

Ανοικτό Πανεπιστήμιο Κύπρου

Σχολή Θετικών και Εφαρμοσμένων Επιστημών

Μεταπτυχιακή Διατριβή στην Ασφάλεια Υπολογιστών και Δικτύων



**Προκλήσεις και Τεχνικές στη Δικανική
Μεγάλων Δεδομένων**

Ανδρούλλα Χαραλάμπους

**Επιβλέπων Καθηγητής
Δρ. Δημήτριος Μανιαδάκης**

Μάιος 2023

Ανοικτό Πανεπιστήμιο Κύπρου

Σχολή Θετικών και Εφαρμοσμένων Επιστημών

Προκλήσεις και Τεχνικές στη Δικανική Μεγάλων Δεδομένων

Ανδρούλλα Χαραλάμπους

Επιβλέπων Καθηγητής
Δρ. Δημήτριος Μανιαδάκης

Η παρούσα μεταπτυχιακή διατριβή υποβλήθηκε
προς μερική εκπλήρωση των απαιτήσεων για απόκτηση

μεταπτυχιακού τίτλου σπουδών
στην Ασφάλεια Υπολογιστών και Δικτύων

από τη Σχολή Θετικών και Εφαρμοσμένων Επιστημών
του Ανοικτού Πανεπιστημίου Κύπρου

Μάιος 2023

Περίληψη

Η παρούσα μεταπτυχιακή διατριβή μελετά τις προκλήσεις και τις τεχνικές επεξεργασίας και ανάλυσης των Μεγάλων Δεδομένων (Big Data) στη Δικανική (Forensics). Η χρήση των ψηφιακών συσκευών και της τεχνολογίας του διαδικτύου αυξάνεται ραγδαία, οδηγώντας σε αύξηση του όγκου (Volume) των παραγόμενων ψηφιακών δεδομένων. Αυτό δημιούργησε νέες προκλήσεις στον τομέα της ψηφιακής δικανικής, ιδίως στο πλαίσιο του χειρισμού των Μεγάλων Δεδομένων. Η διατριβή αυτή μελετά τις προκλήσεις που αντιμετωπίζουν τα Μεγάλα Δεδομένα στην ψηφιακή δικανική. Η κύρια έμφαση δίνεται στις τεχνικές και τα εργαλεία όπως για παράδειγμα MapReduce, η τεχνητή νοημοσύνη (Artificial Intelligence), οι αλγόριθμοι μηχανικής μάθησης (Machine Learning), όπου μπορούν να χρησιμοποιηθούν για την αντιμετώπιση των προκλήσεων όπως η συλλογή-αποθήκευση/επεξεργασία/ανάλυση/ποιότητα/απόρρητο και ασφάλεια των δεδομένων, την ιδιωτικότητα, το Internet of Things (IoT) κ.α. Η μελέτη αυτή εξετάζει επίσης τα νομικά και ηθικά ζητήματα που σχετίζονται με τη χρήση Μεγάλων Δεδομένων στην ψηφιακή δικανική. Τέλος, η εργασία αυτή έχει ως στόχο να παρέχει ιδέες και συστάσεις για ερευνητές που ασχολούνται με την ψηφιακή δικανική και τον χειρισμό των Μεγάλων Δεδομένων. Σημειώνεται ότι, μέχρι στιγμής δεν υπάρχει κάποια καθιερωμένη μεθοδολογία για τα Μεγάλα Δεδομένα στη Δικανική, αλλά μόνο διάφορα πλαίσια καλές πρακτικές, τα οποία εντοπίζονται και αναδεικνύονται στην παρούσα εργασία.

Λέξεις Κλειδιά:

Μεγάλα Δεδομένα, Ψηφιακή Δικανική, Νομικά και Ηθικά ζητήματα, Ιδιωτικότητα, Internet of Things (IoT), Προκλήσεις, Τεχνικές, MapReduce, Τεχνητή Νοημοσύνη, Αλγόριθμοι Μηχανικής Μάθησης

Summary

This Master's thesis studies the challenges and techniques of Big Data processing and analysis in Forensics. The use of digital devices and internet technology is growing rapidly, leading to an increase in the volume of digital data generated. This has created new challenges in the field of digital forensics, especially in the context of Big Data handling. This thesis studies the challenges faced by Big Data in digital forensics. The focus is on the techniques and tools such as for example MapReduce, Artificial Intelligence, Machine Learning algorithms where they can be used to address the challenges such as data collection/storage/processing/analysis/quality/privacy and security, privacy, Internet of Things (IoT) etc. This study also examines the legal and ethical issues related to the use of Big Data in digital forensics. Finally, this paper aims to provide insights and recommendations for researchers involved in digital forensics and Big Data handling. It is noted that, so far, there is no established methodology for Big Data in Forensics, but only several good practice frameworks, which are identified and highlighted in this paper.

Keywords:

Big Data, Digital Forensics, Legal and Ethical issues, Privacy, Internet of Things (IoT), Challenges, Techniques, MapReduce, Artificial Intelligence, Machine Learning Algorithm

Ευχαριστίες

Πριν την παρουσίαση των αποτελεσμάτων της μεταπτυχιακής διατριβής αισθάνομαι την υποχρέωση να εκφράσω τις θερμές μου ευχαριστίες σε όσους συνέβαλαν στην εκπόνηση αυτής της διατριβής άμεσα ή έμμεσα. Θα ήθελα λοιπόν να ευχαριστήσω τον επιβλέπων καθηγητή μου κύριο Δρ. Δημήτριο Μανιαδάκη ο οποίος συνέβαλε με τα σχόλια, την κριτική όπως επίσης την πολύτιμη συνεργασία και καθοδήγηση του στην ολοκλήρωση της μεταπτυχιακής διατριβής.

Τέλος, θερμές ευχαριστίες οφείλω να εκφράσω στην οικογένειά μου, ειδικότερα στους γονείς μου Ευανθία και Νίκο, στα αδέρφια μου Χαράλαμπο και Κωνσταντίνο καθώς επίσης στους φίλους μου και συναδέλφους για την ενθάρρυνση, υποστήριξη, υπομονή και συμπαράσταση τους κατά τη διάρκεια της εκπόνησης της μεταπτυχιακής μου διατριβής.

Περιεχόμενα

Περίληψη.....	iii
Summary	iv
Ευχαριστίες.....	v
Κεφάλαιο 1.....	1
Εισαγωγή	1
1.1 Κίνητρο	2
1.2 Σκοπός και Στόχοι	3
1.3 Σπουδαιότητα και αναγκαιότητα της Έρευνας	4
1.4 Βασικά ερευνητικά ερωτήματα	5
Κεφάλαιο 2.....	6
Μεθοδολογία	6
Κεφάλαιο 3.....	7
Σχετικές Έννοιες.....	7
3.1 Ψηφιακή Δικανική	7
3.2 Μεγάλα Δεδομένα	9
3.2.1 Παραδείγματα Μεγάλων Δεδομένων.....	10
3.2.1.1 Μέσα Κοινωνικής Δικτύωσης.....	10
3.2.1.2 Ηλεκτρονικό Εμπόριο.....	11
3.2.1.3 Υγειονομική Περίθαλψη.....	11
3.2.1.4 Μεταφορές.....	12
3.2.1.5 Πρόγνωση Καιρού	13
3.2.2 Τύποι Ανάλυσης Μεγάλων Δεδομένων	13
3.2.3 Χαρακτηριστικά Μεγάλων Δεδομένων	15
3.3 Internet of Things (IoT).....	17
Κεφάλαιο 4.....	20
Προκλήσεις στη Δικανική Μεγάλων Δεδομένων.....	20
4.1.1 Συλλογή και αποθήκευση δεδομένων	21
4.1.2 Επεξεργασία δεδομένων.....	22
4.1.3 Ανάλυση και ερμηνεία δεδομένων	22
4.1.4 Ποιότητα δεδομένων	22
4.1.5 Απόρρητο και ασφάλεια δεδομένων.....	22
4.2 Ιδιωτικότητα.....	23
4.3 Ηθικά και Νομικά ζητήματα	24
4.3 Ανωνυμία.....	24

4.4 Δικανική Λογιστική	25
4.5 Δικανική Νέφους - Cloud Forensics	26
4.6 Κατανεμημένο σύστημα αποθήκευσης	27
4.7 Εμπόδια στην ανάπτυξη εφαρμογών των Μεγάλων Δεδομένων.....	28
4.7.1 Αναπαράσταση Δεδομένων	28
4.7.2 Μείωση του Πλεονασμού και Συμπύεση Δεδομένων	28
4.7.3 Διαχείριση του κύκλου ζωής.....	29
4.7.4 Αναλυτικός μηχανισμός.....	29
4.7.5 Εμπιστευτικότητα των Δεδομένων.....	29
4.7.6 Διαχείριση της ενέργειας	30
4.7.7 Αναλωσιμότητα και Επεκτασιμότητα	30
4.7.8 Συνεργασία	30
4.8 Ευκαιρίες των Μεγάλων Δεδομένων.....	31
Κεφάλαιο 5.....	32
Τεχνικές Αντιμετώπισης	32
5.1 Cross Drive.....	32
5.2 Τεχνητή Νοημοσύνη	33
5.3 MapReduce	35
5.4 Αλγόριθμοι Machine learning.....	37
5.4.1 Αλγόριθμος Distributed Random Forest.....	38
5.4.1.1 Χαρακτηριστικά αλγόριθμου Distributed Random Forest.....	39
5.4.1.2 Πλεονεκτήματα αλγόριθμου Distributed Random Forest.....	40
5.4.1.3 Μειονεκτήματα αλγόριθμου Distributed Random Forest	41
5.4.2 Αλγόριθμος Decision Tree	41
5.4.3 Αλγόριθμος Gradient Boosting Machines.....	42
5.4.4 Αλγόριθμος Naive Bayes.....	43
5.5 Εξόρυξη Δεδομένων - Data Mining.....	43
5.6 Τεχνολογία ανοικτού κώδικα - Open-Source technology.....	44
5.6.1 Apache Hadoop.....	45
5.6.2 Apache Spark.....	45
5.6.3 Elastic Stack.....	46
5.6.4 H2O.ai.....	46
Κεφάλαιο 6.....	47
Πίνακας Σύνοψης Ευρημάτων.....	47
Κεφάλαιο 7.....	53
Συμπεράσματα	53

Επίλογος.....	57
Βιβλιογραφία	58

Κεφάλαιο 1

Εισαγωγή

Η συνεχής ψηφιοποίηση των ανθρώπων, λόγω της ραγδαίας χρήσης των υπολογιστών και των προσωπικών συσκευών, οδήγησαν στην αύξηση του ηλεκτρονικού εμπορίου και στη δημιουργία αρκετών ψηφιακών εφαρμογών που χρησιμοποιούν οι άνθρωποι στην καθημερινότητα τους. Ταυτόχρονα όμως, αυξήθηκαν και οι περιπτώσεις ηλεκτρονικών εγκλημάτων και απειλών όπως για παράδειγμα ενέργειες κακόβουλων χρηστών που έχουν απώτερο σκοπό την υποκλοπή προσωπικών δεδομένων, με παράνομους και μη ηθικούς τρόπους. Τα Μεγάλα Δεδομένα είναι ένα αναπόσπαστο μέρος των σύγχρονων πληροφοριακών συστημάτων, όπου παράγουν και συλλέγουν τεράστιες ποσότητες δεδομένων από διάφορες πηγές, όπως τα μέσα κοινωνικής δικτύωσης, τις διαδικτυακές συναλλαγές όπως επίσης και τις συσκευές που είναι συνδεδεμένες στο διαδίκτυο. Ωστόσο, με τον αυξανόμενο όγκο και την πολυπλοκότητα των Μεγάλων Δεδομένων, υπάρχει αυξανόμενη ανάγκη για αποτελεσματικές τεχνικές και μεθόδους ανάλυσης, διερεύνησης και εξαγωγής συμπερασμάτων από αυτές τις τεράστιες ποσότητες δεδομένων. Σε αυτό το σημείο έρχεται η δικανική των Μεγάλων Δεδομένων, η οποία είναι η διαδικασία συλλογής, ανάλυσης και ερμηνείας συνόλων δεδομένων μεγάλης κλίμακας για τον εντοπισμό και την πρόληψη του εγκλήματος στον κυβερνοχώρο ή άλλων παράνομων δραστηριοτήτων.

Η αυξανόμενη χρήση ψηφιακών συσκευών και διαδικτύου οδηγεί σε μια τεράστια ποσότητα δεδομένων που κάθε μέρα πολλαπλασιάζεται με ραγδαίους ρυθμούς. Αυτά τα δεδομένα μπορούν να είναι χρήσιμα για ανακρίσεις σε ποινικές υποθέσεις, αλλά ο μεγάλος όγκος και η πολυπλοκότητα τους δημιουργεί σημαντικές προκλήσεις. Η διατριβή αυτή επικεντρώνεται στις προκλήσεις και στις τεχνικές που εμπλέκονται στην δικανική Μεγάλων Δεδομένων, με στόχο τον προσδιορισμό των πιο αποτελεσματικών μεθόδων και στρατηγικών για τη διεξαγωγή εγκληματολογικών ερευνών σε σύνολα δεδομένων μεγάλης κλίμακας. Μέσω μιας ολοκληρωμένης ανασκόπησης της βιβλιογραφίας, η παρούσα διατριβή επιδιώκει να παράσχει μια λεπτομερή ανάλυση της τρέχουσας κατάστασης της δικανικής Μεγάλων Δεδομένων και επισημαίνει τους

βασικούς τομείς στους οποίους απαιτείται περαιτέρω έρευνα. Στόχος της εν λόγω διατριβής, είναι να συμβάλει στην καλύτερη κατανόηση των προκλήσεων και των ευκαιριών στην δικανική Μεγάλων Δεδομένων και να παρέχει πολύτιμες πληροφορίες για τους ερευνητές καθώς και για τους υπεύθυνους χάραξης πολιτικής στον τομέα της κυβερνοασφάλειας.

1.1 Κίνητρο

Τα τελευταία χρόνια, ο πολλαπλασιασμός των ψηφιακών συσκευών και ο τεράστιος όγκος δεδομένων που παράγουν, δημιουργούν σημαντικές προκλήσεις για την ψηφιακή δικανική. Η χρήση των Μεγάλων Δεδομένων στην ψηφιακή δικανική γίνεται όλο και πιο πολύπλοκη, απαιτώντας εξειδικευμένες τεχνικές και εργαλεία για τη διαχείριση τους. Καθώς η τεχνολογία εξελίσσεται, εμφανίζονται παράλληλα νέες ευκαιρίες.

Η αυξανόμενη χρήση ψηφιακών συσκευών στην καθημερινή ζωή δημιουργεί έναν τεράστιο όγκο ψηφιακών δεδομένων, τα οποία μπορούν να χρησιμοποιηθούν για τη διερεύνηση εγκλημάτων, την ταυτοποίηση υπόπτων και την παροχή αποδεικτικών στοιχείων στο δικαστήριο. Ωστόσο, η διαχείριση και η ανάλυση αυτών των δεδομένων μπορεί να είναι ένα δύσκολο έργο, ιδίως όταν πρόκειται για μεγάλες ποσότητες μη δομημένων δεδομένων.

Στόχος της παρούσας έρευνας είναι να παράσχει μια ολοκληρωμένη επισκόπηση των προκλήσεων και των τεχνικών που σχετίζονται με τη διαχείριση και την ανάλυση Μεγάλων Δεδομένων στην ψηφιακή δικανική. Αυτό περιλαμβάνει, τη διερεύνηση των νομικών και δεοντολογικών ζητημάτων που αφορούν τη συλλογή και ανάλυση ψηφιακών δεδομένων, καθώς και των εξειδικευμένων τεχνικών και εργαλείων που είναι διαθέσιμα για τη διαχείριση και ανάλυση των δεδομένων.

Με τη διερεύνηση αυτών των ζητημάτων, η παρούσα έρευνα επιδιώκει να συμβάλει στην ανάπτυξη αποτελεσματικών και υπεύθυνων πρακτικών ψηφιακής δικανικής που μπορούν να χρησιμοποιηθούν για τη διαχείριση και την ανάλυση Μεγάλων Δεδομένων. Αυτό περιλαμβάνει τον εντοπισμό βέλτιστων πρακτικών, τάσεων

και αναδυόμενων τεχνολογιών που μπορούν να χρησιμοποιηθούν για την αποτελεσματική διαχείριση και ανάλυση ψηφιακών δεδομένων με νόμιμο και δεοντολογικό τρόπο.

Εν κατακλείδι, ο σκοπός της έρευνας των προκλήσεων και των τεχνικών των Μεγάλων Δεδομένων στην ψηφιακή δικανική είναι να διασφαλιστεί ότι προσφέρονται οι γνώσεις και τα εργαλεία που απαιτούνται για την αποτελεσματική διαχείριση και ανάλυση ψηφιακών δεδομένων, εξασφαλίζοντας παράλληλα τη συμμόρφωση με τους σχετικούς νόμους και κανονισμούς. Μέσω αυτής της έρευνας, ο τομέας της ψηφιακής δικανικής μπορεί να συνεχίσει να εξελίσσεται και να προσαρμόζεται στις προκλήσεις και τις ευκαιρίες που παρουσιάζει ο πολλαπλασιασμός των ψηφιακών συσκευών και των Μεγάλων Δεδομένων. Στη συνέχεια παρουσιάζονται οι σκοποί και οι στόχοι της μεταπτυχιακής διατριβής.

