

ΠΑΝΕΠΙΣΤΗΜΙΟ ΠΕΙΡΑΙΩΣ

Σχολή Χρηματοοικονομικής και Στατιστικής



Τμήμα Στατιστικής και Ασφαλιστικής Επιστήμης

ΜΕΤΑΠΤΥΧΙΑΚΟ ΠΡΟΓΡΑΜΜΑ ΣΠΟΥΔΩΝ ΣΤΗΝ ΑΝΑΛΟΓΙΣΤΙΚΗ ΕΠΙΣΤΗΜΗ ΚΑΙ ΔΙΑΧΕΙΡΙΣΗ ΚΙΝΔΥΝΩΝ

ΑΣΦΑΛΙΣΗ ΚΑΙ ΧΡΗΣΗ ΜΕΓΑΛΩΝ ΔΕΔΟΜΕΝΩΝ

Παρασκευή Σταμάτη

Διπλωματική Εργασία

Που υποβλήθηκε στο Τμήμα Στατιστικής και
Ασφαλιστικής Επιστήμης του Πανεπιστημίου Πειραιώς
μέρος των απαιτήσεων για την απόκτηση του
Μεταπτυχιακού Διπλώματος Ειδίκευσης στην
Αναλογιστική Επιστήμη και Διαχείριση Κινδύνων

Πειραιάς

Μάρτιος 2024

Η παρούσα Διπλωματική Εργασία εγκρίθηκε ομόφωνα από την Τριμελή Εξεταστική Επιτροπή που ορίστηκε από τη Συνέλευση του Τμήματος Στατιστικής και Ασφαλιστικής Επιστήμης του Πανεπιστημίου Πειραιώς στην υπ' αριθμ. 22/3/2024 συνεδρίασή της, σύμφωνα με τον Εσωτερικό Κανονισμό Λειτουργίας του Προγράμματος Μεταπτυχιακών Σπουδών στην Αναλογιστική Επιστήμη και Διαχείριση Κινδύνων.

Τα μέλη της Επιτροπής ήταν:

- Επίκουρος Καθηγητής, Ξένος Παναγιώτης (Επιβλέπων)
- Καθηγητής, Μπερσίμης Σωτήριος
- Καθηγητής, Χατζηκωνσταντινίδης Ευστάθιος

Η έγκριση της Διπλωματικής Εργασίας από το Τμήμα Στατιστικής και Ασφαλιστικής Επιστήμης του Πανεπιστημίου Πειραιώς δεν υποδηλώνει αποδοχή των γνώμων του συγγραφέα.

UNIVERSITY OF PIRAEUS
School of Finance and Statistics



Department of Statistics and Insurance Science

POSTGRADUATE PROGRAM IN
ACTUARIAL SCIENCE AND RISK MANAGEMENT

INSURANCE AND BIG DATA

By

Paraskevi Stamati

MSc Dissertation

submitted to the Department of Statistics and Insurance
Science of the University of Piraeus in partial
fulfilment of the requirements for the degree of Master
of Science in Actuarial Science and Risk Management

Piraeus, Greece

March 2024

*Στους γονείς μου
Ιωάννη και Αντιόπη*

Ευχαριστίες

Με την εκπόνηση την διπλωματικής μου θα ήθελα να ευχαριστήσω τον επιβλέπων μου καθηγητή μου, κ. Ξένο Παναγιώτη που στάθηκε αρωγός σε όλη την προσπάθεια που πραγματοποίησα για την συγγραφή της διπλωματικής εργασίας. Επιπλέον, θα ήθελα να ευχαριστήσω την οικογένειά μου που μου στάθηκε καθόλη τη διάρκεια των σπουδών μου.

Περίληψη

Όπως υποδηλώνει το όνομα, τα "μεγάλα δεδομένα" ή αλλιώς big data, αναφέρονται σε τεράστιες ποσότητες δεδομένων σε πολλούς διαφορετικούς τομείς, οι οποίες είναι σημαντικά μεγαλύτερες από αυτές που είναι δυνατόν να αποθηκευτούν, να διαχειριστούν και να αναλυθούν σήμερα με τη χρήση τυπικού λογισμικού ή διαδικτυακών πλατφορμών. Η χρήση των μεγάλων δεδομένων μέσω εξειδικευμένων εφαρμογών υπόσχεται σημαντικές προόδους όσον αφορά τη διαχείριση, τη λήψη αποφάσεων, την πρόβλεψη της εμφάνισης ασθενειών και επιδημιών, καθώς και σημαντική μείωση του κόστους της υγειονομικής περίθαλψης, ιδίως στον τομέα της υγείας, ο οποίος παραδοσιακά παράγει τεράστιο όγκο δεδομένων. Στόχος της παρούσας εργασίας είναι να εξετάσει το ρόλο των μεγάλων δεδομένων στον κλάδο της υγείας, μαζί με τις δυσκολίες και τις ευκαιρίες που αντιμετωπίζει και τις βασικές εφαρμογές τους.

Abstract

Big data, as their name suggests, refers to enormous amounts of data produced in a variety of industries that cannot be managed using conventional software or online platforms since they vastly exceed current capabilities in terms of storage, processing, and analytics. The use of Big Data through specialized applications offers substantial advancements in terms of management, decision-making, disease and epidemic predictions, and considerable decreases in healthcare expenses, particularly in the health sector, which usually generates enormous volumes of data. This thesis seeks to examine the importance of big data in the health care industry, as well as the potential and challenges they confront, and their primary uses.

Πίνακας Περιεχομένων

Περίληψη	
Abstract	
Πίνακας Περιεχομένων	
Εισαγωγή.....	1
1. Θεωρητικό Υπόβαθρο.....	3
1.1 Ορισμός Μεγάλων Δεδομένων	3
1.2 Ιστορική Αναδρομή	5
1.3 Αξία Μεγάλων Δεδομένων	8
1.4 Πλεονεκτήματα & Μειονεκτήματα	10
1.5 Χαρακτηριστικά των Μεγάλων Δεδομένων	12
1.6 Ποσοτική Ανάλυση των Big Data	16
1.7 Τύποι των Big Data & Αξιοποίησή τους.....	18
1.8 Τεχνολογική ανάλυση Big Data	20
1.9 Αποθήκευση Big Data.....	23
1.10 Cloud και Big Data	27
2. Τα Big Data στον τομέα Υγείας.....	30
2.1 Χαρακτηριστικά & Πλεονεκτήματα.....	30
2.2 Ηλεκτρονικά συστήματα στην Ελλάδα	35
2.3 Πηγές Big Data στον Τομέα Υγείας.....	45
3. Τομέας Υγείας & Big Data Analytics.....	48
3.1 Ορισμός & Ταξινόμηση	48
3.1.1 Περιγραφική Αναλυτική	50
3.1.2 Διαγνωστική Αναλυτική	51
3.1.3 Προγνωστική Αναλυτική	51
3.1.4 Καθοδηγητική Αναλυτική.....	53
3.2 Η επιρροή & οι εφαρμογές στον Τομέα της Υγείας.....	53
3.3 Ανάλυση από Βιολογικά Δεδομένα	59
3.4 Εμπορικές εφαρμογές	61
3.5 Ηλεκτρονική Υγεία & Οφέλη Big Data στην Υγεία	63

3.6 Παραδείγματα εφαρμογής Μεγάλων Δεδομένων & Ηλεκτρονικών Βάσεων Δεδομένων	69
4. Big Data & Τιμολόγηση ασφαλιστικών προϊόντων Υγείας	72
4.1 Φαρμακοεπαγρύπνιση	72
4.2 Ηλεκτρονική Συνταγογράφηση	74
4.3 ΕΟΠΥΥ	76
4.4 DRGs & ΚΕΝ	79
4.5 ICD 11	81
4.6 Ηλεκτρονικός Φάκελος Υγείας.....	81
4.7 epSOS	83
4.8 ΕΔΑΠΥ	83
4.9 Σύστημα Ηλεκτρονικής Προέγκρισης.....	84
5. Μοντέλο «Θεραπεία ορμονικής υποκατάστασης και θνησιμότητα».....	86
5.1 Σχεδιασμός Μελέτης	86
5.2 Μέθοδοι Ανάλυσης.....	86
5.2.1 Είδος στατιστικής μεθόδου	86
5.2.2 Στατιστικοί έλεγχοι	87
5.3 Αποτελέσματα	87
5.4 Συμπέρασμα	87
6. Μοντέλο «Επιβίωση ατόμων με σακχαρώδη διαβήτη τύπου 2 στο Η.Β.».....	89
6.2 Μέθοδοι Ανάλυσης.....	89
6.2.1 Είδος στατιστικής μεθόδου	89
6.2.2 Στατιστικοί έλεγχοι	89
6.3 Αποτελέσματα	90
6.4 Συμπέρασμα	90
Συμπεράσματα	91
Βιβλιογραφία	95

Εισαγωγή

Τα δεδομένα υγείας παράγονται όλο και περισσότερο σε μαζικές κλίμακες, σε διάφορα επίπεδα φαινοτύπων και από διαφορετικούς τύπους πόρων. Ταυτόχρονα με τις πρόσφατες τεχνολογικές εξελίξεις τόσο στην υποδομή παραγωγής δεδομένων όσο και στις μεθοδολογίες ανάλυσης δεδομένων, έχουν διατυπωθεί πολλοί ισχυρισμοί ότι τα γεγονότα αυτά θα φέρουν επανάσταση στην υγειονομική περίθαλψη, αλλά οι ισχυρισμοί αυτοί εξακολουθούν να αποτελούν αντικείμενο πολλών συζητήσεων. Η αντιμετώπιση των δυνατοτήτων και των προκλήσεων των μεγάλων δεδομένων στην υγειονομική περίθαλψη απαιτεί την κατανόηση των χαρακτηριστικών των δεδομένων. Εδώ, χαρακτηρίζουμε τους διάφορους άξονες των ιατρικών δεδομένων, περιγράφουμε τις εκτιμήσεις και τα συμβιβαστικά μέτρα που λαμβάνονται κατά τη δημιουργία τέτοιων δεδομένων και τους τύπους ανάλυσης που μπορούν να επιτύχουν τα συγκεκριμένα καθήκοντα. Η ανασκόπησή μας έχει ως στόχο να αναλύσει και να συζητήσει αυτά τα θέματα και να συμβάλει στη συνεχιζόμενη συζήτηση σχετικά με τη μετάβαση σε πόρους μεγάλων δεδομένων και τις δυνατότητές της στην προώθηση της κατανόησης της υγείας και της ασθένειας (Pan et al., 2019).

Η υγεία έχει οριστεί ως "μια κατάσταση πλήρους σωματικής, ψυχικής και κοινωνικής ευεξίας και όχι απλώς η απουσία ασθένειας ή αναπηρίας". Ο ορισμός αυτός μπορεί να επεκταθεί ώστε να θεωρηθεί η υγεία όχι ως μια ενιαία κατάσταση, αλλά μάλλον ως μια δυναμική διαδικασία διαφορετικών καταστάσεων σε διαφορετικές χρονικές στιγμές που από κοινού συνθέτουν μια πορεία υγείας. Η ικανότητα κατανόησης της πορείας της υγείας διαφορετικών ατόμων, του τρόπου με τον οποίο θα εξελίσσονταν κατά μήκος διαφορετικών διαδρομών, του τρόπου με τον οποίο το παρελθόν επηρεάζει την παρούσα και τη μελλοντική υγεία και των πολύπλοκων αλληλεπιδράσεων μεταξύ διαφορετικών καθοριστικών παραγόντων της υγείας με την

πάροδο του χρόνου, είναι από τους πιο δύσκολους και σημαντικούς στόχους της ιατρικής. Μετά τις τεχνολογικές, οργανωτικές και μεθοδολογικές εξελίξεις των τελευταίων ετών, έχει αναδειχθεί μια νέα και πολλά υποσχόμενη κατεύθυνση προς την επίτευξη αυτών των στόχων: η ανάλυση μεγάλων δεδομένων στην υγειονομική περίθαλψη. Με τη ραγδαία αύξηση του όγκου των διαθέσιμων ιατρικών πληροφοριών, ο όρος "μεγάλα δεδομένα" έχει γίνει όλο και πιο δημοφιλής στην ιατρική. Η αύξηση αυτή αναμένεται να συνεχιστεί, καθώς τα δεδομένα από τα ηλεκτρονικά αρχεία υγείας (EHR) και άλλες αναδύμενες πηγές δεδομένων, όπως οι φορητές συσκευές και οι πολυεθνικές προσπάθειες για τη συλλογή και αποθήκευση δεδομένων και βιολογικών δειγμάτων σε καθορισμένες Βιοτράπεζες, θα επεκταθούν. Οι αναλύσεις ιατρικών δεδομένων μεγάλης κλίμακας έχουν τη δυνατότητα να αποκαλύψουν νέες και άγνωστες συσχετίσεις, μοτίβα και τάσεις στα δεδομένα που μπορούν να ανοίξουν το δρόμο σε πολλές επιστημονικές ανακαλύψεις στην παθολογία, την ταξινόμηση, τη διάγνωση, τη θεραπεία και την εξέλιξη των ασθενειών. Αυτές περιλαμβάνουν την κατασκευή υπολογιστικών μοντέλων που χρησιμοποιούν τα δεδομένα προκειμένου να προβλέπουν με ακρίβεια τα κλινικά αποτελέσματα και την εξέλιξη της νόσου, τα οποία έχουν τη δυνατότητα να εντοπίζουν άτομα υψηλού κινδύνου και να τα ιεραρχούν για στρατηγικές έγκαιρης παρέμβασης και να αξιολογούν την επίδραση των πολιτικών δημόσιας υγείας σε δεδομένα πραγματικής ζωής. Ωστόσο, εξακολουθούν να υπάρχουν πολλές προκλήσεις για την εκπλήρωση αυτών των φιλόδοξων στόχων (Pan et al., 2019).

Τα μεγάλα δεδομένα, ή απλώς "Big Data", αναφέρονται στον τεράστιο όγκο δεδομένων που προστίθεται τακτικά στις βάσεις δεδομένων των εταιρειών. Χαρακτηρίζονται από το ευρύ φάσμα πληροφοριών και την περίπλοκη δομή από την οποία συνοδεύονται, και λόγω του τεράστιου όγκου τους, η ανάλυσή τους είναι εξαιρετικά δύσκολη. Αν και η διαθεσιμότητα των τεράστιων δεδομένων βρίσκεται στην επικαιρότητα εδώ και αρκετό καιρό, η ιδέα των "Big Data" απογειώθηκε πραγματικά στις αρχές της δεκαετίας του 2000, όταν ο αναλυτής του κλάδου "Doug Laney" την όρισε με βάση τα τρία "V". Προκειμένου να πλαισιωθούν κατάλληλα τα δεδομένα, γίνονται αναφορές στον όγκο τους (value), στην ταχύτητα (velocity) και στην ποικιλία (variety). Πιο συγκεκριμένα, ο όγκος αναφέρεται στους οργανισμούς που συγκεντρώνουν δεδομένα από διάφορες πηγές, συμπεριλαμβανομένων των μέσων κοινωνικής δικτύωσης, του βιομηχανικού εξοπλισμού, των βίντεο, των φωτογραφιών και του ήχου, καθώς και των έξυπνων συσκευών, των συναλλαγών και μιας ποικιλίας άλλων μέσων που βοηθούν στη συλλογή πληροφοριών. Ο τρόπος με τον οποίο

αποθηκεύονται πλέον τα δεδομένα είναι πιο προηγμένος και οι σχετικές διαδικασίες είναι απλούστερες και ταχύτερες όσον αφορά το κόστος, τον χρόνο και την αποθήκευση. Στη συνέχεια, ως αποτέλεσμα της συνεχούς επέκτασης του δικτύου, τα δεδομένα διακινούνται πλέον μέσω των επιχειρήσεων με εξαιρετικά γρήγορους ρυθμούς. Η υποστηρικτική αρχιτεκτονική αυτού του είδους των δεδομένων είναι κατασκευασμένη έτσι ώστε να ανιχνεύει και να αποτρέπει άμεσα πιθανές κακόβουλες απειλές (Garapati & Garapati, 2018).

Όσον αφορά το ζήτημα της ποικιλίας, τα δεδομένα είναι πλέον διαθέσιμα σε ένα ευρύ φάσμα μορφών, όπως μη δομημένο κείμενο, μηνύματα ηλεκτρονικού ταχυδρομείου, βίντεο, αρχεία ήχου, πληροφορίες που σχετίζονται με μετοχές, οικονομικές συναλλαγές και μη δομημένα αριθμητικά δεδομένα που βρίσκονται σε συμβατικές βάσεις δεδομένων. Λόγω του γεγονότος ότι ο σωστός προσδιορισμός των δεδομένων είναι ζωτικής σημασίας, ο αναλυτής πρέπει να γνωρίζει ποια φορτία δεδομένων είναι εποχιακά και ποια όχι, και πρέπει να ενεργεί αναλόγως. Η εγκυρότητα όσον αφορά την ποιότητα των δεδομένων διαδραματίζει επίσης κρίσιμο ρόλο για τον πάροχο πληροφοριών. Δεδομένου ότι όλα τα δεδομένα και οι πληροφορίες προέρχονται από πολυάριθμες πηγές, είναι πρόκληση να εξακριβωθεί η προέλευση και η ακρίβειά τους, ενώ παράλληλα να συνδεθούν, να αντιστοιχηθούν και να μετασχηματιστούν σε όλα τα απαραίτητα συστήματα (Garapati & Garapati, 2018).

1. Θεωρητικό Υπόβαθρο

1.1 Ορισμός Μεγάλων Δεδομένων

Τα Big Data είναι ένας όρος που χρησιμοποιείται για να περιγράψει τον τεράστιο όγκο δεδομένων με τον οποίο έρχεται συνεχώς αντιμέτωπη μια εταιρεία. Ο όρος " Big Data " χρησιμοποιείται συνήθως για να περιγράψει σύνολα δεδομένων που είναι πολύ μεγάλα και περίπλοκα για να μπορούν να διαχειριστούν από τις συνήθεις σύγχρονες δυνατότητες των υπολογιστών. Πριν από λίγα χρόνια, απλώς δεν υπήρχαν σύνολα δεδομένων όπως αυτά που παράγονται από τις αμέτρητες και ηλεκτρονικές συσκευές που κατέχουν σήμερα οι επιχειρήσεις, οι κυβερνήσεις και οι ιδιώτες και διακινούνται μέσω του διαδικτύου. Αυτά τα σύνολα δεδομένων περιλαμβάνουν όλα τα βίντεο στο YouTube, όλα τα γεωγραφικά δεδομένα που αποστέλλονται από τα κινητά τηλέφωνα

στους παροχείς κινητής τηλεφωνίας, όλες τις αναζητήσεις στο Google, όλες τις καθημερινές ηλεκτρονικές συναλλαγές, όλες τις μυστικές καταγραφές της NSA των μεταδεδομένων των τηλεφωνικών κλήσεων και αμέτρητα άλλα σύνολα δεδομένων. Ωστόσο, αυτό που έχει μεγαλύτερη σημασία από την ποσότητα των δεδομένων που εισέρχονται στη βάση δεδομένων μιας εταιρείας είναι η δομή της και τα αποτελέσματα που θα πρέπει τελικά να προσφέρει. Η ανάλυση των μεγάλων δεδομένων μπορεί να παρέχει εσωτερικές πληροφορίες που βοηθούν στη λήψη αποφάσεων και στις τακτικές επιχειρηματικές δράσεις (Shafqat et al., 2018).

Από το 1990, η φράση χρησιμοποιείται συχνά, με τον John Mashey να αποτελεί τον κύριο υποστηρικτή της. Ως αποτέλεσμα, ο τομέας αυτός εξετάζει συστηματικές μεθόδους εξαγωγής και ανάλυσης πληροφοριών. Με άλλα λόγια, χειρίζεται σύνολα δεδομένων που είναι είτε πολύ σύνθετα είτε πολύ μεγάλης κλίμακας για να τα διαχειριστεί το τυπικό λογισμικό εφαρμογών επεξεργασίας δεδομένων. Τα δομημένα, τα αδόμητα και τα ημιδομημένα δεδομένα αποτελούν όλα μέρος της θεμελιώδους φιλοσοφίας τους, αλλά τα αδόμητα δεδομένα είναι η κύρια εστίασή τους. Τα δεδομένα που μπορούν να αποθηκευτούν, να προσπελαστούν και να υποστούν επεξεργασία με καθορισμένο τρόπο αναφέρονται ως δομημένα δεδομένα. Ένα παράδειγμα δομημένων δεδομένων είναι ο πίνακας προσωπικού μιας εταιρείας. Τα αδόμητα δεδομένα είναι μια κατηγορία για πληροφορίες που δεν έχουν ούτε γνωστή μορφή ούτε δομή. Συνήθως είναι πολύ μεγάλα σε μέγεθος, γεγονός που καθιστά δύσκολη την αφομοίωσή τους και τα εμποδίζει να προσφέρουν αξία στον χρήστη. Οι μηχανές αναζήτησης που βρίσκονται στο διαδίκτυο, όπως η Google, αποτελούν αντανάκλαστικά παραδείγματα μη δομημένων δεδομένων. Τα δεδομένα που περιέχουν και τους δύο προαναφερθέντες τύπους δεδομένων αναφέρονται ως ημιδομημένα δεδομένα (Shafqat et al., 2018).

Τα δεδομένα που είναι αποθηκευμένα σε έναν υπολογιστή σε μορφή XML αποτελούν ενδεικτικό παράδειγμα ημιδομημένων πληροφοριών. Μεγαλύτερη στατιστική ισχύ προσφέρουν τα δεδομένα με περισσότερες γραμμές (πολλές περιπτώσεις), ωστόσο τα δεδομένα με περισσότερα χαρακτηριστικά ή στήλες (υψηλότερη πολυπλοκότητα) ενδέχεται να έχουν υψηλότερο ποσοστό ψευδούς ανακάλυψης. Τα μεγάλα δεδομένα, ή απλώς "Big Data", αναφέρονται στον τεράστιο όγκο δεδομένων που αποκτούν οι σύγχρονες επιχειρήσεις και είναι σχεδόν αδύνατο να επεξεργαστούν με απλές, συμβατικές τεχνικές. Ουσιαστικά αποτελούνται από όλα όσα υπάρχουν σήμερα γύρω μας. Κάθε ενασχόληση με τα ψηφιακά μέσα παράγει δεδομένα. Οι πληροφορίες αυτές συλλέγονται γρήγορα, σε μεγάλες ποσότητες και από ένα ευρύ

φάσμα πηγών. Χρειάζονται όμως τα κατάλληλα εργαλεία, η επεξεργαστική ισχύς και η τεχνογνωσία για να βγει κάτι αξιόλογο από αυτά. Η ανάγκη για παραγωγή δεδομένων αυξάνεται περαιτέρω από τις απαιτήσεις του σύγχρονου κόσμου και το διαρκώς μεταβαλλόμενο περιβάλλον του ηλεκτρονικού εμπορίου (Mehta & Pandit, 2018).

Τα Big Data προσδιορίζονται από έξι (6) βασικά χαρακτηριστικά, τα οποία όμως θα αναλυθούν περαιτέρω αργότερα :

1. Όγκος δεδομένων
2. Ταχύτητα παραγωγής, αποθήκευσης και επεξεργασίας
3. Ποικιλία
4. Μεταβλητότητα
5. Ακρίβεια και
6. Αξία (Mehta & Pandit, 2018).

1.2 Ιστορική Αναδρομή

Η προέλευση των μεγάλων δεδομένων ως σύνολα υπήρχε πάντα, παρόλο που ο όρος "Big Data" είναι σχετικά νέος. Για παράδειγμα, τα πρώτα κέντρα δεδομένων και η δημιουργία των σχεσιακών βάσεων δεδομένων τους σηματοδότησαν την αρχή του κόσμου των δεδομένων στις δεκαετίες του 1960 και του 1970. Λόγω των σημαντικών προόδων που είχαν γίνει μέχρι τότε, η 2α Ιουνίου 1986 αναγνωρίστηκε ως ορόσημο για τη δημιουργία του συστήματος Teradata, το οποίο ήταν το πρώτο επιτυχημένο εμπορικό σύστημα παράλληλων βάσεων δεδομένων. Η Kmart παρέλαβε την πρώτη παράλληλη βάση δεδομένων από την Teradata με αποθηκευτικό χώρο 1TB, έτσι ώστε ο μεγαλύτερος λιανοπωλητής στη Βόρεια Αμερική να αυξήσει τη χωρητικότητα αποθήκευσης δεδομένων του. Η επιχείρηση "Big Data" αντιμετώπισε πολλές δυσκολίες ως αποτέλεσμα της επέκτασης των υπηρεσιών διαδικτύου, η οποία προκάλεσε τη γρήγορη αύξηση τόσο του όγκου όσο και της ποικιλίας των πληροφοριών. Ως αποτέλεσμα, οι επιχειρήσεις έπρεπε να προσαρμοστούν σε αυτές τις εξελίξεις προκειμένου να αποφύγουν τη σύγχυση της αγοράς (Mehta & Pandit, 2018).

Η Google, μια μηχανή αναζήτησης και πάροχος υπηρεσιών αποθήκευσης δεδομένων, ανέπτυξε κατευθυντήριες γραμμές προγραμματισμού κατά την ίδια ανταγωνιστική περίοδο για να λύσει τις δυσκολίες που σχετίζονται με τη διαχείριση και την αποθήκευση δεδομένων. Μια σημαντική αλλαγή στην αρχιτεκτονική των υπολογιστών και στους μηχανισμούς επεξεργασίας δεδομένων μεγάλης κλίμακας ήταν

απαραίτητη, καθώς το περιεχόμενο που δημιουργούσαν οι χρήστες και άλλες πιθανές πηγές οδήγησαν τελικά στη δημιουργία μη διαχειρίσιμων ροών δεδομένων. Οι άνθρωποι άρχισαν να συνειδητοποιούν πόσα δεδομένα παράγαν οι χρήστες μέσω κοινωνικών δικτύων όπως το Facebook, το YouTube και άλλες διαδικτυακές υπηρεσίες περίπου το 2005. Το σύστημα Hadoop και η NoSQL είναι δύο πλαίσια ανοικτού κώδικα που αναπτύχθηκαν την ίδια χρονιά ειδικά για την αποθήκευση και την ανάλυση μαζικών συλλογών δεδομένων. Τα μεγάλα δεδομένα αναπτύχθηκαν ως αποτέλεσμα της δημιουργίας των πλαισίων ανοικτού κώδικα, καθώς έκαναν την τεχνολογία πιο απλή στη χρήση και σημαντικά λιγότερο δαπανηρή στη διατήρησή της. Έκτοτε, ο όγκος των μεγάλων δεδομένων έχει πολλαπλασιαστεί. Τεράστιες ποσότητες δεδομένων εξακολουθούν να παράγονται από τους χρήστες, αλλά δεν είναι πλέον μόνο οι άνθρωποι. Αξιοσημείωτη είναι και η τεχνολογική επανάσταση που πραγματοποίησε η επιχείρηση "Apple" το 2007 με την παρουσίαση του πρώτου "έξυπνου" κινητού τηλεφώνου, ενώ την ίδια χρονιά, χάρη στην εξέλιξη της τεχνολογίας, υπήρχαν περισσότερες κινητές συσκευές συνδεδεμένες στο διαδίκτυο από ό,τι άνθρωποι στον πλανήτη (Mehta & Pandit, 2018).

Το 2011, η IBM παρουσίασε τον υπολογιστή "Deer Blue", ένα τεχνολογικό θαύμα βασισμένο στην τεχνητή νοημοσύνη που μπορούσε να προβλέψει και να "χτυπήσει" τις κινήσεις ενός ανθρώπου. Για να είμαστε πιο ακριβείς, η συσκευή αυτή μπορεί να σαρώσει και να επεξεργαστεί 4 terabytes δεδομένων σε μερικά δευτερόλεπτα. Στη συνέχεια, ο Jim Gray βοήθησε στην ανάπτυξη ενός καινοτόμου συστήματος διαχείρισης βάσεων δεδομένων το 2007, το οποίο αναφέρθηκε ως "The Fourth Paradigm". Η ανάπτυξη εργαλείων και μέσων πληροφορικής για τη σωστή και αποτελεσματική διαχείριση και ανάλυση μαζικών δεδομένων θεωρήθηκε τότε ως ο μόνος τρόπος αντιμετώπισης αυτού του παραδείγματος, ενώ η έννοια των μεγάλων δεδομένων εισήχθη στην πραγματικότητα για πρώτη φορά το 2011 μέσω έρευνας με τίτλο εργασίας "Extracting Values from Chaos" (Εξαγωγή αξιών από το χάος). Έτσι, προκειμένου να υποστηριχθεί η ανάπτυξη της διαχείρισης και αποθήκευσης των μεγάλων δεδομένων, στη συνέχεια ελήφθησαν τα προκαταρκτικά μέτρα με βάση το πρώτο ερευνητικό ερώτημα (Zheng & Guo, 2020).

Η πρωτοβουλία έρευνας και ανάπτυξης μεγάλων δεδομένων, η οποία αποτελείται από περισσότερες από 84 πρωτοβουλίες σε έξι διαφορετικούς τομείς, επιλέχθηκε και δημοσιοποιήθηκε από την κυβέρνηση Ομπάμα το 2012. Χρησιμοποίησε τα μεγάλα δεδομένα για πρώτη φορά στην προεκλογική του εκστρατεία για να συλλέξει

μεμονωμένες ψήφους ως αποτέλεσμα αυτής της έρευνας. Ο Gil Press και η υπόθεσή του έκαναν μια προσπάθεια να πλαισιώσουν τα δεδομένα έτσι ώστε να μην υπάρχει ζήτημα στο συνολικό σύστημα το 2013, αλλά ταυτόχρονα, ένα συνδεδεμένο άρθρο το 2013 υποδείκνυε ότι ο ρυθμός ανάπτυξης συνέβαινε με αξιοσημείωτα υψηλό ρυθμό όπου οι πληροφορίες και τα δεδομένα εισέρχονταν στο σύστημα. Συμπερασματικά, η επιστήμη των δεδομένων, η οποία διαδραματίζει συμπληρωματική υποστηρικτική λειτουργία για τα μεγάλα δεδομένα, ήταν η έννοια που δομούσε σωστά τη διαχείριση και την αποθήκευση αυτών των δεδομένων. Λόγω του γεγονότος ότι μέχρι τότε χρησιμοποιούνταν μια μέθοδος που ήταν όχι μόνο χρονοβόρα αλλά και αναξιόπιστη ως προς το αποτέλεσμά της, ένα χρόνο αργότερα εισήχθη η διαδικασία της ψηφιοποίησης των δεδομένων (Press, 2012). Ο όρος "μεγάλα δεδομένα" χρησιμοποιείται σήμερα συχνότερα για να περιγράψει την εφαρμογή της προγνωστικής ανάλυσης, της ανάλυσης της συμπεριφοράς των χρηστών ή άλλων σύγχρονων τεχνικών ανάλυσης δεδομένων που αντλούν αξία από τεράστιες ποσότητες δεδομένων και σπανιότερα από ένα συγκεκριμένο μέγεθος συνόλου δεδομένων. Καθώς τα δεδομένα συλλέγονται από συσκευές όπως αρχεία καταγραφής λογισμικού, κάμερες, εναέρια μέσα, μικρόφωνα, αναγνώριση ραδιοσυχνοτήτων, αναγνώστες, ασύρματα δίκτυα αισθητήρων, το Διαδίκτυο, φορητές ηλεκτρονικές συσκευές και άλλα μέσα, η ποσότητα και ο αριθμός των διαθέσιμων συνόλων δεδομένων έχουν επεκταθεί γρήγορα. Περισσότερες συσκευές και αντικείμενα συνδέονται καθημερινά στο Διαδίκτυο χάρη στην ανάπτυξη του Διαδικτύου των Πραγμάτων (IoT), συγκεντρώνοντας πληροφορίες σχετικά με τις τάσεις χρήσης των καταναλωτών και την απόδοση των προϊόντων (Zheng & Guo, 2020).

Η επακόλουθη ανάπτυξη της "μηχανικής μάθησης", ενός κλάδου της τεχνητής νοημοσύνης που αποτελείται από μεθόδους που επιτρέπουν στους υπολογιστές να αντλούν γνώση από δεδομένα και να παράγουν εφαρμογές τεχνητής νοημοσύνης, είχε ως αποτέλεσμα την παραγωγή ακόμη περισσότερων δεδομένων. Αυτό υποδηλώνει ότι όσο περνάει ο καιρός και η τεχνολογία εξελίσσεται, ο όγκος των παραγόμενων δεδομένων θα αυξάνεται αναλογικά. Από τη δεκαετία του 1980, ο όγκος των πληροφοριών που μπορούν να αποθηκευτούν παγκοσμίως ανά άτομο διπλασιάζεται περίπου κάθε 40 μήνες. Από το 2012, 2,5 exabytes νέων δεδομένων παράγονται κάθε μέρα. Σύμφωνα με μια έρευνα που αναμένεται από την IDC (International Data Group), μέχρι το 2025 θα υπάρχουν 163 zebabytes δεδομένων. Οι σημερινές επιχειρήσεις αποκτούν έναν απίστευτο όγκο δεδομένων. Πάνω από 2,5 τετράκις εκατομμύρια bytes

δεδομένων παράγονται ετησίως και ο αριθμός αυτός αυξάνεται με εκθετικό ρυθμό, υποστηρίζει η IBM (International Business Machines Corporation). Στην πραγματικότητα, μόλις τα δύο προηγούμενα χρόνια είχε παραχθεί το 90% των δεδομένων που υπήρχαν στον κόσμο το 2013. Αυτό δείχνει ότι το σύνολο της γνώσης που έχει δημιουργήσει η ανθρωπότητα από την αρχή του χρόνου είναι μόλις το 10% του συνόλου. Οι σημερινοί άνθρωποι παράγουν δεδομένα με ιλιγγιώδεις ταχύτητες, ωστόσο η ποσότητα που αποθηκεύεται σε αναλογική μορφή είναι αμελητέα μικρή (Zheng & Guo, 2020).

1.3 Αξία Μεγάλων Δεδομένων

Ένα ευρύ φάσμα ανεκμετάλλευτων προοπτικών αξίας καθίσταται διαθέσιμο μέσω της αξιοποίησης και επεξεργασίας τεράστιων ποσοτήτων δεδομένων, όπου οι ίδιες οι εταιρείες όχι μόνο παρέχουν ένα ευεργετικό αποτέλεσμα αλλά και αναλύουν τη μελλοντική τάση. Τώρα που τα "Big Data" χρησιμοποιούνται και επεξεργάζονται, είναι δυνατόν να πειραματίζονται περιοδικά και να βρίσκουν νέες διαδρομές που είτε άμεσα είτε έμμεσα θα οδηγήσουν σε περισσότερες εφευρέσεις. Η σημασία των δεδομένων παρέχει ένα ανταγωνιστικό πλεονέκτημα για τις επερχόμενες αποφάσεις των εταιρειών. Ως εκ τούτου, μια επιχείρηση καθοδηγείται σε έξυπνες επιχειρηματικές λύσεις που μπορούν να παράγουν περισσότερα κέρδη χάρη στη χρήσιμη φύση αυτών των δεδομένων. Ένας μεγάλος όγκος συλλογής δεδομένων δεν οδηγεί, ωστόσο, άμεσα σε οικονομικά κέρδη για έναν οργανισμό. Αντίθετα, χρησιμεύει ως προϋπόθεση για τη βιώσιμη ανάπτυξη ή εντοπίζει ένα πρόβλημα, το οποίο ακολουθείται από την υιοθέτηση του καλύτερου μοντέλου για την επίλυση ενός σχετικού ζητήματος με βάση τη σχετική έρευνα και ανάπτυξη (Zheng & Guo, 2020).

Επιπλέον, έχει αναφερθεί, με βάση περιστασιακά πραγματοποιηθείσα μελέτη, ότι η χρήση των μεγάλων δεδομένων προσθέτει αξία στις επιχειρήσεις μειώνοντας το κόστος, επιταχύνοντας τη λήψη αποφάσεων και καθιστώντας απλούστερη την ανάπτυξη νέων προϊόντων. Πιο συγκεκριμένα, η μείωση του κόστους προκύπτει από την ψηφιοποίηση, η οποία έχει συμβεί κυρίως τα τελευταία χρόνια. Όσον αφορά το κόστος, το σύστημα Hadoop και οι αναλύσεις που βασίζονται στο cloud προσφέρουν πολλά πλεονεκτήματα για την αποθήκευση μεγάλου όγκου δεδομένων. Οι εταιρείες είναι σε θέση να εξετάζουν τις πληροφορίες γρήγορα χάρη στις βελτιώσεις στις ταχύτητες επεξεργασίας και ανάλυσης. Αυτό τους δίνει τη δυνατότητα να σχεδιάζουν

εκ των προτέρων και να λαμβάνουν τις καλύτερες δυνατές αποφάσεις. Επιπλέον, με τη δυνατότητα χρήσης της ανάλυσης για τη χαρτογράφηση των απαιτήσεων του καταναλωτικού κοινού, δίνεται η δυνατότητα να παρέχουν στοιχεία που θα ανταποκρίνονται πληρέστερα στις προσδοκίες του κοινού τους (Zheng & Guo, 2020).

Η σημασία των " Big Data " βασίζεται στον τρόπο χρήσης τους και όχι μόνο στο πόσα δεδομένα είναι ήδη διαθέσιμα. Όπως προαναφέρθηκε, τα δεδομένα μπορούν να συγκεντρωθούν από οποιαδήποτε πηγή προκειμένου να αξιολογηθούν, να παράσχουν πληροφορίες και να επιτρέψουν τη δημιουργία νέων προϊόντων, βελτιωμένων προσφορών και τη λήψη σοφών αποφάσεων. Οι αναλύσεις υψηλής ανάλυσης σε συνδυασμό με τα μεγάλα δεδομένα μπορούν να χρησιμοποιηθούν για την εκτέλεση εργασιών που σχετίζονται με τις επιχειρήσεις, όπως:

- Προσδιορισμός των βασικών αιτιών και αποτυχιών, ζητημάτων και ελαττωμάτων σε σχεδόν πραγματικό χρόνο.
- Δημιουργία κουπονιών στο σημείο πώλησης βάσει των αγοραστικών συνηθειών του πελάτη.
- Επανυπολογισμός ολόκληρων χαρτοφυλακίων κινδύνου σε λίγα λεπτά.
- Εντοπισμός δόλιας συμπεριφοράς προτού επηρεαστεί ο οργανισμός.

Όπως μπορείτε να δείτε, οι επιχειρήσεις εφαρμόζουν τα "μεγάλα δεδομένα" στα συστήματά τους για να βελτιώσουν τις επιχειρηματικές διαδικασίες, να παρέχουν καλύτερη εξυπηρέτηση πελατών, να αναπτύξουν εξατομικευμένες εκστρατείες μάρκετινγκ και να λάβουν άλλα μέτρα για την αύξηση των πωλήσεων και της κερδοφορίας. Είναι σαφές ότι οι επιχειρήσεις που τα χρησιμοποιούν έχουν σημαντικό ανταγωνιστικό πλεονέκτημα έναντι εκείνων που δεν τα χρησιμοποιούν (Palanisamy & Thirunavukarasu, 2019).

Για παράδειγμα, τα " Big Data " προσφέρουν χρήσιμες πληροφορίες για τους πελάτες. Πληροφορίες τις οποίες οι επιχειρήσεις μπορούν να αξιοποιήσουν για να βελτιώσουν σημαντικά το μάρκετινγκ, τη διαφήμιση και τις προωθητικές ενέργειές τους, σε μια προσπάθεια να κερδίσουν πιο πιστούς πελάτες. Οι επιχειρήσεις μπορούν να ανταποκριθούν πιο άμεσα και με μεγαλύτερη ακρίβεια στις απαιτήσεις και τις επιθυμίες των πελατών τους, αναλύοντας τόσο τα ιστορικά δεδομένα όσο και τα δεδομένα που συλλέγονται επί του παρόντος, ώστε να εξαχθούν συμπεράσματα σχετικά με τις μεταβαλλόμενες προτιμήσεις των καταναλωτών (Alexandru et al., 2018).

