



ΠΑΝΕΠΙΣΤΗΜΙΟ ΠΕΙΡΑΙΩΣ
ΣΧΟΛΗ ΤΕΧΝΟΛΟΓΙΩΝ ΠΛΗΡΟΦΟΡΙΚΗΣ ΚΑΙ ΕΠΙΚΟΙΝΩΝΙΩΝ
ΤΜΗΜΑ ΨΗΦΙΑΚΩΝ ΣΥΣΤΗΜΑΤΩΝ

ΠΡΟΓΡΑΜΜΑ ΜΕΤΑΠΤΥΧΙΑΚΩΝ ΣΠΟΥΔΩΝ
“ΠΛΗΡΟΦΟΡΙΑΚΑ ΣΥΣΤΗΜΑΤΑ & ΥΠΗΡΕΣΙΕΣ”

Προβλεπτική Συντήρηση Πλοίων με Χρήση Χρονοσειρών

Από

Παρασκευή Σαμπάνη

Υποβάλλεται για την εκπλήρωση των προϋποθέσεων λήψης Μεταπτυχιακού
Διπλώματος στην ειδίκευση «ΠΠΣ»

του ΠΜΣ “Πληροφορικά Συστήματα & Υπηρεσίες”

στο

ΠΑΝΕΠΙΣΤΗΜΙΟ ΠΕΙΡΑΙΩΣ

Σεπτέμβριος 2024

Επιβλέπων: Μιχαήλ Φιλιππάκης
Ακαδημαϊκή Θέση: Καθηγητής
Πανεπιστήμιο Πειραιώς. Κάτοχος όλων των δικαιωμάτων
Συγγραφέας: Παρασκευή Σαμπάνη

ΣΕΛΙΔΑ ΕΓΚΥΡΟΤΗΤΑΣ

Όνοματεπώνυμο Φοιτήτριας: Παρασκευή Σαμπάνη

Τίτλος Μεταπτυχιακής Διπλωματικής Εργασίας: Προβλεπτική Συντήρηση Πλοίων με Χρήση

Χρονοσειρών

Η παρούσα Μεταπτυχιακή Διπλωματική Εργασία υποβάλλεται ως μερική εκπλήρωση των απαιτήσεων του Προγράμματος Μεταπτυχιακών Σπουδών “Πληροφοριακά Συστήματα & Υπηρεσίες” του Τμήματος Ψηφιακών Συστημάτων του Πανεπιστημίου Πειραιώς και εγκρίθηκε στις 31/10/2024 από τα μέλη της Εξεταστικής Επιτροπής.

Εξεταστική Επιτροπή

Επιβλέπων (Τμήμα Ψηφιακών Συστημάτων, Πανεπιστήμιο Πειραιώς): Μιχαήλ Φιλιππάκης

Μέλος Εξεταστικής Επιτροπής: Μαρία Χαλκίδη, Καθηγήτρια

Μέλος Εξεταστικής Επιτροπής: Δημοσθένης Κυριαζής, Καθηγητής

ΥΠΕΥΘΥΝΗ ΔΗΛΩΣΗ ΑΥΘΕΝΤΙΚΟΤΗΤΑΣ

Η Σαμπάνη Παρασκευή, γνωρίζοντας τις συνέπειες της λογοκλοπής, δηλώνω υπεύθυνα ότι η παρούσα εργασία με τίτλο «Προβλεπτική Συντήρηση Πλοίων με Χρήση Χρονοσειρών», αποτελεί προϊόν αυστηρά προσωπικής εργασίας και όλες οι πηγές που έχω χρησιμοποιήσει, έχουν δηλωθεί κατάλληλα στις βιβλιογραφικές παραπομπές και αναφορές. Τα σημεία όπου έχω χρησιμοποιήσει ιδέες, κείμενο ή/και πηγές άλλων συγγραφέων, αναφέρονται ευδιάκριτα στο κείμενο με την κατάλληλη παραπομπή και η σχετική αναφορά περιλαμβάνεται στο τμήμα των βιβλιογραφικών αναφορών με πλήρη περιγραφή.

Επιπλέον δηλώνω υπεύθυνα ότι η συγκεκριμένη Μεταπτυχιακή Διπλωματική Εργασία έχει συγγραφεί από εμένα προσωπικά και δεν έχει υποβληθεί ούτε έχει αξιολογηθεί στο πλαίσιο κάποιου άλλου μεταπτυχιακού ή προπτυχιακού τίτλου σπουδών, στην Ελλάδα ή στο εξωτερικό.

Παράβαση της ανωτέρω ακαδημαϊκής μου ευθύνης αποτελεί ουσιώδη λόγο για την ανάκληση του πτυχίου μου. Σε κάθε περίπτωση, αναληθούς ή ανακριβούς δηλώσεως, υπόκειμαι στις συνέπειες που προβλέπονται τις διατάξεις που προβλέπει η Ελληνική και Κοινωνική Νομοθεσία περί πνευματικής ιδιοκτησίας.

Η ΔΗΛΟΥΣΑ

Όνοματεπώνυμο: Παρασκευή Σαμπάνη

Αριθμός Μητρώου: me2250

Υπογραφή:

Ευχαριστίες

Η παρούσα μεταπτυχιακή εργασία εκπονήθηκε κατά τη διάρκεια του τελευταίου χρόνου για την απόκτηση του μεταπτυχιακού με τίτλο «Πληροφοριακά Συστήματα και Υπηρεσίες» στην κατεύθυνση Προηγμένα Πληροφοριακά Συστήματα που διοργανώνει το Πανεπιστήμιο Πειραιώς, υπό την επίβλεψη του καθηγητή και διευθυντού του ΠΜΣ Μιχαήλ Φιλιππάκη. Ολοκληρώνοντας την προσπάθειά μου σε αυτό το πρόγραμμα, θα ήθελα να εκφράσω τις ευχαριστίες μου σε μερικούς ανθρώπους που πέρασαν μαζί μου τα τελευταία δύο χρόνια και η στήριξή τους ήταν αρκετά σημαντική.

Αρχικά θα ήθελα να ευχαριστήσω τον επιβλέποντα καθηγητή Κο Μιχαήλ Φιλιππάκη που μου εμπιστεύτηκε το θέμα της παρούσας διατριβής. Μου έδωσε όλη τη βοήθεια και τους πόρους που χρειαζόμουν για να την ολοκληρώσω. Τον ευχαριστώ επίσης και για τον πολύτιμο χρόνο που διέθεσε για να με βοηθήσει σε όποια προβλήματα ή ερωτήσεις είχα πάνω στο θέμα. Επιπλέον, θα ήθελα να ευχαριστήσω όλους τους καθηγητές και τους συμφοιτητές μου σε αυτό το πρόγραμμα. Ήταν δύο σπουδαία χρόνια και έμαθα πολλά από εκείνους.

Είμαι επίσης ευγνώμων στην εταιρεία στην οποία εργάζομαι και στην ναυτιλιακή εταιρεία που μου εμπιστεύτηκαν το σύνολο δεδομένων πάνω στο οποίο βασίστηκε όλη η διπλωματική μου διατριβή.

Θα ήθελα επίσης να ευχαριστήσω την οικογένειά μου, τους γονείς μου Στέλιο και Άννα και τον αδερφό μου Αντώνη, οι οποίοι ήταν πάντα δίπλα μου και με βοήθησαν σε όλες τις σπουδές μου από την αρχή μέχρι σήμερα.

Περιεχόμενα

ΣΕΛΙΔΑ ΕΓΚΥΡΟΤΗΤΑΣ	i
Ευχαριστίες	ii
Περιεχόμενα.....	iii
Περίληψη	v
Abstract.....	vi
Λίστα Εικόνων	vii
Λίστα Πινάκων.....	viii
1. Εισαγωγή.....	1
1.1 Μηχανήματα Πλοίου.....	1
1.2 Σύστημα Προγραμματισμένης Συντήρησης (Planned Maintenance System)	2
1.3 Συνιστώσες ενός Συστήματος Προγραμματισμένης Συντήρησης.....	2
1.4 Είδη Συντήρησης	3
1.4.1 Διορθωτική Συντήρηση	3
1.4.2 Προληπτική Συντήρηση.....	4
1.4.3 Προκαθορισμένη Συντήρηση	4
1.4.4 Συντήρηση με βάση την κατάσταση του μηχανήματος.....	5
1.4.5 Προγνωστική Συντήρηση	6
2. Χρονοσειρές.....	8
2.1 Ορισμός Χρονοσειράς.....	8
2.2 Συνιστώσες Χρονοσειρών	8
2.3 Χαρακτηριστικά Χρονοσειρών	10
2.3.1 Στασιμότητα	10
2.3.2 Αιτιοκρατία και Στοχαστικότητα.....	10
2.3.3 Γραμμικότητα και Μη-Γραμμικότητα	10
3. Γραμμικά Μοντέλα Πρόβλεψης	12
3.1 Autoregressive (AR) Model	12
3.2 Moving Average (MA) Model.....	13
3.3 Autoregressive Moving Average (ARMA) Model.....	14
3.4 AutoRegressive Integrated Moving Average (ARIMA) Model.....	14
3.4.1 Συνάρτηση auto_arima	15

3.5	Seasonal ARIMA (SARIMA) Model	17
4	Κριτήρια Επιλογής Τάξης Μοντέλου ARIMA	19
4.1	Akaike Information Criterion (AIC)	19
4.2	Bayesian Information Criterion (BIC).....	19
5.	Περιγραφή Δεδομένων	21
5.1	Περιγραφή Συνόλου Δεδομένων	21
5.2	Δημιουργία Συνόλου Δεδομένων.....	23
6.	Εφαρμογή αλγοριθμικών μεθόδων – Αποτελέσματα.....	25
6.1	Πρόβλεψη βλαβών	25
6.1.1	Εφαρμογή ARIMA(2,1,2)	31
6.1.2	Εφαρμογή ARIMA(1,1,1)	37
6.1.3	Εφαρμογή ARIMA(0,1,1)	44
6.1.4	Αποτελέσματα Συνάρτησης auto_arima.....	50
7.	Σύγκριση Μοντέλων – Συμπεράσματα	55
	Βιβλιογραφία.....	57
	Παράρτημα Κώδικα	59

Περίληψη

Η συγκεκριμένη διπλωματική εργασία ασχολείται με την ανάλυση χρονοσειρών σε δεδομένα ναυτιλίας. Τα δεδομένα αυτά συλλέχτηκαν από ένα σύστημα προγραμματισμένης συντήρησης μιας ναυτιλιακής εταιρείας και αφορούν τον αριθμό των απρόσμενων βλαβών που μπορεί να προκύψουν κάθε μήνα σε ένα μηχανήμα που βρίσκεται εγκατεστημένο σε όλα τα πλοία της εταιρείας. Το μηχανήμα βρίσκεται στην κύρια μηχανή του πλοίου και το μοντέλο έχει τον κωδικό CE-DE, 12V50DF. Στο παρόν κείμενο, αρχικά εξηγείται τι είναι ένα Σύστημα Προγραμματισμένης Συντήρησης και ο λόγος που είναι υποχρεωμένες όλες οι ναυτιλιακές εταιρίες να το χρησιμοποιούν. Ένας από τους λόγους χρήσης αυτού του συστήματος είναι η συντήρηση των μηχανών του πλοίου. Συνεπώς στο κείμενο γίνεται αναφορά στα τέσσερα είδη συντήρησης που υπάρχουν, τα οποία είναι η Διορθωτική Συντήρηση, η Προληπτική Συντήρηση, η Προκαθορισμένη Συντήρηση και η Συντήρηση με βάση την κατάσταση του μηχανήματος. Για τις ανάγκες της εργασίας και την πρόβλεψη των βλαβών στις μηχανές των πλοίων χρησιμοποιήθηκε ένα σύνολο δεδομένων χρονοσειράς. Αναφέρεται λοιπόν, ο ορισμός ενός συνόλου δεδομένων χρονοσειράς, εξηγείται η μορφή του και αναλύονται τα χαρακτηριστικά του. Η χρονοσειρά η οποία χρησιμοποιείται στην παρούσα εργασία είναι γραμμική χρονοσειρά, επομένως παρουσιάζονται τα γραμμικά μοντέλα τα οποία εφαρμόζονται για πρόβλεψη σε γραμμικές χρονοσειρές.

Σκοπός αυτής της εργασίας είναι η πρόβλεψη του αριθμού των βλαβών που μπορεί να προκύψουν τους επόμενους μήνες στο συγκεκριμένο μηχανήμα. Για το σκοπό αυτό έγινε χρήση μεθόδων ανάλυσης χρονοσειρών και δοκιμάστηκαν τρία μοντέλα ARIMA με διαφορετικές παραμέτρους. Ύστερα από κάθε εφαρμογή ενός από τα μοντέλα ARIMA γινόταν η πρόβλεψη των βλαβών. Έπειτα, σύμφωνα με τα δύο κριτήρια πληροφορίας, τα AIC και BIC, και τους πίνακες προβλέψεων του κάθε μοντέλου, γίνεται η σύγκριση των μοντέλων ARIMA, ώστε να φανεί ποιο από τα τρία μοντέλα ARIMA που χρησιμοποιήθηκαν αποδίδει αποτελεσματικότερα στο σύνολο δεδομένων της χρονοσειράς. Το έγγραφο ολοκληρώνει την αναφορά παρουσιάζοντας το βέλτιστο ARIMA μοντέλο για τη συγκεκριμένη χρονοσειρά που χρησιμοποιήθηκε.

Abstract

This thesis deals with the analysis of time series in shipping data. This data was collected from a scheduled maintenance system of a shipping company and concerns the number of unscheduled failures that can occur each month on a machine installed on all the company's ships. The machine's model name is CE-DE, 12V50DF. First, it is explained what a Planned Maintenance System is and why it is mandatory for all shipping companies to use it. Then, the four types of maintenance are being presented, which are the Corrective Maintenance, Preventive Maintenance, Predictive Maintenance and Condition Based Maintenance. It is explained what a time series dataset is and its characteristics are analysed. The time series which is used in this paper is a linear time series, therefore the linear models which are applied for prediction in linear time series are presented.

The purpose of this thesis is to predict the number of failures that may occur in the upcoming months on this machine. For this purpose, time series analysis methods were used and three ARIMA models with different parameters were tested. After each application of one of the ARIMA models, a prediction of the number of failures was made. Then according to the two information criteria, AIC and BIC, and the prediction tables of each model, the ARIMA models are compared to show which of the three models that were used, performs more effectively on the time series data set. The paper concludes the report by presenting the optimal ARIMA model for the specific time series used.

Λίστα Εικόνων

Εικόνα 1 - Τύποι Συντήρησης	3
Εικόνα 2 - Αναπαράσταση της έννοιας Conditioned Based Maintenance	6
Εικόνα 3- Συνιστώσες Χρονοσειρών Πηγή https://www.analyticsvidhya.com/blog/2021/10/a-comprehensive-guide-to-time-series-analysis/	9
Εικόνα 4 - Χρονοδιάγραμμα βλαβών της μηχανής CE-DE, 12V50DF	25
Εικόνα 5 - Γραφική Παράσταση του Συντελεστή Αυτοσυσχέτισης.....	26
Εικόνα 6 - Γραφική Παράσταση του συντελεστή Μερικής Αυτοσυσχέτισης	27
Εικόνα 7 - Το χρονοδιάγραμμα της σειράς μετά την πρώτη διαφορά.....	28
Εικόνα 8 - Γραφική Παράσταση του Συντελεστή Αυτοσυσχέτισης μετά τις πρώτες διαφορές	29
Εικόνα 9 - Γραφική Παράσταση του Συντελεστή Μερικής Αυτοσυσχέτισης μετά τις πρώτες διαφορές	29
Εικόνα 10 - Σύνολο δεδομένων εκπαίδευσης (train dataset).....	30
Εικόνα 11 - Αποτελέσματα Μοντέλου ARIMA(2,1,2)	31
Εικόνα 12 - Γράφημα προβλέψεων του μοντέλου ARIMA(2,1,2)	34
Εικόνα 13 - Γράφημα των Καταλοίπων και της Πυκνότητας του ARIMA(2,1,2)	35
Εικόνα 14 - Διαγράμματα ACF και PACF των Καταλοίπων για το ARIMA(2,1,2).....	36
Εικόνα 15 - Σύνολο δεδομένων εκπαίδευσης για ARIMA(1,1,1)	37
Εικόνα 16 - Αποτελέσματα μοντέλου ARIMA(1,1,1)	38
Εικόνα 17 - Διάγραμμα Προβλέψεων του ARIMA(1,1,1).....	41
Εικόνα 18 - Διαγράμματα Καταλοίπων και Πυκνότητας του ARIMA(1,1,1)	41
Εικόνα 19 - Διαγράμματα ACF και PACF του ARIMA(1,1,1)	43
Εικόνα 20 - Αποτελέσματα του μοντέλου ARIMA(0,1,1).....	44
Εικόνα 21 - Διάγραμμα Προβλέψεων του ARIMA(0,1,1).....	47
Εικόνα 22 - Διαγράμματα Καταλοίπων και Πυκνότητας του ARIMA(0,1,1)	48
Εικόνα 23 - Διαγράμματα ACF και PACF των Καταλοίπων του ARIMA(0,1,1)	49
Εικόνα 24 - Αποτελέσματα Συνάρτησης <code>auto_arima</code>	52
Εικόνα 25 - Αποτελέσματα του μοντέλου ARIMA(0,1,1).....	54

Λίστα Πινάκων

Πίνακας 1 - Ονομασίες μοντέλων ARIMA.....	15
Πίνακας 2 - Πίνακας τιμών ACF του συνόλου δεδομένων	26
Πίνακας 3 - Πίνακας τιμών PACF	27
Πίνακας 4 - Πίνακας τιμών ACF μετά την 1 ^η διαφορά.....	29
Πίνακας 5 - Πίνακας τιμών PACF μετά την 1 ^η διαφορά.....	29
Πίνακας 6 - Τιμές Πρόβλεψης μοντέλου ARIMA(2,1,2).....	33
Πίνακας 7 - Τιμές Πρόβλεψης μοντέλου ARIMA(1,1,1).....	40
Πίνακας 8 - Τιμές Πρόβλεψης μοντέλου ARIMA(0,1,1).....	46

1. Εισαγωγή

Στον τομέα της ναυτιλίας ένα από τα σημαντικότερα πράγματα που πρέπει να λαμβάνονται υπόψιν είναι η διασφάλιση της αποτελεσματικής και ασφαλούς λειτουργίας των πλοίων. Για να εξασφαλιστεί αυτό, κάθε ναυτιλιακή εταιρία επιβάλλεται να χρησιμοποιεί ένα σύστημα προγραμματισμένης συντήρησης. Το σύστημα προγραμματισμένης συντήρησης είναι ένα πρόγραμμα που πρέπει να είναι εγκατεστημένο στους υπολογιστές και του γραφείου και των πλοίων της εταιρίας, το οποίο έχει σχεδιαστεί για τον προγραμματισμό, τη διαχείριση και την τεκμηρίωση των δραστηριοτήτων συντήρησης στα πλοία. Σκοπός του είναι η ενίσχυση της ασφάλειας, της αξιοπιστίας, και της συμμόρφωσης με τα διεθνή πρότυπα. Στις παρακάτω παραγράφους θα αναλυθούν τα βασικά στοιχεία ενός συστήματος προγραμματισμένης συντήρησης, θα αναφερθούν τα είδη συντήρησης που υπάρχουν και θα εξηγηθούν ορισμένα από τα μηχανήματα ενός πλοίου, τα οποία περιέχονται στη βάση δεδομένων ενός συστήματος συντήρησης.

1.1 Μηχανήματα Πλοίου

Τα μηχανήματα πλοίου είναι το σύνολο του εξοπλισμού ενός πλοίου που το βοηθούν να λειτουργεί ομαλά και με ασφάλεια.

Το πιο σημαντικό μέρος της μηχανής ενός πλοίου ονομάζεται σύστημα πρόωσης. Αυτό το σύστημα έχει κινητήρες, άξονες και έλικες. Ο κύριος κινητήρας, μετατρέπει το καύσιμο σε μηχανική ενέργεια, η οποία κάνει τον άξονα της έλικας να περιστρέφεται, δημιουργώντας έτσι ώθηση προς τα εμπρός [11] [12]. Η αποδοτικότητα και η αξιοπιστία του συστήματος πρόωσης είναι πολύ σημαντικά για την απόδοση του πλοίου, καθώς καθορίζει την ταχύτητα, την κατανάλωση καυσίμου και το συνολικό λειτουργικό κόστος.

Εκτός από το σύστημα πρόωσης, τα μηχανήματα του πλοίου περιλαμβάνουν και συστήματα παραγωγής ενέργειας. Τα συστήματα αυτά περιλαμβάνουν τις γεννήτριες που παράγουν ηλεκτρική ενέργεια για διάφορες λειτουργίες του πλοίου, όπως για παράδειγμα, για το φωτισμό των εσωτερικών χώρων του πλοίου και του καταστρώματος, για τη θέρμανση και για τη λειτουργία διάφορων ηλεκτρονικών συστημάτων και συσκευών επικοινωνίας.

Ένα ακόμη πολύ σημαντικό στοιχείο των μηχανών του πλοίου είναι το σύστημα πλοήγησης. Αυτό περιλαμβάνει ραντάρ, γυροσκόπια, συσκευές GPS και άλλα βοηθήματα που χρησιμεύουν στη χάραξη της πορείας του πλοίου. Τα συστήματα αυτά είναι απαραίτητα για την αποφυγή συγκρούσεων και την πλοήγηση σε δυσμενείς καιρικές συνθήκες όπου δεν υπάρχει ορατότητα.

Στα μηχανήματα ενός πλοίου συγκαταλέγονται επίσης και τα συστήματα έρματος [13] για την ευστάθεια, οι αντλίες υδροσυλλεκτών για την απομάκρυνση της περίσσειας νερού, τα συστήματα HVAC (θέρμανση, εξαερισμός και κλιματισμός) [14] για τη διατήρηση άνετων συνθηκών διαβίωσης μέσα στο πλοίο και ο εξοπλισμός πυρόσβεσης για την ασφάλεια. Τα παραπάνω συστήματα χρησιμοποιούνται για τη λειτουργική απόδοση του πλοίου και για την ασφάλεια του πληρώματος και του φορτίου του.

Η συντήρηση των μηχανημάτων ενός πλοίου είναι υψίστης σημασίας. Τα μηχανήματα πρέπει να ελέγχονται τακτικά και να διορθώνονται τυχόν προβλήματα ώστε να μη χαλάσουν. Ορισμένα πλοία περιλαμβάνουν επίσης αυτοματοποιημένα συστήματα τα οποία παρακολουθούν και αξιολογούν συνεχώς την απόδοση των μηχανημάτων και ειδοποιούν το πλήρωμα για πιθανά προβλήματα, επιτρέποντας την προληπτική συντήρηση.

Συνοψίζοντας, τα μηχανήματα των πλοίων αποτελούνται από πολλά διαφορετικά μέρη που βοηθούν ένα πλοίο να λειτουργεί σωστά και να παραμένει ασφαλές. Αυτά τα μέρη περιλαμβάνουν συστήματα όπως μηχανές, συστήματα πλοήγησης και εξοπλισμό που βοηθούν στην ομαλή λειτουργία του πλοίου.

1.2 Σύστημα Προγραμματισμένης Συντήρησης (Planned Maintenance System)

Σύμφωνα με τον Κώδικα Ασφαλούς Συντήρησης στη ναυτιλία, οι ναυτιλιακές εταιρίες πρέπει να διασφαλίσουν ότι τα μηχανήματα και ο εξοπλισμός του στόλου πρέπει να συντηρούνται και να λειτουργούν σύμφωνα με τους ισχύοντες κανονισμούς, απαιτήσεις, διαδικασίες και πρότυπα που έχουν θεσπιστεί από την εταιρία. Οι απαιτήσεις αυτές ισχύουν τόσο για τη συντήρηση του πλοίου, των μηχανημάτων, τον σωστικό και πυροσβεστικό μηχανισμό όσο και για ολόκληρο το μηχανοστάσιο του πλοίου. Για να επιτευχθεί ο σωστός προγραμματισμός και έλεγχος όλων των συντηρήσεων, χρειάζεται η κάθε ναυτιλιακή εταιρία να έχει εγκατεστημένο σε όλο το στόλο ένα σύστημα προγραμματισμένης συντήρησης.

Το σύστημα προγραμματισμένης συντήρησης είναι ένα σύστημα το οποίο εξυπηρετεί στον προγραμματισμό εργασιών συντήρησης στα μηχανήματα ενός πλοίου. Το σύστημα μπορεί να είναι εξ ολοκλήρου ηλεκτρονικό, εξ ολοκλήρου έντυπο ή συνδυασμός και των δύο. Στις μέρες μας, χρησιμοποιούνται κυρίως ηλεκτρονικά συστήματα συντήρησης για την καλύτερη και ταχύτερη επικοινωνία της εταιρίας με τα πλοία και για ευκολότερο προγραμματισμό των εργασιών ενός πλοίου. Ο προγραμματισμός των εργασιών και κατά συνέπεια η συντήρηση των μηχανημάτων γίνεται σύμφωνα με τις απαιτήσεις του κατασκευαστή για το εκάστοτε μηχανήμα. Σκοπός της προγραμματισμένης συντήρησης είναι να εξασφαλιστεί η ασφαλής και η αξιόπιστη λειτουργία του πλοίου και του εξοπλισμού του.

Μια συστηματική προσέγγιση της συντήρησης ενός πλοίου βασίζεται στην εκτίμηση των κινδύνων και ξεκινά με τη δημιουργία μιας πλήρους βάσης δεδομένων για τα μηχανήματα, τον εξοπλισμό και τα εξαρτήματα του πλοίου. Ένα αποτελεσματικό σύστημα προγραμματισμένης συντήρησης συμβάλει σημαντικά στη συντήρηση των πλοίων μιας ναυτιλιακής εταιρείας καθώς και στην επίτευξη περιβαλλοντικών στόχων μέσω των εργασιών που καταγράφονται στο σύστημα.

1.3 Συνιστώσες ενός Συστήματος Προγραμματισμένης Συντήρησης

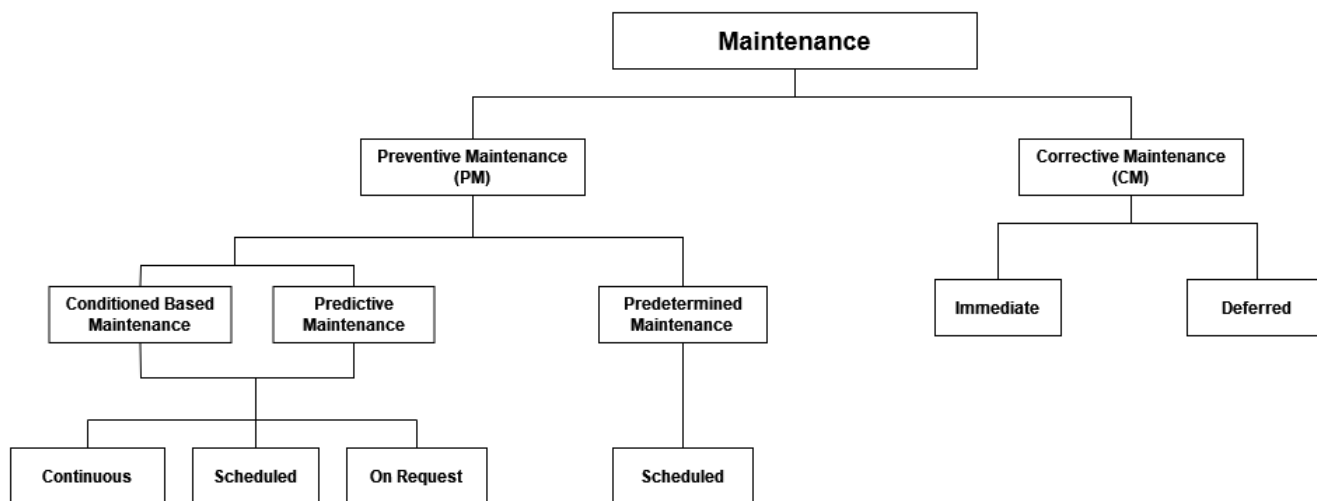
Ένα σύστημα προγραμματισμένης συντήρησης περιλαμβάνει τον προγραμματισμό συντήρησης, τις εντολές εργασίας που πρέπει να γίνουν στο πλοίο, τη διαχείριση των αποθεμάτων και την καταγραφή όλων των δραστηριοτήτων. Αναλυτικότερα, ο προγραμματισμός συντήρησης είναι ένα χρονοδιάγραμμα το οποίο περιγράφει λεπτομερώς ποιες εργασίες συντήρησης πρέπει να γίνουν και πότε. Το χρονοδιάγραμμα αυτό περιλαμβάνει εργασίες που εκτελούνται καθημερινά, εβδομαδιαία, μηνιαία και ετήσια. Επίσης, διασφαλίζει ότι όλος ο εξοπλισμός και τα συστήματα του πλοίου συντηρούνται τακτικά και σύμφωνα με τις συγκεκριμένες απαιτήσεις τους το καθένα. Οι εντολές εργασίας παρέχουν συγκεκριμένες οδηγίες και διαδικασίες για κάθε εργασία συντήρησης. Οι εντολές αυτές έχουν δωθεί είτε από τον κατασκευαστή του εκάστοτε μηχανήματος, είτε από το τεχνικό τμήμα της ναυτιλικής εταιρίας. Σε κάθε καταχώρηση μίας εργασίας στο σύστημα, καταχωρείται και ο τρόπος με τον οποίο πρέπει να γίνει αυτή. Η αποτελεσματική διαχείριση των αποθεμάτων διασφαλίζει ότι τα ανταλλακτικά και τα απαραίτητα εργαλεία είναι διαθέσιμα όταν χρειάζονται. Το σύστημα προγραμματισμένης συντήρησης παρακολουθεί τα επίπεδα αποθεμάτων και μέσω αυτού μπορούν να πραγματοποιηθούν παραγγελίες όταν τα αποθέματα είναι λιγοστά, αποτρέποντας έτσι τις καθυστερήσεις στις εργασίες συντήρησης λόγω μη διαθέσιμων ανταλλακτικών. Επιπλέον, αρκετά σημαντική είναι η πλήρης καταγραφή όλων των

δραστηριοτήτων συντήρησης, η οποία περιλαμβάνει καταγραφές των ημερομηνιών, των διαδικασιών που ακολουθήθηκαν, των ανταλλακτικών που χρησιμοποιήθηκαν και του προσωπικού που συμμετείχε. Η λεπτομερής καταγραφή όλων των παραπάνω στοιχείων είναι ιδιαίτερα χρήσιμη για τους ελέγχους και τις επιθεωρήσεις, διασφαλίζοντας ότι το πλοίο συμμορφώνεται με τους κανόνες της ναυτιλίας.

1.4 Είδη Συντήρησης

Συντήρηση είναι η σειρά διαδικασιών και εργασιών που αποσκοπούν στη μείωση εμφάνισης βλαβών του εξοπλισμού στη ναυτιλιακή βιομηχανία καθώς και σε άλλες βιομηχανίες. Οι κύριοι στόχοι της συντήρησης είναι η αύξηση της ασφάλειας, η μείωση της συχνότητας των απρογραμμάτιστων δυσλειτουργιών που μπορεί να προκύψουν στον εξοπλισμό ενός οργανισμού και τελικά η αύξηση της αξίας του οργανισμού.

Υπάρχουν δύο κύριες κατηγορίες συντήρησης, η διορθωτική (CM) και η προληπτική (PvM). Η κύρια διαφορά τους είναι ότι στην προληπτική συντήρηση, πρωταρχικός στόχος είναι η πρόληψη των βλαβών. Για να επιτευχθεί αυτό, οι εργασίες συντήρησης γίνονται σε προκαθορισμένα χρονικά διαστήματα, ώστε να διασφαλιστεί ότι δε θα προκύψει κάποια βλάβη. Αντίθετα, η διορθωτική προσέγγιση επικεντρώνεται στη διόρθωση ενός εξαρτήματος όταν πάθει βλάβη [2].