1.2 Σκοπός και Στόχοι

Ο σκοπός της παρούσας διατριβής είναι η διερεύνηση των προκλήσεων και των τεχνικών που σχετίζονται με τα Μεγάλα Δεδομένα στην ψηφιακή δικανική. Για την επίτευξη αυτού του σκοπού, προσδιορίζονται οι ακόλουθοι στόχοι. Αρχικά, καθορίζονται οι κύριες προκλήσεις που σχετίζονται με τη διαχείριση και την ανάλυση Μεγάλων Δεδομένων στην ψηφιακή δικανική, συμπεριλαμβανομένης της πολυπλοκότητας των δεδομένων, του όγκου των δεδομένων και της ποικιλίας των πηγών δεδομένων. Επίσης, στόχος της διατριβής αυτής είναι η επισήμανση των νομικών και ηθικών ζητημάτων που αφορούν τη συλλογή και ανάλυση των δεδομένων για εγκληματολογικούς σκοπούς και ο προσδιορισμός των βημάτων για τη διασφάλιση της συμμόρφωσης με τους σχετικούς νόμους και κανονισμούς.

Επιπλέον, γίνεται μελέτη των τρεχουσών τάσεων και των σύγχρονων τεχνικών για τη διαχείριση και ανάλυση Μεγάλων Δεδομένων στην ψηφιακή δικανική, συμπεριλαμβανομένων των πλαισίων, των μοντέλων και των εργαλείων όπως το MapReduce, η μηχανική μάθηση, τεχνητή νοημοσύνη κ.α.. Έπειτα, αξιολογείται η αποτελεσματικότητα αυτών των τεχνικών στη διαχείριση Μεγάλων Δεδομένων στην ψηφιακή δικανική.

Κατά συνέπεια, ο σκοπός και οι στόχοι της παρούσας διατριβής αποσκοπούν στη συμβολή στην ανάπτυξη αποτελεσματικών και υπεύθυνων πρακτικών ψηφιακής δικανικής που μπορούν να χρησιμοποιηθούν για τη διαχείριση καθώς και την ανάλυση Μεγάλων Δεδομένων με νόμιμο και ηθικό τρόπο. Με την επίτευξη αυτών των στόχων, η παρούσα έρευνα μπορεί να συμβάλει στην πρόοδο του τομέα της ψηφιακής δικανικής και στην υποστήριξη της διερεύνησης εγκλημάτων στην ψηφιακή εποχή.

1.3 Σπουδαιότητα και αναγκαιότητα της Έρευνας

Η αναγκαιότητα της έρευνας εστιάζεται στην ανάγκη καταγραφής των σύγχρονων προκλήσεων που απορρέουν από τα Μεγάλα Δεδομένα τα οποία επηρεάζουν την ψηφιακή δικανική. Επίσης, γίνεται αναφορά στην αδυναμία των υφιστάμενων εργαλείων να διαχειριστούν τα Μεγάλα Δεδομένα, καθώς και στην απαίτηση χάραξης ενός οδικού χάρτη για τη δικανική διερεύνηση των Μεγάλων Δεδομένων όπου θα συμπεριλαμβάνονται και προτεινόμενες τεχνικές για τον χειρισμό τους.

Τα Μεγάλα Δεδομένα γίνονται όλο και πιο διαδεδομένα στην καθημερινή μας ζωή και, ως εκ τούτου, η ανάγκη για αποτελεσματική δικανική Μεγάλων Δεδομένων αυξάνεται ραγδαία. Με τις τεράστιες ποσότητες δεδομένων που παράγονται από διάφορες πηγές, υπάρχει μεγαλύτερος κίνδυνος εγκλημάτων στον κυβερνοχώρο και δόλιων δραστηριοτήτων. Συνεπώς, υπάρχει επιτακτική ανάγκη για έρευνα που μπορεί να βοηθήσει στον εντοπισμό και την πρόληψη αυτών των απειλών.

Η έρευνα επίσης επισημάνει την ανάπτυξη νέων πολιτικών και κανονισμών που είναι προσαρμοσμένες στις προκλήσεις της δικανικής Μεγάλων Δεδομένων. Καθώς η χρήση των Μεγάλων Δεδομένων συνεχίζει να αυξάνεται, οι υπεύθυνοι χάραξης πολιτικής πρέπει να αναπτύξουν νέους κανόνες και κανονισμούς για να διασφαλίσουν ότι η διαχείριση των δεδομένων γίνεται με ασφάλεια και νομιμότητα. Η έρευνα σε αυτόν τον τομέα μπορεί να προσφέρει γνώσεις σχετικά με τις προκλήσεις της δικανικής Μεγάλων Δεδομένων και να ενημερώσει για την ανάπτυξη νέων πολιτικών και κανονισμών.

1.4 Βασικά ερευνητικά ερωτήματα

Η μελέτη αυτή στοχεύει να εξετάσει τα ακόλουθα ερευνητικά ερωτήματα:

- Τι προβλήματα δημιουργούνται από τα Μεγάλα Δεδομένα στην ψηφιακή δικανική ανάλυση;
- Ποιες τάσεις έχουν δημιουργηθεί για την “state-of-the-art ” διαχείριση των Μεγάλων Δεδομένων στην ψηφιακή δικανική ανάλυση;
- Ποια τα πλαίσια - μοντέλα - τεχνικές αντιμετώπισης των Μεγάλων Δεδομένων (MapReduce , Machine Learning, κ.λπ.) από τη σκοπιά της ψηφιακής δικανικής;
- Ποιος διαφαίνεται ως ο οδικός χάρτης για τον αποτελεσματικό χειρισμό των Μεγάλων Δεδομένων στην ψηφιακή δικανική;
- Ποιες είναι οι ευκαιρίες των Μεγάλων Δεδομένων στη ψηφιακή δικανική;
- Πώς επηρεάζουν νομικά και ηθικά ζητήματα τη συλλογή και ανάλυση δεδομένων μεγάλου όγκου για εγκληματολογικούς σκοπούς και ποια μέτρα μπορούν να ληφθούν για να εξασφαλιστεί η συμμόρφωση με τους σχετικούς νόμους και κανονισμούς;

Κεφάλαιο 2

Μεθοδολογία

Η μεθοδολογία για τη διατριβή σχετικά με τις προκλήσεις και τις τεχνικές της ψηφιακής δικανικής των Μεγάλων Δεδομένων θα διεξαχθεί με τη χρήση μιας προσέγγισης βιβλιογραφικής ανασκόπησης αφηγηματικού τύπου. Η ανασκόπηση ακολουθεί τις κατευθυντήριες γραμμές και τις συστάσεις που περιγράφονται στο άρθρο της Rossella Ferrari (December, 2015).

Η ανασκόπηση ξεκινά με μία ενδελεχή αναζήτηση σε σχετικές ακαδημαϊκές βάσεις δεδομένων, συμπεριλαμβανομένων των IEEE Xplore, ResearchGate και ScienceDirect, χρησιμοποιώντας τις κατάλληλες λέξεις-κλειδιά και φίλτρα. Επιπλέον, καθορίζονται τα κριτήρια συμπερίληψης και αποκλεισμού ώστε να διασφαλιστεί ότι οι επιλεγμένες μελέτες ανταποκρίνονται στους ερευνητικούς στόχους και απαντούν στα ερευνητικά ερωτήματα.

Οι επιλεγμένες μελέτες αξιολογούνται κριτικά και συνοψίζονται με βάση τα θέματα και ευρήματα που σχετίζονται με τις προκλήσεις και τις τεχνικές της ψηφιακής δικανικής Μεγάλων Δεδομένων. Η βιβλιογραφική ανασκόπηση οργανώνεται σε διάφορες ενότητες, συμπεριλαμβανομένης της εισαγωγής, της μεθοδολογίας, των προκλήσεων στην ψηφιακή δικανική Μεγάλων Δεδομένων, των τεχνικών χειρισμού τους καθώς και των ηθικών και νομικών ζητημάτων και των συμπερασμάτων.

Συνολικά, η παρούσα μεθοδολογία αποσκοπεί στην παροχή μιας ολοκληρωμένης, συστηματικής και κριτικής αξιολόγησης της βιβλιογραφίας σχετικά με τις προκλήσεις και τις τεχνικές της ψηφιακής δικανικής των Μεγάλων Δεδομένων.

Κεφάλαιο 3

Σχετικές Έννοιες

Σε αυτό το κεφάλαιο παρουσιάζονται οι σχετικές έννοιες και το θεωρητικό υπόβαθρο, που αποτέλεσε την βάση για διερεύνηση των ερευνητικών ερωτημάτων της μεταπτυχιακή διατριβής, σε σχέση με τις προκλήσεις και τις τεχνικές στη Δικανική Μεγάλων Δεδομένων. Σημείο αφετηρίας αποτελούν οι σχετικές έννοιες για την Ψηφιακή Δικανική, των Μεγάλων Δεδομένων και το Internet of Things (IoT). Στο κεφάλαιο αυτό δίνεται μια συνοπτική επισκόπηση των ορισμών αυτών.

3.1 Ψηφιακή Δικανική

Σύμφωνα με τον Casey (2011) η δικανική υπολογιστών, γνωστή και ως ψηφιακή δικανική, είναι η διαδικασία συλλογής, ανάλυσης και διατήρησης ηλεκτρονικών δεδομένων με σκοπό τη διερεύνηση και πρόληψη ψηφιακών εγκλημάτων, περιστατικών και παραβιάσεων ασφαλείας. Περιλαμβάνει τη χρήση εξειδικευμένων τεχνικών, εργαλείων και μεθόδων για την ανάκτηση, διατήρηση και ανάλυση ηλεκτρονικών δεδομένων από διάφορες ψηφιακές συσκευές, όπως υπολογιστές, smartphones, tablet, διακομιστές και άλλα ψηφιακά μέσα αποθήκευσης. Ο στόχος της δικανικής υπολογιστών είναι η εξακρίβωση των γεγονότων ενός περιστατικού, ο εντοπισμός των εμπλεκόμενων μερών και η παροχή αποδεικτικών στοιχείων που μπορούν να χρησιμοποιηθούν σε νομικές διαδικασίες.

Επιπλέον, η δικανική υπολογιστών είναι «η επιστημονική εξέταση και ανάλυση των δεδομένων που βρίσκονται σε μέσα αποθήκευσης υπολογιστών ή που ανακτώνται από αυτά, κατά τρόπο ώστε οι πληροφορίες να μπορούν να χρησιμοποιηθούν ως αποδεικτικά στοιχεία σε δικαστήριο» Casey (2011). Ακολουθώντας, οι Nelson, Phillips και Steuart (2014) ορίζουν την δικανική των υπολογιστών ως «την απόκτηση, τη διατήρηση και την ανάλυση ψηφιακών αποδεικτικών στοιχείων με τρόπο που να διαφυλάσσεται η ακεραιότητα των αποδεικτικών στοιχείων και να διατηρείται η αλυσίδα φύλαξης για χρήση σε νομικές διαδικασίες».

Η δικανική υπολογιστών είναι ένας διεπιστημονικός τομέας που συνδυάζει πτυχές της επιστήμης των υπολογιστών, του δικαίου και των ερευνητικών τεχνικών. Χρησιμοποιείται σε διάφορα πλαίσια, όπως ποινικές έρευνες, αστικές δίκες, εταιρικές έρευνες και αντιμετώπιση περιστατικών. Καθώς οι ψηφιακές τεχνολογίες γίνονται όλο και πιο διαδεδομένες, η σημασία της δικανικής υπολογιστών στην πρόληψη και τη διερεύνηση ψηφιακών εγκλημάτων συνεχίζει να αυξάνεται.

Η Ψηφιακή Δικανική απαρτίζεται από το κλάδο της επιστήμης των forensics η οποία σχετίζεται με την ανάκτηση και την διερεύνηση πληροφοριών σε ηλεκτρονικές συσκευές και συνδέεται παράλληλα με το ηλεκτρονικό έγκλημα. Η ορολογία της ψηφιακής δικανικής είναι ουσιαστικά το αντικείμενο που ανιχνεύει αποδεικτικά στοιχεία και τεκμήρια που αφορούν κυρίως εγκληματικές πράξεις. Τα εγκλήματα που διερευνώνται διαχωρίζονται σε ψηφιακά όπως κλοπή κωδικών, hacking τραπεζικών λογαριασμών και τα κλασικά εγκλήματα για παράδειγμα ληστείες, δολοφονίες, απαγωγές που πλέον χρησιμοποιούν ως μέσο τους ηλεκτρονικούς υπολογιστές.

Η ψηφιακή δικανική είναι ένας κλάδος της εγκληματολογικής επιστήμης που περιλαμβάνει τον εντοπισμό, τη διατήρηση, την ανάκτηση και την ανάλυση ηλεκτρονικών αποδεικτικών στοιχείων με σκοπό τη διερεύνηση και την πρόληψη ψηφιακών εγκλημάτων, περιστατικών και παραβιάσεων της ασφάλειας. Πρόκειται για μια διαδικασία χρήσης επιστημονικών και τεχνικών μεθόδων για την απόκτηση, εξέταση και ερμηνεία ψηφιακών δεδομένων με τρόπο που να μπορούν να χρησιμοποιηθούν ως αποδεικτικά στοιχεία σε δικαστήριο.

(Casey, E. (2011)

Σύμφωνα με τον Garfinkel (2010), η ψηφιακή δικανική είναι «η χρήση επιστημονικά τεκμηριωμένων και αποδεδειγμένων μεθόδων προς την κατεύθυνση της διατήρησης, συλλογής, επικύρωσης, ταυτοποίησης, ανάλυσης, ερμηνείας, τεκμηρίωσης και παρουσίασης ψηφιακών αποδεικτικών στοιχείων που προέρχονται από ψηφιακές πηγές με σκοπό τη διευκόλυνση ή την προώθηση της ανακατασκευής γεγονότων που κρίνονται εγκληματικά ή τη συμβολή στην πρόβλεψη μη εξουσιοδοτημένων ενεργειών που αποδεικνύεται ότι διαταράσσουν τις προγραμματισμένες λειτουργίες».

3.2 Μεγάλα Δεδομένα

Είναι γεγονός ότι, υπάρχει μια συνεχής αύξηση των δεδομένων που παράγονται μέσω πολλαπλών πηγών, όπως για παράδειγμα αισθητήρες, διαδικασίες ανάλυσης δεδομένων κ.α.. Με βάση τον Σούμπλη, Α. (2021) τα δεδομένα εξακολουθούν να αποτελούν μια από τις μεγαλύτερες προκλήσεις, επικεντρώνοντας στη συλλογή και συγκέντρωση των δεδομένων για την ανάλυση τους. Υπάρχουν διαφορετικά συστήματα για να αποθηκεύουν δεδομένα σε διαφορετικές μορφές. Η έννοια των Μεγάλων Δεδομένων, αναφέρεται κυρίως στην εξέλιξη καθώς και την χρήση τεχνολογιών που δίνονται στον χρήστη.

Στο άρθρο (Riahi, Y., & Riahi, S. (2018)), αναφέρεται ότι τα Μεγάλα Δεδομένα με την πάροδο του χρόνου και με την συνεχή εξέλιξη της τεχνολογίας, ο όγκος των πληροφοριών αυξάνεται συνεχώς. Ο όρος «Μεγάλα Δεδομένα» παραπέμπει στην εξέλιξη και χρήση των τεχνολογιών που προσφέρουν στον χρήστη, τις σωστές πληροφορίες από ένα πλήθος δεδομένων που αυξάνεται δραματικά σε μεγάλο χρονικό διάστημα. Κατόπιν, εμφανίζονται με απώτερο σκοπό την παροχή πρόσβασης σε πραγματικό χρόνο σε τεράστιες βάσεις δεδομένων.

Επιπλέον, τα Μεγάλα Δεδομένα αναφέρονται στα σύνολα δεδομένων που απαιτούν ειδικά εργαλεία αποθήκευσης και ανάλυσης στον υπολογιστή λόγω του όγκου, της ποικιλίας και της ταχύτητάς τους. Η παραγωγή ψηφιακών δεδομένων αυξάνεται συνεχώς λόγω της ανάπτυξης νέων τεχνολογιών, του διαδικτύου και των κοινωνικών δικτύων. Τα δεδομένα περιλαμβάνουν τόσο δομημένα όσο και αδόμητα δεδομένα και τα χαρακτηριστικά τους απαιτούν εξελιγμένα εργαλεία αποθήκευσης και ανάλυσης σε υπολογιστή. Η ανάλυση Μεγάλων Δεδομένων είναι η διαδικασία συλλογής, οργάνωσης και ανάλυσης τους για την αποκάλυψη μοτίβων και γνώσεων. Οι τέσσερις τύποι ανάλυσης Μεγάλων Δεδομένων είναι η περιγραφική, η διαγνωστική, η προγνωστική και η συντακτική ανάλυση.