1.4 Πλεονεκτήματα & Μειονεκτήματα

Η απόφαση για τη χρήση των "μεγάλων δεδομένων" από μια εταιρεία θα πρέπει να ληφθεί για διάφορους σημαντικούς λόγους. Η υιοθέτηση μιας ολοκληρωμένης λύσης Big Data θα δώσει κατ' αρχάς στην εταιρεία το πλεονέκτημα του χειρισμού των δεδομένων της με τον καλύτερο δυνατό τρόπο. Παρέχουν επίσης βελτιωμένες δυνατότητες όσον αφορά την ταχύτητα και τη χωρητικότητα για τις εταιρείες που χρησιμοποιούν την έννοια του "σύννεφου" για την αποθήκευση των δεδομένων τους. Το γεγονός ότι ο τελικός χρήστης, ο οποίος είναι συνήθως, αν και όχι πάντα, ο πελάτης της εταιρείας, μπορεί να χρησιμοποιήσει άμεσα αυτά τα δεδομένα είναι επίσης πλεονεκτικό. Τα εργαλεία ανάλυσης δεδομένων παρέχουν όλο και περισσότερες κρίσιμες πληροφορίες που, αφενός, καθορίζουν τη χρησιμότητα των δεδομένων και, αφετέρου, βοηθούν την εταιρεία να επιταχύνει τη βελτίωση της ανταγωνιστικότητάς της έναντι άλλων ανταγωνιστών που δραστηριοποιούνται στον ίδιο κλάδο, καθώς αποκτούν αναγνώριση, αναγκαιότητα και ευρεία χρήση. Σε σύγκριση με τον τρόπο με τον οποίο γινόταν ο χειρισμός των δεδομένων στο παρελθόν, οι νέες τεχνολογίες και τα μέσα Big data επιτρέπουν στους αναλυτές να αξιολογούν περισσότερα δεδομένα σε λιγότερο χρόνο. Ως αποτέλεσμα, η παραγωγικότητα της εν λόγω επιχείρησης αυξάνεται ανάλογα με τα συμπεράσματα στα οποία καταλήγουν οι αναλυτές. Ως αποτέλεσμα της ανάλυσης δεδομένων, μια επιχείρηση μπορεί να προσδιορίσει γιατί συνδέεται ή δεν συνδέεται με ορισμένες ομάδες-στόχους, χρησιμοποιώντας τη στρατηγική των μεγάλων δεδομένων για να μάθει τις προτιμήσεις των πελατών και να συγκεντρώσει δυνατότητες για να κερδίσει νέους πελάτες (Alexandru et al., 2018).

Οι οργανισμοί μπορούν επίσης να λάβουν τις καλύτερες δυνατές αποφάσεις χρησιμοποιώντας ποσοτικά δεδομένα για να βοηθήσουν στη λήψη αποφάσεων. Ως αποτέλεσμα, ο εν λόγω οργανισμός μπορεί να θέσει πιο εφικτούς στόχους. Επιπλέον, η χρήση ποιοτικών δεδομένων θα διευκολύνει σημαντικά τις διαδικασίες που μειώνουν την πιθανότητα λαθών και αποτυχιών, διευκολύνοντας τους οργανισμούς να αποκτήσουν σημαντικό ανταγωνιστικό πλεονέκτημα στους αντίστοιχους κλάδους. Η χρήση των μεγάλων δεδομένων είχε επίσης ως αποτέλεσμα τη μείωση του κόστους, η οποία με τη σειρά της οδήγησε σε αντίστοιχη μείωση της πιθανότητας απειλών και κινδύνων για τις επιχειρήσεις, ενώ παράλληλα επέτρεψε τον ταχύτερο εντοπισμό ενδεχόμενης απάτης, ιδίως όσον αφορά τον χρηματοπιστωτικό κλάδο. Αποτέλεσμα της

χρήσης των βέλτιστων πρακτικών στις επιχειρήσεις είναι η αύξηση των εσόδων και η βελτίωση της εξυπηρέτησης των πελατών (Alexandru et al., 2018).

Επιπλέον, ο υψηλότερος ρυθμός καινοτομίας ενθαρρύνει την ανάπτυξη νέων προϊόντων που ικανοποιούν πλήρως τις επιθυμίες του αγοραστικού κοινού, ενώ παράλληλα ενισχύει τον πιο "υγιή" ανταγωνισμό μεταξύ των ανταγωνιστικών επιχειρήσεων του κλάδου. Συμπερασματικά, πολλές διαδικασίες έχουν επιτύχει χωρίς να απαιτείται ο ίδιος χρόνος για την εκτέλεσή τους λόγω της αξιοποίησης των μεγάλων δεδομένων. Έτσι, έχει δημιουργηθεί ένα περιβάλλον που χαρακτηρίζεται από την ταχεία λήψη αποφάσεων υψηλής ποιότητας, η οποία, σε συνδυασμό με τη δημιουργική επιχειρηματικότητα, αυξάνει την κερδοφορία των επιχειρήσεων (Alexandru et al., 2018).

Τα μεγάλα δεδομένα παρουσιάζουν ορισμένες δυσκολίες όσον αφορά τη διαχείριση, την επεξεργασία και την αποθήκευση των δεδομένων. Παρά τα προαναφερθέντα πλεονεκτήματα που μπορούν να προσφέρουν σε έναν τομέα και στη συνέχεια σε μια επιχείρηση, υπάρχει επίσης η πιθανότητα αποκλίσεων. Τα πρωταρχικά ζητήματα προστασίας της ιδιωτικής ζωής, της ασφάλειας και του βαθμού προστασίας της ιδιωτικής ζωής των πολιτών που έχουν τεθεί σε σχέση με αυτού του είδους τα δεδομένα. Η πιθανότητα κατάχρησης δεδομένων έχει αυξηθεί δραματικά τα τελευταία χρόνια μαζί με τη συλλογή και χρήση των "μεγάλων δεδομένων". Η διαφύλαξη της ιδιωτικής ζωής και της ιδιοκτησίας των δεδομένων είναι το κύριο θέμα που προκαλεί τη μεγαλύτερη ανησυχία. Παρά το γεγονός ότι η εποχή των μεγάλων δεδομένων βρίσκεται ακόμη στα σπάργανα, έχουν υπάρξει αρκετές περιπτώσεις παραβίασης της ιδιωτικής ζωής των χρηστών. Το πρόβλημα αυτό είναι αποτέλεσμα της αδυναμίας ορισμένων υπολογιστικών συστημάτων να εντοπίσουν πληροφορίες που θα έπρεπε να θεωρούνται προσωπικά δεδομένα. Λόγω της δυνατότητας άμεσης επιρροής στη συμπεριφορά των καταναλωτών, μπορεί να είναι πρόκληση η διατήρηση της ανωνυμίας όσον αφορά συγκεκριμένους παράγοντες όπως η ηλικία, το φύλο και οι προτιμήσεις των καταναλωτών. Ωστόσο, συστήματα που έχουν να κάνουν με ευαίσθητα προσωπικά δεδομένα, όπως είναι η υγειονομική περίθαλψη, τα κυβερνητικά συστήματα, αλλά και η κοινωνική δικτύωση έχουν υψηλές ποσότητες ευαίσθητων ζητημάτων (Fatt et al., 2018).

Επιπλέον, ο καθένας στον βιομηχανικό κόσμο μπορεί να συνδεθεί με τουλάχιστον ένα περιστατικό σε μια ηλεκτρονική βάση δεδομένων και ο καθένας θα μπορούσε να χρησιμοποιήσει αυτές τις πληροφορίες για εκφοβισμό ή εκβιασμό (Lazer,

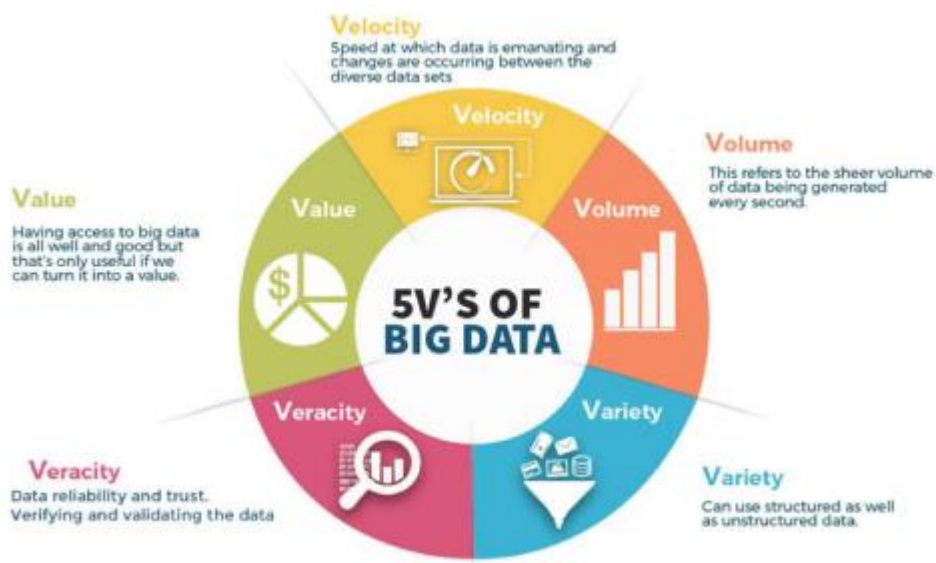
Kennedy, King, King, Vespignani, 2014). Ο Γενικός Κανονισμός για την Προστασία Δεδομένων υιοθετήθηκε από την Ευρωπαϊκή Ένωση ως αποτέλεσμα της δημόσιας κατακραυγής που προκάλεσαν τα Μεγάλα Δεδομένα σχετικά με τις παραβιάσεις δεδομένων και άλλα ζητήματα Γενικού Κανονισμού Προστασίας Δεδομένων (GDPR). Τον Μάιο του 2018 τέθηκε σε ισχύ ένας νόμος για την προστασία των προσωπικών δεδομένων (Fatt et al., 2018).

Ο GDPR θέτει περιορισμούς στα είδη των δεδομένων που μπορούν να συλλέγουν οι οργανισμοί και ζητά τη συγκατάθεση του ατόμου. Επιπλέον, διαθέτει ρήτρα που καλύπτει το δικαίωμα των κατοίκων της Ευρωπαϊκής Ένωσης να ζητούν από τις επιχειρήσεις να διαγράψουν τα δεδομένα τους. Ο νόμος περί απορρήτου της Καλιφόρνιας (CCPA) σκοπεύει να παρέχει στους κατοίκους της Καλιφόρνιας μεγαλύτερο έλεγχο σχετικά με τον τρόπο συλλογής και χρήσης των προσωπικών τους δεδομένων από τις επιχειρήσεις της πολιτείας, παρόλο που δεν υπάρχουν ανάλογοι ομοσπονδιακοί κανόνες στις ΗΠΑ. Ο CCPA εγκρίθηκε την 1η Ιανουαρίου 2020, αφού υπογράφηκε το 2018. Εκτός από το ζήτημα που περιγράφηκε προηγουμένως, υπάρχει επίσης η πιθανότητα οι πληροφορίες που συλλέγονται μέσω των Big Data να έχουν παραπλανητικό χαρακτήρα. Πληροφορίες και δεδομένα παρέχονται συχνά ακόμη και αν είναι συχνά αναληθή και χρησιμεύουν μόνο για την υποστήριξη ενός επιχειρήματος ή δεν έχουν καμία σχέση με τις πληροφορίες που αναζητούνται. Ως εκ τούτου, γίνεται κατανοητό ότι η λειτουργία των Μεγάλων Δεδομένων επιβάλλει την ανάγκη ύπαρξης κατάλληλης υποδομής, καθώς ο αριθμός τους είναι βέβαιο ότι θα αυξηθεί γρήγορα, δεδομένου ότι η επεξεργασία δεδομένων και οι δίσκοι αποθήκευσης αναμένεται να συνεχίσουν να αυξάνονται με την πάροδο του χρόνου (Fatt et al., 2018).

1.5 Χαρακτηριστικά των Μεγάλων Δεδομένων

Τα μεγάλα δεδομένα αναγνωρίζονται από μερικά βασικά χαρακτηριστικά. Τα πρώτα κύρια χαρακτηριστικά των Μεγάλων Δεδομένων, δηλαδή ο "Όγκος - όγκος", η "Ταχύτητα" και η "Ποικιλία", καθορίστηκαν από τον Douglas Laney σε μια ερευνητική εργασία του 2001 με τίτλο "Διαχείριση 3D δεδομένων: Data Volume, Velocity, and Variety" (Έλεγχος του όγκου, της ταχύτητας και της ποικιλίας των δεδομένων). Για περισσότερο από μια δεκαετία, τα προαναφερθέντα τρία χαρακτηριστικά -που συχνά

αναφέρονται ως "3V"- έγιναν οι τρεις ευρέως αποδεκτές διαστάσεις της ποσοτικοποίησης των μεγάλων δεδομένων. Το μοντέλο 3Vs έχει επεκταθεί διαχρονικά καθώς η έρευνα και οι εφαρμογές των μεγάλων δεδομένων έχουν προχωρήσει, προσθέτοντας νέα χαρακτηριστικά (Vs), αλλά για τους σκοπούς του παρόντος εγγράφου, θα επικεντρωθούμε στα πέντε κύρια που φαίνονται στο Σχήμα 1 (Patgiri & Ahmed, 2016).



Εικόνα 1: 5V's of Big Data. Πηγή: <https://www.techentice.com/the-data-veracity-big-data/>

Όγκος – Volume: Ο όγκος, δηλαδή το πόσα δεδομένα παράγονται και διατηρούνται, ουσιαστικά αποφασίζει αν το υλικό θα χαρακτηριστεί ως "μεγάλο" ή όχι. Θα πρέπει να γίνει κατανοητό ότι όταν μιλάμε για το σύνολο δεδομένων, δεν μιλάμε για ένα μόνο δείγμα. Υπάρχει ένας τεράστιος όγκος πληροφοριών που παράγεται καθημερινά από πολυάριθμες πηγές, με αποτέλεσμα η ανάγκη για αποθήκευση και επεξεργαστική ισχύ να αυξάνεται ραγδαία. Στην πραγματικότητα, ο όγκος των δεδομένων αυξάνεται ραγδαία, καθώς παράγονται με εκθετικό ρυθμό χρόνο με το χρόνο (Cano et al., 2017).

Ταχύτητα – Velocity: Ταχύτητα είναι ο χρόνος που απαιτείται για την παραγωγή, αποθήκευση, επεξεργασία και ανάλυση δεδομένων, τα οποία παράγονται συνεχώς και με γρήγορο ρυθμό. Σε αντίθεση με τις εβδομαδιαίες ή μηνιαίες ενημερώσεις που συνέβαιναν στις τυπικές βάσεις δεδομένων αποθήκης δεδομένων, τα

σύνολα μεγάλων δεδομένων ενημερώνονται σε πραγματικό χρόνο ή πολύ κοντά σε αυτόν (Cano et al., 2017).

Ποικιλία - Variety: Η ποικιλομορφία αναφέρεται επίσης στην απαίτηση να συγκεντρώνονται, να οργανώνονται και να χρησιμοποιούνται δεδομένα από πολλές πηγές. Ο όρος Big Data χρησιμοποιείται για να περιγράψει μια ποικιλία ειδών δεδομένων, συμπεριλαμβανομένων εκείνων που έχουν συγκεντρωθεί από πολυάριθμες πηγές και εμπίπτουν στις κατηγορίες εικόνας, ήχου, βίντεο, κειμένου κ.λπ. Ως πηγές πληροφοριών μπορούν να χρησιμοποιηθούν, μεταξύ άλλων, ροές κλικ, αρχεία καταγραφής συστημάτων και συστήματα επεξεργασίας ροών. Αυτές οι πηγές συχνά παράγουν συνεχώς τεράστιες ποσότητες δεδομένων. Ανάλογα με τον τύπο και την πηγή των δεδομένων, μπορούν επίσης να κατηγοριοποιηθούν ως δομημένα, ημιδομημένα ή αδόμητα (Rouse, 2019).

Ποιότητα – Veracity: Μια από τις σημαντικότερες μεταγενέστερες επεκτάσεις των Vs, Ποιότητα Ουσιαστικά, είναι ένα σημάδι του πόσο αξιόπιστο είναι το σύνολο δεδομένων μας ή πόση βεβαιότητα, ορθότητα και ουσία επιδεικνύει. Η πλειονότητα των δεδομένων που παράγονται σήμερα προέρχεται από μη επεξεργασμένες πηγές, όπως ιστότοποι και μέσα κοινωνικής δικτύωσης, αλλά η ακρίβεια και η ποιότητά τους είναι αμφισβητήσιμες. Η αξία των δεδομένων αποδεικνύεται από την ορθότητά τους, γεγονός που αυξάνει τη σημασία αυτού του χαρακτηριστικού, διότι είναι πρόκληση να αναγνωρίζονται δεδομένα από αμφίσημες πηγές, γεγονός που μειώνει το διαμέτρημα της συνολικής συλλογής δεδομένων και της συνολικής ανάλυσης. Μια εφαρμογή που χρησιμοποιεί αυτό το χαρακτηριστικό καθιστά σαφή τη σημασία του. Για παράδειγμα, μια επιχείρηση μπορεί να χρησιμοποιήσει τα ευρήματα μιας μελέτης συμπεριφοράς των πελατών για να βελτιώσει τις ήδη υπάρχουσες προσφορές ή να λανσάρει νέες που να ανταποκρίνονται ειδικά στις επιθυμίες και τις απαιτήσεις των πελατών. Εάν τα δεδομένα που χρησιμοποιούνται είναι ανακριβή, η απόφαση που λαμβάνεται με την πρόθεση που εξάγεται θα είναι επίσης λανθασμένη. Αυτό θα έχει τρομερά αποτελέσματα για την επιχείρηση (Cano et al., 2017).

Αξία – Value: Όταν τα δεδομένα εξάγονται από μεγάλα σύνολα δεδομένων και βρίσκονται, ορίζεται η ιδιότητα Value. Όπως έχει ήδη ειπωθεί, δεν προσθέτουν όλα τα δεδομένα αξία σε μια εταιρεία ή έναν οργανισμό, διότι η χρήση λανθασμένων δεδομένων υποβαθμίζει την ισχύ των ευρημάτων της ανάλυσης. Επιπλέον, τα άχρηστα δεδομένα που μπορεί να κατέχει μια εταιρεία έχουν αξία. Κατά συνέπεια, τα δεδομένα που μένουν αχρησιμοποίητα (γνωστά και ως "dark data") δεν έχουν καμία αξία για τον

οργανισμό και δεν έχουν περαιτέρω χρησιμότητα. Ένας άλλος λόγος για τον οποίο το χαρακτηριστικό Value είναι σημαντικό είναι ο εξής. Μπορεί να μειώσει τον συνολικό όγκο δεδομένων και να αντιμετωπίσει τα ζητήματα που σχετίζονται με την αποθήκευση και τη διαχείριση τεράστιων όγκων δεδομένων. Αυτό συμβαίνει όταν μεταβαίνουμε από δεδομένα χαμηλού επιπέδου (όπως τα ακατέργαστα δεδομένα του αισθητήρα) σε δεδομένα υψηλού επιπέδου (αξίας) (Khattak, et al., 2018).

Μεταβλητότητα – Variability: Το χαρακτηριστικό της μεταβλητότητας προστέθηκε καθώς αναπτύχθηκαν οι εφαρμογές μεγάλων δεδομένων. Ο όρος "μεταβλητότητα" αναφέρεται σε ένα σύνολο δεδομένων που έχουν διττή σημασία, δηλαδή δεδομένα που μπορούν να ερμηνευθούν είτε με θετική είτε με αρνητική έννοια, αλλά των οποίων η σημασία μπορεί επίσης να αλλάξει με την πάροδο του χρόνου- για παράδειγμα, κάτι που σήμερα εκλαμβάνεται ως αρνητικό μπορεί να έχει ουδέτερη ή ακόμη και θετική σημασία στο μέλλον, και το αντίστροφο. Καθώς η σωστή ερμηνεία των απόψεων των χρηστών για ένα θέμα είναι ζωτικής σημασίας για τις επιχειρήσεις ώστε να εξάγουν διορατικά συμπεράσματα και να λαμβάνουν σοφές αποφάσεις, είναι επιτακτική ανάγκη να προσδιορίζεται με ακρίβεια η σημασία μιας φράσης.

Οπτικοποίηση – Visualization: Μια βασική συνιστώσα των μεγάλων δεδομένων είναι η οπτικοποίηση δεδομένων, η οποία περιγράφει τη διαδικασία μετατροπής ενός συνόλου αριθμητικών δεδομένων σε οπτικές πληροφορίες με τη μορφή γραφικών παραστάσεων και διαγραμμμάτων που μπορούν να χρησιμοποιηθούν από τις επιχειρήσεις για την απόκτηση κατανόησης και γνώσης (Dontha, LinkedIn, 2017).

Εγκυρότητα – Validity: Η εγκυρότητα αναφέρεται στο πόσο ακριβή και λεπτομερή είναι τα δεδομένα που θα χρησιμοποιηθούν για την ανάλυση, παρόμοια με την ποιότητα. Σύμφωνα με το περιοδικό Forbes, ένας αναλυτής δεδομένων ξοδεύει περίπου το εξήντα τοις εκατό (60%) του χρόνου του προσπαθώντας να καθαρίσει τα δεδομένα πριν αρχίσει να τα αναλύει, επειδή το σύνολο δεδομένων μπορεί να περιέχει ελλειπείς ή ξεπερασμένες πληροφορίες.

Τρωτότητα – Vulnerability: Η ευπάθεια των μεγάλων δεδομένων μπορεί να χαρακτηριστεί ως τα ελαττώματα ασφαλείας του συνόλου δεδομένων, τα οποία το καθιστούν ευάλωτο σε πιθανές εχθρικές επιθέσεις. Η ευπάθεια των μεγάλων δεδομένων μπορεί να χαρακτηριστεί ως τα ελαττώματα ασφαλείας του συνόλου δεδομένων, τα οποία το καθιστούν ευάλωτο σε ενδεχόμενες επιθέσεις χάκερ και έχουν καταστροφικές συνέπειες για την επωνυμία και την οικονομική υγεία μιας εταιρείας.

Αστάθεια – Volatility: Η αστάθεια που συζητήθηκε παραπάνω συγγέεται συχνά με τη μεταβλητότητα. Η μεταβλητότητα, από την άλλη πλευρά, αντικατοπτρίζει μια μη προγραμματισμένη ή μη αναγνωρισμένη μεταβολή στο σύνολο των δεδομένων που συμβαίνει ξαφνικά ή απότομα. Αντίθετα, η μεταβλητότητα συνδέεται με την αλλοίωση, τη μετάλλαξη και την τροποποίηση του συνόλου δεδομένων, οι οποίες έγιναν με τις καλύτερες προθέσεις και ως αποτέλεσμα της φυσικής εξέλιξης των πραγμάτων (Khattak, et al., 2018).

1.6 Ποσοτική Ανάλυση των Big Data

Τα μεγάλα δεδομένα, που αποκαλούνται επίσης "Big Data", δεν είναι μόνο μια απλή συλλογή δεδομένων. Πρέπει να αναλυθούν προκειμένου η εν λόγω επιχείρηση ή οργανισμός να επωφεληθεί από αυτά, επομένως περνούν από διάφορα στάδια ανάλυσης. Αυτά είναι τα στάδια της ανάλυσής τους:

- α) η απόκτηση και καταγραφή,
- β) ο καθορισμός και η αποθήκευση,
- γ) η ενσωμάτωση στο σύνολο των δεδομένων,
- δ) η απεικόνιση και η ανάλυση και

ε) η ερμηνεία και διάδοση των αποτελεσμάτων. Οι πέντε διαδικασίες που συνθέτουν τη διαδικασία ανάλυσης πληροφοριών περιγράφονται λεπτομερέστερα στις παραγράφους που ακολουθούν.

- Απόκτηση και καταγραφή: Δεδομένα αποκτώνται και συλλέγονται καθημερινά ως αποτέλεσμα της χρήσης των μέσων κοινωνικής δικτύωσης, επειδή όλες οι "έξυπνες" συσκευές αφήνουν το ψηφιακό τους αποτύπωμα. Αυτό επιτρέπει την καταγραφή ποικίλων δεδομένων που μπορούν στη συνέχεια να διατηρηθούν και να αξιολογηθούν λεπτομερέστερα κατά τη διάρκεια της ταξινόμησης και της αξιολόγησης των γενικών δεδομένων.
- Προσδιορισμός και αποθήκευση: Η διαθεσιμότητα της σχετικής πηγής και οι δυνατότητες ανάλυσης είναι τα θέματα του δεύτερου σταδίου, που είναι αυτό του προσδιορισμού. Κατά συνέπεια, όταν επιχειρείται η απόκτηση ενός ακριβούς αποτελέσματος, μια μόνο πηγή είναι συχνά ανεπαρκής για την επίτευξη ενός συγκεκριμένου αποτελέσματος. Ωστόσο, είναι σημαντικό να διπλασιαστεί ο χρόνος ανάλυσης των πηγών πληροφοριών, εάν οι πηγές έχουν σημαντική διαφορά

(Lohr, 2014). Η διαδικασία εξαγωγής πληροφοριών από ένα "σύνολο δεδομένων" αναφέρεται μερικές φορές ως "ανάλυση δεδομένων" με την ευρύτερη έννοια (OECD, 2015). Ωστόσο, η διαδικασία ανάλυσης πραγματοποιείται με συγκεκριμένη μέθοδο όταν οι πληροφορίες συλλέγονται από συγκεκριμένες πηγές. Οι στρατηγικές που χρησιμοποιούνται με βάση ένα συγκεκριμένο κομμάτι μελέτης φαίνεται να ομαδοποιούνται για μια πιο αποτελεσματική διεξαγωγή ταυτόχρονα. Η συγχώνευση δεδομένων, οι οποίες είναι στρατηγικές που χρησιμοποιούνται για τον συνδυασμό του όγκου των δεδομένων από διάφορες πηγές, περιλαμβάνεται ως η πρώτη ταξινόμηση. Το επόμενο βήμα είναι η εξόρυξη δεδομένων, η οποία αναφέρεται σε διαδικασίες και επαναλαμβανόμενα μοτίβα που εμφανίζονται για την εξαγωγή ενός γενικού συμπεράσματος πιο γρήγορα και αποτελεσματικά. Αντίθετα, η βελτιστοποίηση ενσωματώνει προσεγγίσεις των οποίων στόχος είναι η βελτίωση της απόδοσης με βάση ένα σύνολο κριτηρίων. Η κατηγοριοποίηση ολοκληρώνεται με την αναφορά στην οπτικοποίηση, η οποία αναφέρεται σε μεθόδους αναπαραγωγής οπτικών εικόνων και διαγραμμάτων των αποτελεσμάτων της επεξεργασίας δεδομένων. Οι μέθοδοι αυτές επιτρέπουν την πιο άμεση και αποτελεσματική επεξεργασία των πληροφοριών από ό,τι προηγουμένως, ενώ παράλληλα επιτρέπουν την ακόμη πιο εμπειριστατωμένη εξέταση των πληροφοριών.

- Ενσωμάτωση με υφιστάμενα δεδομένα: Η διαδικασία συνδυασμού νέων δεδομένων με παλαιότερα δεδομένα που έχουν συλλεχθεί στο παρελθόν, προκειμένου να δημιουργηθούν αποφάσεις που είναι όσο το δυνατόν πιο ακριβείς.
- Οπτικοποίηση και ανάλυση: Λόγω της πολυπλοκότητας και των 3V's που συζητήθηκαν παραπάνω, οι παραδοσιακές στατιστικές τεχνικές δεν είναι επιτυχείς στην επίλυση των προκλήσεων των μεγάλων δεδομένων. Για την έρευνα των μεγάλων δεδομένων, οι μέθοδοι που χρησιμοποιούν εργαλεία εξόρυξης δεδομένων και οπτικοποίησης είναι πιο αποτελεσματικές. Επειδή δεν βασίζονται σε καθιερωμένες υποθέσεις, οι τεχνικές εξόρυξης δεδομένων διαφέρουν από τις συμβατικές τεχνικές ανάλυσης. Εντοπίζοντας μοτίβα που θα μπορούσαν να οδηγήσουν σε ένα αποτέλεσμα μέσα στα δεδομένα, οι τεχνικές εξόρυξης δεδομένων "αφήνουν τα δεδομένα να μιλήσουν". Οι διαδικασίες που παρατίθενται παρακάτω αποτελούν σημείο αναφοράς για το μέρος της οπτικοποίησης και ανάλυσης δεδομένων της διαδικασίας.

❖ Ταξινόμηση (classification) : Η κατανομή και η κατηγοριοποίηση πληροφοριών ή γεγονότων με βάση κοινά χαρακτηριστικά συνιστούν ταξινόμηση.

❖ Ομαδοποίηση (clustering) : Η ομαδοποίηση είναι η διαδικασία συγκέντρωσης παρόμοιων μοτίβων στα δεδομένα.

❖ Μοντέλα πρόβλεψης (prediction) : Η ανάλυση παλινδρόμησης (regression analysis) ή άλλες μέθοδοι χρησιμοποιούνται για τον εντοπισμό πιθανών συνδέσεων μεταξύ δεδομένων

❖ Σχέσεις μεταξύ των δεδομένων (association) : Η εύρεση συνδέσεων μεταξύ τιμών δεδομένων από ένα ή περισσότερα σύνολα δεδομένων είναι το επίκεντρο της τεχνικής ανάλυσης σχέσεων (συσχέτισης).

❖ Εντοπισμός ανωμαλιών (anomaly detection) : Η εύρεση ακραίων τιμών ή ακραίων περιπτώσεων διακοπής μιας ακολουθίας σε ένα σύνολο δεδομένων είναι γνωστή ως ανίχνευση ανωμαλιών.

❖ Περίληψη : καταγραφή των διαφόρων χαρακτηριστικών ή patterns που εμφανίζονται σε ένα ή περισσότερα data sets.

• Ερμηνεία και διάδοση αποτελεσμάτων : Η λήψη πιο τεκμηριωμένων αποφάσεων καθίσταται δυνατή χάρη στην απεικόνιση των ευρημάτων σε πραγματικό χρόνο για τους κατοίκους και τους αρμόδιους κάθε πόλης.

1.7 Τύποι των Big Data & Αξιοποίησή τους

Υπάρχουν πολυάριθμες πηγές μεγάλων δεδομένων. Άλλες που θα καλυφθούν λεπτομερέστερα στο μάθημα περιλαμβάνουν συστήματα επεξεργασίας συναλλαγών, βάσεις δεδομένων πελατών, έγγραφα, μηνύματα ηλεκτρονικού ταχυδρομείου, ιατρικά αρχεία, αρχεία καταγραφής κλικ στο διαδίκτυο, εφαρμογές κινητών τηλεφώνων, κοινωνικά δίκτυα, δεδομένα που παράγονται από μηχανές, όπως αρχεία καταγραφής δικτύων και διακομιστών, δεδομένα από αισθητήρες σε μηχανήματα παραγωγής, βιομηχανικό εξοπλισμό και συσκευές του διαδικτύου των πραγμάτων κ.λπ. Ορισμένα παραδείγματα από αυτά αναφέρθηκαν παραπάνω, ενώ άλλα θα καλυφθούν λεπτομερέστερα στο μάθημα. γνώσεις σχετικά με τοποθεσίες, επιστήμες και άλλα πράγματα. Τα μεγάλα δεδομένα περιλαμβάνουν επίσης δεδομένα ροής που υποβάλλονται σε επεξεργασία και συλλέγονται σε καθημερινή και συνεχή βάση με τη

μορφή εικόνων, βίντεο και αρχείων ήχου (Ahmad & Tripathi, 2018). Οι άνθρωποι πρέπει να προσαρμόζονται και να μαθαίνουν προκειμένου να συμβαδίζουν με τις νέες εξελίξεις και να αποφεύγουν τις εκπλήξεις. Αυτό συμβαίνει επειδή οι χρήστες έχουν ήδη συνηθίσει να χρησιμοποιούν έξυπνες συσκευές και επειδή οι τεχνικές εξελίξεις έχουν καταστήσει διαθέσιμα νέα δεδομένα. Όλες οι επιχειρήσεις, ειδικότερα, που ανησυχούν για την εξέλιξη και τη μελλοντική ανάπτυξη του οργανισμού τους, θέλουν να υιοθετήσουν στρατηγικές που θα τους προσφέρουν το μεγαλύτερο δυνατό ανταγωνιστικό πλεονέκτημα στον κλάδο τους. Λόγω της πληθώρας των διαθέσιμων πληροφοριών, ο κλάδος των πληροφοριών επεκτείνεται και μεταβάλλεται με ταχείς ρυθμούς, γεγονός που καθιστά επιτακτική την ανάγκη για ανάλυση και επεξήγηση των δεδομένων που εξετάζονται πριν διεξαχθούν ως πληροφορίες. Όλοι οι οργανισμοί εξετάζουν τις πληροφορίες που τους βοηθούν στην επακόλουθη πορεία δράσης τους, προκειμένου να διευθύνουν καλύτερα τις εμπορικές τους δραστηριότητες. Ο τεράστιος όγκος πληροφοριών απαιτεί επίσης ένα συγκρίσιμο επίπεδο επεξεργασίας και ανάλυσης δεδομένων, έτσι ώστε τα ακατέργαστα δεδομένα να μπορούν να μετατραπούν σε πολύτιμα δεδομένα. Ο όρος "Μεγάλα δεδομένα και επιχειρηματική ανάλυση" αναφέρεται σε αυτό το στάδιο, το οποίο περιλαμβάνει τη χρήση λογισμικού επιχειρηματικής ευφυΐας (BI) ή την εφαρμογή ποσοτικής ανάλυσης και τεχνικών ανάλυσης μεγάλων δεδομένων (BDBA). Μια συλλογή μεθόδων και μέσων γνωστών ως "επιχειρηματική ευφυΐα" χρησιμοποιείται για να μετατρέψει τα μη επεξεργασμένα δεδομένα σε πληροφορίες που μπορούν να χρησιμοποιηθούν για την αξιολόγηση κάθε πτυχής των δραστηριοτήτων μιας εταιρείας (Rud, 2012). Μεγάλες ποσότητες δεδομένων μπορούν να διαχειριστούν από την τεχνολογία BI, οι οποίες μπορούν στη συνέχεια να χρησιμοποιηθούν για να βοηθήσουν τις επιχειρήσεις να δημιουργήσουν τρέχουσες ή νέες στρατηγικές ευκαιρίες. Με την εξεύρεση νέων ευκαιριών και την εφαρμογή στρατηγικών με βάση τις πληροφορίες που προσφέρονται από επιχειρήσεις με μακροχρόνια σταθερότητα και παρουσία σε ανταγωνιστικές αγορές, ο στόχος είναι να γίνουν τα δεδομένα ευκολότερα ερμηνεύσιμα (Ahmad & Tripathi, 2018).

Η επιχειρηματική ευφυΐα αποτελεί το κύριο και πρωταρχικό κίνητρο και έχει κάποια οφέλη που προστίθενται για τους πελάτες της, οι οποίοι συνήθως είναι επιχειρήσεις. Σε πρώτο πλάνο βρίσκεται η μείωση του άμεσου κόστους παραγωγής, η οποία αναφέρεται στη μείωση των δαπανών εργασίας και του χρόνου που απαιτείται για τις καθημερινές εργασίες. Επιπλέον, δίνει τη δυνατότητα στους χρήστες να βλέπουν γρήγορα τα αποτελέσματα της εταιρείας χωρίς να χρειάζεται να ξοδεύουν επιπλέον

χρόνο για την εξαγωγή τους, γεγονός που βοηθά τους χρήστες να λαμβάνουν αποφάσεις πιο γρήγορα και με μεγαλύτερη ακρίβεια. Επιπλέον, κάθε χρήστης μπορεί να έχει πρόσβαση στις πληροφορίες που χρειάζεται χρησιμοποιώντας αυτά τα προγράμματα, χωρίς να χρειάζεται πρόσθετη βοήθεια από επαγγελματίες της πληροφορικής. Προκειμένου να τροποποιήσουν τη μορφή, να προσδιορίσουν τον πυρήνα των δεδομένων τους και να προσθέσουν αξία στα περιουσιακά τους στοιχεία - που στην προκειμένη περίπτωση δεν είναι άλλα από τα δεδομένα τους- οι οργανισμοί χρησιμοποιούν εφαρμογές BI, εξετάζουν τα Big Data και διεξάγουν Business Analytics (Ahmad & Tripathi, 2018).

1.8 Τεχνολογική ανάλυση Big Data

Τα μεγάλα δεδομένα είναι αρχικά ανεπεξέργαστα και πρέπει να μετατραπούν σε νέες πληροφορίες για να υποστηρίξουν τους οργανισμούς στις σημερινές και μελλοντικές αναπτυξιακές αποφάσεις τους. Αυτό απαιτεί την κατάλληλη υποδομή επεξεργασίας και τεχνολογίας. Ταυτόχρονα, η αξία των δεδομένων μεγάλων δεδομένων αυξάνεται όταν χρησιμοποιούνται στη λήψη αποφάσεων. Ωστόσο, οι οργανισμοί χρειάζονται αποτελεσματικές μεθόδους για τη μετατροπή τεράστιων ποσοτήτων ταχέως μεταβαλλόμενων και ποικίλων δεδομένων σε πολύτιμες πληροφορίες, προκειμένου να καταστεί δυνατή η λήψη αποφάσεων βάσει στοιχείων. Ωστόσο, η όλη διαδικασία συλλογής δεδομένων από γεννήτριες μεγάλων δεδομένων ορίζει μια κρίσιμη απαίτηση για την ανάλυσή τους σε πέντε φάσεις. Οι δύο επιμέρους δραστηριότητες, η διαχείριση δεδομένων και η ανάλυση δεδομένων, συνδυάζονται σε αυτές τις πέντε φάσεις. Κατ' αρχάς, η διαχείριση των ακατέργαστων δεδομένων συνεπάγεται τη λήψη μέτρων και τη χρήση τεχνολογιών που καθιστούν δυνατή τη συλλογή, την αποθήκευση και, στη συνέχεια, την προετοιμασία και την ανάκτηση δεδομένων για ανάλυση. Το δεύτερο μέρος, η ανάλυση, εξετάζει στη συνέχεια τις μεθόδους και τις διαδικασίες που χρησιμοποιούνται για την εξέταση και την εξαγωγή πληροφοριών από τα μεγάλα δεδομένα. Συμπερασματικά, η "εξαγωγή γνώσης" από τα μεγάλα δεδομένα είναι μια διαδικασία που περιλαμβάνει την ανάλυση μεγάλων δεδομένων ως υποδιαδικασία. Στο επόμενο μέρος εξετάζονται τεχνικές που, δεδομένης της ποικιλίας των μεθοδολογιών, αποτελούν ένα σχετικό υποσύνολο των εργαλείων ανάλυσης που έχουμε στη διάθεσή μας (Labrinidis & Jagadish, 2012). Οι τεχνικές και οι μέθοδοι που χρησιμοποιούνται για την εξαγωγή πληροφοριών από δεδομένα κειμένου αναφέρονται ως ανάλυση δεδομένων κειμένου. Η κατηγορία των κειμενικών

δεδομένων περιλαμβάνει δεδομένα από μέσα κοινωνικής δικτύωσης, μηνύματα ηλεκτρονικού ταχυδρομείου, ιστότοπους, πολυάριθμα ερωτηματολόγια, ημερολόγια και πολλά άλλα μέσα. Τα δεδομένα αυτά αναλύονται με τη χρήση μηχανικής μάθησης, υπολογιστικής γλωσσολογίας και στατιστικής ανάλυσης. Αλλά τα αναλυτικά κείμενα επιτρέπουν στις εταιρείες να μετατρέψουν τεράστιες ποσότητες κειμένων που παράγονται από τον άνθρωπο σε διορατικές περιλήψεις που μπορούν να βοηθήσουν στη λήψη αποφάσεων βάσει στοιχείων (Chung, 2014).

