αποτέλεσμα. Αυτό το μοντέλο μπορεί στη συνέχεια να χρησιμοποιηθεί για να εκτιμηθεί το αποτέλεσμα όταν αυτό δεν είναι γνωστό.

## ΠΑΡΑΡΤΗΜΑ Β

### Α. Νέοι αλγόριθμοι και προγράμματα

Βασικές ομάδες για τα προγράμματα

\*\*\* 1. Data files .

\*\*\* 2. Intermediate steps .

\*\*\* 3. Analysis for total loss .

\*\*\* 1. Data files .

GET

FILE='C:\D.data\DATA ALL\020202dk\99-92-4.sav'.

EXECUTE .

sort cases by aa .

MATCH FILES /FILE=\*

/FILE='C:\D.data\DATA ALL\020216 dko\data\ship1a.sav'

/BY aa.

EXECUTE.

FILTER OFF.

USE ALL.

SELECT IF(aa1 > 0).

EXECUTE .

SAVE OUTFILE='C:\D.data\DATA ALL\020216 dko\data\all1b.sav'

/COMPRESSED.

new file .

GET

FILE='C:\D.data\DATA ALL\020216 dko\data\all1b.sav'.

EXECUTE .

Compute qn = \$casenum .

COMPUTE day92 = YRMODA(daty,datm,datd) - 149462 .

COMPUTE dacc92 = YRMODA(acc2yr ,acc2m,acc2d ) - 149462 .

Compute acc1old = acc1 .

Compute acc2old = acc2 .



Compute acc3old = acc3 .

Compute acc4old = acc4 .

Compute acc5old = acc5 .

Recode acc1 acc2 acc3 acc4 acc5

( 804 = 16)

( 805 = 17)

( 901 = 24)

( 902 = 25)

( 903 = 26)

( 904 = 27)

( 905 = 28)

( 301 = 30)

(66 = 65) .

Recode acc1 acc2 acc3 acc4 acc5

( 0 thru 9 = 0)

(10 thru 19 = 1)

(20 thru 29 = 2)

(30 thru 39 = 3)

(40 thru 49 = 4)

(50 thru 59 = 5)

(60 thru 69 = 6)

(70 thru 79 = 7)

(80 thru 89 = 8)

(90 thru 99 = 9)

into ac1 ac2 ac3 ac4 ac5 .

Recode areacode

(1,2,3 = 1 )

(4,5,6 = 2 )

(7,9,19 = 3)

(23,24,25,26,28 = 4)

(10 = 5)

(8,11 = 6)

(12,13 = 7)

(14,15 = 8)

```
(16,17 = 9)
(18,31 = 10)
(20,21,22 = 11)
(27,29,30 = 12)
```

into narea.

Value labels narea

```
1 ' 1,2,3 '
2 ' 4,5,6 '
3 ' 7,9,19 '
4 ' 23,24,25,26,28 '
5 '10'
6 '8,11'
7 '12,13'
8 '14,15'
9 '16,17'
10 '18,31'
11 ' 20,21,22 '
12 ' 27,29,30 '.
```

String cc1 (a3) .

```
if(substr(class,1,1) = '(') cc1 = substr(class,2,2) .
```

exe .

```
if(substr(class,1,1) ~= '(') cc1 = substr(class,1,2) .
```

exe .

```
SAVE OUTFILE='C:\D.data\DATA ALL\020216 dko\data\all1b.sav'
```

```
/COMPRESSED.
```

\*\*\* 2. Intermediate steps .

```
Define repmis ( !pos !tokens(1) ) .
```

```
Do if(samea = 1) .
```

```
if( missing( !1 ) = 1 ) !1 = lag(!1) .
```

```
End if .
```

exe .

```
!enddefine .
```

```
Define mcd5 ( !pos !tokens(1)
```

```

        /!pos !tokens(1) ) .
!do !i = !1 !to !2 .
Compute !concat(dacc,!i) = 2 .
if ( acc1=!i | acc2=!i | acc3=!i | acc4=!i | acc5=!i ) !concat(dacc,!i) = 1 .
!doend .
exe .
!enddefine .
Define acc5 ( !pos !tokens(1)
        /!pos !tokens(1) ) .
Do if (samea = 1) .
        !do !i = !1 !to !2 .
        if ( acc1 = !i ) !concat(dacc,!i) = 1 .
        if ( lag(acc1) = !i ) !concat(dacc,!i) = 1 .
        !doend .
End if .
exe .
!enddefine .
Define acc6 ( !pos !tokens(1)
        /!pos !tokens(1) ) .
Do if (samea = 1) .
        !do !i = !1 !to !2 .
        if (lag( !concat(dacc,!i) ) = 1 ) !concat(dacc,!i) = 1 .
        !doend .
End if .
exe .
!enddefine .
Define f (!pos !tokens(1)
        /mi !default(2) !TOKENS(1)
        /ma !default(99) !TOKENS(1)
) .
DESCRIPTIVES
VARIABLES= !1
/STATISTICS=MEAN STDDEV MIN MAX .
FREQUENCIES

```

```

VARIABLES=!1
/FORMAT=AFREQ LIMIT(150)
/ORDER ANALYSIS .
FREQUENCIES VARIABLES=!1
/FORMAT LIMIT(50) .
USE ALL.
COMPUTE filter_$=( !1 <= !mi | !1 >= !ma ).
FILTER BY filter_$.
EXECUTE .
SUMMARIZE
/TABLES=aa !1
/FORMAT=VALIDLIST NOCASENUM TOTAL LIMIT=200
/TITLE='Case Summaries'
/MISSING=VARIABLE
/CELLS=COUNT .
FILTER OFF.
USE ALL.
EXECUTE .
!enddefine .
Define bi (!pos !tokens(1)
    /!pos !tokens(1) ) .
SORT CASES BY !1 !2 .
SPLIT FILE LAYERED BY !1 !2 .
SUMMARIZE
/TABLES= qn name
/FORMAT=VALIDLIST NOCASENUM TOTAL LIMIT=1000
/TITLE='Case Summaries'
/MISSING=VARIABLE
/CELLS=COUNT .
SPLIT FILE OFF.
sort cases by qn .
!enddefine .
Define su ( !pos !cmdend) .
USE ALL.