Είναι σημαντικό για τις εταιρείες να έχουν πρόσβαση σε Μεγάλα Δεδομένα, καθώς τους επιτρέπει να λαμβάνουν αποφάσεις βάσει δεδομένων που υποστηρίζουν τις στρατηγικές τους. Με τον ολοένα αυξανόμενο αριθμό χρηστών του Διαδικτύου και των κινητών τηλεφώνων, ο όγκος των ψηφιακών δεδομένων αυξάνεται ραγδαία. Η σημασία των δεδομένων έγκειται στην αξία και τις δυνατότητες που απορρέουν από τα δεδομένα.

Οι αναλύσεις Μεγάλων Δεδομένων μπορούν να βοηθήσουν τις εταιρείες να κατανοήσουν τη συμπεριφορά των πελατών, τις προτιμήσεις και τις τάσεις της αγοράς και να βελτιστοποιήσουν τις λειτουργίες και τις διαδικασίες τους. Συνοψίζοντας, τα Μεγάλα Δεδομένα παρέχουν στις εταιρείες πολύτιμες πληροφορίες που μπορούν να χρησιμοποιηθούν για τη λήψη αποτελεσματικών αποφάσεων που τελικά οδηγούν στην επιχειρηματική επιτυχία.

3.2.1 Παραδείγματα Μεγάλων Δεδομένων

Τα Μεγάλα Δεδομένα αναφέρονται στον τεράστιο όγκο δομημένων, ημιδομημένων και μη δομημένων δεδομένων που παράγονται από οργανισμούς και άτομα σε καθημερινή βάση. Ορισμένα συνήθη παραδείγματα στους τομείς των Μεγάλων Δεδομένων, περιλαμβάνουν δεδομένα όπως κοινωνικών μέσων, το ηλεκτρονικό εμπόριο, η υγειονομική περίθαλψη, οι μεταφορές και η πρόγνωση καιρού.

3.2.1.1 Μέσα Κοινωνικής Δικτύωσης

Στο άρθρο τους οι George, Haas, και Pentland (2014) αναφέρουν, τα μέσα κοινωνικής δικτύωσης ως μία από τις πηγές Μεγάλων Δεδομένων, η οποία παρέχει ευκαιρίες στους ερευνητές της διοίκησης να αναλύσουν τις κοινωνικές τάσεις. Τα κοινοτικά δεδομένα, συμπεριλαμβανομένων των αξιολογήσεων των καταναλωτών για προϊόντα, των κουμπιών ψηφοφορίας (voting buttons) και των ροών Twitter, μπορούν να αποταχθούν για να εξαχθούν μοτίβα κοινωνικής δομής. Οι επιχειρήσεις χρησιμοποιούν μοντέλα κοινωνικών μέσων για να δημιουργήσουν νέες αγορές και κατηγορίες προϊόντων και να ανταποκριθούν γρήγορα στις αντιδράσεις των καταναλωτών. Επιπλέον, οι ροές Twitter σε πραγματικό χρόνο ή οι αναρτήσεις στο Facebook μπορούν να παρέχουν πληροφορίες σχετικά με την εξάπλωση ασθενειών, τα μοτίβα μετακίνησης και το συναίσθημα και το συναίσθημα της κοινότητας που μπορούν να επηρεάσουν την αντίδραση, τα προϊόντα και τις υπηρεσίες ενός οργανισμού και τη στρατηγική του.

3.2.1.2 Ηλεκτρονικό Εμπόριο

Τα Μεγάλα Δεδομένα παράγονται από διάφορες πηγές, συμπεριλαμβανομένων των συναλλαγών ηλεκτρονικού εμπορίου. Καθώς όλο και περισσότερα δεδομένα που δημιουργούνται από τους χρήστες μεταφέρονται και αναλύονται εντός και μεταξύ διαφορετικών τμημάτων, το ηλεκτρονικό εμπόριο εξαρτάται όλο και περισσότερο από ακριβείς και έγκαιρες υπηρεσίες πληροφοριών. Επιπλέον, οι επιχειρηματίες και οι καινοτόμοι χρησιμοποιούν συγκεντρωτικά ανοικτά και δημόσια δεδομένα για να δημιουργήσουν νέα προϊόντα και υπηρεσίες που έχουν τη δυνατότητα να μεταμορφώσουν τον κλάδο του ηλεκτρονικού εμπορίου. Ωστόσο, πτυχές όπως οι συμφωνίες ανταλλαγής δεδομένων και η προστασία των δεδομένων πρέπει να εξεταστούν προσεκτικά και να αναγνωριστούν ρητά, προκειμένου να παρέχεται καλύτερα η λειτουργική υποδομή για τους παρόχους ηλεκτρονικού εμπορίου. Γενικότερα, τα Μεγάλα Δεδομένα από τις συναλλαγές ηλεκτρονικού εμπορίου θα αλλάξουν το τοπίο της πολιτικής και της έρευνας για το ηλεκτρονικό εμπόριο κατά την επόμενη δεκαετία (George, Haas, και Pentland (2014)).

3.2.1.3 Υγειονομική Περίθαλψη

Η υγειονομική περίθαλψη θεωρείται επίσης έναν από τους τομείς όπου τα Μεγάλα Δεδομένα παράγουν πολλές πληροφορίες. Ο κλάδος της υγειονομικής περίθαλψης παράγει μεγάλους όγκους σύνθετων και ετερογενών δεδομένων από πηγές τόσο διαφορετικές όπως οι σημειώσεις του γιατρού, οι εργαστηριακές εξετάσεις, οι εκθέσεις απεικόνισης, οι ασφαλιστικές εταιρείες, τα μέσα κοινωνικής δικτύωσης και οι αλληλουχίες γονιδιώματος (genomic sequences). Ο μεγάλος όγκος και η πολυπλοκότητα αυτών των δεδομένων καθιστά δύσκολη την αποθήκευση, την επεξεργασία και την ανάλυσή τους με τη χρήση παραδοσιακών μεθόδων και τεχνικών. Ωστόσο, η έλευση των μεγάλων δεδομένων απλοποιεί την πληροφορική καθώς και την ιατρική που βασίζεται σε αποδείξεις και συνειδητή αξία, δημιουργώντας ευκαιρίες για καλύτερη πρόληψη και εξατομικευμένη θεραπεία. Για την επεξεργασία, ανάλυση και αξιοποίηση αυτών των δεδομένων σε πραγματικό χρόνο απαιτούνται καινοτόμα εργαλεία και τεχνικές καθώς και ισχυρές υπολογιστικές τεχνικές.

Επιπρόσθετα, η διασφάλιση της ασφάλειας και της εμπιστευτικότητας των δεδομένων υγειονομικής περίθαλψης, η υιοθέτηση τυποποιημένης ορολογίας υγειονομικής περίθαλψης, η ανάπτυξη λειτουργικών συστημάτων για τη διαχείριση μεγάλου όγκου δεδομένων και η εκπαίδευση των παρόχων υγειονομικής περίθαλψης σχετικά με τη σημασία της υιοθέτησης καινοτόμων εργαλείων (Information and Communications Technologies - ICT), αποτελούν παραδείγματα επιτυχημένων εφαρμογών μεγάλης κλίμακας στον τομέα της υγειονομικής περίθαλψης. Πέρα από τις προκλήσεις, όπως η αντίσταση στην αλλαγή, ο κατακερματισμός των ιατρικών δεδομένων, οι ηθικές ανησυχίες, τα ζητήματα χρηστικότητας και οι ανησυχίες για την ασφάλεια και την προστασία της ιδιωτικής ζωής, η υιοθέτηση μεγάλων δεδομένων έχει τη δυνατότητα να φέρει επανάσταση στο σύστημα υγειονομικής περίθαλψης, να μειώσει το κόστος της υγειονομικής περίθαλψης και να βελτιώσει την ποιότητα της παροχής υγειονομικής περίθαλψης (Olaronke, I., & Oluwaseun, O. (2016, December)).

3.2.1.4 Μεταφορές

Ένα παράδειγμα Μεγάλων Δεδομένων που αφορούν τις μεταφορές είναι το σύστημα παρακολούθησης οδικών δικτύων με smartphone (Smartphone Road Monitoring System), το οποίο χρησιμοποιεί μια μεθοδολογία crowdsourcing για τη συλλογή δεδομένων μεταφορών σε πραγματικό χρόνο, συμπεριλαμβανομένων των συνθηκών κυκλοφορίας και της οδηγικής συμπεριφοράς. Το σύστημα αυτό υιοθετεί τόσο ενεργητικούς όσο και παθητικούς τρόπους συλλογής δεδομένων, όπου οι χρήστες μπορούν να αναφέρουν ρητά τα κυκλοφοριακά συμβάντα, ενώ το σύστημα καταγράφει αυτόματα δεδομένα κυκλοφορίας όπως η θέση, η κατεύθυνση, η ταχύτητα και οι λακκούβες (με την ανίχνευση κραδασμών). Τα συγκεντρωμένα δεδομένα μπορούν να αξιοποιηθούν για τη βελτιστοποίηση των συστημάτων μεταφορών μέσω ολοκληρωμένης ανάλυσης, όπως για παράδειγμα ελαχιστοποιώντας τους χρόνους αναμονής των οχημάτων στα φανάρια, βελτιώνοντας έτσι τη ροή της κυκλοφορίας και μειώνοντας τις εκπομπές ρύπων. Επιπλέον, άλλες πηγές μεγάλων δεδομένων μεταφορών περιλαμβάνουν σήματα κυψελοειδών δικτύων, ειδικούς αισθητήρες που είναι ενσωματωμένοι σε οχήματα και αναρτήσεις στα μέσα κοινωνικής δικτύωσης (Neilson, A., Daniel, B., & Tjandra, S. (2019)).

3.2.1.5 Πρόγνωση Καιρού

Η αξιοποίηση των μεγάλων δεδομένων έχει αποδειχθεί ιδιαίτερα επωφελής στον τομέα της πρόγνωσης καιρού. Οι συγγραφείς Hassani και Silva (2015), εξετάζουν τα δυνητικά οφέλη της αξιοποίησης των Μεγάλων Δεδομένων για την απόκτηση ακριβών μετεωρολογικών προβλέψεων και προσδιορίζουν μια αυξημένη ερευνητική εστίαση στον τομέα αυτό. Παρ' όλα αυτά, αναγνωρίζουν ότι οι προβλέψεις καιρού εξακολουθούν να μην είναι ακριβείς πέραν της εβδομάδας. Ειδικότερα, στο άρθρο υπογραμμίζεται ότι η πρόγνωση του καιρού υπήρξε ένας από τους πρωταρχικούς δικαιούχους των Μεγάλων Δεδομένων και δείχνει περαιτέρω πώς η εφαρμογή των μεγάλων δεδομένων διευκόλυνε την ακριβή πρόγνωση του καιρού. Συμπερασματικά, η εφαρμογή των Μεγάλων Δεδομένων έχει αυξήσει σημαντικά την ακρίβεια και τη χρησιμότητα της πρόγνωσης του καιρού.

3.2.2 Τύποι Ανάλυσης Μεγάλων Δεδομένων

Σύμφωνα με τον Ritesh Pathak (January, 2021) οι τύποι ανάλυσης των Μεγάλων Δεδομένων, διαχωρίζονται σε τέσσερις κατηγορίες:

- **Περιγραφική ανάλυση (Descriptive Analytics):** Η περιγραφική ανάλυση είναι ένα ισχυρό εργαλείο στον τομέα της ανάλυσης δεδομένων, ιδίως για τον εντοπισμό προτύπων σε ένα συγκεκριμένο τμήμα πελατών. Απλοποιώντας και συνοψίζοντας δεδομένα του παρελθόντος σε ευανάγνωστη μορφή, παρέχει πληροφορίες σχετικά με το τι συνέβη στο παρελθόν και βοηθά στη δημιουργία αναφορών όπως τα έσοδα, τα κέρδη και οι πωλήσεις μιας εταιρείας. Οι συνοπτικές στατιστικές, η ομαδοποίηση και οι κανόνες συσχέτισης είναι κοινές τεχνικές που χρησιμοποιούνται στην περιγραφική ανάλυση, ιδίως στην ανάλυση καλαθιού αγοράς. Συνολικά, η περιγραφική ανάλυση διαδραματίζει κρίσιμο ρόλο στην εξαγωγή πολύτιμων πληροφοριών από σύνολα Μεγάλων Δεδομένων και στη διευκόλυνση των διαδικασιών λήψης αποφάσεων για τις επιχειρήσεις.
- **Διαγνωστική ανάλυση (Diagnostic Analytics):** Η διαγνωστική ανάλυση είναι ένα ισχυρό εργαλείο που επιτρέπει στις επιχειρήσεις να κατανοήσουν τη βασική αιτία ενός προβλήματος. Παρέχοντας λεπτομερείς και σε βάθος πληροφορίες,

βοηθά τους επιστήμονες δεδομένων να εντοπίσουν τους λόγους που κρύβονται πίσω από ένα συγκεκριμένο συμβάν. Τεχνικές όπως το drill-down, η εξόρυξη δεδομένων και η ανάκτηση δεδομένων, η ανάλυση του λόγου αποχώρησης και η ανάλυση του σκορ υγείας των πελατών αποτελούν παραδείγματα διαγνωστικής ανάλυσης. Όταν πρόκειται για επιχειρήσεις, η διαγνωστική ανάλυση είναι ιδιαίτερα χρήσιμη για την έρευνα των λόγων που οδηγούν σε δείκτες αποχώρησης και τάσεις χρήσης μεταξύ των πιστών πελατών. Για παράδειγμα, μια εταιρεία ηλεκτρονικού εμπορίου μπορεί να χρησιμοποιήσει τη διαγνωστική ανάλυση για να προσδιορίσει γιατί οι πωλήσεις της έχουν μειωθεί, παρόλο που οι πελάτες προσθέτουν προϊόντα στα καλάθια τους. Αναλύοντας τα δεδομένα, η εταιρεία μπορεί να εντοπίσει πιθανούς λόγους, όπως η μη σωστή φόρτωση της φόρμας, τα υψηλά έξοδα αποστολής και οι περιορισμένες μέθοδοι πληρωμής. Οπλισμένη με αυτές τις πληροφορίες, η εταιρεία μπορεί να επικεντρωθεί στην επίλυση του συγκεκριμένου προβλήματος για να βελτιώσει τις πωλήσεις και την ικανοποίηση των πελατών.

- **Προγνωστική ανάλυση (Predictive Analytics):** Στον τομέα των επιχειρήσεων, η προγνωστική ανάλυση έχει τη δυνατότητα να βοηθήσει τους οργανισμούς να λαμβάνουν καλύτερες αποφάσεις, να μειώνουν τους κινδύνους και να αυξάνουν τα κέρδη τους. Τεχνικές όπως η εξόρυξη δεδομένων, η μηχανική μάθηση και η στατιστική μοντελοποίηση χρησιμοποιούνται για την ανάλυση δεδομένων και τη δημιουργία προγνωστικών μοντέλων. Αυτά τα μοντέλα βοηθούν τους οργανισμούς να εντοπίζουν μοτίβα και σχέσεις στα δεδομένα και να κάνουν ακριβείς προβλέψεις για τα μελλοντικά αποτελέσματα. Για παράδειγμα, μια εταιρεία πιστωτικών καρτών μπορεί να χρησιμοποιήσει την προγνωστική ανάλυση για να εντοπίσει τους πελάτες που είναι πιο πιθανό να αθετήσουν τις πληρωμές τους. Αυτό μπορεί να βοηθήσει την εταιρεία να λάβει προληπτικά μέτρα, όπως η μείωση του πιστωτικού ορίου για τους εν λόγω πελάτες. Γενικότερα, η προγνωστική ανάλυση είναι ένα ισχυρό εργαλείο για τις επιχειρήσεις ώστε να βρίσκονται μπροστά από τον ανταγωνισμό και να λαμβάνουν τεκμηριωμένες αποφάσεις με βάση τα δεδομένα.
- **Καθοδηγητική ανάλυση (Prescriptive Analytics):** Η καθοδηγητική ανάλυση είναι η πιο προηγμένη μορφή ανάλυσης και έχει τεράστιες δυνατότητες να προσφέρει ανταγωνιστικό πλεονέκτημα στις επιχειρήσεις. Με την ικανότητά της

να προτείνει την καλύτερη πορεία δράσης, η καθοδηγητική ανάλυση αποκτά ολοένα και μεγαλύτερη σημασία στον σημερινό κόσμο που βασίζεται στα δεδομένα. Αυτός ο τύπος ανάλυσης χρησιμοποιεί έναν συνδυασμό δεδομένων, επιχειρηματικών κανόνων και μαθηματικών αλγορίθμων για την παροχή συστάσεων σχετικά με την καλύτερη δυνατή δράση. Η ανάλυση αυτή επιτρέπει στις εταιρείες να βελτιστοποιήσουν τις δραστηριότητές τους, να ελαχιστοποιήσουν τους κινδύνους και να αυξήσουν τα κέρδη τους. Λαμβάνει υπόψη τόσο τα εσωτερικά όσο και τα εξωτερικά δεδομένα, συμπεριλαμβανομένων των οργανωτικών εισροών και των πληροφοριών από τα μέσα κοινωνικής δικτύωσης, για να κάνει τις πιο τεκμηριωμένες συστάσεις. Επίσης, είναι ιδιαίτερα χρήσιμη όταν συνδυάζεται με την προγνωστική ανάλυση, καθώς μπορεί να βοηθήσει τις επιχειρήσεις να χειριστούν μελλοντικά συμβάντα για να μετριάσουν τους κινδύνους και να επιτύχουν τους στόχους τους.