Οι στρατηγικές συνοπτικής τεκμηρίωσης και σχολιασμού αναπαράγουν ταυτόχρονα αυτόματα μια συμπυκνωμένη περίληψη ενός ή περισσότερων αρχείων. Οι εφαρμογές στην ανάλυση περιλαμβάνουν ειδησεογραφικές και επιστημονικές δημοσιεύσεις, μηνύματα, διαφημίσεις και ιστότοπους. Οι δύο υποκατηγορίες προσέγγισης, η προσέγγιση εξόρυξης και η προσέγγιση περίληψης, σημειώνονται επίσης από τη μέθοδο περίληψης. Στην πρώτη, οι αρχικές μονάδες κειμένου χρησιμοποιούνται για τη δημιουργία περιλήψεων. Ένα τμήμα του αρχικού εγγράφου περιλαμβάνεται στην προκύπτουσα περίληψη. Η διαμόρφωση μιας περίληψης συνεπάγεται τον εντοπισμό των διακριτών στοιχείων ενός κειμένου και τη σύζευξή τους υπό την καθοδήγηση της προσέγγισης εξόρυξης. Η συνάφεια των μονάδων κειμένου καθορίζεται εξετάζοντας πού και πόσο συχνά εμφανίζονται στο κείμενο. Οι αφηρημένες τεχνικές περίληψης περιλαμβάνουν την αφαίρεση σημασιολογικών πληροφοριών από το κείμενο, ενώ οι τεχνικές εξομάλυνσης περίληψης δεν απαιτούν ταυτόχρονη κατανόηση του κειμένου. Ωστόσο, οι περιλήψεις περιλαμβάνουν τμήματα κειμένου που δεν βρίσκονται πάντα στο αρχικό κείμενο. Συμπερασματικά, η αφηρημένη περίληψη χρησιμοποιεί εξελιγμένους αλγόριθμους επεξεργασίας φυσικής γλώσσας για την εξέταση του αρχικού κειμένου κατά την παραγωγή μιας περίληψης. Τα αποσπασματικά συστήματα προτιμούν να αναπαράγουν συνεκτικές περιλήψεις για καλύτερη ανάλυση (Hahn, Mani, 2000).

Τα μη δομημένα δεδομένα ήχου υποβάλλονται σε ανάλυση και εξαγωγή πληροφοριών μέσω της ανάλυσης ήχου. Η φράση "speech analytics" χρησιμοποιείται όταν γίνεται επεξεργασία και ανάλυση του προφορικού ανθρώπινου λόγου. Δεδομένου ότι ο προφορικός ήχος είναι ο πρωταρχικός στόχος αυτών των τεχνικών, οι όροι sound analytics και speech analytics χρησιμοποιούνται συχνά συνώνυμα. Οι δύο κύριες βιομηχανίες στις οποίες χρησιμοποιούνται σήμερα οι τεχνολογίες ανάλυσης ήχου είναι η υγειονομική περίθαλψη και τα κέντρα επαφής για την υποστήριξη πελατών. Η ανάλυση ήχου χρησιμοποιείται από τα κέντρα επαφής για την αποτελεσματική

ανάλυση και επεξεργασία των καταγεγραμμένων κλήσεων. Αυτές οι μέθοδοι βοηθούν στη βελτίωση της εμπειρίας των πελατών, στην αξιολόγηση των επιδόσεων, στις πολιτικές προστασίας της ιδιωτικής ζωής και της ασφάλειας, στη γνώση της συμπεριφοράς των πελατών και στην αξιολόγηση των ζητημάτων προϊόντων ή υπηρεσιών μακροπρόθεσμα. Ωστόσο, τα προγράμματα ανάλυσης ήχου μπορούν να δημιουργηθούν για να εξετάσουν μια ζωντανή κλήση και να δώσουν στους πράκτορες ανατροφοδότηση σε πραγματικό χρόνο. Επιπλέον, τα αυτοματοποιημένα τηλεφωνικά κέντρα χρησιμοποιούν εργαλεία όπως η Interactive Voice Response (IVR) με πρωταρχικό στόχο τον εντοπισμό και τον χειρισμό των εκνευρισμένων και θυμωμένων καλούντων με τον κατάλληλο τρόπο. Επιπλέον, η ανάλυση ήχου χρησιμοποιείται στην υγειονομική περίθαλψη για να βοηθήσει στη διάγνωση και τη διαχείριση συγκεκριμένων ιατρικών καταστάσεων που επηρεάζουν άμεσα το στυλ επικοινωνίας ενός ασθενούς. Επιπλέον, οι αναλύσεις ήχου μπορούν να βοηθήσουν στην ανάλυση των κραυγών μωρών που αποκαλύπτουν λεπτομέρειες σχετικά με την ευημερία και τη συναισθηματική κατάσταση του μωρού. Ο τεράστιος όγκος δεδομένων που συγκεντρώνεται από τα συστήματα κλινικής τεκμηρίωσης με βάση την ομιλία είναι ένας ακόμη παράγοντας που ενθαρρύνει τη χρήση της ανάλυσης ήχου στον ιατρικό τομέα.

Ο όρος "οπτικοακουστική ανάλυση δεδομένων", επίσης γνωστός ως "ανάλυση περιεχομένου βίντεο", αναφέρεται σε μια ευρεία ποικιλία μεθόδων για την παρακολούθηση, την εξέταση και τη μετέπειτα εξαγωγή χρήσιμων δεδομένων - ή αποτελεσμάτων - από ροές βίντεο. Παρά το γεγονός ότι η ανάλυση βίντεο εξακολουθεί να αντιπαραβάλλεται συχνά με άλλες μορφές εξόρυξης δεδομένων (Panigrahi, Abraham, Das, 2010), έχει σημειωθεί σημαντική πρόοδος στην επεξεργασία τόσο του ζωντανού οπτικοακουστικού περιεχομένου όσο και των προηγουμένως καταγεγραμμένων βίντεο. Επιπλέον, οι δύο κύριοι παράγοντες που οδηγούν στην ανάπτυξη της ηλεκτρονικής ανάλυσης βίντεο είναι η αυξανόμενη χρήση συσκευών κλειστού κυκλώματος τηλεόρασης (CCTV) και η δημοτικότητα των ιστότοπων κοινής χρήσης βίντεο. Το μέγεθος αυτών των δεδομένων αποτελεί μείζον πρόβλημα και πρόκληση επεξεργασίας. Το μέγεθος των δεδομένων βίντεο, ωστόσο, αποτελεί σημαντικό εμπόδιο. Για να το θέσουμε σε προοπτική, σχεδόν 2000 σελίδες κειμένου χωράνε σε ένα δευτερόλεπτο βίντεο υψηλής ευκρίνειας (Manyika et al., 2011).

Η ανάλυση οργανωμένων και μη δομημένων δεδομένων από κανάλια κοινωνικών μέσων αναφέρεται ως διερεύνηση και ανάλυση με χρήση κοινωνικών

μέσων. Η φράση "μέσα κοινωνικής δικτύωσης" ή απλώς "μέσα κοινωνικής δικτύωσης" αναφέρεται σε ένα ευρύ φάσμα διαδικτυακών πλατφορμών που επιτρέπουν στους χρήστες να παράγουν και να ανταλλάσσουν υλικό σε τακτική βάση και ανά πάσα στιγμή. Τα κοινωνικά δίκτυα (Facebook, LinkedIn, Instagram), τα ιστολόγια (WordPress), τα μικροσκοπία (Twitter, Tumblr), οι κοινωνικές ειδήσεις (Digg, Reddit), ο διαμοιρασμός πολυμέσων (YouTube), τα wikis (Wikipedia και Wikihow) και οι ιστότοποι ερωτήσεων και απαντήσεων είναι μερικά παραδείγματα από τις διάφορες κατηγορίες στις οποίες υπάγονται (Yahoo Answers, Ask.com, Yelp, TripAdvisor). Το κύριο χαρακτηριστικό των σύγχρονων κοινωνικών αναλύσεων είναι επίσης το είδος των δεδομένων που προσφέρουν. Πολυάριθμοι ακαδημαϊκοί τομείς, όπως η κοινωνιολογία, η ψυχολογία, η ανθρωπολογία, η επιστήμη των υπολογιστών, τα μαθηματικά, η φυσική, ακόμη και τα οικονομικά, έχουν διεξάγει έρευνες σχετικά με τα social media analytics. Ωστόσο, το μάρκετινγκ είναι ο πρωταρχικός και σημαντικότερος πυλώνας που στηρίζει την ανάπτυξη των μέσων κοινωνικής δικτύωσης και επεκτείνεται γρήγορα τα τελευταία χρόνια. Ως εκ τούτου, είναι απαραίτητο να υπάρχουν και να χρησιμοποιούνται τα κατάλληλα εργαλεία για την άμεση επεξεργασία και αποσαφήνιση των δεδομένων πριν αυτά μετατραπούν σε χρήσιμες πληροφορίες. Συμπερασματικά, χρησιμοποιούνται στρατηγικές και διαδικασίες που μπορούν να παρέχουν γρήγορα το απαιτούμενο αποτέλεσμα, ανάλογα με το είδος των δεδομένων που πρέπει να αναλυθούν και να μετατραπούν σε χρήσιμες πληροφορίες. Λόγω του διαρκώς μεταβαλλόμενου και αναπτυσσόμενου περιβάλλοντος, είναι απαραίτητο να αναλύονται γρήγορα και αποτελεσματικά όλα τα είδη δεδομένων, διότι η αποτυχία να το πράξει θα μπορούσε να καταστήσει τις πληροφορίες άχρηστες. Σε γενικές γραμμές, οι οργανισμοί έχουν πλέον πρόσβαση σε μια παγκόσμια εικόνα πληροφοριών χάρη στην εξαγωγή πληροφοριών από όλα τα μέσα, γεγονός που επιτρέπει τον επανασχεδιασμό ή την προσαρμογή των στρατηγικών κερδοφορίας (Manyika et al., 2011).

1.9 Αποθήκευση Big Data

Όταν υπήρχαν αρχικά δεδομένα, υπήρχε ανάγκη να αποθηκευτούν. Η απαίτηση για την αποτελεσματική αποθήκευση και διαχείριση του όγκου των εισερχόμενων

δεδομένων αυξάνεται με την πάροδο του χρόνου, καθώς οι ανάγκες της εταιρείας και των άλλων επιχειρήσεων διαφοροποιούνται και εξελίσσονται.

Πριν από τριάντα περίπου χρόνια, η έννοια της αποθήκευσης δεδομένων εμφανίστηκε για πρώτη φορά ως ένα κρίσιμο εργαλείο για την ταχεία επίλυση επειγόντων ζητημάτων πελατών μιας επιχείρησης. Εξαιτίας αυτού, απαιτήθηκε ένα πιο σύγχρονο, εξορθολογισμένο, αρχιτεκτονικά άρτιο μοντέλο που θα αντικαθιστούσε τις ήδη αποτελεσματικές μεθόδους λήψης αποφάσεων και θα βοηθούσε πραγματικά τις επιχειρήσεις να λαμβάνουν καλύτερες και ταχύτερες αποφάσεις. Ωστόσο, σύμφωνα με τους Marz & Warren (2015), ένα σύστημα που χρησιμοποιείται σε εφαρμογές Big Data θα πρέπει να περιορίζεται από συγκεκριμένα χαρακτηριστικά :

- Να είναι ανθεκτικό και ανεκτικό σε σφάλματα (Robustness and fault tolerance)
- Να διαβάζει και να ανανεώνει γρήγορα τις εγγραφές (Low latency reads and updates)
- Να είναι επεκτάσιμο (Scalability)
- Να είναι γενικευμένο (Generalization)
- Να μπορεί να δεχθεί νέες λειτουργίες (Extensibility)
- Να μπορεί να απαντάει σε ερωτήματα (Ad hoc queries)
- Να απαιτεί μικρή συντήρηση (Minimal maintenance)
- Να μπορεί να δίνει λύση σε παρουσιαζόμενα σφάλματα (Debuggability)

Η διατήρηση των δεδομένων μετά τη συλλογή και τη διατήρησή τους με τρόπο που να καθιστά απλή και αποτελεσματική την ανάκτηση πληροφοριών σε μεταγενέστερο χρόνο αναφέρεται ως "αποθήκευση δεδομένων". Σε αντίθεση με τα συμβατικά λειτουργικά συστήματα και τις εφαρμογές βάσεων δεδομένων, οι επιχειρήσεις, τόσο οι κερδοσκοπικές όσο και οι μη κερδοσκοπικές, είναι οι κύριοι χρήστες της αποθήκευσης δεδομένων (Ahmad & Tripathi, 2018).

Τα δεδομένα λαμβάνονται από διάφορες πηγές σχεσιακών βάσεων δεδομένων και στη συνέχεια διασφαλίζεται η εγκυρότητα και η ακρίβειά τους, με τα δομημένα δεδομένα να αποτελούν το κύριο αντικείμενο της αποθήκευσης και να υποστηρίζονται από αυτήν. Η συμβατική μέθοδος αποθήκευσης δεδομένων θα αντέξει και θα ευημερήσει όσο η ανάλυση δεδομένων για τη λήψη αποφάσεων αποτελεί σημαντικό θέμα. Ωστόσο, καθώς ο όγκος των δεδομένων για τις ενεργές επιχειρήσεις αυξάνεται και υπάρχει ανάγκη εκσυγχρονισμού, οι μορφές αποθήκευσης δεδομένων

προσαρμόζονται επί του παρόντος στα Big Data. Ωστόσο, η μετάβαση από ένα σύστημα αποθήκευσης σε ένα άλλο είναι πάντα σταδιακή (Chen et al., 2012).

Το θεμέλιο για την επεξεργασία δεδομένων και τη μετέπειτα αποθήκευση τεράστιων όγκων πληροφοριών είναι η αρχιτεκτονική αποθήκευσης δεδομένων. Περιλαμβάνει τις ακόλουθες φάσεις :

- Συστήματα Προέλευσης: τα οποία, αποτελούν τις διάφορες πηγές μέσω των οποίων λαμβάνονται τα δεδομένα.
- Μετακίνηση Δεδομένων: που αφορά τις κύριες τεχνικές προγραμματισμού για τη μετακίνηση των δεδομένων εντός του οικοσυστήματος.
- Βάσεις Δεδομένων: που δημιουργούνται και αναπτύσσονται εντός της αποθήκευσης των δεδομένων.
- Στάδια Προγραμματισμού: για όλα τα δεδομένα που πρέπει να αποσταλούν στην αποθήκη δεδομένων, αυτές οι βάσεις δεδομένων χρησιμεύουν ως ο κύριος ρόλος επεξεργασίας και προεπεξεργασίας.

Η λειτουργική αποθήκευση δεδομένων, η οποία χρησιμεύει κυρίως ως εργαλείο για την καθημερινή τους επεξεργασία, απεικονίζει τη δομή της βάσης δεδομένων. Οι Datamarts είναι εξειδικευμένες βάσεις δεδομένων που δημιουργούνται και αναπτύσσονται για χρήση από συγκεκριμένες επιχειρηματικές μονάδες. Οι εταιρείες δημιουργούν μια μεγάλη ποικιλία από αυτές για να τις ενσωματώσουν στην αποθήκευση δεδομένων (Chen et al., 2012).

Όπως έχει ήδη διαπιστωθεί, η αποθήκευση δεδομένων συχνά συνδυάζει πίνακες σχεσιακών βάσεων δεδομένων, αρχεία και ένα ευρύ φάσμα διαφορετικών πηγών. Προκειμένου να καταστεί δυνατή και να διευκολυνθεί η επεξεργασία ερωτημάτων με μεγαλύτερη ακρίβεια και συνέπεια, δημιουργείται μια καλά σχεδιασμένη τεχνική αποθήκευσης μεγάλων δεδομένων με ρητό στόχο την τοποθέτηση των δεδομένων σε μια τυποποιημένη μορφή. Τα μετατρεπόμενα εξαγόμενα αρχεία πρέπει να ακολουθούν τις κατευθυντήριες γραμμές και τις πρακτικές για τις οποίες προορίζονται για τη φύλαξη των δεδομένων. Ωστόσο, προκειμένου οι μεγάλες πηγές δεδομένων να συνεργαστούν με ασφάλεια και να παραδώσουν πολύτιμα αποτελέσματα, τα δεδομένα πρέπει στη συνέχεια να εξαχθούν από τις πηγές. Οι πληροφορίες μεγάλων δεδομένων αποθηκεύονται εντελώς διαφορετικά από τα δεδομένα που διατηρούνται σε ένα πιο συμβατικό παράδειγμα αποθήκευσης δεδομένων. Στη συμβατική αποθήκευση δεδομένων, τα δεδομένα δεν αλλάζουν ποτέ μορφή μετά τη διαδικασία κωδικοποίησης. Μια τυπική αποθήκευση δεδομένων παράγει δεδομένα ως απάντηση στην απαίτηση

αξιολόγησης ενός συγκεκριμένου προβλήματος, το οποίο απαιτεί παρακολούθηση, ανάλυση και, κατά συνέπεια, λύση (Chen et al., 2012).

Στα μεγάλα δεδομένα, η αποθήκευση πληροφοριών μπορεί κατά συνέπεια να ποικίλλει σε μεγάλο βαθμό. Η κατανεμημένη φύση των μεγάλων δεδομένων συχνά αναγκάζει τους χρήστες να αποθηκεύουν πρώτα τα δεδομένα πριν κάνουν εξαγωγή και μετασχηματισμό δεδομένων, πάντα σύμφωνα με την άμεση, παραγόμενη ζήτηση. Λόγω των προαναφερθέντων παραγόντων, είναι σαφές ότι η αποθήκευση μεγάλων δεδομένων πρέπει να αντιμετωπίζεται διαφορετικά από άλλους τύπους δεδομένων. Ο προσδιορισμός των ομοιοτήτων και των διαφορών μεταξύ της διαχείρισης των δεδομένων στην παραδοσιακή αποθήκευση δεδομένων και της αποθήκευσης δεδομένων όταν ενσωματώνονται τα μεγάλα δεδομένα είναι πλέον πολύτιμος αλλά και σημαντικός. Όσον αφορά τους παραλληλισμούς μεταξύ των δύο τεχνικών διαχείρισης δεδομένων, αυτοί περιλαμβάνουν (Chen et al., 2012):

- Απαιτήσεις για κοινούς ορισμούς δεδομένων.
- Απαιτήσεις για την εξαγωγή και μετατροπή βασικών πηγών δεδομένων.

Συγχρόνως, ως προς τις διαφορές ανάμεσα στη παραδοσιακή αποθήκευση των δεδομένων και αυτή των Μεγάλων Δεδομένων, περιλαμβάνονται :

- Η κατανεμημένη υπολογιστική προσέγγιση των μεγάλων δεδομένων, η οποία απαιτείται για την αποθήκευση.

- Ενώ η παραδοσιακή αποθήκευση μεγάλων δεδομένων χρησιμοποιείται για την επέκταση των βάσεων δεδομένων και τον εκσυγχρονισμό των συστημάτων αποθήκευσης, η ανάλυση μεγάλων δεδομένων είναι το κύριο μέλημα των νέων, εξελισσόμενων τεχνολογιών (Chen et al., 2012).

Οι προσεγγίσεις για τη δημιουργία και τη διαχείριση δεδομένων αλλάζουν δραστικά ως αποτέλεσμα της εμφάνισης και της επέκτασης των μεγάλων δεδομένων. Λόγω του κόστους της παραδοσιακής αποθήκευσης δεδομένων, η οποία συχνά γίνεται σε ένα μόνο σύστημα εντός της βάσης δεδομένων, οι επιχειρήσεις βελτιστοποίησαν αυτές τις τεχνικές αποθήκευσης και μείωσαν δραστικά τον όγκο των δεδομένων που διαχειρίζονται σε καθημερινή βάση. Οι παραδοσιακές τεχνικές αποθήκευσης, ωστόσο, δεν λειτουργούν όταν οι επιχειρήσεις επιδιώκουν να αξιοποιήσουν τον τεράστιο όγκο

πληροφοριών που παράγονται από τις πηγές μεγάλων δεδομένων. Προκειμένου να διευκολυνθεί η μετάβαση στη νέα συνιστώσα διαχείρισης πληροφοριών της επιχείρησης, η αποθήκευση των δεδομένων τους έχει γίνει, επομένως, ένα μέσο για τη δημιουργία ενός βέλτιστου περιβάλλοντος. Επιπλέον, ο όγκος των πληροφοριών που εισέρχονται σε έναν οργανισμό απαιτεί μια νέα στρατηγική για την αποθήκευση δεδομένων λόγω της διαχείρισης αυτών των πρόσθετων πληροφοριών (Jina et al., 2015).

Το μοντέλο αποθήκευσης υλικού είναι συνήθως η καλύτερη επιλογή όταν οι επιχειρήσεις απαιτείται να συνδυάσουν τη δομή της αποθήκευσης δεδομένων με τα Big Data. Πρόκειται για ένα ολοκληρωμένο σύστημα που συνδυάζει υλικό που έχει σχεδιαστεί για τη διαχείριση και την αποθήκευση δεδομένων. Πρακτικά, το σύστημα αυτό θα μπορούσε να είναι ένα δωμάτιο γεμάτο με διακομιστές ή κάποιο άλλο εργαλείο για την αποθήκευση και την επεξεργασία μεγάλων δεδομένων (Jina et al., 2015).

1.10 Cloud και Big Data

Τα μεγάλα δεδομένα έχουν πολυάριθμα και ποικίλα πλεονεκτήματα για τους οργανισμούς, όπως έχει ήδη περιγραφεί στα παραπάνω κεφάλαια. Ωστόσο, η τεράστια ποσότητα επεξεργαστικής ισχύος και υπηρεσιών λογισμικού που απαιτούνται για τη λειτουργία των Μεγάλων Δεδομένων μπορεί να επηρεάσει σημαντικά αρνητικά ακόμη και την οικονομική και πνευματική υπόσταση των μεγαλύτερων επιχειρήσεων. Σε αυτή την περίπτωση, το Cloud μπήκε στο προσκήνιο για να αντιμετωπίσει το ζήτημα που θα προέκυπτε από την κάλυψη των απαιτήσεων των Big Data. Το Cloud είναι μια πλατφόρμα που υποστηρίζει και προωθεί τη διαχείριση των Big Data και μπορεί να εφαρμοστεί σε ένα ευρύ φάσμα εργασιών που σχετίζονται με την αποθήκευση. Μπορεί να προσφέρει σχεδόν απεριόριστους υπολογιστικούς πόρους και υπηρεσίες που επιτρέπουν στις επιχειρήσεις να χρησιμοποιούν τα μεγάλα δεδομένα για να αναλαμβάνουν δράσεις. Το Cloud θεωρείται ως μία από τις μεγαλύτερες πλατφόρμες για την αποθήκευση δεδομένων, με μεγάλη ευκολία πρόσβασης σε πραγματικό χρόνο και χωρίς απαίτηση για φυσικό χώρο ή εξοπλισμό αποθήκευσης, με βάση τις ανακαλύψεις στις δυσκολίες αποθήκευσης και μεταφοράς δεδομένων. Αυτοί οι παράγοντες συμβάλλουν στην αυξανόμενη δημοτικότητα της αποθήκευσης δεδομένων

σε περιβάλλοντα νέφους, τα οποία προσφέρουν επίσης εξαιρετική ασφάλεια και μειωμένο κόστος (Jina et al., 2015).

Επιπλέον, η χρήση του νέφους είναι μια απλή διαδικασία, δεδομένου ότι ο χρήστης μπορεί να συγκροτήσει γρήγορα την απαιτούμενη υποδομή υπολογιστικών δεδομένων και αποθηκευτικών πόρων που βασίζονται στο νέφος. Μπορούν επίσης να ανεβάζουν σύνολα δεδομένων, να συνδέουν υπηρεσίες Cloud και να εκτελούν αναλύσεις. Ταυτόχρονα, οι χρήστες έχουν πρόσβαση σε σχεδόν απεριόριστους πόρους, τους οποίους μπορούν να δεσμεύσουν, να χρησιμοποιήσουν για απεριόριστο χρονικό διάστημα και στη συνέχεια να πετάξουν στα σκουπίδια το περιβάλλον, πληρώνοντας μόνο για τις υπηρεσίες και τους πόρους που πραγματικά καταναλώθηκαν (Jina et al., 2015).

Ανεξάρτητα από το μέγεθός τους, όλοι οι οργανισμοί μπορούν να επωφεληθούν σημαντικά από το Cloud. Ορισμένα από τα πιο αισθητά και άμεσα πλεονεκτήματα της χρήσης μεγάλων δεδομένων στο νέφος είναι τα εξής :

- **Επεκτασιμότητα:** Για να αγοράσει και να υλοποιήσει τον τεράστιο όγκο εξοπλισμού που απαιτείται για να επιτρέψει μια υποδομή μεγάλων δεδομένων, ένα συμβατικό επιχειρηματικό κέντρο δεδομένων περιορίζεται από τον χώρο, την ισχύ και τους οικονομικούς πόρους. Από την άλλη πλευρά, ένα νέφος μπορεί να ελέγχει εκατομμύρια διακομιστές. Οι χρήστες μπορούν να συγκροτήσουν την υποδομή για ένα έργο μεγάλων δεδομένων επειδή τόσο η υποδομή όσο και οι υπηρεσίες λογισμικού είναι ήδη διαθέσιμες (σχεδόν ανεξαρτήτως μεγέθους).
- **Ευελιξία:** Διαφορετικοί τύποι δεδομένων μπορεί να απαιτούν διαφορετικό αριθμό διακομιστών. Μέσω του νέφους, οι χρήστες μπορούν να χρησιμοποιούν όσους πόρους απαιτούνται για την ολοκλήρωση μιας εργασίας και στη συνέχεια να απελευθερώνουν τους πόρους αυτούς μετά την ολοκλήρωση της εργασίας.
- **Κόστος:** Ένα επιχειρησιακό κέντρο δεδομένων αποτελεί σημαντική κεφαλαιουχική δαπάνη για μια εταιρεία. Το Cloud, αντίθετα, ενσωματώνει όλες τις δαπάνες σε ένα ευέλικτο μοντέλο συντήρησης με πρόσβαση σε πόρους κατά παραγγελία και υπηρεσίες πληρωμής ανά χρήση (Jina et al., 2015).
- **Προσβασιμότητα:** Πολλά Clouds προσφέρουν παγκόσμιο αποτύπωμα, επιτρέποντας την ανάπτυξη πόρων και υπηρεσιών στην πλειονότητα των σημαντικότερων περιοχών του κόσμου (Ahmad & Tripathi, 2018).

- **Ελαστικότητα:** Τα Clouds αναπαράγουν δεδομένα ως δεδομένο για τη διατήρηση υψηλής διαθεσιμότητας στους αποθηκευτικούς πόρους, ενώ ακόμη πιο ισχυρές επιλογές αποθήκευσης είναι προσβάσιμες στα Clouds.

Με την πάροδο των ετών, τα νέφη και πολλοί άλλοι εξωτερικοί πάροχοι μεγάλων δεδομένων έχουν αποδείξει την αξία τους. Ωστόσο, παρά τα πλεονεκτήματά τους, τα οποία έχουν καλυφθεί εκτενώς στην παρούσα διατριβή, οι επιχειρήσεις πρέπει επίσης να λαμβάνουν υπόψη τα μειονεκτήματά τους, μερικά από τα οποία είναι τα εξής:

- **Εξάρτηση από το Διαδίκτυο:** Για τη χρήση του νέφους απαιτείται πλήρης δικτυακή πρόσβαση από το τοπικό δίκτυο μέσω του Διαδικτύου στο δίκτυο του παρόχου του νέφους. Οποιαδήποτε καθυστέρηση κατά μήκος αυτής της διαδρομής δικτύου μπορεί να προκαλέσει αυξημένη καθυστέρηση ή, στη χειρότερη περίπτωση, πλήρη απώλεια της προσβασιμότητας στο νέφος. Οποιαδήποτε χρήση Μεγάλων Δεδομένων εντός των νεφών θα πρέπει να λαμβάνει δεόντως υπόψη τις επιπτώσεις των διακοπών (Jina et al., 2015).
- **Κόστος αποθήκευσης:** μπορεί να αυξηθεί με την πάροδο του χρόνου όταν χρησιμοποιείται το νέφος για την αποθήκευση δεδομένων. Η αποθήκευση δεδομένων, η μετανάστευση δεδομένων και η διατήρηση δεδομένων είναι τα τρία βασικά προβλήματα. Μεγάλες ποσότητες δεδομένων πρέπει να φορτώνονται αργά στο σύννεφο και υπάρχει μηνιαία χρέωση για αυτές τις περιπτώσεις αποθήκευσης. Ωστόσο, μπορεί να υπάρξουν επιπλέον χρεώσεις εάν τα δεδομένα μεταφερθούν. Συνεπώς, τα μεγάλα σύνολα δεδομένων είναι συχνά πολύ ευαίσθητα στον χρόνο. Αυτό ουσιαστικά υποδηλώνει ότι ορισμένα δεδομένα μπορεί να είναι άχρηστα για διάφορες αναλύσεις, ακόμη και ώρες αργότερα. Οι επιχειρήσεις θα πρέπει να αναπτύσσουν διεξοδικές πολιτικές διατήρησης και διαγραφής για τη διαχείριση τόσο της διατήρησης όσο και της διαγραφής δεδομένων, καθώς σχετίζονται με τα έξοδα αποθήκευσης στο cloud, επειδή η διατήρηση πλεοναζόντων δεδομένων κοστίζει χρήματα.
- **Ασφάλεια:** Οι δομές μεγάλων δεδομένων ενδέχεται να περιέχουν εμπιστευτικές ή προσωπικά αναγνωρίσιμες πληροφορίες, οι οποίες μπορεί να διέπονται από νόμους ή άλλους περιορισμούς που έχουν θεσπιστεί από τις επιχειρήσεις ή την κυβέρνηση. Οι χρήστες του νέφους θα πρέπει να έχουν λάβει τις απαιτούμενες προφυλάξεις για την προστασία της ασφάλειάς τους (Jina et al., 2015).

- Έλλειψη τυποποίησης: Αυτό μπορεί να οδηγήσει σε υποδεέστερες επιδόσεις ή ενδεχομένως να εκθέσει την εταιρεία σε προβλήματα ασφάλειας. Θα πρέπει να υπάρχει μια καθορισμένη τεχνική για την κατασκευή, την υλοποίηση και την ανάπτυξη δεδομένων σε ένα νέφος. Οι επιχειρησιακοί χρήστες θα πρέπει να καταγράφουν την αρχιτεκτονική των μεγάλων δεδομένων και τυχόν εφαρμοστέες πολιτικές και διαδικασίες (Ahmad & Tripathi, 2018).

2. Τα Big Data στον τομέα Υγείας

2.1 Χαρακτηριστικά & Πλεονεκτήματα

Δεδομένου ότι ο όγκος των δεδομένων υγείας αναμένεται να αυξηθεί απότομα στο μέλλον, είναι προφανές ότι η ανάγκη για καινοτόμες και έξυπνες τεχνικές διαχείρισης δεδομένων είναι πιο επιτακτική από ποτέ. Κανείς δεν θα μπορούσε να αμφισβητήσει τη σημασία της μέγιστης δυνατής αξιοποίησης των τεράστιων ποσοτήτων δεδομένων και πληροφοριών που παράγονται και καταναλώνονται κάθε δευτερόλεπτο στον τομέα της υγείας, προκειμένου να εξαχθούν εύστοχα συμπεράσματα. Το γεγονός ότι τα μοντέλα αποζημίωσης των παρόχων υγειονομικής περίθαλψης εξελίσσονται είναι μια άλλη εξέλιξη. Η αμοιβή για την απόδοση και η ουσιαστική χρήση των υπηρεσιών υγειονομικής περίθαλψης αναδεικνύονται ως νέα κρίσιμα στοιχεία στο σημερινό περιβάλλον της υγειονομικής περίθαλψης. Οι επιχειρήσεις υγειονομικής περίθαλψης πρέπει να αποκτήσουν τα διαθέσιμα εργαλεία, τις υποδομές και τις μεθοδολογίες για την αποτελεσματική χρήση των μεγάλων δεδομένων, ακόμη και όταν το κέρδος δεν είναι και δεν θα έπρεπε να είναι το κύριο κίνητρό τους. Διαφορετικά, οι επιχειρήσεις κινδυνεύουν να χάσουν τεράστια ποσά σε χαμένες πωλήσεις και έσοδα (Παρασύρη, 2014).

Επιπλέον, η εστίαση έχει μετατοπιστεί από την πρόληψη των ασθενειών στη θεραπεία, ως αποτέλεσμα της εκρηκτικής ανόδου της βιομηχανίας δεδομένων υγείας. Οι επιστήμονες έχουν επικεντρωθεί στην ενίσχυση της αξιοπιστίας και της αποτελεσματικότητας των συστημάτων υγειονομικής περίθαλψης για τη μείωση του κόστους θεραπείας και την παροχή καλύτερων φαρμάκων στους ασθενείς. Τα δεδομένα κλινικών δοκιμών, τα ευρήματα απεικονιστικών εξετάσεων και τα ιατρικά αρχεία ασθενών αποτελούν άφθονες πηγές μεγάλων δεδομένων στα νοσοκομεία και τα εθνικά συστήματα υγείας. Ο μεγάλος όγκος των επί του παρόντος διαθέσιμων (αλλά

ανεπαρκώς αναλυμένων) ιατρικών δεδομένων κάθε είδους μπορεί να υποβληθεί σε υφιστάμενες αναλυτικές μεθοδολογίες προκειμένου να αποκτηθεί μεγαλύτερη κατανόηση των αποτελεσμάτων, τα οποία στη συνέχεια θα μεταδοθούν και θα εφαρμοστούν στο σημείο περίθαλψης. Η γνώση τόσο του πληθυσμού όσο και του ατόμου θα πρέπει να υποστηρίζει και να παρέχει χρήσιμες πληροφορίες στον θεράποντα ιατρό και τον ασθενή κατά τη διαδικασία λήψης αποφάσεων και να διευκολύνει την επιλογή της καταλληλότερης πορείας δράσης για τον συγκεκριμένο ασθενή (Παρασύρη, 2014).

Τα τρία V από τον ορισμό των μεγάλων δεδομένων, δηλαδή η ταχύτητα, η ποικιλία και ο όγκος, αποτελούν εγγενές στοιχείο των δεδομένων που παράγουν, και η υγειονομική περίθαλψη αποτελεί μια εξαιρετική εικόνα αυτού. Πολυάριθμα συστήματα υγειονομικής περίθαλψης, ασφαλιστικές εταιρείες, ερευνητές, κυβερνητικοί οργανισμοί κ.λπ. έχουν πρόσβαση σε αυτά τα δεδομένα. Καθεμία από αυτές τις πηγές δεδομένων είναι επίσης υπερφορτωμένη και, ως εκ τούτου, δεν είναι σε θέση να προσφέρει μια διαφανή πλατφόρμα σε παγκόσμια κλίμακα. Επιπλέον, δεδομένου ότι η ακρίβεια των δεδομένων υγείας είναι ζωτικής σημασίας για τη χρήσιμη εφαρμογή τους στην πρόοδο της μεταγραφικής έρευνας, η έννοια της εγκυρότητας αποτελεί κρίσιμη προσθήκη στα προαναφερθέντα στοιχεία (Παρασύρη, 2014).

Τα μεγάλα δεδομένα στον κλάδο της υγειονομικής περίθαλψης αποτελούν σημαντική πρόκληση λόγω του όγκου τους, της ποικιλομορφίας των ειδών δεδομένων και του επείγοντος χαρακτήρα με τον οποίο πρέπει να διαχειριστούν τα δεδομένα αυτά. Ενώ η καινοτόμος χρήση του Google Glass από τους γιατρούς θα έρθει επίσης να προσθέσει στις κοινωνικές και συμπεριφορικές συνιστώσες των μεγάλων δεδομένων, η αυξανόμενη χρήση της τηλεϊατρικής θα δοκιμάσει περαιτέρω την ικανότητα αποθήκευσης δεδομένων ασθενών. Τα ηλεκτρονικά ιατρικά αρχεία (EMR) είναι γνωστό ότι βοηθούν σημαντικά στην ομαλή ροή της υγειονομικής περίθαλψης και περιέχουν πληθώρα πληροφοριών, συμπεριλαμβανομένων των δημογραφικών στοιχείων των ασθενών, των κλινικών δεδομένων και των γονιδιωματικών δεδομένων. Οι μελλοντικές εφαρμογές δεδομένων σε πραγματικό χρόνο θα μπορούσαν να μειώσουν τη νοσηρότητα και τη θνησιμότητα των ασθενών και ακόμη και να σταματήσουν την ανάπτυξη νοσοκομειακών ασθενειών με τον έγκαιρο εντοπισμό λοιμώξεων, την έγκαιρη ανίχνευση λοιμώξεων και τη χορήγηση των κατάλληλων θεραπειών. Τα δεδομένα ροής σε πραγματικό χρόνο χρησιμοποιούνται ήδη για να

παρακολουθούν τα βρέφη στη ΜΕΘ και να προειδοποιούν τους θεράποντες ιατρούς για τυχόν απειλητικές για τη ζωή μολύνσεις στο νοσηλευόμενο βρέφος. Η υγειονομική περίθαλψη θα υποστεί πιθανότατα μια επανάσταση ως αποτέλεσμα της δυνατότητας ανάλυσης τεράστιων ποσοτήτων δεδομένων σε όλες τις ειδικότητες σε πραγματικό χρόνο (Παρασύρη, 2014).

Στον κλάδο της υγειονομικής περίθαλψης χρησιμοποιούνται δομημένα, ημιδομημένα και αδόμητα δεδομένα. Ένα λογισμικό υπολογιστή μπορεί γρήγορα να αποθηκεύει, να αναζητά, να ανακτά, να αναλύει και να χειρίζεται δομημένα δεδομένα. Όπως αναφέρθηκε προηγουμένως, τα ημιδομημένα δεδομένα περιέχουν πληροφορίες αποθηκευμένες σε μορφή διεπαφής XML3 (Extensible Markup Language), ενώ τα δομημένα δεδομένα περιλαμβάνουν πληροφορίες από κλινικούς αισθητήρες, εργαστήρια και σχεσιακές βάσεις δεδομένων. Οι ηλεκτρονικοί ιατρικοί φάκελοι, οι χειρόγραφες σημειώσεις, τα γραφήματα, οι ακτινογραφίες, οι περιλήψεις των περιλήψεων των εξιτηρίων των ασθενών, τα φυσιολογικά δεδομένα και τα δεδομένα ελεύθερου κειμένου χωρίς συγκεκριμένο σχεδιασμό αποτελούν παραδείγματα μη δομημένων δεδομένων. Το 90% του συνόλου των μη δομημένων δεδομένων της υγειονομικής περίθαλψης αποτελεί μεγάλο όγκο δεδομένων (Παρασύρη, 2014).

Ο τομέας της υγειονομικής περίθαλψης κατακλύζεται ήδη από νέες ροές δεδομένων, τόσο δομημένων όσο και αδόμητων, από πηγές όπως ο εξοπλισμός γυμναστικής, η γονιδιωματική και γενετική έρευνα, οι έρευνες στο διαδίκτυο και άλλες πηγές. Ο όγκος των δεδομένων που μπορούν να διαβαστούν, να αποθηκευτούν και να δομηθούν με τρόπο ώστε τα υπολογιστικά συστήματα να μπορούν να τα ερμηνεύσουν και να εξάγουν πληροφορίες από αυτά είναι, ωστόσο, περιορισμένος. Ειδικότερα, οι εφαρμογές που σχετίζονται με την υγεία θα πρέπει να διαθέτουν πιο αποτελεσματικές μεθόδους συγχώνευσης και αλλαγής ετερογενών δεδομένων, όπως η μετάβαση από δομημένα σε μη δομημένα δεδομένα (Wang et al., 2018).