```



```

FILTER BY filter .
EXECUTE .
SUMMARIZE
  /TABLES= aa !1
  /FORMAT=VALIDLIST NOCASENUM TOTAL LIMIT=100
  /TITLE='Case Summaries'
  /MISSING=VARIABLE
  /CELLS=COUNT .
compute filter = 0 .
FILTER OFF.
USE ALL.
EXECUTE .
!enddefine .
define m34(!pos !tokens(1)
  /!pos !tokens(1)
  /!pos !tokens(1)
  /!pos !tokens(1)
).
sort cases by !1 !2 !3 !4 .
CREATE
  /z41=LEAD(!1 1) .
CREATE
  /z42=LEAD(!2 1) .
CREATE
  /z43=LEAD(!3 1) .
CREATE
  /z44=LEAD(!4 1) .
if( !1 = z41 and
  !2 = z42 and
  !3 = z43 and
  !4 = -1
)
!4 = z44 .
exe.

```

```

!enddefine .
define m24 (!pos !tokens(1)
    /!pos !tokens(1)
    /!pos !tokens(1)
    /!pos !tokens(1)
) .
sort cases by !1 !2 !3 !4 .
CREATE
/z41=LEAD(!1 1) .
CREATE
/z42=LEAD(!2 1) .
CREATE
/z43=LEAD(!3 1) .
CREATE
/z44=LEAD(!4 1) .
Do if (
    !1 = z41 and
    !2 = z42 and
    !3 = -1 and
    !4 = -1
) .
Compute !3 = z43 .
Compute !4 = z44 .
End if .
exe.
!enddefine .
define m14 (!pos !tokens(1)
    /!pos !tokens(1)
    /!pos !tokens(1)
    /!pos !tokens(1)
) .
sort cases by !1 !2 !3 !4 .
CREATE
/z41=LEAD(!1 1) .

```

```

CREATE
  /z42=LEAD(!2 1) .
CREATE
  /z43=LEAD(!3 1) .
CREATE
  /z44=LEAD(!4 1) .
Do if (
  !1 = z41  and
  !2 = -1  and
  !3 = -1   and
  !4 = -1
) .
  Compute !2 = z42 .
  Compute !3 = z43 .
  Compute !4 = z44 .
End if .
exe .
!enddefine .
define r34 (!pos !tokens(1)
  !/pos !tokens(1)
  !/pos !tokens(1)
  !/pos !tokens(1)
) .
sort cases by !1 !2 !3 !4 .
CREATE
  /z41=LEAD(!1 1) .
CREATE
  /z42=LEAD(!2 1) .
CREATE
  /z43=LEAD(!3 1) .
CREATE
  /z44=LEAD(!4 1) .
Compute filter = 0 .
Do if (

```

```

!1 = z41 and
!2 = z42 and
!3 = z43 and
!4 ~ z44
).

```

Compute filter = 1 .

End if .

exe.

CREATE

```
/zfilter=Lag(filter 1) .
```

```
if (zfilter = 1) filter = 1 .
```

```
Compute z61 = aa .
```

```
Compute z62 = !1 .
```

```
Compute z63 = !2 .
```

```
Compute z64 = !3 .
```

```
Compute z65 = !4 .
```

```
Compute z66 = filter .
```

exe.

!enddefine .

```
define r24 (!pos !tokens(1)
```

```
  !/pos !tokens(1)
```

```
  !/pos !tokens(1)
```

```
  !/pos !tokens(1)
```

) .

```
sort cases by !1 !2 !3 !4 .
```

CREATE

```
/z41=LEAD(!1 1) .
```

CREATE

```
/z42=LEAD(!2 1) .
```

CREATE

```
/z43=LEAD(!3 1) .
```

CREATE

```
/z44=LEAD(!4 1) .
```

```
Compute filter = 0 .
```



Do if (

```
!1 = z41 and
!2 = z42 and
!3 ~ = z43 and
!4 ~ = z44
```

).

Compute filter = 1 .

End if .

exe.

CREATE

```
/zfilter=Lag(filter 1) .
```

```
if (zfilter = 1) filter = 1 .
```

```
Compute z61 = aa .
```

```
Compute z62 = !1 .
```

```
Compute z63 = !2 .
```

```
Compute z64 = !3 .
```

```
Compute z65 = !4 .
```

```
Compute z66 = filter .
```

exe.

```
!enddefine .
```

```
define r14 (!pos !tokens(1)
```

```
!pos !tokens(1)
```

```
!pos !tokens(1)
```

```
!pos !tokens(1)
```

).

```
sort cases by !1 !2 !3 !4 .
```

CREATE

```
/z41=LEAD(!1 1) .
```

CREATE

```
/z42=LEAD(!2 1) .
```

CREATE

```
/z43=LEAD(!3 1) .
```

CREATE

```
/z44=LEAD(!4 1) .
```

Compute filter = 0 .

Do if(

```
    !1 = z41    and
      !2 ~= z42  and
      !3 ~= z43  and
      !4 ~= z44
```

),

Compute filter = 1 .

End if .

exe.

CREATE

```
/zfilter=Lag(filter 1) .
```

```
if (zfilter = 1 ) filter = 1 .
```

```
Compute z61 = aa .
```

```
Compute z62 = !1 .
```

```
Compute z63 = !2 .
```

```
Compute z64 = !3 .
```

```
Compute z65 = !4 .
```

```
Compute z66 = filter .
```

exe.

```
!enddefine .
```

```
define fil (!pos !tokens(1)
```

```
    !/pos !tokens(1)
```

```
    !/pos !tokens(1)
```

```
    !/pos !tokens(1) ) .
```

```
sort cases by !1 !2 !3 !4 .
```

CREATE

```
/z41=LEAD(!1 1) .
```

CREATE

```
/z42=LEAD(!2 1) .
```

CREATE

```
/z43=LEAD(!3 1) .
```

CREATE

```
/z44=LEAD(!4 1) .
```

Compute filt1 = 1 .

Do if (

!1 = z41 and

!2 = z42 and

!3 = z43 and

!4 = z44

).

Compute filt1 = 2 .

End if .

exe.

Do if (

!1 ~= z41 or

!2 ~= z42 or

!3 ~= z43 or

!4 ~= z44

).

Compute filt1 = 4 .

End if .

exe.

Do if ( sysmis(!1) = 1 or

sysmis(!2) = 1 or

sysmis(!3) = 1 or

sysmis(!4) = 1

).

Compute filt1 = 3 .

End if .

Do if ( sysmis(!1) = 1 and

sysmis(!2) = 1 and

sysmis(!3) = 1 and

sysmis(!4) = 1

).

Compute filt1 = 0 .

End if .