3.2.3 Χαρακτηριστικά Μεγάλων Δεδομένων

Τα χαρακτηριστικά των Μεγάλων Δεδομένων ποικίλουν, καθώς αποτελούνται από μια διαφοροποιημένη μάζα δεδομένων. Σύμφωνα με τους Youssra Riahi και Sara Riahi (2018) τα Μεγάλα Δεδομένα είναι ένα ολοένα και πιο διαδεδομένο φαινόμενο στη σημερινή εποχή, το οποίο χαρακτηρίζεται από μοναδικά χαρακτηριστικά που το διαφοροποιούν από τα παραδοσιακά σύνολα δεδομένων. Σε αυτή την παράγραφο, αναφέρονται τα βασικά χαρακτηριστικά τους:

- **Όγκος (Volume):** Ένα από τα σημαντικότερα χαρακτηριστικά των Μεγάλων Δεδομένων είναι ο όγκος τους. Τα Μεγάλα Δεδομένα αναφέρονται σε τεράστια σύνολα δεδομένων που δεν μπορούν να υποστούν αποτελεσματική επεξεργασία με τις παραδοσιακές τεχνικές επεξεργασίας δεδομένων λόγω του μεγέθους τους. Ο όγκος αυτός παράγεται από διάφορες πηγές, συμπεριλαμβανομένων των μέσων κοινωνικής δικτύωσης, των αισθητήρων και του Διαδικτύου των Πραγμάτων (IoT).
- **Ποικιλομορφία (Variety):** Ένα άλλο βασικό χαρακτηριστικό των Μεγάλων Δεδομένων είναι η ποικιλομορφία τους. Τα δεδομένα έρχονται σε πολλές

διαφορετικές μορφές, συμπεριλαμβανομένων των δομημένων, των ημιδομημένων και των μη δομημένων δεδομένων. Αυτή η ποικιλομορφία απαιτεί εξειδικευμένα εργαλεία και τεχνολογίες για την αποτελεσματική αποθήκευση, διαχείριση και ανάλυση των δεδομένων. Η ποικιλομορφία των Μεγάλων Δεδομένων αποτελεί μια πρόκληση για τους οργανισμούς, καθώς πρέπει να έχουν την ικανότητα να επεξεργάζονται και να ερμηνεύουν δεδομένα από ένα ευρύ φάσμα πηγών.

- **Ταχύτητα (Velocity):** Η ταχύτητα των Μεγάλων Δεδομένων είναι ένα άλλο κρίσιμο χαρακτηριστικό, καθώς τα δεδομένα παράγονται και συλλέγονται με υψηλές ταχύτητες, συχνά σε πραγματικό χρόνο. Αυτό απαιτεί γρήγορη και αποτελεσματική επεξεργασία για την εξαγωγή πολύτιμων πληροφοριών από τα δεδομένα. Η ταχύτητα των Μεγάλων Δεδομένων είναι απαραίτητη για τους οργανισμούς που πρέπει να λαμβάνουν έγκαιρα αποφάσεις με βάση τα δεδομένα.
- **Ακρίβεια/ Αξιοπιστία (Veracity):** Η ακρίβεια είναι ένα άλλο σημαντικό χαρακτηριστικό των Μεγάλων Δεδομένων. Τα Μεγάλα Δεδομένα είναι συχνά αδόμητα και ελλιπή, γεγονός που μπορεί να οδηγήσει σε προβλήματα με την ποιότητα των δεδομένων. Οι ανακρίβειες, οι ασυνέπειες και οι προκαταλήψεις μπορούν να επηρεάσουν την ανάλυση και την ερμηνεία των δεδομένων. Είναι ζωτικής σημασίας να υπάρχουν διαδικασίες που να διασφαλίζουν την ακρίβεια των δεδομένων.
- **Αξία (Value):** Ο απώτερος στόχος των Μεγάλων Δεδομένων είναι η άντληση πολύτιμων πληροφοριών και γνώσεων που μπορούν να χρησιμοποιηθούν για τη βελτίωση των επιχειρηματικών λειτουργιών, των διαδικασιών και της λήψης αποφάσεων. Η αξία των Μεγάλων Δεδομένων έγκειται στην ικανότητά τους να παρέχουν στους οργανισμούς ένα ανταγωνιστικό πλεονέκτημα αξιοποιώντας τις γνώσεις από τους όγκους των δεδομένων.

5 Most Critical V's of Big Data



Σχήμα 1: Τα κρίσιμα χαρακτηριστικά των Μεγάλων Δεδομένων

(Ghandan Gaur (Oct 24, 2010))

Όπως απεικονίζεται στο πιο πάνω σχήμα, τα Μεγάλα Δεδομένα χαρακτηρίζονται από την αξία (Value), την αξιοπιστία (Veracity), την ποικιλομορφία (Variety), τον όγκο (Volume) και την ταχύτητα (Velocity) τους. Τα χαρακτηριστικά αυτά έχουν σημαντικές επιπτώσεις για τους οργανισμούς που πρέπει να διαθέτουν τα απαραίτητα εργαλεία και τεχνικές για την αποτελεσματική επεξεργασία, διαχείριση και ανάλυση όγκων Μεγάλων Δεδομένων. Αξιοποιώντας τις γνώσεις και τη γνώση που αποκτώνται από τα Μεγάλα Δεδομένα, οι οργανισμοί μπορούν να αποκτήσουν ανταγωνιστικό πλεονέκτημα στους αντίστοιχους κλάδους τους.

3.3 Internet of Things (IoT)

Σύμφωνα με το άρθρο (Conti, M., Dehghantanha, A., Franke, K., & Watson, S. 2018), το Internet of Things (IoT) ενσωματώνει αισθητήρες, αντικείμενα καθώς και έξυπνους κόμβους που να μπορούν να επικοινωνούν μεταξύ τους χωρίς την ανθρώπινη παρέμβαση. Πιο συγκεκριμένα, τα αντικείμενα λειτουργούν με αυτόματο τρόπο σε σχέση με άλλα αντικείμενα. Οι κόμβοι IoT είναι σε θέση να παρέχουν δεδομένα, να έχουν πρόσβαση και να εξουσιοδοτούν τους πόρους με τους οποίους βασίζονται για τη συλλογή και την εξαγωγή δεδομένων καθώς επίσης και για τη λήψη αποφάσεων με την ανάλυση των συλλεγόμενων δεδομένων. Πρόσθετα, το IoT οδηγεί σε μια σύνδεση ανθρώπων, υπηρεσιών, αισθητήρων και αντικειμένων. Ωστόσο, οι συσκευές IoT αναπτύσσονται σε πολλές εφαρμογές, από έξυπνα δίκτυα μέχρι και συστήματα υγειονομικής περίθαλψης και μεταφοράς των πληροφοριών. Ωστόσο, οι ευκαιρίες που υπάρχουν στον τομέα IoT αυξήσαν σημαντικά τον αριθμό των έξυπνων συσκευών όπως και των έξυπνων, αυτόνομων υπηρεσιών που προσφέρονται στα δίκτυα IoT. Αξίζει να σημειωθεί ότι, τα θέματα ασφάλειας όπως για παράδειγμα το απόρρητο, ο έλεγχος πρόσβασης, η ασφαλής επικοινωνία καθώς και αποθήκευση δεδομένων δημιουργούν σημαντικές προκλήσεις στο περιβάλλον IoT. Βέβαια, κάθε μεμονωμένη συσκευή που δημιουργείται, κάθε νέος αισθητήρας που αναπτύσσεται όπως και κάθε μεμονωμένο

byte που συγχρονίζεται στο περιβάλλον IoT υπάρχει η περίπτωση να χρειαστεί να τεθεί υπό έλεγχο κατά τη διάρκεια μιας έρευνας. Αφενός, η γρήγορη ανάπτυξη των συσκευών και των υπηρεσιών IoT έχει ως αποτέλεσμα την ανάπτυξη πολλών ευάλωτων και ανασφαλών κόμβων.

Το IoT έχει δύο κατευθύνσεις, η πρώτη είναι οι έξυπνες φυσικές συσκευές και η δεύτερη κατεύθυνση τα περιβάλλοντα και η παρούσα διασύνδεση. Σημειώνεται επίσης ότι, το IoT δεν σημαίνει ότι τα αποτελέσματα είναι μόνο θετικά, αλλά η ασφάλεια και τα ζητήματα εμπιστευτικότητας που προκαλούνται από έναν αρκετά μεγάλο αριθμό συσκευών συνδεδεμένων στο Διαδίκτυο, που αυτό έχει ως αποτέλεσμα να καθιστούν αδύνατη την επιτήρηση όλου του δικτύου. Πολλά αντικείμενα IoT εμφανίστηκαν καιρό πριν από την ίδια την ιδέα, αλλά δεν είχαν μια παγκόσμια διάσταση. Επιπλέον, ένα άλλο χαρακτηριστικό του IoT αναφέρεται στα μοντέλα επικοινωνίας τα οποία περιγράφονται από το συμβούλιο Αρχιτεκτονικής Διαδικτύου. Πιο κάτω ακολουθεί μια σειρά μοντέλων επικοινωνίας: Το μοντέλο επικοινωνίας Device-to-Device το οποίο αντιπροσωπεύει δύο ή περισσότερες συσκευές όπου συνδέονται άμεσα καθώς και επικοινωνούν μεταξύ τους ή μέσω μιας ενδιάμεσης εφαρμογής διακομιστή (Server), σε πολλούς τύπους δικτύων, συμπεριλαμβανομένων των δικτύων IP ή του Διαδικτύου. Το μοντέλο επικοινωνίας Device-to-Cloud εφαρμόζει την ιδέα ότι οι συσκευές συνδέονται απευθείας σε μια υπηρεσία Internet Cloud. Το μοντέλο Device-to-Gateway λειτουργεί όταν κάτι ενδιάμεσο μεταξύ της συσκευής Διαδικτύου και τις υπηρεσίες υπολογιστικού νέφους (Cloud Computing), τα οποία αποτελούνται από μια εφαρμογή λογισμικού, λειτουργεί σε τοπικό επίπεδο συσκευή πύλης και παρέχει λειτουργίες, όπως για παράδειγμα μετάφραση δεδομένων και πρωτοκόλλου και ασφάλειας. Το μοντέλο Back-end-Data-sharing είναι μια αρχιτεκτονική επικοινωνία που επιτρέπει στους χρήστες να αναλύουν και να εξάγουν αντικείμενα δεδομένων, μέσα από μια υπηρεσία Cloud και αυτό έρχεται σε συνδυασμό με δεδομένα από άλλες πηγές (Banica, L., Burtescu, E., & Enescu, F. (2017)).

Το Internet of Things (IoT) αναφέρεται στο δίκτυο φυσικών συσκευών, οχημάτων, κτιρίων και άλλων αντικειμένων που διαθέτουν αισθητήρες, λογισμικό και συνδεσιμότητα, επιτρέποντάς τους να συλλέγουν και να ανταλλάσσουν δεδομένα με άλλες συσκευές και συστήματα μέσω του Διαδικτύου. Πρόκειται για έναν ταχέως αναπτυσσόμενο τομέα που έχει τη δυνατότητα να μεταμορφώσει διάφορες βιομηχανίες, επιτρέποντας νέες μορφές αυτοματοποίησης, βελτιστοποίησης και καινοτομίας.

Σύμφωνα με τους Atzori, Iera και Morabito (2010), το IoT είναι "μια παγκόσμια υποδομή για την κοινωνία της πληροφορίας, που επιτρέπει προηγμένες υπηρεσίες με τη διασύνδεση (φυσικών και εικονικών) πραγμάτων με βάση τις υπάρχουσες και εξελισσόμενες διαλειτουργικές τεχνολογίες πληροφοριών και επικοινωνιών". Ομοίως, οι Gubbi, Buyya, Marusic και Palaniswami (2013) ορίζουν το IoT ως «ένα δίκτυο μοναδικά αναγνωρίσιμων και διευθυνσιοδοτούμενων αντικειμένων που μπορούν να επικοινωνούν μεταξύ τους και να συνεργάζονται με τους γείτονές τους για την επίτευξη ενός κοινού στόχου». Το IoT χαρακτηρίζεται από την ικανότητά του να επιτρέπει την παρακολούθηση και τον έλεγχο φυσικών συστημάτων σε πραγματικό χρόνο, καθώς και από την ικανότητά του να παράγει τεράστιες ποσότητες δεδομένων που μπορούν να αναλυθούν και να χρησιμοποιηθούν για τη βελτίωση της λήψης αποφάσεων και των επιχειρηματικών διαδικασιών. Έχει εφαρμογές σε ένα ευρύ φάσμα τομέων, συμπεριλαμβανομένων των έξυπνων σπιτιών, των έξυπνων πόλεων, της υγειονομικής περίθαλψης, της γεωργίας, των μεταφορών και της μεταποίησης.

Κεφάλαιο 4

Προκλήσεις στη Δικανική Μεγάλων Δεδομένων

Η εκθετική ανάπτυξη των ψηφιακών δεδομένων αλλάζει ριζικά τον τρόπο με τον οποίο οι άνθρωποι και οι επιχειρήσεις συλλέγουν, χρησιμοποιούν και μοιράζονται πληροφορίες. Η αυξανόμενη χρήση των ψηφιακών gadget συνέβαλε επίσης, στην αύξηση του εγκλήματος στον κυβερνοχώρο και άλλων διαδικτυακών εγκλημάτων. Η αναγνώριση, η ανάλυση και η παρουσίαση ψηφιακών αποδεικτικών στοιχείων στο δικαστήριο απαιτούν τη χρήση ψηφιακών εγκληματολογικών ερευνών. Υπάρχουν όμως αρκετές δυσκολίες με τις ψηφιακές ιατροδικαστικές έρευνες, ειδικότερα στον καθορισμό δεδομένων. Τεράστιες ποσότητες ψηφιακών δεδομένων πρέπει να συγκεντρωθούν, να αναλυθούν και να ερμηνευθούν ως μέρος της ψηφιακής δικανικής Μεγάλων Δεδομένων, κάτι το οποίο μπορεί να δημιουργήσει ειδικές δυσκολίες στους ερευνητές. Σε αυτή τη μελέτη, συζητούνται τα προβλήματα με την ψηφιακή δικανική Μεγάλων Δεδομένων, μαζί με πιθανές τεχνικές.

Μία από τις κύριες προκλήσεις στην ψηφιακή δικανική είναι ο συνεχής αυξανόμενος όγκος δεδομένων. Αυτό το πρόβλημα έχει γίνει ακόμη πιο έντονο με την εμφάνιση Μεγάλων Δεδομένων και απαιτεί επανεξέταση του τρόπου με τον οποίο χειρίστηκαν οι έρευνες ψηφιακής δικανικής τα τελευταία χρόνια. Στο άρθρο (Adedayo, O. M, 2016) συζητούνται οι προκλήσεις και οι ανάγκες της ψηφιακής δικανικής ενόψει των τρεχουσών τάσεων και απαιτήσεων των διαφορετικών ερευνών. Προτείνεται ένα πλαίσιο ανάλυσης ψηφιακής δικανικής που λαμβάνει υπόψη τις υπάρχουσες τεχνικές καθώς και τις τρέχουσες προκλήσεις. Σκοπός του πλαισίου είναι να επαναξιολογήσει τα διάφορα στάδια της διαδικασίας εξέτασης της ψηφιακής δικανικής και να εισαγάγει σε κάθε στάδιο τις απαιτούμενες τεχνικές για τη βελτίωση της καλύτερης συλλογής,

ανάλυσης, διατήρησης και παρουσίασης ενόψει των Μεγάλων Δεδομένων και άλλων προκλήσεων που αντιμετωπίζει η ψηφιακή δικανική.