Δεδομένου ότι προέρχονται από διάφορες πηγές και έχουν διάφορους τύπους και μορφές, τα δεδομένα υγειονομικής περίθαλψης είναι χαοτικά και κατακερματισμένα. Τα κλινικά δεδομένα μπορούν να αντληθούν από ηλεκτρονικούς φακέλους υγείας, συστήματα καταχώρησης ιατρικών εντολών (CPOE4), συστήματα υποστήριξης κλινικών αποφάσεων (CDSS5), αρχεία χορήγησης φαρμάκων, εργαστηριακά και φαρμακευτικά αρχεία μελετών κοόρτης, κυβερνητικές έρευνες και κλινικές δοκιμές. Άλλες πηγές περιλαμβάνουν αρχεία χορήγησης φαρμάκων, αρχεία χορήγησης φαρμάκων, αρχεία χορήγησης φαρμάκων και αρχεία χορήγησης φαρμάκων.

Η διάγνωση του ασθενούς σύμφωνα με τη Διεθνή Στατιστική Ταξινόμηση των Ασθενειών και των Συναφών Προβλημάτων Υγείας (ICD6), η κατάσταση της υγείας του ασθενούς κατά την έξοδο, καθώς και πληροφορίες για τις οικονομικές απαιτήσεις, συμπεριλαμβανομένων, μεταξύ άλλων, των χρεώσεων της επίσκεψης και του ποιος πρέπει να πληρώσει την αμοιβή για την υπηρεσία που έλαβε, περιέχονται στα διοικητικά δεδομένα. Από την άλλη πλευρά, τα κλινικά δεδομένα περιλαμβάνουν δημογραφικές πληροφορίες των ασθενών και πληροφορίες από τις επισκέψεις που πραγματοποιούν στον γιατρό τους (Wang et al., 2018).

Οι πηγές μεγάλων δεδομένων υγειονομικής περίθαλψης μπορούν να κατηγοριοποιηθούν χονδρικά σε εσωτερικές (όπως ηλεκτρονικά αρχεία υγείας, συστήματα υποστήριξης κλινικών αποφάσεων, CPOE κ.λπ.) και εξωτερικές (όπως πηγές δεδομένων τρίτων) (κυβερνητικές πηγές, εργαστήρια, φαρμακεία, ασφαλιστικές εταιρείες κ.λπ.). Οι πληροφορίες αυτές αποθηκεύονται σε διάφορες μορφές (επίπεδα αρχεία, σχεσιακούς πίνακες, ASCII/κείμενο, κ.λπ.), φυλάσσονται σε διάφορες τοποθεσίες (τόσο φυσικά όσο και με την έννοια ότι φιλοξενούνται στις εγκαταστάσεις διαφόρων παροχών υγειονομικής περίθαλψης) και φυλάσσονται σε διάφορες εφαρμογές (εφαρμογές επεξεργασίας συναλλαγών, βάσεις δεδομένων, κ.λπ.). Όπως φαίνεται από τα παραπάνω, τα μεγάλα δεδομένα στην υγειονομική περίθαλψη μπορεί να αναφέρονται στη φυσιολογία, τη συμπεριφορά, τη βιολογία, την κλινική πρακτική, την περιβαλλοντική έκθεση, την ιατρική απεικόνιση, τη διαχείριση ασθενειών, το ιστορικό συνταγογραφούμενων φαρμάκων, τη διατροφή ή ακόμη και παράγοντες άσκησης. Στη διεθνή βιβλιογραφία δεν υπάρχει κοινά αποδεκτή ταξινόμηση των πηγών μεγάλων δεδομένων στην υγειονομική περίθαλψη. Σύμφωνα με μια άποψη, υπάρχουν δύο κύριες πηγές μεγάλων δεδομένων: η γονιδιωματική έρευνα (δεδομένα που προέρχονται από την έρευνα σχετικά με τον γονότυπο, τη γονιδιακή έκφραση και τα δεδομένα αλληλούχισης) και οι αγοραστές και οι πάροχοι υπηρεσιών υγειονομικής περίθαλψης (δεδομένα που προέρχονται από ηλεκτρονικά αρχεία υγείας, ασφαλιστικά αρχεία, συνταγές φαρμάκων, ανατροφοδότηση ασθενών κ.λπ.) (Wang et al., 2018).

Τα πλεονεκτήματα των μεγάλων δεδομένων στην υγειονομική περίθαλψη έχουν αποδειχθεί κυρίως στους τομείς της πρόληψης ασθενειών, του εντοπισμού τροποποιήσιμων παραγόντων κινδύνου για ασθένειες και της δημιουργίας παρεμβάσεων για τη μεταβολή της συμπεριφοράς ενός ατόμου. Σύμφωνα με τους Rumsfeld και συν., η ανάλυση μεγάλων δεδομένων στην υγεία μπορεί να προωθήσει σημαντικά τους ακόλουθους τομείς:

1. Ανάπτυξη των προγνωστικών μοντέλων αναφορικά με τους κινδύνους και τη χρήση των πόρων,
2. Διαχείριση πληθυσμού,
3. Επιτήρηση της ασφάλειας των φαρμάκων και του τεχνολογικού εξοπλισμού,
4. Ετερογένεια των ασθενειών και των θεραπειών,
5. Ιατρική ακριβείας και συστήματα υποστήριξης κλινικών αποφάσεων,
6. Ποιότητα φροντίδας και μέτρηση της απόδοσης (performance measurement),
7. Δημόσια υγεία, και
8. Εφαρμογές έρευνας (Wang et al., 2018).

Για όλο το φάσμα των οργανισμών υγειονομικής περίθαλψης, από μικρές ιδιωτικές κλινικές έως τεράστια νοσοκομειακά δίκτυα, η ψηφιοποίηση, η συσχέτιση και η αποτελεσματική χρήση των μεγάλων δεδομένων συνεπάγεται μια σειρά σημαντικών πλεονεκτημάτων. Τα πιθανά πλεονεκτήματα περιλαμβάνουν τη διαχείριση τόσο της ατομικής όσο και της δημόσιας υγείας, τον έγκαιρο εντοπισμό της απάτης στον τομέα της υγειονομικής περίθαλψης και την αναγνώριση ασθενειών σε πρώιμο στάδιο, όταν αυτές μπορούν να αντιμετωπιστούν ευκολότερα και αποτελεσματικότερα. Η εξέταση αυτών των δεδομένων, η οποία αποτελεί ένα διαφορετικό επιστημονικό αντικείμενο που ονομάζεται ανάλυση μεγάλων δεδομένων (big data analytics), μπορεί επίσης να δώσει απαντήσεις σε πολυάριθμα ζητήματα. Τεράστιες ποσότητες ιστορικών δεδομένων μπορούν να χρησιμοποιηθούν για την πρόβλεψη ή ακόμη και τον υπολογισμό ορισμένων εξελίξεων ή αποτελεσμάτων, όπως η διάρκεια παραμονής (LOS11), οι ασθενείς που θα επέλεγαν να χειρουργηθούν, οι ασθενείς που πιθανώς δεν θα ήθελαν να χειρουργηθούν, οι επιπλοκές, οι ασθενείς που κινδυνεύουν από σήψη, οι ενδονοσοκομειακές ασθένειες, η εξέλιξη της νόσου και οι αιτιολογικοί της παράγοντες και οι συνοσηρότητες (Wang et al. , 2018).

Σύμφωνα με προβλέψεις του Διεθνούς Ινστιτούτου McKinsey, η χρήση της ανάλυσης μεγάλων δεδομένων μπορεί να εξοικονομήσει στο σύστημα υγειονομικής περίθαλψης των ΗΠΑ 300 δισεκατομμύρια δολάρια ετησίως, εκ των οποίων τα δύο τρίτα θα προέλθουν από τη μείωση του κόστους υγειονομικής περίθαλψης. Δύο από τους βασικούς τομείς στους οποίους θα μπορούσαν να εξοικονομηθούν πόροι είναι οι κλινικές εφαρμογές και ο τομέας της έρευνας και ανάπτυξης (E&A ή R&D), όπου οι

απώλειες από τη σπατάλη πόρων έχουν φθάσει τα 165 δισεκατομμύρια δολάρια και τα 108 δισεκατομμύρια δολάρια, αντίστοιχα (Wang et al., 2018).

2.2 Ηλεκτρονικά συστήματα στην Ελλάδα

Οι πληροφορίες για την υγεία των πολιτών, οι οποίες αποτελούνται κυρίως από κλινικά, βιομετρικά και συνταγογραφικά δεδομένα, συλλέγονται από διάφορες πηγές. Η συλλογή και η επεξεργασία των ιατρικών και νοσηλευτικών δεδομένων γίνεται από έναν σημαντικό αριθμό οργανισμών και ιδρυμάτων, κυρίως του δημόσιου τομέα. Το ακόλουθο σχήμα παρέχει μια σχηματική επισκόπηση των πρωτογενών πηγών ψηφιακών δεδομένων υγείας στην Ελλάδα (ΣΕΒ, 2020) :



Διάγραμμα 1: Ψηφιακά δεδομένα υγείας. Πηγή: ΣΕΒ, 2020

Σχεδόν 11 εκατομμύρια άνθρωποι στην Ελλάδα χρησιμοποιούν το ηλεκτρονικό σύστημα υγείας, το οποίο περιλαμβάνει επίσης 66.058 γιατρούς που εκτελούν συνταγές στον δημόσιο και εμπορικό τομέα, 11.368 φαρμακοποιούς, εξωτερικά ιατρεία νοσοκομείων, παρόχους διαγνωστικών υπηρεσιών και επαγγελματίες υγείας (ΕΛ.ΣΤΑΤ, 2019).

Ηλεκτρονική Διακυβέρνηση Κοινωνικής Ασφάλισης (ΗΔΙΚΑ)

Από την ίδρυσή της το 2007, η μη κερδοσκοπική ανώνυμη εταιρεία με την επωνυμία "Ηλεκτρονική Διακυβέρνηση Κοινωνικής Ασφάλισης" (ΗΔΙΚΑ Α.Ε.) έχει εργαστεί σε πρωτοβουλίες πληροφορικής και επικοινωνιών για φορείς κοινωνικής ασφάλισης, καθώς και για τις επεκτάσεις των εν λόγω οργανισμών. Στόχος της ΗΔΙΚΑ είναι να προσφέρει ολοκληρωμένες, υψηλής ποιότητας λύσεις πληροφορικής και επικοινωνιών που θα υποστηρίξουν τη μακροπρόθεσμη, ακριβή, πλήρη και αποτελεσματική λειτουργία των φορέων κοινωνικής ασφάλισης και υγείας, καθώς και την παροχή σύγχρονων ηλεκτρονικών υπηρεσιών και πληροφοριών προς τους πολίτες. Με την ολοκλήρωση συγκεκριμένων στόχων όπως αυτοί, είναι σε θέση να εκπληρώσει την αποστολή της.

-Η δημιουργία και συντήρηση ηλεκτρονικών υπηρεσιών εθνικής εμβέλειας στους τομείς της υγείας και της κοινωνικής ασφάλισης με παράλληλη βέλτιστη αξιοποίηση των πόρων του ΕΣΠΑ (Μίνου et al., 2020).

-Ο εκσυγχρονισμός, η τυποποίηση και η επικαιροποίηση των εφαρμογών που χρησιμοποιούνται από δημόσιους φορείς όπως οι φορείς κοινωνικής ασφάλισης.

-Η εξέλιξη και ενδυνάμωση της επιχείρησης μέσω της επέκτασης των δυνατοτήτων της και της δημιουργίας μιας νέας εταιρικής ταυτότητας που θα την τοποθετεί ως "νησίδα πληροφόρησης/εκπαίδευσης/γνώσης της ηλεκτρονικής διακυβέρνησης της κοινωνικής ασφάλισης και της υγείας".

Το πλαίσιο δραστηριοποίησης της ΗΔΙΚΑ περιλαμβάνει τους ακόλουθους βασικούς τομείς: □

-Σχεδιασμός, ανάπτυξη, λειτουργία, διαχείριση και συντήρηση συστημάτων πληροφοριών και επικοινωνιών, συμπεριλαμβανομένου του υλικού, του λογισμικού και των υπηρεσιών, για χρήση από όλους τους φορείς που σχετίζονται με την κοινωνική ασφάλιση, την υγειονομική περίθαλψη και το ευρύ κοινό.

-Εκτέλεση και διαχείριση έργων στους τομείς της πληροφορικής, της επικοινωνίας και των νέων τεχνολογιών του ΕΣΠΑ, καθώς και της διοικητικής μεταρρύθμισης για την ενίσχυση των υπηρεσιών κοινωνικής ασφάλισης και υγείας (Μίνου et al., 2020).

-Διεξάγονται διαγωνισμοί για έργα ενίσχυσης του πλαισίου διοικητικής και οικονομικής διαχείρισης και των προτύπων παροχής υπηρεσιών των φορέων κοινωνικής ασφάλισης και υγείας.

-Η χρήση ηλεκτρονικών υπηρεσιών από τους πολίτες κατά την αλληλεπίδραση με τους φορείς κοινωνικής ασφάλισης (ηλεκτρονική διακυβέρνηση).

-Διασφάλιση και προώθηση της διαλειτουργικότητας των συστημάτων πληροφορικής και επικοινωνίας των φορέων που εμπλέκονται στην υγεία, την πρόνοια και την κοινωνική πολιτική, καθώς και των φορέων κοινωνικής ασφάλισης που υπάγονται στο Υπουργείο Εργασίας & Κοινωνικής Ασφάλισης (Μίνου et al., 2020).

-Συνδυασμός των δεδομένων στους τομείς της κοινωνικής ασφάλισης και της υγείας στην Ελλάδα.

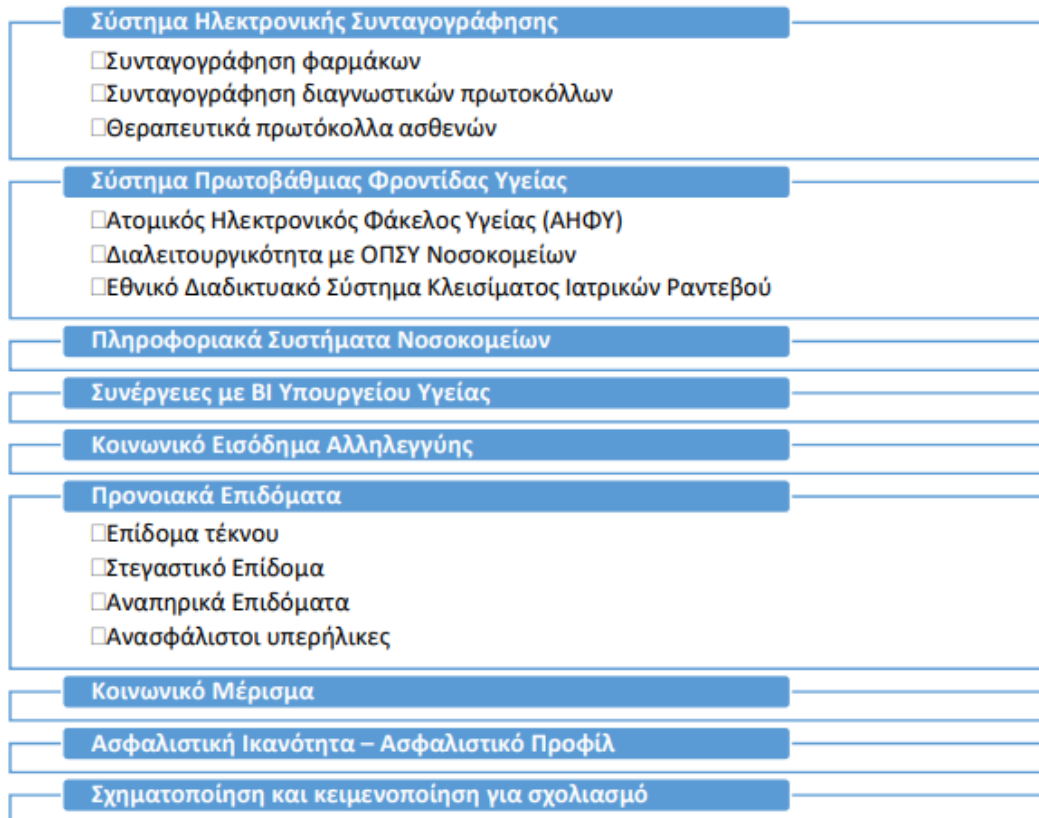
-Παροχή καθοδήγησης στους φορείς κοινωνικής ασφάλισης.

-Συνεργασία και επικοινωνία με τους εταίρους της ΕΕ.

-Παροχή αξιολογήσεων και στατιστικών στοιχείων για το σύστημα κοινωνικής ασφάλισης της Ελλάδας.

Σύμφωνα με το νόμο 4600 (ΦΕΚ 43/Α/09.03.2020), πρόσφατα εφαρμόστηκε και συντηρείται από την ΗΔΙΚΑ ένας ατομικός ηλεκτρονικός φάκελος υγείας (ΑΗΦΥ). Ο Ηλεκτρονικός Φάκελος Υγείας είναι μια τυποποιημένη μέθοδος τεκμηρίωσης του τρόπου με τον οποίο κάθε Έλληνας χρησιμοποιεί τις υπηρεσίες υγείας. Συνδέεται με τον ΑΜΚΑ του ασθενούς και με τη σύνδεση των ιατρικών δεδομένων του ασθενούς, οι αξιολογήσεις σχετικά με τη χρήση των υπηρεσιών υγείας και οι ιατρικές αποφάσεις μπορούν να ληφθούν από μια πιο ολοκληρωμένη σκοπιά, καθώς οι κλινικοί ιατροί έχουν πρόσβαση σε όλα τα ιατρικά δεδομένα. Όλες οι πληροφορίες που προκύπτουν από μια εξέταση ή μια επίσκεψη που απαιτείται για τη διάγνωση, τη θεραπεία, τη νοσηλεία ή την αποκατάσταση της σωματικής ή ψυχικής υγείας πρέπει να εισάγονται στο Σύστημα Ηλεκτρονικής Συνταγογράφησης (ΣΗΣ) από γιατρούς, οδοντιάτρους και άλλους επαγγελματίες υγείας που είναι πιστοποιημένοι και εξουσιοδοτημένοι χρήστες του συστήματος (Μίνου et al., 2020).

Εν κατακλείδι, το ΗΔΙΚΑ προσφέρει μεγάλα δεδομένα και ηλεκτρονικές υπηρεσίες για τα ακόλουθα στον τομέα της υγείας:



BI-Health (Υπουργείο Υγείας)

Το Σύστημα Επιχειρηματικής Ευφυΐας (BI) του Εθνικού Συστήματος Υγείας (ΕΣΥ), γνωστό και ως BI-Health, είναι ένα πληροφοριακό σύστημα αιχμής που βελτιώνει τις δράσεις των συνεχιζόμενων διοικητικών πληροφοριών του Υπουργείου Υγείας. Μέσω του εξορθολογισμού των διαδικασιών διοικητικής πληροφόρησης, της αποτελεσματικότερης διαχείρισης των πόρων και του βαθύτερου ελέγχου των επιχειρησιακών και οικονομικών δεδομένων, το BI-Health αποτελεί βασικό παράγοντα στον οργανωτικό, επιχειρησιακό και οικονομικό εκσυγχρονισμό του ΕΣΥ. Το σύστημα BI-Health διασφαλίζει ότι τα αναλυτικά και συγκεντρωτικά δεδομένα από τις μονάδες δημόσιας υγείας της επικράτειας συγκεντρώνονται και επεξεργάζονται σε κεντρικό επιχειρησιακό επίπεδο. Επιτρέπει επίσης τη διάχυση των πληροφοριών στους διοικητικούς μηχανισμούς με απώτερο στόχο την αύξηση του επιπέδου των προσφερόμενων υπηρεσιών υγείας (Μίνου et al., 2020).

Μέσω της ηλεκτρονικής επεξεργασίας και διαχείρισης των διαθέσιμων δεδομένων, το BI-Health στοχεύει στη βελτίωση των επιχειρησιακών δυνατοτήτων του

Υπουργείου Υγείας, παρέχοντας παράλληλα στους πολίτες και σε άλλα μέρη τη δυνατότητα πρόσβασης σε πληροφορίες και δεδομένα μέσω μιας διαδικτυακής πύλης. Το Σύστημα επιτρέπει την ηλεκτρονική μεταφορά εγγράφων και δεδομένων σε διάφορους τύπους δεδομένων. Πρόκειται για μια ολοκληρωμένη λύση MIS-BI για τον κλάδο της υγειονομικής περίθαλψης που συνδέει το Υπουργείο Υγείας, επτά περιφερειακά συστήματα πληροφοριών υγείας, 131 νοσοκομεία, 225 κέντρα υγείας, περιφερειακές κλινικές υγείας και περιφερειακά κέντρα υγείας πολλαπλών χρήσεων με σκοπό τη συλλογή, την ομογενοποίηση και την επεξεργασία δεδομένων. Συμμορφώνεται επίσης με όλες τις τεχνικές απαιτήσεις.

Οι υπηρεσίες που προσφέρονται από το υλικό και το λογισμικό που δημιουργήθηκε στο πλαίσιο του έργου και οι οποίες βοηθούν άμεσα στην επίτευξη των στόχων που έχει θέσει το Υπουργείο Υγείας: □

-Κατασκευή υποδομής αποθήκης δεδομένων με εργαλεία για τη σύνδεση με εφαρμογές για δεδομένα και εξαγωγή δεδομένων. Πρωταρχικός στόχος αυτής της υποδομής είναι να παρέχει στα στελέχη του Υπουργείου ακριβείς και ενημερωμένες πληροφορίες, καθώς και να τα βοηθήσει στη λήψη αποφάσεων με τη χρήση τεχνολογίας αιχμής και αποτελεσματικής πληροφορικής (Μίνου et al., 2020).

-Προκειμένου να αξιολογηθεί το έργο και ο αντίκτυπος των πολιτικών που εφαρμόζονται, να μελετηθούν οι βασικοί οικονομικοί δείκτες, να ανατροφοδοτηθεί η παραγωγική διαδικασία με τα αποτελέσματα των ερευνών και να βελτιωθεί η διαχείριση των υπηρεσιών, παρέχονται επιχειρησιακές πληροφορίες για το σύνολο της δραστηριότητας των υπηρεσιών του Υπουργείου, των νοσοκομείων, των κέντρων υγείας και των λοιπών μονάδων υγείας, καθώς και για το σύνολο της οικονομικής δραστηριότητας γενικότερα, αξιοποιώντας τον πληροφοριακό πλούτο των επιμέρους παραγωγικών υποσυστημάτων. Το σύστημα προσφέρει τη δυνατότητα υποβολής ερωτημάτων, διενέργειας ερευνών και παραγωγής αναφορών και αναλύσεων ανά λογαριασμό με δυναμικά κριτήρια και φίλτρα που καθορίζονται κατά περίπτωση από τους χρήστες των Κεντρικών Υπηρεσιών.

-Στατιστική επεξεργασία και ανάλυση των δεδομένων της αποθήκης δεδομένων για τη διερεύνηση των συσχετίσεων μεταξύ των διαφόρων οικονομικών μεταβλητών, τον εντοπισμό των βασικών παραγόντων που επηρεάζουν τις τιμές των βασικών οικονομικών μεταβλητών και τη διερεύνηση των τάσεων και των κατανομών. Για τη δημιουργία αυτών των εκθέσεων θα χρησιμοποιηθεί λογισμικό που δημιουργήθηκε ειδικά για το σκοπό αυτό. Οι χρήστες καθορίζουν δυναμικά τους

παράγοντες που λαμβάνονται υπόψη και το επίπεδο διερεύνησης (λεπτομερή ή συγκεντρωτικά δεδομένα και αντίστοιχα επίπεδα). Χρησιμοποιούνται γραφήματα για την παρουσίαση των αποτελεσμάτων της επεξεργασίας στην οθόνη του χρήστη (ράβδοι, πίτες, ιστογράμματα κ.λπ.). Σύμφωνα με τις αξιολογήσεις ασφαλείας των χρηστών, είναι επίσης δυνατή η εκτύπωση αναφορών και η εξαγωγή των δεδομένων.

-Η δημιουργία, η διερεύνηση και η προβολή των πιθανών επιπτώσεων των οικονομικών μέτρων και πολιτικών που λαμβάνονται υπόψη με τη μορφή σεναρίων για την υποβοήθηση του προσωπικού των κεντρικών υπηρεσιών στη διατύπωση συστάσεων και στη διαδικασία λήψης αποφάσεων. Για το σκοπό αυτό έχουν αξιοποιηθεί οικονομομετρικά μοντέλα, μοντέλα προσομοίωσης και στατιστικά μοντέλα σε συνδυασμό με τη σύγχρονη τεχνολογία (OLAP, Data Mining κ.λπ.). Προκειμένου να εκτιμηθούν τα αναμενόμενα αποτελέσματα, οι υποθετικές καταστάσεις που προκύπτουν από την εξέταση διαφόρων μέτρων μπορούν να επεξεργαστούν με αυτόν τον τρόπο.

-Αποθήκευση και χρήση δεδομένων από τουλάχιστον πέντε (5) προηγούμενα έτη, προκειμένου να αυξηθεί η αξιοπιστία των στατιστικών ευρημάτων.

-Παρακολούθηση της ακρίβειας των δεδομένων που αποστέλλουν οι μονάδες υγείας στο σύστημα MIS-BI. Στόχος είναι η αποστολή ειδοποιήσεων στις μονάδες που δεν αποστέλλουν δεδομένα στο σύστημα ή αποστέλλουν ελλιπή δεδομένα σύμφωνα με τις απαιτήσεις του συστήματος MIS-BI (Μίνου et al., 2020).

Εθνικό Δίκτυο Υποδομών Τεχνολογίας και Έρευνας (ΕΔΥΤΕ)

Η ΕΔΥΤΕ Α.Ε. συνδράμει το έργο του Υπουργείου Υγείας τα τελευταία 7 χρόνια, παρέχοντας υποδομές δικτύου, υπολογιστών και αποθήκευσης αιχμής, καθώς και αναπτύσσοντας υπηρεσίες με σκοπό την υποβοήθηση του κλινικού έργου, την ενίσχυση της ερευνητικής δραστηριότητας και την ενίσχυση της ανταγωνιστικότητας σημαντικών τομέων της υγείας. Πιο συγκεκριμένα, 31 νοσοκομειακές μονάδες σε όλη την επικράτεια έχουν ήδη λάβει τις ακόλουθες υποδομές και υπηρεσίες χάρη στο έργο που έχουν επιτελέσει η ΕΔΥΤΕ Α.Ε. και το Υπουργείο Υγείας (και το οποίο επεκτείνεται επί του παρόντος σε όλες τις νοσοκομειακές μονάδες της χώρας).

-Σύνδεση όλων των δημόσιων νοσοκομείων της χώρας με το δίκτυο οπτικών ινών της ΕΔΕΤ και, στη συνέχεια, με το διαδίκτυο, σε εξαιρετικά γρήγορους ρυθμούς έως και 10Gbps.

-Η προμήθεια και εγκατάσταση σημείων ασύρματης πρόσβασης (Wi-Fi) εντός των νοσοκομείων.

-Η παροχή υπηρεσίας εφεδρικής αποθήκευσης των δεδομένων που παράγονται από τις απεικονιστικές συσκευές των νοσοκομείων στο Κέντρο Δεδομένων Υγείας, το οποίο έχει εγκατασταθεί και λειτουργεί σήμερα από την ΕΔΥΤΕ Α.Ε. στο Ηράκλειο Κρήτης, με βάση το πρότυπο DICOM (υπηρεσία HARMONI).

-Η επέκταση του Κέντρου Δεδομένων Υγείας για την κάλυψη των αναγκών όλων των νοσοκομείων.

-Η δημιουργία λίστας χρηστών για τους εργαζόμενους των νοσοκομείων, ώστε με ένα λογαριασμό χρήστη να έχουν πρόσβαση σε όλα τα πληροφοριακά συστήματα των νοσοκομείων και στις νέες υπηρεσίες της ΕΔΥΤΕ Α.Ε., ανάλογα με την εργασία τους.

-Προσφορά υπηρεσίας για την πρόσβαση των επιστημόνων των νοσοκομείων στην υπηρεσία δια-δανεισμού ιατρικών ειδών ΔΙΑΔΟΣΙΣ. Με τη χρήση των συνδρομών HEALink, το επιστημονικό προσωπικό όλων των νοσοκομείων μπορεί πλέον να διαβάζει ιατρικά άρθρα που, μέχρι πρόσφατα, ήταν διαθέσιμα αποκλειστικά στους συναδέλφους τους στα πανεπιστημιακά νοσοκομεία.

-Μια στοχευμένη αναβάθμιση των περιφερειακών νοσοκομειακών δικτύων ώστε να καταστεί δυνατή η παροχή νέων υπηρεσιών και υποδομών.

-Παροχή δικτυακών, υπολογιστικών και αποθηκευτικών πόρων (Cloud) του Κέντρου Δεδομένων Υγείας στο Υπουργείο Υγείας και σε νοσοκομειακές μονάδες για την υποστήριξη ερευνητικών δραστηριοτήτων στους τομείς της ιατρικής, της βιοτεχνολογίας και της βιοπληροφορικής, οι οποίες έχουν υψηλές απαιτήσεις για την αποθήκευση και την επεξεργασία του τεράστιου όγκου δεδομένων που παράγονται από τη σύγχρονη απεικόνιση (για παράδειγμα, ιατρικές εικόνες και βίντεο υψηλής ανάλυσης) και πειραματικές συσκευές (για παράδειγμα, τεχνολογίες υβριδοποίησης μικροσυστοιχιών DNA με βάση τα γονίδια) (Minou et al, 2020).

Το Εθνικό Μητρώο Αιμοδοτών(EMA) δημιουργήθηκε, υλοποιήθηκε και λειτουργεί από την ΕΔΥΤΕ Α.Ε. υπό την ηγεσία του Εθνικού Κέντρου Αιμοδοσίας (ΕΚΑ). Το EMA αποτελείται από ένα ολοκληρωμένο πληροφοριακό σύστημα αιχμής που επιτρέπει τη διατήρηση του μητρώου αιμοδοτών και την κατασκευή σύγχρονων,

φιλικών προς το χρήστη διαδικασιών προς όφελος τόσο των αιμοδοτών όσο και του συνόλου των υπηρεσιών αιμοδοσίας της χώρας. Απώτερος στόχος είναι να καταστεί το αίμα και τα παράγωγα του αίματος ασφαλέστερα. Εκτός από το Εθνικό Μητρώο Αιμοδοτών, εκδίδονται επίσης κάρτες αναγνώρισης εθελοντών αιμοδοτών, απλουστεύοντας σημαντικά τη διαδικασία αιμοδοσίας. Η δεύτερη φάση του έργου, η οποία προβλέπει τη δημιουργία ενός σύγχρονου και ολοκληρωμένου εθνικού συστήματος αιμοδοσίας που θα καλύπτει ολόκληρο τον κύκλο της αιμοδοσίας, από τη φλέβα του αιμοδότη έως τη φλέβα του ασθενούς, υλοποιείται επί του παρόντος στο πλαίσιο της συνεχιζόμενης συνεργασίας μεταξύ της ΕΔΥΤΕ Α.Ε. και του Ε.ΚΕ.Α (vein-to-vein) (Minou et al., 2020).

Εθνικός Οργανισμός Παροχής Υπηρεσιών Υγείας (ΕΟΠΥΥ)

Τόσο οι πολίτες όσο και οι επαγγελματίες υγείας μπορούν να υποβάλουν ηλεκτρονική αίτηση μέσω του Εθνικού Οργανισμού Παροχής Υπηρεσιών Υγείας (ΕΟΠΥΥ). Τα δεδομένα για συγκεκριμένες διαγνωστικές εξετάσεις και την αποζημίωσή τους πρόκειται να συγκεντρωθούν από το σύστημα e-ΔΑΠΥ. Τα στοιχεία αφορούν πληροφορίες για τους παρόχους διαγνωστικών εξετάσεων και πληροφορίες για τις διαγνωστικές εξετάσεις που διενεργούνται για κάθε ασθενή και πάροχο (Minou et al., 2020).

Υπηρεσίες για τους πολίτες

- Φάκελος ασφάλισης υγείας
- Εγγραφή σε Οικογενειακό ιατρό,
- Αναζήτηση παρόχου υγείας
- Αναζήτηση φαρμακείου ΕΟΠΥΥ
- Αναζήτηση φαρμάκου
- Ραντεβού με ιατρούς ΕΟΠΥΥ
- Χάρτης υγείας
- Παροχές ΕΚΠΥ και συμβεβλημένοι πάροχοι (Αναζήτηση Εγκεκριμένων Προϊόντων ΕΟΠΥΥ)
- Παραλαβή ΦΥΚ (φάρμακα υψηλού κόστους)
- Ηλεκτρονική Υποβολή ατομικών αιτημάτων
- Εφαρμογή ραντεβού σε Περιφερειακές Διευθύνσεις (ΠΕ.ΔΙ.) του ΕΟΠΥΥ, για την εξυπηρέτησή σε θέματα που αφορούν στην απόδοση δαπανών Παροχών Ασφαλισμένων

Υπηρεσίες για τους παρόχους υγείας

- Έλεγχος και εκκαθάριση συνταγών (ΚΜΕΣ)
- Ηλεκτρονική συνταγογραφηση
- Εοργynet (Μητρώο Αποζημιούμενων προϊόντων ΕΟΠΥΥ/συμβάσεις παρόχων)
- Παροχές ΕΚΠΥ και συμβεβλημένοι πάροχοι
- Σύμβαση Ιδιώτη Οικογενειακού Ιατρού
- Ασφαλιστική ικανότητα
- Ιατρικές υπηρεσίες
- Εδαφυ Τεχνικές προδιαγραφές αρχείων
- Κινήσεις παρόχων
- Ληξιπρόθεσμες οφειλές
- Αιμοκαθάρσεις
- Ακτινοθεραπείες
- Αναγγελία Εισόδου/Εξόδου
- ΕΚΑΒ/Διαχείριση ΜΕΘ
- Μητρώο Ηπατίτιδας C
- Μητρώο Χρόνιας Μυελογενούς Λευχαιμίας
- Ενημέρωση Συναλλασσόμενων
- Σύστημα Ηλεκτρονικής Προέγκρισης Φαρμάκων
- Μητρώο Απομυελινωτικών παθήσεων του ΚΝΣ
- Μητρώο Σακχαρώδη διαβήτη
- Μητρώο Περιτοναϊκής Κάθαρσης
- Επιτροπές φαρμάκων Εξωτερικού και Υψηλού κόστους ΕΟΠΥΥ
- Μητρώο Θαλασσαιμίας
- Μητρώο Εγκαυματιών
- Διάθεση Φαρμάκων Υψηλού κόστους από ιδιωτικά φαρμακεία

Εθνικός Οργανισμός Φαρμάκων (ΕΟΦ)

Με το νόμο 1316/1983 ιδρύθηκε ο Εθνικός Οργανισμός Φαρμάκων ως Νομικό Πρόσωπο Δημοσίου Δικαίου που υπάγεται στο Υπουργείο Υγείας. Το πεδίο εφαρμογής του ΕΟΦ περιλαμβάνει όλα τα φαρμακευτικά προϊόντα για ανθρώπινη και κτηνιατρική χρήση, τα βιολογικά προϊόντα για ανθρώπινη και κτηνιατρική χρήση, τις φαρμακούχες ζωοτροφές και τα πρόσθετα ζωοτροφών, τα ειδικά διαιτητικά τρόφιμα και τα συμπληρώματα διατροφής, τα βιοκτόνα, τον ιατρικό εξοπλισμό και τα καλλυντικά (Μίνου et al., 2020).

Ο ΕΟΦ λειτουργεί σύμφωνα με την εθνική και ευρωπαϊκή νομοθεσία, με πλήρη διαφάνεια και σε συνεργασία με την Ευρωπαϊκή Ένωση: □

-Αξιολογεί και αποδέχεται αντικείμενα.

-Αξιολογεί συνεχώς την αποτελεσματικότητα, την ασφάλεια και την ποιότητα των φαρμακευτικών προϊόντων καθ' όλη τη διάρκεια της εμπορικής τους ζωής.

-Παρακολουθεί, ελέγχει και εποπτεύει την αγορά για όλα τα αγαθά για τα οποία είναι υπεύθυνη.

-Ελέγχει την τήρηση των προτύπων άριστης παρασκευής, εργαστηριακής και κλινικής πρακτικής, καθώς και την ορθή εφαρμογή των νόμων που διέπουν την εμπορία, τη διανομή, την πώληση και τη διαφήμιση των φαρμακευτικών προϊόντων.

-Δημιουργεί και προωθεί τη φαρμακευτική έρευνα και μελέτη.

-Να παρέχει στο κοινό, στους αρμόδιους φορείς και στους επαγγελματίες υγείας τις πιο πρόσφατες πληροφορίες σχετικά με τα είδη που είναι επιφορτισμένη με την ευθύνη της, ώστε να διασφαλίζεται η ασφαλής και κατάλληλη χρήση τους.

Όταν πωλείται ένα φαρμακευτικό προϊόν, ο ΕΟΦ παρακολουθεί την τιμή και τον αριθμό των συσκευασιών που πωλούνται κάθε μήνα (Μίνου et al., 2020).

Εθνικός Οργανισμός Δημόσιας Υγείας (ΕΟΔΥ)

Ο Εθνικός Οργανισμός Δημόσιας Υγείας (ΕΟΔΥ), ο οποίος ιδρύθηκε το 2019, διοικείται από τον υπουργό Υγείας. Τα προηγούμενα Κέντρα Ελέγχου και Πρόληψης Νοσημάτων καταργήθηκαν το 2019 και ο ΕΟΔΥ είναι ο παγκόσμιος αντικαταστάτης.

Οι βασικές αρμοδιότητες του ΕΟΔΥ περιλαμβάνουν πράγματα όπως η επιδημιολογική επιτήρηση, η εκτίμηση κινδύνων, η παροχή γνωμοδοτήσεων επί επιστημονικών θεμάτων, η ετοιμότητα και η αντιμετώπιση, η παροχή αξιόπιστων και συγκρίσιμων επιδημιολογικών δεδομένων και στατιστικών σε αρμόδιους φορείς σε εθνικό, ευρωπαϊκό και διεθνές επίπεδο, η εκπαίδευση και κατάρτιση στον τομέα της δημόσιας υγείας, η ενημέρωση του κοινού και των επαγγελματιών του ιατρικού κλάδου για τους κινδύνους από σοβαρές απειλές για την υγεία και η προώθηση δραστηριοτήτων ευαισθητοποίησης. Με ετοιμότητα για την αντιμετώπιση εκτάκτων καταστάσεων υγείας, ο ΕΟΔΥ αποτελεί τον επιχειρησιακό κόμβο για το σχεδιασμό και την εκτέλεση δράσεων προστασίας της δημόσιας υγείας προληπτικού, επιχειρησιακού και παρεμβατικού χαρακτήρα. Συντονίζει τις δραστηριότητές του με τις ανάγκες του κράτους και των διεθνών οργανισμών με τους οποίους συνεργάζεται.

Ο ΕΟΔΥ καταγράφει επιδημιολογικά – στατιστικά δεδομένα επιτήρησης των λοιμωδών νοσημάτων με στόχο:

-Καθορισμός των μεταβλητών κινδύνου που σχετίζονται με μια ασθένεια, ένα ζήτημα ή ένα συμβάν και μέτρηση του επιπολασμού του.