!enddefine .

```

*** samea .
sort cases by ship1 aa .
CREATE
  /s1a =LEAD(ship1 1) .
CREATE
  /dayaccl =LEAD(dayacc92 1) .
Compute sameacc = 0 .
Do if ( s1a = ship1 ) .
  if ( abs(dayaccl -dayacc92) < 61 ) sameacc = lag(sameacc) + 1 .
End if .
Compute samea = lag(sameacc) + 1 .
exe .
*** sameb .
sort cases by ship1 .
Compute sameb = 1 .
if (ship1 = lag( ship1 ) ) sameb = 1+lag(sameb) .
exe .
*** filt1 .
sort cases by ship1 .
CREATE
  /samebl =LEAD(sameb 1) .
Compute filt = 2 .
Do if ( sameb = samebl ) .
  Compute filt = 1 .
end if .
exe .
Missing valid name1 BUILT74t1 grt1 type1 ( -1 ) .
if ( NVALID(name1, built1, grt1, type1) < 4 ) filt = 3 .
EXECUTE .
Missing valid name1 built1 grt1 type1 ( ) .
Value labels filt1
1 'Unique - OK' .
2 'Multiple - OK'
3 'With missing values' .

```



SORT CASES BY

ship1 (A) narea (A) dacc92 (A) .

CREATE

/z51=LEAD(ship1 1) .

CREATE

/z52=LEAD(narea 1) .

CREATE

/z53=LEAD(dacc92 1) .

Compute x = 0 .

Do if ( (z51= ship1) and (z52 = narea) and (abs(z53-dacc92)<60) ) .

    Compute x = 1 .

Else if ( (z51= ship1 ) and ( (z52 ~= narea) or (abs(z53-dacc92)>=60) ) ) .

    Compute x = 2 .

End if .

Value labels x

0 'Next case different'

1 'Next case same ship area day'

2 'Only ship same' .

freq x .

Compute seqacc = 1 .

exe .

SAVE OUTFILE='C:\D.data\DATA ALL\020217dk\accident a-3 abcd.sav'

  /keep

aa name ship1 datd datm datyr

areacode areaname latid longd

day92 dacc92 x acc1 acc2 acc3 acc4 acc5

name

  /COMPRESSED.

new file .

GET

  FILE='C:\D.data\DATA ALL\020217dk\accident a-3 abcd.sav'.

EXECUTE .

\*\*\* 3. Analysis for total loss .

Define o (!pos !tokens(1)) .

ONEWAY

!1 BY totloss

/STATISTICS DESCRIPTIVES

/MISSING ANALYSIS .

!enddefine .

Define t (!pos !tokens(1)) .

TABLES

/FORMAT BLANK MISSING(,.)

/GBASE=CASES

/FTOTAL= \$t000001 "Total"

/TABLE=!1 + \$t000001 BY ena + totloss

/STATISTICS

count( totloss( F5.0 ))

cpct( totloss( PCT5.1 ) 'Row %': !1 )

cpct( totloss( PCT5.1 ) 'Col %':ena totloss ).

!enddefine .

f type .

f code .

f datd .

f datm .

f datyr .

f publ .

f ent .

f name .

f flag .

f type .

f prname .

f owner .

f lastown .

f lastmang .

f yearown .

f grt .

f class .

f built .  
f age .  
f shipyard .  
f design .  
f designat .  
f power .  
f engbuild .  
f speed .  
f desport .  
f despord .  
f desporm .  
f desporyr .  
f arrport .  
f arrpord .  
f arrporm .  
f arrporyr .  
f load .  
f acc1 .  
f acc1d .  
f acc1m .  
f acc1yr .  
f acc2 .  
f acc2d .  
f acc2m .  
f acc2yr .  
f acc3 .  
f acc3d .  
f acc3m .  
f acc3yr .  
f acc4 .  
f acc4d .  
f acc4m .  
f acc4yr .  
f acc5 .

Πανεπιστήμιο Πειραιώς

f acc5d .  
f acc5m .  
f acc5yr .  
f areatype .  
f areaname .  
f areacode .  
f latid .  
f longd .  
f voyagest .  
f missing .  
f injured .  
f dead .  
f cost .  
f reptime .  
f timelost .  
f tloss .  
f tloss1 .  
f loadlost .  
f costld1 .  
f pollutio .  
f chempol .  
f chemqty .  
f clean .  
f wayclean .  
f cleancos .  
f weath .  
t samecl .  
t datyr .  
t month .  
t narea .  
t voyage .  
t flag2 .  
t weath .  
t cc2 .

Πανεπιστήμιο Πειραιώς



t areatype .

t grtl grou.

t owners .

o owner .

o seqacc .

o day92 .

o dacc92 .

o age .

o built1 .

o grtl .

o noface.

Πανεπιστήμιο Πειραιώς

B. Ο αλγόριθμος C5.0, στα δεδομένα του αρχείου σε γλώσσα C

\* C Code File For Rule Induction Model.