Στην παρούσα ενότητα παρουσιάζεται μια επισκόπηση του ευρύτερου πεδίου της ψηφιακής δικανικής με στόχο τον εντοπισμό των εξελίξεων και των βέλτιστων πρακτικών της τεχνολογίας, καθώς και των άλυτων ερευνητικών προκλήσεων. Ο Raghavan (2013) επισημάνει την ανάγκη τυποποίησης της έννοιας των ψηφιακών αποδεικτικών στοιχείων, ώστε οι ερευνητές να μπορούν να εκτελούν δικανική ανάλυση χρησιμοποιώντας μια κοινή πλατφόρμα. Σύμφωνα με τη διαδικασία απόκτησης φυσικών αποδεικτικών στοιχείων, ο Raghavan προτείνει την υιοθέτηση μιας μεθοδολογίας που είναι παρόμοια με τον τρόπο αποθήκευσης και οργάνωσης των φυσικών αποδεικτικών στοιχείων. Ωστόσο, δεδομένου ότι τα ψηφιακά αποδεικτικά στοιχεία μπορούν να αλλοιωθούν, να αντιγραφούν ή να διαγραφούν, προτείνει τις αρχές των τεσσάρων σημείων της αυθεντικοποίησης, της ακεραιότητας, του ελέγχου πρόσβασης και της μη αποκήρυξης ως βασικές εκτιμήσεις κατά το χειρισμό των ψηφιακών αποδεικτικών στοιχείων.

Επίσης, ο Raghavan εντόπισε ορισμένες από τις σημαντικότερες προκλήσεις που αντιμετωπίζει ο τομέας της ψηφιακής δικανικής, συμπεριλαμβανομένης της ανάγκης κλιμάκωσης της τεχνολογίας και προσαρμογής κλιμακούμενων αρχιτεκτονικών, της απαίτησης για ομοιόμορφα προγράμματα πιστοποίησης και μαθήματα στην ψηφιακή δικανική και της ανάγκης για αλλαγές στους νόμους που διέπουν το επιτρεπτό των ψηφιακών αποδεικτικών στοιχείων στα δικαστήρια. Οι προκλήσεις αυτές είναι κρίσιμες για την πρόοδο της ψηφιακής δικανικής και την αποτελεσματική αξιοποίηση των ψηφιακών αποδεικτικών στοιχείων στις έρευνες.

Σύμφωνα με τον Guarino (2013) οι δικανικές έρευνες συχνά παράγουν όγκους Μεγάλων Δεδομένων και ως εκ τούτου, ο τομέας της ψηφιακής δικανικής αντιμετωπίζει πολλές προκλήσεις που σχετίζονται με τον χειρισμό και την ανάλυση Μεγάλων Δεδομένων. Μερικές από τις σημαντικότερες προκλήσεις περιλαμβάνουν:

4.1.1 Συλλογή και αποθήκευση δεδομένων

Με το αυξανόμενο μέγεθος των ψηφιακών συσκευών αποθήκευσης, ο όγκος των δεδομένων που πρέπει να συλλεχθούν και να αναλυθούν κατά τη διάρκεια των

εγκληματολογικών ερευνών έχει αυξηθεί εκθετικά. Η συλλογή και η αποθήκευση αυτών των δεδομένων μπορεί να είναι μια πρόκληση, ειδικά όταν πρόκειται για τη διατήρηση της ακεραιότητας των δεδομένων.

4.1.2 Επεξεργασία δεδομένων

Η επεξεργασία ποσοτήτων Μεγάλων Δεδομένων είναι χρονοβόρα και εντατική σε πόρους. Οι δικανικοί ερευνητές χρησιμοποιήσουν εξειδικευμένα εργαλεία και τεχνικές για την αναζήτηση και ανάλυση των δεδομένων, τα οποία μπορεί είναι περίπλοκα και απαιτούν σημαντική εμπειρία.

4.1.3 Ανάλυση και ερμηνεία δεδομένων

Ο τεράστιος όγκος των δεδομένων που παράγονται κατά τη διάρκεια των εγκληματολογικών ερευνών δυσκολεύει τον εντοπισμό σχετικών πληροφοριών και προτύπων. Οι ερευνητές εντοπίζουν και ερμηνεύουν μοτίβα μέσα στα δεδομένα, και αυτό θεωρείται μια πολύπλοκη και χρονοβόρα διαδικασία.

4.1.4 Ποιότητα δεδομένων

Τα σύνολα Μεγάλων Δεδομένων παρουσιάζουν προκλήσεις που σχετίζονται με την ποιότητα των δεδομένων. Τα δεδομένα αυτά είναι ελλιπή, χωρίς ακρίβεια ή και ακόμα περιέχουν σφάλματα, τα οποία επηρεάζουν την αξιοπιστία των αποτελεσμάτων της έρευνας.

4.1.5 Απόρρητο και ασφάλεια δεδομένων

Τα Μεγάλα Δεδομένα στις ιατροδικαστικές έρευνες περιέχουν συχνά ευαίσθητες προσωπικές πληροφορίες, επομένως η διατήρηση του απορρήτου και της ασφάλειας των δεδομένων είναι κρίσιμης σημασίας. Επίσης, πρέπει να διασφαλίζεται ότι τα δεδομένα προστατεύονται από μη εξουσιοδοτημένη πρόσβαση, κλοπή ή λάθος χρήση. (Guarino, A. (2013).

4.2 Ιδιωτικότητα

Σύμφωνα με τους Sangeetha και Sudha Sadasivam (2019) ο όρος των Μεγάλων Δεδομένων (Big Data) έχει κερδίσει σημαντική προσοχή τα τελευταία χρόνια λόγω των τεράστιων ποσοτήτων δομημένων, ημιδομημένων και αδόμητων δεδομένων που παράγονται στην ψηφιακή εποχή. Ωστόσο, η εκμετάλλευση αυτών των δεδομένων μεγάλης κλίμακας μπορεί να απειλήσει την ιδιωτική ζωή των ατόμων, καθιστώντας τη διατήρηση της ιδιωτικής ζωής έναν κρίσιμο τομέα έρευνας. Αναφορικά, οι ακόλουθες είναι μερικές από τις προκλήσεις που θέτουν τα Μεγάλα Δεδομένα για την προστασία της ιδιωτικής ζωής:

1. **Παραβιάσεις δεδομένων:** Τα Μεγάλα Δεδομένα εμπεριέχουν τον κίνδυνο μαζικών παραβιάσεων δεδομένων, τα οποία μπορεί να οδηγήσουν στην έκθεση ευαίσθητων πληροφοριών και να προκαλέσουν σημαντική βλάβη στα άτομα.
2. **Ανωνυμία και εκ νέου ταυτοποίηση:** Τα Μεγάλα Δεδομένα μπορεί να διευκολύνουν την ταυτοποίηση ατόμων ακόμη και όταν τα δεδομένα τους είναι ανώνυμα. Η ταυτοποίηση εκ νέου των ατόμων αποτελεί πρόβλημα προστασίας της ιδιωτικής ζωής όταν τα δεδομένα μοιράζονται ή συνδυάζονται σε διαφορετικά σύνολα δεδομένων.
3. **Έλλειψη ελέγχου:** Τα άτομα έχουν συχνά ελάχιστο έλεγχο των δεδομένων τους από τη στιγμή που αυτά συλλέγονται και μοιράζονται στο οικοσύστημα των Μεγάλων Δεδομένων. Αυτή η έλλειψη ελέγχου μπορεί να οδηγήσει σε απώλεια της ιδιωτικής ζωής και σε δυνητικά επιβλαβείς συνέπειες.
4. **Περιορισμένοι έλεγχοι πρόσβασης:** Τα συστήματα Μεγάλων Δεδομένων έχουν συχνά περιορισμένους ελέγχους πρόσβασης, γεγονός που καθιστά δύσκολο να διασφαλιστεί ότι μόνο εξουσιοδοτημένα άτομα μπορούν να έχουν πρόσβαση σε ευαίσθητα δεδομένα.
5. **Ανεπαρκείς κανονισμοί:** Οι ισχύοντες κανονισμοί προστασίας της ιδιωτικής ζωής ενδέχεται να μην είναι επαρκείς για την προστασία της ιδιωτικής ζωής των ατόμων στο πλαίσιο των Μεγάλων Δεδομένων. Καθώς τα Μεγάλα Δεδομένα συνεχίζουν να εξελίσσονται, ενδέχεται να χρειαστεί να αναπτυχθούν νέοι κανονισμοί για την αντιμετώπιση αυτών των προκλήσεων.

4.3 Ηθικά και Νομικά Ζητήματα

Η συλλογή και ανάλυση μεγάλου όγκου δεδομένων για σκοπούς δικανικής παρουσιάζει σύνθετα νομικά και δεοντολογικά ζητήματα που πρέπει να λαμβάνονται υπόψη για να διασφαλίζεται η συμμόρφωση με τους σχετικούς νόμους και κανονισμούς. Σύμφωνα με τους συγγραφείς Balogun, και Zuva (2017) η ανάπτυξη και η εφαρμογή των ψηφιακών δικανικών συστημάτων εγείρουν διάφορα ηθικά ζητήματα, όπως η προστασία της ιδιωτικής ζωής και η εμπιστευτικότητα, η απόρριψη και η συσκοτίση, η ακρίβεια και η αποτελεσματικότητα, η χειραγώγηση, η παραποίηση, η αντικατάσταση και η μείωση. Τα μοναδικά χαρακτηριστικά των ψηφιακών δικανικών συστημάτων, τα οποία υπερβαίνουν τα τυπικά συστήματα πληροφοριών, επιτείνουν τις ηθικές ανησυχίες. Στο έγγραφο εντοπίζονται πρόσθετα ηθικά ζητήματα που σχετίζονται με τα ψηφιακά δικανικά συστήματα, συμπεριλαμβανομένης της εκμετάλλευσης των εργαλείων, της ανεπάρκειας των νομικών μέτρων για την αντιμετώπιση σύνθετων ζητημάτων, της ανάγκης για συνεχή αναθεώρηση των νομοθετικών πλαισίων και της εφαρμογής τεχνικών μέτρων, όπως η κατάρτιση, η αναδιαμόρφωση και η αναδιάρθρωση των συστημάτων. Για την αντιμετώπιση των δεοντολογικών ανησυχιών που προκύπτουν από την ανάπτυξη και την εφαρμογή των ψηφιακών δικανικών συστημάτων, συνιστώνται προσωρινά μέτρα που αντιμετωπίζουν άμεσα τις ανησυχίες των ενδιαφερομένων για την εμπιστευτικότητα που προκύπτουν από την τεχνολογική πρόοδο, τις νομικές παρεμβάσεις και τα τεχνικά μέτρα, όπως η κατάρτιση, η επανασχεδίαση και η αναδιάρθρωση των συστημάτων. Είναι σημαντικό να σημειωθεί ότι τα μέτρα αυτά πρέπει να εφαρμοστούν εγκαίρως και αποτελεσματικά, ώστε να μετριαστούν οι πιθανές ηθικές παραβιάσεις.

4.3 Ανωνυμία

Η ανωνυμία είναι μία από τις σημαντικότερες προκλήσεις για την προστασία της ιδιωτικής ζωής που σχετίζονται με τα Μεγάλα Δεδομένα. Η ανωνυμία αναφέρεται στην ικανότητα προστασίας της ταυτότητας των ατόμων των οποίων οι προσωπικές πληροφορίες συλλέγονται και υποβάλλονται σε επεξεργασία σε σύνολα δεδομένων

μεγάλης κλίμακας. Με την ευρεία υιοθέτηση των τεχνολογιών Μεγάλων Δεδομένων, τα προσωπικά δεδομένα μπορούν να συλλεχθούν από διάφορες πηγές, όπως πλατφόρμες μέσω κοινωνικής δικτύωσης, διαδικτυακές συναλλαγές και κινητές συσκευές. Η ανωνυμία δεδομένων είναι ζωτικής σημασίας για την προστασία των δικαιωμάτων ιδιωτικότητας των ατόμων, την πρόληψη της κλοπής ταυτότητας και τη μείωση του κινδύνου κατάχρησης προσωπικών πληροφοριών. Ωστόσο, η επίτευξη της ανωνυμίας στα Μεγάλα Δεδομένα είναι ένα περίπλοκο έργο λόγω του τεράστιου όγκου (volume) και της ποικιλομορφίας (variety) των δεδομένων, της έλλειψης τυποποίησης στις τεχνικές ανωνυμίας και της πιθανής εκ νέου ταυτοποίησης ατόμων μέσω της διασταύρωσης των συνόλων δεδομένων. Επίσης, οι τεχνικές όπως η απόκρυψη δεδομένων και ο κατακερματισμός, μπορεί να οδηγήσουν στην απώλεια πολύτιμων πληροφοριών, μειώνοντας έτσι τη χρησιμότητα των δεδομένων για αναλυτικούς σκοπούς. Ωστόσο, η επίτευξη ισορροπίας μεταξύ της ανωνυμίας και της χρησιμότητας των δεδομένων αποτελεί κρίσιμη πρόκληση στην προστασία της ιδιωτικής ζωής των Μεγάλων Δεδομένων (Sangeetha, S., & Sudha Sadasivam, G. (2019))

4.4 Δικανική Λογιστική

Κατά την ανάπτυξη εργαλείων ανάλυσης δεδομένων για τη διερεύνηση της απάτης, η δικανική λογιστική στις αναδυόμενες οικονομίες αντιμετωπίζει σημαντικές προκλήσεις λόγω της έλλειψης των απαιτούμενων δεξιοτήτων και εμπειρογνωμοσύνης. Σύμφωνα με τους ερευνητές Akinbowale, Mashigo, και Zerihun (2023), η επιτυχία της διαδικασίας διερεύνησης εξαρτάται σε μεγάλο βαθμό από τις δεξιότητες και την εμπειρογνωμοσύνη του εγκληματολογικού ερευνητή και την επιλογή του εργαλείου διερεύνησης της εγκληματολογικής λογιστικής. Ωστόσο, η διερεύνηση και η ανίχνευση της απάτης στην ψηφιακή εποχή είναι πολύπλοκη λόγω της συνωμοσίας, της εξαπάτησης και της απόκρυψης, γεγονός που καθιστά δύσκολο τον εντοπισμό της αιτίας και των δραστών. Για να ξεπεραστούν οι προκλήσεις αυτές, χρησιμοποιούνται προηγμένες και γνωστικές αναλύσεις που παρέχουν μεγαλύτερη ευελιξία και αυτοματοποίηση στη φάση της ανάλυσης δεδομένων, εξαλείφοντας έτσι την αυστηρή χειρωνακτική εργασία και μειώνοντας τον κίνδυνο ανθρώπινου λάθους.

Οι προηγμένες και γνωστικές αναλύσεις προσφέρουν υψηλή υπολογιστική

απόδοση και αφήγηση δεδομένων, η οποία μπορεί να καθοδηγήσει τους εγκληματολογικούς λογιστές να βγάλουν τεκμηριωμένα συμπεράσματα σχετικά με ύποπτες υποθέσεις απάτης. Αποτελεί σημαντικό, για τους εγκληματολογικούς λογιστές να ενημερώνονται για τις τελευταίες εξελίξεις στην ανάλυση δεδομένων, ώστε να αποφεύγεται η υπερβολική εξάρτηση από την παραδοσιακή ανάλυση δεδομένων. Με την ενσωμάτωση της δικανικής λογιστικής και της τεχνολογίας Μεγάλων Δεδομένων, μπορεί να ενισχυθεί η φάση ανάλυσης δεδομένων του πλαισίου της δικανικής λογιστικής. Για να επιτευχθεί αυτό, θα πρέπει να καθοριστεί η ικανότητα και των δύο τεχνικών και να προσδιοριστεί το εύρος των δεδομένων και των αποδεικτικών στοιχείων που είναι διαθέσιμα για ανάλυση. Η αντιμετώπιση αυτών των προκλήσεων μπορεί να βοηθήσει την εγκληματολογική λογιστική να διαδραματίσει σημαντικό ρόλο στον μετριασμό της απάτης, ιδίως στον τραπεζικό κλάδο των αναδυόμενων οικονομιών.

4.5 Δικανική Νέφους - Cloud Forensics

Η δικανική του νέφους αποτελεί επίσης πρόκληση λόγω της εξαιρετικά κατανεμημένης και πολύπλοκης αρχιτεκτονικής του νέφους, του μοντέλου χρήσης πολλαπλών μισθωτών, της εικονικοποίησης και της ευμετάβλητης φύσης των ίδιων των δεδομένων. Παρουσιάζεται ως ένα υποπεδίο της ψηφιακής δικανικής που ασχολείται με τη συλλογή, ανάλυση και αποθήκευση ηλεκτρονικών δεδομένων που είναι αποθηκευμένα στο cloud (εικονική αποθήκευση) για τη διερεύνηση ποινικών αδικημάτων. Η κατανεμημένη φύση των πόρων (εφαρμογές, αποθήκευση κ.λπ.) στο νέφος αμφισβητεί τις βασικές έννοιες των ορίων και της ιδιοκτησίας του τόπου του εγκλήματος στην ψηφιακή εγκληματολογία και αποτελεί ένα επιπλέον ζήτημα στη διαχείριση μιας αλυσίδας επιτήρησης (chain of custody) σε ένα περιβάλλον νέφους. Τέλος, η κρυπτογράφηση δημιουργεί έναν συμβιβασμό μεταξύ της ενισχυμένης ιδιωτικής ζωής των καταναλωτών και της επιτυχούς έρευνας, καθιστώντας δύσκολη την ανάκτηση δεδομένων από τους ερευνητές ακόμη και με νομική και κυβερνητική υποστήριξη (Arshad, H., Jantan, A. B., & Abiodun, O. I. (2018)).