-Καθορισμός του επιπολασμού μιας ασθένειας και παρακολούθηση των τάσεων της με την πάροδο του χρόνου.

-Η ανίχνευση επιδημιών.

-Η αξιολόγηση πρωτοβουλιών και σχεδίων δημόσιας υγείας.

Η επιδημιολογική επιτήρηση γίνεται κυρίως μέσω των 3 συστημάτων επιτήρησης (Ειδικά Συστήματα Επιτήρησης/Δήλωσης Κρουσμάτων), Σύστημα Εργαστηριακής Επιτήρησης, Σύστημα Παρατηρητών Νοσηρότητας (Sentinel), Σύστημα Υποχρεωτικής Δήλωσης (Minou et al., 2020).

2.3 Πηγές Big Data στον Τομέα Υγείας

Τα δεδομένα υγειονομικής περίθαλψης είναι ασύνδετα και κατακερματισμένα και προέρχονται από διάφορες πηγές με διάφορους τύπους και μορφές. Τα ηλεκτρονικά αρχεία υγείας, τα συστήματα καταχώρησης ιατρικών εντολών (CPOE), τα συστήματα υποστήριξης κλινικών αποφάσεων (CDSS), τα αρχεία χορήγησης φαρμάκων, τα εργαστηριακά και φαρμακευτικά αρχεία των μελετών κοόρτης, οι κυβερνητικές έρευνες και οι κλινικές δοκιμές μπορούν να χρησιμοποιηθούν για τη συλλογή κλινικών δεδομένων, όπως τα ζωτικά σημεία, το ιατρικό ιστορικό του ασθενούς, τα φάρμακα, οι εμβολιασμοί και οι απεικονιστικές εξετάσεις. Η ημερομηνία εισαγωγής του ασθενούς, η ημερομηνία εξιτηρίου, η διάγνωση σύμφωνα με τη Διεθνή Στατιστική Ταξινόμηση των Ασθενειών και των Συναφών Προβλημάτων Υγείας (ICD), η κατάσταση της υγείας του ασθενούς κατά την έξοδο, καθώς και δεδομένα οικονομικών απαιτήσεων, συμπεριλαμβανομένων, μεταξύ άλλων, των χρεώσεων επίσκεψης, ποιος πρέπει να πληρώσει την αμοιβή για το θα πρέπει να δει έναν γιατρό, περιλαμβάνονται όλα στα διοικητικά δεδομένα.

Οι πηγές μεγάλων δεδομένων υγειονομικής περίθαλψης μπορούν να κατηγοριοποιηθούν χονδρικά σε εσωτερικές (όπως ηλεκτρονικά αρχεία υγείας,

συστήματα υποστήριξης κλινικών αποφάσεων, CPOE κ.λπ.) και εξωτερικές (όπως πηγές δεδομένων τρίτων) (κυβερνητικές πηγές, εργαστήρια, φαρμακεία, ασφαλιστικές εταιρείες κ.λπ.). Οι πληροφορίες αυτές αποθηκεύονται σε διάφορες μορφές (επίπεδα αρχεία, σχεσιακοί πίνακες, ASCII/κείμενο κ.λπ.), φυλάσσονται σε διάφορες τοποθεσίες (τόσο φυσικά όσο και με την έννοια ότι φιλοξενούνται σε διάφορες εγκαταστάσεις παρόχων υγειονομικής περίθαλψης) και φυλάσσονται σε διάφορες εφαρμογές (εφαρμογές επεξεργασίας συναλλαγών, βάσεις δεδομένων κ.λπ.). Όπως φαίνεται από τα παραπάνω, τα μεγάλα δεδομένα στην υγειονομική περίθαλψη μπορεί να αναφέρονται στη φυσιολογία, τη συμπεριφορά, τη βιολογία, την κλινική πρακτική, την περιβαλλοντική έκθεση, την ιατρική απεικόνιση, τη διαχείριση ασθενειών, το ιστορικό συνταγογραφούμενων φαρμάκων, τη διατροφή ή ακόμη και παράγοντες άσκησης. Στη διεθνή βιβλιογραφία δεν υπάρχει κοινά αποδεκτή ταξινόμηση των πηγών μεγάλων δεδομένων στην υγειονομική περίθαλψη. Σύμφωνα με μια άποψη, η γενετική έρευνα (δεδομένα που παράγονται από τη μελέτη του γονότυπου, της γονιδιακής έκφρασης και των δεδομένων αλληλουχίας) και ο καταναλωτής-πάροχος υπηρεσιών υγείας είναι οι δύο βασικές πηγές Μεγάλων Δεδομένων (δεδομένα που προέρχονται από ηλεκτρονικά αρχεία υγείας, ασφαλιστικά αρχεία, συνταγές φαρμάκων, ανατροφοδότηση ασθενών κ.λπ.) (Alguliyev & Hajirahimova, 2014).

Από την άλλη πλευρά, σε μελέτη του ο Swan ταξινόμησε τις μεγάλες ροές των Big Data στην υγεία σε:

Παραδοσιακά δεδομένα υγείας που λαμβάνονται από κλινικές βάσεις δεδομένων, διοικητικές βάσεις δεδομένων (ασφαλιστικές απαιτήσεις και φάρμακα), δεδομένα ηλεκτρονικού φακέλου υγείας και δεδομένα από περιφερειακά εργαστηριακά συστήματα πληροφοριών. Οι πληροφορίες αυτές συμβάλλουν σημαντικά στη βελτίωση της παροχής υγειονομικής περίθαλψης και στην κατανόηση των αποτελεσμάτων των ασθενειών.

Πληροφορίες που προέρχονται από όλο το φάσμα των βιολογικών δεδομένων (συμπεριλαμβανομένων των δεδομένων γονιδιωματικής, πρωτεομικής και μεταβολομικής). Οι πληροφορίες αυτές είναι απαραίτητες για την κατανόηση των μηχανισμών των ασθενειών και την προώθηση της εξατομίκευσης των προσφερόμενων θεραπειών (Alguliyev & Hajirahimova, 2014).

Τα λεγόμενα βιομετρικά δεδομένα (πληροφορίες από συσκευές που φορούν οι ασθενείς ή πληροφορίες που συλλαμβάνονται από αισθητήρες) και τα δεδομένα από τα μέσα κοινωνικής δικτύωσης αποτελούν πρόσθετες πηγές πληροφοριών. Αυτά τα

στατιστικά στοιχεία προσφέρουν διορατικές λεπτομέρειες σχετικά με τις συνήθειες και τη συμπεριφορά του ατόμου. Τα δεδομένα στον κλάδο της υγειονομικής περίθαλψης, σύμφωνα με τους Belle et al, είναι διασκορπισμένα μεταξύ διαφόρων συστημάτων υγειονομικής περίθαλψης, ασφαλιστικών φορέων, ακαδημαϊκών ερευνητών και κυβερνητικών οργανισμών. Τα μεγάλα δεδομένα στην ιατρική ακριβείας προέρχονται από τέσσερις κύριους ενδιαφερόμενους φορείς, σύμφωνα με τη μελέτη των Huang και συν.:

- i. Κυβέρνηση και μεγάλες επιχειρήσεις,
- ii. Μικρότερους φορείς όπως ακαδημαϊκές ομάδες και νεοφυείς επιχειρήσεις (start-ups) στους τομείς της τεχνολογίας, βιοτεχνολογίας και των ιατροτεχνολογικών συσκευών,
- iii. Αγοραστές και παρόχους υπηρεσιών υγείας, και
- iv. Μη κερδοσκοπικούς οργανισμούς, ομάδες υπεράσπισης ασθενών (Alguliyev & Hajirahimova, 2014).

3. Τομέας Υγείας & Big Data Analytics

3.1 Ορισμός & Ταξινόμηση

"Η πληροφορία είναι το καύσιμο του 21ου αιώνα και η ανάλυσή της είναι ο κινητήρας", δήλωσε ο Peter Sondergaard. Ο μόνος τρόπος με τον οποίο τα δεδομένα που παράγονται από άτομα και αποθηκεύονται σε διάφορα αποθετήρια από πολλούς οργανισμούς μπορούν να έχουν αντίκτυπο είναι αν αναλυθούν και εφαρμοστούν σωστά. Για να το θέσουμε αλλιώς, τα δεδομένα θα χρησιμεύουν απλώς ως πόρος για τον οργανισμό χωρίς να αναλύονται σωστά. Θα πρέπει να τονιστεί ότι ο όρος "Big Data" σχετίζεται επίσης με την ισχύ των δεδομένων και όχι μόνο με την ποσότητά τους. Η πολυπλοκότητα και το μέγεθος των συνόλων δεδομένων καθιστούν εξαιρετικά δύσκολη για τις τρέχουσες μεθόδους την ανάλυσή τους, την καταγραφή των ευρημάτων και τον εντοπισμό πιθανών συνδέσεων (Alguliyev & Hajirahimova, 2014).

Η φράση "ανάλυση μεγάλων δεδομένων" ή "big data analytics" είναι μια φράση που χρησιμοποιείται συχνά σήμερα. Πρόκειται για μια μέθοδο μετατροπής μη επεξεργασμένων δεδομένων σε γνώση. Η φράση αυτή αναφέρεται κυρίως στην απαίτηση για την εξαγωγή διεισδυτικών συμπερασμάτων και προβλέψεων από τα δεδομένα, προκειμένου να υποστηριχθεί η λήψη αποφάσεων αφενός και να ενισχυθούν οι οργανωτικές λειτουργίες αφετέρου. Η ανάλυση μεγάλων δεδομένων καταβάλλει προσπάθειες για την αποκάλυψη κρυμμένων μοτίβων, την εύρεση ανεξερεύνητων συσχετίσεων, την κατανόηση των τάσεων της αγοράς, των προτιμήσεων των πελατών και την άντληση άλλων ωφέλιμων επιχειρηματικών πληροφοριών, προκειμένου να ενισχυθεί η λήψη αποφάσεων στον ταχέως αναπτυσσόμενο επιχειρηματικό τομέα. Η συστηματική χρήση ιατρικών δεδομένων και σχετικών πληροφοριών διαχείρισης μέσω της εφαρμογής αναλυτικών μεθόδων και εργαλείων, καθώς και ποσοτικών και ποιοτικών στατιστικών στοιχείων, ανάλυσης πλαισίου και πρόβλεψης για την εξαγωγή συμπερασμάτων που μπορούν να αποτελέσουν τη βάση για τη δράση και τη δημιουργία μιας στρατηγικής και επιχειρησιακής διαχείρισης με βάση την πληροφορία για καλύτερη υγειονομική περίθαλψη. Αυτό είναι γνωστό ως ανάλυση δεδομένων στην υγειονομική περίθαλψη ή ανάλυση υγείας (Berger, 2014).

Επιπλέον, ο όρος "Health Analytics" ή "Data Analytics in Healthcare" αναφέρεται σε ένα ευρύ φάσμα εφαρμογών επιχειρηματικής ευφυΐας και ανάλυσης μεγάλων δεδομένων. Μια ευρεία ποικιλία συστημάτων πληροφοριών υγείας,

συμπεριλαμβανομένων των ηλεκτρονικών ιατρικών φακέλων, των συστημάτων αρχειοθέτησης και επικοινωνίας εικόνων, των εργαστηριακών συστημάτων πληροφοριών και των συστημάτων αποθήκευσης δεδομένων υγειονομικής περίθαλψης back-end, έχουν ενσωματωθεί καλά και είναι διαλειτουργικά, επιτρέποντας τη διαθεσιμότητα και την προσβασιμότητα των δεδομένων και των πληροφοριών. Οι ερευνητές της Frost & Sullivan προώθησαν την ιδέα της προηγμένης ανάλυσης υγείας κινούμενοι προς την ίδια κατεύθυνση (Berger, 2014).

Λόγω της αυξανόμενης υιοθέτησης τεχνολογιών αισθητήρων για έξυπνα τηλέφωνα και φορητά συστήματα, ο αριθμός των πηγών δεδομένων στον κλάδο της υγείας έχει πρόσφατα επεκταθεί γρήγορα. Ως αποτέλεσμα, είναι πρόκληση η ανάλυση αυτών των δεδομένων με τη χρήση συμβατικών αναλυτικών τεχνικών, οι οποίες θεωρούνται ανεπαρκείς για την αντιμετώπιση του όγκου των ποικίλων ιατρικών δεδομένων. Υπό αυτό το πρίσμα, τα ιδρύματα υγειονομικής περίθαλψης εργάζονται για την ορθή αξιοποίηση του τεράστιου όγκου δεδομένων και πληροφοριών που έχουν αποθηκεύσει, χρησιμοποιώντας την ανάλυση μεγάλων δεδομένων (Big Data Analytics). Η ενσωμάτωση ετερογενών δεδομένων, η διασφάλιση της ποιότητας των δεδομένων, η ανάλυση, η μοντελοποίηση, η ερμηνεία και η επαλήθευση περιλαμβάνονται στην ανάλυση μεγάλων δεδομένων. Η χρήση τους στην ιατρική και την υγειονομική περίθαλψη γενικότερα, ειδικότερα, επιτρέπει τη μελέτη μεγάλων συνόλων δεδομένων από εκατοντάδες ασθενείς, τον εντοπισμό συστάδων και τη συσχέτιση συνόλων δεδομένων, καθώς και τη δημιουργία μοντέλων πρόβλεψης με τη χρήση τεχνικών εξόρυξης δεδομένων. Η βιοπληροφορική, η ιατρική απεικόνιση, η υπολογιστική αισθητήρων, η ιατρική πληροφορική και η πληροφορική υγείας είναι μερικοί μόνο από τους κλάδους που αναλύονται στο πλαίσιο της ανάλυσης μεγάλων δεδομένων στους τομείς της ιατρικής και της υγειονομικής περίθαλψης. Οι προσεγγίσεις της ανάλυσης μεγάλων δεδομένων προβλέπεται να παράγουν νέα γνώση που θα είναι εξαιρετικά επωφελής για τους ασθενείς, τους γιατρούς και τους υπεύθυνους χάραξης πολιτικής για την υγεία (Berger, 2014).

Λόγω της κλίμακας του αντίκτυπου της διαχείρισης με βάση τις πληροφορίες στην επίλυση προβλημάτων και στη λήψη αποφάσεων, η ανάλυση δεδομένων στην υγειονομική περίθαλψη έχει γίνει ένα κρίσιμο πεδίο μελέτης τόσο για τους ερευνητές όσο και για τους επαγγελματίες της υγειονομικής περίθαλψης κατά τη διάρκεια των δύο τελευταίων δεκαετιών. Η υιοθέτηση των πληροφοριακών συστημάτων υγείας επιταχύνεται παγκοσμίως, γεγονός που θα διευρύνει σημαντικά την ποσότητα και θα

αναβαθμίσει το επίπεδο των διαθέσιμων δεδομένων υγείας. Ταυτόχρονα, οι πρωτοποριακές εξελίξεις στις αναλυτικές τεχνικές επέτρεψαν την ανάλυση τεράστιου όγκου δεδομένων και την ανακάλυψη νέων γνώσεων. Ως αποτέλεσμα, υπάρχουν τώρα περισσότερες ευκαιρίες από ποτέ άλλοτε για την εφαρμογή αυτών των τεχνικών για την αύξηση του επιπέδου της θεραπείας και τη μείωση του κόστους της.

Η περιγραφική (Descriptive Analytics), η διαγνωστική (Diagnostic Analytics), η προγνωστική (Predictive Analytics) και η περιγραφική (Prescriptive Analytics) είναι οι τέσσερις τύποι ανάλυσης μεγάλων δεδομένων που χρησιμοποιούνται στην υγειονομική περίθαλψη (Guided Analytics) (Berger, 2014).

3.1.1 Περιγραφική Αναλυτική

Με τη χρήση ιστορικών δεδομένων, επιχειρηματικής ευφυΐας και εξόρυξης δεδομένων, η περιγραφική ανάλυση βοηθά στη δημιουργία μιας εικόνας των επιδόσεων του παρελθόντος και περιγράφει την τρέχουσα κατάσταση. Η χρήση αυτού του αναλυτικού στοιχείου θα χρησιμεύσει ως ακρογωνιαίος λίθος για την ανάπτυξη της καλύτερης στρατηγικής για την επίτευξη του στόχου. Για τη διεξαγωγή αυτού του επιπέδου ανάλυσης χρησιμοποιούνται διάφορες στρατηγικές (Berger, 2014).

Η μάθηση χωρίς επίβλεψη είναι ένα άλλο όνομα για την περιγραφική ανάλυση. Συγκεκριμένα συνοψίζει:

- Τι συνέβει στη διοίκηση των υπηρεσιών υγείας;
- Ποιος είναι ο αντίκτυπος μιας παραμέτρου στο σύστημα;

Το πιο εύκολο επίπεδο κατανόησης και εφαρμογής είναι η περιγραφική ανάλυση. Ελέγχει πλήρως τα δεδομένα και τα περιγράφει μόνο, χωρίς περαιτέρω διερεύνηση, ανάλυση ή συσχετίσεις των μεταβλητών ή των πληροφοριών. Ο στόχος της περιγραφικής ανάλυσης είναι να βοηθήσει τους επαγγελματίες του τομέα της υγείας να κατανοήσουν και να αναλύσουν τις αποφάσεις, τις επιδόσεις και τα αποτελέσματα με την κατηγοριοποίηση, τον χαρακτηρισμό, τη συγκέντρωση και την ταξινόμηση των δεδομένων. Απλά γραφήματα και πίνακες χρησιμοποιούνται συνήθως για την απεικόνιση στατιστικών στοιχείων, υποδεικνύοντας πράγματα όπως τα ποσοστά πληρότητας των νοσοκομείων, τις εξόδους, τη μέση διάρκεια παραμονής και άλλους σχετικούς δείκτες.

Η οπτικοποίηση δεδομένων χρησιμοποιείται για την παροχή μιας πιο ολοκληρωμένης εικόνας της τεκμηριωμένης κλινικής πρακτικής, βοηθώντας στην ανακάλυψη μοτίβων περίθαλψης ή παρέχοντας απαντήσεις σε συγκεκριμένα ερωτήματα. Επιτρέπουν τον χειρισμό δεδομένων σε πραγματικό ή σχεδόν πραγματικό χρόνο, ενώ παράλληλα αποκτούν οπτικές πληροφορίες για κάθε ασθενή ή ηλεκτρονικό ιατρικό φάκελο. Αυτή η ικανότητα καθιστά δυνατό τον εντοπισμό μοτίβων ασθενών που συνδέονται με επανεισαγωγές σε νοσοκομεία και δεν είχαν εντοπιστεί προηγουμένως (Berger, 2014).

3.1.2 Διαγνωστική Αναλυτική

Η διάγνωση και η πρωταρχική αιτία του προβλήματος προβλέπονται με τη χρήση δεδομένων του παρελθόντος. Η διαγνωστική ανάλυση επιδιώκει να εντοπίσει τις αιτίες συγκεκριμένων περιστατικών και τους λόγους που συνέβησαν. Για παράδειγμα, η διαγνωστική ανάλυση χρησιμοποιεί πολυάριθμες τεχνικές, όπως η ομαδοποίηση και τα δέντρα αποφάσεων, για να προσπαθήσει να κατανοήσει τις αιτίες των επαναλαμβανόμενων επανεισαγωγών ορισμένων ασθενών. Για τον εντοπισμό της προέλευσης ενός προβλήματος και τη βοήθεια των χρηστών στην κατανόηση της φύσης και του αντίκτυπου των προβλημάτων, απαιτείται εκτεταμένη εξέταση και καθοδηγούμενη ανάλυση των υφιστάμενων δεδομένων με τη χρήση εργαλείων όπως οι τεχνικές απεικόνισης. Η κατανόηση του τρόπου με τον οποίο οι εισροές και οι διαδικασίες του συστήματος επηρεάζουν την απόδοση μπορεί να αποτελέσει μέρος αυτού. Για παράδειγμα, οι μεγαλύτεροι χρόνοι αναμονής για την παροχή ορισμένων υπηρεσιών υγειονομικής περίθαλψης μπορεί να σχετίζονται με διάφορους σημαντικούς παράγοντες, συμπεριλαμβανομένων εκείνων που σχετίζονται με τον ασθενή, τον πάροχο ή τον οργανισμό (Bachhety et al., 2021).

3.1.3 Προγνωστική Αναλυτική

Δεδομένου ότι δίνει έμφαση στη χρήση πληροφοριών και όχι μόνο στατιστικών στοιχείων, η προγνωστική ανάλυση λειτουργεί με πιο σύνθετο τρόπο από

την περιγραφική ανάλυση. Για την πρόβλεψη μελλοντικών αποτελεσμάτων, χρησιμοποιεί ιστορικούς δείκτες και τρέχουσες κρίσεις. Επιδεικνύει την ικανότητα να κάνει προβλέψεις για το μέλλον, ενώ παράλληλα βοηθά στην ανίχνευση τάσεων και πιθανών διαφορούμενων συνεπειών. Για παράδειγμα, μπορεί να ζητηθεί να προβλέψει εάν ένας ασθενής θα αντιμετωπίσει δυσκολίες ή όχι. Οι μέθοδοι μηχανικής μάθησης χρησιμοποιούνται συχνά για την κατασκευή μοντέλων πρόβλεψης. Σε σύγκριση με τις παραδοσιακές τεχνικές της εταιρείας, η προγνωστική ανάλυση παράγει καλύτερα αποτελέσματα αξιοποιώντας τα τεράστια σύνολα δεδομένων για τη βελτίωση της εμπειρίας των καταναλωτών. Χρησιμοποιούνται για την αξιολόγηση τεράστιων ποσοτήτων δεδομένων, συμπεριλαμβανομένων των μη δομημένων δεδομένων, και παράγουν πληροφορίες που μπορούν να αξιοποιηθούν για την πρόβλεψη μελλοντικών γεγονότων (Bachhety et al., 2021).

Από τη σκοπιά της επιστήμης των υπολογιστών, η πρόβλεψη μελλοντικών εξελίξεων με βάση τα παρόντα σύνολα δεδομένων είναι ένα δύσκολο ζήτημα. Αυτού του είδους τα λογισμικά επιχειρηματικής ευφυΐας βοηθούν στον υπολογισμό ροών δεδομένων σε ευρύτερη κλίμακα, όπως το περιεχόμενο των μέσων κοινωνικής δικτύωσης, η αγοραστική συμπεριφορά, οι τακτικές ενέργειες των χρηστών και οι έρευνες. Αναφέρονται επίσης ως επιβλεπόμενη μάθηση και εξετάζουν τόσο ιστορικά δεδομένα όσο και δεδομένα πραγματικού χρόνου. Επειδή όλα τα πιθανά μελλοντικά γεγονότα είναι πιθανά, μπορούν μόνο να υποθέσουν τι μπορεί να συμβεί- δεν μπορούν να προβλέψουν το μέλλον (Bachhety et al., 2021).

Ειδικότερα, επιχειρούν να δώσουν απαντήσεις στα ακόλουθα ερωτήματα:

- Τι θα συμβεί; Ποιες είναι οι μελλοντικές τάσεις;
- Ποια απόφαση θα ληφθεί βάσει του προηγούμενου ιστορικού;

Για να προετοιμαστεί για μια επιδημία, ένας φαρμακοποιός, για παράδειγμα, μπορεί να χρειάζεται να γνωρίζει πόση ποσότητα ενός φαρμακευτικού σκευάσματος πρέπει να αποθηκεύσει. Με βάση τις τεράστιες ποσότητες δεδομένων που έχουν συγκεντρωθεί κατά την προηγούμενη περίοδο, θα μπορούσαν να προβλεφθούν και να αξιολογηθούν ορισμένες εξελίξεις ή κλινικές εκβάσεις ασθενών, όπως η διάρκεια της παραμονής ενός ασθενούς στο νοσοκομείο, οι ασθενείς που μπορεί να επιλέξουν να υποβληθούν σε χειρουργική επέμβαση, οι ασθενείς που είναι απίθανο να ωφεληθούν

από μια τέτοια επέμβαση ή να παρουσιάσουν επιπλοκές, ή ακόμη και η θνησιμότητα (Bachhety et al., 2021).

3.1.4 Καθοδηγητική Αναλυτική

Το Guiding Analytics-Καθοδηγητική Αναλυτική αναλαμβάνει δράση όταν πρέπει να ληφθούν αποφάσεις σχετικά με ένα ευρύ φάσμα πρακτικών εναλλακτικών λύσεων, δίνοντας τη δυνατότητα στους υπεύθυνους λήψης αποφάσεων σε έναν οργανισμό όχι μόνο να σκεφτούν τις επιπτώσεις και τα αναμενόμενα αποτελέσματα των αποφάσεων, αλλά και να διαγνώσουν νέες ευκαιρίες ή ζητήματα και να προτείνουν την καλύτερη δυνατή πορεία δράσης προκειμένου να αξιοποιηθεί έγκαιρα η παρεχόμενη ανάλυση. Πριν από τη λήψη μιας απόφασης, αυτή η αναλυτική τεχνική συνθέτει αυτόματα τα μεγάλα δεδομένα και προσφέρει συμβουλές για ένα ευρύ φάσμα πιθανών αποτελεσμάτων. Ο υπεύθυνος λήψης αποφάσεων μπορεί να χρησιμοποιήσει αυτές τις πληροφορίες ως καθοδήγηση και να αναλάβει την κατάλληλη δράση (Bachhety et al., 2021).

Η Καθοδηγητική Αναλυτική παρέχει καθοδήγηση ως προς:

- Τι θα πρέπει να κάνουμε;
- Ποιο είναι το καλύτερο αποτέλεσμα και πως μπορούμε να το πετύχουμε;

Η χρήση υβριδικών δεδομένων, συμπεριλαμβανομένων τόσο δομημένων όσο και μη δομημένων δεδομένων, η ενσωμάτωση της πρόβλεψης και της καθοδηγητικής ανάλυσης με ταυτόχρονη συνεκτίμηση όλων των πιθανών παρενεργειών, η χρήση προσαρμοστικών αλγορίθμων που προσαρμόζονται εύκολα σε κάθε κατάσταση και η χρήση ισχυρών και αξιόπιστων μηχανισμών ανατροφοδότησης είναι οι πέντε βασικές παράμετροι που καθορίζουν την επιτυχία της καθοδηγούμενης ανάλυσης (Bachhety et al., 2021).

3.2 Η επιρροή & οι εφαρμογές στον Τομέα της Υγείας

Οι αναλύσεις μεγάλων δεδομένων έχουν τη δύναμη να αλλάξουν ριζικά τα τρέχοντα επιχειρηματικά και κλινικά μοντέλα για την παροχή ευφούς και αποτελεσματικής υγειονομικής περίθαλψης. Προκειμένου να διευκολυνθούν οι δευτερογενείς χρήσεις των δεδομένων, επιτρέπουν την ενσωμάτωση των αποπροσδιορισμένων πληροφοριών υγείας. Επιπλέον, μπορούν να βοηθήσουν στην αυτόνομη λήψη αποφάσεων με τον εντοπισμό μοτίβων και την κατανόηση των σχέσεων. Οι αναλύσεις μεγάλων δεδομένων μπορούν να βοηθήσουν στην έγκαιρη ανίχνευση της νόσου, στην ακριβή πρόβλεψη της πορείας της, στον εντοπισμό αποκλίσεων από μια υγιή κατάσταση, στην παρουσίαση επιπλοκών, καθώς και στην ανίχνευση της απάτης στην κλινική πρακτική. Βοηθούν τις εταιρείες υγειονομικής περίθαλψης στην εξατομίκευση των προβλέψεων, στην προσφορά εστιασμένης φροντίδας λαμβάνοντας υπόψη τη σχέση κόστους-αποτελεσματικότητας της περίθαλψης, στη μείωση της σπατάλης πόρων και στην ενθάρρυνση των ανθρώπων να διατηρήσουν την υγεία τους, κάνοντας σχετικές συστάσεις (Bachhety et al., 2021).

Οι αναλύσεις μεγάλων δεδομένων δίνουν στους ανθρώπους την ευκαιρία να βρουν γεγονότα που συμβαίνουν σπάνια, αλλά θα μπορούσαν ωστόσο να έχουν μεγάλο θεραπευτικό αντίκτυπο. Επιπλέον, η ομοιογένεια των κλινικών δεδομένων και η αποτελεσματική εφαρμογή τους υποστηρίζουν ένα ευρύ φάσμα εφαρμογών, όπως η παρακολούθηση ασθενειών, τα συστήματα υποστήριξης κλινικών αποφάσεων, η διαχείριση της προσωπικής υγειονομικής περίθαλψης, η ενίσχυση της αποτελεσματικότητας και της ποιότητας της υγειονομικής περίθαλψης και η μείωση του κόστους της υγειονομικής περίθαλψης. Σύμφωνα με την έρευνα των Sukumar και συν., η χρήση της ανάλυσης μεγάλων δεδομένων στην υγειονομική περίθαλψη μπορεί να δώσει απαντήσεις σε 8 σημαντικά προβλήματα:

1. Πώς μπορεί να αυξηθεί το κόστος για διάφορες πτυχές της υγειονομικής περίθαλψης στο μέλλον;
2. Πως επηρεάζουν ορισμένες αλλαγές σε όρους πολιτικής το κόστος και την συμπεριφορά;
3. Πως διαφοροποιούνται τα κόστη της υγειονομικής περίθαλψης ανά γεωγραφική περιοχή;
4. Μπορούν να εντοπιστούν ψευδείς απαιτήσεις;
5. Ποιες επιλογές θεραπείας φαίνονται πιο αποτελεσματικές για διάφορες ασθένειες;

6. Γιατί φαίνεται ότι κάποιοι πάροχοι έχουν καλύτερα αποτελέσματα για την υγεία;
7. Γιατί οι ασθενείς επιλέγουν έναν πάροχο έναντι άλλου;
8. Υπάρχουν ενδείξεις κάποιας επιδημίας; (Bachhety et al., 2021).

Οι τομείς στους οποίους η ανάλυση δεδομένων αναμένεται να επιφέρει τα καλύτερα αποτελέσματα είναι: ο εντοπισμός των ασθενών που χρησιμοποιούν συχνότερα υπηρεσίες υγείας ή έχουν περισσότερες πιθανότητες να εμφανίσουν αρνητικά αποτελέσματα- η παροχή στους ασθενείς των πληροφοριών που χρειάζονται για να λαμβάνουν τεκμηριωμένες αποφάσεις και, ως εκ τούτου, να διαχειρίζονται καλύτερα την υγεία τους- η απλούστευση της υιοθέτησης και παρακολούθησης υγιεινών συμπεριφορών- και ο εντοπισμός θεραπειών, προγραμμάτων και διαδικασιών που δεν αποδίδουν.

Ο τρόπος διαχείρισης αυτού του τεράστιου όγκου δεδομένων είναι η κύρια δυσκολία των μεγάλων δεδομένων, όπως έχει ήδη διαπιστωθεί. Τα δεδομένα πρέπει να αποθηκεύονται σε μορφή αρχείου που να είναι εύκολα προσβάσιμη και κατανοητή, ώστε να αναλαμβάνεται αποτελεσματική ανάλυση και να διατίθενται στην επιστημονική κοινότητα. Η χρήση προηγμένων εργαλείων, μεθόδων και τεχνολογιών πληροφορικής στο πλαίσιο της υγειονομικής περίθαλψης αποτελεί μια από τις μεγαλύτερες προκλήσεις του κλάδου. Και για να επιτευχθεί αυτό, θα πρέπει να συνεργαστούν επιστήμονες από διάφορα πεδία -όπως η βιολογία, η πληροφορική, η στατιστική και τα μαθηματικά-. Μια άλλη δυσκολία στην ανάλυση μεγάλων δεδομένων είναι η ετερογένεια των δεδομένων. Είναι πιο δύσκολο να φτάσει κανείς σε αποτελέσματα όταν χρησιμοποιεί παραδοσιακές τεχνολογίες ανάλυσης λόγω του τεράστιου όγκου και της πολύ διαφορετικής φύσης των μεγάλων δεδομένων στον τομέα της υγείας. Οι συστάδες υψηλής υπολογιστικής ισχύος που συνδέονται με υποδομές υπολογιστικού πλέγματος είναι οι πιο δημοφιλείς πλατφόρμες για την εκτέλεση λογισμικού που επιτρέπει την ανάλυση μεγάλων δεδομένων (Belle et al., 2015).

Ένα τέτοιο σύστημα είναι το νέφος, το οποίο προσφέρει αξιόπιστες υπηρεσίες και τεχνολογίες εικονικής αποθήκευσης. Παρέχει δυναμική ανακάλυψη και διαθεσιμότητα πόρων μαζί με υψηλή αξιοπιστία, επεκτασιμότητα και αυτονομία. Αυτές οι πλατφόρμες μπορούν να χρησιμεύσουν ως δέκτες δεδομένων από ευρέως διασκορπισμένους αισθητήρες, ως υπολογιστές που αναλύουν και ερμηνεύουν τα

δεδομένα και ως πηγές οπτικοποίησης δεδομένων για τον χρήστη. Επιπλέον, χρησιμοποιώντας υπηρεσίες κινητού υπολογισμού και υπολογισμού ομίχλης, η επεξεργασία και η ανάλυση μεγάλων δεδομένων μπορεί να γίνει πιο κοντά στην αρχική πηγή των δεδομένων στο λεγόμενο Διαδίκτυο των πραγμάτων. Για να εξεταστούν οι συστάδες διακομιστών που κατέχουν Big Data, απαιτούνται εξελιγμένοι αλγόριθμοι για την εφαρμογή τεχνικών μηχανικής μάθησης και τεχνητής νοημοσύνης. Τέτοιοι αλγόριθμοι ή λογισμικό θα μπορούσαν να δημιουργηθούν χρησιμοποιώντας μια γλώσσα προγραμματισμού (όπως η Python, η R ή μια άλλη γλώσσα) κατάλληλη για την επεξεργασία μεγάλων δεδομένων. Επομένως, η διαχείριση των μεγάλων δεδομένων από τη βιοϊατρική έρευνα απαιτεί μια καλή κατανόηση της βιολογίας και της επιστήμης των υπολογιστών. Το Hadoop και το Apache Spark είναι οι δύο πιο συχνά χρησιμοποιούμενες τεχνολογίες για την επεξεργασία μεγάλων δεδομένων.

Apache Hadoop

Τα μεγάλα δεδομένα δεν μπορούν να διαχειριστούν αποτελεσματικά με την απλή φόρτωση τεράστιων ποσοτήτων (μεγάλων) δεδομένων στη μνήμη ακόμη και των πιο ισχυρών υπολογιστικών συστημάτων. Ως εκ τούτου, η κατανομή και η παράλληλη επεξεργασία μεγάλων ποσοτήτων περίπλοκων μεγάλων δεδομένων σε διάφορους υπολογιστικούς κόμβους είναι η πιο φυσική στρατηγική για τον σκοπό αυτό. Για να επεξεργαστούν τα δεδομένα σε αποδεκτό χρονικό διάστημα, πρέπει να κατανεμηθούν μεταξύ τους χιλιάδες υπολογιστικές συσκευές λόγω του συχνά τεράστιου μεγέθους των δεδομένων. Πρέπει να αντιμετωπίσει κανείς προβληματισμούς όπως το πώς θα επιτύχει τον παραλληλισμό ή την παράλληλη επεξεργασία, την κατανομή των δεδομένων και τον τρόπο χειρισμού πιθανών σφαλμάτων όταν εργάζεται με εκατοντάδες ή χιλιάδες κόμβους. Το Apache Hadoop είναι ένα από τα πιο γνωστά προγράμματα ανοικτού κώδικα που διατίθενται για αυτή τη χρήση (Belle et al., 2015).

Το Apache Hadoop ειδικότερα έχει τη δυνατότητα να επεξεργάζεται εξαιρετικά τεράστιους όγκους δεδομένων, κυρίως μέσω της κατανομής της επεξεργασίας τεράστιων συνόλων δεδομένων σε διάφορους διακομιστές (κόμβους), καθένας από τους οποίους επιλύει ορισμένες πτυχές του συνολικού προβλήματος και στη συνέχεια τις ενσωματώνει για την εξαγωγή της τελικής εξόδου. Οι εφαρμογές που χρησιμοποιούν το Hadoop μπορούν να αναλύσουν τεράστιες ποσότητες δεδομένων, είτε αυτά είναι δομημένα είτε όχι. Ο αλγόριθμος MapReduce, μια τεχνική

προγραμματισμού για το χειρισμό και την παραγωγή μαζικών συνόλων δεδομένων, υλοποιείται από το Hadoop. Οι Map και Reduce είναι οι δύο λειτουργίες που συνθέτουν το MapReduce. Η Map είναι μια λειτουργία χαρτογράφησης που έχει ως αποστολή την ταξινόμηση και την οργάνωση των δεδομένων σε διακριτούς κόμβους. Μια συνάρτηση μείωσης που ονομάζεται Reduce είναι κυρίως υπεύθυνη για την εκτέλεση των πράξεων καταμέτρησης που συμβαίνουν σε κάθε ξεχωριστό κόμβο. Επιπλέον, οργανώνει την επικοινωνία μεταξύ μαζικών υπολογιστικών συστάδων και μιμείται τις αποτυχίες των υπολογιστικών χειρισμών. Το στοιχείο του συστήματος που επιτρέπει την κλιμακούμενη και αποτελεσματική αποθήκευση δεδομένων σε πολλούς κόμβους συστάδων ονομάζεται Hadoop File Distribution System (HDFS), το οποίο αποτελεί το τμήμα αποθήκευσης των κατανεμημένων δεδομένων (Belle et al., 2015).

Πολλές τεράστιες επιχειρήσεις, όπως η Yahoo, το Facebook και άλλες, έσπευσαν να υιοθετήσουν το Hadoop επειδή προσφέρει πρόσθετα εργαλεία που βελτιώνουν τα τμήματα αποθήκευσης και επεξεργασίας. Οι ερευνητές μπορούν να χρησιμοποιούν σύνολα δεδομένων που διαφορετικά δεν θα μπορούσαν να χειριστούν χάρη στο Hadoop. Η πλατφόρμα Hadoop χρησιμοποιείται για την εκτέλεση πολυάριθμων μεγάλων εργασιών, όπως η εύρεση του κατά πόσον υπάρχει σχέση μεταξύ των δεδομένων για την ποιότητα του αέρα και των εισαγωγών σε νοσοκομεία για άσθμα, η ανακάλυψη φαρμάκων με τη χρήση γενετικών και πρωτεομικών δεδομένων και άλλα παρόμοια στοιχεία της υγειονομικής περίθαλψης (Burghard, 2012).

MongoDB

Μια από τις πιο δημοφιλείς βάσεις δεδομένων, η MongoDB, η οποία είναι επίσης ανοικτού κώδικα και βασίζεται σε έγγραφα³⁸, έχει δημιουργηθεί και διατίθεται στην αγορά από την 10Gen. Για την αποθήκευση αρχείων, η MongoDB βασίζεται στη μορφή JSON³⁹. Σε αντίθεση με τις συμβατικές μορφές, όπως οι γραμμές και οι πίνακες, έχει μια ανοικτή μορφή που είναι αναγνώσιμη τόσο από τη μηχανή όσο και από τον άνθρωπο. Αυτό προωθεί την ανταλλαγή δεδομένων. Επιπλέον, επειδή οι σχετικές πληροφορίες για ένα συγκεκριμένο αρχείο περιέχονται σε ένα μόνο έγγραφο JSON, η κλιμάκωση της μορφής JSON είναι απλούστερη, επειδή δεν απαιτούνται ξεχωριστές αναζητήσεις. Η αυτόματη κλιμάκωση, η προσαρμοστικότητα, η απλότητα, η μεγάλη απόδοση και η διαθεσιμότητα είναι μερικά μόνο από τα πλεονεκτήματά του.