\* Filename : TL.c

\* Date : 30/Apr/2002 20:02:34

```
#include "TL.h"
```

```
int TL(
```

```
    struct exTL example,
```

```
    double *confidence )
```

```
{
```

```
    switch (example.VOYAGE) {
```

```
        case VOYAGE__ :
```

```
            *confidence = 0;
```

```
            return(No);
```

```
            break;
```

```
        case VOYAGE_No :
```

```
            *confidence = 0.952375;
```

```
            return(No);
```

```
            break;
```

```
        case VOYAGE_Yes :
```

```
            switch (example.ACCIDE0) {
```

```
                case ACCIDE0_COLLISION_WITH_SHIP_ :
```

```
                    if (example.NOFACC <= 1) {
```

```
                        *confidence = 0.979189;
```

```
                        return(No);
```

```
                    }
```

```
                else {
```

```
                    switch (example.GRT1GROU) {
```

```
                        case GRT1GROU__1 :
```

```
                            if (example.ACCSER <= 1) {
```

```
if (example.GRT1 <= 1547) {
    switch (example.CLASOC1) {
        case CLASOC1_N_A :
            *confidence = 0;
            return(No);
            break;
        case CLASOC1_AB :
            *confidence = 0.666666;
            return(Yes);
            break;
        case CLASOC1_BR :
        case CLASOC1_CRS :
        case CLASOC1_DSRK :
        case CLASOC1_IRS :
        case CLASOC1_JR :
        case CLASOC1_PR :
        case CLASOC1_RCB :
        case CLASOC1_RI :
        case CLASOC1_PR :
        case CLASOC1_RP :
            *confidence = 0;
            return(No);
            break;
        case CLASOC1_BV :
            switch (example.AREATYPE) {
                case AREATYPE__ :
                    *confidence = 0;
                    return(Yes);
                    break;
                case AREATYPE_controlled :
                    *confidence = 0.824271;
                    return(No);
                    break;
                case AREATYPE_open_sea :
```

```
*confidence = 1.0;
return(Yes);
break;
case AREATYPE_port :
    *confidence = 0.587865;
    return(No);
    break;
default :
    *confidence = 0.0;
    return(Yes);
}
break;
case CLASOC1_GL :
    *confidence = 0.793614;
    return(No);
    break;
case CLASOC1_CC :
    *confidence = 1.0;
    return(Yes);
    break;
case CLASOC1_KI :
    *confidence = 1.0;
    return(Yes);
    break;
case CLASOC1_KR :
    *confidence = 0.649041;
    return(Yes);
    break;
case CLASOC1_NK :
    *confidence = 1.0;
    return(No);
    break;
case CLASOC1_NV :
    *confidence = 0.701072;
```



```

return(Yes);
break;
case CLASOC1_RS :
    *confidence = 0.5;
    return(No);
    break;

case CLASOC1_other :
    switch (example.WEATH) {
    case WEATH_CALM_WEATHER :
        *confidence = 0.561242;
        return(No);
        break;
    case WEATH_NOT_REPORTED :
        switch (example.NTYPE2) {
        case NTYPE2_N_A :
            *confidence = 0;
            return(No);
            break;
        case NTYPE2_ORE_CARRIER :
            *confidence = 1.0;
            return(No);
            break;
        case NTYPE2__ :
            case NTYPE2_BARGE_CARRIER :
            case NTYPE2_BULK_CARRIER :
            case NTYPE2_BULK_DRY_CARGO :
            case NTYPE2_BULK_OIL_CARRIER :
            case NTYPE2_CATAMARAN :
            case NTYPE2_CHEM_TANKER :
            case NTYPE2_DECK_CARGO :
            case
NTYPE2_LIQ_NOT_PETR_GAS_TANKER__WATER_MOLASSES_FRUIT_ :
            case NTYPE2_LIQ__NAT__PET_GAS_TANKER :

```

```

case NTYPE2_LIVESTOCK_CARRIER :
case NTYPE2_ORE_OIL_CARRIER :
case NTYPE2_PALLETISED_CARGO :
case NTYPE2_RO_RO_CARGO :
case NTYPE2_UREA_CARRIER :
case NTYPE2_VEHICLES_CARRIER :
    *confidence = 0;
    return(No);
    break;
case NTYPE2_CEMENT_CARRIER :
    *confidence = 1.0;
    return(No);
    break;
case NTYPE2_CHEM_OIL_TANKER :
    *confidence = 1.0;
    return(No);
    break;
case NTYPE2_CONTAINER :
    *confidence = 0.587865;
    return(Yes);
    break;
case NTYPE2_FERRY_PASS_RO_RO_GENER :
    *confidence = 1.0;
    return(No);
    break;
case NTYPE2_GENERAL_CARGO :
    *confidence = 0.603033;
    return(No);
    break;
case NTYPE2_LIQ_PETR_GAS_TANKER_ :
    *confidence = 1.0;
    return(Yes);
    break;
case NTYPE2_LIQ_GAS_CHEM__TANKER :

```

```

*confidence = 1.0;
return(No);
break;
case NTYPE2_OIL__TANKER :
*confidence = 1.0;
return(Yes);
break;
case NTYPE2_OTHER_DRY_CARGO :
if (example.BUILT1 <= 1976) {
*confidence = 1.0;
return(Yes);
}
else {
*confidence = 0.829141;
return(No);
}
break;
case NTYPE2_Other :
if (example.DAY92 <= 1973) {
if (example.GRT1 <= 146) {
*confidence = 1.0;
return(Yes);
}
else {
if (example.DAY92 <= 1336) {
*confidence = 0.83383;
return(No);
}
else {
*confidence = 1.0;
return(Yes);
}
}
}
}
}

```

```

else {
    if (example.DAY92 <= 2708) {
        if (example.BUILT1 <= 1990) {
            *confidence = 0.914086;
            return(No);
        }
        else {
            *confidence = 0.75;
            return(Yes);
        }
    }
    else {
        *confidence = 1.0;
        return(Yes);
    }
}
break;
case NTYPE2_REFRIGERATED :
    *confidence = 0.701072;
    return(Yes);
    break;
case NTYPE2_TANKER :
    *confidence = 0.844392;
    return(No);
    break;
case NTYPE2_WOODCHIPS_CARRIER :
    *confidence = 1.0;
    return(Yes);
    break;
default :
    *confidence = 0.0;
    return(No);
}
break;