Εικόνα 1 - Τύποι Συντήρησης

Πηγή: <https://doi.org/10.1016/j.joes.2020.03.003>

1.4.1 Διορθωτική Συντήρηση

Η διορθωτική συντήρηση, επίσης γνωστή ως «run-to-failure (RTF)» και «συντήρηση λόγω βλάβης» αναφέρεται στην πρακτική της αντιμετώπισης των βλαβών του εξοπλισμού μόνο όταν αυτές συμβαίνουν. Η βασική φιλοσοφία είναι να «το διορθώνουμε όταν χαλάει», δηλαδή οι ενέργειες συντήρησης ξεκινούν μετά τη βλάβη. Σε γενικές γραμμές, η διορθωτική συντήρηση δεν θεωρείται αποτελεσματική για τις επιχειρήσεις, καθώς αναγκάζει τις εταιρείες να δρουν μόνο όταν υπάρχει πρόβλημα, αντί να επιτρέπει την προληπτική πρόβλεψη και τον προγραμματισμό. Ωστόσο, η στρατηγική αυτή δεν είναι εντελώς αδικαιολόγητη. Η διορθωτική συντήρηση μπορεί να είναι βιώσιμη σε συγκεκριμένες περιπτώσεις, όπως όταν ο εν λόγω εξοπλισμός είναι χαμηλής

κρισιμότητας ή όταν η βλάβη του δεν επηρεάζει τη βασική λειτουργία του κύριου συστήματος. Με άλλα λόγια, η διορθωτική συντήρηση είναι αποδεκτή εάν η βλάβη του εξοπλισμού δεν διαταράσσει την κανονική λειτουργία ή δεν επηρεάζει τους γενικότερους στόχους του οργανισμού [2].

Γενικά, οι ναυτιλιακές εταιρίες αποφεύγουν να βασίζονται αποκλειστικά στη διορθωτική συντήρηση, καθώς δεν θεωρείται αποτελεσματική στρατηγική. Η προσέγγιση αυτή απαιτεί δραστικά μέτρα σε απρόσμενες βλάβες, αντί της χρήσης τακτικής ανάλυσης και παρακολούθησης του συστήματος. Ωστόσο, ένα πλεονέκτημα της διορθωτικής συντήρησης είναι ότι τα εξαρτήματα των μηχανημάτων χρησιμοποιούνται μέχρι να φτάσουν στο τέλος της διάρκειας ζωής τους, μεγιστοποιώντας τη χρησιμότητά τους. Παρά τους περιορισμούς της, η διορθωτική συντήρηση μπορεί να προτιμάται σε ορισμένες περιπτώσεις, όπως οι παρακάτω:

- Όταν η βλάβη του εξοπλισμού δεν επηρεάζει κρίσιμα τις κανονικές λειτουργίες.
- Όταν το κόστος της προγραμματισμένης συντήρησης υπερβαίνει την αξία του ίδιου του εξοπλισμού.
- Όταν οι περιοδικές επιθεωρήσεις ενός περιουσιακού στοιχείου δεν παρέχουν σημαντική αξία στον οργανισμό.

1.4.2 Προληπτική Συντήρηση

Η προληπτική συντήρηση ορίζεται ως η συντήρηση που πραγματοποιείται σε προκαθορισμένα διαστήματα ή σύμφωνα με καθορισμένα κριτήρια και αποσκοπεί στη μείωση της πιθανότητας βλάβης. Η φιλοσοφία της προληπτικής συντήρησης είναι να «το φτιάχνουμε πριν χαλάσει», δηλαδή η συντήρηση πραγματοποιείται πριν ο εξοπλισμός φτάσει σε κατάσταση βλάβης. Η συντήρηση γίνεται από εξειδικευμένο προσωπικό για τη διατήρηση του εξοπλισμού και των εγκαταστάσεων σε ικανοποιητική κατάσταση λειτουργίας με τη συστηματική επιθεώρηση, ανίχνευση και διόρθωση αρχικών βλαβών είτε πριν εμφανιστούν είτε πριν εξελιχθούν σε σημαντικές βλάβες. Η συντήρηση, περιλαμβάνει δοκιμές, μετρήσεις, ρυθμίσεις και αντικατάσταση εξαρτημάτων, που εκτελούνται ειδικά για την πρόληψη της εμφάνισης βλαβών. Η προληπτική συντήρηση είναι μια περιοδική επιθεώρηση προκειμένου να παρατηρείται οτιδήποτε ασυνήθιστο το οποίο να μπορεί να προκαλέσει μελλοντική βλάβη στον εξοπλισμό του πλοίου. Στόχος είναι η πρόληψη της βλάβης του εξοπλισμού, η ελαχιστοποίηση της πιθανότητας της μη λειτουργίας του, και η διατήρηση υψηλού επιπέδου αξιοπιστίας με την εκτέλεση τακτικών εργασιών συντήρησης. Η βασική παραδοχή είναι ότι όλος ο εξοπλισμός και τα μηχανήματα θα παρουσιάσουν βλάβες και προβλήματα με την πάροδο του χρόνου λόγω της χρήσης. Η προληπτική συντήρηση κατηγοριοποιείται σε τρεις τύπους, την Προκαθορισμένη Συντήρηση, την Συντήρηση με βάση την κατάσταση του μηχανήματος και την Προγνωστική Συντήρηση [2].

1.4.3 Προκαθορισμένη Συντήρηση

Η προκαθορισμένη συντήρηση είναι ένα είδος προληπτικής συντήρησης. Εκτελείται σε συγκεκριμένα χρονικά διαστήματα με βάση ιστορικά δεδομένα και την εκτιμώμενη διάρκεια ζωής των εξαρτημάτων. Η προσέγγιση αυτή βασίζεται σε πληροφορίες από παρόμοιο ή πανομοιότυπο εξοπλισμό, χωρίς να λαμβάνεται υπόψη η πραγματική απόδοση ή η τρέχουσα κατάσταση του μηχανήματος. Κατά συνέπεια, η κατάσταση του εξοπλισμού, είτε λειτουργεί κανονικά είτε βρίσκεται σε κατάσταση βλάβης, δεν λαμβάνεται υπόψη και η συντήρηση πραγματοποιείται αυστηρά σύμφωνα με το καθορισμένο χρονοδιάγραμμα. Η προκαθορισμένη συντήρηση είναι πιο αποτελεσματική όταν ο εξοπλισμός παρουσιάζει σχετικά προβλέψιμα πρότυπα βλαβών. Ωστόσο, η αυξανόμενη πολυπλοκότητα του σύγχρονου εξοπλισμού και η προηγμένη τεχνολογία κατά τις τελευταίες δεκαετίες, έχει καταστήσει πιο δύσκολο τον εντοπισμό σταθερών προτύπων βλάβης. Η μέθοδος αυτή

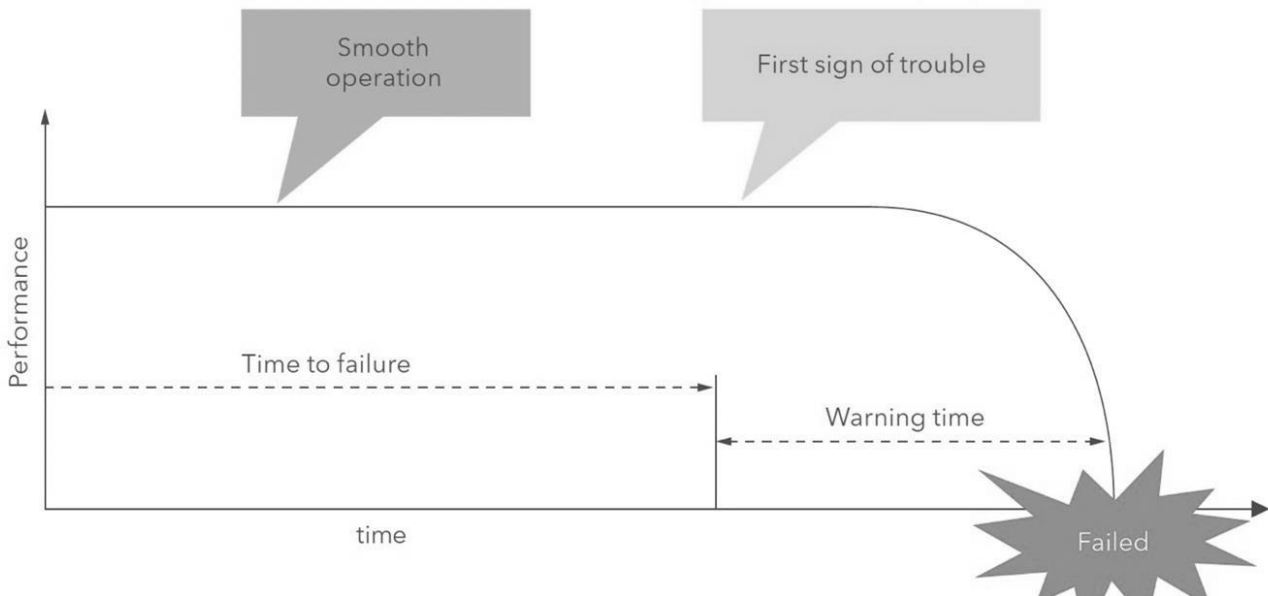
επικρίνεται συχνά ως σπάταλη, καθώς μπορεί να οδηγήσει σε «υπερσυντήρηση» και να προκαλέσει υπερβολικό κόστος χωρίς να ενισχύσει σημαντικά την αξιοπιστία του μηχανήματος [2].

1.4.4 Συντήρηση με βάση την κατάσταση του μηχανήματος

Η συντήρηση με βάση την κατάσταση του μηχανήματος (Conditioned Based Maintenance (CBM)) ορίζεται ως μια διαδικασία προληπτικής συντήρησης που περιλαμβάνει την παρακολούθηση της κατάστασης στην οποία βρίσκεται ένα μηχάνημα, την επιθεώρηση των εξαρτημάτων, τις δοκιμές, την ανάλυση και τις απαραίτητες διορθωτικές ενέργειες. Ουσιαστικά, η CBM επικεντρώνεται στη συντήρηση του εξοπλισμού τη σωστή στιγμή. Ο πυρήνας αυτής της στρατηγικής είναι η συνεχής παρακολούθηση των συστημάτων και των εξαρτημάτων, η οποία καθορίζει το πότε απαιτούνται εργασίες συντήρησης. Για να έχει επιτυχία η μέθοδος της συντήρησης με βάση την κατάσταση του μηχανήματος, απαιτείται συνεχής παρακολούθηση του εξοπλισμού, παρέχοντας δεδομένα σε πραγματικό ή σχεδόν πραγματικό χρόνο για τον εντοπισμό βλάβης και την ένδειξη της ακριβούς στιγμής κατά την οποία πρέπει να πραγματοποιηθεί συντήρηση.

Η παρακολούθηση στο CBM πραγματοποιείται με τη χρήση αισθητήρων που συλλέγουν συνεχώς δεδομένα σχετικά με την κατάσταση του συστήματος. Επειδή οι αισθητήρες αυτοί μπορεί να είναι ακριβοί, πολλές εταιρίες παρακολουθούν μόνο τα κρίσιμα εξαρτήματά τους. Αναμένεται όμως ότι, καθώς το κόστος των αισθητήρων θα μειώνεται, η χρήση της CBM θα επεκταθεί, επιτρέποντας την παρακολούθηση περισσότερων εξαρτημάτων, συμπεριλαμβανομένων των μη κρίσιμων.

Οι πρωταρχικοί στόχοι της CBM είναι η μείωση του κόστους με την αποφυγή της περιττής συντήρησης και η αύξηση της ασφάλειας με την εκτέλεση ενεργειών συντήρησης στον κατάλληλο χρόνο. Η CBM επικεντρώνεται στη λειτουργική κατάσταση του εξοπλισμού, επιτρέποντας την έγκαιρη ανίχνευση πιθανών βλαβών και τον προγραμματισμό της απαραίτητης συντήρησης. Ωστόσο, δεν παρέχει τα μέσα για την καθυστέρηση στο χρόνο εμφάνισης των βλαβών ή τη βελτίωση της συνολικής λειτουργικότητας του μηχανήματος [2].



Εικόνα 2 - Αναπαράσταση της έννοιας Conditioned Based Maintenance

1.4.5 Προγνωστική Συντήρηση

Η προγνωστική συντήρηση (Predictive Maintenance (PdM)) είναι μια προληπτική προσέγγιση της συντήρησης του εξοπλισμού, η οποία περιλαμβάνει την παρακολούθηση της κατάστασης και της απόδοσης των μηχανημάτων για την πρόβλεψη πιθανών βλαβών πριν από την εμφάνισή τους. Αναλύοντας τα δεδομένα που συλλέγονται από την παρακολούθηση των μηχανημάτων, η προγνωστική συντήρηση επιτρέπει στις ομάδες συντήρησης να εντοπίζουν μοτίβα, να ανιχνεύουν πρώιμα σημάδια φθοράς ή δυσλειτουργίας και να εκτιμούν πότε ένα εξάρτημα είναι πιθανό να παρουσιάσει βλάβη. Αυτό επιτρέπει τον προγραμματισμό των δραστηριοτήτων συντήρησης ακριβώς όταν χρειάζεται, ελαχιστοποιώντας τις απρόσμενες βλάβες και βελτιστοποιώντας τη χρήση των πόρων.

Η προγνωστική συντήρηση χρησιμοποιεί ιστορικά δεδομένα και προηγμένες αναλύσεις, όπως η μηχανική μάθηση ή στατιστικά μοντέλα, για να προβλέψει πότε ο εξοπλισμός θα παρουσιάσει πιθανή βλάβη. Η PdM παρακολουθεί την κατάσταση του εξοπλισμού και αναλύει τις τάσεις και τα πρότυπα στα δεδομένα για να προβλέψει πότε είναι πιθανό να εμφανιστούν μελλοντικά προβλήματα. Αυτό επιτρέπει τον προγραμματισμό της συντήρησης με βάση τις προβλεπόμενες μελλοντικές συνθήκες, αντί να βασίζεται αποκλειστικά στην κατάσταση του εξοπλισμού σε πραγματικό χρόνο. Η προγνωστική συντήρηση ουσιαστικά «προβλέπει» πότε μπορεί να εμφανιστεί βλάβη, επιτρέποντας προληπτικές ενέργειες συντήρησης. Οι τεχνικές που χρησιμοποιούνται συνήθως στην προγνωστική συντήρηση περιλαμβάνουν την ανάλυση κραδασμών, τη θερμική απεικόνιση, την ανάλυση λαδιού και τις δοκιμές υπερήχων. Αυτές οι μέθοδοι βοηθούν στην αξιολόγηση της υγείας του εξοπλισμού και στον εντοπισμό παρατυπιών που θα μπορούσαν να υποδηλώνουν επικείμενη βλάβη.

Η προγνωστική συντήρηση προσφέρει πολλά οφέλη. Συμβάλει στην αντιμετώπιση θεμάτων πριν αυτά κλιμακωθούν σε μεγαλύτερα προβλήματα, παρατείνει τη διάρκεια ζωής του εξοπλισμού με την πρόληψη καταστροφικών βλαβών και μειώνει το κόστος συντήρησης επιτρέποντας επισκευές ή αντικαταστάσεις μόνο

όταν είναι απαραίτητο. Επιπλέον, βελτιώνει την ασφάλεια μειώνοντας την πιθανότητα απροσδόκητων δυσλειτουργιών του εξοπλισμού. Ωστόσο, η προγνωστική συντήρηση απαιτεί επενδύσεις σε εργαλεία παρακολούθησης, δυνατότητες ανάλυσης δεδομένων και εξειδικευμένο προσωπικό για την αποτελεσματική ερμηνεία των δεδομένων. Αυτός ο τύπος συντήρησης είναι ιδιαίτερα χρήσιμος, καθώς ενισχύει την αξιοπιστία και τη λειτουργική αποδοτικότητα διατηρώντας τον εξοπλισμό στα βέλτιστα επίπεδα απόδοσης.

2. Χρονοσειρές

Σε αυτό το κεφάλαιο εξηγείται ένα μέγεθος που αλλάζει τιμές με το χρόνο, και μελετάται η εξάρτηση ενός μεγέθους X σε κάποια χρονική στιγμή t (X_t), από το ίδιο μέγεθος σε προηγούμενες χρονικές στιγμές X_{t-1} , X_{t-2} , Ένα σύνολο τέτοιων παρατηρήσεων λέγεται χρονική σειρά ή χρονοσειρά.

2.1 Ορισμός Χρονοσειράς

Με τον όρο Χρονοσειρά εννοούμε μια σειρά από παρατηρήσεις που λαμβάνονται με την πάροδο του χρόνου σε τακτά χρονικά διαστήματα ή σε περιόδους που ισαπέχουν μεταξύ τους. Αυτός ο τύπος δεδομένων χρησιμοποιείται για την παρακολούθηση των μεταβολών με την πάροδο του χρόνου.

Ένα από τα βασικά χαρακτηριστικά των χρονοσειρών είναι ότι τα δεδομένα είναι διατεταγμένα χρονολογικά, δηλαδή η σειρά των σημείων δεδομένων είναι σημαντική. Αυτή η χρονική διάταξη διακρίνει τα δεδομένα χρονολογικών σειρών από άλλους τύπους δεδομένων. Ένα δεύτερο χαρακτηριστικό είναι ότι κάθε παρατήρηση σε μια χρονοσειρά είναι μια τιμή που καταγράφεται σε μια συγκεκριμένη χρονική στιγμή. Τέλος, ένα τρίτο χαρακτηριστικό είναι πως τα δεδομένα χρονολογικών σειρών συλλέγονται συνήθως σε τακτά χρονικά διαστήματα (π.χ. ωριαία, ημερήσια, μηνιαία, ετήσια), αν και μπορεί να υπάρχουν και ακανόνιστες χρονολογικές σειρές.

Έστω Y_i οι τιμές των παρατηρήσεων και X_i οι αντίστοιχες χρονικές στιγμές μετρημένες σε έτη ή μήνες ή ημέρες ή ώρες, δευτερόλεπτα, κτλ. Τα ζεύγη της μορφής $M(X_i, Y_i)$ αποτελούν παρατηρήσεις χρονοσειρών και μπορούν να παρασταθούν στο καρτεσιανό σύστημα αξόνων. Με τη σύνδεση αυτών των σημείων δημιουργείται ένα χρονοδιάγραμμα που δείχνει τη συνολική εξέλιξη του υπό μελέτη φαινομένου ή χαρακτηριστικού.

Οι χρονοσειρές μπορούν να ταξινομηθούν σε γενικές γραμμές σε δύο ομάδες, στις διακριτές χρονοσειρές και στις συνεχείς χρονοσειρές. Συνεχής θεωρείται μια χρονοσειρά όπου η τιμή του φαινομένου καταγράφεται συνεχώς. Για παράδειγμα, τα δεδομένα καταγραφής ημερήσιων θερμοκρασιών αποτελούν μια συνεχή χρονοσειρά. Διακριτή χρονοσειρά είναι εκείνη στην οποία η τιμή του φαινομένου παρατηρείται μόνο σε προκαθορισμένα χρονικά διαστήματα, όπως για παράδειγμα τα επιτόκια καταθέσεων τραπεζής ανά μήνα. Ο εκάστοτε αναλυτής καθορίζει ποια τιμή θα χρησιμοποιήσει για αυτά τα χρονικά διαστήματα ή τη δειγματοληψία, τα οποία ποικίλλουν ανάλογα με την εφαρμογή. Μια πρόσθετη ταξινόμηση βασίζεται στον αριθμό των μεταβλητών που απαιτούν ανάλυση. Υπάρχουν δύο κατηγορίες με βάση αυτή την ταξινόμηση, οι μονοδιάστατες και οι πολυδιάστατες χρονοσειρές [8]. Η μονοδιάστατη χρονοσειρά αποτελείται από μεμονωμένες παρατηρήσεις που καταγράφονται με την πάροδο του χρόνου. Για παράδειγμα, οι ημερήσιες τιμές κλεισίματος μιας μετοχής αποτελούν μονοδιάστατη χρονοσειρά. Η πολυδιάστατη χρονοσειρά αποτελείται από πολλαπλές παρατηρήσεις που καταγράφονται ταυτόχρονα με την πάροδο του χρόνου. Για παράδειγμα, η ταυτόχρονη καταγραφή της θερμοκρασίας, της υγρασίας και της ταχύτητας του ανέμου κατά τη διάρκεια μιας περιόδου αποτελεί μια πολυδιάστατη χρονοσειρά.

2.2 Συνιστώσες Χρονοσειρών







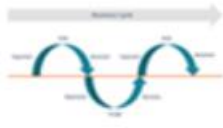
Τέσσερις συνιστώσες συνθέτουν συνήθως μια χρονοσειρά και αυτές μπορούν να εξαχθούν από τα δεδομένα για να χρησιμοποιηθούν στη λήψη αποφάσεων. Οι συνιστώσες αυτές είναι οι ακόλουθες:

- Τάση (Tend):

Η συνιστώσα της τάσης αντιπροσωπεύει τη μακροπρόθεσμη κίνηση ή κατεύθυνση των δεδομένων. Οι τάσεις μπορεί να είναι ανοδικές, καθοδικές ή επίπεδες, υποδεικνύοντας εάν τα δεδομένα γενικά αυξάνονται, μειώνονται ή παραμένουν σταθερά με την πάροδο του χρόνου.

- **Εποχικότητα (Seasonality):**
Η εποχικότητα αναφέρεται στις περιοδικές διακυμάνσεις των δεδομένων που συμβαίνουν σε τακτά χρονικά διαστήματα λόγω εποχικών παραγόντων. Για παράδειγμα, οι λιανικές πωλήσεις μπορεί να κορυφώνονται κατά την περίοδο των διακοπών κάθε χρόνο.
- **Κυκλικά Μοτίβα:**
Τα κυκλικά μοτίβα είναι παρόμοια με την εποχικότητα, αλλά εμφανίζονται σε ακανόνιστα διαστήματα. Τα μοτίβα αυτά συνήθως οφείλονται σε οικονομικούς κύκλους ή σε άλλους μη εποχικούς παράγοντες, όπου δηλαδή τα δεδομένα παρουσιάζουν αυξήσεις ή πτώσεις που δεν έχουν σταθερή συχνότητα. Η διάρκεια αυτών των μοτίβων είναι συνήθως παραπάνω από ένα χρόνο.
- **Ακανόνιστες Διακυμάνσεις (Θόρυβος):**
Οι ακανόνιστες διακυμάνσεις, ή αλλιώς θόρυβος, αναφέρονται στις απότομες μεταβολές των δεδομένων της χρονοσειράς που δε μπορούν να επαναληφθούν ποτέ σε μια χρονοσειρά. Αποτελούν τυχαίες διακυμάνσεις στα δεδομένα που δε μπορούν να αποδοθούν σε τάση, εποχικότητα ή κυκλικά μοτίβα. Για παράδειγμα, οι απεργίες είναι ένας τύπος ακανόνιστης διακύμανσης.

Στον παρακάτω πίνακα παρουσιάζονται συγκεντρωτικά όλες οι συνιστώσες των χρονοσειρών [24].

	Trend	Seasonality	Cyclical	Irregularity
Time	Fixed Time Interval	Fixed Time Interval	Not Fixed Time Interval	Not Fixed Time Interval
Duration	Long and Short Term	Short Term	Long and Short Term	Regular/Irregular
Visualization				
Nature - I	Gradual	Swings between Up or Down	Repeating Up and Down	Errored or High Fluctuation
Nature – II	Upward/Down Trend	Pattern repeatable	No fixed period	Short and Not repeatable
Prediction Capability	Predictable	Predictable	Challenging	Challenging
Market Model				Highly random/Unforeseen Events – along with white noise.

Εικόνα 3- Συνιστώσες Χρονοσειρών

Πηγή <https://www.analyticsvidhya.com/blog/2021/10/a-comprehensive-guide-to-time-series-analysis/>

Μια χρονοσειρά μπορεί να περιέχει μόνο ορισμένες από τις παραπάνω συνιστώσες και όχι όλες. Προκειμένου να εξεταστούν αυτές οι συνιστώσες, χρησιμοποιούνται δύο μαθηματικά μοντέλα που δείχνουν πώς

αυτές οι συνιστώσες προβλέπουν τις παρατηρήσεις της χρονοσειράς. Τα μοντέλα αυτά είναι το πολλαπλασιαστικό μοντέλο (multiplicative model) και το προσθετικό μοντέλο (additive model) [26].

2.3 Χαρακτηριστικά Χρονοσειρών

Η εκτίμηση του συστήματος που παράγει τη χρονοσειρά και η πιθανή πρόβλεψη μελλοντικών τιμών της ποσότητας που παρατηρούμε είναι οι προκλήσεις της ανάλυσης χρονοσειρών. Για να έχει νόημα η ανάλυση της χρονοσειράς, θα πρέπει πρώτα να απορρίψουμε την υπόθεση ότι η διακύμανση των τιμών της ποσότητας που παρατηρούμε είναι τυχαία, δηλαδή το σύστημα που παρατηρούμε είναι λευκός θόρυβος. Εάν οι παρατηρήσεις στη χρονοσειρά δεν είναι ανεξάρτητες, η πληροφορία που περιέχεται στη χρονοσειρά μπορεί να παρουσιάζεται με διαφορετικές μορφές. Συνεπώς, πριν προχωρήσουμε στην προσαρμογή ενός μοντέλου στη χρονοσειρά, θα πρέπει να εξετάσουμε τα ακόλουθα βασικά χαρακτηριστικά: Στασιμότητα της σειράς, Αιτιοκρατία και Στοχαστικότητα και Γραμμικότητα και Μη-γραμμικότητα της χρονοσειράς. Καθένα από αυτά τα χαρακτηριστικά αναλύεται στις παρακάτω παραγράφους.

2.3.1 Στασιμότητα

Η στασιμότητα είναι η ιδιότητα που δείχνει ότι οι διακυμάνσεις των τιμών της χρονοσειράς δεν μεταβάλλονται με την πάροδο του χρόνου. Πιο συγκεκριμένα, στασιμότητα στις χρονοσειρές σημαίνει ότι οι στατιστικές ιδιότητες των δεδομένων, όπως η μέση τιμή, η διακύμανση και η αυτοσυσχέτιση, παραμένουν σταθερές με την πάροδο του χρόνου. Αυτό σημαίνει ότι η χρονοσειρά δεν παρουσιάζει τάσεις, εποχιακά μοτίβα ή άλλες δομές που μεταβάλλονται με την πάροδο του χρόνου. Οι στάσιμες χρονοσειρές είναι ευκολότερο να μοντελοποιηθούν και να προβλεφθούν επειδή η συμπεριφορά τους δε μεταβάλλεται, καθιστώντας δυνατή την αποτελεσματική εφαρμογή στατιστικών τεχνικών ανάλυσης. Ουσιαστικά, μια στάσιμη χρονοσειρά μοιάζει ίδια σε κάθε χρονική στιγμή, χωρίς συστηματικές αλλαγές στο μοτίβο της.

2.3.2 Αιτιοκρατία και Στοχαστικότητα

Το αιτιοκρατικό μέρος του συστήματος που παράγει η χρονοσειρά αναφέρεται σε πρότυπα ή συμπεριφορές που είναι προβλέψιμες και ακολουθούν ένα συγκεκριμένο σύνολο κανόνων. Η στοχαστικότητα αναφέρεται στην τυχαιότητα ή στο απρόβλεπτο μιας χρονοσειράς. Μια στοχαστική χρονοσειρά περιλαμβάνει τυχαίες συνιστώσες που καθιστούν αδύνατη την πρόβλεψη μελλοντικών τιμών με απόλυτη βεβαιότητα. Αντ' αυτού, οι προβλέψεις γίνονται με όρους πιθανοτήτων. Όλες οι χρονοσειρές από πραγματικά μεγέθη περιέχουν θόρυβο, οπότε όλες οι πραγματικές χρονοσειρές είναι στοχαστικές. Στην ανάλυση των πραγματικών χρονοσειρών ο εντοπισμός του αιτιοκρατικού μέρους του συστήματος απαιτεί πρόκληση. Όταν αυτό είναι κρυμμένο μέσα στο θόρυβο, το σύστημα θεωρείται στοχαστικό.

2.3.3 Γραμμικότητα και Μη-Γραμμικότητα

Η γραμμικότητα στις χρονοσειρές αναφέρεται σε μια σχέση όπου οι μεταβολές των σημείων δεδομένων ακολουθούν μια ευθεία γραμμή όταν απεικονίζονται σε ένα γράφημα. Σε μια γραμμική χρονοσειρά, η τρέχουσα τιμή της σειράς είναι μια γραμμική συνάρτηση των προηγούμενων τιμών της και ενδεχομένως άλλων

παραγόντων. Αυτό σημαίνει ότι τυχόν διακυμάνσεις ή αλλαγές στη σειρά μπορούν να εκφραστούν ως άθροισμα των επιδράσεων των παρελθοντικών σημείων δεδομένων. Ένα παράδειγμα γραμμικών μοντέλων είναι το αυτοπαλίνδρομο μοντέλο, στο οποίο η τιμή σε μια συγκεκριμένη χρονική στιγμή είναι ένας γραμμικός συνδυασμός των προηγούμενων τιμών, όπως φαίνεται στην εξίσωση:

$$y_t = c + \varphi_1 y_{t-1} + \varphi_2 y_{t-2} + \dots + \varphi_p y_{t-p} + \varepsilon_t$$

Η μη γραμμικότητα στις χρονοσειρές αναφέρεται σε σχέσεις που δε μπορούν να περιγραφούν επαρκώς με τη χρήση μιας ευθείας γραμμής ή μιας απλής γραμμικής συνάρτησης. Στις μη γραμμικές χρονοσειρές, η τρέχουσα τιμή εξαρτάται από τις παρελθούσες τιμές με έναν πιο σύνθετο τρόπο, που συχνά περιλαμβάνει αλληλεπιδράσεις, εκθετικές επιδράσεις ή περίπλοκα μοτίβα που δε μπορούν να αποτυπωθούν από γραμμικά μοντέλα [25]. Παράδειγμα μη γραμμικού μοντέλου χρονοσειράς αποτελεί το Threshold Autoregressive (TAR) Models:

$$y_t = \begin{cases} \mu_1 + \varphi_1 y_{t-1} + u_{1t}, & \text{αν } s_{t-k} < r \\ \mu_2 + \varphi_2 y_{t-1} + u_{2t}, & \text{αν } s_{t-k} \geq r \end{cases}$$

Η κατανόηση του κατά πόσον μια χρονοσειρά είναι γραμμική ή μη γραμμική είναι σημαντική για την επιλογή των κατάλληλων τεχνικών μοντελοποίησης. Τα γραμμικά μοντέλα είναι κατάλληλα για πιο απλά πρότυπα δεδομένων, ενώ τα μη γραμμικά μοντέλα είναι απαραίτητα για την αποτύπωση πιο σύνθετων σχέσεων και συμπεριφορών στα δεδομένα.

3. Γραμμικά Μοντέλα Πρόβλεψης

Για την πρόβλεψη μελλοντικών τάσεων στα δεδομένα χρονοσειρών συνήθως χρησιμοποιούνται δύο μοντέλα, το Αυτοπαλίνδρομο (Autoregressive (AR)) και το Κινητού Μέσου Όρου (Moving Average -MA) [15]. Με τον συνδυασμό αυτών των δύο μοντέλων παράγεται το μοντέλο Αυτοπαλίνδρομου Κινητού Μέσου Όρου (Autoregressive Moving Average - ARMA) και το μοντέλο Αυτοπαλίνδρομου Ολοκληρωμένου Κινητού Μέσου Όρου (Autoregressive Integrated Moving Average - ARIMA). Μια παραλλαγή του ARIMA ονομάζεται Εποχιακός Αυτοπαλίνδρομος Ολοκληρωμένος Κινητός Μέσος Όρος (Seasonal Autoregressive Integrated Moving Average - SARIMA) και χρησιμοποιείται για δεδομένα που έχουν εποχικότητα, δηλαδή που παρουσιάζουν επαναλαμβανόμενα μοτίβα μέσα στο χρόνο. Αυτά τα μοντέλα βασίζονται στην προσέγγιση Box-Jenkins και είναι γνωστά για την εύκολη κατανόηση και ερμηνεία τους. Ωστόσο, η γραμμική φύση αυτών των μοντέλων περιορίζει την ικανότητά τους να προβλέπουν μη γραμμικά προβλήματα, τα οποία συνιστούν την πλειοψηφία των χρονοσειρών [16].

3.1 Autoregressive (AR) Model

Το αυτοπαλίνδρομο μοντέλο (AR) είναι ένα θεμελιώδες στατιστικό μοντέλο που χρησιμοποιείται για την ανάλυση και την πρόβλεψη δεδομένων χρονοσειρών. Σε ένα μοντέλο AR, προβλέπεται η τιμή ενδιαφέροντος χρησιμοποιώντας το γραμμικό συνδυασμό των προηγούμενων τιμών της (υστερήσεις) και ενός στοχαστικού όρου σφάλματος. Η προσέγγιση αυτή αποτυπώνει τη σχέση μεταξύ των προηγούμενων και των παρόντων τιμών, καθιστώντας την χρήσιμη για τον εντοπισμό και την πρόβλεψη προτύπων σε δεδομένα χρονοσειρών. Ο όρος αυτοπαλινδρόμηση υποδηλώνει ότι πρόκειται για παλινδρόμηση της μεταβλητής έναντι του εαυτού της [7].