4.6 Κατανεμημένο σύστημα αποθήκευσης

Σύμφωνα με τους Chen, Mao και Liu, (2014), μια από τις προκλήσεις που θέτουν τα Μεγάλα Δεδομένα είναι η ανάπτυξη ενός κατανεμημένου συστήματος αποθήκευσης (Distributed Storage System) μεγάλης κλίμακας που επεξεργάζεται και να αναλύει αποτελεσματικά τα δεδομένα. Κατά τη χρήση ενός κατανεμημένου συστήματος για την αποθήκευση τεράστιου όγκου δεδομένων, πρέπει να λαμβάνονται υπόψη διάφοροι παράγοντες. Οι παράγοντες αυτοί περιλαμβάνουν τη συνέπεια, τη διαθεσιμότητα και την ανοχή κατακερματισμών.

Η συνέπεια είναι κρίσιμη σε ένα κατανεμημένο σύστημα αποθήκευσης που απαιτεί πολλαπλούς διακομιστές για τη συνεργατική αποθήκευση δεδομένων. Καθώς αυξάνεται ο αριθμός των διακομιστών (servers) αυξάνεται και η πιθανότητα αποτυχίας του διακομιστή. Για να εξασφαλιστεί η διαθεσιμότητα σε περίπτωση βλάβης διακομιστή, τα δεδομένα συνήθως χωρίζονται σε πολλαπλά κομμάτια και αποθηκεύονται σε διαφορετικούς διακομιστές. Ωστόσο, οι αστοχίες διακομιστών και η παράλληλη αποθήκευση μπορεί να οδηγήσουν σε ασυνέπειες μεταξύ διαφορετικών αντιγράφων των ίδιων δεδομένων. Ως εκ τούτου, η συνέπεια αναφέρεται στη διασφάλιση ότι πολλαπλά αντίγραφα των ίδιων δεδομένων είναι πανομοιότυπα.

Επίσης στο άρθρο τους οι Chen, Mao και Liu αναφέρουν ότι η διαθεσιμότητα είναι ένας άλλος βασικός παράγοντας που πρέπει να ληφθεί υπόψη σε ένα κατανεμημένο σύστημα αποθήκευσης που λειτουργεί σε πολλαπλά σύνολα διακομιστών. Καθώς χρησιμοποιούνται περισσότεροι διακομιστές, οι αποτυχίες διακομιστών γίνονται πιο πιθανές. Ιδανικά, ολόκληρο το σύστημα δεν θα πρέπει να επηρεάζεται σημαντικά για την ικανοποίηση των αιτημάτων των πελατών όσον αφορά την ανάγνωση και την εγγραφή. Η ιδιότητα αυτή ονομάζεται διαθεσιμότητα.

Η ανοχή κατακερματισμών είναι εξίσου σημαντική επειδή οι πολλαπλοί διακομιστές σε ένα κατανεμημένο σύστημα αποθήκευσης συνδέονται με ένα δίκτυο που μπορεί να παρουσιάσει βλάβες συνδέσεων ή και ακόμα προσωρινή συμφόρηση. Επομένως, το κατανεμημένο σύστημα πρέπει να έχει ένα ορισμένο επίπεδο ανοχής σε προβλήματα που προκαλούνται από αστοχίες του δικτύου. Θα ήταν επιθυμητό το κατανεμημένο σύστημα αποθήκευσης να συνεχίσει να λειτουργεί σωστά ακόμη και όταν το δίκτυο κατατεμαχιστεί.

4.7 Εμπόδια στην ανάπτυξη εφαρμογών των Μεγάλων Δεδομένων

Η ανάπτυξη των Μεγάλων Δεδομένων αντιμετωπίζει διάφορα εμπόδια που μπορεί να εμποδίσουν την επιτυχή εφαρμογή και αξιοποίησή τους.

4.7.1 Αναπαράσταση Δεδομένων

Στο άρθρο τους οι Chen, Mao και Liu, (2014), δηλώνουν ότι η έννοια της αναπαράστασης των δεδομένων (Data representation) θεωρείται ένα κρίσιμο στοιχείο στην ανάλυση Μεγάλων Δεδομένων, καθώς πολλά σύνολα δεδομένων έχουν ορισμένα επίπεδα ετερογένειας ως προς τον τύπο, τη δομή (structure), τη σημασιολογία (semantics), την οργάνωση (organization), το granularity (Ιδιότητα να αποτελείται από πολλά επιμέρους στοιχεία.) και την προσβασιμότητα (accessibility). Ο πρωταρχικός στόχος της αναπαράστασης δεδομένων είναι να καταστήσει τα δεδομένα πιο ουσιαστικά για την ανάλυση τους με υπολογιστή και την ερμηνεία τους από τον χρήστη. Είναι σημαντικό να σημειωθεί ότι μια ακατάλληλη αναπαράσταση δεδομένων μπορεί να μειώσει την αξία των αρχικών δεδομένων και να εμποδίσει ακόμη και την αποτελεσματική ανάλυση δεδομένων. Επομένως, η αποτελεσματική αναπαράσταση δεδομένων πρέπει να αντικατοπτρίζει τη δομή, την τάξη και τον τύπο των δεδομένων, καθώς και τις ολοκληρωμένες τεχνολογίες, ώστε να επιτρέπει αποτελεσματικές λειτουργίες σε διαφορετικά σύνολα δεδομένων.

4.7.2 Μείωση του Πλεονασμού και Συμπύεση Δεδομένων

Η μείωση του πλεονασμού (Redundancy reduction) και η συμπύεση δεδομένων (data compression) είναι αποτελεσματικές μέθοδοι για τη μείωση του υψηλού επιπέδου πλεονασμού που γενικά συναντάται στα σύνολα δεδομένων. Ωστόσο, είναι ζωτικής σημασίας να διασφαλιστεί ότι οι μέθοδοι αυτές δεν επηρεάζουν τις δυνητικές αξίες των δεδομένων. Για παράδειγμα, τα δεδομένα που παράγονται από δίκτυα αισθητήρων είναι εξαιρετικά πλεοναστικά και οι τεχνικές φιλτραρίσματος και συμπύεσης μπορούν να εφαρμοστούν για να μειώσουν το μέγεθός τους κατά τάξεις μεγέθους.

4.7.3 Διαχείριση του κύκλου ζωής

Η διαχείριση του κύκλου ζωής των δεδομένων (Data life cycle management) αποτελεί σημαντική πρόκληση στην ανάλυση Μεγάλων Δεδομένων, καθώς η διάχυτη ανίχνευση και η χρήση υπολογιστή (computing) παράγουν δεδομένα σε πρωτοφανείς ρυθμούς και κλίμακες. Τα τρέχοντα συστήματα αποθήκευσης ενδέχεται να μην υποστηρίζουν τέτοια μαζικά δεδομένα και θα πρέπει να αναπτυχθεί μια αρχή σπουδαιότητας των δεδομένων που σχετίζεται με την αναλυτική αξία για να αποφασιστεί ποια δεδομένα θα πρέπει να αποθηκευτούν και ποια δεδομένα θα πρέπει να απορριφθούν.

4.7.4 Αναλυτικός μηχανισμός

Το αναλυτικό σύστημα (Analytical mechanism) των Μεγάλων Δεδομένων πρέπει να επεξεργάζεται μάζες δεδομένων σε περιορισμένο χρόνο. Τα παραδοσιακά RDBMS (Σύστημα διαχείρισης σχεσιακών βάσεων δεδομένων), δεν έχουν σχεδιαστεί με δυνατότητα κλιμάκωσης και επέκτασης, τα οποία ενδέχεται να μην ανταποκρίνονται στις απαιτήσεις απόδοσης της ανάλυσης Μεγάλων Δεδομένων. Οι μη σχεσιακές βάσεις δεδομένων επιδιώκουν μοναδικά πλεονεκτήματα στην επεξεργασία μη δομημένων δεδομένων και έχουν γίνει κυρίαρχες στην ανάλυση Μεγάλων Δεδομένων. Ωστόσο, εξακολουθούν να υπάρχουν ορισμένα ζητήματα με τις μη σχεσιακές βάσεις δεδομένων όσον αφορά τις επιδόσεις τους και τις ειδικές εφαρμογές. Ως εκ τούτου, απαιτείται μια συμβιβαστική λύση μεταξύ RDBMS και μη-σχεσιακών βάσεων δεδομένων, όπως μια μικτή αρχιτεκτονική βάσεων δεδομένων που ενσωματώνει τα πλεονεκτήματα και των δύο τύπων βάσεων δεδομένων.

4.7.5 Εμπιστευτικότητα των Δεδομένων

Η εμπιστευτικότητα των δεδομένων (Data confidentiality) αποτελεί κρίσιμη ανησυχία στην ανάλυση Μεγάλων Δεδομένων, καθώς οι περισσότεροι πάροχοι

υπηρεσιών Μεγάλων Δεδομένων ή οι ιδιοκτήτες μπορεί να μην έχουν την ικανότητα να διατηρούν και να αναλύουν τέτοια τεράστια σύνολα δεδομένων. Ως εκ τούτου, μπορεί να απαιτούνται επαγγελματίες ή εργαλεία για την ανάλυση τέτοιων δεδομένων, γεγονός που αυξάνει τους πιθανούς κινδύνους ασφάλειας. Θα πρέπει να λαμβάνονται κατάλληλα προληπτικά μέτρα για την προστασία των ευαίσθητων δεδομένων, ώστε να διασφαλίζεται η ασφάλειά τους κατά την παράδοση της ανάλυσης Μεγάλων Δεδομένων σε τρίτους για επεξεργασία.

4.7.6 Διαχείριση της ενέργειας

Η διαχείριση της ενέργειας (Energy management) είναι μια άλλη σημαντική πρόκληση στην ανάλυση Μεγάλων Δεδομένων, καθώς η επεξεργασία, η αποθήκευση και η μετάδοση Μεγάλων Δεδομένων καταναλώνουν αναπόφευκτα περισσότερη ηλεκτρική ενέργεια. Ως εκ τούτου, πρέπει να δημιουργηθεί ένας μηχανισμός ελέγχου και διαχείρισης της κατανάλωσης ενέργειας σε επίπεδο συστήματος για τα Μεγάλα Δεδομένα, διασφαλίζοντας παράλληλα την επεκτασιμότητα και την προσβασιμότητα.

4.7.7 Αναλωσιμότητα και Επεκτασιμότητα

Η αναλωσιμότητα και η επεκτασιμότητα (Expendability and scalability) αποτελούν κρίσιμες απαιτήσεις για το σύστημα ανάλυσης Μεγάλων Δεδομένων. Ο αναλυτικός αλγόριθμος πρέπει να είναι σε θέση να επεξεργάζεται ολοένα και πιο διερευνόμενα και πιο σύνθετα σύνολα δεδομένων για να υποστηρίξει τα σημερινά και μελλοντικά σύνολα δεδομένων.

4.7.8 Συνεργασία

Η ανάλυση των Μεγάλων Δεδομένων είναι μια διεπιστημονική έρευνα που απαιτεί τη συνεργασία (cooperation) εμπειρογνομώνων σε διαφορετικούς τομείς για τη

συγκομιδή του δυναμικού των Μεγάλων Δεδομένων. Πρέπει να δημιουργηθεί μια ολοκληρωμένη αρχιτεκτονική δικτύου Μεγάλων Δεδομένων που να βοηθά τους επιστήμονες και τους μηχανικούς διαφόρων τομέων να έχουν πρόσβαση σε διαφορετικά είδη δεδομένων και να αξιοποιούν πλήρως την εμπειρογνωμοσύνη τους για να συνεργαστούν και να επιτύχουν αναλυτικούς στόχους (Chen, Mao και Liu, (2014)).

4.8 Ευκαιρίες των Μεγάλων Δεδομένων

Το έγγραφο των Zawoad, και Hasan (2015, August), υπογραμμίζει ότι τα Μεγάλα Δεδομένα μπορούν να προσφέρουν ευκαιρίες για ψηφιακές δικανικές έρευνες, αποκαλύπτοντας άγνωστες προηγούμενες γνώσεις μέσω της ανάλυσης μεγάλων συνόλων δεδομένων. Με τα κατάλληλα εργαλεία και τεχνολογίες, ο εντοπισμός, η συλλογή και η ανάλυση ενός μεγάλου συνόλου δεδομένων μπορεί να οδηγήσει σε νέα δεδομένα σχετικά με εγκληματικά περιστατικά. Επιπλέον, το έγγραφο προτείνει ένα εννοιολογικό μοντέλο για την δικανική Μεγάλων Δεδομένων, το οποίο μπορεί να προσφέρει νέες γνώσεις για τον προσδιορισμό γεγονότων σχετικά με εγκληματικά περιστατικά. Συνολικά, η διαθεσιμότητα Μεγάλων Δεδομένων στην ψηφιακή εγκληματολογία μπορεί να ανοίξει ευκαιρίες για τους ερευνητές να αντλήσουν σημαντικές πληροφορίες από τα συσσωρευμένα δεδομένα.

Κεφάλαιο 5

Τεχνικές Αντιμετώπισης

Οι τεχνικές μεγάλων δεδομένων μπορούν να βοηθήσουν τους ιατροδικαστές να αναλύσουν μεγάλα και πολύπλοκα σύνολα δεδομένων, γεγονός που μπορεί να οδηγήσει σε πιο αποδοτικές και αποτελεσματικές έρευνες.

5.1 Cross Drive

Οι Patterson και Hargreaves (2012) συζητούν τις δυνατότητες χρήσης αυτοματοποιημένων ψηφιακών εγκληματολογικών χρονοδιαγραμμάτων στην ανάλυση πολλαπλών δίσκων, η οποία περιλαμβάνει την ανάλυση δεδομένων από πολλαπλές συσκευές αποθήκευσης για τον εντοπισμό σχέσεων και πιθανών αποδεικτικών στοιχείων σε ποινικές έρευνες. Οι συγγραφείς υποστηρίζουν ότι τα ψηφιακά εγκληματολογικά χρονοδιαγράμματα μπορούν να μειώσουν σημαντικά τον χρόνο και την προσπάθεια που απαιτείται για την ανάλυση, βελτιώνοντας παράλληλα την ακρίβεια και την αξιοπιστία τους. Προτείνουν, ένα εργαλείο λογισμικού με την ονομασία Timeline Explorer που επιτρέπει στους ερευνητές να απεικονίζουν και να διερευνούν τις σχέσεις μεταξύ δεδομένων από πολλαπλούς δίσκους, καθώς και να αναζητούν συγκεκριμένα αρχεία. Ωστόσο, αναγνωρίζουν τους πιθανούς περιορισμούς και τις προκλήσεις που σχετίζονται με τη χρήση αυτοματοποιημένων ψηφιακών εγκληματολογικών χρονολογικών γραμμών στην ανάλυση μεταξύ δίσκων, όπως η επίδραση της ποιότητας των δεδομένων, της διαθεσιμότητας μεταδεδομένων και των μέτρων ασφαλείας στην ακρίβεια της ανάλυσης.

Οι συγγραφείς σημειώνουν ότι τα ψηφιακά εγκληματολογικά χρονοδιαγράμματα μπορούν να βοηθήσουν τους ερευνητές να εντοπίσουν σχετικά δεδομένα που διαφορετικά θα μπορούσαν να διαφύγουν, καθώς και να παρέχουν ένα σαφές και συνοπτικό χρονοδιάγραμμα των γεγονότων που σχετίζονται με μια συγκεκριμένη

υπόθεση. Αυτό μπορεί να είναι ιδιαίτερα χρήσιμο σε πολύπλοκες υποθέσεις που περιλαμβάνουν ποσότητες Μεγάλων Δεδομένων από πολλαπλές πηγές. Με τη χρήση αυτοματοποιημένων ψηφιακών εγκληματολογικών χρονοδιαγραμμάτων, οι ερευνητές μπορούν να εντοπίζουν γρήγορα βασικά στοιχεία και να δημιουργούν ένα χρονοδιάγραμμα γεγονότων που μπορεί να χρησιμοποιηθεί για την υποστήριξη της υπόθεσής τους στο δικαστήριο.