```



```
case WEATH_POOR_VISIBILITY :
    *confidence = 0.815715;
    return(Yes);
    break;
case WEATH_SNOWSTORM :
    *confidence = 1.0;
    return(No);
    break;
case WEATH_STORM :
    *confidence = 0.723696;
    return(Yes);
    break;
case WEATH_TYPHOON :
    *confidence = 0;
    return(No);
    break;
default :
    *confidence = 0.0;
    return(No);
}
break;
default :
    *confidence = 0.0;
    return(No);
}
}
else {
    *confidence = 0.802086;
    return(No);
}
}
else {
    *confidence = 0.959516;
    return(No);
}
```

```
}  
break;  
case GRT1GROU__2 :  
    *confidence = 0.766498;  
    return(No);  
    break;  
case GRT1GROU__3 :  
    *confidence = 0.908901;  
    return(No);  
    break;  
case GRT1GROU__4 :  
    *confidence = 1.0;  
    return(No);  
    break;  
case GRT1GROU__5 :  
    *confidence = 1.0;  
    return(No);  
    break;  
case GRT1GROU__6 :  
    *confidence = 1.0;  
    return(No);  
    break;  
case GRT1GROU__7 :  
    *confidence = 1.0;  
    return(No);  
    break;  
case GRT1GROU__9999_00 :  
    *confidence = 0.560163;  
    return(Yes);  
    break;  
default :  
    *confidence = 0.0;  
    return(No);  
}
```

```
}  
break;  
case ACCIDE0_CONTACT_INC_STRIKING_ICE :  
    *confidence = 0.909264;  
    return(No);  
    break;  
case ACCIDE0_FIRE_EXPLOSION :  
    if (example.NOFACC <= 1) {  
        *confidence = 0.889811;  
        return(No);  
    }  
    else {  
        if (example.GRT1 <= 869) {  
            if (example.ACCSER <= 1) {  
                switch (example.NTYPE2) {  
                    case NTYPE2_N_A :  
                        *confidence = 0;  
                        return(Yes);  
                        break;  
                    case NTYPE2_ORE_CARRIER :  
                    case NTYPE2__ :  
                    case NTYPE2_BULK_OIL_CARRIER :  
                    case NTYPE2_CEMENT_CARRIER :  
                    case NTYPE2_DECK_CARGO :  
                    case NTYPE2_LIQ_PETR_GAS_TANKER :  
                    case NTYPE2_LIQ__NAT__PET_GAS_TANKER :  
                    case NTYPE2_LIQ_GAS_CHEM_TANKER :  
                    case NTYPE2_LIVESTOCK_CARRIER :  
                    case NTYPE2_PALLETISED_CARGO :  
                    case NTYPE2_UREA_CARRIER :  
                    case NTYPE2_VEHICLES_CARRIER :  
                    case NTYPE2_WOODCHIPS_CARRIER :  
                        *confidence = 0;  
                        return(Yes);
```

```
break;
case NTYPE2_BARGE_CARRIER :
    *confidence = 1.0;
    return(No);
    break;
case NTYPE2_BULK_CARRIER :
    *confidence = 0.5;
    return(No);
    break;
case NTYPE2_BULK_DRY_CARGO :
    *confidence = 1.0;
    return(Yes);
    break;
case NTYPE2_CATAMARAN :
    *confidence = 1.0;
    return(No);
    break;
case NTYPE2_CHEM_TANKER :
    *confidence = 0.645275;
    return(No);
    break;
case NTYPE2_CHEM_OIL_TANKER :
    *confidence = 1.0;
    return(No);
    break;
case NTYPE2_CONTAINER :
    *confidence = 1.0;
    return(Yes);
    break;
case NTYPE2_FERRY_PASS_RO_RO_GENER :
    *confidence = 0.60008;
    return(Yes);
    break;
case NTYPE2_GENERAL_CARGO :
```



```
*confidence = 0.725725;
return(No);
break;
case
NTYPE2_LIQ_NOT_PETR_GAS_TANKER__WATER_MOLASSES_FRUIT_ :
    *confidence = 1.0;
    return(No);
    break;
case NTYPE2_OIL__TANKER :
    *confidence = 0.574332;
    return(No);
    break;
case NTYPE2_ORE_OIL_CARRIER :
    *confidence = 1.0;
    return(No);
    break;
case NTYPE2_OTHER_DRY_CARGO :
    *confidence = 0.790604;
    return(Yes);
    break;
case NTYPE2_Other :
    *confidence = 0.71825;
    return(Yes);
    break;
case NTYPE2_REFRIGERATED :
    *confidence = 0.810952;
    return(No);
    break;
case NTYPE2_RO_RO_CARGO :
    *confidence = 1.0;
    return(Yes);
    break;
case NTYPE2_TANKER :
    *confidence = 0.846897;
```

```

        return(No);
        break;
    default :
        *confidence = 0.0;
        return(Yes);
    }
}
else {
    *confidence = 0.748068;
    return(No);
}
}
else {
    *confidence = 0.699671;
    return(No);
}
}
break;
case ACCIDE0_FOUNDERED :
    if (example.DAY92 <= 797) {
        *confidence = 0.823617;
        return(Yes);
    }
    else {
        if (example.DAY92 <= 846) {
            if (example.BUILT1 <= 1990) {
                *confidence = 1.0;
                return(No);
            }
        }
        else {
            switch (example.MONTH) {
                case MONTH_N_A :
                    *confidence = 0;
                    return(Yes);

```

```

    break;
case MONTH_Apr :
    *confidence = 1.0;
    return(No);
    break;
case MONTH_Aug :
case MONTH_Dec :
case MONTH_Feb :
case MONTH_Jan :
case MONTH_Jul :
case MONTH_Jun :
case MONTH_May :
case MONTH_Nov :
case MONTH_Oct :
case MONTH_Sep :
    *confidence = 0;
    return(Yes);
    break;
case MONTH_Mar :
    *confidence = 1.0;
    return(Yes);
    break;
default :
    *confidence = 0.0;
    return(Yes);
}
}
}
else {
    switch (example.GRT1GROU) {
    case GRT1GROU__1 :
        switch (example.NTYPE2) {
        case NTYPE2_N_A :
            *confidence = 0;

```

```

return(Yes);
break;
case NTYPE2__ORE_CARRIER :
    *confidence = 0.769866;
    return(No);
    break;
case NTYPE2__ :
case NTYPE2__BULK_OIL_CARRIER :
case
NTYPE2_LIQ_NOT_PETR_GAS_TANKER__WATER_MOLASSES_FRUIT_ :
case NTYPE2_LIQ_NAT_PET_GAS_TANKER :
case NTYPE2_LIQ_GAS_CHEM_TANKER :
case NTYPE2_UREA_CARRIER :
    *confidence = 0;
    return(Yes);
    break;
case NTYPE2_BARGE_CARRIER :
    if (example.DATYR <= 97) {
        *confidence = 0.857143;
        return(No);
    }
    else {
        *confidence = 1.0;
        return(Yes);
    }
    break;
case NTYPE2_BULK_CARRIER :
    switch (example.FLAG2) {
case FLAG2_N_A :
        *confidence = 0;
        return(Yes);
        break;
case FLAG2__ :
case FLAG2_GR_BR :

```



```

case FLAG2_RUSSIA :
    *confidence = 0;
    return(Yes);
    break;
case FLAG2_ELSE :
    *confidence = 0.940155;
    return(Yes);
    break;
case FLAG2_FLAG :
    *confidence = 0.660784;
    return(Yes);
    break;
case FLAG2_GRECEE :
    *confidence = 1.0;
    return(No);
    break;
case FLAG2_USA_CAN_JAP :
    if (example.GRT1 <= 47) {
        *confidence = 1.0;
        return(Yes);
    }
    else {
        *confidence = 1.0;
        return(No);
    }
    break;
default :
    *confidence = 0.0;
    return(Yes);
}
break;
case NTYPE2_BULK_DRY_CARGO :
    switch (example.WEATH) {
    case WEATH_N_A :