Ένα μοντέλο αυτοπαλίνδρομης τάξης p μπορεί να γραφεί ως

$$y_t = c + \varphi_1 y_{t-1} + \varphi_2 y_{t-2} + \dots + \varphi_p y_{t-p} + \varepsilon_t$$

όπου:

- y_t είναι η τιμή της χρονοσειράς τη χρονική στιγμή t
- c είναι ένας σταθερός όρος
- $\varphi_1, \varphi_2, \dots, \varphi_p$ είναι οι συντελεστές του μοντέλου, οι οποίοι αντιπροσωπεύουν τα βάρη των τιμών με χρονική υστέρηση
- ε_t είναι ένας όρος σφάλματος λευκού θορύβου με μηδενικό μέσο και σταθερή διακύμανση σ^2

Αυτό είναι σαν μια πολλαπλή παλινδρόμηση, αλλά με τις υστερημένες τιμές του y_t ως προγνωστικούς παράγοντες. Αναφερόμαστε σε αυτό το μοντέλο ως AR(p). Η τάξη p του μοντέλου καθορίζει τον αριθμό των προηγούμενων παρατηρήσεων που χρησιμοποιούνται για την πρόβλεψη της τρέχουσας τιμής.

Για παράδειγμα, το πρώτης τάξης μοντέλο αυτοπαλινδρόμησης AR(1) γράφεται:

$$y_t = c + \varphi_1 y_{t-1} + \varepsilon_t$$

Ένα αυτοπαλίνδρομο μοντέλο εκφράζει μια πρόβλεψη σαν μία συνάρτηση των προηγούμενων τιμών των χρονοσειρών.

Για να είναι έγκυρο το μοντέλο AR, η χρονοσειρά πρέπει να είναι στάσιμη, δηλαδή οι στατιστικές της ιδιότητες (μέσος όρος, διακύμανση, αυτοσυσχέτιση) δεν μεταβάλλονται με την πάροδο του χρόνου.

Για τον προσδιορισμό της τάξης p του μοντέλου γίνονται τα διαγράμματα αυτοσυσχέτισης (ACF) και μερικής αυτοσυσχέτισης (PACF). Η συνάρτηση αυτοσυσχέτισης (ACF) ενός μοντέλου AR(p) παρουσιάζει σημαντική συσχέτιση μέχρι την υστέρηση p και στη συνέχεια τείνει στο μηδέν. Η συνάρτηση μερικής αυτοσυσχέτισης (PACF) διακόπτεται μετά την υστέρηση p . Αυτό σημαίνει ότι μόνο οι πρώτες υστερήσεις p έχουν σημαντική μερική συσχέτιση με την τρέχουσα τιμή [6][18].

3.2 Moving Average (MA) Model

Το μοντέλο Κινητού Μέσου Όρου (MA) χρησιμοποιείται κυρίως για την εξομάλυνση των βραχυπρόθεσμων διακυμάνσεων και την ανάδειξη μακροπρόθεσμων τάσεων. Σε αντίθεση με το αυτοπαλίνδρομο (AR) μοντέλο, το οποίο χρησιμοποιεί παρελθοντικές τιμές της ίδιας της χρονοσειράς, το μοντέλο MA χρησιμοποιεί παρελθονικά σφάλματα πρόβλεψης για την πρόβλεψη μελλοντικών τιμών. Τα σφάλματα αυτά είναι οι διαφορές μεταξύ των παρατηρούμενων τιμών και των προβλεπόμενων τιμών σε προηγούμενες χρονικές στιγμές.

Ένα μοντέλο κινητού μέσου τάξης q μπορεί να παρασταθεί μαθηματικά ως εξής:

$$y_t = c + \varepsilon_t + \theta_1 \varepsilon_{t-1} + \theta_2 \varepsilon_{t-2} + \dots + \theta_q \varepsilon_{t-q}$$

όπου:

- $\theta_1, \theta_2, \dots, \theta_q$ είναι οι συντελεστές που αντιπροσωπεύουν τα βάρη των παρελθοντικών όρων σφάλματος
- ε_t είναι λευκός θόρυβος τη χρονική στιγμή t

Αναφερόμαστε σε αυτό το μοντέλο ως MA(q). Κάθε τιμή του y_t μπορεί να θεωρηθεί ως ένας σταθμισμένος κινητός μέσος όρος των q τελευταίων σφαλμάτων πρόβλεψης. Για να είναι έγκυρο το μοντέλο MA, η χρονοσειρά πρέπει να είναι στάσιμη.

Για τον προσδιορισμό της τάξης q του μοντέλου γίνονται τα διαγράμματα αυτοσυσχέτισης (ACF) και μερικής αυτοσυσχέτισης (PACF). Η συνάρτηση αυτοσυσχέτισης ACF ενός μοντέλου MA(q) παρουσιάζει σημαντικές συσχετίσεις μέχρι την υστέρηση q και στη συνέχεια τείνει προς το μηδέν. Αυτό το πρότυπο βοηθά στον προσδιορισμό της τάξης q του υποδείγματος. Σε αντίθεση με τα μοντέλα AR, η συνάρτηση μερικής αυτοσυσχέτισης (PACF) ενός μοντέλου Κινητού Μέσου (MA) δεν παρουσιάζει ένα απλό μοτίβο αποκοπής. Οι τιμές της PACF μειώνονται σταδιακά και δεν παρέχουν σαφείς πληροφορίες σχετικά με την τάξη q του μοντέλου [18].

Για παράδειγμα, το πρώτης τάξης μοντέλο κινητού μέσου MA(1) γράφεται:

$$y_t = c + \varepsilon_t + \theta_1 \varepsilon_{t-1}$$

Σε αυτό το υπόδειγμα, η τρέχουσα τιμή y_t επηρεάζεται από το σταθερό όρο c , το τρέχον σφάλμα ε_t και το σφάλμα της προηγούμενης περιόδου ε_{t-1} . Ο συντελεστής θ_1 υποδεικνύει την ισχύ και την κατεύθυνση της σχέσης μεταξύ y_t και ε_{t-1} .

Η αλλαγή των παραμέτρων $\theta_1, \theta_2, \dots, \theta_q$ οδηγεί σε διαφορετικά πρότυπα χρονοσειρών. Όπως και στα αυτοπαλίνδρομα μοντέλα, η διακύμανση του όρου σφάλματος ε_t θα αλλάξει μόνο την κλίμακα της σειράς και όχι τα πρότυπα [7].

3.3 Autoregressive Moving Average (ARMA) Model

Το μοντέλο αυτοπαλίνδρομου κινητού μέσου (ARMA) είναι ένας συνδυασμός του αυτοπαλίνδρομου (AR) μοντέλου και του κινητού μέσου (MA). Ο λόγος δημιουργίας του ήταν ορισμένα προβλήματα που απαιτούσαν μοντέλα AR ή MA υψηλής τάξης, με πολλές παραμέτρους. Για να ξεπεραστεί αυτό το πρόβλημα δημιουργήθηκε το ARMA, ο συνδυασμός των δύο αυτών μοντέλων, έτσι ώστε ο αριθμός των χρησιμοποιούμενων παραμέτρων να διατηρείται μικρός [19]. Χάρη στην ενσωμάτωση των χαρακτηριστικών των δύο μοντέλων, το ARMA μπορεί να συλλάβει ένα ευρύτερο φάσμα προτύπων σε δεδομένα χρονοσειρών, καθιστώντας το ένα ισχυρό εργαλείο για ανάλυση και πρόβλεψη [17].

Ένα μοντέλο ARMA ορίζεται από δύο παραμέτρους p και q . Το p αντιπροσωπεύει την τάξη του αυτοπαλίνδρομου μέρους και το q την τάξη του κινητού μέσου όρου. Το μοντέλο ARMA(p,q) μπορεί να παρασταθεί μαθηματικά ως εξής:

$$y_t = c + \varphi_1 y_{t-1} + \theta_1 \varepsilon_{t-1} + \varphi_2 y_{t-2} + \theta_2 \varepsilon_{t-2} + \dots + \varphi_p y_{t-p} + \theta_q \varepsilon_{t-q} + \varepsilon_t$$

όπου φ_i και θ_i είναι οι συντελεστές του μοντέλου AR και MA αντίστοιχα.

Το τμήμα AR του μοντέλου χρησιμοποιεί παρελθοντικές τιμές της χρονοσειράς για να προβλέψει μελλοντικές τιμές. Αυτό το στοιχείο αποτυπώνει τη σχέση μεταξύ της τρέχουσας τιμής και των υστερήσεων της. Το τμήμα MA του μοντέλου χρησιμοποιεί παρελθοντικά σφάλματα πρόβλεψης για την πρόβλεψη μελλοντικών τιμών. Αυτό το στοιχείο λαμβάνει υπόψη τις τυχαίες διαταραχές στη χρονοσειρά. Το υπόδειγμα ARMA απαιτεί η χρονοσειρά να είναι στάσιμη, δηλαδή οι στατιστικές της ιδιότητες μέσος όρος, διακύμανση, και αυτοσυσχέτιση να μην μεταβάλλονται με την πάροδο του χρόνου.

3.4 AutoRegressive Integrated Moving Average (ARIMA) Model

Μια αρκετά δημοφιλής και ευρέως χρησιμοποιούμενη στατιστική μέθοδος για την πρόβλεψη χρονοσειρών είναι το μοντέλο ARIMA. Το όνομα του μοντέλου ARIMA προέρχεται από τα αρχικά των αγγλικών λέξεων AutoRegressive Integrated Moving Average. Το μοντέλο ARIMA συνδυάζει τρία βασικά στοιχεία, την αυτοπαλινδρόμηση (AR), τη διαφορά (I) και τον κινητό μέσο όρο (MA). Με την ενσωμάτωση όλων αυτών των στοιχείων, το ARIMA μπορεί να χειριστεί αρκετά δεδομένα χρονοσειρών, συμπεριλαμβανομένων και εκείνων με τάση και εποχικότητα, δηλαδή των μη-στάσιμων χρονοσειρών. Τα μοντέλα που αναλύθηκαν παραπάνω, AR – MA – ARMA, χρησιμοποιούνται σε στάσιμες χρονοσειρές, δηλαδή ο μέσος και η διακύμανση δεν εξαρτώνται από τον χρόνο t , αλλά παραμένουν σταθερά. Για να μετατραπεί ένα μη-στάσιμο μοντέλο σε στάσιμο, πρέπει να αφαιρεθούν τα μη-στάσιμα χαρακτηριστικά. Το ARIMA μπορεί να εφαρμοστεί και σε μη-στάσιμες χρονοσειρές χάρη στην παράμετρο της διαφοράς (I) που διαθέτει [16].

Ένα μοντέλο ARIMA συμβολίζεται ως ARIMA(p,d,q) όπου:

- p είναι η τάξη του αυτοπαλίνδρομου τμήματος (AR)
- d είναι ο βαθμός διαφοράς που απαιτείται για να καταστεί η σειρά στάσιμη (I)
- q είναι η τάξη του κινητού μέσου όρου (MA)

Επομένως το μοντέλο ARIMA θα μπορούσε να σπάσει σε ARIMA(AR, I, MA) όπου:

- AR (Autoregression): Αυτό δίνει έμφαση στην εξαρτημένη σχέση μεταξύ μιας παρατήρησης και των προηγούμενων ή «καθυστερημένων» παρατηρήσεων.
- I (Integrated): Για την επίτευξη μιας στάσιμης χρονοσειράς, δηλαδή μιας χρονοσειράς που δεν παρουσιάζει τάση ή εποχικότητα, εφαρμόζεται διαφορά. Συνήθως περιλαμβάνει την αφαίρεση μιας παρατήρησης από την προηγούμενη παρατήρησή της. Για παράδειγμα, η πρώτης τάξης διαφορά πραγματοποιείται αντικαθιστώντας το y_t με τη σχέση:

$$y'_t = y_t - y_{t-1}$$

- MA (κινητός μέσος όρος): Αυτή η συνιστώσα μηδενίζει τη σχέση μεταξύ μιας παρατήρησης και του υπολειπόμενου σφάλματος από ένα μοντέλο κινητού μέσου όρου που βασίζεται σε καθυστερημένες παρατηρήσεις [6][5].

Η γενική μορφή ενός υποδείγματος ARIMA(p,d,q) μπορεί να εκφραστεί από την παρακάτω σχέση:

$$y'_t = c + \varphi_1 y'_{t-1} + \dots + \varphi_p y'_{t-p} + \theta_1 \varepsilon_{t-1} + \dots + \theta_q \varepsilon_{t-q} + \varepsilon_t$$

όπου y'_t είναι η σειρά μετά τις διαφορές, οι οποίες μπορεί να έχουν γίνει περισσότερες από μία φορές. Οι «παράγοντες πρόβλεψης» στο δεξί μέλος περιλαμβάνουν και τιμές υστέρησης y_t και σφάλματα υστέρησης.

Παρακάτω παρουσιάζεται ένας πίνακας με τις ειδικές περιπτώσεις μοντέλων ARIMA [4].

Πίνακας 1 - Ονομασίες μοντέλων ARIMA

Λευκός θόρυβος	ARIMA(0,0,0) χωρίς σταθερά
Τυχαίος περίπατος	ARIMA(0,1,0) χωρίς σταθερά
Τυχαίος περίπατος με περιπλάνηση	ARIMA(0,1,0) με σταθερά
Αυτοπαλινδρόμηση	ARIMA(p,0,0)
Κινητός μέσος	ARIMA(0,0,q)

3.4.1 Συνάρτηση auto_arima

Ο προσδιορισμός των βέλτιστων παραμέτρων p , d και q για το μοντέλο ARIMA μπορεί να είναι μια πολύπλοκη και χρονοβόρα διαδικασία. Για την αντιμετώπιση αυτής της πρόκλησης, η βιβλιοθήκη `pmdarima` προσφέρει τη συνάρτηση `auto_arima`, η οποία αυτοματοποιεί τη διαδικασία επιλογής του μοντέλου ARIMA, βελτιώνοντας έτσι τη διαδικασία κατασκευής του μοντέλου. Η συνάρτηση αυτή αυτοματοποιεί τη διαδικασία εντοπισμού των καλύτερων παραμέτρων του μοντέλου ARIMA (p , d , q) με αναζήτηση σε πολλαπλούς συνδυασμούς. Η συνάρτηση αυτή διερευνά ένα εύρος πιθανών παραμέτρων του μοντέλου και επιλέγει αυτό που εξισορροπεί βέλτιστα την πολυπλοκότητα του μοντέλου και την καλή προσαρμογή, όπως μετράται με τα κριτήρια πληροφορίας όπως το κριτήριο πληροφορίας Akaike (AIC) και το κριτήριο πληροφορίας Bayes (BIC). Όσο πιο χαμηλοί οι συντελεστές AIC και BIC, τόσο καλύτερο το μοντέλο.

Για να χρησιμοποιήσουμε τη βιβλιοθήκη της Python “`pmdarima`” πρώτα χρειάστηκε να την κατεβάσουμε μέσω της εντολής `pip`:

```
pip install pmdarima
```


Η συνάρτηση `auto_arima` παρουσιάζεται παρακάτω με όλες τις παραμέτρους που περιέχει.

```
pmdarima.arima.auto_arima(y, X=None, start_p=2, d=None, start_q=2, max_p=5, max_d=2, max_q=5,
start_P=1, D=None, start_Q=1, max_P=2, max_D=1, max_Q=2, max_order=5, m=1, seasonal=True,
stationary=False, information_criterion='aic', alpha=0.05, test='kpss', seasonal_test='ocsb',
stepwise=True, n_jobs=1, start_params=None, trend=None, method='lbfgs', maxiter=50,
offset_test_args=None, seasonal_test_args=None, suppress_warnings=True, error_action='trace',
trace=False, random=False, random_state=None, n_fits=10, return_valid_fits=False, out_of_sample_size=0,
scoring='mse', scoring_args=None, with_intercept='auto', sarimax_kwargs=None, **fit_args)
```

Η συνάρτηση `auto_arima` είναι ιδιαίτερα ευέλικτη, επιτρέποντας στους χρήστες να καθορίσουν το εύρος των παραμέτρων που πρέπει να ληφθούν υπόψη κατά τη διαδικασία επιλογής μοντέλου. Οι βασικές παράμετροι της συνάρτησης `auto_arima` περιγράφονται λεπτομερώς παρακάτω:

- `y`: Τα δεδομένα της χρονοσειράς εισόδου.
- `start_p`: Η αρχική τιμή p για την τάξη AR (AutoRegressive). Η ρύθμιση αυτής της παραμέτρου σε 0 υποδηλώνει ότι η αναζήτηση της βέλτιστης τάξης AR ξεκινά από το 0.
- `start_q`: Η αρχική τιμή q για την τάξη MA (Moving Average). Η ρύθμιση αυτής της παραμέτρου σε 0 υποδηλώνει ότι η αναζήτηση της βέλτιστης εντολής MA ξεκινά από το 0.
- `max_p`: Η μέγιστη τιμή p για την τάξη AR (p) που θα ληφθεί υπόψη στην αναζήτηση.
- `max_q`: Η μέγιστη τιμή q για την τάξη MA (q) που θα ληφθεί υπόψη στην αναζήτηση.
- `m`: Ο αριθμός των περιόδων σε κάθε εποχιακό κύκλο. Αυτή η παράμετρος είναι σχετική για δεδομένα εποχικών χρονοσειρών, όταν `seasonal=True`.
- `Seasonal`: Μια δυαδική παράμετρος που υποδεικνύει αν θα εφαρμοστεί ένα εποχιακό μοντέλο ARIMA. Εάν τεθεί σε `True`, το μοντέλο θα εξετάσει την εποχιακή διαφορά.
- `d`: Η τάξη της διαφοράς (d). Εάν δεν υπάρχει, η συνάρτηση θα καθορίσει αυτόματα την κατάλληλη τάξη διαφοράς.
- `Stationary`: Μια δυαδική παράμετρος που προσδιορίζει τη στασιμότητα της χρονοσειράς. Αν η χρονοσειρά είναι στάσιμη, τότε παίρνει την τιμή `True` και το `d` πρέπει να μηδενιστεί. Η προεπιλεγμένη τιμή είναι `False`.
- `trace`: Δυαδική παράμετρος. Αν είναι `True`, η συνάρτηση θα εκτυπώσει πληροφορίες σχετικά με τη διαδικασία προσαρμογής.
- `error_action`: Ελέγχει τη συμπεριφορά κατά την προσαρμογή των μοντέλων με παραμέτρους που δε συγκλίνουν. Η προεπιλογή είναι "warn", δηλαδή προειδοποιεί κάθε φορά που υπάρχει ένα σφάλμα.
- `Suppress_warnings`: Δυαδική παράμετρος, η οποία όταν είναι `True` καταστέλλει τις προειδοποιήσεις για να διατηρήσει καθαρή την έξοδο του αλγορίθμου.
- `Stepwise`: Δυαδική παράμετρος. Αν είναι `True`, χρησιμοποιείται ένας αλγόριθμος βηματικής αναζήτησης για την αναζήτηση του βέλτιστου μοντέλου. Αυτό μπορεί να είναι γρηγορότερο αλλά λιγότερο ενδεδειγμένο από μια διεξοδική αναζήτηση [9].

Επιπλέον παράμετροι που χρησιμοποιεί η `auto_arima` συνάρτηση όταν η χρονοσειρά παρουσιάζει εποχικότητα είναι οι παρακάτω:

- `max_P`, `max_Q`: Παράμετροι για εποχικούς όρους AR και MA.
- `seasonal_order`: Ορίζει την εποχιακή συνιστώσα του μοντέλου ως (P,D,Q,s).
- `information_criterion`: Κριτήριο που χρησιμοποιείται για την επιλογή μοντέλου, συνήθως «aic» ή «bic».
- `max_d` και `max_D`: Μέγιστες τάξεις για μη εποχιακή και εποχιακή διαφορά.

Η συνάρτηση `auto_arima` λειτουργεί μέσω μιας σειράς συστηματικών βημάτων που αποσκοπούν στον προσδιορισμό του βέλτιστου μοντέλου ARIMA για ένα δεδομένο σύνολο δεδομένων χρονοσειράς. Η διαδικασία αυτή ξεκινά με την αρχική διαφορά και προχωρά μέσω της επιλογής μοντέλου, της αξιολόγησης και της τελικής εξόδου.

Αρχικά, η συνάρτηση καθορίζει τη τάξη της διαφοράς (d) που είναι απαραίτητη για να καταστήσει τη χρονοσειρά στάσιμη. Αυτό το βήμα είναι κρίσιμο, διότι μια στάσιμη σειρά έχει συνεπείς στατιστικές ιδιότητες με την πάροδο του χρόνου, καθιστώντας την κατάλληλη για μοντελοποίηση ARIMA. Η σειρά της διαφοράς προσδιορίζεται συνήθως με τη χρήση στατιστικών δοκιμών, όπως η δοκιμασία Augmented Dickey-Fuller (ADF), η οποία αξιολογεί την παρουσία μοναδιαίων ριζών στα δεδομένα.

Μόλις καθοριστεί η κατάλληλη σειρά διαφοράς, η συνάρτηση `auto_arima` αναλαμβάνει το έργο της επιλογής μοντέλου. Κατά τη διάρκεια αυτής της φάσης, η συνάρτηση πραγματοποιεί αναζήτηση σε καθορισμένα εύρη των παραμέτρων των αυτοπαλίνδρομων (p) και των κινητών μέσων (q), καθώς και των εποχικών αντίστοιχων παραμέτρων τους (P και Q), εφόσον υπάρχουν. Η αναζήτηση μπορεί να διεξαχθεί εξαντλητικά, αξιολογώντας όλους τους δυνατούς συνδυασμούς εντός των συγκεκριμένων περιοχών των παραμέτρων. Εναλλακτικά, μπορεί να χρησιμοποιηθεί μια πιο αποτελεσματική σταδιακή αναζήτηση. Η βηματική αναζήτηση προσθέτει ή αφαιρεί παραμέτρους με βάση τη στατιστική τους σημασία, μειώνοντας έτσι τον υπολογιστικό φόρτο χωρίς να διακυβεύεται η πληρότητα της αναζήτησης.

Για κάθε υποψήφιο μοντέλο, η συνάρτηση `auto_arima` προσαρμόζει το μοντέλο ARIMA στα δεδομένα και υπολογίζει κριτήρια πληροφοριών όπως το κριτήριο πληροφορίας Akaike (AIC) και το κριτήριο πληροφορίας Bayes (BIC). Αυτά τα κριτήρια χρησιμεύουν ως μέτρα ποιότητας του μοντέλου, εξισορροπώντας την καλή προσαρμογή με την πολυπλοκότητα του μοντέλου. Επιβάλλοντας κυρώσεις σε μοντέλα με περισσότερες παραμέτρους, το AIC και το BIC συμβάλλουν στην πρόληψη της υπερβολικής προσαρμογής, διασφαλίζοντας ότι το επιλεγμένο μοντέλο γενικεύεται καλά σε νέα δεδομένα.

Το βέλτιστο μοντέλο επιλέγεται με βάση την τιμή του κριτηρίου πληροφορίας, με το μοντέλο που παρουσιάζει τη χαμηλότερη τιμή να επιλέγεται ως το μοντέλο με την καλύτερη προσαρμογή. Η προσέγγιση αυτή διασφαλίζει ότι το επιλεγμένο μοντέλο επιτυγχάνει ισορροπία μεταξύ προσαρμογής και απλότητας, ελαχιστοποιώντας τον κίνδυνο υπερπροσαρμογής και αποτυπώνοντας παράλληλα τα βασικά πρότυπα των δεδομένων.

Τέλος, η συνάρτηση `auto_arima` εξάγει το μοντέλο με την καλύτερη προσαρμογή, παρέχοντας μια περιεκτική σύνοψη των παραμέτρων του, τη στατιστική τους σημασία και διάφορες διαγνωστικές μετρήσεις. Αυτή η λεπτομερής έξοδος επιτρέπει στους χρήστες να αξιολογήσουν περαιτέρω και να βελτιώσουν το μοντέλο, όπως απαιτείται, διασφαλίζοντας ότι ανταποκρίνεται στις συγκεκριμένες απαιτήσεις της εργασίας πρόβλεψης. Μέσω αυτής της συστηματικής και αυτοματοποιημένης διαδικασίας, το `auto_arima` απλοποιεί την επιλογή του μοντέλου ARIMA, καθιστώντας την προσιτή και αποτελεσματική.

3.5 Seasonal ARIMA (SARIMA) Model

Το μοντέλο SARIMA είναι μια επέκταση του μοντέλου ARIMA που έχει σχεδιαστεί για να διαχειρίζεται δεδομένα χρονολογικών σειρών με εποχιακά πρότυπα. Με την ενσωμάτωση εποχικών στοιχείων στο μοντέλο, το SARIMA μπορεί να συλλάβει και να προβλέψει αποτελεσματικά δεδομένα που παρουσιάζουν τακτικές εποχικές διακυμάνσεις.

Στο μοντέλο SARIMA υπάρχουν τρία στοιχεία τάσεων που απαιτούν διαμόρφωση και είναι τα ίδια με το μοντέλο ARIMA:

- p : Τάξη της μη εποχικής αυτοπαλίνδρομης (AR) συνιστώσας
- d : Βαθμός της μη εποχικής διαφοράς
- q : Τάξη της συνιστώσας του μη εποχικού κινητού μέσου (MA)

Υπάρχουν επίσης και τέσσερα εποχικά στοιχεία που δεν αποτελούν μέρος του ARIMA, τα οποία είναι τα παρακάτω:

- P : Το πλήθος των εποχικών αυτοπαλίνδρομων όρων (SAR)
- D : Βαθμός εποχικής διαφοροποίησης
- Q : Το πλήθος των εποχικών όρων του κινητού μέσου (SMA)
- m : Ο αριθμός των χρονικών βημάτων για μια εποχική περίοδο.

Ένα μοντέλο SARIMA συμβολίζεται ως $SARIMA(p,d,q)(P,D,Q)m$ [16][20].

4 Κριτήρια Επιλογής Τάξης Μοντέλου ARIMA

Κατά την επιλογή των καλύτερων τάξεων p , d και q για ένα μοντέλο ARIMA(p,d,q), χρησιμοποιούνται συχνά δύο σημαντικά κριτήρια. Ένα είναι το κριτήριο πληροφορίας Akaike (AIC) και το άλλο είναι το κριτήριο πληροφορίας Bayes (BIC). Και τα δύο συμβάλλουν στην εξισορρόπηση της καλής προσαρμογής του μοντέλου με την πολυπλοκότητά του, με στόχο την αποφυγή της υπερπροσαρμογής, διασφαλίζοντας παράλληλα ότι το μοντέλο αποτυπώνει τα υποκείμενα πρότυπα των δεδομένων.

4.1 Akaike Information Criterion (AIC)

Το κριτήριο πληροφορίας Akaike μετρά την καλή προσαρμογή ενός στατιστικού μοντέλου. Περιγράφει τον συμβιβασμό μεταξύ μεροληψίας και διακύμανσης κατά την κατασκευή του μοντέλου, δηλαδή αυτόν της ακρίβειας και της πολυπλοκότητας του μοντέλου.

Το AIC δεν αποτελεί έλεγχο του μοντέλου με την έννοια του ελέγχου υποθέσεων. Αντίθετα, παρέχει ένα μέσο σύγκρισης μεταξύ μοντέλων. Αποτελεί δηλαδή ένα εργαλείο για την επιλογή μοντέλων. Δεδομένου ενός συνόλου δεδομένων, διάφορα υποψήφια μοντέλα μπορούν να καταταχθούν σύμφωνα με το AIC τους, με το μοντέλο που έχει το ελάχιστο AIC να είναι το καλύτερο.

Το AIC ορίζεται ως εξής:

$$AIC = 2k - 2 \cdot \ln(L)$$

όπου k είναι ο αριθμός των παραμέτρων του μοντέλου και $\ln(L)$ είναι η συνάρτηση λογαριθμικής πιθανοφάνειας για το στατιστικό μοντέλο.

Για μικρότερα σύνολα δεδομένων το AIC εφαρμόζει διόρθωση 2^{ης} τάξης:

$$AICc = AIC + \frac{2k(k+1)}{N-k-1} = \frac{2 \cdot N \cdot k}{N-k-1} - 2 \cdot \ln(L)$$

όπου N είναι το μέγεθος του δείγματος των δεδομένων [21].

4.2 Bayesian Information Criterion (BIC)

Το κριτήριο πληροφορίας Bayes, είναι παρόμοιο με το AIC, αλλά περιλαμβάνει ισχυρότερη ποινή για μοντέλα με περισσότερες παραμέτρους.

Το κριτήριο πληροφορίας BIC ορίζεται ως εξής:

$$BIC = -2 \cdot \ln(L) + 2 \cdot \ln(N) \cdot k$$

όπου L είναι η τιμή της πιθανοφάνειας, N είναι ο αριθμός των δεδομένων και k είναι ο αριθμός των εκτιμώμενων παραμέτρων [23].

Όπως και το κριτήριο AIC, έτσι και το BIC εξισορροπεί την προσαρμογή και την πολυπλοκότητα του μοντέλου, αλλά τείνει να ευνοεί πιο έντονα τα απλούστερα μοντέλα λόγω του όρου $\ln(N)$ στην ποινή. Παρόμοια με το AIC, το BIC χρησιμοποιείται για σύγκριση μοντέλων. Προτιμάται το μοντέλο με το χαμηλότερο BIC [8][22].

Συνεπώς, το AIC χρησιμοποιείται συχνά όταν ο πρωταρχικός στόχος είναι η ακρίβεια πρόβλεψης, εστιάζοντας στην εύρεση ενός μοντέλου που εξηγεί καλύτερα τα δεδομένα με ενδεχομένως περισσότερες παραμέτρους. Το BIC προτιμάται όταν η απλότητα του μοντέλου είναι ζωτικής σημασίας και υπάρχει ανάγκη αποφυγής της υπερπροσαρμογής, ιδίως με μεγάλα σύνολα δεδομένων.

Τα AIC και BIC είναι βασικά εργαλεία για την επιλογή του καλύτερου μοντέλου ARIMA, βοηθώντας στην εξισορρόπηση του συμβιβασμού μεταξύ της προσαρμογής του μοντέλου και της πολυπλοκότητας. Συγκρίνοντας αυτά τα κριτήρια σε διαφορετικά μοντέλα, προσδιορίζεται το καταλληλότερο μοντέλο για τα δεδομένα της χρονοσειράς, εξασφαλίζοντας ακριβείς και αξιόπιστες προβλέψεις.

5. Περιγραφή Δεδομένων

Το αντικείμενο του παρόντος κεφαλαίου είναι η περιγραφή της μεθοδολογίας που ακολουθήθηκε στα πειράματα της παρούσας εργασίας. Πιο συγκεκριμένα, το κεφάλαιο 5 περιγράφει τα σύνολα δεδομένων που χρησιμοποιήθηκαν, την προ-επεξεργασία, την ανάλυση των δεδομένων και την ανάπτυξη των δοκιμαστικών συνόλων που θα χρησιμοποιηθούν από τα μοντέλα, όπως τα σύνολα δεδομένων εκπαίδευσης και ελέγχου.

5.1 Περιγραφή Συνόλου Δεδομένων

Τα δεδομένα ελήφθησαν από μια βάση δεδομένων προγραμματισμού εργασιακών πόρων ERP μιας ναυτιλιακής εταιρείας η οποία κατέχει 40 πλοία. Γενικά, η βάση δεδομένων περιείχε διάφορα δεδομένα σχετικά με τη συντήρηση των πλοίων. Για την εργασία αυτή, συγκεντρώθηκαν ιστορικά δεδομένα όπως οι δραστηριότητες συντήρησης που έχουν οριστεί για το μηχάνημα CE-DE, 12V50DF και οι απροσδόκητες εργασίες που πραγματοποιήθηκαν στο συγκεκριμένο μηχάνημα όποτε ήταν απαραίτητο.

Τα αρχικά σύνολα δεδομένων που χρησιμοποιήθηκαν για την παρούσα εργασία είναι δύο. Και τα δύο αφορούν τα ιστορικά αρχεία ενός συστήματος προγραμματισμένης συντήρησης μιας πραγματικής ναυτιλιακής εταιρείας. Το ένα σύνολο δεδομένων αφορούσε τις προγραμματισμένες εργασίες που γίνονται στα πλοία της εταιρείας, και το δεύτερο σύνολο αφορούσε τις απρογραμματίστες εργασίες που προκύπτουν σε ένα πλοίο, οι οποίες όμως καταγράφονται από τους ναυτικούς μέσα στο σύστημα.