Επιπλέον, δίνει στον χρήστη πρόσβαση σε αναζήτηση κειμένου και διαθέτει σύνδεση με το Apache Hadoop. Όσον αφορά τη λειτουργία της MongoDB στον κλάδο της υγειονομικής περίθαλψης, προσφέρει πολυάριθμες απαντήσεις σε κρίσιμα προβλήματα. Για παράδειγμα, χρησιμοποιώντας τα υπάρχοντα εργαλεία εξόρυξης δεδομένων, δίνει τη δυνατότητα στον χειριστή να αναπτύξει ένα ολοκληρωμένο προφίλ για κάθε ασθενή που περιλαμβάνει όλες τις διαγνωστικές εξετάσεις στις οποίες έχει υποβληθεί, καθώς και να κάνει προβλέψεις για πιθανές σχέσεις. Επιπλέον, έχει τη δυνατότητα ενημέρωσης των προφίλ με την αλλαγή ή την προσθήκη δεδομένων, καθώς και την αλλαγή των συγκρίσεων με την πάροδο του χρόνου. Δεδομένου ότι προσφέρει τη δυνατότητα συσχέτισης δεδομένων, την οποία ένας γιατρός δεν θα ήταν σε θέση να εκτελέσει, ανεξάρτητα από το επίπεδο των δεξιοτήτων του, η MongoDB είναι επίσης ζωτικής σημασίας για την έγκαιρη ανακάλυψη και διάγνωση ασυνήθιστων ασθενειών. Επιπλέον, επιτρέπει τον έλεγχο της εξάπλωσης μιας επιδημικής έξαρσης σε πρώιμο στάδιο. Τέλος, μπορεί να προχωρήσει άμεσα στη διάγνωση μιας ασθένειας σε πραγματικό χρόνο (Burghard, 2012).

Άλλες NoSQL βάσεων δεδομένων

Η CouchBase είναι μια άλλη βάση δεδομένων NoSQL που χρησιμοποιεί μια ευέλικτη αρχιτεκτονική JSON για να επιτρέπει στους χρήστες να διαμορφώνουν τις εφαρμογές τους χωρίς να περιορίζονται από ένα συγκεκριμένο σχήμα βάσης δεδομένων. Οι υψηλές ταχύτητες ανάγνωσης και εγγραφής, καθώς και η απλή επεκτασιμότητα αποτελούν περαιτέρω πλεονεκτήματα. Επιπλέον, η CouchBase επιτρέπει την ταυτόχρονη εξυπηρέτηση πολλαπλών χρηστών, ενώ κατανέμει τα δεδομένα και το φορτίο εξίσου σε όλους τους προσβάσιμους διακομιστές. Επιπλέον, η βάση δεδομένων ανοικτού κώδικα Apache Cassandra, η οποία δημιουργήθηκε από την επιχείρηση DataStax, διατίθεται στην αγορά ως ανταγωνιστής της βάσης δεδομένων Oracle. Επιτρέπει τη διαχείριση σημαντικών ποσοτήτων ημιδομημένων, αδόμητων και δομημένων δεδομένων που φυλάσσονται σε διάφορα κέντρα δεδομένων ή ακόμη και στο νέφος. Τέλος, για τη συγχώνευσή του μπορεί να χρησιμοποιηθεί το Hadoop (Burghard, 2012).

Apache Spark

Μια άλλη επιλογή λογισμικού ανοικτού κώδικα για το Hadoop είναι το Apache Spark. Πρόκειται για μια ενιαία μηχανή για κατανεμημένη επεξεργασία δεδομένων με βιβλιοθήκες υψηλότερου επιπέδου για την υποστήριξη λειτουργιών όπως η χρήση της γλώσσας SQL για την υποβολή ερωτημάτων σε δεδομένα, η επεξεργασία δεδομένων ροής, η μηχανική μάθηση και η επεξεργασία γραφημάτων, μεταξύ άλλων (GraphX). Επειδή η διεπαφή προγραμματισμού μπορεί να ενσωματωθεί εύκολα και απαιτεί λιγότερη εργασία για τον κώδικα, αυτές οι βιβλιοθήκες βοηθούν στην ενίσχυση της παραγωγικότητας των προγραμματιστών. Η επεξεργασία δεδομένων στη μνήμη υποβοηθείται από την ανάπτυξη του λεγόμενου Elastic Distributed Dataset (RDD42), το οποίο μπορεί να κάνει το Spark περίπου 100 φορές πιο γρήγορο από το Hadoop.

Όταν το μέγεθος των δεδομένων είναι μικρότερο από το μέγεθος της προσβάσιμης μνήμης, αυτό ισχύει περισσότερο. Αυτό απεικονίζει πόση μνήμη RAM θα χρειαζόταν για την επεξεργασία μεγάλων δεδομένων με τη χρήση του Apache Spark. Το MapReduce αναμένεται να είναι πιο αποδοτικό για μεγάλα σύνολα δεδομένων σε σύγκριση με το Apache Spark, επειδή το κόστος της μνήμης είναι υψηλότερο. Παρόμοια, το πρόγραμμα Apache Storm δημιουργήθηκε για να προσφέρει ένα πλαίσιο πραγματικού χρόνου για την επεξεργασία δεδομένων ροής. Οι περισσότερες γλώσσες προγραμματισμού υποστηρίζονται από αυτή την πλατφόρμα. Επιπλέον, διαθέτει ενσωματωμένη ανοχή σε σφάλματα και ισχυρή οριζόντια επεκτασιμότητα, καθιστώντας δυνατή τη μελέτη μεγάλων δεδομένων (Das et al., 2010).

3.3 Ανάλυση από Βιολογικά Δεδομένα

Οι επιστήμονες βιοπληροφορικής έχουν ένα εντελώς νέο είδος ζητήματος στην αντιμετώπιση μεγάλων δεδομένων από την έρευνα βιολογικών δεδομένων. Μια τέτοια περίπλοκη ανάλυση δεδομένων βιολογικών συστημάτων απαιτεί ισχυρούς αλγορίθμους. Ο τελικός στόχος είναι να μετατραπούν αυτές οι τεράστιες ποσότητες δεδομένων σε μια χρήσιμη βάση γνώσης. Η μεταφραστική βιοπληροφορική⁴⁴ είναι η εφαρμογή προσεγγίσεων βιοπληροφορικής για τη μετατροπή βιοϊατρικών και γενετικών δεδομένων σε μεθόδους πρόβλεψης και πρόληψης της υγείας. Όσον αφορά την υγειονομική

περίθαλψη με βάση τα δεδομένα, βρίσκεται στην πρώτη γραμμή. Για την εξεύρεση νέων μεταδεδομένων που μπορούν να υποστηρίξουν τα φάρμακα ακριβείας, μπορούν να συγχωνευθούν διάφοροι τύποι ποσοτικών δεδομένων στην υγειονομική περίθαλψη, όπως εργαστηριακές μετρήσεις, δεδομένα φαρμακευτικής αγωγής και γονιδιωματικά προφίλ (Das et al., 2010).

Συνεπώς, απαιτούνται αναδυόμενες νέες τεχνολογίες που θα βοηθήσουν στην ανάλυση αυτού του ψηφιακού πλούτου. Ακολουθούν ορισμένα εργαλεία για την ανάλυση μεγάλων δεδομένων βιολογικών δεδομένων:

1. SparkSeq: Αυτή η αποτελεσματική πλατφόρμα, που βασίζεται στο πλαίσιο Apache Spark και τη βιβλιοθήκη Hadoop, αξιοποιείται για την ακριβή ανάλυση νουκλεοτιδίων σε δυναμικά γονιδιωματικά δεδομένα.
2. SAMQA: ελέγχει την ακρίβεια και την ακεραιότητα τεράστιων ποσοτήτων γενετικών δεδομένων.
3. ART: Έχει τη δυνατότητα να μοντελοποιεί προφίλ σφαλμάτων ανάγνωσης και μήκη ανάγνωσης για δεδομένα που παράγονται από τεχνολογίες αλληλούχισης υψηλής απόδοσης, όπως τα συστήματα SOLiD και Illumina.
4. DistMap: Χρησιμοποιώντας υπολογιστικές συστάδες Hadoop, το DistMap είναι ένα πρόσθετο εργαλείο για κατανεμημένη χαρτογράφηση μικρού μήκους που επιδιώκει να αντιμετωπίσει μεγαλύτερη ποικιλία εφαρμογών αλληλουχίας. Ο χαρτογράφος BWA, για παράδειγμα, είναι μία από τις εφαρμογές του και μπορεί να επεξεργαστεί 500 εκατομμύρια ζεύγη αναγνώσεων σε περίπου έξι ώρες, δηλαδή περίπου 13 φορές ταχύτερα από έναν παραδοσιακό χαρτογράφο ενός κόμβου.
5. Το SeqWare είναι μια μηχανή ερωτημάτων που βασίζεται στο σύστημα βάσεων δεδομένων Apache HBase και ενσωματώνει προγράμματα περιήγησης γονιδιώματος και εργαλεία για την παροχή πρόσβασης σε εκτεταμένα δεδομένα σε όλο το γονιδίωμα.
6. Το CloudBurst είναι μια προσέγγιση υπολογιστικού παραλληλισμού που ενισχύει την επεκτασιμότητα της ανάγνωσης μεγάλων δεδομένων ακολουθιών σε έργα χαρτογράφησης γονιδιώματος.
7. Το BlueSNP, ένα πρόγραμμα R με βάση το Hadoop για μελέτες συσχέτισης ολόκληρου του γονιδιώματος (GWAS45), αποσκοπεί στη στατιστική ανάλυση συνόλων δεδομένων γονότυπων-φαινοτύπων και στην εύρεση στατιστικά σημαντικών συνδέσεων μεταξύ τους.

8. Myrna: μια πλατφόρμα βασισμένη στο cloud που χρησιμοποιεί ομαλοποίηση δεδομένων και στατιστική μοντελοποίηση για την παροχή πληροφοριών σχετικά με τις διακυμάνσεις στα επίπεδα γονιδιακής έκφρασης. (Das et al., 2010).

3.4 Εμπορικές εφαρμογές

Πολυάριθμες επιχειρήσεις έχουν χρησιμοποιήσει τεχνητή νοημοσύνη για την ανάλυση δημοσιευμένων αποτελεσμάτων, δεδομένων κειμένου και δεδομένων εικόνας, προκειμένου να αντιμετωπίσουν τις δυσκολίες των μεγάλων δεδομένων και να εκτελέσουν πιο αξιόπιστες αξιολογήσεις. Μία από τις μεγαλύτερες και πιο έμπειρες εταιρείες που προσφέρουν εμπορικές υπηρεσίες ανάλυσης στην υγειονομική περίθαλψη είναι η IBM Corporation. Μια πλατφόρμα τεχνητής νοημοσύνης από την IBM που ονομάζεται Watson Health επιτρέπει στα νοσοκομεία, τους επαγγελματίες υγείας και τους ερευνητές να μοιράζονται και να αναλύουν δεδομένα ασθενών. Παρόμοια με αυτό, η Flatiron Health προσφέρει ψηφιακές λύσεις για την ανάλυση της υγειονομικής περίθαλψης που στοχεύουν ιδιαίτερα στην έρευνα για τον καρκίνο. Άλλες σημαντικές επιχειρήσεις, όπως η Oracle Corporation και η Google Inc., επικεντρώνονται στη δημιουργία πλατφορμών για κατανεμημένη υπολογιστική ισχύ και αποθήκευση με βάση το νέφος. Είναι ενδιαφέρον να σημειωθεί ότι πολυάριθμες επιχειρήσεις και νεοφυείς επιχειρήσεις έχουν ξεκινήσει πρόσφατα πρωτοβουλίες που επικεντρώνονται στην προσφορά αναλύσεων και λύσεων για τον κλάδο της υγειονομικής περίθαλψης (Das et al., 2010).

AYASDI

Η Ayasdi είναι ένας σημαντικός προμηθευτής που επικεντρώνεται σε προσεγγίσεις που βασίζονται στη μηχανική μάθηση για να προσφέρει μια πλατφόρμα για τη μηχανική νοημοσύνη μαζί με ένα πλαίσιο εφαρμογών με αποδεδειγμένη λειτουργική επεκτασιμότητα. Προσφέρει μια ποικιλία εφαρμογών για την ανάλυση της υγειονομικής περίθαλψης, όπως αυτές για τη διαχείριση της κλινικής απόκλισης και την κατανόηση του τρόπου με τον οποίο αυτή επηρεάζει το κόστος της κλινικής θεραπείας. Μπορεί επίσης να αναλύσει και να ελέγξει τον τρόπο με τον οποίο έχουν

συσταθεί τα νοσοκομεία, τον τρόπο με τον οποίο αλληλεπιδρούν μεταξύ τους οι κλινικοί γιατροί, τον τρόπο με τον οποίο κάνουν επιλογές βάσει κινδύνου και τον τρόπο με τον οποίο θεραπεύουν τους ασθενείς. Επιπλέον, προσφέρει μια εφαρμογή αξιολόγησης και διαχείρισης της υγείας του πληθυσμού, μια προληπτική προσέγγιση που υπερβαίνει τις συμβατικές τεχνικές ανάλυσης κινδύνου (Dash et al., 2019).

Linguamatics

Πρόκειται για έναν αλγόριθμο που βασίζεται σε έναν αλγόριθμο για την εξαγωγή γνώσης από διαδραστικό κείμενο που χρησιμοποιεί επεξεργασία φυσικής γλώσσας (NLP) (I2E). Το I2E έχει τη δυνατότητα να εξάγει και να αναλύει μια ποικιλία δεδομένων. Δεδομένου ότι δεν απαιτούνται εξειδικευμένες γνώσεις για την κατανόηση των δεδομένων, τα αποτελέσματα που επιτυγχάνονται με τη χρήση αυτής της τεχνικής είναι δέκα φορές καλύτερα από εκείνα που επιτυγχάνονται με τη χρήση άλλων τεχνολογιών. Μέσω αυτής της μεθόδου, τα μη δομημένα δεδομένα μπορούν να χρησιμοποιηθούν για την αποκάλυψη πληροφοριών σχετικά με γενετικές συνδέσεις και γεγονότα. Όλοι γνωρίζουμε ότι η μάθηση χρειάζεται δεδομένα υψηλής ποιότητας προκειμένου να δημιουργηθούν σαφή, φιλτραρισμένα αποτελέσματα. Ωστόσο, η NLP επιτρέπει την εξαγωγή καθαρών και δομημένων πληροφοριών που συχνά παραμένουν κρυμμένες σε μη δομημένα δεδομένα εισόδου όταν αυτά ενσωματώνονται στον ΗΦΥ ή σε άλλα κλινικά αρχεία γενικότερα (Dash et al., 2019).

IBM Watson

Πρόκειται για μια εφαρμογή τεχνολογίας αιχμής από την IBM που επικεντρώνεται στην ανάλυση μεγάλων δεδομένων σε όλους σχεδόν τους κλάδους. Για την απόκτηση των περισσότερων πληροφοριών από τις λιγότερες εισροές, η πλατφόρμα αυτή κάνει σημαντική χρήση αλγορίθμων που βασίζονται στην τεχνητή νοημοσύνη και τη μηχανική μάθηση. Με τη βοήθεια του IBM Watson, ενός υπερυπολογιστή, μπορούν να παραχθούν σχετικά και δομημένα δεδομένα συνδυάζοντας διάφορους τομείς της υγειονομικής περίθαλψης. Το IBM Watson και η Pfizer συνεργάστηκαν με επιτυχία για την επιτάχυνση του εντοπισμού νέων συνδυασμών ανοσο-ογκολογίας σε μια προσπάθεια να βρεθούν νέα φάρμακα για συγκεκριμένα μοντέλα καρκινικών ασθενειών (Dash et al., 2019).

3.5 Ηλεκτρονική Υγεία & Οφέλη Big Data στην Υγεία

Η χρήση των δικτύων πληροφοριών και επικοινωνιών για σκοπούς που σχετίζονται με την υγεία με στόχο τη μείωση των ιατρικών σφαλμάτων, την αύξηση της ποιότητας των υπηρεσιών και τη μείωση του κόστους της υγειονομικής περίθαλψης είναι γνωστή ως ηλεκτρονική υγεία, σύμφωνα με τον Παγκόσμιο Οργανισμό Υγείας. Οι εγκαταστάσεις ηλεκτρονικής υγείας, σε συνεργασία με εταίρους σε παγκόσμιο, περιφερειακό και εθνικό επίπεδο, προωθούν και ενισχύουν τη χρήση αυτών των συστημάτων για τη βελτίωση της ποιότητας της υγείας, την παροχή ισότιμης πρόσβασης και μια σειρά άλλων χρήσεων. Οι κινητές εφαρμογές υγείας, που συνήθως αναφέρονται ως "κινητή εφαρμογή υγείας", έχουν ενταχθεί από τον Παγκόσμιο Οργανισμό Υγείας στο ευρύτερο θέμα της ηλεκτρονικής υγείας λόγω της ευρείας αποδοχής και της απλότητας χρήσης τους (Kaminskiy, 2017). Τα κύρια τεχνικά χαρακτηριστικά που επιτρέπουν σημαντικές αλλαγές με τη χρήση των πληροφοριακών και τηλεπικοινωνιακών συστημάτων περιλαμβάνουν: α) την ταυτόχρονη διαθεσιμότητα των ίδιων πληροφοριών σε διάφορα επίπεδα ασφαλείας, β) τη δυνατότητα άσκησης ελέγχου και συντονισμού από απομακρυσμένα μέρη, γ) την τεράστια υπολογιστική ικανότητα που επιτρέπει την ολοκλήρωση πολύπλοκων υπολογισμών σε ελάχιστο χρόνο με ελάχιστη πιθανότητα λάθους και δ) τη μεγάλη ικανότητα τόξου. Η ανάλυση μεγάλων δεδομένων μπορεί να χρησιμοποιηθεί στην υγειονομική περίθαλψη για τη βελτίωση της ποιότητας της θεραπείας, τη μείωση των συνολικών δαπανών και τη μείωση των σφαλμάτων και της σπατάλης πόρων (Kaminskiy, 2017).

α.Φαρμακοεπαγρύπνιση:

Μπορεί να αναλύσει κλινικές μελέτες, αρχεία ασθενών και άλλες αυτόκλητες πηγές και να αναγνωρίσει ανεπιθύμητες ενέργειες (Kaminskiy, 2017).

β.Βιοϊατρική τεχνολογία:

Η βιοϊατρική έρευνα και η έρευνα πληροφορικής στον τομέα της υγειονομικής περίθαλψης χρησιμοποιούν όλο και περισσότερο την τεχνολογία των μεγάλων δεδομένων. Παράγονται και συγκεντρώνονται πρωτοφανείς όγκοι βιολογικών και κλινικών δεδομένων με πρωτοφανείς ρυθμούς. Για παράδειγμα, οι πιο πρόσφατες τεχνολογίες αλληλουχίας DNA επιτρέπουν την επεξεργασία δισεκατομμυρίων αλληλουχιών DNA κάθε μέρα, και οι εφαρμογές ηλεκτρονικών αρχείων υγείας (EHR) συλλέγουν πολλά δεδομένα ασθενών.

Με τη βοήθεια τεχνολογικών εξελίξεων όπως η εμφάνιση νέων μηχανών αλληλουχίας, η ανάπτυξη εφευρετικού υλικού και λογισμικού για παράλληλους υπολογισμούς και η εκτεταμένη επέκταση των ηλεκτρονικών φακέλων υγείας, το κόστος συλλογής και ανάλυσης βιοϊατρικών δεδομένων αναμένεται να μειωθεί σημαντικά (EHR). Οι ακόλουθοι βιοϊατρικοί τομείς προσφέρουν νέες προοπτικές για εφαρμογές μεγάλων δεδομένων με σκοπό την παροχή νέων γνώσεων και την ανάπτυξη νέων τεχνικών για τη βελτίωση της ποιότητας της υγειονομικής περίθαλψης (Kaminskiy, 2017).

- (α) βιοπληροφορική,
- (β) κλινική πληροφορική,
- (γ) πληροφορική απεικόνισης, και
- (δ) πληροφορική δημόσιας υγείας.

γ. Υποστηρίζοντας την έρευνα -OMICS

Η βάση για την εξατομικευμένη περίθαλψη είναι η ανάλυση γονιδιώματος, η οποία είναι επίσης ο πιο υποσχόμενος επαγγελματίας της εξατομικευμένης ιατρικής. Επιπλέον, τα μεγάλα δεδομένα έχουν δημιουργήσει νέες προοπτικές για το πώς διαφέρουν οι άνθρωποι, ιδίως όταν πρόκειται για τις επιπτώσεις μιας ασθένειας ή μιας θεραπείας. Έχουμε πλέον στη διάθεσή μας εκτεταμένα μοριακά δεδομένα για την έρευνά μας λόγω των πολυάριθμων αναλύσεων γονιδιώματος. Επιτρέποντας την έγκαιρη διάγνωση, αυξάνοντας την αποτελεσματικότητα της θεραπείας ή επιτρέποντας την αλλαγή της δόσης με εξατομικευμένο τρόπο για τη μεγιστοποίηση της ανταπόκρισης και την ελαχιστοποίηση των παρενεργειών, η εξατομίκευση της θεραπείας συμβάλλει στη βελτίωση της υγείας. Τα μεγάλα δεδομένα μπορούν να χρησιμοποιηθούν για την ανάλυση μοτίβων ασθενειών και πολλοί οργανισμοί και

επιχειρήσεις επικεντρώνονται στα omics με την ελπίδα να προωθήσουν την έρευνα, να φέρουν επανάσταση στην ανάπτυξη θεραπειών και να αλλάξουν τις ιατρικές διαδικασίες. Οι πρωτοβουλίες Big Data to Knowledge (BD2K) και Precision Medicine από τα Εθνικά Ινστιτούτα Υγείας των ΗΠΑ αντικατοπτρίζουν αυτή την ανάγκη (NIH) (Fernandes et al., 2012).

δ. Μετατρέποντας δεδομένα σε πληροφορίες και αντίστροφα

Οι καλύτερες κλινικές και οικονομικότερες μέθοδοι για τη διάγνωση και τη θεραπεία των ασθενών καθοδηγούνται από την κλινική αξία των μεγάλων δεδομένων. Ο τρόπος κατάλληλης αξιοποίησης αυτών των δεδομένων αποτελεί κρίσιμη ανεκπλήρωτη ανάγκη, δεδομένου του τεράστιου όγκου δεδομένων στον κλάδο της υγειονομικής περίθαλψης. Ο μετασχηματισμός των δεδομένων σε αξιοποιήσιμες πληροφορίες αποτελεί κρίσιμο παράγοντα. Η υγειονομική περίθαλψη βάσει αποδείξεων απαιτεί τη μετατροπή των μη δομημένων δεδομένων σε δομημένα δεδομένα, η οποία αποτελεί ένα κρίσιμο πρώτο βήμα. (Ιατρική βασισμένη σε αποδείξεις). Η εξόρυξη δεδομένων, η εκτεταμένη ανάλυση συστάδων, η στατιστική ανάλυση, οι φιλικές προς τον χρήστη οπτικοποιήσεις, η τεχνητή νοημοσύνη και τα πρότυπα επεξεργασίας φυσικής γλώσσας (NLP) είναι μερικές από τις βάσεις στις οποίες βασίζονται τα εργαλεία (Fernandes et al., 2012).

ε. Υποστηρίζοντας την αυτοφροντίδα

Πολλές επιχειρήσεις χρησιμοποιούν τα μεγάλα δεδομένα για να μας δώσουν τη δυνατότητα να ενδυναμώσουμε τους εαυτούς μας, συνδυάζοντας τη χρησιμότητα των κινητών συσκευών με τη δύναμη των μεγάλων δεδομένων. Για την καλύτερη διαχείριση χρόνιων ασθενειών όπως ο διαβήτης, οι καρδιαγγειακές παθήσεις και το άσθμα, καθώς και για την καλύτερη κατανόηση των προτύπων και των μηχανισμών που ενθαρρύνουν τους ασθενείς να αλλάξουν τις συνήθειές τους προς όφελος της πρόληψης, μπορεί κανείς να συλλέγει ιατρικά δεδομένα και, για παράδειγμα, να καταγράφει τις συνθήκες ύπνου από τους ασθενείς (Fernandes et al., 2012).

στ. Υποστηρίζοντας τους παρόχους

Η έλλειψη χρόνου και πόρων, αφενός, και ο καταγιγισμός πληροφοριών, αφετέρου, ασκούν όλο και μεγαλύτερη πίεση στο υγειονομικό προσωπικό, το οποίο δεν επιτρέπεται να κάνει λάθη ή κρίσεις που θα έθεταν σε κίνδυνο την ποιότητα της εργασίας του. Τα μεγάλα δεδομένα μπορούν να διαδραματίσουν σημαντικό ρόλο στην υποβοήθηση των παρόχων υγειονομικής περίθαλψης, αλλά πρέπει ακόμη να δημιουργήσουμε λογισμικό που να είναι ακόμη πιο φιλικό προς τον χρήστη (Fernandes et al., 2012).

ζ. Δημιουργία ευαισθητοποίησης για το ευρύ κοινό & ενημέρωση

Τα μεγάλα δεδομένα είναι ένα προφανές μέσο που μπορεί να χρησιμοποιηθεί για να αρχίσουν να επιλύονται διάφορα ζητήματα που προκαλούνται από τη διαθεσιμότητα ή την απουσία δεδομένων. Για παράδειγμα, για την ανίχνευση κακών αντιδράσεων ή πλαστών φαρμάκων. για την εύρεση περιβαλλοντικών στοιχείων που μπορεί να κάνουν πιο πιθανή την εμφάνιση ορισμένων παθήσεων, όπως το άσθμα. για την πρόβλεψη της εκδήλωσης πανδημιών. για τη βοήθεια των αναπτυσσόμενων χωρών στην καλύτερη ιεράρχηση των θεμάτων δημόσιας υγείας και στη λήψη καλύτερων πολιτικών αποφάσεων. για τη βελτίωση της γρήγορης αντίδρασης του συστήματος δημόσιας υγείας, επιτρέποντάς μας να εντοπίζουμε επιδημίες και τη μετάδοση ασθενειών και επιτρέποντάς μας να αναλύουμε τις τάσεις των ασθενειών.

Θα μπορούσε να οδηγήσει σε πιο στοχευμένη και ταχύτερη ανάπτυξη εμβολίων από την πλευρά των εμβολίων. Θα μπορούσε επίσης να δώσει στους ασθενείς και το ευρύτερο κοινό πολλές πληροφορίες για την πρόληψη των επιληπτικών κρίσεων (Fernandes et al., 2012).

η. Δημιουργία νέου οικοσυστήματος στο χώρο της υγείας

Τα μεγάλα δεδομένα μπορούν επίσης να χρησιμοποιηθούν για την ενσωμάτωση διαφόρων πηγών δεδομένων για πρώτη φορά με τρόπους που θα μπορούσαν να βελτιώσουν νέους τύπους αναλύσεων και, γιατί όχι, να διευκολύνουν την εξεύρεση απαντήσεων σε σημαντικά ερωτήματα, πολλά από τα οποία μπορεί να μην έχουν τεθεί ακόμη. Μπορεί κανείς να εντοπίσει ποιοι ασθενείς είναι πιο πιθανό να παρουσιάσουν υποτροπή και να τους παράσχει αποτελεσματικότερη φροντίδα συνδυάζοντας και

αναλύοντας έναν συνδυασμό οργανωμένων και μη δομημένων κλινικών και οικονομικών δεδομένων. Επιπλέον, τεχνολογίες όπως η τηλεϊατρική, οι εφαρμογές που προσφέρουν διαγνωστικούς αλγόριθμους και οι εικονικοί γιατροί και νοσηλευτές μπορούν να αλλάξουν τον τρόπο με τον οποίο βλέπουμε την υγειονομική περίθαλψη. Ενισχύουμε την υγειονομική περίθαλψη αξιοποιώντας αλληλεπικαλυπτόμενες δεξαμενές πρωτογενών πληροφοριών.

Αυτά είναι μερικά παραδείγματα πηγών: πληροφορίες φαρμακευτικής έρευνας και ανάπτυξης σχετικά με τον μηχανισμό δράσης, την τοξικότητα, τις αξιώσεις ανεπιθύμητων ενεργειών, δεδομένα από παρόχους υγειονομικής περίθαλψης, και τέλος οι συμπεριφορές των ασθενών περιλαμβάνονται στις ιατρικές εικόνες. Μπορούμε να προσδιορίσουμε ποιοι ασθενείς θα ωφεληθούν εάν υιοθετήσουν συγκεκριμένες τροποποιήσεις στον τρόπο ζωής τους, χρησιμοποιώντας τα δεδομένα και τις αναλύσεις που μπορεί να έχουμε για τα προφίλ των ασθενών, εξασφαλίζοντας την πρόληψη. Ένα παράδειγμα για το πώς τα μεγάλα δεδομένα έχουν ωφελήσει τον ιατρικό τομέα είναι η χρήση των δεδομένων της Google για την πρόβλεψη της εξάπλωσης της γρίπης το 2009. Σύμφωνα με έρευνα που διεξήχθη από τον Mc Kinsey το 2011, οι Ηνωμένες Πολιτείες μπορούν να εξοικονομήσουν ένα εντυπωσιακό ποσό ύψους 1 δισεκατομμυρίου δολαρίων ετησίως από τις δαπάνες υγειονομικής περίθαλψης με την ανάλυση τεράστιων ποσοτήτων δεδομένων. Επιπλέον, επειδή χρησιμοποιούνται αλγόριθμοι για τη λήψη αποφάσεων, μειώνονται δραματικά τα ιατρικά λάθη (Garapati & Garapati, 2018).

θ. Εφαρμογές της ανάλυσης Big Data στον τομέα της έρευνας

Ο φαρμακευτικός χώρος, ένα υποσύνολο του χημικού χώρου που περιλαμβάνει πιθανές μοριακές δομές, είναι επίσης τεράστιος και χρησιμοποιείται για το σχεδιασμό νέων φαρμάκων. Ωστόσο, αναλύοντας τα τεράστια δεδομένα, μπορούμε να αξιοποιήσουμε καλύτερα τα χημικά μόρια που θα μπορούσαν να προχωρήσουν στο επόμενο στάδιο και να γίνουν φάρμακα. Η ανάλυση των mega data μπορεί επίσης να χρησιμοποιηθεί για την ανάπτυξη ενός μοντέλου που προβλέπει τη συμπεριφορά των υποψήφιων φαρμάκων, αξιοποιώντας το σύνολο των δεδομένων σε βάσεις δεδομένων όπως η PubChem, η ChEMBL κ.λπ. (Garapati & Garapati, 2018).

Επιπλέον, μπορεί να προσφέρει μοντέλα που προβλέπουν το βέλτιστο προφίλ ασθενών για κάθε φάρμακο ή ιατροτεχνολογικό προϊόν, καθώς και στατιστικά

εργαλεία και αλγόριθμους για τη βελτίωση του σχεδιασμού κλινικών δοκιμών και την επιλογή των επιλέξιμων συμμετεχόντων. Με τον τρόπο αυτό, μειώνονται τα σφάλματα σχεδιασμού μελετών και επιταχύνεται η διαδικασία προώθησης ενός φαρμάκου στην αγορά. Από τη βιοϊατρική έρευνα έχουν παραχθεί και θα συνεχίσουν να παράγονται μεγάλες ποσότητες δεδομένων σε διάφορες μορφές και επίπεδα. Για την ενίσχυση της γνώσης και της έρευνας, υπάρχει αυξανόμενη ανάγκη καλύτερης κατανόησης και αξιοποίησης των δεδομένων. Για να αξιοποιηθούν στο έπακρο τα βιοϊατρικά δεδομένα, τα Εθνικά Ινστιτούτα Υγείας της Αμερικής (NIH-NIH) έχουν ξεκινήσει μια πρωτοβουλία με την ονομασία Big Data to Knowledge (BD2K). Το BD2K αποσκοπεί στην ανάπτυξη των απαιτούμενων αναλυτικών εργαλείων και των εκπαιδευμένων επιστημόνων που θα τα υποστηρίξουν, καθώς και στον καλύτερο προσδιορισμό του τρόπου με τον οποίο θα αντληθεί αξία από τα δεδομένα, τόσο για τον ερευνητή όσο και για την ευρύτερη επιστημονική κοινότητα (Jared, 2014).

ι. Ο ρόλος των Big Data στην εξοικονόμηση πόρων

Η χρήση των μεγάλων δεδομένων στα συστήματα υγειονομικής περίθαλψης στις ΗΠΑ και σε άλλα έθνη έχει σημειώσει κάποια πρώτη επιτυχία. Οι οικονομικές προκλήσεις που προκλήθηκαν από την αύξηση των δαπανών υγειονομικής περίθαλψης στην Αμερική έχουν προκαλέσει την ανάπτυξη ενός πολλά υποσχόμενου φάσματος πιλοτικών πρωτοβουλιών που αξιοποιούν πλήρως τα μεγάλα δεδομένα:

- Μέσω της τηλεϊατρικής ή άλλων τεχνολογιών απομακρυσμένης παρακολούθησης, το Τμήμα Απόμαχων Αμερικής (VA) στην Αμερική θέτει σε εφαρμογή πολλά προγράμματα παρακολούθησης ασθενών.

- Στην Καλιφόρνια, ήταν η κοινοπραξία Kaiser Permanente που συνέδεσε το συντομότερο δυνατό τα κλινικά δεδομένα και τα δεδομένα κόστους, παρέχοντας το βασικό σύνολο δεδομένων που πυροδότησε τον εντοπισμό των παρενεργειών του Vioxx και την επακόλουθη ανάκληση του φαρμάκου.

- Το Εθνικό Ινστιτούτο Υγείας και Κλινικής Αριστείας υπήρξε πρωτοπόρος στη χρήση εκτεταμένων βάσεων δεδομένων για τη διερεύνηση της σχέσης κόστους-αποτελεσματικότητας νέων φαρμάκων και ακριβών παλαιών θεραπειών, καθώς και για τη συμμετοχή σε διαπραγματεύσεις με τη φαρμακοβιομηχανία (Jared, 2014).

Στην Αυστραλία έχουν αναπτυχθεί δύο εφαρμογές για την ανάλυση των απαιτήσεων των ασφαλιστικών ταμείων με τη χρήση μεγάλων δεδομένων για την ανεύρεση απάτης, κατάχρησης, σπατάλης και σφαλμάτων που τα συστήματα επεξεργασίας συναλλαγών δεν είναι σε θέση να εντοπίσουν. Η πρώτη εφαρμογή, CMC-I+ Plus, προσφέρει εξελιγμένη ανάλυση επιδόσεων χρησιμοποιώντας ακριβή ποσά απαιτήσεων νοσοκομείων και ασφαλιστικών ταμείων, καθώς και μεθοδολογίες για προγνωστικά μοντέλα σε ιατρικά και νοσοκομειακά δεδομένα. Η δεύτερη εφαρμογή, CMC Health Insurance Business (CMC-HIBIS), χρησιμοποιεί έναν πρωτοποριακό συνδυασμό δεδομένων αγοράς, επιχειρηματικών κανόνων, τεχνολογιών και δεδομένων οικονομικών απαιτήσεων από νοσοκομεία, ιατρούς και ασφαλιστικά ταμεία για τον εντοπισμό σφαλμάτων, απάτης και παραλείψεων, ενώ παράλληλα ενημερώνει τους ενδιαφερόμενους μέσω ειδοποιήσεων (Jared, 2014).

3.6 Παραδείγματα εφαρμογής Μεγάλων Δεδομένων & Ηλεκτρονικών Βάσεων Δεδομένων

Η υγειονομική περίθαλψη γνώρισε επανάσταση τα τελευταία χρόνια χάρη στην ψηφιοποίηση, αλλά είναι σημαντικό να θυμόμαστε ότι αυτό ήταν το αποτέλεσμα μιας μακράς εξέλιξης κατά την οποία τα δεδομένα που συλλέγονταν χρησιμοποιούνταν για ιατρικές μελέτες. Για παράδειγμα, η πρώτη αναφορά για τη λέπρα έγινε στη Νορβηγία το 1856. Το πρώτο μητρώο νεοπλασιών δημιουργήθηκε στη Δανία το 1943, και άλλα σκανδιναβικά έθνη ακολούθησαν μεταξύ 1960 και 2000. Η δυνατότητα ακριβούς απάντησης κλινικών ερωτημάτων σχετικά με υποομάδες ασθενών και ο γρήγορος εντοπισμός προειδοποιητικών σημάτων πιθανών ανεπιθύμητων ενεργειών είναι δύο πλεονεκτήματα της ύπαρξης τεράστιων συνόλων δεδομένων. Τα μεγάλα μεγέθη δειγμάτων επέτρεψαν τον γρήγορο εντοπισμό πιθανών παρενεργειών των νέων θεραπειών που ήταν άγνωστες κατά τη διάρκεια των μελετών φάσης III για τα φάρμακα. Ένας γρήγορος υπολογισμός αποκαλύπτει ότι, στην περίπτωση της απόσυρσης της rofecoxib (Vioxx της Merck) το 2004, θα είχαμε εντοπίσει τα ανεπιθύμητα καρδιαγγειακά συμβάντα για το Vioxx σε τρεις μήνες σε αντίθεση με πέντε χρόνια, αν είχαμε δεδομένα για 100 εκατομμύρια ασθενείς (Jianqing et al., 2014).

Στη Σκανδιναβία, τα δημόσια νοσοκομεία, τα πανεπιστήμια και οι ερευνητικές εγκαταστάσεις συγκεντρώνουν και διατηρούν μεγάλα δεδομένα σε ένα περιβάλλον όπου όλοι οι πολίτες έχουν πρόσβαση σε φαρμακευτικά προϊόντα και ιατρική περίθαλψη, ανεξάρτητα από τη δυνατότητά τους να πληρώσουν. Ενώ συλλέγονται μεγάλα δεδομένα, η παρακολούθηση δεν είναι ισόβια και η πρόσβαση του ασφαλισμένου στο σύστημα υγείας μπορεί να τερματιστεί αυτόματα. Αντίθετα, για παράδειγμα, στην Αμερική, οι δημογραφικές και οικονομικές ανισότητες εμφανίζονται έντονα. Μέσω της ανοικτής βάσης δεδομένων OpenMRS, το Αμερικανικό Εθνικό Ινστιτούτο επέλεξε να υποστηρίξει μια επιδημιολογική μελέτη για τη φυματίωση στο Περού το 2006. Μέχρι τον Αύγουστο του 2011, 9.256 άτομα, 102.274 φόρμες εγγραφής, 13.829 παράμετροι εργαστηριακών εξετάσεων και 208 χρήστες είχαν προστεθεί στο μητρώο, το οποίο είχε ξεκινήσει τον Σεπτέμβριο του 2009. Για το έργο χρειαζόταν ένα ασφαλές πληροφοριακό σύστημα για τη διαχείριση των δεδομένων από έναν πληθυσμό-στόχο 16.000 ατόμων που θα παρακολουθούνταν για τουλάχιστον ένα έτος. Τα ευρήματα έδειξαν ότι το OpenMRS μπορεί να αποτελέσει ένα χρήσιμο σύστημα αναζήτησης και έρευνας με εφαρμογές τόσο στην κλινική πρακτική όσο και στην έρευνα, ιδίως σε τοποθεσίες με περιορισμένους ανθρώπινους και υλικούς πόρους (Jianqing et al., 2014).