```

```

case FLAG2_RUSSIA :
    *confidence = 0;
    return(Yes);
    break;
case FLAG2_ELSE :
    *confidence = 0.940155;
    return(Yes);
    break;
case FLAG2_FLAG :
    *confidence = 0.660784;
    return(Yes);
    break;
case FLAG2_GREECE :
    *confidence = 1.0;
    return(No);
    break;
case FLAG2_USA_CAN_JAP :
    if (example.GRT1 <= 47) {
        *confidence = 1.0;
        return(Yes);
    }
    else {
        *confidence = 1.0;
        return(No);
    }
    break;
default :
    *confidence = 0.0;
    return(Yes);
}
break;
case NTYPE2_BULK_DRY_CARGO :
    switch (example.WEATH) {
        case WEATH_N_A :

```

```

*confidence = 0;
return(Yes);
break;
case WEATH_CALM_WEATHER :
case WEATH_POOR_VISIBILITY :
case WEATH_SNOWSTORM :
case WEATH_TYPHOON :
    *confidence = 0;
    return(Yes);
    break;
case WEATH_NOT_REPORTED :
    if (example.ACCSER <= 1) {
        *confidence = 1.0;
        return(No);
    }
    else {
        *confidence = 1.0;
        return(Yes);
    }
    break;
case WEATH_STORM :
    *confidence = 1.0;
    return(Yes);
    break;
default :
    *confidence = 0.0;
    return(Yes);
}
break;
case NTYPE2_CATAMARAN :
    *confidence = 1.0;
    return(No);
    break;
case NTYPE2_CEMENT_CARRIER :

```

```

*confidence = 0.869971;
return(Yes);
break;
case NTYPE2_CHEM_TANKER :
    switch (example.AREATYPE) {
        case AREATYPE_N_A :
            *confidence = 0;
            return(No);
            break;
        case AREATYPE__ :
        case AREATYPE_controlled :
            *confidence = 0;
            return(No);
            break;
        case AREATYPE_open_sea :
            *confidence = 0.740446;
            return(Yes);
            break;
        case AREATYPE_port :
            *confidence = 0.916915;
            return(No);
            break;
        default :
            *confidence = 0.0;
            return(No);
    }
    break;
case NTYPE2_CHEM_OIL_TANKER :
    *confidence = 0.88933;
    return(No);
    break;
case NTYPE2_CONTAINER :
    *confidence = 0.660687;
    return(Yes);

```



```

break;
case NTYPE2_DECK_CARGO :
    *confidence = 1.0;
    return(Yes);
    break;
case NTYPE2_FERRY_PASS_RO_RO_GENER :
    *confidence = 0.8103;
    return(Yes);
    break;
case NTYPE2_GENERAL_CARGO :
    if (example.DATYR <= 97) {
        switch (example.MONTH) {
            case MONTH_Apr :
                *confidence = 0.793932;
                return(Yes);
                break;
            case MONTH_Aug :
                switch (example.FLAG2) {
                    case FLAG2_N_A :
                        *confidence = 0;
                        return(Yes);
                        break;
                    case FLAG2__ :
                    case FLAG2_GR_BR :
                    case FLAG2_GRECEE :
                    case FLAG2_USA_CAN_JAP :
                        *confidence = 0;
                        return(Yes);
                        break;
                    case FLAG2_ELSE :
                        *confidence = 0.704218;
                        return(No);
                        break;
                    case FLAG2_FLAG :

```

```
*confidence = 0.733691;
return(Yes);
break;
case FLAG2_RUSSIA :
    *confidence = 1.0;
    return(Yes);
    break;
default :
    *confidence = 0.0;
    return(Yes);
}
break;
case MONTH_Dec :
    if (example.NOFAcc <= 1) {
        *confidence = 0.763601;
        return(No);
    }
    else {
        *confidence = 0.817136;
        return(Yes);
    }
    break;
case MONTH_Feb :
    *confidence = 0.793932;
    return(No);
    break;
case MONTH_Jan :
    *confidence = 0.882426;
    return(No);
    break;
case MONTH_Jul :
    if (example.NOFAcc <= 1) {
        *confidence = 0.927148;
        return(Yes);
```

```

}
else {
    *confidence = 0.658281;
    return(No);
}
break;
case MONTH_Jun :
    *confidence = 0.82136;
    return(No);
    break;
case MONTH_Mar :
    *confidence = 0.948499;
    return(Yes);
    break;
case MONTH_May :
    *confidence = 0.638152;
    return(Yes);
    break;
case MONTH_Nov :
    *confidence = 0.817136;
    return(Yes);
    break;
case MONTH_Oct :
    if (example.GRT1 <= 766) {
        *confidence = 1.0;
        return(No);
    }
    else {
        *confidence = 0.895133;
        return(Yes);
    }
    break;
case MONTH_Sep :
    if (example.ACCSER <= 1) {

```

```

        *confidence = 0.666667;
        return(Yes);
    }
    else {
        *confidence = 0.923288;
        return(No);
    }
    break;
default :
    *confidence = 0.0;
    return(Yes);
}
}
else {
    *confidence = 0.762573;
    return(Yes);
}
break;
case NTYPE2_LIQ_PETR_GAS_TANKER_ :
    *confidence = 0.829141;
    return(Yes);
    break;
case NTYPE2_LIVESTOCK_CARRIER :
    *confidence = 0.824271;
    return(Yes);
    break;
case NTYPE2_OIL__TANKER :
    switch (example.AREATYPE) {
    case AREATYPE__ :
        *confidence = 0;
        return(Yes);
        break;
    case AREATYPE_controlled :
        *confidence = 0.774082;

```



```
    return(Yes);
    break;
case AREATYPE_open_sea :
    *confidence = 0.936413;
    return(Yes);
    break;
case AREATYPE_port :
    *confidence = 1.0;
    return(No);
    break;
default :
    *confidence = 0.0;
    return(Yes);
}
break;
case NTYPE2_ORE_OIL_CARRIER :
    *confidence = 1.0;
    return(No);
    break;
case NTYPE2_OTHER_DRY_CARGO :
    *confidence = 0.716404;
    return(Yes);
    break;
case NTYPE2_Other :
    *confidence = 0.748087;
    return(Yes);
    break;
case NTYPE2_PALLETISED_CARGO :
    *confidence = 1.0;
    return(No);
    break;
case NTYPE2_REFRIGERATED :
    *confidence = 0.73369;
    return(Yes);
```

```

    break;
case NTYPE2_RO_RO_CARGO :
    *confidence = 0.638972;
    return(No);
    break;
case NTYPE2_TANKER :
    *confidence = 0.605182;
    return(No);
    break;
case NTYPE2_VEHICLES_CARRIER :
    *confidence = 0.5;
    return(No);
    break;
case NTYPE2_WOODCHIPS_CARRIER :
    *confidence = 0.857143;
    return(Yes);
    break;
default :
    *confidence = 0.0;
    return(Yes);
}
break;
case GRT|GROU__2 :
    switch (example.AREATYPE) {
    case AREATYPE__ :
        *confidence = 0;
        return(No);
        break;
    case AREATYPE_controlled :
        *confidence = 0.602769;
        return(No);
        break;
    case AREATYPE_open_sea :
        *confidence = 0.590754;