Τα δεδομένα που περιείχαν και τα δύο αυτά σύνολα δεδομένων περιγράφονται παρακάτω:

- **JobName:** το όνομα της κάθε εργασίας που αφορά το μηχάνημα CE-DE, 12V50DF.
- **ShipComponent:** το όνομα του μηχανήματος CE-DE, 12V50DF πάνω στο οποίο γίνεται όλη η ανάλυση.
- **ShipComponentID:** το μοναδικό αναγνωριστικό (GUID) του μηχανήματος αυτού το οποίο για κάθε πλοίο είναι μοναδικό.
- **DueDate:** η ημερομηνία λήξης της κάθε εργασίας συντήρησης.
- **DateDone:** η ημερομηνία εκτέλεσης της κάθε εργασίας συντήρησης.
- **DueRunningHours:** οι απαιτούμενες ώρες λειτουργίας του μηχανήματος στις οποίες θα έπρεπε να κλείσει η εργασία συντήρησης.
- **DoneRunningHours:** οι ώρες λειτουργίας του μηχανήματος στις οποίες εν τέλει έκλεισε η εργασία συντήρησης.
- **LastDoneRunningHours:** οι ώρες λειτουργίας του μηχανήματος όπως είχαν δηλωθεί στο σύστημα την τελευταία φορά που είχε κλείσει η συγκεκριμένη εργασία συντήρησης.
- **On-time Completion:** αυτή είναι δυαδική μεταβλητή. Υπολογίζεται με την αφαίρεση της ημερομηνίας εκτέλεσης μείον την ημερομηνία λήξης ($DateDone - DueDate$). Αν η διαφορά αυτή είναι θετική σημαίνει ότι η εργασία έγινε στην ώρα της, δηλαδή πριν την ημερομηνία λήξης της, ενώ όταν η διαφορά είναι αρνητική σημαίνει ότι η εργασία καθυστέρησε να γίνει, δηλαδή ξεπέρασε την ημερομηνία λήξης. Όταν, λοιπόν η παράμετρος αυτή έχει την τιμή 0 σημαίνει ότι η εργασία συντήρησης δεν έγινε έγκαιρα, ενώ όταν έχει την τιμή 1 σημαίνει πως έγινε έγκαιρα, πριν την ημερομηνία λήξης της εργασίας.
- **DefectNoted:** Δυαδική μεταβλητή. Όταν έχει τιμή 1 σημαίνει ότι το μηχάνημα CE-DE, 12V50DF παρουσίασε απροσδόκητη βλάβη και κατά συνέπεια δημιουργήθηκε στη βάση απρογραμματίστη εργασία, ενώ αν η τιμή είναι 0 σημαίνει πως το συγκεκριμένο μηχάνημα δεν παρουσίασε κάποια βλάβη. Αυτό υπολογίζεται από μια στήλη του γενικού συνόλου δεδομένων που ονομάζεται *UnscheduledJobType*, η οποία όταν δεν υπάρχει βλάβη είναι NULL, ενώ όταν υπάρχει γράφει τον κωδικό του τύπου της απρογραμματίστης

δουλειάς. Επομένως για τη δημιουργία της δικής μας στήλης, γίνεται έλεγχος αν το `UnscheduledJobType` είναι `NULL` τότε να πάρει την τιμή 0, αλλιώς 1.

- `Vessel`: το όνομα του πλοίου
- `PreviousDateDone`: η προηγούμενη ημερομηνία στην οποία έγινε η εργασία συντήρησης του μηχανήματος.
- `IntervalDays`: οι ημέρες που μεσολάβησαν από την προηγούμενη ημερομηνία λήξης της εργασίας συντήρησης μέχρι την παρούσα ημερομηνία λήξης. Ο ακέραιος αριθμός των ημερών αυτών υπολογίζεται από τη διαφορά $DueDate - PreviousDueDate$.

Τις ίδιες ακριβώς στήλες, με αυτές που εξηγήθηκαν παραπάνω, περιλαμβάνουν και το σύνολο δεδομένων των προγραμματισμένων και των απρογραμματίστων εργασιών.

Ύστερα από αρκετή έρευνα ως προς την αξιοποίηση αυτού του συνόλου δεδομένων, αποφασίστηκε ότι εφόσον το πιο σημαντικό στοιχείο του συνόλου είναι οι ημερομηνίες, θα ήταν ενδιαφέρον να μετατραπεί το παραπάνω σύνολο δεδομένων σε χρονοσειρά. Ο τρόπος μετατροπής των δεδομένων σε δεδομένα χρονοσειράς θα αναλυθεί στο κεφάλαιο 5.2.

Το νέο σύνολο δεδομένων περιέχει μηνιαίες εγγραφές από τον Ιανουάριο του 2017 έως τον Ιούνιο του 2023, με διάφορα στατιστικά στοιχεία που σχετίζονται με τις καθυστερήσεις των εργασιών συντήρησης, τις έγκαιρες εργασίες και τις ημέρες που μεσολαβούν μεταξύ των εργασιών. Ακολουθεί λεπτομερής περιγραφή της κάθε στήλης:

- `YearMonth`: Αυτή η στήλη προσδιορίζει το έτος και τον μήνα για κάθε εγγραφή στο σύνολο δεδομένων. Είναι στη μορφή `YYYY-MM` (π.χ. `2017-01` για τον Ιανουάριο του 2017).
- `NumOfDefects`: Ο ακέραιος αριθμός των ελαττωμάτων που αναφέρθηκαν σε ένα μήνα.
- `MeanDelayedJobs`: Ο μέσος αριθμός καθυστερημένων εργασιών ανά μήνα.
- `MeanOnTimeJobs`: Ο μέσος αριθμός των εργασιών που γίνανε στην ώρα τους, δηλαδή πριν την καταληκτική ημερομηνία τους, ανά μήνα.
- `StdDelayedJobs`: Αυτή η στήλη παρέχει την τυπική απόκλιση του αριθμού των καθυστερημένων εργασιών ανά μήνα. Η τυπική απόκλιση μετρά το μέγεθος της διακύμανσης στις καθυστερημένες εργασίες.
- `StdOnTimeJobs`: Η τυπική απόκλιση του αριθμού των έγκαιρων εργασιών ανά μήνα.
- `MaxDelayedJobs`: Ο μέγιστος αριθμός των καθυστερημένων εργασιών που καταγράφηκαν σε ένα μήνα.
- `MinOnTimeJobs`: Ο ελάχιστος αριθμός των εργασιών που έγιναν στην ώρα τους, ανά μήνα.
- `MeanIntervalDays`: Ο μέσος αριθμός ημερών μεταξύ των εργασιών κάθε μήνα. Η στήλη αυτή αντιπροσωπεύει το μέσο διάστημα μεταξύ διαδοχικών εργασιών.
- `StdIntervalDays`: Η τυπική απόκλιση του αριθμού των ημερών μεταξύ των εργασιών για κάθε μήνα.
- `MaxIntervalDays`: Ο μέγιστος αριθμός των ημερών μεταξύ των εργασιών που καταγράφονται σε ένα μήνα.
- `MinIntervalDays`: Ο ελάχιστος αριθμός ημερών μεταξύ των εργασιών που καταγράφονται σε ένα μήνα. Αυτή η στήλη αντιπροσωπεύει το μικρότερο διάστημα μεταξύ διαδοχικών εργασιών κατά τη διάρκεια του μήνα.

Αυτό το σύνολο δεδομένων προσφέρει μια ολοκληρωμένη εικόνα των μετρικών απόδοσης των εργασιών για αρκετά χρόνια, συμπεριλαμβανομένου του αριθμού των ελαττωμάτων, του μέσου όρου των καθυστερημένων και των έγκαιρων εργασιών, των τυπικών αποκλίσεων τους και των διαστημάτων μεταξύ των εργασιών.

Σκοπός αυτής τη εργασίας είναι η πρόβλεψη του αριθμού των βλαβών στο μηχάνημα CE-DE, 12V50DF, για τους επόμενους μήνες, ώστε η εταιρεία να είναι προετοιμασμένη για κάθε απροσδόκητη βλάβη. Να έχει παραγγείλει δηλαδή παραπάνω εξαρτήματα, ώστε να μπορέσει να διαχειριστεί εγκαίρως τον προβλεπόμενο αριθμό βλαβών που μπορεί να προκύψουν στο συγκεκριμένο μηχάνημα. Συνεπώς, από αυτό το σύνολο δεδομένων, για την παρούσα εργασία, θα αξιοποιηθούν μόνο οι στήλες YearMonth και NumOfDefects για την ανάλυση και την πρόβλεψη τιμών της χρονοσειράς. Οι ίδιες αναλύσεις πάντως, μπορούν κατ'επέκταση να γίνουν και στις υπόλοιπες στήλες του συνόλου δεδομένων.

5.2 Δημιουργία Συνόλου Δεδομένων

Όπως αναφέρθηκε προηγουμένως, επειδή με το αρχικό σύνολο δεδομένων του συστήματος προγραμματισμένης συντήρησης, δε γινόταν να γίνει κάποια ανάλυση δεδομένων και να εφαρμοστεί κάποιο προβλεπτικό μοντέλο, αποφασίστηκε να μετατραπεί αυτό το σύνολο δεδομένων σε χρονοσειρά. Στόχος του νέου συνόλου δεδομένων ήταν να εφαρμοστούν κατάλληλα μοντέλα χρονοσειρών, με σκοπό να προβλέπεται για τους επόμενους έξι μήνες πόσες βλάβες περίπου μπορεί να προκύψουν συγκεκριμένα στα μηχανήματα CE-DE, 12V50DF, τα οποία υπάρχουν σε όλο το στόλο. Επειδή ενδιαφερόμαστε για την πρόβλεψη βλάβης, αποφασίστηκε ότι το σύνολο δεδομένων που περιέχει μόνο τις προγραμματισμένες εργασίες συντήρησης δε χρειάζεται για την παρούσα εργασία. Ο λόγος είναι ότι τα δεδομένα με τις προγραμματισμένες εργασίες δεν περιέχουν δεδομένα όπου προέκυψε κάποια απρόσμενη βλάβη, διότι είναι καθαρά προγραμματισμένες εργασίες οι οποίες πρέπει να γίνονται ανά συγκεκριμένα χρονικά διαστήματα η καθεμία. Αφού θέλουμε να συγκεντρώσουμε το ενδιαφέρον μας στις απρόβλεπτες βλάβες που προκύπτουν, καθώς αυτό τον αριθμό θα προσπαθήσουμε να προβλέψουμε μελλοντικά, μας είναι χρήσιμο μόνο το σύνολο δεδομένων με τις απρόσμενες εργασίες συντήρησης. Αυτό το σύνολο δεδομένων περιέχει τις εργασίες συντήρησης και την ημερομηνία που έγινε η καθεμία, για κάθε βλάβη που προέκυψε στο συγκεκριμένο μηχάνημα.

Επομένως από το σύνολο δεδομένων των απρογραμματίστων εργασιών συντήρησης κρατήσαμε μόνο τις στήλες ShipComponentID, DueDate, DateDone, PreviousDueDate, IntervalDays, Vessel και δημιουργήσαμε μια παραπάνω στήλη με όνομα DateDifference η οποία δείχνει τη διαφορά μεταξύ της ημέρας που έγινε η εργασία (DateDone) με την ημέρα που έληγε το χρονικό περιθώριο να γίνει αυτή η δουλειά (DueDate). Άρα $DateDifference = DateDone - DueDate$.

Αν $DateDifference = DateDone - DueDate < 0$ σημαίνει ότι η εργασία έγινε έγκαιρα.

Αν $DateDifference = DateDone - DueDate > 0$ σημαίνει ότι η εργασία καθυστέρησε να γίνει.

Ύστερα χρησιμοποιώντας τη βιβλιοθήκη pandas της Python, υπολογίστηκαν οι στήλες της χρονοσειράς, όπως παρουσιάστηκαν στην παραπάνω παράγραφο. Αρχικά, εξήχθη το έτος και ο μήνας από τη στήλη DateDone και δημιουργήθηκε μια νέα στήλη η YearMonth, η οποία περιέχει σε κάθε γραμμή μόνο το μήνα και το χρόνο και αδιαφορούμε για την ημέρα (π.χ. 2017-01). Έπειτα ομαδοποιήθηκαν τα δεδομένα του αρχικού συνόλου ανά μήνα και χρόνο και αθροίστηκαν οι αριθμοί βλαβών για κάθε μήνα ξεχωριστά, και έτσι δημιουργήθηκε η στήλη NumOfDefects. Για τη δημιουργία της στήλης MeanDelayedJobs, έγινε φιλτράρισμα στα δεδομένα των απρογραμματίστων εργασιών, μόνο για τις καθυστερημένες εργασίες (όπου DateDifference > 0) και στη συνέχεια υπολογίστηκε η μέση τιμή της στήλης DateDifference για κάθε μήνα. Για τη στήλη MeanOnTimeJobs, φιλτράρονται τα δεδομένα ώστε να περιλαμβάνονται μόνο οι εργασίες που εκτελούνται στην ώρα τους (όπου DateDifference ≤ 0) και στη συνέχεια υπολογίζεται η μέση τιμή της στήλης DateDifference για κάθε μήνα. Για τον

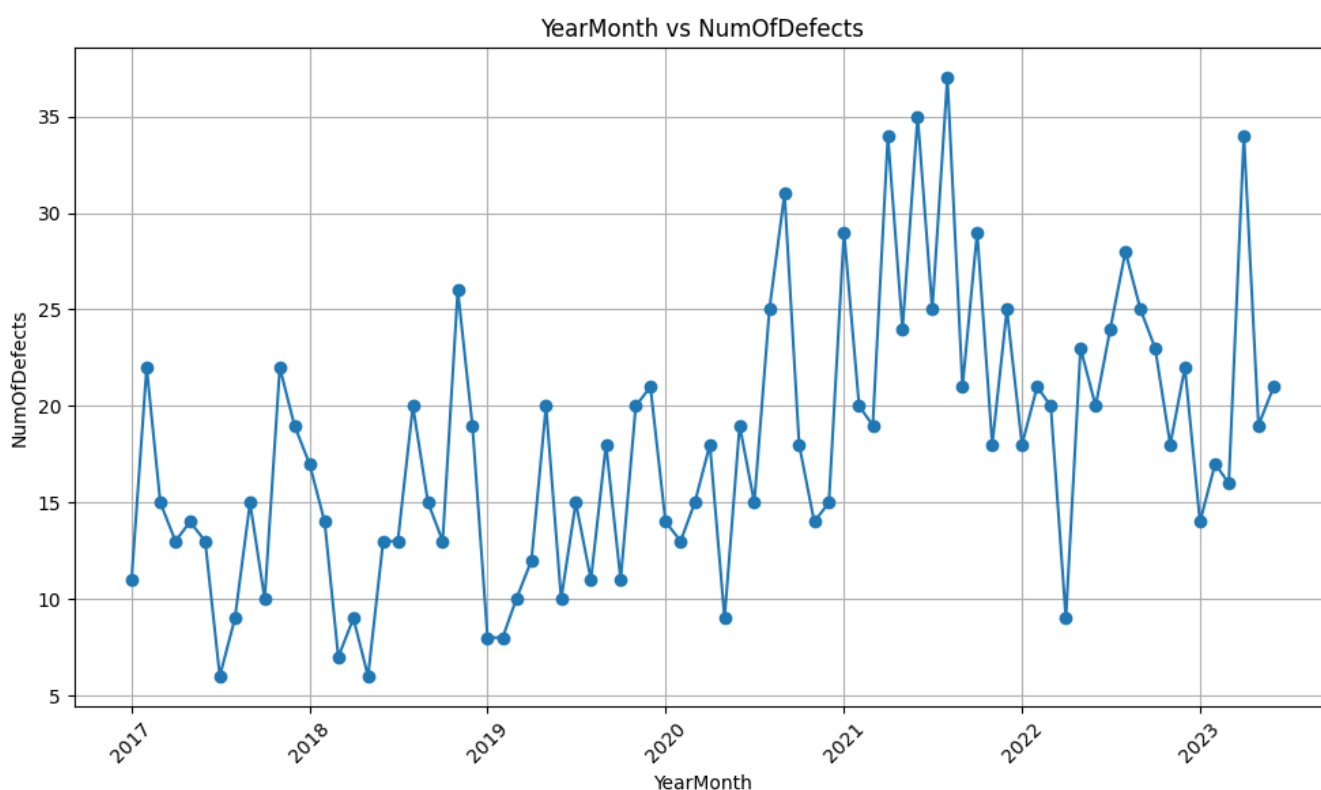
υπολογισμό της στήλης StdDelayedJobs φιλτράρονται τα δεδομένα ώστε να περιλαμβάνονται μόνο οι καθυστερημένες εργασίες (όπου DateDifference > 0) και υπολογίζεται η τυπική απόκλιση της στήλης DateDifference για κάθε μήνα. Αντίστοιχα, για τη στήλη StdOnTimeJobs υπολογίζεται η τυπική απόκλιση των εργασιών που εκτελούνται εγκαίρως (όπου DateDifference ≤ 0). Για τον υπολογισμό της στήλης MaxDelayedJobs φιλτράρονται τα δεδομένα ώστε να περιλαμβάνονται μόνο οι γραμμές των καθυστερημένων εργασιών, δηλαδή όπου DateDifference > 0, ύστερα ομαδοποιούνται τα δεδομένα με βάση το YearMonth και υπολογίζεται το μέγιστο DateDifference για κάθε ομάδα. Αντίστοιχα για τη στήλη MinDelayedJobs, φιλτράρονται τα δεδομένα ώστε να περιλαμβάνονται μόνο οι γραμμές των εργασιών που εκτελούνται στην ώρα τους, δηλαδή όπου DateDifference ≤ 0, έπειτα ομαδοποιούνται τα φιλτραρισμένα δεδομένα ανά YearMonth και υπολογίζεται το ελάχιστο της στήλης DateDifference για κάθε ομάδα. Κατά τον ίδιο τρόπο όπως με τις παραπάνω αντίστοιχες στήλες, υπολογίζονται η μέση τιμή, η τυπική απόκλιση, το μέγιστο και το ελάχιστο της στήλης IntervalDays για κάθε μήνα. Τέλος, γίνεται η ενοποίηση όλων των παραπάνω στηλών και έτσι δημιουργείται το νέο σύνολο δεδομένων που αποτελεί πλέον χρονοσειρά.

6. Εφαρμογή αλγοριθμικών μεθόδων – Αποτελέσματα

Στο κεφάλαιο αυτό παρουσιάζεται η εφαρμογή των αλγοριθμικών μεθόδων για τη μελέτη του συνόλου δεδομένων, καθώς και τα αποτελέσματα που προκύπτουν μέσα από αυτές. Πιο συγκεκριμένα, εφαρμόζεται το μοντέλο ARIMA με διαφορετικές παραμέτρους που επηρεάζουν τις μηνιαίες προβλέψεις των τιμών.

6.1 Πρόβλεψη βλαβών

Σε αυτό το κεφάλαιο εφαρμόζουμε τον αλγόριθμο ARIMA για την πρόβλεψη βλαβών στη μηχανή CE-DE, 12V50DF, η οποία υπάρχει κοινή σε όλο το στόλο. Για να καταλάβουμε τι χρονοσειρά έχουμε, αρχικά κατασκευάζουμε το χρονοδιάγραμμα των βλαβών που έχει εμφανίσει η συγκεκριμένη μηχανή σε όλα τα πλοία συνολικά, από το 2017, που πρωτοεγκαταστάθηκε, έως το 2023.



Εικόνα 4 - Χρονοδιάγραμμα βλαβών της μηχανής CE-DE, 12V50DF

Επιπλέον υπολογίζουμε τη μέση τιμή των αριθμών των βλαβών, η οποία είναι 17.9.

Από το παραπάνω χρονοδιάγραμμα φαίνεται ότι η χρονοσειρά δεν είναι στάσιμη διότι οι τιμές δεν είναι κοντά στη μέση τιμή. Υπάρχει δηλαδή μη στασιμότητα στη μέση τιμή. Επίσης, οι τιμές της σειράς σε «κοντινά» χρονικά σημεία διαφέρουν λίγο, κάτι που σημαίνει υψηλές συσχετίσεις για μικρές υστερήσεις. Αυτό αποτελεί ένδειξη μη στασιμότητας της παραπάνω χρονοσειράς.

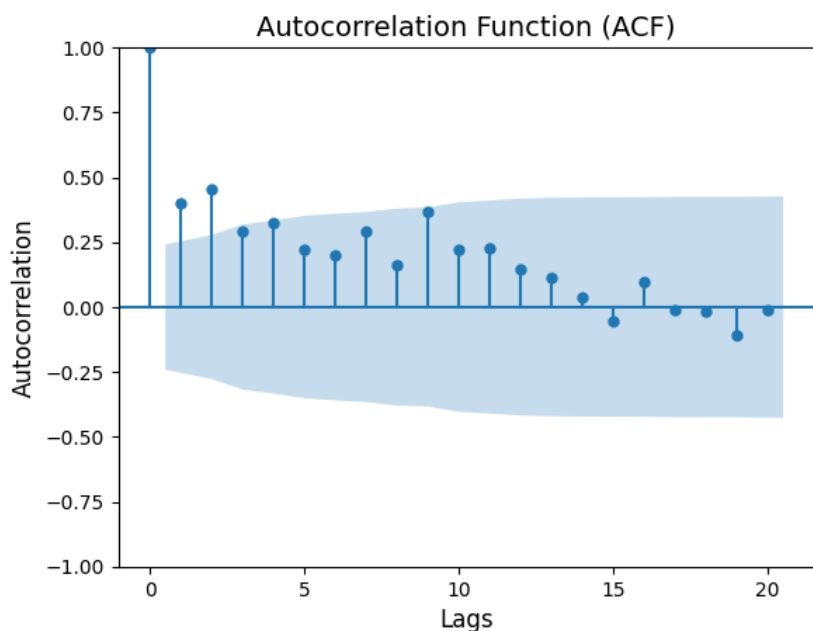
Η μη-στασιμότητα μπορεί να γίνει αντιληπτή και από τους συντελεστές αυτοσυσχέτισης και μερικής αυτοσυσχέτισης που υπολογίζονται παρακάτω.

Για να υπολογίσουμε τους παραπάνω συντελεστές, αρχικά χωρίζουμε το σύνολο δεδομένων σε σύνολο δεδομένων εκπαίδευσης (train dataset) και σε σύνολο δεδομένων ελέγχου (test dataset). Το σύνολο εκπαίδευσης αποτελείται από το 85% του συνόλου δεδομένων, ενώ το σύνολο ελέγχου από το υπόλοιπο 15%.

Ο συντελεστής αυτοσυσχέτισης του συνόλου δεδομένων εκπαίδευσης φαίνεται στο παρακάτω γράφημα

Πίνακας 2 - Πίνακας τιμών ACF του συνόλου δεδομένων

και οι τιμές του παρουσιάζονται στον παρακάτω πίνακα:



Εικόνα 5 - Γραφική Παράσταση του Συντελεστή Αυτοσυσχέτισης

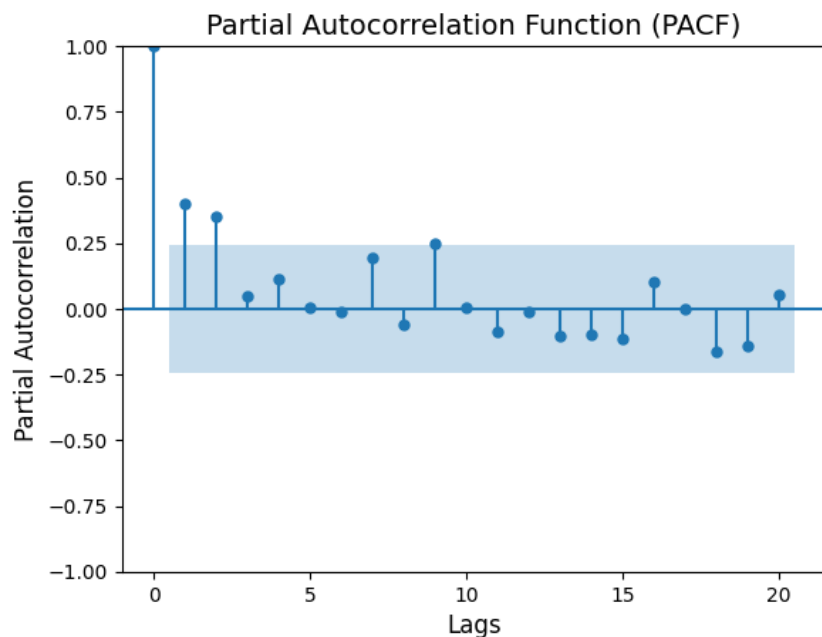
Lag	ACF
0	1.
1	0.40238412
2	0.45499005
3	0.2941763
4	0.32571941
5	0.22373306
6	0.2011576
7	0.29334208
8	0.1622193
9	0.36765993
10	0.22039932
11	0.22722345
12	0.14793533
13	0.1136019
14	0.03975571
15	-0.0543356
16	0.09934556
17	-0.01107156
18	-0.0191728
19	0.11082981
20	-0.0106657

Όπως φαίνεται στο παραπάνω διάγραμμα και στον πίνακα, οι συντελεστές αυτοσυσχέτισης «φθίνουν» πολύ αργά στο μηδέν, ενδεικτικό της μη-στασιμότητας της σειράς.

Ανάλογα υπολογίζονται και οι συντελεστές μερικής αυτοσυσχέτισης του συνόλου δεδομένων εκπαίδευσης όπως φαίνεται στο παρακάτω γράφημα και πίνακα.

Πίνακας 3 - Πίνακας τιμών PACF

Lag	PACF
0	1.
1	0.40857464
2	0.36284637
3	0.05223794
4	0.12078219
5	0.00809632
6	-0.01017326
7	0.21973126
8	-0.06376315
9	0.29859787
10	0.00791888
11	-0.10027078
12	-0.00767999
13	-0.13537334
14	-0.13019589
15	-0.16445077
16	0.13431173
17	-0.00963414
18	-0.24417646
19	-0.22615511
20	0.06647265

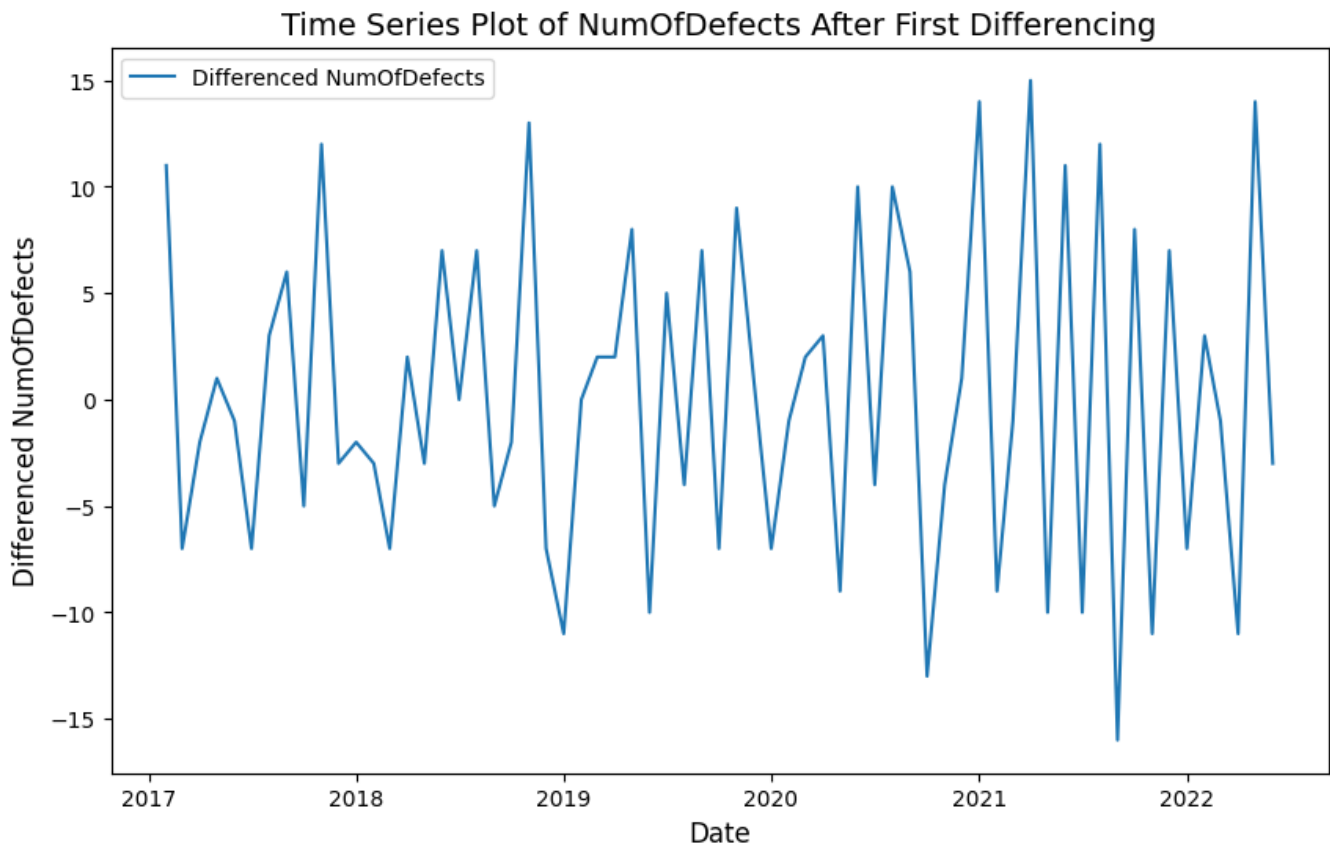


Εικόνα 6 - Γραφική Παράσταση του συντελεστή Μερικής Αυτοσυσχέτισης

Όμοια παρατηρούμε πως και οι συντελεστές μερικής αυτοσυσχέτισης «φθίνουν» αργά στο μηδέν, επομένως συμπεραίνουμε και με αυτόν τον τρόπο πως η χρονοσειρά δεν είναι στάσιμη [3].

Αφού η σειρά δεν είναι στάσιμη, σχηματίζουμε τη σειρά των πρώτων διαφορών. Η διαφορά μπορεί να συμβάλει στη σταθεροποίηση του μέσου όρου μιας χρονοσειράς αφαιρώντας τις μεταβολές στο επίπεδο μιας χρονοσειράς και κατά συνέπεια την εξάλειψη (ή τη μείωση) της τάσης και της εποχικότητας [4].

Το χρονοδιάγραμμα της σειράς μετά την πρώτη διαφορά παρουσιάζεται παρακάτω.



Εικόνα 7 - Το χρονοδιάγραμμα της σειράς μετά την πρώτη διαφορά

Με μια πρώτη ματιά παρατηρούμε ότι η χρονοσειρά τώρα φαίνεται στάσιμη, διότι οι τιμές βρίσκονται γύρω από τη μέση τιμή, η οποία μετά τις πρώτες διαφορές είναι 0.14, και οι διασπορές είναι περίπου ίσες.

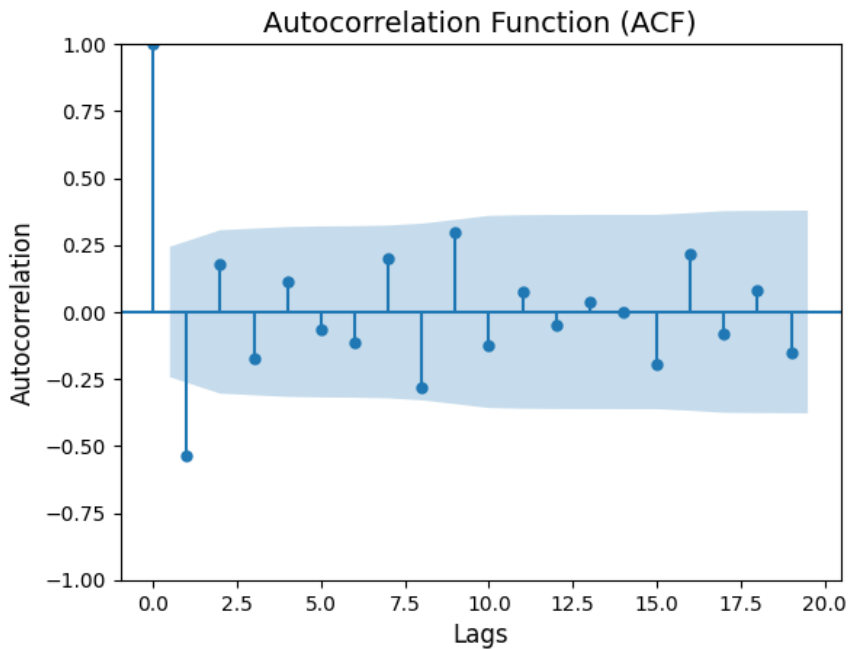
Για να εξετάσουμε την στασιμότητά της κατασκευάζουμε πάλι τα διαγράμματα των συντελεστών αυτοσυσχέτισης και μερικής αυτοσυσχέτισης, όπως φαίνονται παρακάτω.