Η χρησιμότητα των μεγάλων δεδομένων και τα αποτελέσματα των ασθενών συσχετίζονται άμεσα στον τομέα της αναισθησιολογίας, γεγονός που είναι λογικό δεδομένου του όγκου των επεμβάσεων που πραγματοποιούνται καθημερινά σε όλο τον κόσμο. Τα πρώτα βήματα έγιναν το 1990 και έκτοτε έχουμε πρόσβαση σε πληθώρα πληροφοριών σχετικά με την αναισθησία λόγω των ηλεκτρονικών αρχείων καταγραφής δεδομένων. Προκειμένου να αυτοματοποιηθεί η συλλογή δεδομένων από τις διαδικασίες, το 67% των αμερικανικών πανεπιστημιακών νοσοκομείων από τον Μάιο του 2013 είχαν εγκαταστήσει το AIMS (Anaesthesia Information Management System). Από αυτά τα νοσοκομεία, σε συνδυασμό με τα δημόσια νοσοκομεία, τα τμήματα επειγόντων περιστατικών και τα δεδομένα από τους ηλεκτρονικούς φακέλους υγείας, μπορεί να προέλθει ένας απίστευτος όγκος μεγάλων δεδομένων. Το διαδικτυακό σύστημα TrialViz, το οποίο βασίζεται στο "Clinical Practice Research Datalink", είναι αξιοσημείωτο διότι επιτρέπει στους χρήστες να επιλέγουν τις κατάλληλες ιατρικές πρακτικές με βάση δύο κριτήρια: την καταλληλότητα των ασθενών για ένταξη σε μια μελέτη και την ποιότητα των δεδομένων, και όλα αυτά

σχεδόν σε πραγματικό χρόνο και με τη χρήση πρωτοποριακών οπτικοποιήσεων και εργαλείων (Jianqing et al., 2014).

Η μέθοδος των "μεγάλων δεδομένων" παρουσιάζει μια ιδιαίτερη ευκαιρία στον τομέα της καρδιο-ογκολογίας, παρά τις δυσκολίες της. Οι προσεγγίσεις των μεγάλων δεδομένων για τη φαρμακοεπαγρύπνηση μπορούν να βοηθήσουν στην έγκαιρη ανίχνευση συμπτωμάτων καρδιαγγειακής τοξικότητας που μπορεί να συνδέονται με μια ποικιλία νέων φαρμάκων και θεραπευτικών σχημάτων που ερευνώνται σε αυτόν τον τομέα. Επιπλέον, θα μπορούσε να αποσαφηνίσει την επικράτησή τους και ενδεχομένως να ενθαρρύνει τη μελέτη των πιθανών παραγόντων που συμβάλλουν. Τέλος, η χρήση των μεγάλων δεδομένων μπορεί να διαδραματίσει κρίσιμο ρόλο στη δημόσια υγεία, βοηθώντας στην εκπαίδευση του πληθυσμού και συνιστώντας κυβερνητικές πρωτοβουλίες για την αποφυγή τόσο του καρκίνου όσο και των καρδιαγγειακών παθήσεων (Lohr, 2014).

Οι μεγάλες βάσεις δεδομένων περίθαλψης χρησιμοποιούνται όλο και περισσότερο σε επιδημιολογικές μελέτες για την καρδιακή ανεπάρκεια (HF), δεδομένου ότι παρέχουν μια ανεκτίμητη ευκαιρία να εντοπιστούν οι παράγοντες κινδύνου για έναν πληθυσμό που πάσχει από μια πάθηση με κακή πρόγνωση και κακή ποιότητα ζωής. Όσον αφορά την ασφάλεια, η καρδιακή ανεπάρκεια αποτελεί συχνά πρωταρχικό ή δευτερεύοντα στόχο για πολλές δοκιμές. Οι μεγάλες βάσεις δεδομένων μπορεί να αποτελέσουν την κύρια πηγή πληροφοριών λόγω της πολυπλοκότητας της πάθησης και της διάγνωσης, της ποικίλης χρήσης της σχετικής ορολογίας στην κλινική πρακτική και των περιορισμών των διαφόρων συστημάτων κωδικοποίησης ασθενειών (Lohr, 2014).

Η χρήση φαρμακευτικών βοτάνων κατά τη διάρκεια της εγκυμοσύνης για τις αφρικανικές γυναίκες ως παράγοντας κινδύνου για τη γέννηση εμβρύου με συγγενή ανωμαλία είναι ένα ενδιαφέρον παράδειγμα του τρόπου με τον οποίο τα μεγάλα δεδομένα μπορούν να χρησιμοποιηθούν σε μη παραδοσιακούς τομείς όπως η εθνική φαρμακολογία. Στην πραγματικότητα, αποδείχθηκε ότι οι έγκυες γυναίκες χρησιμοποιούσαν αυτά τα φυτά κατά μέσο όρο από 2% έως 100% του χρόνου, κυρίως για να ανακουφίσουν την πρωινή ναυτία, να τονώσουν τον τοκετό και να διευκολύνουν τον τοκετό.

Τα μεγάλα δεδομένα στη δερματολογία μπορούν να χρησιμοποιηθούν για την ενίσχυση των μοντέλων πρόβλεψης κινδύνου νόσου, όπως για το μελάνωμα, για την υποστήριξη στοχευμένου ελέγχου για άτομα υψηλού κινδύνου (όπως ο πληθυσμιακός

έλεγχος για τη στοχευμένη διάγνωση του καρκίνου του δέρματος), για τη βελτίωση της διαχείρισης διαφόρων δερματικών νόσων, για την εξέταση της γενετικής προδιάθεσης μιας δερματικής νόσου και για την εξέταση διαφόρων φαινοτύπων νόσου που μπορεί να επωφεληθούν από εξατομικευμένες θεραπείες (π.χ. ψωρίαση ή έκζεμα). Θεωρείται επίσης ότι τα μεγάλα δεδομένα μπορούν να παρουσιάσουν εξαιρετικές ευκαιρίες για τους ερευνητές να μεγιστοποιήσουν τις δυνατότητες των ήδη υπάρχουσών πηγών δεδομένων και να δημιουργήσουν νέες για επερχόμενη μελέτη, καθώς και να αποτελέσουν μια ιδανική προσέγγιση για την παρακολούθηση και την αξιολόγηση της ασφάλειας των φαρμάκων (Lohr, 2014).

Παρόμοια με άλλους τομείς, η ρευματολογία έχει σημειώσει σημαντική βελτίωση στην κατανόηση των ζητημάτων που αφορούν την επιδημιολογία των ρευματολογικών ασθενειών, τις παθήσεις, τη θεραπευτική αποτελεσματικότητα και ασφάλεια, την ποιότητα της φροντίδας των ασθενών κ.λπ (Lohr, 2014).

4. Big Data & Τιμολόγηση ασφαλιστικών προϊόντων Υγείας

4.1 Φαρμακοεπαγρύπνιση

Πρόσφατα, μια επιστολή της Ευρωπαϊκής Επιτροπής (ΕΕ), του Ευρωπαϊκού Οργανισμού Φαρμάκων (EMA) και τις Διοικήσεις των Φαρμακευτικών Οργανισμών (ΗΜΑ), την οποία συνυπέγραψαν όλοι οι χορηγοί κλινικών δοκιμών που διεξάγονται στην Ευρωπαϊκή Ένωση, τους υπενθύμισε την υποχρέωσή τους να δημοσιεύουν περίληψη των ευρημάτων κάθε δοκιμής στη βάση δεδομένων κλινικών δοκιμών της ΕΕ (EudraCT). Στην επιστολή υποστηρίζεται ότι η διασφάλιση και η προαγωγή της δημόσιας υγείας εξαρτώνται σε κρίσιμο βαθμό από τη διαφάνεια και την πρόσβαση του κοινού στα ευρήματα των κλινικών δοκιμών, ανεξάρτητα από το αν αυτά είναι ευνοϊκά ή δυσμενή. Όλες οι κλινικές δοκιμές υποβάλλουν τα στοιχεία των πρωτοκόλλων τους και τα ευρήματά τους στη βάση δεδομένων κλινικών δοκιμών της ΕΕ (EudraCT). Το μητρώο κλινικών δοκιμών της ΕΕ καθιστά τις πληροφορίες αυτές διαθέσιμες στο ευρύ κοινό (CTR της ΕΕ). Οι χορηγοί πρέπει να δημοσιοποιούν τα

δεδομένα της δοκιμής εντός ενός έτους από την ολοκλήρωση της μελέτης (ή 6 μήνες για παιδιατρική δοκιμή). Η Διεθνής Πλατφόρμα Μητρώου Κλινικών Δοκιμών (ICTRP) του Παγκόσμιου Οργανισμού Υγείας (WHO), του οποίου το EU CTR είναι το σημαντικότερο μητρώο, λαμβάνει αυτές τις πληροφορίες. Τον Απρίλιο του 2019 υπήρχαν συνολικά 57.68 κλινικές δοκιμές στη βάση δεδομένων EudraCT, εκ των οποίων οι 27.093 είχαν ολοκληρωθεί. Τα αποτελέσματα για 18.432 από αυτές τις ολοκληρωμένες δοκιμές θα έπρεπε να έχουν καταγραφεί. Για το 68,2% των δοκιμών, οι χορηγοί πληρούσαν τις προϋποθέσεις για τη δημοσίευση- ωστόσο, το 31,8% των αποτελεσμάτων των δοκιμών παρέμενε ανολοκλήρωτο (Raghupathi & Raghupathi, 2014).

Ένας υπεύθυνος φαρμακοεπαγρύπνησης πρέπει να διορίζεται σύμφωνα με το ευρωπαϊκό κανονιστικό και νομικό πλαίσιο, και πιο συγκεκριμένα τον τόμο 9A EudraLex, προκειμένου να διασφαλίζεται ότι ο κάτοχος άδειας κυκλοφορίας φαρμακευτικών προϊόντων συμμορφώνεται με τις απαιτήσεις φαρμακοεπαγρύπνησης στην Ευρωπαϊκή Ένωση και τον Ευρωπαϊκό Οικονομικό Χώρο (EEA Qualified Person Pharmacovigilance - EEA QPPV). Το σύστημα που χρησιμοποιεί ο Κάτοχος Άδειας Κυκλοφορίας (ΚΑΚ) για τη συλλογή, αξιολόγηση και αναφορά πληροφοριών σχετικά με τυχόν πιθανές ανεπιθύμητες ενέργειες που αρχειοθετούνται, τουλάχιστον σε μία κεντρική τοποθεσία εντός της Ευρωπαϊκής Ένωσης, θα πρέπει να είναι απολύτως οικείο στον υπεύθυνο φαρμακοεπαγρύπνησης. Για την υποβολή ηλεκτρονικών αναφορών στη βάση δεδομένων EudraVigilance, ένα τέτοιο σύστημα απαιτεί συχνά τη χρήση μιας ηλεκτρονικής βάσης δεδομένων. Το σύστημα της κίτρινης κάρτας καταγράφει ηλεκτρονικά τις ανεπιθύμητες ενέργειες. Ο μηχανισμός που χρησιμοποιείται για τη μεταφορά και την επεξεργασία δεδομένων φαρμακοεπαγρύπνησης θα πρέπει να διασφαλίζει ότι όλες οι διαδικασίες σύνθεσης και επαναξιολόγησης είναι ανιχνεύσιμες και εμπιστευτικές τόσο κατά την εσωτερική διατμηματική ανταλλαγή πληροφοριών όσο και κατά τη συνεργασία με τρίτους εξωτερικούς αποδέκτες. Τόσο ο αποστολέας όσο και ο προορισμός πρέπει να χρησιμοποιούν διακριτά αναγνωριστικά για να είναι σε θέση να πιστοποιούν κάθε επεξεργασία, σύνθεση και ανταλλαγή δεδομένων, προκειμένου η διαδικασία της διασύνδεσης και της ανταλλαγής δεδομένων να διεξάγεται σωστά για να επιτευχθεί μια βιώσιμη λύση (Raghupathi & Raghupathi, 2014).

4.2 Ηλεκτρονική Συνταγογράφηση

Από τον Μάιο του 2010, όλοι οι γιατροί στην Ελλάδα, ανεξάρτητα από το αν έχουν σύμβαση με τον ΕΟΠΥΥ ή όχι, έχουν τη δυνατότητα να συνταγογραφούν ηλεκτρονικά φάρμακα στους ασθενείς τους. Η δυνατότητα αυτή τέθηκε ως προϋπόθεση στην πρώτη δανειακή σύμβαση της Ελλάδας (Μνημόνιο Ι). Η κρατική εταιρεία ΗΔΙΚΑ ΑΕ ανέλαβε την τεχνική υποδομή και την υποστήριξη του ιατρικού λογισμικού για την ηλεκτρονική συνταγογράφηση. Το Σύστημα Ηλεκτρονικής Συνταγογράφησης της ΗΔΙΚΑ ΑΕ μόλις εγκρίθηκε για χρήση στα νοσοκομεία για την υποστήριξη των διαδικασιών συνταγογράφησης από τους θεράποντες ιατρούς τόσο για τους εσωτερικούς όσο και για τους εξωτερικούς ασθενείς. Με την αξιοποίηση όλων των υπηρεσιών που παρέχει η ηλεκτρονική συνταγογράφηση, αυτό συνεπάγεται και την αναβάθμιση του επιπέδου ιατρικής φροντίδας, υποστήριξης και προστασίας του ασθενούς (φίλτρα SPC, Πρωτόκολλα θεραπευτικής συνταγογράφησης, αντενδείξεις, αλληλεπιδράσεις, συσχέτιση διαγνώσεων με θεραπείες, εκτέλεση συνταγών, ιχνηλασιμότητα φαρμάκων κ.α.).

Στην απόφαση αναφέρεται ότι το Υπουργείο Υγείας θεωρεί ότι η σύνδεση των πληροφοριακών υποσυστημάτων του νοσοκομείου με τους πρωτογενείς μηχανισμούς διαλειτουργικότητας του Συστήματος Ηλεκτρονικής Συνταγογράφησης είναι η καλύτερη επιλογή για την επίτευξη αυτής της λειτουργικής ικανότητας (ePrescription API services). Οι ασθενείς, οι επαγγελματίες υγείας, οι οργανισμοί κοινωνικής ασφάλισης και η Ευρωπαϊκή Ένωση κερδίζουν ταυτόχρονα από αυτό (Raghupathi & Raghupathi, 2014).

A. Για τους ασθενείς

Η πιθανότητα λάθους κατά την εκτέλεση της συνταγής μειώνεται με τη χρήση της κωδικοποίησης. Επιπλέον, οι αλληλεπιδράσεις ενός συνταγογραφούμενου σκευάσματος βρίσκονται υπό τη διαχείριση ενός ολοκληρωμένου συστήματος ηλεκτρονικής συνταγογράφησης για τη φαρμακοεπιδημιολογία. Δεδομένου ότι το πληροφοριακό σύστημα καταγράφει, αποθηκεύει και ανακτά όλες τις ηλεκτρονικές συνταγές ανάλογα με τις ανάγκες, ο ασθενής δεν χρειάζεται να παρακολουθεί άσκοπη γραφειοκρατία ή διαβουλεύσεις. Τα δεδομένα συμπεριλαμβάνονται στις πληροφορίες

σε πραγματικό χρόνο και βοηθούν τον θεράποντα ιατρό να αποκτήσει μια εμπειριστατωμένη εικόνα του ιστορικού του ασθενούς. Επιπλέον, ο ασθενής αναλαμβάνει ενεργό ρόλο στη θεραπεία που τον αφορά, επειδή η ψηφιοποίηση της υγειονομικής περίθαλψης τον διευκολύνει να έχει πρόσβαση στους ηλεκτρονικούς φακέλους του και να συμμετέχει στη λήψη αποφάσεων, ενώ παράλληλα λαμβάνει ιατρική περίθαλψη υψηλής ποιότητας (Roy, 2017).

B. Για τους επαγγελματίες υγείας

Καθώς ο γιατρός καλείται να επιλέξει από μια βάση δεδομένων, τα ιατρικά λάθη μειώνονται σημαντικά και το EY έχει μια πλήρη εικόνα του ασθενούς. Το σύστημα ηλεκτρονικής συνταγογράφησης έχει τη δυνατότητα εξοικονόμησης μεγάλου χρόνου, γεγονός που επιτρέπει στον γιατρό να αφιερώσει περισσότερη από την προσοχή του στον ασθενή και λιγότερο χρόνο σε διοικητικά καθήκοντα.

Ο φαρμακοποιός κερδίζει από τη χρήση της ηλεκτρονικής συνταγογράφησης, καθώς η ηλεκτρονική εκτέλεση της συνταγής τον φέρνει πιο κοντά στον ασθενή και επιταχύνει την πληρωμή των λογαριασμών. Ο φαρμακοποιός είναι επίσης σε θέση να διαχειρίζεται αποτελεσματικότερα τα αποθέματά του, αφού, χάρη στη στατιστική ανάλυση των ηλεκτρονικών συνταγών, μπορεί να έχει σαφή εικόνα της ροής των διαφόρων φαρμακευτικών σκευασμάτων και να εκμεταλλεύεται τις οικονομίες κλίμακας. Αφαιρώντας την πιθανότητα ανωμαλιών κατά την εκτέλεσή της, η διαφάνεια του συστήματος ηλεκτρονικής συνταγογράφησης δεν αφήνει περιθώρια αμφισβήτησης (Roy, 2017).

Γ. Για την Πολιτεία και τους Φορείς Κοινωνικής Ασφάλισης

Η ηλεκτρονική συνταγογράφηση έχει οδηγήσει σε σημαντική εξοικονόμηση κόστους, επειδή οι κατευθυντήριες γραμμές που έχουν καθοριστεί σύμφωνα με την Ιατρική βασισμένη σε στοιχεία ("Evidence-Based Medicine", EBM) έχουν μειώσει σημαντικά το πρόβλημα της σπατάλης στη συνταγογράφηση. Λόγω των μέτρων ασφαλείας που εφαρμόζονται, αποτρέπονται οι διπλές ηλεκτρονικές συνταγές. Το σύστημα μπορεί να γίνει το καλύτερο δυνατό για τη λήψη αποφάσεων χάρη στη δυνατότητα καταγραφής δεδομένων και στατιστικής επεξεργασίας τους. Τα δεδομένα που παράγονται από την ηλεκτρονική συνταγογράφηση δίνουν στην επιστημονική

κοινότητα τη δυνατότητα πρόβλεψης μελλοντικών γεγονότων και δίνουν στην κυβέρνηση την ευκαιρία να εφαρμόσει μέτρα δημόσιας υγείας που προάγουν την υγεία των πολιτών (Rud, 2012).

Δ. Για την Ευρωπαϊκή κοινότητα

Με την ανάπτυξη της διασυνοριακής ηλεκτρονικής συνταγογράφησης, τα ευρωπαϊκά άτομα θα μπορούν να παραγγέλνουν τα φάρμακά τους σε ένα κράτος και να εκτελούν τη συνταγή σε άλλο κράτος από το οποίο μπορεί να ταξιδεύουν, όπως προβλέπεται από την πρωτοβουλία eρSOS, επιτρέποντάς τους να έχουν πρόσβαση σε διασυνοριακές υπηρεσίες υγείας υψηλής ποιότητας. Επιπλέον, ορισμένες θεραπείες και φάρμακα που περιλαμβάνονται στον "αρνητικό κατάλογο" του Εθνικού Οργανισμού Παροχής Υπηρεσιών Υγείας (ΕΟΠΥΥ) δεν καλύπτονται από τις ηλεκτρονικές συνταγές (π.χ. λογοθεραπεία). Λόγω της επικείμενης κατασκευής του ηλεκτρονικού φακέλου υγείας, ο οποίος θα πρέπει να περιέχει όλες τις ιατρικές υπηρεσίες και τα φάρμακα που έχουν χορηγηθεί προκειμένου να είναι πλήρης, αυτό πρέπει να εξεταστεί. Πρόσφατα κατασκευάστηκε η ηλεκτρονική πλατφόρμα eΔΑΠΥ, η οποία επιτρέπει τη διακίνηση διαβητικών ειδών, ορθοπεδικού και αναπνευστικού εξοπλισμού, καθώς και ειδών διατροφής και δίαιτας, προκειμένου να αποζημιώνονται από την ΕΟΠΥΥ. Αν και πρόκειται για μια διαφορετική πλατφόρμα, ακολουθεί την ίδια λογική (πιστοποίηση των επαγγελματιών υγείας, αριθμοί κοινωνικής ασφάλισης των ασθενών κ.λπ.) με την πλατφόρμα της ηλεκτρονικής συνταγογράφησης (Wang, 2015).

4.3 ΕΟΠΥΥ

Οι αρμοδιότητες του αντίστοιχου τμήματος του ΕΟΠΥΥ περιγράφονται ακολούθως :

- Προκειμένου να διασφαλιστεί το επίπεδο της φαρμακευτικής περίθαλψης που παρέχεται στα καλυπτόμενα πρόσωπα του Οργανισμού και η υπεύθυνη χρήση των

διαθέσιμων φαρμακευτικών πόρων, απαιτείται παρακολούθηση και αξιολόγηση των αποτελεσμάτων που επιτυγχάνονται μέσω των συστημάτων ηλεκτρονικής συνταγογράφησης και επεξεργασίας συνταγών.

- Δημιουργία και τήρηση μητρώου φαρμάκων που περιλαμβάνει επιπλέον στοιχεία σχετικά με τις πολιτικές πληρωμής.

- Συλλογή και επεξεργασία δεδομένων σχετικά με τη χρήση συνταγογραφούμενων φαρμάκων και φαρμακευτικών σκευασμάτων, προκειμένου να εκδίδονται οι κατάλληλες αποφάσεις για rebate και clawback.

- Θέσπιση των απαιτούμενων κανονισμών ηλεκτρονικής συνταγογράφησης και επίβλεψη της εφαρμογής τους. Συμμετοχή στη διαδικασία δημιουργίας και εφαρμογής στον οργανισμό μητρώων παρακολούθησης ακριβών και ορφανών θεραπειών, καθώς και παρακολούθησης της τήρησης των θεραπευτικών πρωτοκόλλων. Με τους φαρμακευτικούς παραγωγούς, να εφαρμόζετε συμφωνίες για τον όγκο των τιμών και τον επιμερισμό του κινδύνου και να συνδράμετε στη διαδικασία αξιολόγησης τεχνολογιών υγείας.

- Να συνδράμει την Επιτροπή Διαπραγμάτευσης του ΕΟΠΥΥ παρέχοντας πληροφορίες σχετικά με τις τιμές και τη χρήση των φαρμάκων (Παρασύρη, 2017).

- Υποβολή εισηγήσεων προς την Επιτροπή Διαπραγμάτευσης σχετικά με την προστιθέμενη αξία των νέων φαρμακευτικών προϊόντων, συγκριτικές αναλύσεις των προϊόντων που κυκλοφορούν ήδη στην αγορά και επεξεργασία δεδομένων που τίθενται στη διάθεση της Διεύθυνσης ως αποτέλεσμα της παρακολούθησης της ανάλυσης της φαρμακευτικής κατανάλωσης και της χρήσης των δεδομένων που τίθενται στη διάθεση ως αποτέλεσμα της συμμετοχής σε ευρωπαϊκά και διεθνή δίκτυα τιμολόγησης και αποζημίωσης.

- Υποβοήθηση του έργου των επιτροπών για τον καθορισμό του κατά πόσον ένας ασθενής πρέπει να θεραπευτεί με ακριβά ή διαφορετικά μη διαθέσιμα φάρμακα.

- Χειρισμός ερωτημάτων σχετικά με φάρμακα OFF LABEL

- Στις αρμοδιότητες του Τμήματος Επεξεργασίας και Ελέγχου Συνταγών του ΕΟΠΥΥ περιλαμβάνεται η διαχείριση των ακυρώσεων συνταγών και η εξαίρεση του ΑΜΚΑ από την ηλεκτρονική συνταγογράφηση.

- Παραλαβή τιμολογίων από συμβεβλημένα ιδιωτικά φαρμακεία, φαρμακεία κρατικών νοσοκομείων, στρατιωτικά νοσοκομεία και φαρμακεία του ΕΟΠΥΥ και ηλεκτρονική επεξεργασία τους.

- Για να διαπιστωθεί αν τηρήθηκε η φαρμακευτική νομοθεσία κατά την έκδοση και εκτέλεση των συνταγών για φάρμακα, διενεργήθηκε έλεγχος των εκτελεσμένων συνταγών που ελήφθησαν από το μηχανογραφημένο σύστημα συνταγογράφησης και των χειρόγραφων συνταγών που υποβλήθηκαν από τα φαρμακεία.

- Ελέγχθηκαν οι επισκοπήσεις των σχετικών συνταγογραφούμενων φαρμάκων και των εξειδικευμένων συμπληρωμάτων διατροφής για την αναζήτηση δόλιων συνταγών.

- Αποδέσμευση των ευρημάτων του ελέγχου ώστε να είναι δυνατή η πραγματοποίηση πληρωμών προς τα στρατιωτικά νοσοκομεία, τα φαρμακεία των κρατικών νοσοκομείων και τα ιδιωτικά φαρμακεία.

- Επεξεργασία των δεδομένων από την εξαγωγή του ελέγχου για clawback και rebate, καθώς και των δεικτών συνταγογράφησης και φαρμακευτικής κατανάλωσης, προκειμένου να καθοριστούν οι στόχοι και να ληφθούν μέτρα για την ορθή συνταγογράφηση και την ορθή χορήγηση, καθώς και για την προετοιμασία εκθέσεων και αναφορών (Παρασύρη, 2017).

- Παροχή μηνιαίας ανάλυσης των ευρημάτων του ελέγχου του λογαριασμού συνταγογράφησης για τους συμβεβλημένους φαρμακοποιούς και στατιστικών πληροφοριών σχετικά με την επιστροφή των προϊόντων τους στους κατόχους άδειας κυκλοφορίας, προκειμένου να ενημερώνονται οι συνταγογραφούντες γιατροί.

- Κωδικοποίηση των κανονισμών φαρμάκων για το μηχανογραφημένο σύστημα ελέγχου και επιχειρηματικής ευφυΐας του Οργανισμού.

- Επεξήγηση του τρόπου έκδοσης και εκτέλεσης των συνταγών του ΕΟΠΥΥ

- Εποπτεία και έλεγχος των διαδικασιών και πρακτικών διασφάλισης ποιότητας που χρησιμοποιούνται για την προμήθεια των ασφαλισμένων του ΕΟΠΥΥ με τα σχετικά φάρμακα και τα εξειδικευμένα διαιτητικά σκευάσματα. διαχείριση των αιτημάτων ΑΥΣ από τους ασφαλισμένους του Οργανισμού. Εφαρμογή των συμφωνιών επιμερισμού όγκου τιμών και κινδύνου με τους παραγωγούς ιατρικού εξοπλισμού και τους προμηθευτές εξειδικευμένων σκευασμάτων διατροφής (Παρασύρη, 2017).

- Συνδρομή στη διαδικασία εξασφάλισης περαιτέρω εκπτώσεων από τους προμηθευτές και παρακολούθηση της είσπραξης του κέρδους χονδρικής πώλησης.

- Δημιουργία και τήρηση μητρώου των σχετικών ειδικών τροφίμων και φαρμακευτικών προϊόντων.

- Η παρακολούθηση της συμμόρφωσης με τους όρους και τις προϋποθέσεις των συμβάσεων μεταξύ του φαρμακείου και του παρόχου θεραπευτικού παράγοντα.

- Υποστήριξη της Επιτροπής Διαπραγμάτευσης του ΕΟΠΥΥ με την παροχή στοιχείων τιμών και κατανάλωσης σχετικών φαρμακευτικών προϊόντων και ειδικών διαιτητικών σκευασμάτων, καθώς και εισήγηση προς την Επιτροπή για την προσθήκη νέων προϊόντων στην αγορά και τη διενέργεια συγκριτικών αναλύσεων της αγοράς.

- Ανάλυση των πληροφοριών που αντλούνται από το μηχανογραφημένο σύστημα ελέγχου και εκτέλεσης συνταγών σε περιπτώσεις που γιατροί και φαρμακοποιοί δεν τηρούν τη νομοθεσία. παραγωγή σχετικών εισηγήσεων προς τα όργανα λήψης αποφάσεων των κυρωτικών οργανισμών.

- Αντιμετώπιση των προσφυγών κατά των αποφάσεων του Οργανισμού από φαρμακοποιούς, γιατρούς και φαρμακευτικές εταιρείες

- Διεκπεραίωση υποθέσεων που αφορούν κυρώσεις και χρέη που συνδέονται με τον ΕΟΠΥΥ

- Δημιουργία και τήρηση καταλόγου κρατικών και ιδιωτικών φαρμακείων για τη διευκόλυνση της ηλεκτρονικής συνταγογράφησης, πληρωμής λογαριασμών και συστημάτων σάρωσης

- Επεξεργασία των ευρημάτων από το σύστημα ηλεκτρονικής συνταγογράφησης για την επαλήθευση ότι οι συνταγές εκτελούνται σύμφωνα με τις αποφάσεις του Οργανισμού σε υποθέσεις που αφορούν φαρμακοποιούς και φαρμακευτικές εταιρείες Επεξεργασία υποθέσεων κυρώσεων για την ολοκληρωμένη χρηματοδότηση του ΕΟΠΥΥ (Παρασύρη, 2017).

4.4 DRGs & KEN

Οι νοσοκομειακές επισκέψεις που απαιτούν την παροχή εξειδικευμένης θεραπείας, την κατανομή οικονομικών, ανθρώπινων και τεχνολογικών πόρων και τη δυναμική χρήση ιατρικών διαδικασιών για την επίλυση του προβλήματος κατηγοριοποιούνται και ομαδοποιούνται χρησιμοποιώντας το σύστημα DRG. Διαφέρει σε όλο τον κόσμο και η χρήση του εισήχθη στα νοσοκομεία προκειμένου να ολοκληρωθεί η επικοινωνία και ο συντονισμός με τους ασφαλιστικούς φορείς, ώστε να παρέχονται αποζημιώσεις για το κόστος νοσηλείας που θα μπορούσαν να υπολογιστούν με διαφανή τρόπο (Watson et al., 2020).

Τα οφέλη από τη χρήση των DRGs περιλαμβάνουν τη μείωση της τυπικής διάρκειας παραμονής και νοσηλείας στα νοσοκομεία, καθώς και τη βελτίωση της αποτελεσματικότητας και της διαφάνειας στη χρήση και την κατανομή των πόρων. Λαμβάνονται γρήγορες αποφάσεις και με την αποτελεσματική κατανομή των πόρων μεγιστοποιείται η συνολική ποιότητα των υπηρεσιών. Ωστόσο, η οικονομική παραμετροποίηση της νόσου δεν διασφαλίζει ότι ο ασθενής θα κερδίσει κλινικά, επειδή δίνεται προτεραιότητα στα οικονομικά κίνητρα που προκύπτουν από τη γρήγορη έξοδο από την ιατρική μονάδα (Watson et al., 2020).

Στην Ευρώπη παρατηρούνται τρεις βασικές τάσεις DRG:

- 1) Οι χώρες προσπαθούν να χρησιμοποιήσουν τα συστήματα πληρωμών ως κίνητρα για τη βελτίωση της ποιότητας
- 2) Τα συστήματα DRGs αναβαθμίζονται για να προσαρμόζονται σε αυτές τις περιπτώσεις, δεδομένου ότι τα νοσοκομεία βλέπουν αύξηση της καθημερινής δραστηριότητας, ιδίως στα εξωτερικά ιατρεία. Ακολουθώντας το αμερικανικό μοντέλο, πολλά ευρωπαϊκά έθνη, συμπεριλαμβανομένης της Αγγλίας, της Γερμανίας και της Σουηδίας, επεκτείνουν τη χρήση των DRGs στα νοσοκομεία ψυχικής υγείας και αποκατάστασης.
- 3) Τα συστήματα ομαδοποιημένων πληρωμών αναπτύσσονται με μεγάλο ενδιαφέρον προκειμένου να βελτιωθεί η συνεργασία και η ποιότητα και να ευθυγραμμιστούν καλύτερα τα κίνητρα μεταξύ των διαφόρων παρόχων υγειονομικής περίθαλψης. Ωστόσο, είναι αρκετά δύσκολο να δημιουργηθούν συσκευές όπως τα DRGs για αυτά (Watson et al., 2020).

Με βάση τα διεθνώς αναγνωρισμένα DRGs, το ελληνικό σύστημα χρέωσης και διαχείρισης Ελληνικών Κλειστών Νοσηλείων (KEN) χρησιμοποιείται για τις εισαγωγές νοσοκομειακών ασθενών. Στόχος της διαδικασίας είναι η ταξινόμηση όλων των νοσηλείων ασθενών σε περίπου 700 κατηγορίες, προκειμένου να εξορθολογιστούν οι διαδικασίες χρέωσης, ελέγχου, εκκαθάρισης τιμολογίων και πληρωμής από τους ασφαλιστικούς οργανισμούς των νοσοκομείων. Οι KEN, οι οποίοι χρησιμοποιούνται για τη γρήγορη και απλή τιμολόγηση περιστατικών, είναι ουσιαστικά ένας τυποποιημένος κατάλογος νοσηλείων που καθορίζει το κόστος και την τυπική διάρκεια παραμονής για κάθε κατηγορία νοσηλείας. Εν κατακλείδι, παρά τις δυσκολίες, τα DRGs είναι πολύ δημοφιλή σε όλο τον κόσμο, δεδομένου ότι παρέχουν τα καλύτερα ποσοτικοποιήσιμα στοιχεία για το τι πραγματικά επιτυγχάνουν τα νοσοκομεία. Και

εξαιτίας αυτού, είναι ζωτικής σημασίας για τα ευρωπαϊκά έθνη να ανταλλάσσουν γνώσεις, σοφία και βέλτιστες πρακτικές προκειμένου να επωφεληθούν το ένα από το άλλο (Yousef, 2021).

4.5 ICD 11

Ο Παγκόσμιος Οργανισμός Υγείας υποστηρίζει τη Διεθνή Ταξινόμηση των Ασθενειών, ένα παγκόσμιο πρότυπο για την αναφορά ασθενειών και προβλημάτων υγείας, το οποίο αποσκοπεί κυρίως στην κωδικοποίηση ασθενειών και την οργάνωση των πληροφοριών για την υγεία. Η ICD-11, η πιο πρόσφατη και επικαιροποιημένη έκδοση της διεθνούς κατηγοριοποίησης, κυκλοφόρησε στις 18 Ιουνίου 2018 (79). Ακολουθεί ένας κατάλογος με τα πλεονεκτήματα αυτού του κωδικού: □

- ανταλλαγή και αξιολόγηση δεδομένων υγείας σε διάφορα συστήματα υγείας, περιοχές και χώρες
- συγκρίσεις μεταξύ δεδομένων από την ίδια περιοχή □
- έλεγχος της επίπτωσης και επιπολασμού των ασθενειών □
- διαφάνεια στη διαχείριση της κατανομής των πόρων και της αμοιβής των παροχών □
- παρακολούθηση των οδηγιών ασφάλειας και ποιότητας □
- υπολογισμός και μέτρηση ασθενειών, θανάτων, συμπτωμάτων και άλλων μεταβλητών που σχετίζονται με την υγεία (Yousef, 2021).
- η ενδεδειγμένη ταξινόμηση των διαταραχών διευκολύνει τις διάφορες απαιτήσεις των επαγγελματιών υγείας για τις ασθένειες. □
- οι πληροφορίες υγείας μπορούν εύκολα να αποθηκεύονται, να ανακτώνται και να αναλύονται για να βοηθούν τους γιατρούς να λαμβάνουν αποφάσεις (Yousef, 2021).

4.6 Ηλεκτρονικός Φάκελος Υγείας

Ο προσωπικός ηλεκτρονικός φάκελος υγείας περιλαμβάνει το συγκεκριμένο ιατρικό ιστορικό του λήπτη, καθώς και όλα τα δεδομένα, τις αξιολογήσεις και άλλες

πληροφορίες σχετικά με την κατάσταση του ασθενούς και τις αλλαγές κατά τη διάρκεια της θεραπείας. Το περιεχόμενο του ΑΗΦΥ πρέπει να διατηρείται για πάντα και να απαιτείται ενιαία σε ομοσπονδιακό επίπεδο. Το άρθρο 84 του σχεδίου νόμου "Εκσυγχρονισμός και μεταρρύθμιση του θεσμικού πλαισίου των ιδιωτικών κλινικών, ίδρυση του Εθνικού Οργανισμού Δημόσιας Υγείας, ίδρυση του Εθνικού Ινστιτούτου Νεοπλασιών και άλλες διατάξεις" περιγράφει αυτά. Η δημιουργία, η διαχείριση και η επεξεργασία του φακέλου εμπίπτει στην αρμοδιότητα του Υπουργείου Υγείας. Ο γενικός ιατρός, ο παθολόγος ή ο παιδίατρος που είναι πιστοποιημένος και εξουσιοδοτημένος από το σύστημα αρχειοθέτησης του ΑΗΦΥ μπορεί να ενεργοποιήσει το ΑΗΦΥ για λογαριασμό του εγγεγραμμένου λήπτη υπηρεσιών υγείας. Κάθε πληροφορία που συλλέγεται κατά τη διάρκεια ενός ελέγχου ή μιας επίσκεψης και η οποία είναι απαραίτητη για τη διάγνωση, τη θεραπεία, τη νοσηλεία ή την αποκατάσταση της σωματικής ή ψυχικής υγείας του ασθενούς πρέπει να καταχωρείται στο ΑΗΦΥ από όλους τους γιατρούς και άλλους παρόχους υγειονομικής περίθαλψης. Οι κανόνες του Γενικού Κανονισμού για την Προστασία Δεδομένων θα εφαρμόζονται στα Ατομικά Ηλεκτρονικά Αρχεία Υγείας. Η Ηλεκτρονική Συνταγογράφηση, το ηλεκτρονικό αρχείο δαπανών δικαιούχων ασφάλισης υγείας που αναπτύχθηκε και τηρείται στον ΑΗΦΥ, καθώς και οι πολυάριθμες εγγραφές συγκαταλέγονται στα συστήματα αρχειοθέτησης του ΕΟΠΥΥ (π.χ. μητρώα ασθενών που δημιουργούνται επί του παρόντος). Για τη διενέργεια επιδημιολογικών, στατιστικών, οικονομικών, διοικητικών και διαχειριστικών αναλύσεων με σκοπό την ενίσχυση των δεικτών υγείας και του διαμετρήματος των προσφερόμενων υπηρεσιών, η ΗΔΙΚΑ έχει πλέον τη δυνατότητα να διαβιβάζει ανώνυμα δεδομένα στο Υπουργείο Υγείας. Για τη διεξαγωγή επιστημονικής ή ιστορικής έρευνας ή για στατιστικούς λόγους, το Υπουργείο Υγείας, ως υπεύθυνος επεξεργασίας, επιτρέπεται να δημοσιεύει ή να διαθέτει, έναντι συνδρομής ή καθορισμένης χρέωσης, συγκεντρωτικά δεδομένα στατιστικού τύπου από τα οποία δεν είναι δυνατή η ταυτοποίηση των υποκειμένων των δεδομένων. Ο πολίτης-αποδέκτης υπηρεσιών υγείας έχει το δικαίωμα να βλέπει τα δεδομένα εντός της ΑΗΦΥ (Yousef, 2021).

Το Υπουργείο Υγείας και η ΗΔΙΚΑ Α.Ε. τονίζουν ότι, με βάση τις πληροφορίες που προκύπτουν από την καταγραφή των δεδομένων του χρήστη που εισέρχεται στο σύστημα, είναι δυνατόν να ενημερωθεί με ειδική καταχώρηση στο ΑΗΦΥ, με ηλεκτρονικό ταχυδρομείο ή SMS, για το ποιος έχει πρόσβαση στις πληροφορίες του φακέλου του και για τον χρόνο αυτής της πρόσβασης. Τέλος, τονίζεται ότι τρίτοι, όπως

εργοδότες, ασφαλιστικοί φορείς ή τράπεζες, δεν μπορούν να αξιοποιήσουν τα ευαίσθητα προσωπικά δεδομένα που συλλέγονται στο ΑΗΦΥ για άλλους σκοπούς. Στην ουσία, η ΗΔΙΚΑ θα εξελιχθεί σε ένα μεγάλο κέντρο δεδομένων, στο οποίο θα αποθηκεύονται όλες οι πληροφορίες για την υγεία ή τουλάχιστον εκείνες για τις οποίες το άτομο έχει δώσει ρητή έγκριση (Yousef, 2021).