```

```

    return(Yes);
    break;
case AREATYPE_port :
    *confidence = 0.534086;
    return(No);
    break;
default :
    *confidence = 0.0;
    return(Yes);
}
break;
case GRT1GROU__3 :
    switch (example.FLAG2) {
    case FLAG2__ :
        *confidence = 0;
        return(No);
        break;
    case FLAG2_ELSE :
        *confidence = 0.5802;
        return(Yes);
        break;
    case FLAG2_FLAG :
        *confidence = 0.820784;
        return(No);
        break;
    case FLAG2_GR_BR :
        *confidence = 1.0;
        return(Yes);
        break;
    case FLAG2_GRECEE :
        *confidence = 1.0;
        return(Yes);
        break;
    case FLAG2_RUSSIA :

```

```

    *confidence = 1.0;
    return(Yes);
    break;
case FLAG2_USA_CAN_JAP :
    *confidence = 0.661921;
    return(No);
    break;
default :
    *confidence = 0.0;
    return(No);
}
break;
case GRT1GROU__4 :
    *confidence = 0.879103;
    return(No);
    break;
case GRT1GROU__5 :
    switch (example.FLAG2) {
    case FLAG2_N_A :
        *confidence = 0;
        return(No);
        break;
    case FLAG2__ :
    case FLAG2_GR_BR :
    case FLAG2_GREECEE :
    case FLAG2_RUSSIA :
        *confidence = 0;
        return(No);
        break;
    case FLAG2_ELSE :
        *confidence = 1.0;
        return(Yes);
        break;
    case FLAG2_FLAG :

```



```

    *confidence = 1.0;
    return(No);
    break;
case FLAG2_USA_CAN_JAP :
    *confidence = 1.0;
    return(No);
    break;
default :
    *confidence = 0.0;
    return(No);
}
break;
case GRT1GROU__6 :
    *confidence = 0.769866;
    return(Yes);
    break;
case GRT1GROU__7 :
    *confidence = 1.0;
    return(No);
    break;
case GRT1GROU__9999_00 :
    *confidence = 0.695328;
    return(Yes);
    break;
default :
    *confidence = 0.0;
    return(Yes);
}
}
break;
case ACCIDE0_GROUNDED:
    *confidence = 0.8702;
    return(No);

```

```

break;
case ACCIDE0_HULL_MCHY_EQUIP_DAMAGED :
    if (example.NOFACC <= 3) {
        *confidence = 0.940233;
        return(No);
    }
else {
    switch (example.NTYPE2) {
    case NTYPE2_N_A :
        *confidence = 0;
        return(No);
        break;
    case NTYPE2__ORE_CARRIER :
    case NTYPE2__ :
    case NTYPE2_BARGE_CARRIER :
    case NTYPE2_BULK_OIL_CARRIER :
    case NTYPE2_CATAMARAN :
    case NTYPE2_CEMENT_CARRIER :
    case NTYPE2_DECK_CARGO :
    case
NTYPE2_LIQ_NOT_PETR_GAS_TANKER_WATER_MOLASSES_FRUIT_ :
    case NTYPE2_LIQ_NAT_PET_GAS_TANKER :
    case NTYPE2_LIQ_GAS_CHEM_TANKER :
    case NTYPE2_ORE_OIL_CARRIER :
    case NTYPE2_PALLETISED_CARGO :
    case NTYPE2_UREA_CARRIER :
        *confidence = 0;
        return(No);
        break;
    case NTYPE2_BULK_CARRIER :
        if (example.BUILT1 <= 1974) {
            *confidence = 1.0;
            return(Yes);
        }

```

```

else {
    *confidence = 0.925268;
    return(No);
}
break;
case NTYPE2_BULK_DRY_CARGO :
    *confidence = 1.0;
    return(No);
    break;
case NTYPE2_CHEM_TANKER :
    *confidence = 1.0;
    return(No);
    break;
case NTYPE2_CHEM_OIL_TANKER :
    *confidence = 1.0;
    return(No);
    break;
case NTYPE2_CONTAINER :
    *confidence = 1.0;
    return(No);
    break;
case NTYPE2_FERRY_PASS_RO_RO_GENER :
    *confidence = 0.90481;
    return(No);
    break;
case NTYPE2_GENERAL_CARGO :
    *confidence = 0.646793;
    return(Yes);
    break;
case NTYPE2_LIQ_PETR_GAS_TANKER_ :
    *confidence = 0.666667;
    return(Yes);
    break;
case NTYPE2_LIVESTOCK_CARRIER :

```

```

*confidence = 1.0;
return(Yes);
break;
case NTYPE2_OIL_TANKER :
*confidence = 0.701072;
return(Yes);
break;
case NTYPE2_OTHER_DRY_CARGO :
switch (example.GRT1GROU) {
case GRT1GROU_N_A :
*confidence = 0;
return(No);
break;
case GRT1GROU_1 :
*confidence = 0.776305;
return(No);
break;
case GRT1GROU_2 :
*confidence = 1.0;
return(Yes);
break;
case GRT1GROU_3 :
case GRT1GROU_4 :
case GRT1GROU_5 :
case GRT1GROU_6 :
case GRT1GROU_7 :
*confidence = 0;
return(No);
break;
case GRT1GROU_9999_00 :
*confidence = 1.0;
return(Yes);
break;
default :

```



```

    *confidence = 0.0;
    return(No);
}
break;
case NTYPE2_Other :
    switch (example.AREATYPE) {
    case AREATYPE__ :
        *confidence = 0;
        return(Yes);
        break;
    case AREATYPE_controlled :
        *confidence = 1.0;
        return(Yes);
        break;
    case AREATYPE_open_sea :
        switch (example.NTYPE1) {
        case NTYPE1_N_A :
            *confidence = 0;
            return(Yes);
            break;
        case NTYPE1__ :
        case NTYPE1_BULK_DRY_CARGO :
        case NTYPE1_CONTAINER :
        case NTYPE1_FERRY_PASS_RO_RO_GENERAL :
        case NTYPE1_OTHER_DRY_CARGO :
        case NTYPE1_PASSENGER :
        case NTYPE1_TANKER :
            *confidence = 0;
            return(Yes);
            break;
        case NTYPE1_FISHING :
            *confidence = 0.906591;
            return(Yes);
            break;