Πίνακας 4 - Πίνακας τιμών ACF μετά την 1^η διαφορά

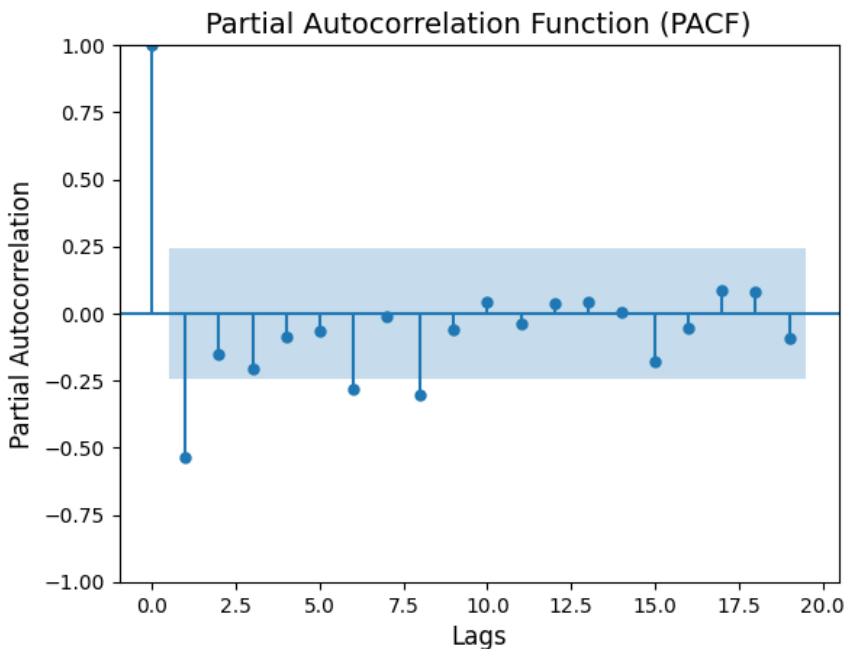
Lag	ACF
0	1.
1	-0.53502084
2	0.17967332
3	-0.17447969
4	0.11376303
5	-0.06694815
6	-0.11449658
7	0.1988094
8	-0.28380496
9	0.29594777
10	-0.12304228
11	0.07645222
12	-0.04932594
13	0.03609433
14	-0.0026294
15	-0.19619561
16	0.21750088
17	-0.08238612
18	0.08076665
19	-0.15017616
20	-0.00310026

Πίνακας 5 - Πίνακας τιμών PACF μετά την 1^η διαφορά

Lag	PACF
0	1.
1	-0.54338054
2	-0.15592354
3	-0.2199107
4	-0.09922637
5	-0.07667298
6	-0.31797098
7	-0.02086466
8	-0.37274424
9	-0.10226355
10	0.01814209
11	-0.08010612
12	0.03325322
13	0.04003921
14	0.01033509
15	-0.25791038
16	-0.10665685
17	0.0983385
18	0.10807202
19	-0.14717655
20	-0.53821839



Εικόνα 8 - Γραφική Παράσταση του Συντελεστή Αυτοσυσχέτισης μετά τις πρώτες διαφορές



Εικόνα 9 - Γραφική Παράσταση του Συντελεστή Μερικής Αυτοσυσχέτισης μετά τις πρώτες διαφορές

Από τα παραπάνω γραφήματα φαίνεται πως οι τιμές φθίνουν άμεσα στο μηδέν, άρα η σειρά γίνεται στάσιμη μετά τις πρώτες διαφορές.

Από τα ACF και PACF γραφήματα της χρονοσειράς, πριν εφαρμοστούν οι πρώτες διαφορές, φαίνεται ότι ο αριθμός των lags ξεκινάει από το 0. Το lag0 δεν το λαμβάνουμε υπόψιν γιατί δείχνει την αυτοσυσχέτιση της τιμής y (NumOfDefects) με τον εαυτό της, και για αυτό το λόγο πάντα το Lag0 είναι 1.

Το μπλε πλαίσιο σε αυτά τα δύο γραφήματα είναι το διάστημα εμπιστοσύνης. Για να βρούμε ποιες παραμέτρους p και q θα χρησιμοποιήσουμε για το μοντέλο ARIMA(p,d,q) κοιτάζουμε πόσα lags βρίσκονται εκτός τους διαστήματος εμπιστοσύνης στα γραφήματα PACF και ACF αντίστοιχα. Στην περίπτωση μας, και στα δύο γραφήματα, υπάρχουν δύο γραμμές, εξαιρουμένου του Lag0, οι οποίες προεξέχουν από το διάστημα εμπιστοσύνης. Επομένως, θεωρούμε ότι για το ARIMA μοντέλο, καλές παράμετροι μπορεί να είναι $p=2$ και $q=2$.

Από τη βιβλιοθήκη statsmodels χρησιμοποιούμε το μοντέλο ARIMA στο σύνολο δεδομένων εκπαίδευσης (`df_train`), το οποίο αποτελείται από 70 δεδομένα. Πιο συγκεκριμένα, το σύνολο εκπαίδευσης αποτελείται από δύο στήλες, το μήνα και έτος και τον αριθμό των βλαβών στο μηχανήμα του πλοίου που μελετάμε.

NumOfDefects	
YearMonth	
2017-01	11
2017-02	22
2017-03	15
2017-04	13
2017-05	14
...	...
2022-06	20
2022-07	24
2022-08	28
2022-09	25
2022-10	23

Εικόνα 10 - Σύνολο δεδομένων εκπαίδευσης (train dataset)

Το σύνολο δεδομένων ελέγχου (`df_test`) περιέχει μόνο ένα δεδομένο. Λόγω του μικρού πλήθους δεδομένων και του ίδιου του μοντέλου, το ARIMA δε μπορεί να προβλέψει πολύ μελλοντικές τιμές. Στην συγκεκριμένη εργασία, για τους υπόλοιπους 12 μήνες προβλέπει περίπου τον ίδιο αριθμό βλαβών, με μικρές διαφορές στα δεκαδικά ψηφία του αριθμού πρόβλεψης. Επομένως, σαν μέθοδο πρόβλεψης αποφασίστηκε να προβλέπουμε μόνο μία τιμή, δηλαδή κάθε μήνα ξεχωριστά. Για να το κάνουμε αυτό, «γεμίζουμε» το σύνολο εκπαίδευσης με 69 τιμές, δηλαδή με τα δεδομένα από το Γενάρη του 2017 έως τον Οκτώβρη του 2022, και το σύνολο ελέγχου με την 70^η τιμή, δηλαδή τον Νοέμβρη του 2022. Σκοπός είναι το μοντέλο ARIMA να εκπαιδευτεί με το σύνολο των δεδομένων του συνόλου εκπαίδευσης και ύστερα να προβλέψει όσο πιο πλησιέστερη τιμή γίνεται για το σύνολο ελέγχου. Για την επόμενη πρόβλεψη, προσθέτουμε τον Νοέμβρη του 2022 στο σύνολο εκπαίδευσης και στο σύνολο ελέγχου βάζουμε την 71^η τιμή του συνόλου δεδομένων, που είναι ο Δεκέμβρης του 2022.

6.1.1 Εφαρμογή ARIMA(2,1,2)

Όπως έχουμε προαναφέρει, το ARIMA δέχεται τρεις παραμέτρους, τις p , d και q . Αυτές οι παράμετροι μπορούν να εξηγηθούν ως εξής:

- p είναι ο αριθμός των αυτοπαλίνδρομων όρων (AR)
- d είναι ο αριθμός των μη εποχικών διαφορών (I)
- q είναι ο αριθμός των καθυστερημένων σφαλμάτων πρόβλεψης (MA)

Συνεπώς το ARIMA(p,d,q) θα μπορούσε κι αλλιώς να γραφτεί ως ARIMA(AR, I, MA).

Σε αυτή την παράγραφο θα εφαρμόσουμε στο σύνολο δεδομένων εκπαίδευσης το μοντέλο ARIMA(2,1,2), όπου $p=2$, $d=1$ και $q=2$.

Τα αποτελέσματα αυτού το μοντέλου παρουσιάζονται στην παρακάτω εικόνα:

```
SARIMAX Results
=====
Dep. Variable:      NumOfDefects      No. Observations:      70
Model:             ARIMA(2, 1, 2)      Log Likelihood         -220.525
Date:              Sun, 14 Jul 2024      AIC                    451.051
Time:              17:22:05           BIC                    462.221
Sample:            01-01-2017           HQIC                   455.483
                  - 10-01-2022
Covariance Type:   opg
=====
              coef      std err      z      P>|z|      [0.025      0.975]
-----
ar.L1         -0.1453      0.540      -0.269      0.788      -1.203      0.913
ar.L2          0.2663      0.164      1.622      0.105      -0.056      0.588
ma.L1         -0.5904      0.574     -1.028      0.304      -1.716      0.535
ma.L2         -0.2557      0.469     -0.545      0.586      -1.175      0.664
sigma2        34.4469      6.655      5.176      0.000      21.403      47.491
=====
Ljung-Box (L1) (Q):      0.01      Jarque-Bera (JB):      1.82
Prob(Q):                 0.94      Prob(JB):              0.40
Heteroskedasticity (H): 1.52      Skew:                  0.40
Prob(H) (two-sided):    0.32      Kurtosis:              3.06
=====
```

Εικόνα 11 - Αποτελέσματα Μοντέλου ARIMA(2,1,2)

Ας αναλύσουμε τα αποτελέσματα από τη σύνοψη του μοντέλου ARIMA(2,1,2):

Περίληψη μοντέλου:

- Dep. Variable: NumOfDefects
Αυτή είναι η εξαρτημένη μεταβλητή που προσπαθεί να προβλέψει το μοντέλο.

- No. Observations: 70
Ο αριθμός των σημείων δεδομένων που χρησιμοποιούνται για την προσαρμογή του μοντέλου.
- Μοντέλο: ARIMA(2, 1, 2)
Δείχνει ότι το μοντέλο είναι ένα μοντέλο ARIMA με 2 αυτοπαλίνδρομους όρους (AR), οι διαφορές των δεδομένων έχουν γίνει μία φορά (I) και υπάρχουν 2 όροι κινητού μέσου όρου (MA).
- Log Likelihood (Λογική πιθανοφάνεια): -220.525
Αυτή η τιμή χρησιμοποιείται στον υπολογισμό των AIC και BIC. Υψηλότερες τιμές υποδηλώνουν γενικά καλύτερη προσαρμογή.
- AIC (Akaike Information Criterion): 451.051
Μια χαμηλότερη τιμή AIC υποδηλώνει καλύτερη προσαρμογή του μοντέλου στα δεδομένα, με μια ποινή για τον αριθμό των παραμέτρων που χρησιμοποιούνται. Χρησιμοποιείται για τη σύγκριση μοντέλων.
- BIC (Bayesian Information Criterion): 462.221
Παρόμοια με το AIC, αλλά με μεγαλύτερη ποινή για μοντέλα με περισσότερες παραμέτρους. Χρησιμοποιείται επίσης για τη σύγκριση μοντέλων.
- HQIC (Κριτήριο πληροφοριών Hannan-Quinn): 455.483
Ένα άλλο κριτήριο για την επιλογή μοντέλων, το οποίο επίσης επιβάλλει ποινές για τον αριθμό των παραμέτρων.

Συντελεστές:

- ar.L1: -0.1453
Ο συντελεστής για την πρώτη υστέρηση (lag 1) του αυτοπαλίνδρομου μέρους του μοντέλου. Το τυπικό σφάλμα είναι 0.540 και η τιμή p είναι 0.788, γεγονός που υποδηλώνει ότι ο συντελεστής αυτός δεν είναι στατιστικά σημαντικός.
- ar.L2: 0.2663
Ο συντελεστής για τη δεύτερη υστέρηση του αυτοπαλίνδρομου μέρους. Το τυπικό σφάλμα είναι 0.164 και η τιμή p είναι 0.105, γεγονός που υποδηλώνει οριακή σημαντικότητα.
- ma.L1: -0.5904
Ο συντελεστής για την πρώτη υστέρηση του μέρους του κινητού μέσου όρου. Το τυπικό σφάλμα είναι 0.574 και η τιμή p είναι 0.304, υποδεικνύοντας ότι ο συντελεστής αυτός δεν είναι στατιστικά σημαντικός.
- ma.L2: -0.2557
Ο συντελεστής για τη δεύτερη υστέρηση του μέρους του κινητού μέσου όρου. Το τυπικό σφάλμα είναι 0.469 και η τιμή p είναι 0.586, υποδεικνύοντας ότι ο συντελεστής αυτός δεν είναι στατιστικά σημαντικός.
- sigma2: 34.4469
Η διακύμανση των υπολοίπων (σφαλμάτων) του μοντέλου (residuals). Μια χαμηλότερη τιμή υποδηλώνει μικρότερη διακύμανση των σφαλμάτων.

Διαγνωστικά:

- Ljung-Box (L1) (Q): 0.01

Η στατιστική του Ljung-Box Q-test στην υστέρηση 1. Μια υψηλή τιμή p (Prob(Q) = 0.94) υποδηλώνει ότι δεν υπάρχει σημαντική αυτοσυσχέτιση στα κατάλοιπα στην υστέρηση 1.

- Jarque-Bera (JB): 1.82

Το στατιστικό ελέγχου Jarque-Bera για την κανονικότητα των καταλοίπων. Μια υψηλή τιμή p (Prob(JB) = 0.40) υποδηλώνει ότι τα κατάλοιπα κατανέμονται κανονικά.

- Heteroskedasticity (Ετεροσκεδαστικότητα (H)): 1.52

Το στατιστικό ελέγχου ετεροσκεδαστικότητας. Μια υψηλή p-τιμή (Prob(H) = 0.32) υποδηλώνει ότι η διακύμανση των καταλοίπων είναι σταθερή με την πάροδο του χρόνου (δεν υπάρχει ετεροσκεδαστικότητα).

- Skew (Στρέβλωση): 0.40

Ένα μέτρο της ασυμμετρίας της κατανομής των καταλοίπων.

- Kyrptosis (Κύρτωση): 3.06

Ένα μέτρο της "ουράς" της κατανομής των καταλοίπων. Μια κύρτωση κοντά στο 3 υποδηλώνει κανονική κατανομή.

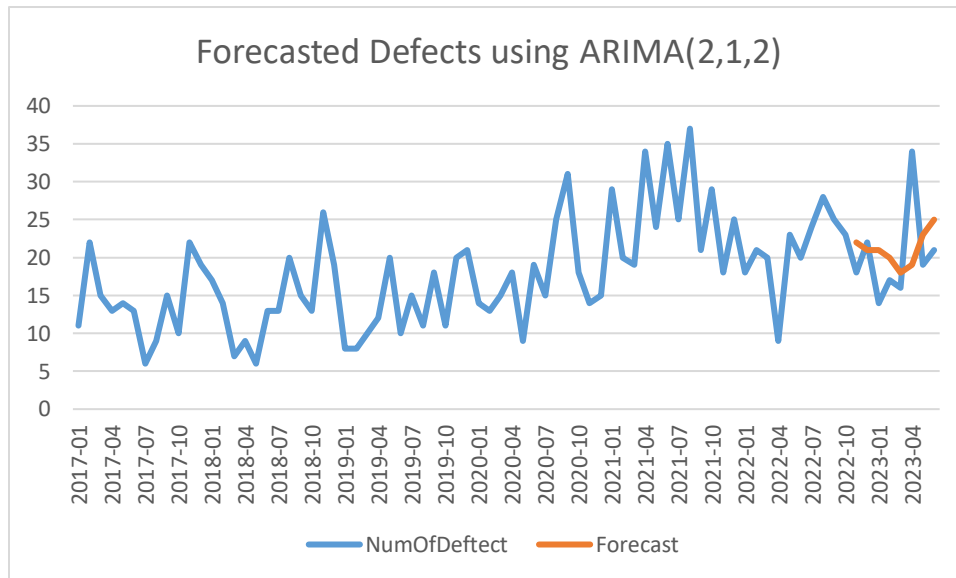
Το μοντέλο ARIMA(2,1,2) προσαρμόστηκε στα δεδομένα και εκτιμήθηκαν οι συντελεστές για τους όρους AR και MA. Οι περισσότεροι από τους συντελεστές AR και MA δεν είναι στατιστικά σημαντικοί (p-values > 0.05), γεγονός που υποδηλώνει ότι μπορεί να μην είναι απαραίτητοι. Τα κατάλοιπα (residuals) του μοντέλου φαίνεται να είναι κανονικά κατανεμημένα και δεν παρουσιάζουν σημαντική αυτοσυσχέτιση ή ετεροσκεδαστικότητα, υποδεικνύοντας καλή προσαρμογή. Τα κριτήρια πληροφοριών (AIC, BIC, HQIC) μπορούν να χρησιμοποιηθούν για τη σύγκριση αυτού του μοντέλου με άλλα πιθανά μοντέλα για την επιλογή του καλύτερου. Δεδομένου ότι οι όροι AR και MA δεν είναι στατιστικά σημαντικοί, ίσως θα έπρεπε να εξεταστούν απλούστερα μοντέλα (όπως ARIMA(1,1,1) ή ακόμη και ARIMA(0,1,1)) και να συγκριθούν οι τιμές AIC και BIC ώστε να βρεθεί ένα δυνητικά καλύτερο μοντέλο.

Ο αριθμός των βλαβών της μηχανής για το μήνα Νοέμβριο του 2022 ήταν 18 για όλα τα πλοία συνολικά. Το μοντέλο ARIMA(2,1,2) προέβλεψε για την ίδια περίοδο 22 βλάβες. Παρακάτω παρουσιάζεται ένας πίνακας με τις κανονικές τιμές και τις τιμές προβλεψής.

Πίνακας 6 - Τιμές Πρόβλεψης μοντέλου ARIMA(2,1,2)

Μήνας	Αρ. Βλαβών	Πρόβλεψη Αρ. Βλαβών
11/2022	18	22
12/2022	22	21
01/2023	14	21
02/2023	17	20
03/2023	16	18
04/2023	34	19
05/2023	19	23
06/2023	21	25

Παρακάτω παρουσιάζεται το γράφημα των πραγματικών τιμών και αυτών που προβλέφθηκαν απο το Νοέμβρη του 2022 μέχρι και τον Ιούνη του 2023.



Εικόνα 12 - Γράφημα προβλέψεων του μοντέλου ARIMA(2,1,2)

Όπως φαίνεται και από τον παραπάνω πίνακα και απο το γράφημα, υπάρχουν ασυνέπειες στην ακρίβεια των προβλέψεων, με ορισμένους μήνες να παρουσιάζουν μεγαλύτερες αποκλίσεις. Η ανάλυση των πραγματικών και των προβλεπόμενων τιμών αποκαλύπτει ότι ενώ το μοντέλο παρέχει λογικές προβλέψεις, υπάρχουν περιπτώσεις σημαντικών αποκλίσεων, ιδίως τον Απρίλιο του 2023. Το μοντέλο αποτυπώνει τη γενική τάση, αλλά χρειάζεται βελτίωση για να βελτιώσει την ακρίβειά του για τους μήνες με μεγάλη μεταβλητότητα στην εμφάνιση βλαβών.

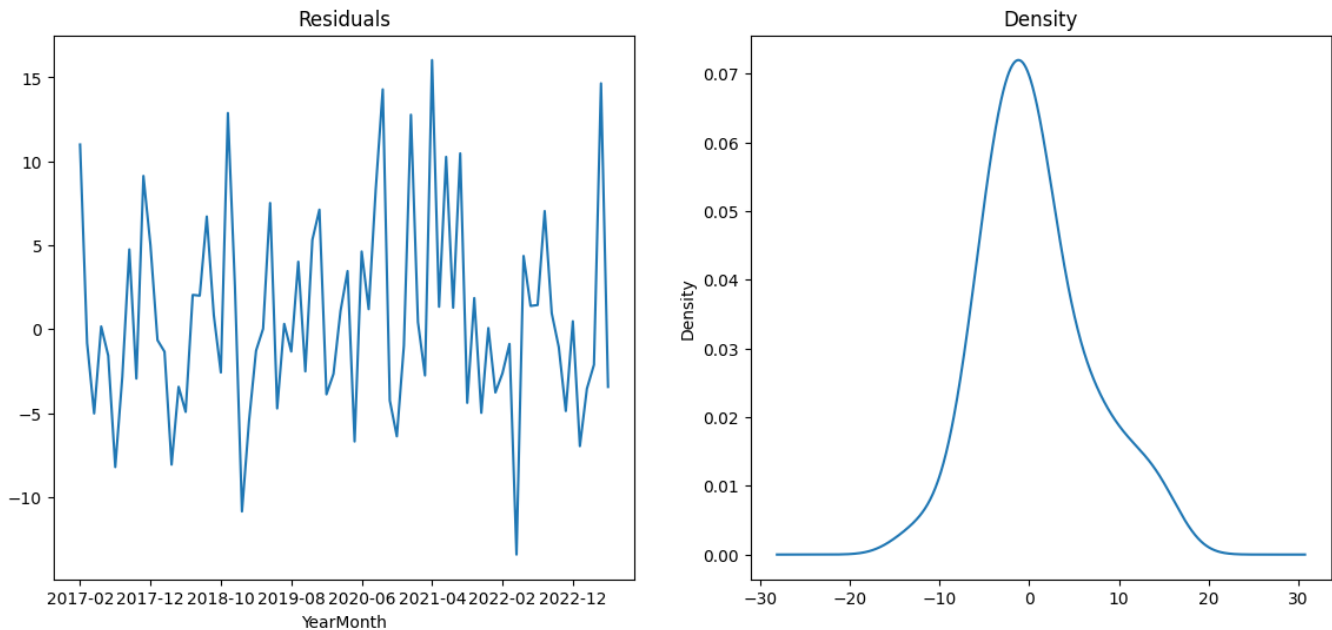
Τα κατάλοιπα μετά από ένα μοντέλο χρονοσειράς αντιπροσωπεύουν τις διαφορές που απομένουν μετά την προσαρμογή ενός μοντέλου. Τα κατάλοιπα είναι οι διαφορές μεταξύ των παρατηρούμενων τιμών και των αντίστοιχων προσαρμοσμένων τιμών τους, και ο μαθηματικός τους τύπος είναι:

$$\varepsilon_t = y_t - \hat{y}_t$$

Είναι σημαντικά για την αξιολόγηση του κατά πόσον ένα μοντέλο έχει καταγράψει αποτελεσματικά την πληροφορία των δεδομένων. Μια αποτελεσματική μέθοδος πρόβλεψης θα παράγει κατάλοιπα που θα παρουσιάζουν τα ακόλουθα χαρακτηριστικά:

1. Τα κατάλοιπα είναι ασυσχέτιστα. Εάν υπάρχουν συσχετίσεις μεταξύ των καταλοίπων, αυτό υποδηλώνει ότι κάποια πληροφορία έχει διαφύγει από το μοντέλο και θα πρέπει να ενσωματωθεί στις προβλέψεις.
2. Τα κατάλοιπα έχουν μηδενική μέση τιμή. Εάν ο μέσος όρος των καταλοίπων αποκλίνει από το μηδέν, αυτό σημαίνει ότι οι προβλέψεις είναι μεροληπτικές [7].

Για το μοντέλο ARIMA(2,1,2) τα γραφήματα των καταλοίπων και της πυκνότητάς τους παρουσιάζονται στις ακόλουθες δύο εικόνες.



Εικόνα 13 - Γράφημα των Καταλοίπων και της Πυκνότητας του ARIMA(2,1,2)

Αριστερά φαίνεται το γράφημα των καταλοίπων (Residuals). Στον άξονα x είναι οι χρονολογίες που περιέχονται στο σύνολο εκπαίδευσης, ενώ στον άξονα y είναι οι τιμές των καταλοίπων, δηλαδή οι διαφορές μεταξύ των πραγματικών αριθμών των βλαβών μείον των προβλεπόμενων τιμών από το μοντέλο ARIMA(2,1,2).

Δεξιά φαίνεται το γράφημα πυκνότητας των καταλοίπων (Density). Στον άξονα x είναι οι τιμές των καταλοίπων, ενώ ο άξονας y αντιπροσωπεύει την πυκνότητα, η οποία είναι ένα μέτρο της κατανομής πιθανότητας των υπολοίπων. Αυτό το γράφημα βοηθά να ελεγχθεί αν τα κατάλοιπα είναι κανονικά κατανομημένα. Για να ακολουθούν κανονική κατανομή πρέπει να έχουν καμπύλη περίπου σε σχήμα καμπάνας. Το σχήμα αυτού του διαγράμματος μπορεί να δώσει πληροφορίες για το αν τα κατάλοιπα έχουν ασυμμετρία ή κύρτωση (ουρά).

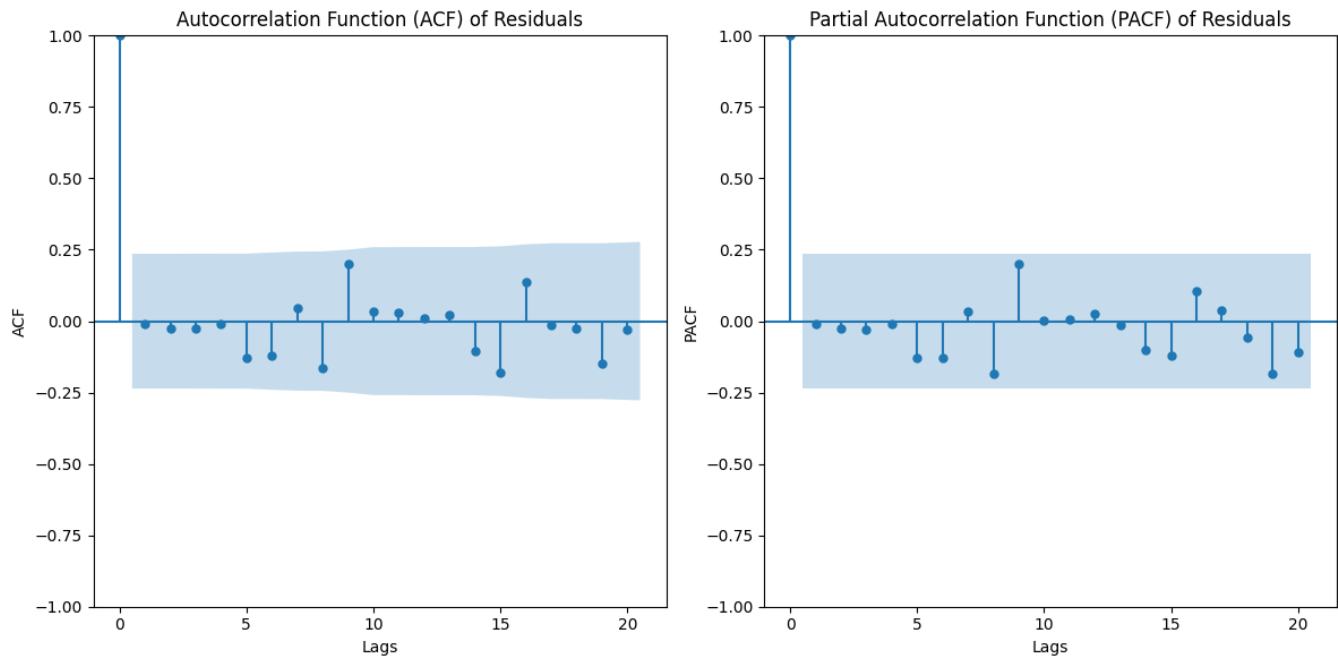
Συνοπτικά, το διάγραμμα καταλοίπων δείχνει τη χρονοσειρά των καταλοίπων, βοηθώντας στη διάγνωση εάν υπάρχουν πρότυπα ή αυτοσυσχέτιση στα κατάλοιπα, ενώ το διάγραμμα πυκνότητας βοηθά να εκτιμηθεί εάν τα κατάλοιπα είναι περίπου κανονικά κατανομημένα, γεγονός που υποστηρίζει τις υποθέσεις του μοντέλου ARIMA.

Σύμφωνα με τα παραπάνω γραφήματα, τα κατάλοιπα φαίνονται τυχαία και κανονικά κατανομημένα, γεγονός που υποδηλώνει ένα καλό μοντέλο ARIMA. Επιπλέον, η μέση τιμή των κατάλοιπων είναι 0,76, δηλαδή αρκετά κοντά στο μηδέν. Επομένως, το μοντέλο αποδίδει αρκετά καλά, αλλά θα μπορούσε να επωφεληθεί από βελτιώσεις για την αντιμετώπιση περιόδων υψηλότερης μεταβλητότητας και πιθανών μη εξεταζόμενων προτύπων.

Δεν αρκούν όμως μόνο τα παραπάνω γραφήματα για να καταλάβουμε αν το ARIMA μοντέλο μας είναι καλό. Πρέπει επιπλέον να μελετήσουμε και τα διαγράμματα Αυτοσυσχέτισης (ACF) και Μερικής Αυτοσυσχέτισης (PACF) των καταλοίπων, τα οποία μας βοηθούν να προσδιορίσουμε αν τα κατάλοιπα από το μοντέλο ARIMA μοιάζουν με «λευκό θόρυβο», δηλαδή αν δεν εμφανίζουν καθόλου αυτοσυσχέτιση [4]. Πιο συγκεκριμένα, όταν μια χρονοσειρά είναι λευκός θόρυβος, κάθε αυτοσυσχέτιση θα είναι κοντά στο μηδέν. Φυσικά, δεν θα είναι

ακριβώς ίσες με το μηδέν, καθώς υπάρχει και κάποια τυχαία διακύμανση. Για μια σειρά λευκού θορύβου, αναμένουμε ότι το 95% των κορυφών (μέγιστων τιμών) στο ACF θα βρίσκεται εντός του διαστήματος $\pm 2/\sqrt{T}$, όπου T είναι το μήκος της χρονικής σειράς. Αυτά τα όρια στην γραφική παράσταση του ACF είναι τα όρια του μπλε πλαισίου στις παρακάτω εικόνες. Εάν μία ή περισσότερες μεγάλες κορυφές βρίσκονται εκτός αυτών των ορίων, ή εάν σημαντικά περισσότερες από το 5% των κορυφών βρίσκονται εκτός αυτών των ορίων, τότε η σειρά πιθανότητα δεν είναι λευκός θόρυβος.

Παρακάτω φαίνονται τα διαγράμματα ACF και PACF των καταλοίπων.



Εικόνα 14 - Διαγράμματα ACF και PACF των Καταλοίπων για το ARIMA(2,1,2)

Το αριστερά διάγραμμα ACF δείχνει τη συσχέτιση των καταλοίπων με τον εαυτό τους σε διαφορετικές χρονικές υστερήσεις (lags). Η συσχέτιση στην υστέρηση 0 είναι πάντα 1, καθώς υπάρχει τέλεια συσχέτιση με τον εαυτό της. Οι υπόλοιπες υστερήσεις βρίσκονται κοντά στο μηδέν και εντός του μπλε διαστήματος εμπιστοσύνης, υποδεικνύοντας ότι δεν υπάρχει σημαντική αυτοσυσχέτιση στις διάφορες υστερήσεις. Αυτό υποδηλώνει ότι τα κατάλοιπα είναι πιθανότατα ασυσχέτιστα, γεγονός που αποτελεί καλή ένδειξη ότι το υπόδειγμα έχει συλλάβει τα υποκείμενα μοτίβα στα δεδομένα.

Το δεξί διάγραμμα PACF δείχνει τη μερική συσχέτιση των καταλοίπων με τον εαυτό τους σε διαφορετικές υστερήσεις, ελέγχοντας τις τιμές των ενδιάμεσων υστερήσεων. Όπως και με την ACF, η συσχέτιση στην υστέρηση 0 είναι πάντα 1. Οι μερικές συσχετίσεις στις υπόλοιπες υστερήσεις είναι κοντά στο μηδέν και εντός του μπλε διαστήματος εμπιστοσύνης, οπότε τα κατάλοιπα θεωρούνται λευκός θόρυβος.

Συνεπώς, τα κατάλοιπα από το μοντέλο ARIMA φαίνεται να μοιάζουν με λευκό θόρυβο, όπως φαίνεται από τα διαγράμματα ACF και PACF. Αυτό υποδηλώνει ότι το υπόδειγμα ARIMA έχει καταγράψει επαρκώς την υποκείμενη δομή της χρονοσειράς. Εφόσον δεν υπάρχουν σημαντικές αυτοσυσχετίσεις στα κατάλοιπα, το μοντέλο ARIMA(2,1,2) φαίνεται να ταιριάζει καλά στα δεδομένα.

6.1.2 Εφαρμογή ARIMA(1,1,1)

Για να βρούμε ποιοι παράμετροι p , d , q ταιριάζουν καλύτερα στο ARIMA μοντέλο που εφαρμόζεται στο σύνολο δεδομένων μας, πρέπει να γίνουν αρκετές δοκιμές, με διαφορετικές παραμέτρους. Σε αυτή την παράγραφο θα παρουσιάσουμε τα αποτελέσματα του μοντέλου ARIMA(1,1,1), δηλαδή του μοντέλου με παραμέτρους $p=1$, $d=1$ και $q=1$.

Ακολουθείται ακριβώς η ίδια διαδικασία όπως με το ARIMA(2,1,2). Χωρίζουμε το σύνολο δεδομένων σε σύνολο εκπαίδευσης (train set) και σε σύνολο ελέγχου (test set). Το σύνολο δεδομένων εκπαίδευσης περιέχει τον αριθμό των βλαβών της μηχανής CE-DE, 12V50DF από τον Ιανουάριο του 2017 έως και τον Οκτώβριο του 2022, όπως φαίνεται στην εικόνα.