Τέλος, οι συγγραφείς υπογραμμίζουν τη σημασία της διασφάλισης ότι τα ψηφιακά εγκληματολογικά χρονοδιαγράμματα είναι αξιόπιστα και παραδεκτά στο δικαστήριο. Σημειώνουν ότι τα αυτοματοποιημένα ψηφιακά εγκληματολογικά χρονοδιαγράμματα θα πρέπει να σχεδιάζονται έτσι ώστε να ανταποκρίνονται στις αποδεικτικές απαιτήσεις του νομικού συστήματος και να βασίζονται σε υγιείς εγκληματολογικές αρχές. Αυτό περιλαμβάνει τη διασφάλιση ότι τα δεδομένα που χρησιμοποιούνται για τη δημιουργία του χρονοδιαγράμματος είναι ακριβή και αξιόπιστα και ότι η μεθοδολογία που χρησιμοποιείται για την ανάλυση των δεδομένων είναι καλά τεκμηριωμένη και διαφανής. Οι συγγραφείς προτείνουν επίσης ότι οι ιατροδικαστές θα πρέπει να είναι προετοιμασμένοι να υπερασπιστούν την εγκυρότητα του ψηφιακού εγκληματολογικού χρονολογίου στο δικαστήριο και ότι οι δικαστές και οι επαγγελματίες του νομικού κλάδου θα πρέπει να λάβουν κατάρτιση σχετικά με τη χρήση και την ερμηνεία των ψηφιακών εγκληματολογικών αποδεικτικών στοιχείων. Με την αντιμετώπιση αυτών των ζητημάτων, οι συγγραφείς προτείνουν ότι τα αυτοματοποιημένα ψηφιακά εγκληματολογικά χρονοδιαγράμματα μπορούν να αποτελέσουν ένα ισχυρό εργαλείο για την υποστήριξη των ποινικών ερευνών και την προώθηση της δικαιοσύνης (Patterson, J., & Hargreaves, C. J. (2012)).

5.2 Τεχνητή Νοημοσύνη

Σύμφωνα με το άρθρο, οι Saxena, Usha, Vinoth, Veena, και Nancy (2023) η τεχνητή νοημοσύνη (Artificial Intelligence) έχει καταστεί απαραίτητο εργαλείο για τους ερευνητές ψηφιακών αποδεικτικών στοιχείων λόγω των ποικίλων τεχνικών και των ιδιόκτητων εφαρμογών της που τους επιτρέπουν να εξετάζουν κρίσιμες πηγές πληροφοριών. Οι εγκληματολόγοι υπολογιστών χρησιμοποιούν διάφορες τεχνικές τεχνητής νοημοσύνης, την τεχνογνωσία τους και τις γνώσεις των εμπειρογνομόνων

τους για να αναζητήσουν κρυμμένα αρχεία, να αξιολογήσουν τον μη διατιθέμενο αποθηκευτικό χώρο του δίσκου σε αναζήτηση τυχόν αντιγράφων σκόπιμα διαγραμμένων ή κρυπτογραφημένων αρχείων και να βρουν απαντήσεις στα ερωτήματά τους. Οι πιο δημοφιλείς τυπικές τεχνικές τεχνητής νοημοσύνης που χρησιμοποιούνται στην ψηφιακή δικανική είναι η αντίστροφη στεγανογραφία (Steganography) Artificial Intelligence, η στοχαστική δικανική, η ανάλυση διασταυρούμενων δίσκων, η ζωντανή ανάλυση και η ανάκτηση διαγραμμένων αρχείων. Στην αντίστροφη στεγανογραφία, οι εμπειρογνώμονες δικανικής υπολογιστών χρησιμοποιούν λειτουργίες κατακερματισμού δεδομένων για να εντοπίσουν πιθανές πληροφορίες που έχουν αλλοιωθεί ή κρυφτεί κρυφά. Η δικανική αποσκοπεί στην αξιολόγηση και ανάκτηση πληροφοριών χωρίς τη χρήση τεχνουργημάτων οποιουδήποτε τύπου. Η ανάλυση πολλαπλών δίσκων, γνωστή και ως anomaly detection, συγκρίνει ύποπτα συμβάντα ή πληροφορίες σε διαφορετικά συστήματα και συσκευές υπολογιστών για την αποκάλυψη και αξιολόγηση κρίσιμων δεδομένων. Η Real-Time ανάλυση εξετάζει αρχεία ή δεδομένα στο εσωτερικό ενός συστήματος σε πραγματικό χρόνο, ενώ η ανάκτηση διαγραμμένων αρχείων περιλαμβάνει την αναζήτηση σε απορρίμματα, θραύσματα μνήμης ή κάδους ανακύκλωσης του υπολογιστή για κάθε αρχείο που έχει αφαιρεθεί από μια θέση αλλά έχει αφήσει τα ίχνη του αλλού. Η εφαρμογή αυτών των τεχνικών στην ψηφιακή δικανική είναι η άμεση ανάγκη της ώρας, καθώς μπορούν να αυτοματοποιήσουν τη διαδικασία εγκληματολογικής ανάλυσης, να προσθέσουν μια έξυπνη γνώμη εμπειρογνώμονα και να διαδραματίσουν θετικό ρόλο στην απόδοση δικαιοσύνης και στη διατήρηση του νόμου και της τάξης στον ψηφιακό τομέα.

Επιπλέον, η χρήση της τεχνητής νοημοσύνης στην ψηφιακή δικανική έχει πολλά οφέλη, όπως η μείωση του χρόνου και του κόστους των ερευνών, η αύξηση της ακρίβειας και της ποιότητας των αποτελεσμάτων και η υποβοήθηση ερευνών που αφορούν τον όγκο Μεγάλων Δεδομένων. Η τεχνητή νοημοσύνη μπορεί επίσης να διαδραματίσει κρίσιμο ρόλο στον εντοπισμό και την πρόληψη του εγκλήματος στον κυβερνοχώρο και των παραβιάσεων δεδομένων, καθώς και στη διατήρηση των ψηφιακών αποδεικτικών στοιχείων με αδιάβλητο και αξιόπιστο τρόπο. Ωστόσο, η υιοθέτηση της τεχνητής νοημοσύνης στην ψηφιακή δικανική θέτει επίσης αρκετές προκλήσεις, όπως η ανάγκη για εξειδικευμένο προσωπικό, ο κίνδυνος μεροληψίας στη λήψη αποφάσεων και η

δυνατότητα εκμετάλλευσης των συστημάτων τεχνητής νοημοσύνης από επιτιθέμενους. Για την αντιμετώπιση αυτών των προκλήσεων, οι ερευνητές διερευνούν νέες τεχνικές που βασίζονται στην τεχνητή νοημοσύνη, όπως η βαθιά μάθηση, η ενισχυτική μάθηση και η επεξεργασία φυσικής γλώσσας, και αναπτύσσουν δεοντολογικές κατευθυντήριες γραμμές και πρότυπα για τη χρήση της στην ψηφιακή δικανική. Έτσι, με αυτό τον τρόπο η ενσωμάτωση της τεχνητής νοημοσύνης και της τεχνολογίας blockchain μπορεί να ενισχύσει την αξιοπιστία και την αξιοπιστία των ψηφιακών εγκληματολογικών ερευνών, διασφαλίζοντας την απονομή δικαιοσύνης και την τήρηση του κράτους δικαίου στην ψηφιακή εποχή.

Έπειτα η χρήση τεχνικών τεχνητής νοημοσύνης στην ψηφιακή δικανική έχει καταστεί απαραίτητο εργαλείο για τους ερευνητές ψηφιακών αποδεικτικών στοιχείων. Οι συνήθεις τεχνικές της τεχνητής νοημοσύνης, όπως για παράδειγμα η αντίστροφη στεγανογραφία, η στοχαστική δικανική, η ανάλυση cross disc, η real-time ανάλυση και η ανάκτηση διαγραμμένων αρχείων, έχουν αποδειχθεί αποτελεσματικές στην αποκάλυψη και αξιολόγηση κρίσιμων δεδομένων. Η εφαρμογή αυτών των τεχνικών μπορεί να αυτοματοποιήσει τη διαδικασία εγκληματολογικής ανάλυσης, να προσθέσει μια έξυπνη γνώμη εμπειρογνώμονα και να διαδραματίσει θετικό ρόλο στην απόδοση δικαιοσύνης και στη διατήρηση του νόμου και της τάξης στον ψηφιακό τομέα (Saxena, I., Usha, G., Vinoth, N. A. S., Veena, S., & Nancy, M. (2023)).

5.3 MapReduce

Η τεχνική προγραμματισμού MapReduce αποτελείται από δύο στοιχεία, δηλαδή τον mapper και τον reducer, τα οποία επιτρέπουν την παράλληλη επεξεργασία δεδομένων, διευκολύνοντας έτσι τον χειρισμό δεδομένων μεγάλης κλίμακας με την κατανομή της επεξεργασίας σε πολλαπλούς χαρτογράφους και μειωτές. Στο πλαίσιο της ανίχνευσης εγκλημάτων στον κυβερνοχώρο στο Διαδίκτυο των Πραγμάτων - Internet of Things (IoT), η εργασία επιλογής χαρακτηριστικών πραγματοποιείται στη φάση mapper, ενώ η πραγματική ανίχνευση εγκλημάτων στον κυβερνοχώρο πραγματοποιείται στη φάση reducer, χρησιμοποιώντας το πλαίσιο MapReduce. Η φάση

mapper περιλαμβάνει το φιλτράρισμα και την προεπεξεργασία των δεδομένων, με στόχο την εξαγωγή ουσιαστικών χαρακτηριστικών που μπορούν να χρησιμοποιηθούν για την ανίχνευση εγκλημάτων στον κυβερνοχώρο. Τα χαρακτηριστικά αυτά διαβιβάζονται στη συνέχεια στη φάση reducer, όπου πραγματοποιείται η πραγματική διαδικασία ανίχνευσης. Η διαδικασία αυτή περιλαμβάνει τη χρήση των εξαχθέντων χαρακτηριστικών για την εκπαίδευση ενός μοντέλου ανίχνευσης και στη συνέχεια την εφαρμογή του εκπαιδευμένου μοντέλου για τον εντοπισμό πιθανών συμβάντων κυβερνοεγκλήματος. Η προσέγγιση MapReduce για την ανίχνευση εγκλημάτων στον κυβερνοχώρο προτιμάται λόγω της ικανότητάς της να χειρίζεται μεγάλους όγκους δεδομένων και της επεκτασιμότητάς της, επιτρέποντας την αποτελεσματική ανίχνευση συμβάντων εγκλημάτων στον κυβερνοχώρο σε πραγματικό χρόνο. (Tharaliya, S., & Sharma, P. K. (2023)).

Το πλαίσιο MapReduce, το οποίο αναπτύχθηκε αρχικά από την Google, είναι ένα ιδιόκτητο πλαίσιο που χρησιμοποιείται για την παράλληλη επεξεργασία τεράστιων συνόλων δεδομένων σε ένα καταναμημένο δίκτυο υπολογιστών. Το Hadoop, μια υλοποίηση αυτού του πλαισίου ανοικτού κώδικα, ενσωματώνει το καταναμημένο σύστημα αρχείων Hadoop (HDFS) για την αποθήκευση σημαντικών συνόλων δεδομένων και τη ροή τους στις εφαρμογές των χρηστών. Πρωταρχικός στόχος του Hadoop είναι η αποτελεσματική εκτέλεση στατιστικής ανάλυσης δεδομένων. Στον τομέα της δικανικής Μεγάλων Δεδομένων, το πλαίσιο MapReduce χρησιμοποιείται συνήθως για την εκτέλεση διαδικασιών σε τεράστιο όγκο δυνητικών αποδεικτικών στοιχείων. Το MapReduce είναι μια ισχυρή τεχνολογία Μεγάλων Δεδομένων που μπορεί να αξιοποιηθεί για ψηφιακές εγκληματολογικές έρευνες. Επιτρέποντας την επεξεργασία και την ανάλυση μεγάλου όγκου δεδομένων με καταναμημένο και παράλληλο τρόπο, το MapReduce μπορεί να βοηθήσει τους ερευνητές να εντοπίσουν μοτίβα, και άλλες σημαντικές πληροφορίες που μπορεί να διαφύγουν από τις παραδοσιακές εγκληματολογικές τεχνικές (Zawoad, S., & Hasan, R. (2015, August)).

5.4 Αλγόριθμοι Machine learning

Σύμφωνα με τους Balushi, Shaker, και Kumar, B. (2023, January), τα τελευταία χρόνια, η αύξηση των εγκλημάτων στον κυβερνοχώρο έχει οδηγήσει σε αυξημένη εστίαση στην ψηφιακή δικανική ως μέσο απόκτησης αποδεικτικών στοιχείων υψηλής ποιότητας. Ωστόσο, οι ιατροδικαστές συχνά αντιμετωπίζουν προκλήσεις που σχετίζονται με τη συλλογή και την ανάλυση δεδομένων όταν προσπαθούν να ανακατασκευάσουν γεγονότα. Οι τεχνικές μηχανικής μάθησης έχουν αναδειχθεί ως ένα δυναμικό πολύτιμο εργαλείο για τους ερευνητές, δεδομένου του τεράστιου όγκου ανθρώπινων αλληλεπιδράσεων που συμβαίνουν σε καθημερινή βάση. Η μηχανική μάθηση είναι ένα εξειδικευμένο υποπεδίο της τεχνητής νοημοσύνης που περιλαμβάνει την ανάπτυξη υπολογιστικών μοντέλων και αλγορίθμων ικανών να εκτελούν συγκεκριμένες εργασίες χωρίς ρητό προγραμματισμό, όπως η εκπαίδευση και ο έλεγχος συνόλων δεδομένων.

Αυτή η τεχνολογία έχει τη δυνατότητα να βοηθήσει σημαντικά στις έρευνες και ως εκ τούτου, το παρόν έγγραφο παρέχει μια ολοκληρωμένη επισκόπηση των διαφόρων τεχνικών μηχανικής μάθησης που μπορούν να χρησιμοποιηθούν για την εξέταση και ανάλυση ψηφιακών αποδεικτικών στοιχείων κατά τη διαδικασία έρευνας. Κάθε αλγόριθμος μηχανικής μάθησης έχει σχεδιαστεί για την αντιμετώπιση ενός συγκεκριμένου τομέα της ψηφιακής δικανικής με βάση την ικανότητά του να ξεπερνά τις προκλήσεις που σχετίζονται με την πολυπλοκότητα τους, τον όγκο των δεδομένων, τη χρονική ευθυγράμμιση, τη συσχέτιση, τη συνέπεια και άλλα σχετικά χαρακτηριστικά. Επιπλέον, το άρθρο εξετάζει τη διαθεσιμότητα διαφόρων αλγορίθμων μηχανικής μάθησης που μπορούν να χρησιμοποιηθούν για διάφορους τομείς της ψηφιακής δικανικής.

Οι αλγόριθμοι μηχανικής μάθησης παρέχουν διάφορα οφέλη στην ψηφιακή δικανική, όπως η αποδοτική και αποτελεσματική επεξεργασία τεράστιων όγκων δεδομένων, η αυτοματοποίηση των διαδικασιών έρευνας και η ικανότητα αναγνώρισης μοτίβων που μπορεί να μην είναι εύκολα ανιχνεύσιμα από τους ανθρώπινους ερευνητές. Συνοπτικά, η αξιοποίηση της μηχανικής μάθησης στην ψηφιακή δικανική μπορεί να βοηθήσει στον εντοπισμό και την πρόληψη της μη εξουσιοδοτημένης πρόσβασης σε πληροφορίες και να διευκολύνει τον προσδιορισμό της ακριβούς φύσης ενός

εγκλήματος. Είναι σημαντικό να αξιολογούνται προσεκτικά τα πλεονεκτήματα και οι περιορισμοί κάθε αλγορίθμου κατά την επιλογή του καταλληλότερου για μια συγκεκριμένη εργασία στην ψηφιακή δικανική (Balushi, Shaker, και Kumar, B. (2023, January)).

5.4.1 Αλγόριθμος Distributed Random Forest

Η δικανική είναι η πρακτική της συλλογής, ανάλυσης και ερμηνείας ψηφιακών στοιχείων για την υποστήριξη της διερεύνησης εγκλημάτων στον κυβερνοχώρο. Με την αύξηση του όγκου, της ταχύτητας και της ποικιλίας των ψηφιακών δεδομένων, οι παραδοσιακές τεχνικές καθίστανται λιγότερο αποτελεσματικές. Τα τελευταία χρόνια, τεχνικές καταναμημένου υπολογισμού, όπως η Random Forest (RF), έχουν εφαρμοστεί στην δικανική Μεγάλων Δεδομένων για την αντιμετώπιση αυτών των προκλήσεων. Στην παρούσα διατριβή συζητείται η χρήση της τεχνικής Distributed Random Forest (DRF) για την δικανική των Μεγάλων Δεδομένων. Ο αλγόριθμος Random Forest είναι μια ιδιαίτερα διαδεδομένη και ευρέως χρησιμοποιούμενη μέθοδος από τους επιστήμονες δεδομένων. Είναι ένας αλγόριθμος μηχανικής μάθησης με επίβλεψη που βρίσκει εκτεταμένη χρήση τόσο σε προβλήματα ταξινόμησης όσο και σε προβλήματα παλινδρόμησης. Ο αλγόριθμος κατασκευάζει δέντρα αποφάσεων σε διάφορα τυχαία δείγματα των δεδομένων εισόδου και λαμβάνει την ψήφο πλειοψηφίας για ταξινόμηση και τον μέσο όρο για παλινδρόμηση.