```

```

case NTYPE1_MISCELLANEOUS_ACTIVITES :
    *confidence = 0.567025;
    return(Yes);
    break;
case NTYPE1_OTHER :
    switch (example.WEATH) {
    case WEATH_N_A :
        *confidence = 0;
        return(Yes);
        break;
    case WEATH_CALM_WEATHER :
    case WEATH_POOR_VISIBILITY :
    case WEATH_SNOWSTORM :
    case WEATH_TYPHOON :
        *confidence = 0;
        return(Yes);
        break;
    case WEATH_NOT_REPORTED :
        *confidence = 1.0;
        return(No);
        break;
    case WEATH_STORM :
        *confidence = 1.0;
        return(Yes);
        break;
    default :
        *confidence = 0.0;
        return(Yes);
    }
    break;
case NTYPE1_PONTOON_BARGE :
    *confidence = 1.0;
    return(No);
    break;

```

```

default :
    *confidence = 0.0;
    return(Yes);
}
break;
case AREATYPE_port :
    *confidence = 0.681745;
    return(No);
    break;
default :
    *confidence = 0.0;
    return(Yes);
}
break;
case NTYPE2_REFRIGERATED :
    *confidence = 1.0;
    return(No);
    break;
case NTYPE2_RO_RO_CARGO :
    *confidence = 0.774082;
    return(No);
    break;
case NTYPE2_TANKER :
    *confidence = 0.763951;
    return(No);
    break;
case NTYPE2_VEHICLES_CARRIER :
    *confidence = 1.0;
    return(No);
    break;
case NTYPE2_WOODCHIPS_CARRIER :
    *confidence = 1.0;
    return(Yes);
    break;

```

```

default :
    *confidence = 0.0;
    return(No);
}
}
break;
case ACCIDE0_MISCELLANEOUS :
    *confidence = 0.777641;
    return(No);
    break;
case ACCIDE0_MISSING :
    if (example.NOFACC <= 1) {
        *confidence = 0.819358;
        return(No);
    }
    else {
        if (example.ACCSER <= 1) {
            *confidence = 0.641268;
            return(Yes);
        }
        else {
            if (example.NOFACC <= 2) {
                *confidence = 0.920701;
                return(No);
            }
            else {
                if (example.ACCSER <= 2) {
                    switch (example.NTYPE2) {
                        case NTYPE2_N_A :
                            *confidence = 0;
                            return(Yes);
                            break;
                        case NTYPE2__0RE_CARRIER :
                        case NTYPE2__ :

```



```

case NTYPE2_BARGE_CARRIER :
case NTYPE2_BULK_OIL_CARRIER :
case NTYPE2_CATAMARAN :
case NTYPE2_CEMENT_CARRIER :
case NTYPE2_CHEM_TANKER :
case NTYPE2_CHEM_OIL_TANKER :
case NTYPE2_DECK_CARGO :
case
NTYPE2_LIQ_NOT_PETR_GAS_TANKER__WATER_MOLASSES_FRUIT_ :
case NTYPE2_LIQ_PETR_GAS_TANKER_ :
case NTYPE2_LIQ_NAT_PET_GAS_TANKER :
case NTYPE2_LIQ_GAS_CHEM_TANKER :
case NTYPE2_LIVESTOCK_CARRIER :
case NTYPE2_OIL_TANKER :
case NTYPE2_ORE_OIL_CARRIER :
case NTYPE2_PALLETISED_CARGO :
case NTYPE2_UREA_CARRIER :
case NTYPE2_VEHICLES_CARRIER :
case NTYPE2_WOODCHIPS_CARRIER :
    *confidence = 0;
    return(Yes);
    break;
case NTYPE2_BULK_CARRIER :
    *confidence = 0.565014;
    return(Yes);
    break;
case NTYPE2_BULK_DRY_CARGO :
    *confidence = 0.618204;
    return(Yes);
    break;
case NTYPE2_CONTAINER :
    *confidence = 1.0;
    return(No);
    break;

```

```
case NTYPE2_FERRY_PASS_RO_RO_GENER :
    *confidence = 0.818359;
    return(No);
    break;
case NTYPE2_GENERAL_CARGO :
    if (example.DAY92 <= 2155) {
        *confidence = 0.815715;
        return(No);
    }
    else {
        *confidence = 0.840743;
        return(Yes);
    }
    break;
case NTYPE2_OTHER_DRY_CARGO :
    *confidence = 0.525804;
    return(Yes);
    break;
case NTYPE2_Other :
    *confidence = 0.75;
    return(Yes);
    break;
case NTYPE2_REFRIGERATED :
    *confidence = 1.0;
    return(No);
    break;
case NTYPE2_RO_RO_CARGO :
    *confidence = 1.0;
    return(No);
    break;
case NTYPE2_TANKER :
    *confidence = 0.5;
    return(No);
    break;
```

```

default :
    *confidence = 0.0;
    return(Yes);
}
}
else {
    switch (example.WEATH) {
    case WEATH_CALM_WEATHER :
        *confidence = 0.649464;
        return(No);
        break;
    case WEATH_NOT_REPORTED :
        *confidence = 0.794299;
        return(No);
        break;
    case WEATH_POOR_VISIBILITY :
        *confidence = 0.666667;
        return(No);
        break;
    case WEATH_SNOWSTORM :
        *confidence = 1.0;
        return(Yes);
        break;
    case WEATH_STORM :
        if (example.NOFACC <= 3) {
            *confidence = 0.677753;
            return(No);
        }
        else {
            *confidence = 1.0;
            return(Yes);
        }
        break;
    case WEATH_TYPHOON :

```

```

        *confidence = 0.701072;
        return(Yes);
        break;
    default :
        *confidence = 0.0;
        return(No);
    }
}
}
}
}
}
break;
case ACCIDE0_WAR_LOSS_HOSTILITIES :
    *confidence = 0.848405;
    return(No);
    break;
    default :
        *confidence = 0.0;
        return(No);
    }
    break;
default :
    *confidence = 0.0;
    return(No);
}
*confidence = 0.0;
return(No);
}

```