YearMonth	NumOfDefects
2017-01	11
2017-02	22
2017-03	15
2017-04	13
2017-05	14
...	...
2022-06	20
2022-07	24
2022-08	28
2022-09	25
2022-10	23

Εικόνα 15 - Σύνολο δεδομένων εκπαίδευσης για ARIMA(1,1,1)

Το σύνολο δεδομένων ελέγχου περιέχει μόνο μία τιμή, τον Νοέμβριο του 2022.

NumOfDefects	18
Name:	2022-11, dtype: int64

Όπως έχει εξηγηθεί και παραπάνω, για κάθε μήνα που προβλέπουμε, το σύνολο εκπαίδευσης παίρνει όλα τα δεδομένα μέχρι και τον προηγούμενο μήνα από αυτόν που θέλουμε να προβλέψουμε, και το σύνολο ελέγχου περιέχει μόνο το μήνα που θέλουμε να προβλέψουμε.

Εφαρμόζουμε το μοντέλο ARIMA(1,1,1) στο σύνολο δεδομένων ελέγχου και τα αποτελέσματα παρουσιάζονται στην παρακάτω εικόνα.

SARIMAX Results						
Dep. Variable:	NumOfDefects	No. Observations:	70			
Model:	ARIMA(1, 1, 1)	Log Likelihood	-222.122			
Date:	Sat, 20 Jul 2024	AIC	450.245			
Time:	15:48:17	BIC	456.947			
Sample:	01-01-2017	HQIC	452.904			
	- 10-01-2022					
Covariance Type:	opg					
	coef	std err	z	P> z	[0.025	0.975]
ar.L1	0.0962	0.187	0.513	0.608	-0.271	0.464
ma.L1	-0.8234	0.086	-9.623	0.000	-0.991	-0.656
sigma2	36.0999	6.924	5.214	0.000	22.530	49.670
Ljung-Box (L1) (Q):		0.06	Jarque-Bera (JB):	1.88		
Prob(Q):		0.80	Prob(JB):	0.39		
Heteroskedasticity (H):		1.70	Skew:	0.39		
Prob(H) (two-sided):		0.21	Kurtosis:	2.78		

Εικόνα 16 - Αποτελέσματα μοντέλου ARIMA(1,1,1)

Ας αναλύσουμε τα αποτελέσματα από τη σύνοψη του μοντέλου ARIMA(1,1,1):

Περίληψη μοντέλου:

- Dep. Variable: NumOfDefects
Αυτή είναι η εξαρτημένη μεταβλητή που προσπαθεί να προβλέψει το μοντέλο.
- No. Observations: 70
Ο αριθμός των σημείων δεδομένων που χρησιμοποιούνται για την προσαρμογή του μοντέλου.
- Μοντέλο: ARIMA(1, 1, 1)
Δείχνει ότι το μοντέλο είναι ένα μοντέλο ARIMA με 1 αυτοπαλίνδρομο όρο (AR), μία σειρά διαφορών των δεδομένων (I) και υπάρχει 1 όρος κινητού μέσου όρου (MA).
- Log Likelihood (Λογική πιθανοφάνεια): -222.122
Αυτή η τιμή χρησιμοποιείται στον υπολογισμό των AIC και BIC. Υψηλότερες τιμές υποδηλώνουν γενικά καλύτερη προσαρμογή.
- AIC (Akaike Information Criterion): 450.245
Μια χαμηλότερη τιμή AIC υποδηλώνει καλύτερη προσαρμογή του μοντέλου στα δεδομένα, λαμβάνοντας υπόψη την πολυπλοκότητα του μοντέλου.

- BIC (Bayesian Information Criterion): 456.904
Παρόμοιο με το AIC, αλλά με μεγαλύτερη ποινή για μοντέλα με περισσότερες παραμέτρους. Χαμηλότερες τιμές BIC υποδηλώνουν ένα καλύτερο μοντέλο.
- HQIC (Κριτήριο πληροφοριών Hannan-Quinn): 452.904
Ένα άλλο κριτήριο για τη σύγκριση μοντέλων, που εξισορροπεί την προσαρμογή και την πολυπλοκότητα.

Συντελεστές:

- ar.L1: 0.0962
Ο συντελεστής για την πρώτη υστέρηση (lag 1) του αυτοπαλίνδρομου μέρους του μοντέλου. Το τυπικό σφάλμα (std err) είναι 0.187, η στατιστική z για το συντελεστή είναι 0.513 και η τιμή p είναι 0.608. Το διάστημα επιστοσύνης 95% για τον συντελεστή είναι [-0.271, 0.464]. Εφόσον το p-value είναι μεγαλύτερο του διαστήματος εμπιστοσύνης, τότε ότι ο συντελεστής αυτός δεν είναι στατιστικά σημαντικός.
- ma.L1: -0.8234
Ο συντελεστής για την πρώτη υστέρηση του μέρους του κινητού μέσου όρου. Το τυπικό σφάλμα (std err) είναι 0.086 και η τιμή p είναι 0, υποδεικνύοντας ότι ο συντελεστής αυτός είναι στατιστικά σημαντικός, διότι p-value < 0.05.
- sigma2: 36.0999
Η διακύμανση των καταλοίπων (σφαλμάτων) του μοντέλου (residuals). Μια χαμηλότερη τιμή υποδηλώνει μικρότερη διακύμανση των σφαλμάτων.

Διαγνωστικά:

- Ljung-Box (L1) (Q): 0.06
Η στατιστική του Ljung-Box Q-test στην υστέρηση 1. Αυτό είναι ένας έλεγχος αυτοσυσχέτισης στα κατάλοιπα. Μια υψηλή τιμή p (Prob(Q) = 0.80) υποδηλώνει ότι δεν υπάρχει σημαντική αυτοσυσχέτιση στα κατάλοιπα στην υστέρηση 1.
- Jarque-Bera (JB): 1.88
Το στατιστικό ελέγχου Jarque-Bera για την κανονικότητα των καταλοίπων. Μια υψηλή τιμή p (Prob(JB) = 0.39) υποδηλώνει ότι τα κατάλοιπα κατανέμονται κανονικά.
- Heteroskedasticity (Ετεροσκεδαστικότητα (H)): 1.70
Το στατιστικό ελέγχου ετεροσκεδαστικότητας. Μια υψηλή p-τιμή (Prob(H) = 0.21) υποδηλώνει ότι η διακύμανση των καταλοίπων είναι σταθερή με την πάροδο του χρόνου (δεν υπάρχει σημαντική ετεροσκεδαστικότητα).
- Skew (Στρέβλωση): 0.39
Ένα μέτρο της ασυμμετρίας της κατανομής των καταλοίπων. Μια στρέβλωση κοντά στο μηδέν υποδηλώνει συμμετρικά κατάλοιπα.
- Kyrptosis (Κύρτωση): 2.78
Ένα μέτρο της «ουράς» της κατανομής των καταλοίπων. Μια κύρτωση κοντά στο 3 υποδηλώνει κανονική κατανομή.

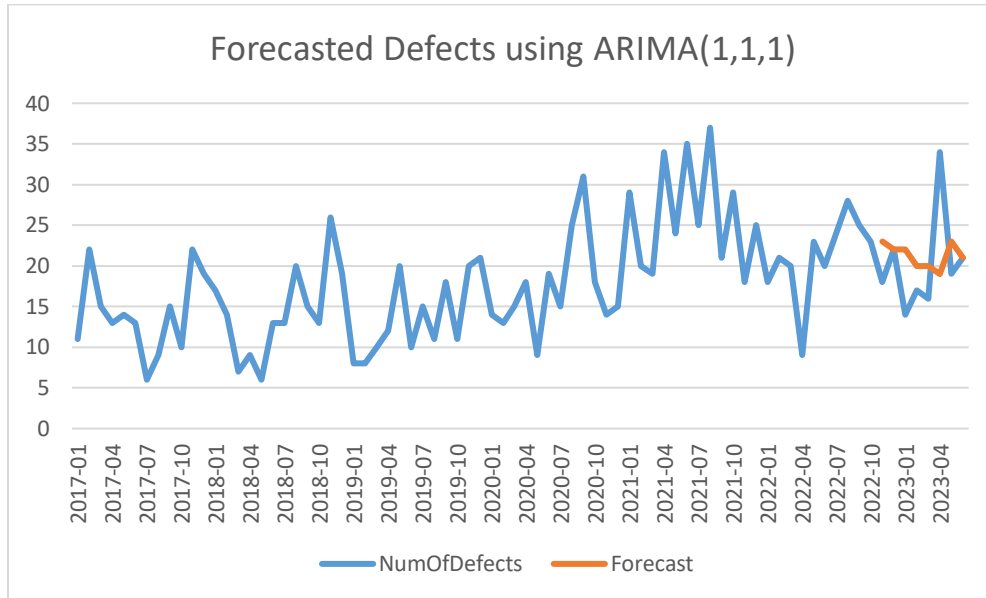
Συνοψίζοντας όλα τα παραπάνω, οι τιμές AIC, BIC και HQIC υποδηλώνουν ότι το μοντέλο είναι αρκετά καλά προσαρμοσμένο. Ο συντελεστή AR (ar.L1) δεν είναι στατιστικά σημαντικός αφού $p\text{-value}=0.608 > 0.05$ γεγονός που υποδηλώνει ότι η αυτοπαλίνδρομη συνιστώσα ρ μπορεί να μην είναι απαραίτητη. Αντιθέτως, ο συντελεστής MA (ma.L1) είναι εξαιρετικά σημαντικός αφού $p\text{-value}=0.000 < 0.05$ υποδεικνύοντας ότι η συνιστώσα του κινητού μέσου όρου είναι σημαντική. Τα κατάλοιπα δεν παρουσιάζουν σημαντική αυτοσυσχέτιση (δοκιμή Ljung-Box), είναι κανονικά κατανεμημένα (δοκιμή Jarque-Bera) και έχουν σταθερή διακύμανση (δοκιμή ετεροσκεδαστικότητας). Αυτές οι διαγνωστικές δοκιμές δείχνουν ότι τα κατάλοιπα συμπεριφέρονται καλά, ικανοποιώντας τις υποθέσεις του ARIMA. Επομένως, ενώ το ARIMA(1,1,1) αποδίδει αρκετά καλά, μπορεί να υπάρχουν περιθώρια βελτίωση με την επανεκτίμηση της αναγκαιότητας της συνιστώσας AR, το οποίο θα αναλυθεί στην επόμενη παράγραφο.

Ο αριθμός των βλαβών της μηχανής για το μήνα Νοέμβριο του 2022 ήταν 18 για όλα τα πλοία συνολικά. Το μοντέλο ARIMA(1,1,1) προέβλεψε για την ίδια περίοδο 23 βλάβες. Παρακάτω παρουσιάζεται ένας πίνακας με τις κανονικές τιμές και τις τιμές προβλεψης.

Πίνακας 7 - Τιμές Πρόβλεψης μοντέλου ARIMA(1,1,1)

Μήνας	Αρ. Βλαβών	Πρόβλεψη Αρ. Βλαβών
11/2022	18	23
12/2022	22	22
01/2023	14	22
02/2023	17	20
03/2023	16	20
04/2023	34	19
05/2023	19	23
06/2023	21	21

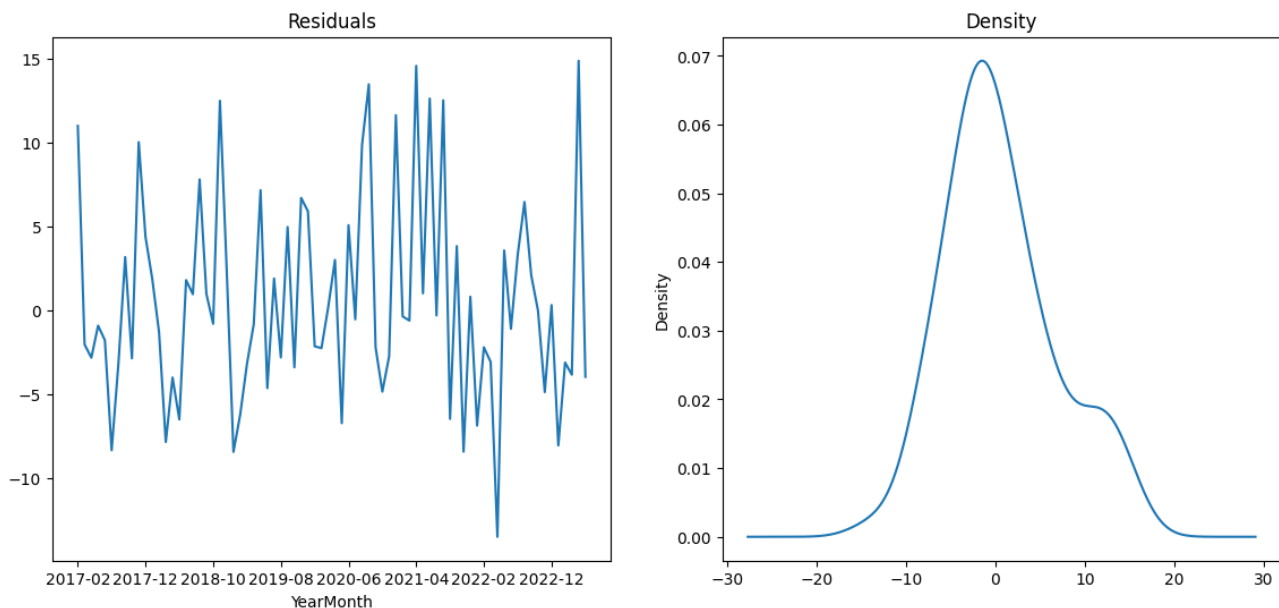
Παρακάτω παρουσιάζεται το γράφημα των πραγματικών τιμών και αυτών που προβλέφθηκαν απο το Νοέμβρη του 2022 μέχρι και τον Ιούνη του 2023.



Εικόνα 17 - Διάγραμμα Προβλέψεων του ARIMA(1,1,1)

Όπως φαίνεται και από τον παραπάνω πίνακα και από το γράφημα, το μοντέλο ARIMA(1,1,1) είναι αποτελεσματικό στην αποτύπωση της γενικής τάσης των δεδομένων. Οι προβλέψεις ευθυγραμμίζονται αρκετά καλά με τις πραγματικές τιμές, υποδεικνύοντας την ικανότητα του μοντέλου να παρέχει χρήσιμες προβλέψεις. Ωστόσο, ορισμένες αποκλίσεις υποδεικνύουν πιθανές περιοχές για βελτίωση του μοντέλου ώστε να αντιμετωπιστεί καλύτερα η υψηλή μεταβλητότητα και οι περιστασιακές κορυφές των δεδομένων.

Σημαντικό κριτήριο για την αξιολόγηση του κατά πόσον ένα μοντέλο έχει καταγράψει αποτελεσματικά την πληροφορία των δεδομένων αποτελούν τα κατάλοιπα. Για το μοντέλο ARIMA(1,1,1) το γράφημα των καταλοίπων (Residuals) και της πυκνότητάς τους (Density) παρουσιάζονται στις δύο ακόλουθες εικόνες.



Εικόνα 18 - Διαγράμματα Καταλοίπων και Πυκνότητας του ARIMA(1,1,1)

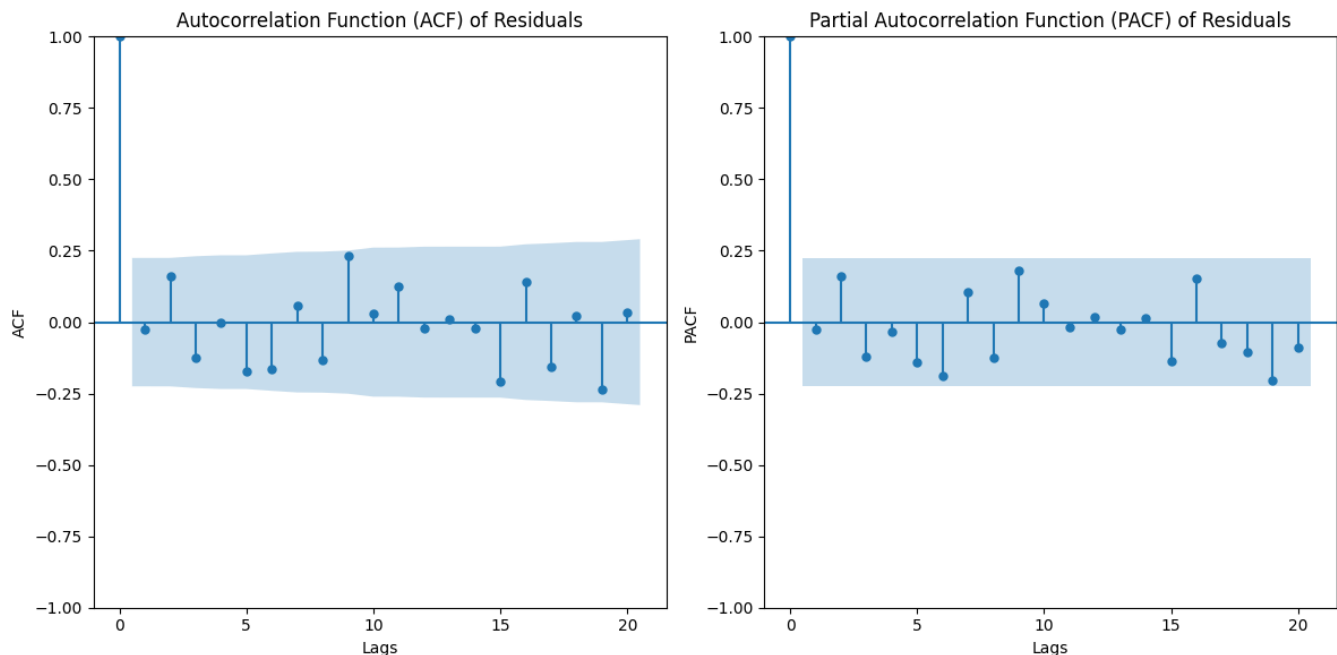
Το διάγραμμα των καταλοίπων δείχνει τις διαφορές μεταξύ των παρατηρούμενων τιμών και των τιμών που προβλέπει το μοντέλο ARIMA(1,1,1) με την πάροδο του χρόνου. Ο άξονας x αντιπροσωπεύει το χρόνο, καλύπτοντας την περίοδο από τις αρχές του 2017 έως το τέλος του 2022, δηλαδή τα δεδομένα που περιέχονται στο σύνολο εκπαίδευσης. Ο άξονας y αντιπροσωπεύει το μέγεθος των καταλοίπων, δηλαδή τις αποκλίσεις των προβλέψεων του μοντέλου από τις πραγματικές προηγούμενες τιμές. Ιδανικά, τα κατάλοιπα θα πρέπει να κατανέμονται τυχαία γύρω από το μηδέν, υποδεικνύοντας ότι δεν υπάρχουν συστηματικά σφάλματα στο μοντέλο. Εδώ, τα κατάλοιπα κυμαίνονται γύρω από το μηδέν, υποδεικνύοντας ότι οι προβλέψεις του μοντέλου είναι άλλοτε πάνω και άλλοτε κάτω από τις πραγματικές τιμές. Ενώ τα κατάλοιπα εμφανίζονται γενικά τυχαία, υπάρχουν περιόδοι όπου τα κατάλοιπα παρουσιάζουν μια «επιμονή» πάνω ή κάτω από το μηδέν, υποδηλώνοντας πιθανή αυτοσυσχέτιση. Η μεταβλητότητα των καταλοίπων φαίνεται αρκετά σταθερή, αλλά υπάρχουν περιόδοι αυξημένης μεταβλητότητας, όπως γύρω στα μέσα του 2020 και στα τέλη του 2022. Η μέση τιμή των καταλοίπων είναι 0.57, αρκετά κοντά στο μηδέν γεγονός που υποδεικνύει πως το ARIMA(1,1,1) είναι ένα καλό μοντέλο για τα δεδομένα μας.

Το διάγραμμα της πυκνότητας δείχνει την κατανομή των καταλοίπων, δίνοντας μια ιδέα της συνολικής διασποράς και της κεντρικής τάσης τους. Ο άξονας x αντιπροσωπεύει την τιμή των καταλοίπων, ενώ ο άξονας y αντιπροσωπεύει την πυκνότητα, η οποία είναι μια εξομαλυμένη εκτίμηση της κατανομής πιθανότητας των καταλοίπων. Το διάγραμμα δείχνει μια καμπύλη σε σχήμα καμπάνας με κέντρο γύρω από το μηδέν, γεγονός που υποδηλώνει ότι τα κατάλοιπα είναι περίπου κανονικά κατανεμημένα. Η κορυφή γύρω από το μηδέν δείχνει ότι τα περισσότερα κατάλοιπα είναι κοντά στο μηδέν, πράγμα που σημαίνει ότι οι προβλέψεις του μοντέλου είναι γενικά ακριβείς. Η κατανομή φαίνεται να είναι αρκετά συμμετρική γύρω από το μηδέν, με μια μικρή λοξότητα προς τα δεξιά. Υπάρχουν ουρές και στις δύο πλευρές της κορυφής, που υποδηλώνουν την παρουσία μεγαλύτερων καταλοίπων, αλλά αυτά δεν είναι υπερβολικά βαριά, γεγονός που υποδηλώνει ότι οι ακραίες τιμές δεν αποτελούν σημαντικό ζήτημα. Η ελαφρά δεξιά κλίση υποδηλώνει ότι υπάρχουν ορισμένα μεγαλύτερα θετικά κατάλοιπα, αλλά αυτά δεν είναι αρκετά ακραία ώστε να προκαλούν ανησυχία για σημαντική μεροληψία του μοντέλου ή για μη υπολογίσιμα πρότυπα.

Συνεπώς, το μοντέλο ARIMA(1,1,1) αποδίδει αρκετά καλά με γενικά ακριβείς προβλέψεις και κανονικά κατανεμημένα κατάλοιπα. Ωστόσο, υπάρχουν περιθώρια βελτίωσης όσον αφορά την καταγραφή περιόδων υψηλότερης μεταβλητότητας και την αντιμετώπιση περιστασιακών μεγαλύτερων καταλοίπων.

Δεν αρκούν μόνο τα παραπάνω γραφήματα για να αξιολογήσουμε αν το ARIMA μοντέλο μας είναι καλό. Πρέπει επίσης να εξετάσουμε τα διαγράμματα Αυτοσυσχέτισης (ACF) και Μερικής Αυτοσυσχέτισης (PACF) των καταλοίπων, τα οποία μας βοηθούν να διαπιστώσουμε αν τα κατάλοιπα του μοντέλου ARIMA παρουσιάζουν χαρακτηριστικά «λευκού θορύβου», δηλαδή αν δεν εμφανίζουν καμία αυτοσυσχέτιση. Τα διαγράμματα αυτά χρησιμοποιούνται για να διαγνώσουν αν ένα μοντέλο χρονοσειράς έχει καταγράψει επαρκώς τα υποκείμενα πρότυπα δεδομένων.

Παρακάτω φαίνονται τα διαγράμματα ACF και PACF των καταλοίπων του μοντέλου ARIMA(1,1,1).



Εικόνα 19 - Διαγράμματα ACF και PACF του ARIMA(1,1,1)

Το διάγραμμα ACF απεικονίζει την αυτοσυσχέτιση των καταλοίπων σε διαφορετικές χρονικές υστερήσεις (lags). Ο άξονας x αντιπροσωπεύει την υστέρηση, δηλαδή των αριθμό των χρονικών βημάτων και ο άξονας y αντιπροσωπεύει τον συντελεστή συσχέτισης. Οι μπλε ράβδοι αντιπροσωπεύουν τους συντελεστές αυτοσυσχέτισης για κάθε υστέρηση και η σκιασμένη περιοχή αναπαριστά το διάστημα εμπιστοσύνης 95%. Η αυτοσυσχέτιση στην υστέρηση 0 είναι πάντα 1, καθώς αντιπροσωπεύει τη συσχέτιση της σειράς με τον εαυτό της. Οι υπόλοιποι συντελεστές αυτοσυσχέτισης βρίσκονται εντός του διαστήματος εμπιστοσύνης, υποδεικνύοντας ότι τα κατάλοιπα δεν παρουσιάζουν σημαντική αυτοσυσχέτιση στις περισσότερες υστερήσεις.

Το διάγραμμα PACF δείχνει τη μερική αυτοσυσχέτιση των καταλοίπων σε διάφορες χρονικές υστερήσεις. Ο άξονας x αντιπροσωπεύει την υστέρηση και ο άξονας y αντιπροσωπεύει τον συντελεστή μερικής συσχέτισης. Οι μπλε ράβδοι αντιπροσωπεύουν τους συντελεστές μερικής αυτοσυσχέτισης για κάθε υστέρηση και η σκιασμένη περιοχή αντιπροσωπεύει το διάστημα εμπιστοσύνης. Η μερική αυτοσυσχέτιση στην υστέρηση 0 είναι 1, παρόμοια με το διάγραμμα ACF. Για να είναι τα κατάλοιπα λευκός θόρυβος, οι συντελεστές μερικής αυτοσυσχέτισης για υστερήσεις μεγαλύτερες του μηδενός θα πρέπει επίσης να είναι κοντά στο μηδέν και εντός του διαστήματος εμπιστοσύνης. Σε αυτό το διάγραμμα, οι περισσότεροι συντελεστές μερικής αυτοσυσχέτισης βρίσκονται εντός του διαστήματος εμπιστοσύνης, υποδηλώνοντας ότι δεν υπάρχει σημαντική μερική αυτοσυσχέτιση στις περισσότερες υστερήσεις.

Τόσο τα διαγράμματα ACF όσο και τα διαγράμματα PACF υποδηλώνουν ότι τα κατάλοιπα είναι λευκός θόρυβος, καθώς οι περισσότεροι συντελεστές αυτοσυσχέτισης και μερικής αυτοσυσχέτισης εμπίπτουν στα διαστήματα εμπιστοσύνης. Αυτό δείχνει ότι το μοντέλο ARIMA(1,1,1) έχει καταγράψει επαρκώς τα υποκείμενα πρότυπα στα δεδομένα, αφήνοντας κατάλοιπα που δεν παρουσιάζουν σημαντική αυτοσυσχέτιση.

6.1.3 Εφαρμογή ARIMA(0,1,1)

Στην παράγραφο 6.1.2 είδαμε ότι το μοντέλο ARIMA(1,1,1) εφαρμόζει καλά στα δεδομένα αλλά ο όρος AR (ar.L1) είχε συντελεστή 0.0962 με υψηλό p-value το οποίο ήταν 0,608, γεγονός που υποδηλώνει ότι ο συντελεστής αυτός δεν είναι στατιστικά σημαντικός. Αντιθέτως, ο όρος MA (ma.L1) είχε συντελεστή -0.8234 με μηδενικό p-value, που υποδηλώνει υψηλή στατιστική σημαντικότητα.

Δεδομένου ότι ο όρος AR δεν είναι στατιστικά σημαντικός, ενδέχεται να συμβάλλει ελάχιστα στην προβλεπτική ικανότητα του μοντέλου. Ως εκ τούτου, ένα απλούστερο μοντέλο χωρίς τον όρο AR θα μπορούσε ενδεχομένως να αποδώσει εξίσου καλά ή και καλύτερα.

Για αυτό το λόγο σε αυτή την παράγραφο θα μελετήσουμε τα αποτελέσματα του μοντέλου ARIMA(0,1,1), το οποίο αφαιρεί τον ασήμαντο όρο AR, αφήνοντας μόνο τη διαφορά I και τον όρο MA.

Όπως και στα δύο προηγούμενα μοντέλα ARIMA, έτσι και σε αυτό ακολουθείται ακριβώς η ίδια διαδικασία. Χωρίζουμε το σύνολο δεδομένων σε σύνολο εκπαίδευσης (train set) και σε σύνολο ελέγχου (test set). Το σύνολο δεδομένων εκπαίδευσης περιέχει τον αριθμό των βλαβών της μηχανής CE-DE, 12V50DF από τον Ιανουάριο του 2017 έως και τον Οκτώβριο του 2022, και το σύνολο δεδομένων ελέγχου περιέχει μόνο μία τιμή, τον Νοέμβριο του 2022. Όπως έχει εξηγηθεί και στα δύο παραπάνω μοντέλα, για κάθε μήνα που προβλέπουμε, το σύνολο εκπαίδευσης παίρνει όλα τα δεδομένα μέχρι και τον προηγούμενο μήνα απο αυτόν που θέλουμε να προβλέψουμε, και το σύνολο ελέγχου περιέχει μόνο το μήνα που θέλουμε να προβλέψουμε.

Εφαρμόζουμε το μοντέλο ARIMA(0,1,1) στο σύνολο δεδομένων ελέγχου και τα αποτελέσματα φαίνονται στην παρακάτω εικόνα.

SARIMAX Results						
=====						
Dep. Variable:	NumOfDefects	No. Observations:	70			
Model:	ARIMA(0, 1, 1)	Log Likelihood	-222.291			
Date:	Sun, 21 Jul 2024	AIC	448.583			
Time:	19:50:20	BIC	453.051			
Sample:	01-01-2017	HQIC	450.355			
	- 10-01-2022					
Covariance Type:	opg					
=====						
	coef	std err	z	P> z	[0.025	0.975]

ma.L1	-0.7736	0.086	-9.047	0.000	-0.941	-0.606
sigma2	36.3116	7.033	5.163	0.000	22.527	50.096
=====						
Ljung-Box (L1) (Q):		0.11	Jarque-Bera (JB):	1.52		
Prob(Q):		0.74	Prob(JB):	0.47		
Heteroskedasticity (H):		1.57	Skew:	0.33		
Prob(H) (two-sided):		0.29	Kurtosis:	2.68		
=====						

Εικόνα 20 - Αποτελέσματα του μοντέλου ARIMA(0,1,1)

Παρακάτω αναλύονται τα αποτελέσματα του μοντέλου ARIMA(0,1,1):

Περίληψη μοντέλου:

- Dep. Variable: NumOfDefects
Αυτή είναι η εξαρτημένη μεταβλητή που προσπαθεί να προβλέψει το μοντέλο.
- No. Observations: 70
Ο αριθμός των σημείων δεδομένων που χρησιμοποιούνται για την προσαρμογή του μοντέλου.
- Μοντέλο: ARIMA(0, 1, 1)
Δείχνει ότι το μοντέλο είναι ένα μοντέλο ARIMA με 0 αυτοπαλίνδρομους όρους (AR), υπολογίζεται η πρώτη σειρά διαφορών (I) και υπάρχει 1 όρος κινητού μέσου όρου (MA).
- Log Likelihood (Λογική πιθανοφάνεια): -222.291
Αποτελεί μέτρο προσαρμογής του μοντέλου. Αυτή η τιμή χρησιμοποιείται στον υπολογισμό των AIC και BIC. Υψηλότερες τιμές υποδηλώνουν γενικά καλύτερη προσαρμογή.
- AIC (Akaike Information Criterion): 448.583
Ένα χαμηλότερο AIC υποδηλώνει καλύτερη προσαρμογή του μοντέλου στα δεδομένα, λαμβάνοντας υπόψη την πολυπλοκότητα του μοντέλου.
- BIC (Bayesian Information Criterion): 453.051
Παρόμοιο με το AIC, αλλά με μεγαλύτερη ποινή για μοντέλα με περισσότερες παραμέτρους. Χαμηλότερες τιμές BIC υποδηλώνουν ένα καλύτερο μοντέλο.
- HQIC (Κριτήριο πληροφοριών Hannan-Quinn): 450.355
Ένα άλλο κριτήριο για τη σύγκριση μοντέλων, που εξισορροπεί την προσαρμογή και την πολυπλοκότητα.

Συντελεστές:

- $ma.L1$: -0.7736
Ο συντελεστής για την πρώτη υστέρηση του τμήματος του κινητού μέσου όρου. Το τυπικό σφάλμα του συντελεστή (std err) είναι 0,086 και το p-value είναι 0, υποδεικνύοντας ότι ο συντελεστής αυτός είναι στατιστικά σημαντικός, διότι p-value < 0.05.
- σ^2 : 36.3116
Η διακύμανση των καταλοίπων (σφαλμάτων) του μοντέλου. Μια χαμηλότερη τιμή υποδηλώνει μικρότερη διακύμανση των σφαλμάτων.

Διαγνωστικά:

- Ljung-Box (L1) (Q): 0.11
Η στατιστική του Ljung-Box Q-test στην υστέρηση 1. Αυτό είναι ένας έλεγχος αυτοσυσχέτισης στα κατάλοιπα. Η τιμή p (Prob(Q) = 0.74) υποδηλώνει ότι δεν υπάρχει σημαντική αυτοσυσχέτιση στα κατάλοιπα στην υστέρηση 1.
- Jarque-Bera (JB): 1.52

Το στατιστικό ελέγχου Jarque-Bera για την κανονικότητα των καταλοίπων. Μια υψηλή τιμή p ($\text{Prob}(JB) = 0.47$) υποδηλώνει ότι τα κατάλοιπα κατανέμονται κανονικά.