4.7 epSOS

Το ευρωπαϊκό έργο epSOS (Smart Open Services for European Patients) λειτουργεί σε ένα πολύπλοκο περιβάλλον και επικεντρώνει τη συλλογιστική του σε δύο σημαντικές διασυνοριακές υπηρεσίες περίθαλψης, την περίληψη ασθενούς (patient summary), την ηλεκτρονική συνταγογράφηση (eprescription) και την εκτέλεση της ιατρικής συνταγής (edispensing) (2008-2014). Πρωταρχικός στόχος του έργου είναι να δείξει ότι όλα τα κράτη μέλη που παρέχουν σήμερα αυτές τις υπηρεσίες μπορούν να το κάνουν επ' αόριστον. Η επίτευξη της διαλειτουργικότητας και η τυποποίηση των διαδικασιών, ώστε οι πληροφορίες των ασθενών να είναι ορατές και κατανοητές από όλους, αποτελούν προϋποθέσεις για την επίτευξη αυτού του στόχου. Ταυτόχρονα, διασφαλίζεται ότι η ανταλλαγή ιατρικών δεδομένων μεταξύ των μελών της ΕΕ θα γίνεται με ασφαλή και αποτελεσματικό τρόπο. Για να διασφαλιστεί ότι η διαλειτουργικότητα δεν θα οδηγήσει σε αλλαγές στο υφιστάμενο νομικό και κανονιστικό πλαίσιο κάθε χώρας, είναι ζωτικής σημασίας να ληφθούν υπόψη οι νομικές ιδιαιτερότητες και διακρίσεις κάθε κράτους μέλους της ΕΕ (Weill & Woerner, 2015).

4.8 ΕΔΑΠΥ

(Δήλωση Αναλυτικών Παραστατικών Υγείας): Η εΔΑΠΥ είναι ένα ολοκληρωμένο σύστημα που αυτοδιαχειρίζεται από τους άμεσα εμπλεκόμενους, συμπεριλαμβανομένων των ιατρών, των φαρμακείων, των παρόχων υγειονομικής περίθαλψης και των κατασκευαστών ιατρικού εξοπλισμού και συσκευών. Χρησιμοποιείται για την καταγραφή και τον έλεγχο, σε επίπεδο σύμβασης, των παρεχόμενων υπηρεσιών υγείας και, τελικά, της αποζημίωσης για τις υπηρεσίες αυτές. Ο Οργανισμός είναι επίσης σε θέση να διαχειρίζεται τις απαιτήσεις των πολιτών εξίσου

σε όλες τις κατηγορίες παροχών υγείας (φάρμακα, ειδικές θεραπείες, νοσηλεία, ιατρικές εξετάσεις, υλικά, διατροφικά σκευάσματα, θεραπευτικά είδη κ.λπ.) μέσω του ΔΑΠΥ. Επιπλέον, ο Οργανισμός είναι επίσης σε θέση να διασφαλίσει ενιαίους κανόνες για την αντιμετώπιση των χιλιάδων επαγγελματιών υγείας που συνεργάζονται με τον Οργανισμό σε καθημερινή βάση στους διάφορους τομείς ενδιαφέροντός τους. Το σύστημα του ΔΑΠΥ παρακολουθείται καθημερινά από ιατρούς και άλλους επαγγελματίες υγείας, καθώς και από τους ίδιους τους πολίτες, οι οποίοι έχουν τη δυνατότητα να συνδεθούν στο σύστημα χρησιμοποιώντας τον προσωπικό τους φάκελο ασφάλισης υγείας για να δουν πώς έχουν αντιδράσει οι ασφαλιστικές τους εισφορές στις υπηρεσίες, τα φάρμακα, τα υλικά και τα θεραπευτικά βοηθήματα που έχουν λάβει μέσω του ΕΟΠΥΥ από ιατρούς, φαρμακεία, δημόσιες και ιδιωτικές δομές, προμηθευτές επαγγελματιών υγείας (Παρασύρη, 2017).

4.9 Σύστημα Ηλεκτρονικής Προέγκρισης

Ο Ε.Ο.Π.Υ.Υ.Υ. έχει δημιουργήσει ενιαίο Ηλεκτρονικό Σύστημα Προέγκρισης (Σ.Η.Π) για την ηλεκτρονική διαχείριση και εξέταση των αιτημάτων αναγκαιότητας αποζημίωσης φαρμάκων για τα οποία ο Ε.Ο.Π.Υ.Υ. προβαίνει σε διαπίστωση σύμφωνα με τις σχετικές διατάξεις και τα οποία εμπίπτουν στις ακόλουθες κατηγορίες:

- α) Ακριβά φάρμακα για συγκεκριμένες παθήσεις παρ. Δαπάνες των ακόλουθων ειδών ακριβών φαρμάκων: Ακριβά φάρμακα για συγκεκριμένες παθήσεις που εμπίπτουν στην κατηγορία υψηλού κόστους του άρθρου 2 του άρθρου 12 του ν. 12.
- β) Φάρμακα που δεν είναι διαθέσιμα στην Ελλάδα (εκτός Ελλάδας) και θα χορηγηθούν με επείγουσες εισαγωγές ή ειδικά αιτήματα.
- γ) Φάρμακα που χρησιμοποιούνται για σκοπούς μη εγκεκριμένους από τον ΕΟΦ.
- δ) Φάρμακα που δεν περιλαμβάνονται στον κατάλογο των αποζημιούμενων φαρμάκων (θετικός κατάλογος), τα οποία δεν έχουν υποβληθεί σε αξιολόγηση, ζητείται να χορηγηθούν επειγόντως για ασθένειες ή παθολογικές καταστάσεις που απειλούν άμεσα τη ζωή του ασθενούς ή έχουν τη δυνατότητα να προκαλέσουν μόνιμη βλάβη στην υγεία του.
- ε) Φάρμακα έγκαιρης πρόσβασης για τα οποία ο Ε.Ο.Φ. ζητά προσωρινή ατομική άδεια αλλά δεν παραδίδονται δωρεάν από τον ΚΑΚ ή τον τοπικό αντιπρόσωπο.

2. Η υποβολή αιτημάτων στον Σ.Η.Π και η επεξεργασία τους εμπίπτουν στην κατηγορία της επεξεργασίας ειδικών κατηγοριών δεδομένων προσωπικού χαρακτήρα χωρίς τη

συγκατάθεση του υποκειμένου των δεδομένων που απαιτείται για λόγους δημόσιου συμφέροντος στον τομέα της δημόσιας κοινωνικής ασφάλισης. Η εν λόγω επεξεργασία υπόκειται σε κατάλληλες και ειδικές εγγυήσεις που διασφαλίζουν την προστασία των δικαιωμάτων και των ελευθεριών των φυσικών προσώπων. Απαγορεύεται ρητά σε οποιονδήποτε τρίτο, εκτός του Οργανισμού, να επεξεργάζεται δεδομένα προσωπικού χαρακτήρα για άλλους σκοπούς ως αποτέλεσμα της επεξεργασίας των προαναφερθέντων δεδομένων που σχετίζονται με την υγεία για λόγους δημόσιου συμφέροντος. Οι υπεύθυνοι επεξεργασίας ή οι εκτελούντες την επεξεργασία των δεδομένων πρέπει να τηρούν το επαγγελματικό απόρρητο (Παρασύρη, 2017).

5. Μοντέλο «Θεραπεία ορμονικής υποκατάστασης και θνησιμότητα»

Η μελέτη με τίτλο «Θεραπεία ορμονικής υποκατάστασης και θνησιμότητα» αφορά μια αναδρομική μελέτη “συνομάδας” (Retrospective Cohort Study) και έχει αγγλικό τίτλο "The effect of hormone replacement therapy on the survival of UK women: a retrospective cohort study 1984–2017". Πραγματοποιήθηκε από ερευνητές του University of East Anglia στο Ηνωμένο Βασίλειο.

5.1 Σχεδιασμός Μελέτης

Ο τύπος μελέτης που πραγματοποιήθηκε για τον σχεδιασμό της μελέτης ονομάζεται Matched Cohort Study και ο πληθυσμός που έλαβε μέρος ήταν ίσος με 105.199 χρήστες ορμονικής αντικατάστασης (Hormone Replacement Therapy - HRT) όσον αφορά τα περιστατικά και 224.643 μη χρήστες όσον αφορά τα ελεγκτικά περιστατικά, που ταιριάχτηκαν με βάση την ηλικία και τη γενική πρακτική.

5.2 Μέθοδοι Ανάλυσης

Το στατιστικό μοντέλο που πραγματοποιήθηκε ως μεθοδολογία ανάλυσης ονομάζεται Weibull Double-Cox και αφορά μοντέλα παλινδρόμησης (regression models). Οι μεταβλητές που έλαβαν μέρος στην ανάλυση ήταν η ηλικία στην πρώτη θεραπεία, η συνομάδα γεννήσεως, ο διαβήτης τύπου 2, η υπέρταση και θεραπεία για υπέρταση, η στεφανιαία νόσος, η οοφορεκτομή/υστερεκτομή, ο δείκτης μάζας σώματος, το κάπνισμα και η κατάσταση ένδειας. Τα κύρια κριτήρια αξιολόγησης αποτελούνταν από την συνολική θνησιμότητα εξαιτίας όλων των αιτιών.

5.2.1 Είδος στατιστικής μεθόδου

Η μελέτη χρησιμοποιεί μια αναδρομική μελέτη συνομάδας για να εκτιμήσει τον κίνδυνο θνησιμότητας σε γυναίκες που υποβλήθηκαν σε ορμονική αντικατάσταση (HRT). Αυτός ο τύπος μελέτης επιτρέπει την ακριβή εκτίμηση των σχετικών κινδύνων, λαμβάνοντας υπόψη πολλαπλούς παράγοντες που μπορεί να επηρεάσουν το αποτέλεσμα.

5.2.2 Στατιστικοί έλεγχοι

Η ανάλυση χρησιμοποίησε μοντέλα παλινδρόμησης Weibull Double-Cox για την εκτίμηση των κινδύνων. Αυτά τα μοντέλα απαιτούν την επαλήθευση ορισμένων υποθέσεων, όπως η προφορικότητα των κινδύνων και η ανεξαρτησία των παρατηρήσεων. Η παλινδρόμηση Weibull Double-Cox, που μερικές φορές αναφέρεται ως μοντέλο Double Weibull, είναι μια στατιστική μέθοδος που χρησιμοποιείται για την ανάλυση επιβίωσης, ιδιαίτερα στην ιατρική έρευνα και τη μηχανική αξιοπιστίας. Επεκτείνει το μοντέλο αναλογικών κινδύνων Cox επιτρέποντας δύο διαφορετικές βασικές λειτουργίες κινδύνου. Σε αυτό το μοντέλο, η συνάρτηση κινδύνου ορίζεται ως το γινόμενο δύο κατανομών Weibull, η καθεμία με τις δικές της παραμέτρους σχήματος και κλίμακας. Η συνάρτηση επιβίωσης για το μοντέλο παλινδρόμησης Weibull Double-Cox μπορεί να εκφραστεί ως:

$$S(t) = \left(-\left(\frac{t}{\lambda_1}\right)^{a_1} - \left(\frac{t}{\lambda_2}\right)^{a_2} \right)$$

$S(t)$ είναι η πιθανότητα επιβίωσης τη δεδομένη στιγμή λ_1 και λ_2 , οι παράμετροι κλίμακας για τις δύο κατανομές Weibull, και a_1 και a_2 , οι παράμετροι σχήματος για τις δύο κατανομές Weibull.

5.3 Αποτελέσματα

Τα αποτελέσματα που προκύπτουν από την στατιστική ανάλυση αφορούν τους ρυθμιζόμενους συντελεστές κινδύνου (Hazard Ratios - HRs), όπου εντοπίστηκαν διαφορετικοί συντελεστές για διαφορετικές ηλικιακές ομάδες χρηστών και διαφορετικούς τύπους θεραπείας ορμονικής αντικατάστασης (HRT).

5.4 Συμπέρασμα

Το μοντέλο παλινδρόμησης που χρησιμοποιείται σε αυτή τη μελέτη είναι ιδιαίτερα χρήσιμο για δεδομένα από το χρόνο μέχρι το συμβάν όπου η συνάρτηση κινδύνου έχει

ένα συγκεκριμένο σχήμα που μπορεί να μοντελοποιηθεί από την κατανομή Weibull. Το τμήμα Double-Cox επιτρέπει τη συμπερίληψη χρονικά μεταβλητών συμμεταβλητών, καθιστώντας το έτσι πιο ευέλικτο στο χειρισμό κλινικών δεδομένων πραγματικού κόσμου. Το συγκεκριμένο μοντέλο είναι κατάλληλο για την εκτίμηση του κινδύνου θνησιμότητας σε γυναίκες που υποβάλλονται σε θεραπεία ορμονικής υποκατάστασης (HRT), καθώς λαμβάνει υπόψη τόσο τους βιολογικούς όσο και τους χρονικούς παράγοντες. Το μοντέλο συζητείται λεπτομερώς στην αρχική δημοσίευση, όπου χρησιμεύει ως ακρογωνιαίος λίθος για ολόκληρη την ανάλυση.

6. Μοντέλο «Επιβίωση ατόμων με σακχαρώδη διαβήτη τύπου 2 στο Η.Β.»

Η μελέτη με τίτλο «Επιβίωση ατόμων με σακχαρώδη διαβήτη τύπου 2 στο Ηνωμένο Βασίλειο» αφορά ακόμη μια αναδρομική μελέτη “συνομάδας” (Retrospective Cohort Study) και έχει αγγλικό τίτλο " On the survival of individuals diagnosed with type 2 diabetes mellitus in the United Kingdom: A retrospective matched cohort study". Κι αυτή η μελέτη πραγματοποιήθηκε από ερευνητές του University of East Anglia στο Ηνωμένο Βασίλειο.

6.1 Σχεδιασμός Μελέτης

Ο τύπος μελέτης που πραγματοποιήθηκε για τον σχεδιασμό της μελέτης ονομάζεται Retrospective Matched Cohort Study και ο πληθυσμός που έλαβε μέρος αφορούσε άτομα που ήταν γεννημένα μεταξύ του 1930 και 1960, τα οποία είχαν διαγνωστεί με διαβήτη τύπου 2 (T2DM), μεταξύ του 2000 και 2016, και είχαν ηλικία μεταξύ 50 και 74 ετών.

6.2 Μέθοδοι Ανάλυσης

Το στατιστικό μοντέλο που πραγματοποιήθηκε ως μεθοδολογία ανάλυσης ονομάζεται Gompertz Double-Cox. Οι μεταβλητές που έλαβαν μέρος στην ανάλυση ήταν το ιατρικό ιστορικό, οι κοινωνιο-δημογραφικοί παράγοντες και ο τρόπος ζωής.

6.2.1 Είδος στατιστικής μεθόδου

Η μελέτη χρησιμοποιεί μια αναδρομική μελέτη συνομάδας για να εκτιμήσει τους κινδύνους θνησιμότητας σε άτομα που έχουν διαγνωστεί με διαβήτη τύπου 2 (T2DM). Αυτός ο τύπος μελέτης παρέχει τη δυνατότητα για ακριβείς εκτιμήσεις των σχετικών κινδύνων, λαμβάνοντας υπόψη πολλαπλούς επηρεαστικούς παράγοντες.

6.2.2 Στατιστικοί έλεγχοι

Η ανάλυση χρησιμοποίησε μοντέλα παλινδρόμησης Gompertz Double-Cox για την εκτίμηση των κινδύνων θνησιμότητας. Το μοντέλο παλινδρόμησης Gompertz Double-Cox, που μερικές φορές αναφέρεται ως μοντέλο Gompertz-Makeham, είναι μια επέκταση του κλασικού μοντέλου Gompertz. Το μοντέλο Gompertz χρησιμοποιείται στη μοντελοποίηση των ποσοστών θνησιμότητας και στην ανάλυση παραγόντων που επηρεάζουν τον κίνδυνο θνησιμότητας στους πληθυσμούς.

Σε αυτό το μοντέλο, η συνάρτηση κινδύνου (η στιγμιαία πιθανότητα να συμβεί ένα συμβάν σε μια δεδομένη στιγμή) εκφράζεται ως το άθροισμα δύο όρων: του όρου Gompertz και του όρου Makeham. Ο όρος Gompertz περιγράφει την εκθετική αύξηση του κινδύνου θνησιμότητας με την ηλικία, ενώ ο όρος Makeham αντιπροσωπεύει έναν επιπλέον βασικό κίνδυνο θνησιμότητας που δεν εξαρτάται από την ηλικία.

Μαθηματικά, η συνάρτηση κινδύνου ($h(t)$) στο μοντέλο παλινδρόμησης Gompertz Double-Cox δίνεται από:

$$h(t) = ae^{\beta t} + \gamma$$

Όπου: Τα a και β είναι παράμετροι του όρου Gompertz, που καθορίζουν το ρυθμό αύξησης της θνησιμότητας με την ηλικία.

Το γ είναι η παράμετρος του όρου Makeham, που αντιπροσωπεύει τον βασικό κίνδυνο θνησιμότητας και το t , είναι η μεταβλητή ηλικίας ή χρόνου.

6.3 Αποτελέσματα

Τα αποτελέσματα που προκύπτουν από την στατιστική ανάλυση αφορούν τους συντελεστές κινδύνου (Hazard Ratios - HRs), όπου εντοπίστηκαν διαφορετικοί συντελεστές για διαφορετικές ηλικιακές ομάδες χρηστών και συνομάδες γεννήσεων.

6.4 Συμπέρασμα

Το μοντέλο που χρησιμοποιείται σε αυτή τη μελέτη είναι ιδιαίτερα χρήσιμο για την εκτίμηση κινδύνου σε αυτήν την αναδρομική μελέτη. Η λειτουργία Gompertz είναι γνωστή για τη μοντελοποίηση των ποσοστών θνησιμότητας και των διαδικασιών γήρανσης, καθιστώντας την κατάλληλη για μια μελέτη που επικεντρώνεται στον Σακχαρώδη Διαβήτη Τύπου 2 και τη θνησιμότητα. Το τμήμα Double-Cox επιτρέπει

την προσαρμογή των χρονικά μεταβλητών συμμεταβλητών, καθιστώντας το μοντέλο εξαιρετικά προσαρμόσιμο σε πολύπλοκα σύνολα δεδομένων που περιλαμβάνουν πολλούς παράγοντες. Το μοντέλο αναλύεται στην αρχική δημοσίευση, λειτουργώντας ως η βάση για την εκτίμηση κινδύνου και τη στατιστική ανάλυση.

Συμπεράσματα

Η ουσιαστική και επείγουσα ανάγκη μετασχηματισμού όλων των λειτουργιών σε μια πιο αυτοματοποιημένη ψηφιακή εποχή έχει προκύψει από την εποχή της ψηφιοποίησης. Πιο συγκεκριμένα, όλες οι τεχνολογικές εξελίξεις συνεισφέρουν πληροφορίες, οι οποίες στη συνέχεια συγκεντρώνονται σε μια ενιαία βασική πηγή δεδομένων για τη δημιουργία αυτού που είναι γνωστό ως Big Data, μια τεχνολογική εξέλιξη που έχει βοηθήσει πολλές επιχειρήσεις να επιτύχουν τους στόχους τους πιο γρήγορα και εύκολα από ό,τι είχαν αρχικά σχεδιάσει. Παράλληλα, επισημαίνεται και η αποτελεσματική λειτουργία των κρατών που χρησιμοποιούν πληροφορίες για τη λειτουργία και τη διαχείριση ολοκληρωμένων κρατών, εκτός από τη συμβολή και τη βοήθειά τους σε γενικούς οργανισμούς. Με βάση αυτές τις εκθέσεις, είναι ζωτικής σημασίας να καλύπτονται όλα τα όρια, διότι με τον τρόπο αυτό οι πηγές δεδομένων θα μπορέσουν να συγκεντρώσουν δεδομένα που, στην αρχή, θα είναι χρήσιμα για τη λειτουργία ενός κράτους, ενός οργανισμού ή ακόμη και σε επίπεδο βελτίωσης της απόδοσης των εργαζομένων. Προκειμένου να αποφευχθούν πιθανοί περιορισμοί και κίνδυνοι, το σύστημα που βοηθά στην επεξεργασία και ανάλυση δεδομένων πρέπει να βρίσκεται συνεχώς ένα βήμα μπροστά. Όλα τα στελέχη που συμβάλλουν στην αποτελεσματική διοίκηση της επιχείρησης συμμετέχουν στη στρατηγική διαχείριση, όπως και όλα τα στοιχεία και οι στατιστικές που θα δημιουργήσουν κέρδη αμέσως, αντιμετωπίζοντας όλες τις οργανωτικές επιθυμίες και στόχους. Η μελλοντική παροχή υγειονομικής περίθαλψης θα επηρεαστεί σίγουρα σε μεγάλο βαθμό από τα μεγάλα δεδομένα, ιδίως από την ανάλυσή τους, η οποία επιτρέπει τη χρήση ενός ευρέος φάσματος διαφορετικών, οργανωμένων και μη δομημένων πηγών δεδομένων. Μια ποικιλία δυνατοτήτων ανάλυσης χρησιμοποιείται ήδη, βοηθώντας στη λήψη αποφάσεων και ενισχύοντας την αποτελεσματικότητα του προσωπικού υγειονομικής περίθαλψης και την πορεία της υγείας των ασθενών. Τα μεγάλα δεδομένα έχουν πολλά

να προσφέρουν στον κλάδο της υγειονομικής περίθαλψης, αλλά παρουσιάζουν και ορισμένα μειονεκτήματα, όπως ανησυχίες σχετικά με την προστασία της ιδιωτικής ζωής, την ακρίβεια και την αξιοπιστία. Επιπλέον, θα πρέπει να αναγνωριστεί ότι τα ηλεκτρονικά δεδομένα υγείας είναι ακόμη ως επί το πλείστον ανεξερεύνητα, καθιστώντας ζωτικής σημασίας τη μετατροπή των επεξεργασμένων δεδομένων σε τεκμηριωμένες αποφάσεις. Πρέπει να δημιουργηθεί μια νέα γενιά μεθόδων επεξεργασίας, εκτός από την προφανή ανάγκη για περισσότερη μελέτη σχετικά με την ομογενοποίηση και την εναρμόνιση των συνεχών και διακριτών μορφών ιατρικών δεδομένων. Υπάρχουν τεράστιες δυνατότητες για αξιοποιήσιμες πληροφορίες που μπορούν να παρασχεθούν από την έρευνα σχετικά με την εξαγωγή βιολογικών δεικτών και "λανθανόντων" προτύπων για την κατανόηση και την πρόβλεψη της εκδήλωσης ασθενειών. Είναι δυνατή η ανάπτυξη αλγορίθμων για την αντιμετώπιση του φιλτραρίσματος των δεδομένων, της παρεμβολής, του μετασχηματισμού, της εξαγωγής χαρακτηριστικών, της επιλογής χαρακτηριστικών και άλλων διαδικασιών. Επιπλέον, υπάρχουν πιθανότητες να ενισχυθούν και να κατασκευαστούν ισχυρά CDSS για κλινική πρόβλεψη, συνταγογράφηση και διάγνωση χάρη στην αξιοπιστία και την πρόοδο των αλγορίθμων μηχανικής μάθησης.

Τα Big Data, τα Open Data και οι βάσεις δεδομένων θα μπορούσαν να αξιοποιηθούν με τους ακόλουθους τρόπους:

- Σημαντικές δυσκολίες αποθήκευσης δεδομένων θα επιλυθούν με τη βελτιστοποίηση της τεχνολογίας υπολογιστικού νέφους, αλλά τα ζητήματα ασφάλειας και εμπιστευτικότητας των δεδομένων πρέπει να αντιμετωπιστούν συνολικά. Ως βασικό δόγμα της ηλεκτρονικής υγείας, ο ηλεκτρονικός φάκελος υγείας θα διασφαλίσει την ποιότητα και θα κάνει συνετή χρήση των ήδη περιορισμένων πόρων ενοποιώντας τις διάσπαρτες πληροφορίες και επιτρέποντας τη δημιουργία μιας διεθνούς και πλήρους παγκόσμιας κοινότητας.

- Προκειμένου να προσεγγιστεί η διαλειτουργικότητα, η δημιουργία παγκόσμιων προτύπων και η ανάπτυξη δεξιοτήτων, θα πρέπει να ενισχυθούν οι διεθνείς σχέσεις στον τομέα της ηλεκτρονικής υγείας.

- Οι μελέτες επιδημιολογίας φαρμάκων συνήθως διεξάγονται ανεξάρτητα με διάφορες μεθοδολογίες και στη συνέχεια τα αποτελέσματα συνδυάζονται με μετα-ανάλυση. Δυστυχώς, πολλές πληροφορίες χάνονται κατά τη διαδικασία αυτή. Η ικανότητά μας να συνδυάζουμε μελέτες και να αξιολογούμε τα αποτελέσματα μπορεί να περιορίζεται από την ετερογένεια του σχεδιασμού και της μεθοδολογίας των

μελετών. Η χρήση ενός τυποποιημένου πρωτοκόλλου για τις πολλές βάσεις δεδομένων θα ήταν μια ασφαλέστερη και πιο αισθητικά ευχάριστη στρατηγική. Στη συνέχεια να πραγματοποιήσουμε τοπική προσαρμογή και ανάλυση. Η στρατηγική αυτή μειώνει τη διαφοροποίηση μεταξύ των επιμέρους μελετών και διευκολύνει τη μετάφραση των αποτελεσμάτων πολλαπλών ερευνών. Επιπλέον, είναι απλούστερος ο συγχρονισμός με τους εθνικούς κανονισμούς που διέπουν τη συλλογή, την αποθήκευση και την επεξεργασία των ευρημάτων. Από την άλλη πλευρά, έχει τον περιορισμό ότι κάθε πηγή δεδομένων πρέπει να είναι σε θέση να διαχειρίζεται και να αναλύει τα δεδομένα και, το πιο κρίσιμο, ότι το υποτιθέμενο κοινό πρωτόκολλο μπορεί να είναι γραμμένο σε διαφορετικές γλώσσες προγραμματισμού ή/και μεταφράσεις.

- Υπάρχουν πολυάριθμες προσεγγίσεις για την προσέγγιση του θέματος του συνδυασμού δεδομένων από ηλεκτρονικές βάσεις δεδομένων σε διάφορα έθνη με απώτερο στόχο την εύρεση πιθανών αρνητικών συνεπειών από τη χρήση ναρκωτικών: i) ομοιόμορφες διαδικασίες, αποτελέσματα και περιγραφές των μεταβλητών παραμέτρων ii) ομοιογένεια της τοπικής επεξεργασίας δεδομένων, όπως αυτή που γίνεται από κάθε εθνικό οργανισμό και iii) κοινή χρήση δεδομένων μέσω μιας κοινής πλατφόρμας μέσω κοινών συγκεντρωτικών αναλύσεων. Πρωτοβουλίες όπως οι EU-ADR, Sentinel, OMOP και PROTECT έχουν δείξει τις τεράστιες δυνατότητες της συγχώνευσης διαφόρων βάσεων δεδομένων, όχι μόνο για τη διεξαγωγή μελετών ασφάλειας αλλά και για τον εντοπισμό ενδείξεων σχετικά με αναμενόμενες ή μη αναμενόμενες επιβλαβείς επιπτώσεις.

- Η έννοια της "έξυπνης υγείας" συνδέεται στενά με εκείνη της "ευεξίας και της φυσικής κατάστασης" και περιλαμβάνει σημαντικό όγκο δεδομένων από κλινικές μελέτες, βιομετρικούς αισθητήρες (όπως θερμοκρασία, καρδιακός ρυθμός, αρτηριακή πίεση κ.λπ.), μεγάλα δεδομένα που σχετίζονται με το γονιδίωμα (όπως αλληλουχία βάσεων DNA, γονιδιακή έκφραση, γονότυποι κ.λπ.), μεγάλα δεδομένα που σχετίζονται με τους παρόχους-πληρωτές (όπως ηλεκτρονικά ιατρικά αρχεία, συνταγές φαρμάκων, ασφαλιστικά δεδομένα, διαγνωστικές εξετάσεις κ.λπ.). Η έξυπνη υγεία είναι, πάνω απ' όλα, ένα στυλ σκέψης και μια προσέγγιση που αναζητά τις συνδέσεις μεταξύ πληροφοριών που φαίνονται να είναι άσχετες μεταξύ τους και τις συνέργειες που, αν δεν υπάρχουν ήδη, πρέπει να οικοδομήσουμε.

- Τα μεγάλα δεδομένα θα είναι πιθανώς πρώτα χρήσιμα σε όσους είναι αφοσιωμένοι στην καινοτομία. Υπάρχουν ορισμένες τακτικές που μπορούν να χρησιμοποιήσουν οι πληρωτές, οι πάροχοι, οι κατασκευαστές φαρμάκων και οι

έμποροι προϊόντων για να ωφελήσουν τον κλάδο της υγειονομικής περίθαλψης χρησιμοποιώντας τα μεγάλα δεδομένα.

- Πρέπει πρώτα να δημιουργηθούν τα θεμέλια για τυποποιημένους ορισμούς. Επί του παρόντος, οι ορισμοί των "αποδείξεων" και της "αξίας" ποικίλλουν ευρέως εντός και μεταξύ των διαφόρων κλάδων της υγειονομικής περίθαλψης. Ως αποτέλεσμα, οι πληρωτές, οι πάροχοι και άλλοι ενδιαφερόμενοι χρησιμοποιούν πολλαπλές μεθόδους ανάλυσης μεγάλων δεδομένων. Ακόμη και οι επιστήμονες ερμηνεύουν ή παρουσιάζουν τα δεδομένα με τρόπο που εξυπηρετεί καλύτερα τους σκοπούς τους.

Βιβλιογραφία

Ahmad Z, Tripathi MM. (2018). Approaches of Big Data in Healthcare: A Critical Review. *International Journal of Advanced Research in Computer Science* ; 9(2): 122-127.

Alexandru AG, Radu MI, Bizon ML. (2018). Big Data in Healthcare - Opportunities and Challenges. *Informatica Economică*; 22(2):43-58.

Alguliyev R.M., Hajirahimova M. Sh., (2014), « “Big Data” phenomenon: Challenges and Opportunities”, *Problems of Information Technology*, No2, Page. 3-16. Barbier G., Liu H., (2011), «Data mining in social media. In Social network data analytics», Page 327-352

Bachhety, S., Kapania, S., & Jain, R. (2021). Big Data Analytics for healthcare: Theory and applications. In *Applications of big data in healthcare* (pp. 45-67). Academic Press.

Belle A, & Thiagarajan R, Soroushmehr R, Navidi F, Beard D, Najarian K. (2015). Big Data Analytics in Healthcare. *BioMed Research International*. Διαθέσιμο στο: https://www.researchgate.net/publication/279198958_Big_Data_Analytics_in_7_Healthcare

Berger ML, Doban V. (2014). Big Data, advanced analytics and the future of comparative effectiveness research. *Journal of Comparative Effectiveness Research*; 167–176.

Burghard Cynthia, (2012), «Big Data and Analytics Key to Accountable Care Success», *IDC Health Insights*.

Cano, I., Tenyi, A., Vela, E., Miralles, F., & Roca, J. (2017). Perspectives on big data applications of health information. *Current Opinion in Systems Biology*, 3, 36-42.

Chen H., Chiang R. H., Storey, V. C., (2012), «Business intelligence and analytics: from big data to big impact, *MIS quarterly*, Page 1165-1188.

Chung W., (2014). «BizPro: Extracting and categorizing business intelligence factors from textual news articles», *International Journal of Information Management*, Vol 34 (Issue 2), Page 272-284.

Chung W., (2014). «BizPro: Extracting and categorizing business intelligence factors from textual news articles», *International Journal of Information Management*, Vol 34 (Issue 2), Page 272-284.

Das, S., Abraham, A., & Panigrahi, B. K. (2010). Computational intelligence: Foundations, perspectives, and recent trends. *Computational Intelligence and Pattern Analysis in Biological Informatics*, 1-37.

Dash S, Shakyawar SK, Sharma M, Kaushik S. (2019). Big Data in healthcare: management, analysis and future prospects. *J Big Data*; 6:54. Διαθέσιμο στο: <https://doi.org/10.1186/s40537-019-0217-0>

Dash, S., Shakyawar, S. K., Sharma, M., & Kaushik, S. (2019). Big data in healthcare: management, analysis and future prospects. *Journal of Big Data*, 6(1), 1-25.

Denis Kaminskiy, (2017), «What's the difference between "Big Data" and Data».

Dontha, R. (2017). LinkedIn. Ανάκτηση από What I Always Wanted To Know About Big Data* (*but was afraid to ask): <https://www.linkedin.com/pulse/what-i-always-wantedknow-big-data-afraid-ask-ramesh-dontha?trk=mp-author-card>

Fatt QK, Ramadas A. (2018). The Usefulness and Challenges of Big Data in Healthcare. *Journal of Healthcare Communications*, 3(2):21. Διαθέσιμο στο: <http://healthcare-communications.imedpub.com/archive.php>

Fernandes L, O'Connor M, Weaver V. (2012). Big Data, bigger outcomes: Healthcare is embracing the Big Data movement, hoping to revolutionize HIM by distilling vast collection of data for specific analysis. *J AHIMA.*; 83(10):38-43.

Garapati, S. L., & Garapati, S. (2018). Application of big data analytics: an innovation in health care. *Comput Intell*, 14(1), 15-27.

Gil Press, (2013), «A very short history of Big Data», Forbes. (Article).

Hahn U., Mani I., (2000), «The challenges of automatic summarization. Computer», Vol 33 (Issue 11), Page 29-36.

Institute for Health Technology Transformation. (2013). Transforming Health Care through Big Data Strategies for leveraging Big Data in the health care industry. Διαθέσιμο στο: http://c4fd63cb482ce6861463-bc6183f1c18e748a49b87a25911a0555.r93.cf2.rackcdn.com/iHT2_BigData_2_013.pdf

Jared Dean, (2014), «Big Data, Data Mining and Machine Learning», Εκδόσεις: John Wiley & Sons, Inc., Hoboken, New Jersey.

Jianqing Fan, Fang Han, Han Liu, (2014), «Challenges of Big Data Analysis», *National Science review*, Vol 1 (Issue 2), Page 293-314.

Jina X., Benjamin W. W., Chenga X., Wanga Y., (2015), Significance and Challenges of Big Data Research, *Big Data Research*, vol 2 (Issue 2), Page 59–64.

Khattak, A. Khan, K. Khan, W. Khan, M. Kamran, and M. Fahad, (2018). “Empirical Analysis of Recent Advances, Characteristics and Challenges of Big Data”, *EAI Endorsed Scal Inf Syst*, vol. 6, no. 23, p. e3.

Kumar, H., & Singh, N. (2017). Review paper on Big Data in healthcare informatics. *International Research Journal of Engineering and Technology*, 4(2), 197-201.

Labrinidis A., Jagadish H. V., (2012), «Challenges and opportunities with big data», Proceedings of the VLDB Endowment, Vol 5 (Issue 12), page 2032-2033.

Lohr Steve, (2014), «For Big-Data Scientists, ‘Janitor Work’ Is Key Hurdle to Insights», New York Times.

Manyika J., Chui M., Brown B., Bughin J., Dobbs R., Roxburgh C., Byers A. H., (2011), «Big data: The next frontier for innovation, competition, and productivity».

Mehta N, Pandit A. (2018). Concurrence of Big Data Analytics and healthcare: A systematic review. *International Journal of Medical Informatics*; 114:57–65.

Minou, J., Mantas, J., Malamateniou, F., & Kaitelidou, D. (2020). Health Professionals’ Perception about Big Data Technology in Greece. *Acta Informatica Medica*, 28(1), 48.

Palanisamy V, Thirunavukarasu R. (2019). Implications of big data analytics in developing healthcare frameworks – A review. *Journal of King Saud University – Computer and Information Sciences*; 31:415–425.

Pan, J., Ding, S., Wu, D., Yang, S., Yang, J., (2019). Exploring behavioural intentions toward smart healthcare services among medical practitioners: a technology transfer perspective. *Int. J. Prod. Res.* 57 (18), 5801–5820.

Patgiri, R., & Ahmed, A. (2016). Big data: The v's of the game changer paradigm. In *2016 IEEE 18th international conference on high performance computing and communications; IEEE 14th international conference on smart city; IEEE 2nd international conference on data science and systems (HPCC/SmartCity/DSS)* (pp. 17-24). IEEE.

Raghupathi W, Raghupathi V. (2014). Big Data Analytics in healthcare: promise and potential. *Health Inf Sci Syst.* ;2:3. doi:10.1186/2047-2501-2-3.

Rouse, M. (2019). *Search Data Management*. Ανάκτηση από Big data: <https://searchdatamanagement.techtarget.com/definition/big-data>

Roy, A. K. (2017). Advances and Scope in Big Data Analytics in Healthcare. *Curr. Trends Biomed. Eng. Biosci*, 9(2), 1-7.

Rud Parr Olivia, (2012), «Business Intelligence Success Factors: Tools for Aligning your business in the global economy», Hoboken, N.J.: Wiley & Sons.

Sarwar MU, Hanify MK, Talibz R, Mobeenx A, Aslam M.(2017). A Survey of Big Data Analytics in Healthcare. (IJACSA) International Journal of Advanced Computer Science and Applications; 8(6):355-359.

Senthilkumar SA, Bharatendara K Rai, Amruta A Meshram, Angappa Gunasekaran, Chandrakumarmangalam S. (2018). Big Data in Healthcare Management: A Review of Literature. *American Journal of Theoretical and Applied Business*; 4(2):57-69.

Shafqat S, Kishwer S, Rasool R, Junaid Q, Tehmina A, Hafiz FA. (2018). Big Data Analytics enhanced healthcare systems: a review. *J Supercomput.* Διαθέσιμο στο: <https://doi.org/10.1007/s11227-017-2222-4>

Subagja, I. K., Amaliyah, N., Hiermy, U., Rahardjo, B. T., Laxmi, E., Lydia, K. S., & Nguyen, P. T. (2019). Evaluation of Big Data analytics in medical science. *Int. J. Eng. Adv. Technol.*, 8(6), 717-720.

Wang L, Alexander CA. (2015). Big Data in Medical Applications and Health Care. *Am. Med. J.*; 6:1–8. doi:10.3844/amjsp.2015.1.8

Wang, Y., L. Kung, and T. A. Byrd. (2018). “Big Data Analytics: Understanding Its Capabilities and Potential Benefits for Healthcare Organizations.” *Technological Forecasting and Social Change* 126: 3– 13. doi:10.1016/j.techfore.2015.12.019.

Watson B. Richard, Ryan J. Peter, (2020), «Big Data Analytics in Australian Local Government», *Smart Cities*, MDPI.

Weill P., Woerner S., (2015), «Thriving in an increasingly digital ecosystem», *MIT Sloan Management Review*, Vol 56 (Issue4), Page 27–34.

Wheelen Thomas. L., Hunger David J., (2012) «Strategic Management and Business Policy», *Toward global sustainability*, 13th edition, Pearson Prentice Hall.

Yousef, M. M. (2021). Big data analytics in health care: a review paper. *International Journal of Computer Science & Information Technology (IJCSIT) Vol, 13*.

Zheng, L., & Guo, L. (2020). Application of Big Data Technology in Insurance Innovation. *Research Gate*. Ανάκτηση από: https://www.researchgate.net/publication/340690220_Application_of_Big_Data_Technology_in_Insurance_Innovation

Παρασύρη, Ε. (2014). *Εξόρυξη Γνώσης και Δεδομένων, Πλεονεκτήματα & Μειονεκτήματα σε μία επιχείρηση*. Ηράκλειο: Τεχνολογικό Εκπαιδευτικό Ίδρυμα Κρήτης.

περιέχει πληροφορίες υγείας για ανώνυμους ασθενείς στην πρωτοβάθμια περίθαλψη του Ηνωμένου Βασιλείου που χρονολογούνται από τη δεκαετία του 1960.