- Heteroskedasticity (Ετεροσκεδαστικότητα (H)): 1.57

Το στατιστικό ελέγχου ετεροσκεδαστικότητας. Η τιμή p -value ($\text{Prob}(H) = 0,29$) υποδηλώνει ότι η διακύμανση των καταλοίπων είναι σταθερή με την πάροδο του χρόνου, δηλαδή δεν υπάρχει σημαντική ετεροσκεδαστικότητα.

- Skew (Στρέβλωση): 0.33

Ένα μέτρο της ασυμμετρίας της κατανομής των καταλοίπων. Μια στρέβλωση κοντά στο μηδέν υποδηλώνει συμμετρικά κατάλοιπα.

- Kyrptosis (Κύρτωση): 2.68

Ένα μέτρο της «ουράς» της κατανομής των καταλοίπων. Μια κύρτωση κοντά στο 3 υποδηλώνει κανονική κατανομή.

Από τα παραπάνω αποτελέσματα παρατηρούμε ότι οι τιμές AIC, BIC και HQIC του μοντέλου ARIMA(0,1,1) είναι χαμηλότερες από εκείνες του μοντέλου ARIMA(1,1,1), γεγονός που υποδηλώνει καλύτερη προσαρμογή, λαμβάνοντας υπόψη την πολυπλοκότητα του μοντέλου. Ο όρος MA είναι εξαιρετικά σημαντικός αφού το p -value=0 < 0.05, υποδεικνύοντας ότι συμβάλλει αρκετά στο μοντέλο. Επιπλέον, οι διαγνωστικοί έλεγχοι (Ljung-Box, Jarque-Bera, Heteroskedasticity) υποδεικνύουν ότι τα κατάλοιπα συμπεριφέρονται σαν λευκός θόρυβος, γεγονός που αποτελεί θετικό σημάδι για την επάρκεια του μοντέλου.

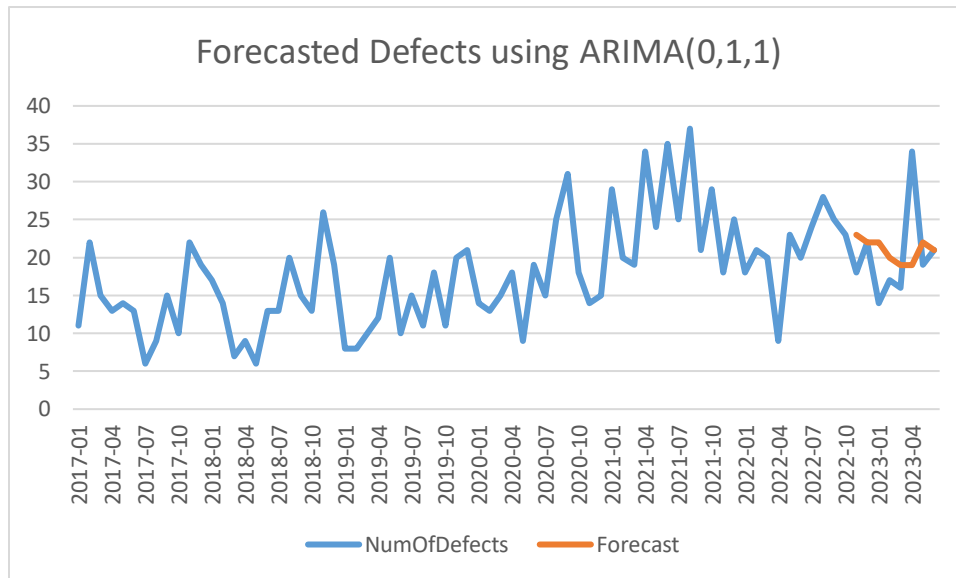
Συνεπώς, το μοντέλο ARIMA(0,1,1) φαίνεται να αποδίδει καλά, με σημαντικούς όρους MA και κατάλοιπα που μοιάζουν με λευκό θόρυβο. Αυτό υποδηλώνει ότι το μοντέλο αποτυπώνει επαρκώς τα υποκείμενα πρότυπα στα δεδομένα χωρίς να συμβαίνει υπερπροσαρμογή.

Το επόμενο βήμα είναι να δούμε τις προβλέψεις που κάνει το μοντέλο ARIMA(0,1,1). Ο αριθμός των βλαβών της μηχανής για το μήνα Νοέμβριο του 2022 ήταν 18 για όλα τα πλοία συνολικά. Το μοντέλο ARIMA(0,1,1) προέβλεψε για την ίδια περίοδο 23 βλάβες. Παρακάτω παρουσιάζεται ένας πίνακας με τις κανονικές τιμές και τις τιμές προβλεψής.

Πίνακας 8 - Τιμές Πρόβλεψης μοντέλου ARIMA(0,1,1)

Μήνας	Αρ. Βλαβών	Πρόβλεψη Αρ. Βλαβών
11/2022	18	23
12/2022	22	22
01/2023	14	22
02/2023	17	20
03/2023	16	19
04/2023	34	19
05/2023	19	22
06/2023	21	21

Παρακάτω παρουσιάζεται το γράφημα των πραγματικών τιμών και αυτών που προβλέφθηκαν απο το Νοέμβριο του 2022 μέχρι και τον Ιούνιο του 2023.



Εικόνα 21 - Διάγραμμα Προβλέψεων του ARIMA(0,1,1)

Όπως φαίνεται ο πίνακας πρόβλεψης και κατά συνέπεια το διάγραμμα των πραγματικών και των προβλεπόμενων τιμών είναι πολύ παρόμοια και για τα δύο μοντέλα ARIMA(1,1,1) και ARIMA(0,1,1). Αυτό συμβαίνει επειδή και τα δύο μοντέλα αποτυπώνουν τα ίδια θεμελιώδη μοτίβα στα δεδομένα, με μικρές μόνο διαφορές στην αυτοπαλίνδρομη (AR) συνιστώσα. Η ομοιότητα στις προβλεπόμενες τιμές υποδηλώνει ότι η συνιστώσα του κινητού μέσου όρου (MA) είναι ο κυρίαρχος παράγοντας στη δομή των δεδομένων.

Πιο αναλυτικά, το μοντέλο ARIMA(0,1,1) δεν περιλαμβάνει κανένα αυτοπαλίνδρομο όρο (AR = 0), περιλαμβάνει όμως ένα βήμα διαφοράς (I = 1) και έναν όρο κινητού μέσου όρου (MA = 1).

Η εξίσωση αυτού το μοντέλου είναι η εξής:

$$y_t - y_{t-1} = e_t + \theta_1 e_{t-1}$$

όπου y_t και y_{t-1} είναι η τρέχουσα και η προηγούμενη παρατήρηση αντίστοιχα,

θ_1 είναι ο συντελεστής του κινητού μέσου (MA),

e_t και e_{t-1} είναι ο τρέχων και ο προηγούμενος όρος σφάλματος αντίστοιχα.

Το μοντέλο ARIMA(1,1,1) περιλαμβάνει έναν αυτοπαλίνδρομο όρο (AR = 1), ένα βήμα διαφοράς (I = 1) και έναν όρο κινητού μέσου (MA = 1).

Η εξίσωση αυτού του μοντέλου είναι η εξής:

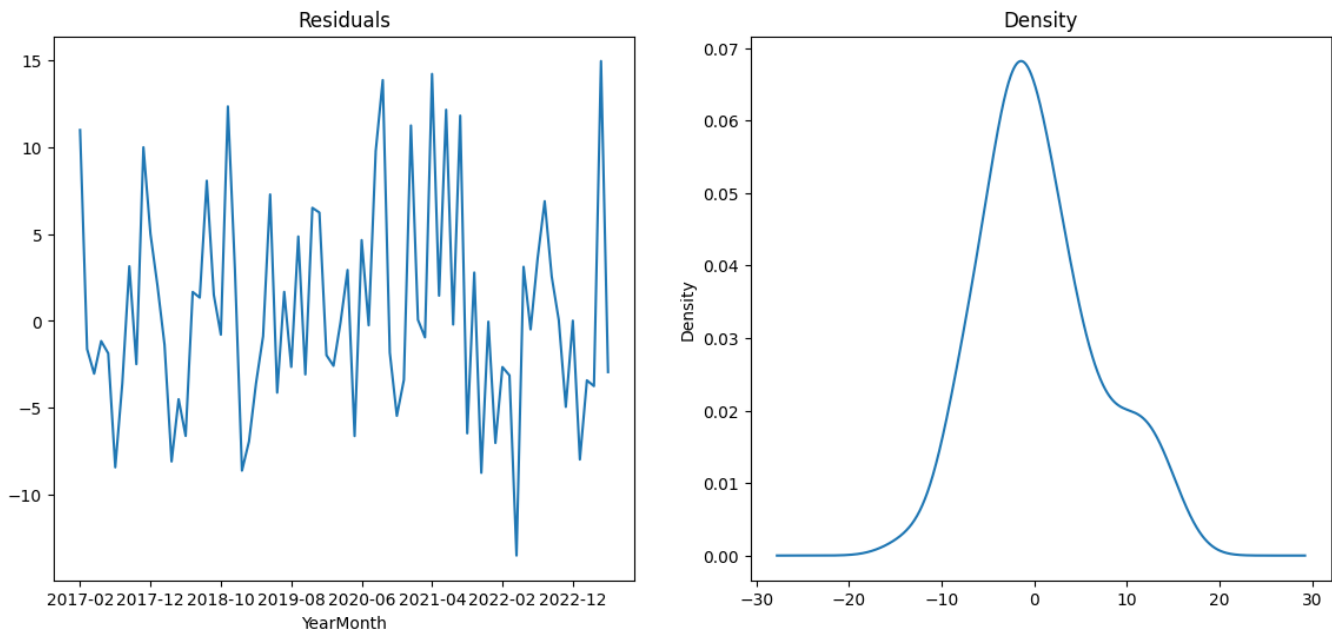
$$y_t - y_{t-1} = \varphi_1(y_{t-1} - y_{t-2}) + \theta_1 e_{t-1} + e_t$$

όπου φ_1 είναι ο συντελεστής αυτοπαλινδρόμησης AR [7].

Και τα δύο μοντέλα έχουν μια σημαντική συνιστώσα κινητού μέσου όρου. Ο όρος MA(1) καταγράφει την πρωταρχική δομή στα κατάλοιπα, οδηγώντας σε παρόμοιες τιμές πρόβλεψης. Ενώ, η συνιστώσα AR στο μοντέλο ARIMA(1,1,1) δεν είναι σημαντική (p-value = 0.608), γεγονός που υποδηλώνει ότι δεν προσθέτει μεγάλη προγνωστική δύναμη πέρα από αυτήν που ήδη καταγράφει η συνιστώσα MA. Εφόσον ο όρος AR δεν είναι

σημαντικός στο υπόδειγμα $ARIMA(1,1,1)$, η επιρροή του στην πρόβλεψη είναι ελάχιστη. Αυτό έχει ως αποτέλεσμα προβλέψεις που είναι πολύ κοντά σε αυτές που παράγονται από το μοντέλο $ARIMA(0,1,1)$, όπου ο όρος AR απουσιάζει. Επομένως, τα μοντέλα $ARIMA(1,1,1)$ και $ARIMA(0,1,1)$ παράγουν σχεδόν πανομοιότυπες προβλέψεις, επειδή ο πρόσθετος όρος AR στο μοντέλο $ARIMA(1,1,1)$ δεν μεταβάλλει σημαντικά τις τιμές πρόβλεψης. Αυτό επιβεβαιώνει ότι το απλούστερο μοντέλο είναι επαρκές για την αποτύπωση των βασικών προτύπων στα δεδομένα.

Στην εικόνα που ακολουθεί παρουσιάζονται τα διαγράμματα των καταλοίπων και πυκνότητας του μοντέλου $ARIMA(0,1,1)$.



Εικόνα 22 - Διαγράμματα Καταλοίπων και Πυκνότητας του $ARIMA(0,1,1)$

Η ομοιότητα μεταξύ των διαγραμμάτων καταλοίπων και πυκνότητας για τα μοντέλα $ARIMA(0,1,1)$ και $ARIMA(1,1,1)$ ενισχύει το προηγούμενο συμπέρασμα ότι ο όρος AR στο μοντέλο $ARIMA(1,1,1)$ δεν προσθέτει σημαντική προγνωστική αξία. Ας αναλύσουμε αυτά τα γραφήματα στο πλαίσιο του μοντέλου $ARIMA(0,1,1)$.

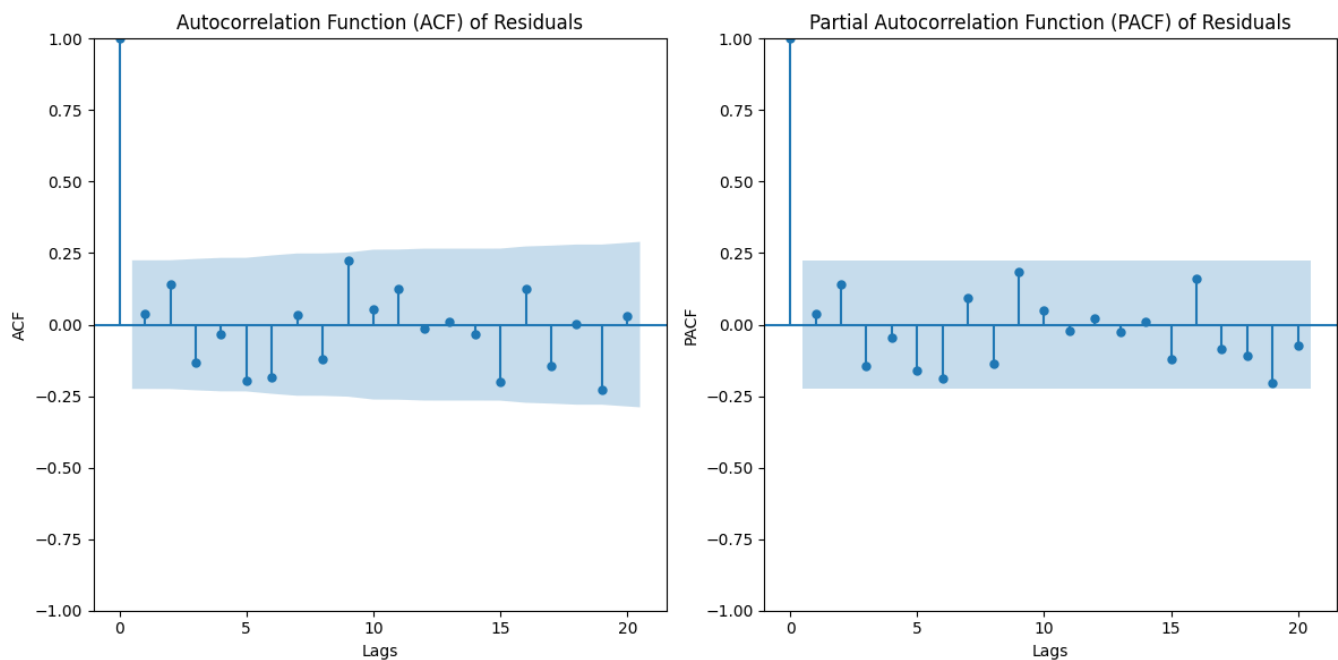
Το διάγραμμα καταλοίπων, στα αριστερά, δείχνει τις διαφορές μεταξύ παρατηρούμενων τιμών και των τιμών που προβλέπει το μοντέλο $ARIMA(0,1,1)$ με την πάροδο του χρόνου. Ο άξονας x αντιπροσωπεύει το χρόνο καλύπτοντας την περίοδο από τις αρχές του 2017 έως το τέλος του 2022. Ο άξονας y αντιπροσωπεύει το μέγεθος των καταλοίπων, δηλαδή τις αποκλίσεις των προβλέψεων του μοντέλου από τις πραγματικές παρατηρούμενες τιμές. Τα κατάλοιπα κυμαίνονται γύρω από τη γραμμή μηδέν, υποδεικνύοντας ότι οι προβλέψεις του μοντέλου είναι άλλοτε πάνω και άλλοτε κάτω από τις πραγματικές τιμές. Ενώ τα κατάλοιπα εμφανίζονται γενικά τυχαία, υπάρχουν περίοδοι όπου τα κατάλοιπα παρουσιάζουν επιμονή πάνω ή κάτω από το μηδέν, υποδηλώνοντας πιθανή αυτοσυσχέτιση. Η διακύμανση των υπολοίπων φαίνεται αρκετά σταθερή, αλλά υπάρχουν περίοδοι αυξημένης διακύμανσης, όπως γύρω στα μέσα του 2020 και στα τέλη του 2022. Η συνεπής διακύμανση αποτελεί θετικό δείκτη, αλλά οι αιχμές στα κατάλοιπα υποδηλώνουν ότι το μοντέλο μπορεί να μην καταγράφει όλες τις υποκείμενες δυναμικές.

Το διάγραμμα πυκνότητας, στα δεξιά, δείχνει την κατανομή των καταλοίπων, δίνοντας μια ιδέα της συνολικής διασποράς και της κεντρικής τάσης τους. Ο άξονας x αντιπροσωπεύει την τιμή των καταλοίπων. Ο

άξονας y αντιπροσωπεύει την πυκνότητα, η οποία είναι μια εξομαλυμένη εκτίμηση της κατανομής πιθανότητας των υπολοίπων. Το διάγραμμα δείχνει μια καμπύλη σε σχήμα καμπάνας με κέντρο γύρω από το μηδέν, γεγονός που υποδηλώνει ότι τα κατάλοιπα είναι περίπου κανονικά κατανομημένα. Η κορυφή γύρω από το μηδέν δείχνει ότι τα περισσότερα κατάλοιπα είναι κοντά στο μηδέν, πράγμα που σημαίνει ότι οι προβλέψεις του μοντέλου είναι γενικά ακριβείς. Επιλέον, η κατανομή φαίνεται να είναι αρκετά συμμετρική γύρω από το μηδέν, με μια μικρή λοξότητα προς τα δεξιά. Υπάρχουν ουρές και στις δύο πλευρές της κορυφής, που υποδηλώνουν την παρουσία μεγαλύτερων καταλοίπων, αλλά αυτά δεν είναι υπερβολικά βαριά, γεγονός που υποδηλώνει ότι οι ακραίες τιμές δεν αποτελούν σημαντικό ζήτημα. Η ελαφρά δεξιά κλίση υποδηλώνει ότι υπάρχουν ορισμένα μεγαλύτερα θετικά κατάλοιπα, δηλαδή υποεκτιμήσεις από το μοντέλο.

Από τα παραπάνω, συμπεραίνουμε ότι το μοντέλο $ARIMA(0,1,1)$ αποτυπώνει αρκετά καλά τη γενική τάση και τα εποχικά πρότυπα των δεδομένων. Το διάγραμμα υπολειμμάτων δείχνει ότι ενώ οι προβλέψεις του μοντέλου είναι γενικά κοντά στις πραγματικές τιμές, υπάρχουν περίοδοι όπου το μοντέλο δυσκολεύεται να συλλάβει τη μεταβλητότητα.

Εκτός από τα διαγράμματα των καταλοίπων και της πυκνότητας, εξετάστηκαν και τα διαγράμματα Αυτοσυσχέτισης (ACF) και Μερικής Αυτοσυσχέτισης (PACF) των καταλοίπων. Παρακάτω φαίνονται τα διαγράμματα ACF και PACF των καταλοίπων του μοντέλου $ARIMA(0,1,1)$.



Εικόνα 23 - Διαγράμματα ACF και PACF των Καταλοίπων του $ARIMA(0,1,1)$

Το διάγραμμα ACF δείχνει την αυτοσυσχέτιση των καταλοίπων σε διαφορετικές χρονικές υστερήσεις. Ο άξονας x αντιπροσωπεύει την υστέρηση, δηλαδή τον αριθμό των χρονικών βημάτων, και ο άξονας y αντιπροσωπεύει τον συντελεστή συσχέτισης. Οι μπλε ράβδοι αντιπροσωπεύουν τους συντελεστές αυτοσυσχέτισης για κάθε υστέρηση και η σκιασμένη περιοχή αντιπροσωπεύει το διάστημα εμπιστοσύνης 95%. Η αυτοσυσχέτιση στην υστέρηση 0 είναι πάντα 1, καθώς αντιπροσωπεύει τη συσχέτιση της σειράς με τον εαυτό της. Για να μοιάζουν τα κατάλοιπα με λευκό θόρυβο, οι συντελεστές αυτοσυσχέτισης για υστερήσεις μεγαλύτερες του μηδενός θα πρέπει να είναι κοντά στο μηδέν και εντός του διαστήματος εμπιστοσύνης. Σε αυτό

το διάγραμμα, οι περισσότεροι συντελεστές αυτοσυσχέτισης βρίσκονται εντός του διαστήματος εμπιστοσύνης, υποδεικνύοντας ότι τα κατάλοιπα δεν παρουσιάζουν σημαντική αυτοσυσχέτιση στις περισσότερες υστερήσεις.

Το διάγραμμα PACF δείχνει τη μερική αυτοσυσχέτιση των καταλοίπων σε διαφορετικές χρονικές υστερήσεις. Ο άξονας x αντιπροσωπεύει την υστέρηση και ο άξονας y αντιπροσωπεύει τον συντελεστή μερικής συσχέτισης. Οι μπλε ράβδοι αντιπροσωπεύουν τους συντελεστές μερικής αυτοσυσχέτισης για κάθε υστέρηση και η σκιασμένη περιοχή αντιπροσωπεύει το διάστημα εμπιστοσύνης. Αντίστοιχα και εδώ η μερική αυτοσυσχέτιση στην υστέρηση 0 είναι 1. Για να μοιάζουν τα κατάλοιπα με λευκό θόρυβο, οι συντελεστές μερικής αυτοσυσχέτισης για υστερήσεις μεγαλύτερες του μηδενός θα πρέπει επίσης να είναι κοντά στο μηδέν και εντός του διαστήματος εμπιστοσύνης. Σε αυτό το διάγραμμα, οι περισσότεροι συντελεστές μερικής αυτοσυσχέτισης βρίσκονται εντός του διαστήματος εμπιστοσύνης, υποδηλώνοντας ότι δεν υπάρχει σημαντική μερική αυτοσυσχέτιση στις περισσότερες υστερήσεις.

Συνεπώς τα διαγράμματα ACF και PACF υποδηλώνουν ότι τα κατάλοιπα είναι λευκός θόρυβος, καθώς οι συντελεστές αυτοσυσχέτισης και μερικής αυτοσυσχέτισης εμπίπτουν εντός των διαστημάτων εμπιστοσύνης. Αυτό δείχνει ότι το μοντέλο ARIMA(0,1,1) έχει καταγράψει επαρκώς τα υποκείμενα πρότυπα στα δεδομένα, αφήνοντας κατάλοιπα που δεν παρουσιάζουν σημαντική αυτοσυσχέτιση. Αυτό ευθυγραμμίζεται με την προηγούμενη ανάλυση ότι το μοντέλο ARIMA(0,1,1) αποτυπώνει αποτελεσματικά την ουσιαστική δυναμική των δεδομένων.

6.1.4 Αποτελέσματα Συνάρτησης `auto_arima`

Όπως παρατηρήθηκε στις προηγούμενες παραγράφους, η εύρεση των καλύτερων παραμέτρων p, d και q για το μοντέλο ARIMA(p, d, q) μπορεί να είναι δύσκολη και χρονοβόρα διαδικασία. Για αυτό το λόγο, υπάρχει η συνάρτηση `auto_arima` η οποία μπορεί να βοηθήσει στην αυτοματοποίηση αυτής της διαδικασίας. Η `auto_arima` εξετάζει διάφορους συνδυασμούς των τριών παραμέτρων και επιλέγει το βέλτιστο συνδυασμό για το μοντέλο. Η επιλογή αυτή γίνεται συγκρίνοντας τα κριτήρια πληροφορίας AIC και BIC όλων των συνδυασμών, και επιλέγει εκείνο με τις χαμηλότερες τιμές σε AIC και BIC.

Εφόσον δοκιμάστηκαν τρία διαφορετικά μοντέλα ARIMA, τα ARIMA(2,1,2), ARIMA(1,1,1) και ARIMA(0,1,1), εφαρμόστηκε και η συνάρτηση `auto_arima` πάνω στο σύνολο εκπαίδευσης του συνόλου δεδομένων, με σκοπό μήπως βρεθεί κάποιος καλύτερος συνδυασμός των p, d, q συγκριτικά με τους παραπάνω.

Η συνάρτηση `auto_arima` που χρησιμοποιήθηκε στο δικό μας σύνολο δεδομένων περιέχει τις παρακάτω παραμέτρους:

- `df_train['NumOfDefects']`: Τα δεδομένα της χρονοσειράς στα οποία θα προσαρμοστεί το μοντέλο ARIMA. Αυτό είναι το σύνολο δεδομένων εκπαίδευσης που περιέχει τον αριθμό των ελαττωμάτων.
- `start_p = 0` : Η αρχική τιμή για την τάξη AR (p) ορίστηκε να είναι το 0.
- `start_q = 0` : Η αρχική τιμή για την τάξη MA (q) ορίστηκε να είναι το 0.
- `max_p = 2` : Η μέγιστη τιμή για την τάξη AR (p) ορίστηκε να είναι το 2.
- `max_q = 2` : Η μέγιστη τιμή για την εντολή MA (q) ορίστηκε να είναι το 2.

Επομένως η `auto_arima` θα δοκιμάσει όλους τους συνδυασμούς παραμέτρων από το 0 έως και το 2. Αυτό το εύρος επιλέχθηκε λόγω των διαγραμμάτων ACF και PACF του συνόλου δεδομένων που είχε το πολύ δύο χρονικές υστερήσεις (lags) εκτός του διαστήματος εμπιστοσύνης.

- $m = 12$: Ο αριθμός των περιόδων σε κάθε εποχιακό κύκλο. Ακόμη και αν η χρονοσειρά δεν έχει εποχικότητα, δηλαδή αν `seasonal = False`, η παράμετρος αυτή διατηρείται αλλά δεν έχει καμία επίδραση.
- `Seasonal = False`: Υποδεικνύει ότι τα δεδομένα της χρονοσειράς δεν έχουν εποχικότητα.
- `d = None`: Η σειρά της διαφοράς συμβολίζεται με d . Αν δεν υπάρχει, η συνάρτηση `auto_arima` καθορίζει την κατάλληλη τάξη διαφοράς για να καταστήσει τη σειρά στάσιμη.
- `trace = True`: Ενεργοποιεί τη λεπτομερή έξοδο της διαδικασίας προσαρμογής. Δηλαδή, θα εμφανιστεί ο κατάλογος των εξεταζόμενων μοντέλων ARIMA.
- `error_action = 'ignore'`: Αγνοεί τα σφάλματα από μοντέλα που δε συγκλίνουν.
- `suppress_warnings = True`: Καταστέλλει τις προειδοποιήσεις κατά τη διάρκεια της διαδικασίας προσαρμογής.
- `Stepwise = True`: Χρησιμοποιεί έναν αλγόριθμο βηματικής διαδικασίας για την αναζήτηση των καλύτερων παραμέτρων του μοντέλου.

Η συνάρτηση ξεκινά με τον προσδιορισμό της τάξης της διαφοράς (d) που απαιτείται για να καταστεί η σειρά στάσιμη. Αυτό γίνεται με την εκτέλεση στατιστικών ελέγχων όπως ο έλεγχος Augmented Dickey-Fuller (ADF). Εάν η σειρά διαπιστωθεί ότι δεν είναι στάσιμη, εφαρμόζεται η διαφορά. Στη συνέχεια, η συνάρτηση αναζητά στα καθορισμένα εύρη των p και q για να βρει το καλύτερο μοντέλο ARIMA. Η αναζήτηση αυτή μπορεί να είναι εξαντλητική, λόγω της αξιολόγησης όλων των συνδυασμών, ή σταδιακή, δηλαδή να γίνεται η προσθήκη/αφαίρεση παραμέτρων επαναλαμβανόμενα με βάση τη στατιστική σημαντικότητα. Σε αυτή την περίπτωση, ενεργοποιείται η βηματική αναζήτηση (`stepwise=True`), η οποία καθιστά τη διαδικασία πιο αποτελεσματική μειώνοντας τον υπολογιστικό φόρτο. Η βηματική προσέγγιση συνήθως περιλαμβάνει:

1. Έναρξη με ένα απλό μοντέλο.
2. Προσθήκη ή αφαίρεση παραμέτρων μία προς μία.
3. Αξιολόγηση της προσαρμογής του μοντέλου με τη χρήση κριτηρίων πληροφορίας όπως το AIC και BIC.
4. Επανάληψη της διαδικασίας έως ότου βρεθεί το καλύτερο μοντέλο.

Για κάθε υποψήφιο μοντέλο, η συνάρτηση προσαρμόζει το μοντέλο ARIMA στο σύνολο δεδομένων εκπαίδευσης και υπολογίζει τις τιμές AIC και BIC. Αυτές οι τιμές εξισορροπούν την προσαρμογή του μοντέλου και την πολυπλοκότητα, τιμωρώντας τα μοντέλα με περισσότερες παραμέτρους. Το μοντέλο με τη χαμηλότερη τιμή του κριτηρίου πληροφορίας (είτε AIC είτε BIC) θεωρείται το καλύτερο μοντέλο. Η συνάρτηση επιλέγει το μοντέλο με τη χαμηλότερη τιμή AIC/BIC ως βέλτιστο μοντέλο, εξασφαλίζοντας ισορροπία μεταξύ ακρίβειας και απλότητας. Τέλος, επιστρέφει το μοντέλο που ταιριάζει καλύτερα, συμπεριλαμβανομένης μιας σύνοψης των παραμέτρων του, της στατιστικής τους σημαντικότητας και διαφόρων διαγνωστικών μετρήσεων. Αυτό επιτρέπει την περαιτέρω αξιολόγηση και τελειοποίηση.

Παρακάτω παρατίθενται τα λεπτομερή αποτελέσματα της συνάρτησης `auto_arima`.

```

Performing stepwise search to minimize aic
ARIMA(0,1,0)(0,0,0)[0] intercept : AIC=477.311, Time=0.04 sec
ARIMA(1,1,0)(0,0,0)[0] intercept : AIC=456.064, Time=0.03 sec
ARIMA(0,1,1)(0,0,0)[0] intercept : AIC=449.717, Time=0.03 sec
ARIMA(0,1,0)(0,0,0)[0] : AIC=475.348, Time=0.01 sec
ARIMA(1,1,1)(0,0,0)[0] intercept : AIC=inf, Time=0.10 sec
ARIMA(0,1,2)(0,0,0)[0] intercept : AIC=inf, Time=0.10 sec
ARIMA(1,1,2)(0,0,0)[0] intercept : AIC=inf, Time=0.09 sec
ARIMA(0,1,1)(0,0,0)[0] : AIC=448.583, Time=0.02 sec
ARIMA(1,1,1)(0,0,0)[0] : AIC=450.245, Time=0.09 sec
ARIMA(0,1,2)(0,0,0)[0] : AIC=450.383, Time=0.04 sec
ARIMA(1,1,0)(0,0,0)[0] : AIC=454.133, Time=0.02 sec
ARIMA(1,1,2)(0,0,0)[0] : AIC=inf, Time=0.09 sec

Best model: ARIMA(0,1,1)(0,0,0)[0]
Total fit time: 0.682 seconds

```

Εικόνα 24 - Αποτελέσματα Συνάρτησης *auto_arima*

Η συνάρτηση *auto_arima* πραγματοποίησε μια σταδιακή αναζήτηση για την ελαχιστοποίηση του AIC, επαναλαμβάνοντας διάφορους συνδυασμούς των παραμέτρων p , d και q . Η προσέγγιση της βηματικής αναζήτησης περιλαμβάνει την επαναληπτική προσθήκη και αφαίρεση παραμέτρων με βάση τη στατιστική τους σημαντικότητα, περιορίζοντας έτσι αποτελεσματικά το βέλτιστο μοντέλο.

Η διαδικασία αναζήτησης αξιολόγησε τα ακόλουθα μοντέλα:

1. ARIMA(0,1,0)(0,0,0)[0] intercept:
 - AIC = 477.311

Αυτό το μοντέλο είναι ένα απλό μοντέλο στο οποίο έχουν υπολογιστεί μία φορά οι διαφορές μεταξύ των δεδομένων, χωρίς στοιχεία AR ή MA και περιλαμβάνει μία διατομή.
2. ARIMA(1,1,0)(0,0,0)[0] intercept:
 - AIC = 456.064

Αυτό το υπόδειγμα περιλαμβάνει έναν αυτοπαλίνδρομο όρο μαζί με διαφορά και μια διατομή.
3. ARIMA(0,1,1)(0,0,0)[0] intercept:
 - AIC = 449.717

Αυτό το μοντέλο περιλαμβάνει έναν όρο κινητού μέσου όρου μαζί με διαφορά και μια διατομή.
4. ARIMA(0,1,0)(0,0,0)[0]:
 - AIC = 475.348

Αυτό το υπόδειγμα είναι ένα απλό μοντέλο το οποίο περιλαμβάνει μόνο τη σειρά των πρώτων διαφορών χωρίς καμία διατομή.
5. ARIMA(1,1,1)(0,0,0)[0] intercept:
 - AIC = inf

Αυτό το υπόδειγμα, το οποίο περιλαμβάνει όρους AR και MA μαζί με διαφορά και μια διατομή, δεν συγκλίνει σωστά, οδηγώντας σε άπειρη τιμή AIC.

6. ARIMA(0,1,2)(0,0,0)[0] intercept:

- AIC = inf

Όμοια με το προηγούμενο μοντέλο, αυτό το μοντέλο δεν συγκλίνει σωστά.

7. ARIMA(1,1,2)(0,0,0)[0] intercept:

- AIC = inf

Αυτό το μοντέλο επίσης δεν κατάφερε να συγκλίνει, με αποτέλεσμα την άπειρη τιμή AIC.

8. ARIMA(0,1,1)(0,0,0)[0]:

- AIC = 448.583

Αυτό το μοντέλο περιλαμβάνει έναν όρο κινητού μέσου όρου μαζί με τη διαφορά, χωρίς παρεμβολή. Πέτυχε τη χαμηλότερη τιμή AIC μεταξύ των μοντέλων που αξιολογήθηκαν.

9. ARIMA(1,1,1)(0,0,0)[0]:

- AIC = 450.245

Αυτό το υπόδειγμα περιλαμβάνει όρους AR και MA μαζί με διαφορά, χωρίς καμία διατομή.

10. ARIMA(0,1,2)(0,0,0)[0]:

- AIC = 450.383

Αυτό το μοντέλο περιλαμβάνει δύο όρους κινητού μέσου όρου μαζί με διαφορά, χωρίς διατομή.

11. ARIMA(1,1,0)(0,0,0)[0]:

- AIC = 454.133

Αυτό το υπόδειγμα περιλαμβάνει έναν αυτοπαλίνδρομο όρο μαζί με διαφορά, χωρίς διατομή.

12. ARIMA(1,1,2)(0,0,0)[0]:

- AIC = inf

Αυτό το μοντέλο απέτυχε να συγκλίνει, με αποτέλεσμα την άπειρη τιμή AIC.

Η συνάρτηση `auto_arima` προσδιόρισε το μοντέλο ARIMA(0,1,1)(0,0,0)[0] ως το καλύτερο μοντέλο, με τη χαμηλότερη τιμή AIC 448.583. Αυτό το μοντέλο περιλαμβάνει έναν όρο κινητού μέσου όρου και διαφοράς, χωρίς όρο διατομής. Η απουσία σύγκλισης σε ορισμένα μοντέλα υποδεικνύει ότι οι συγκεκριμένοι συνδυασμοί παραμέτρων δεν παρήγαγαν σταθερές λύσεις, με αποτέλεσμα να προκύπτουν άπειρες τιμές AIC.

Το επιλεγμένο μοντέλο, ARIMA(0,1,1), υποδηλώνει ότι τα δεδομένα της χρονοσειράς επωφελούνται από ένα επίπεδο διαφοράς για την επίτευξη στασιμότητας και περιλαμβάνει ένα μόνο στοιχείο κινητού μέσου όρου για να ληφθεί υπόψη η δομή αυτοσυσχέτισης στα κατάλοιπα.

Η λεπτομερής διαδικασία βηματικής αναζήτησης διασφαλίζει ότι το επιλεγμένο μοντέλο είναι τόσο λιτό όσο και αποτελεσματικό στην αποτύπωση της υποκείμενης δομής των δεδομένων, όπως αποδεικνύεται από την ελαχιστοποιημένη τιμή AIC. Ο συνολικός χρόνος που απαιτήθηκε για τη βηματική αναζήτηση ήταν 0.682 δευτερόλεπτα, γεγονός που αντικατοπτρίζει την αποτελεσματικότητα της συνάρτησης `auto_arima` στην επιλογή μοντέλου.

Συμπερασματικά, η συνάρτηση `auto_arima` προσδιόρισε αποτελεσματικά το μοντέλο ARIMA(0,1,1) ως τη βέλτιστη προσαρμογή για το σύνολο δεδομένων εκπαίδευσης, εξασφαλίζοντας ισορροπία μεταξύ της απλότητας του μοντέλου και της καλής προσαρμογής.

Στην ακόλουθη εικόνα φαίνονται αναλυτικά και τα αποτελέσματα του μοντέλου ARIMA(0,1,1):

SARIMAX Results						
Dep. Variable:	y	No. Observations:	70			
Model:	SARIMAX(0, 1, 1)	Log Likelihood	-222.291			
Date:	Tue, 23 Jul 2024	AIC	448.583			
Time:	23:10:31	BIC	453.051			
Sample:	01-01-2017	HQIC	450.355			
	- 10-01-2022					
Covariance Type:	opg					
	coef	std err	z	P> z	[0.025	0.975]
ma.L1	-0.7736	0.086	-9.047	0.000	-0.941	-0.606
sigma2	36.3116	7.033	5.163	0.000	22.527	50.096
Ljung-Box (L1) (Q):	0.11	Jarque-Bera (JB):	1.52			
Prob(Q):	0.74	Prob(JB):	0.47			
Heteroskedasticity (H):	1.57	Skew:	0.33			
Prob(H) (two-sided):	0.29	Kurtosis:	2.68			

Εικόνα 25 - Αποτελέσματα του μοντέλου ARIMA(0,1,1)

Προφανώς είναι ακριβώς τα ίδια αποτελέσματα που είχαμε ήδη αναλύσει στην παράγραφο 6.1.3, όπου είχαμε αναλύσει το μοντέλο ARIMA(0,1,1) και τις προβλέψεις του.

7. Σύγκριση Μοντέλων – Συμπεράσματα

Σε αυτό το κεφάλαιο γίνεται η σύγκριση των τριών μοντέλων ARIMA που εφαρμόστηκαν στο σύνολο δεδομένων, τα ARIMA(2,1,2), ARIMA(1,1,1) και ARIMA(0,1,1), ώστε να προσδιοριστεί ποιο μοντέλο ταιριάζει καλύτερα στα δεδομένα της χρονοσειράς που αντιπροσωπεύουν τον αριθμό των βλαβών ανά μήνα στη μηχανή CE-DE, 12V50DF. Η σύγκριση βασίζεται στις μετρικές απόδοσης, συμπεριλαμβανομένου του κριτηρίου πληροφορίας Akaike (AIC), και του κριτηρίου πληροφορίας Bayes (BIC) και των διαγνωστικών ελέγχων για την επάρκεια του μοντέλου.

Το μοντέλο ARIMA(2,1,2) περιλαμβάνει δύο αυτοπαλίνδρομους (AR) όρους, μία διαφορά (d) και δύο όρους κινητού μέσου όρου (MA). Αυτό το μοντέλο μπορεί να συλλάβει πιο σύνθετα μοτίβα στα δεδομένα λόγω του μεγαλύτερου αριθμού παραμέτρων. Αυτό το μοντέλο παρήγαγε τιμή AIC = 451.051 και τιμή BIC = 462.221. Τα τυπικά σφάλματα, οι τιμές z και οι τιμές p (p-value) για τους συντελεστές δείχνουν ότι ορισμένες παράμετροι, όπως για παράδειγμα η α_1 και α_2 , δεν είναι στατιστικά σημαντικές, καθώς οι τιμές p είναι πολύ υψηλότερες από το συμβατικό όρο του 0.05. Οι διαγνωστικοί έλεγχοι των κατάλοιπων, συμπεριλαμβανομένου του τεστ Ljung-Box, δεν υποδεικνύουν σημαντική αυτοσυσχέτιση στα κατάλοιπα ($\text{Prob}(Q) = 0.94$). Ο έλεγχος Jarque-Bera δείχνει ότι τα κατάλοιπα είναι περίπου κανονικά κατανομημένα ($\text{Prob}(JB) = 0.40$) και ο έλεγχος ετεροσκεδαστικότητας (H) δεν δείχνει σημαντική ετεροσκεδαστικότητα ($\text{Prob}(H) = 0.32$). Ωστόσο, ο σχετικά μεγάλος αριθμός παραμέτρων στο μοντέλο ARIMA(2,1,2) θα μπορούσε να οδηγήσει σε υπερπροσαρμογή.

Το μοντέλο ARIMA(1,1,1) περιλαμβάνει έναν αυτοπαλίνδρομο όρο (AR), έναν όρο διαφοράς και έναν όρο κινητού μέσου (MA). Η διαδικασία προσαρμογής του μοντέλου δεν αντιμετώπισε προβλήματα σύγκλισης και παρήγαγε πεπερασμένες τιμές AIC και BIC. Συγκεκριμένα, η τιμή AIC για αυτό το μοντέλο ήταν 450,245. Οι διαγνωστικοί έλεγχοι έδειξαν ότι τα κατάλοιπα ήταν περίπου ασυσχέτιστα και κανονικά κατανομημένα, γεγονός που υποδηλώνει ότι το μοντέλο αποτύπωσε τα βασικά πρότυπα των δεδομένων αρκετά καλά. Παρά την απλότητά του σε σύγκριση με το μοντέλο ARIMA(2,1,2), το μοντέλο ARIMA(1,1,1) είχε ελαφρώς υψηλότερη τιμή AIC, υποδεικνύοντας ότι το μοντέλο ARIMA(1,1,1) είχε ελαφρώς χειρότερη προσαρμογή στα δεδομένα.

Το μοντέλο ARIMA(0,1,1) που αποτελείται από έναν όρο διαφοράς και έναν όρο κινητού μέσου (MA) χωρίς στοιχεία αυτοπαλινδρόμησης (AR), αναδείχθηκε ως το μοντέλο με την καλύτερη προσαρμογή σύμφωνα με τη συνάρτηση `auto_arima`. Αυτό το μοντέλο είχε τη χαμηλότερη τιμή AIC = 448.583, υποδεικνύοντας καλύτερη ισορροπία μεταξύ καλής προσαρμογής και πολυπλοκότητας του μοντέλου σε σύγκριση με τα άλλα μοντέλα. Οι διαγνωστικοί έλεγχοι για αυτό το μοντέλο έδειξαν ότι τα κατάλοιπα ήταν ασυσχέτιστα και κανονικά κατανομημένα, παρόμοια με το μοντέλο ARIMA(1,1,1), γεγονός που υποδηλώνει καλή προσαρμογή στα δεδομένα.

Συνοψίζοντας όλα τα παραπάνω, το μοντέλο ARIMA(2,1,2), ενώ είναι ικανό να αποτυπώσει πολύπλοκα πρότυπα, παρουσίασε ορισμένες μη σημαντικές παραμέτρους και ελαφρώς υψηλότερη τιμή AIC από τα μοντέλα ARIMA(1,1,1) και ARIMA(0,1,1). Το μοντέλο ARIMA(1,1,1) είχε καλή απόδοση αλλά και υψηλότερη τιμή AIC σε σύγκριση με το μοντέλο ARIMA(0,1,1).

Επομένως, το μοντέλο ARIMA(0,1,1) είναι το μοντέλο με τις καλύτερες επιδόσεις μεταξύ των τριών που αξιολογήθηκαν. Επιτυγχάνει τη χαμηλότερη τιμή AIC, υποδεικνύοντας καλύτερη προσαρμογή, διατηρώντας παράλληλα την απλότητα του μοντέλου. Το μοντέλο ARIMA(1,1,1) έχει επίσης καλές επιδόσεις, αλλά είναι οριακά λιγότερο βέλτιστο με βάση τη σύγκριση του κριτηρίου πληροφορίας Akaike (AIC). Το μοντέλο ARIMA(2,1,2), παρά τις δυνατότητές του για την αποτύπωση πιο σύνθετων προτύπων, περιλαμβάνει μη σημαντικές παραμέτρους και έχει υψηλότερη τιμή AIC, γεγονός που το καθιστά λιγότερο επιθυμητό. Ως εκ τούτου, το μοντέλο ARIMA(0,1,1) συνιστάται για την πρόβλεψη του αριθμού των βλαβών σε αυτό το σύνολο δεδομένων, παρέχοντας μια αξιόπιστη και λιτή προσέγγιση στη μοντελοποίηση της χρονοσειράς.

Βιβλιογραφία

1. Papoutsidakis, M., Sfyroera, E., & Alafodimos, N. (n.d.). Information systems (ERP) for shipping companies. Jmest.org. Retrieved September 15, 2024, from <https://www.jmest.org/wp-content/uploads/JMESTN42353221.pdf>
2. Jimenez, V. J., Bouhmala, N., & Gausdal, A. H. (2020). Developing a predictive maintenance model for vessel machinery. *Journal of Ocean Engineering and Science*, 5(4), 358–386. <https://doi.org/10.1016/j.joes.2020.03.003>
3. How to choose the parameters for the model — Arauto 0.1.0 documentation. (n.d.). Readthedocs.io. Retrieved September 15, 2024, from https://arauto.readthedocs.io/en/latest/how_to_choose_terms.html
4. 9.1 Στασιμότητα και διαφορά. (n.d.). Otexts.com. Retrieved September 15, 2024, from <https://otexts.com/fppgr/stationarity.html>
5. (N.d.). Machinelearningmastery.com. Retrieved September 15, 2024, from <https://machinelearningmastery.com/arima-for-time-series-forecasting-with-python/>
6. Bora, N., Principal Associate, & Data Science. (n.d.). Understanding ARIMA models for machine learning. Capital One. Retrieved September 15, 2024, from <https://www.capitalone.com/tech/machine-learning/understanding-arima-models/>
7. 3.3 Residual diagnostics. (n.d.). Otexts.com. Retrieved September 15, 2024, from <https://otexts.com/fpp2/residuals.html>
8. (N.d.-b). Unipi.Gr. Retrieved September 15, 2024, from [https://oldsite.unipi.gr/faculty/mbouts/Boutsikas\(2020\)Time_Series.pdf](https://oldsite.unipi.gr/faculty/mbouts/Boutsikas(2020)Time_Series.pdf)
9. Pmdarima.arima.auto_arima — pmdarima 2.0.4 documentation. (n.d.). Alkaline-ml.com. Retrieved September 15, 2024, from https://alkaline-ml.com/pmdarima/modules/generated/pmdarima.arima.auto_arima.html
10. auto.arima function - RDocumentation. (n.d.). Rdocumentation.org. Retrieved September 15, 2024, from <https://www.rdocumentation.org/packages/forecast/versions/8.23.0/topics/auto.arima>
11. (N.d.-c). Hcg.Gr. Retrieved September 15, 2024, from <https://maredu.hcg.gr/modules/document/file.php/MAK264/%CE%9D%CE%B1%CF%85%CF%80%CE%B7%CE%B3%CE%AF%CE%B1/%CE%9C%CE%AD%CF%83%CE%B1%20%CF%80%CF%81%CF%8C%CF%89%CF%83%CE%B7%CF%82%20%CF%83%CF%84%CE%B1%20%CF%80%CE%BB%CE%BF%CE%AF%CE%B1%20%CE%BA%CE%B1%CE%B9%20%CE%B7%20%CE%B5%CE%BE%CE%AD%CE%BB%CE%B9%CE%BE%CE%AE%20%CF%84%CE%BF%CF%85%CF%82.pdf>
12. (N.d.-d). Ntua.Gr. Retrieved September 15, 2024, from https://www.lme.ntua.gr/academic-info-1/prospheromena-mathemata/egkatastaseis-prooses/files/basikes_arxes2.pdf
13. (N.d.-e). Uniwa.Gr. Retrieved September 15, 2024, from <https://polynoe.lib.uniwa.gr/xmlui/bitstream/handle/11400/3579/%ce%a0%ce%9c%ce%a3%20%ce%94%ce%99%ce%a0%ce%9b%ce%a9%ce%9c%ce%91%ce%a4%ce%99%ce%9a%ce%97%20%ce%95%ce%a1%ce%93%ce%91%ce%a3%ce%99%ce%91%20%ce%9c%ce%91%ce%a1%ce%99%ce%9f%ce%a3%20%ce%9c%ce%a0%ce%a1%ce%9f%ce%a5%ce%96%ce%99%ce%a9%ce%a4%ce%97%ce%a3%20signed.pdf?sequence=1&isAllowed=y>

14. Acar, A., Uryan, M., Doğrul, A., Karakurt, A. S., & Çelik, D. D. C. (2023). Numerical investigation of hvac systems of a naval ship compartment: Natural ventilation and air-conditioning. *Journal of Naval Sciences and Engineering*, 19(1), 77–100. <https://doi.org/10.56850/jnse.1274350>
15. Third Edition. (n.d.). *To time series and forecasting*. Warin.Ca. Retrieved September 15, 2024, from https://warin.ca/ressources/books/2016_Book_IntroductionToTimeSeriesAndFor.pdf
16. Chatfield, C. (2003). *The analysis of time series*. Chapman and Hall/CRC.
17. (N.d.-f). *Sciencedirect.com*. Retrieved September 15, 2024, from <https://www.sciencedirect.com/topics/mathematics/autoregressive-moving-average-model>
18. Shetty, C. (2020, September 22). *Time series models*. Towards Data Science. <https://towardsdatascience.com/time-series-models-d9266f8ac7b0>
19. (N.d.-g). *Sciencedirect.com*. Retrieved September 15, 2024, from <https://www.sciencedirect.com/topics/economics-econometrics-and-finance/arma-model>
20. (N.d.-h). *Machinelearningmastery.com*. Retrieved September 15, 2024, from <https://machinelearningmastery.com/sarima-for-time-series-forecasting-in-python/>
21. (N.d.-i). *Sciencedirect.com*. Retrieved September 15, 2024, from <https://www.sciencedirect.com/topics/medicine-and-dentistry/akaike-information-criterion>
22. (N.d.-j). *Iop.org*. Retrieved September 15, 2024, from <https://iopscience.iop.org/article/10.1088/1742-6596/2449/1/012027/meta>
23. (N.d.-k). *Sciencedirect.com*. Retrieved September 15, 2024, from <https://www.sciencedirect.com/topics/medicine-and-dentistry/bayesian-information-criterion>
24. Pandian, S. (2021, October 23). *Time Series Analysis: Definition, components, methods, and applications*. Analytics Vidhya. <https://www.analyticsvidhya.com/blog/2021/10/a-comprehensive-guide-to-time-series-analysis/>
25. Introduction, 18 1. (n.d.). *Nonlinear Time Series Models*. Washington.edu. Retrieved September 15, 2024, from <https://faculty.washington.edu/ezivot/econ584/notes/nonlinear.pdf>
26. Shailesh. (2023, February 20). *Various techniques to detect and isolate time series components using python*. Analytics Vidhya. <https://www.analyticsvidhya.com/blog/2023/02/various-techniques-to-detect-and-isolate-time-series-components-using-python/>

Παράρτημα Κώδικα

SQL script για τη δημιουργία του αρχικού συνόλου δεδομένων:

```
SELECT scm.id as ShipComponentID, sjl.DueDate,
       sjl.DateDone,
       LAG(sjl.DateDone) OVER (PARTITION BY scm.id ORDER BY sjl.DateDone) AS PreviousDueDate,
       DATEDIFF(day, LAG(sjl.DateDone) OVER (PARTITION BY scm.id ORDER BY sjl.DateDone),
       sjl.DateDone) AS IntervalDays
       , DATEDIFF(day, sjl.DateDone, sjl.DueDate) AS DateDifference
       , s.Name AS Vessel
from tblshipjoblist sjl
inner join tblShipJobContext sjc on sjl.id=sjc.ShipJobID
inner join tblShipComponentModel scm on scm.id=sjc.ShipComponentID
inner join tblcompanycomponenttree cct on scm.companycomponenttreeid = cct.id
inner join tblship s on scm.ShipID=s.ID
inner join tblUnscheduledJobType ujt on ujt.ID=sjl.UnscheduledJobType
where
scm.Name = 'CE-DE, 12V50DF'
and DoneRunningHours<>0
and sjl.IsDeleted=0
and scm.IsDeleted=0
and cct.IsDeleted=0
and cct.IsDeleted=0
ORDER BY scm.id, sjl.DateDone;
```

Python Κώδικας για την εύρεση του μέσου όρου των καθυστερημένων εργασιών και των εργασιών που έγιναν στην ώρα τους.

```
# Extract year and month from the DateDone column
df['YearMonth'] = pd.to_datetime(df['DateDone']).dt.to_period('M')

# Calculate the mean of delayed and on-time jobs for each month of each year
mean_delayed_by_month = df[df['DateDifference'] >
0].groupby('YearMonth')['DateDifference'].mean().reset_index()
mean_on_time_by_month = df[df['DateDifference'] <=
0].groupby('YearMonth')['DateDifference'].mean().reset_index()

# Rename columns for clarity
mean_delayed_by_month.columns = ['YearMonth', 'MeanDelayedJobs']
mean_on_time_by_month.columns = ['YearMonth', 'MeanOnTimeJobs']

# Merge the two dataframes
mean_jobs_by_month = pd.merge(mean_delayed_by_month, mean_on_time_by_month,
on='YearMonth', how='outer').sort_values('YearMonth')

mean_jobs_by_month
```

Έυρεση της τυπικής απόκλισης των καθυστερημένων εργασιών και των εργασιών που έγιναν στην ώρα τους.

```
# Calculate the standard deviation of delayed and on-time jobs for each month of each year
std_delayed_by_month = df[df['DateDifference'] >
0].groupby('YearMonth')['DateDifference'].std().reset_index()
std_on_time_by_month = df[df['DateDifference'] <=
0].groupby('YearMonth')['DateDifference'].std().reset_index()

# Rename columns for clarity
std_delayed_by_month.columns = ['YearMonth', 'StdDelayedJobs']
std_on_time_by_month.columns = ['YearMonth', 'StdOnTimeJobs']

# Merge the two dataframes
std_jobs_by_month = pd.merge(std_delayed_by_month, std_on_time_by_month,
on='YearMonth', how='outer').sort_values('YearMonth')

std_jobs_by_month
```

Υπολογισμός του μέγιστου αριθμού των αργοπορημένων εργασιών και του ελάχιστου αριθμού των εργασιών που έγιναν στην ώρα τους.

```
# Calculate the max of delayed jobs and min of on-time jobs for each month of each year
max_delayed_by_month = df[df['DateDifference'] >
0].groupby('YearMonth')['DateDifference'].max().reset_index()
min_on_time_by_month = df[df['DateDifference'] <=
0].groupby('YearMonth')['DateDifference'].min().reset_index()

# Rename columns for clarity
max_delayed_by_month.columns = ['YearMonth', 'MaxDelayedJobs']
min_on_time_by_month.columns = ['YearMonth', 'MinOnTimeJobs']

# Merge the two dataframes
extreme_jobs_by_month = pd.merge(max_delayed_by_month, min_on_time_by_month,
on='YearMonth', how='outer').sort_values('YearMonth')

extreme_jobs_by_month
```

Ενοποίηση όλων των παραπάνω υπολογιζόμενων στηλών σε ένα dataframe:

```
# Combine all the calculations into one dataframe
combined_stats_by_month = pd.merge(mean_jobs_by_month, std_jobs_by_month,
on='YearMonth', how='outer')
```

```
combined_stats_by_month = pd.merge(combined_stats_by_month, extreme_jobs_by_month,
on='YearMonth', how='outer')
```

```
combined_stats_by_month
```

Υπολογισμός του μέσου όρου, της τυπικής απόκλισης, του μέγιστου και του ελάχιστου αριθμού των ημερών από την προηγούμενη συντήρηση του μηχανήματος.

```
# Extract year and month from the DateDone column for IntervalDays analysis
df['YearMonth'] = pd.to_datetime(df['DateDone']).dt.to_period('M')

# Calculate the mean, std, max, and min of IntervalDays for each month of each year
interval_stats = df.groupby('YearMonth')['IntervalDays'].agg(['mean', 'std', 'max',
'min']).reset_index()

# Rename columns for clarity
interval_stats.columns = ['YearMonth', 'MeanIntervalDays', 'StdIntervalDays',
'MaxIntervalDays', 'MinIntervalDays']

# Merge with the previously combined stats dataframe
combined_stats_with_interval = pd.merge(combined_stats_by_month, interval_stats,
on='YearMonth', how='outer').sort_values('YearMonth')

combined_stats_with_interval
```

Από τους παραπάνω κώδικες δημιουργήθηκε το τελικό σύνολο δεδομένων με όνομα `combined_stats_with_interval_days.xlsx`, το οποίο έχει την παρακάτω μορφή:

YearMonth	umOfDefect	MeanDelayedJobs	MeanOnTimeJobs	StdDelayedJobs	StdOnTimeJobs	MaxDelayedJobs	MinOnTimeJobs	MeanIntervalDays	StdIntervalDays	MaxIntervalDays	MinIntervalDays
2017-01	11	3,6	0	1,816590212	0	6	0	1,25	1,258305739	3	0
2017-02	22	38	-0,75	70,29935988	1,183215957	180	-3	11,05882353	10,39513461	32	0
2017-03	15	11,28571429	0	16,97757625	0	49	0	15,16666667	11,23171915	39	4
2017-04	13	39,33333333	0	45,74203027	0	91	0	12,36363636	10,41415121	30	1
2017-05	14	6,4	-0,1	11,52388823	0,316227766	27	-1	164,2222222	391,5312177	1204	1
2017-06	13	64,66666667	-0,142857143	147,2449207	0,377964473	365	-1	29,55555556	35,92739592	108	1
2017-07	6	25,8	-1	31,49126863		72	-1	44,33333333	35,72487462	89	1

Ο κώδικας που χρησιμοποιήθηκε για τη δημιουργία των γραφικών παραστάσεων κάθε στήλης με βάση τη στήλη YearMonth:

```
import matplotlib.pyplot as plt

# Convert YearMonth to datetime for proper plotting
combined_df['YearMonth'] = combined_df['YearMonth'].astype(str)
combined_df['YearMonth'] = pd.to_datetime(combined_df['YearMonth'], errors='coerce')
```

```

# List of columns to plot against YearMonth
columns_to_plot = combined_df.columns[1:]

# Create a plot for each column
for column in columns_to_plot:
    plt.figure(figsize=(10, 6))
    plt.plot(combined_df['YearMonth'], combined_df[column], marker='o')
    plt.title(f'YearMonth vs {column}')
    plt.xlabel('YearMonth')
    plt.ylabel(column)
    plt.grid(True)
    plt.xticks(rotation=45)
    plt.tight_layout()
    plt.show()

```

Δημιουργία των συνόλων δεδομένων ελέγχου (df_test) και εκπαίδευσης (df_train). Το df_train αποτελείται από 70 εγγραφές ενώ το df_test αποτελείται από μόνο μία. Για κάθε μηνιαία πρόβλεψη αυξάνεται το df_train κατά μία εγγραφή και το df_test παίρνει σαν εγγραφή τον επόμενο μήνα από αυτόν που είχε στην προηγούμενη πρόβλεψη.

```

df_test = combined_df[['NumOfDefects']].iloc[70]
df_train = combined_df[['NumOfDefects']].iloc[:70]

```

Δημιουργία των γραφικών παραστάσεων του συντελεστή αυτοσυσχέτισης (ACF) και μερικής αυτοσυσχέτισης (PACF) του συνόλου δεδομένων:

```

import matplotlib.pyplot as plt
from statsmodels.graphics.tsaplots import plot_acf, plot_pacf

# Plot ACF
acf_original = plot_acf(df_train, lags=20) # Adjust lags as necessary
plt.gca().set_title('Autocorrelation Function (ACF)', fontsize=14)
plt.gca().set_xlabel('Lags', fontsize=12)
plt.gca().set_ylabel('Autocorrelation', fontsize=12)
plt.show()

# Plot PACF
pacf_original = plot_pacf(df_train, lags=20) # Adjust lags as necessary
plt.gca().set_title('Partial Autocorrelation Function (PACF)', fontsize=14)
plt.gca().set_xlabel('Lags', fontsize=12)
plt.gca().set_ylabel('Partial Autocorrelation', fontsize=12)
plt.show()

```

Εξαγωγή των τιμών του συντελεστή αυτοσυσχέτισης (ACF) και μερικής αυτοσυσχέτισης (PACF) του συνόλου δεδομένων:

```
from statsmodels.tsa.stattools import acf, pacf

# Calculate ACF and PACF values
acf_values = acf(df_train['NumOfDefects'], nlags=20)
pacf_values = pacf(df_train['NumOfDefects'], nlags=20)
# Print ACF and PACF values
print("ACF values:\n", acf_values)
print("\nPACF values:\n", pacf_values)
```

Εφαρμογή της πρώτης διαφοράς στο σύνολο δεδομένων εκπαίδευσης (train df) και η δημιουργία της γραφικής παράστασης του συνόλου ύστερα από την πρώτη διαφορά:

```
df_train_diff = df_train.diff().dropna()

# Plot the differenced time series
plt.figure(figsize=(10, 6))
plt.plot(df_train_diff, label='Differenced NumOfDefects')
plt.title('Time Series Plot of NumOfDefects After First Differencing', fontsize=14)
plt.xlabel('Date', fontsize=12)
plt.ylabel('Differenced NumOfDefects', fontsize=12)
plt.legend()
plt.show()
```

Δημιουργία των γραφικών παραστάσεων ACF και PACF ύστερα από την πρώτη διαφορά:

```
acf_diff = plot_acf(df_train_diff)
plt.gca().set_title('Autocorrelation Function (ACF)', fontsize=14)
plt.gca().set_xlabel('Lags', fontsize=12)
plt.gca().set_ylabel('Autocorrelation', fontsize=12)
plt.show()

pacf_diff = plot_pacf(df_train_diff)
plt.gca().set_title('Partial Autocorrelation Function (PACF)', fontsize=14)
plt.gca().set_xlabel('Lags', fontsize=12)
plt.gca().set_ylabel('Partial Autocorrelation', fontsize=12)
plt.show()
```


Η συνάρτηση auto_arima εφαρμοσμένη στο σύνολο δεδομένων εκπαίδευσης (train_df):

```
from pmdarima import auto_arima

stepwise_fit = auto_arima(df_train['NumOfDefects'],
                          start_p=0, start_q=0,
                          max_p=2, max_q=2,
                          m=12,
                          seasonal=False,          # seasonal=True when seasonal data
                          d=None, trace=True,
                          error_action='ignore',    # don't want to know if an order
does not work
                          suppress_warnings=True,  # don't want convergence warnings
                          stepwise=True)          # set to stepwise

stepwise_fit.summary()
```

Εφαρμογή του μοντέλου ARIMA(2,1,2):

```
from statsmodels.tsa.arima.model import ARIMA

model = ARIMA(df_train, order=(2,1,2))
model_fit = model.fit()
print(model_fit.summary())
```

Εφαρμογή του μοντέλου ARIMA(1,1,1):

```
from statsmodels.tsa.arima.model import ARIMA

model = ARIMA(df_train, order=(1,1,1))
model_fit = model.fit()
print(model_fit.summary())
```

Εφαρμογή του μοντέλου ARIMA(0,1,1):

```
from statsmodels.tsa.arima.model import ARIMA

model = ARIMA(df_train, order=(0,1,1))
model_fit = model.fit()
print(model_fit.summary())
```

Πρόβλεψη του αριθμού των βλαβών για τον επόμενο μήνα:

```
# Forecast the next values based on the length of the test set
forecast_steps = len(df_test)
```

```
forecast = model_fit.forecast(steps=forecast_steps)
```

```
# Print the forecasted values  
print(forecast)
```

Υπολογισμός της πυκνότητας και των καταλοίπων και οι γραφικές παραστάσεις τους συγκεντρωτικά με τα ACF και PACF γραφήματα:

```
import matplotlib.pyplot as plt  
from statsmodels.graphics.tsaplots import plot_acf, plot_pacf  
  
# Calculate the residuals from the model  
residuals = model_fit.resid[1:]  
  
# Plot the residuals and their density  
fig, ax = plt.subplots(1, 2, figsize=(14, 6))  
residuals.plot(title='Residuals', ax=ax[0])  
residuals.plot(title='Density', kind='kde', ax=ax[1])  
plt.show()  
  
# Plot the ACF and PACF of the residuals  
fig, ax = plt.subplots(1, 2, figsize=(12, 6))  
  
# ACF plot  
plot_acf(residuals, lags=20, ax=ax[0])  
ax[0].set_title('Autocorrelation Function (ACF) of Residuals')  
ax[0].set_xlabel('Lags')  
ax[0].set_ylabel('ACF')  
  
# PACF plot  
plot_pacf(residuals, lags=20, ax=ax[1])  
ax[1].set_title('Partial Autocorrelation Function (PACF) of Residuals')  
ax[1].set_xlabel('Lags')  
ax[1].set_ylabel('PACF')  
  
# Show the plots  
plt.tight_layout()  
plt.show()
```