Για την κατανόηση της λειτουργία του αλγορίθμου Random Forest στη μηχανική μάθηση (Machine Learning), είναι απαραίτητο να κατανοήσουμε πρώτα την έννοια της τεχνικής μάθησης συνόλου. Η μάθηση συνόλου αναφέρεται στην πρακτική του συνδυασμού πολλαπλών μοντέλων για την πραγματοποίηση προβλέψεων, αντί να βασίζεται σε ένα μεμονωμένο μοντέλο. Συνεπώς, αντί να χρησιμοποιείται ένα μόνο μοντέλο, χρησιμοποιείται μια συλλογή μοντέλων για να γίνουν ακριβέστερες προβλέψεις. Η τεχνική μάθησης συνόλου χρησιμοποιεί δύο κύριες μεθόδους για τη βελτίωση της ακρίβειας των προβλέψεων με το συνδυασμό πολλαπλών μοντέλων:

1. Η πρώτη μέθοδος είναι το Bagging, το οποίο περιλαμβάνει τη δημιουργία διαφορετικών υποσυνόλων εκπαίδευσης από το δείγμα δεδομένων εκπαίδευσης με αντικατάσταση. Κάθε υποσύνολο χρησιμοποιείται για την εκπαίδευση ενός διαφορετικού μοντέλου και το τελικό αποτέλεσμα βασίζεται στην ψηφοφορία πλειοψηφίας. Ένα παράδειγμα μεθόδου bagging που χρησιμοποιείται στη μάθηση συνόλου είναι το Random Forest.
2. Η δεύτερη μέθοδος είναι η Boosting, η οποία περιλαμβάνει το συνδυασμό αδύναμων εκπαιδευτών για τη δημιουργία ισχυρών εκπαιδευτών με τη δημιουργία διαδοχικών μοντέλων. Τα μοντέλα δημιουργούνται με μια σειρά έτσι ώστε το τελικό μοντέλο να έχει την υψηλότερη ακρίβεια. Παραδείγματα μεθόδων boosting που χρησιμοποιούνται στη μάθηση συνόλου περιλαμβάνουν το ADA BOOST και το XG BOOST. (Sruthi E R, 2017)

5.4.1.1 Χαρακτηριστικά αλγόριθμου Distributed Random Forest

Βασικό χαρακτηριστικό του αλγορίθμου Random Forest είναι η ικανότητά του να χειρίζεται σύνολα δεδομένων που περιέχουν τόσο συνεχείς όσο και κατηγορικές μεταβλητές σε εργασίες παλινδρόμησης και ταξινόμησης, αντίστοιχα. Ο αλγόριθμος έχει βρεθεί ότι αποδίδει σημαντικά καλά και στους δύο αυτούς τομείς. Ο αλγόριθμος Random Forest διαθέτει αρκετά σημαντικά χαρακτηριστικά που τον καθιστούν δημοφιλή επιλογή μεταξύ των επιστημόνων δεδομένων. Ένα από αυτά τα χαρακτηριστικά είναι η ποικιλομορφία, καθώς δεν λαμβάνονται υπόψη όλα τα χαρακτηριστικά, οι μεταβλητές ή τα χαρακτηριστικά κατά τη δημιουργία ενός μεμονωμένου δέντρου. Κάθε δέντρο κατασκευάζεται με διαφορετικό τρόπο, και αυτό έχει ως αποτέλεσμα την αυξημένη ποικιλομορφία. Επιπλέον, ο αλγόριθμος είναι απρόσβλητος στην διάστασης, καθώς κάθε δέντρο δεν εξετάζει όλα τα χαρακτηριστικά, μειώνοντας έτσι τον χώρο των χαρακτηριστικών. Ένα άλλο αξιοσημείωτο χαρακτηριστικό είναι ο παραλληλισμός (Parallelization), καθώς κάθε δέντρο δημιουργείται ανεξάρτητα χρησιμοποιώντας διαφορετικά δεδομένα και χαρακτηριστικά. Αυτό επιτρέπει την πλήρη αξιοποίηση της CPU για τη δημιουργία Random Forests. Επιπλέον, ο αλγόριθμος Random Forest δεν απαιτεί διαχωρισμό των δεδομένων για τα σύνολα εκπαίδευσης και δοκιμής, καθώς το 30% των δεδομένων είναι αθέατο από το δέντρο απόφασης. Τέλος, ο αλγόριθμος

προσφέρει σταθερότητα, καθώς το αποτέλεσμα βασίζεται στην ψηφοφορία πλειοψηφίας ή στον μέσο όρο (Sruthi E R, 2017).

5.4.1.2 Πλεονεκτήματα αλγόριθμου Distributed Random Forest

Σύμφωνα με την Liu, Y. (2014), η μέθοδος του Random Forest είναι ένας αλγόριθμος ταξινόμησης που συνδυάζει τα Decision Tree και τα πλεονεκτήματα του bagging και της τυχαίας επιλογής χαρακτηριστικών. Γενικότερα, τα πλεονεκτήματά του αλγόριθμου περιλαμβάνουν:

- Το bagging επιτρέπει την εκτίμηση της σημασίας των χαρακτηριστικών και του σφάλματος γενίκευσης.
- Τα δέντρα που κατασκευάζονται με τη μέθοδο του Random Forest χρησιμοποιούν τον αλγόριθμο CART, ο οποίος χειρίζεται τόσο συνεχή όσο και διακριτά χαρακτηριστικά
- Η μέθοδος μπορεί να αντιμετωπίσει αποτελεσματικά το πρόβλημα της μη ισορροπημένης ταξινόμησης
- Έχει υψηλή ανοχή στο θόρυβο και στην ακρίβεια ταξινόμησης

Επιπλέον, στο πλαίσιο των Μεγάλων Δεδομένων, η μέθοδος του τυχαίου δάσους έχει τα ακόλουθα πλεονεκτήματα:

- Ικανότητα να χειρίζεται μεγάλες ποσότητες δεδομένων
- Τα δέντρα αποφάσεων (Decision Tree) που παράγονται από τη μέθοδο είναι σχετικά απλά και εύκολα στην ερμηνεία από τους επιχειρηματικούς αναλυτές
- Η μέθοδος είναι κατάλληλη για καταναμημένα και παράλληλα περιβάλλοντα και παρουσιάζει καλή κλιμάκωση
- Ο απλός ταξινομητής που δημιουργείται από τα δέντρα αποφάσεων (Decision Tree) μπορεί να επεξεργαστεί αποτελεσματικά τα δεδομένα, γεγονός που είναι κατάλληλο για τον γρήγορο ρυθμό ανανέωσης των δεδομένων που χαρακτηρίζει τα περιβάλλοντα Μεγάλων Δεδομένων.

Αυτά τα πλεονεκτήματα καθιστούν τη μέθοδο του Random Forest ένα πολύτιμο εργαλείο για την ανάλυση δεδομένων σε περιβάλλοντα Μεγάλων Δεδομένων.

5.4.1.3 Μειονεκτήματα αλγόριθμου Distributed Random Forest

Παρά τα πλεονεκτήματα του αλγορίθμου Random Forest που αναφέρθηκαν προηγουμένως, η έλευση των Μεγάλων Δεδομένων παρουσιάζει αρκετά μειονεκτήματα:

Σε ένα περιβάλλον των Μεγάλων Δεδομένων, ο ρυθμός ενημέρωσης των δεδομένων είναι ταχύς, το ίδιο και οι ενημερώσεις των χαρακτηριστικών των δεδομένων και των μη ορατών λειτουργιών στα δεδομένα. Ως αποτέλεσμα, τα δέντρα αποφάσεων που βασίζονται σε σύνολα εκπαίδευσης μπορεί γρήγορα να ξεπεραστούν και να καταστούν λιγότερο αποτελεσματικά στην ταξινόμηση δεδομένων. Συνεπώς, είναι απαραίτητο οι αλγόριθμοι στο πλαίσιο Μεγάλων Δεδομένων να διαθέτουν προσαρμοστικότητα δεδομένων, χωρίς να παρεμποδίζουν την κανονική ροή των δεδομένων μέσω του ταξινομητή.

Τα δέντρα αποφάσεων τείνουν να είναι άπληστοι αλγόριθμοι που μπορούν εύκολα να οδηγήσουν σε αστάθεια και υπερπροσαρμογή. Αυτό μπορεί να έχει σημαντικό αντίκτυπο στην ακρίβεια των δέντρων απόφασης, καθιστώντας ζωτικής σημασίας την αντιμετώπιση αυτού του ζητήματος σε σενάρια Μεγάλων Δεδομένων.

Η κλίμακα Forest για το Random Forest δεν έχει ακόμη καθοριστεί. Μια μεγάλη κλίμακα μπορεί να οδηγήσει σε πλεονασμό, ο οποίος θα μπορούσε να μειώσει την αποτελεσματικότητα και την ακρίβεια της ταξινόμησης (Liu, Y. (2014)).

5.4.2 Αλγόριθμος Decision Tree

Το δέντρο αποφάσεων (Decision Tree) είναι μια ευέλικτη μέθοδος μάθησης που μπορεί να χρησιμοποιηθεί τόσο για εργασίες παλινδρόμησης όσο και για εργασίες ταξινόμησης. Πρόκειται για μια απλή προσέγγιση που επιτρέπει την αντιστοίχιση των αποτελεσμάτων από δοκιμές στην ταξινόμηση στοιχείων δεδομένων. Το μοντέλο δέντρου αποφάσεων συλλαμβάνει τη λογική των διαφόρων αποφάσεων και τις μεταφράζει σε μια δενδροειδή δομή, με τον ανώτερο κόμβο να είναι ο κόμβος ρίζας. Οι εσωτερικοί κόμβοι του δέντρου αποφάσεων αντιπροσωπεύουν δοκιμές που σχετίζονται με τις μεταβλητές εισόδου ή τα χαρακτηριστικά, και ο αλγόριθμος ταξινόμησης διακλαδίζεται στον κατάλληλο κόμβο-παιδί μετά την ολοκλήρωση της δοκιμής. Η

διαδικασία αυτή συνεχίζεται μέχρι να φτάσουμε στον κόμβο φύλλο, ο οποίος είναι υπεύθυνος για την τελική απόφαση ταξινόμησης. Η απλότητα και η ερμηνευσιμότητα του δέντρου αποφάσεων το καθιστούν δημοφιλή επιλογή στη μηχανική μάθηση και την εξόρυξη δεδομένων, επιτρέποντας τη σαφή κατανόηση της διαδικασίας λήψης αποφάσεων (Sarker, I. H. (2021)).

Η πρόβλεψη είναι απαραίτητη για την κατανόηση της τάσης και τη λήψη τεκμηριωμένων επενδυτικών αποφάσεων. Η ανάλυση Μεγάλων Δεδομένων είναι μια διαδικασία συλλογής, διαχείρισης και ανάλυσης μεγάλων συνόλων δεδομένων για τον εντοπισμό μοτίβων και τάσεων. Ο στόχος της ανάλυσης Μεγάλων Δεδομένων είναι να προσδιοριστεί ο βέλτιστος χρόνος εισόδου ή εξόδου από την αγορά για την επένδυση χρημάτων. Παρόλο που διατίθενται διάφορα εργαλεία, όπως εργαλεία επιχειρηματικής ευφυΐας (Business Intelligence), στατιστικά εργαλεία και εργαλεία οπτικοποίησης δεδομένων, για την εκτέλεση αναλύσεων, τα περισσότερα από αυτά δεν μπορούν να χειριστούν τα σύνολα Μεγάλων Δεδομένων ή χρειάζονται πολύ χρόνο για την επεξεργασία των δεδομένων. Η ανάλυση Μεγάλων Δεδομένων μπορεί να είναι χρήσιμη για την εξόρυξη δεδομένων, την πρόβλεψη και την πρόβλεψη, μεταξύ άλλων εφαρμογών. Κατά την πρόβλεψη της τιμής του χρυσού, διάφοροι παράγοντες, συμπεριλαμβανομένης της τιμής ανοίγματος, της τιμής κλεισίματος, της χαμηλότερης τιμής, της υψηλότερης τιμής και της αξίας του χρυσού, μπορούν να ληφθούν υπόψη για τον προσδιορισμό της ποσοστιαίας μεταβολής της τιμής (Hou, Y. (2018, December)).

5.4.3 Αλγόριθμος Gradient Boosting Machines

Σύμφωνα με τους Shrivastav, και Kumar (2022), ο αλγόριθμος μηχανής ενίσχυσης κλίσης (GBM), κατάλληλη για την επεξεργασία και ανάλυση στοχαστικών δεδομένων υψηλής συχνότητας χρηματιστηριακών αγορών. Η χρήση του αλγόριθμου Gradient Boosted Machine (GBM) στην ανάλυση Μεγάλων Δεδομένων έχει αποδειχθεί ότι είναι ένα αποτελεσματικό μοντέλο μηχανικής μάθησης για εργασίες παλινδρόμησης. Η ερμηνευσιμότητα και τα προσαρμόσιμα χαρακτηριστικά του παράγουν ακριβή αποτελέσματα. Το μοντέλο παράγει την τιμή πρόβλεψης σε κάθε επανάληψη με βάση την προηγούμενη επανάληψη και εξασφαλίζει βέλτιστα αποτελέσματα

χρησιμοποιώντας ένα μέσο βάρος. Ένας πρόσθετος ταξινομητής μπορεί να κληθεί σε κάθε στάδιο για να βελτιώσει τη συνολική απόδοση. Η τροποποιημένη έκδοση του boosting μπορεί να ταξινομηθεί ως μη γραμμική ταξινόμηση, η οποία βελτιστοποιεί την ακρίβεια του δέντρου χωρίς να επηρεάζει την ταχύτητά του. Η GBM παρέχει ένα εύκολα διανεμητέο και παραλληλοποιήσιμο χαρακτηριστικό, το οποίο δημιουργεί ένα αβίαστο περιβάλλον για τη ρύθμιση και την επιλογή μοντέλων. Αυτή η έκδοση της GBM είναι ικανή να χειρίζεται Μεγάλα Δεδομένα με βέλτιστη ακρίβεια και δεν χρησιμοποιείται συνήθως στην πρόβλεψη της χρηματιστηριακής αγοράς. Ωστόσο, η αποτελεσματικότητά της είναι σημαντική στο σημερινό σενάριο των Μεγάλων Δεδομένων.

5.4.4 Αλγόριθμος Naive Bayes

Η παρούσα ενότητα επικεντρώνεται στον αλγόριθμο ταξινόμησης Naive Bayes. Τα μοντέλα Naive Bayes είναι γνωστά για την απλότητα και την ταχύτητά τους, γεγονός που τα καθιστά κατάλληλα για σύνολα δεδομένων υψηλής διάστασης. Απαιτούν πολύ λίγες ρυθμιζόμενες παραμέτρους, γεγονός που τα καθιστά χρήσιμα ως μια γρήγορη και βρώμικη βασική γραμμή για προβλήματα ταξινόμησης. Αυτή η ενότητα έχει ως στόχο να παράσχει μια διαισθητική εξήγηση του τρόπου λειτουργίας των ταξινομητών naive Bayes, ακολουθούμενη από παραδείγματα εφαρμογής τους σε διάφορα σύνολα δεδομένων (VanderPlas, J. (2016)).

5.5 Εξόρυξη Δεδομένων - Data Mining

Οι Sangeetha και Sudha Sadasivam (2019) αναφέρουν ότι στην εξόρυξη δεδομένων (Data mining), ο στόχος είναι η εξαγωγή χρήσιμων πληροφοριών από τα δεδομένα μέσω της χρήσης διαφόρων τεχνικών. Οι τεχνικές αυτές αποκαλύπτουν πολύτιμα επιχειρηματικά πρότυπα και γνώσεις που συχνά κρύβονται σε τεράστιες ποσότητες δεδομένων. Ωστόσο, η χρήση τέτοιων τεχνικών αποτελεί απειλή για την ιδιωτική ζωή των ατόμων, καθιστώντας τη διατήρηση της ιδιωτικής ζωής στην εξόρυξη δεδομένων ουσιαστικό μέλημα. Ο κύριος στόχος της διατήρησης της ιδιωτικής ζωής στην

εξόρυξη δεδομένων είναι η προστασία των δεδομένων να αποκαλυφθούν. Για την επίτευξη αυτού του στόχου, χρησιμοποιούνται διάφοροι αλγόριθμοι για την τροποποίηση των δεδομένων κατά τρόπο ώστε τα ευαίσθητα δεδομένα και η γνώση να παραμένουν ιδιωτικά ακόμη και μετά τη διαδικασία εξόρυξης. Η διατήρηση της ιδιωτικότητας περιλαμβάνει την προστασία των δεδομένων τόσο κατά τη διάρκεια της συλλογής δεδομένων όσο και κατά τη διάρκεια των διαδικασιών εξόρυξης. Η ιδιωτικότητα εισόδου (Input Privacy) αναφέρεται στην προστασία των δεδομένων κατά τη συλλογή, ενώ η ιδιωτικότητα εξόδου (Output Privacy) αναφέρεται στην προστασία των δεδομένων κατά τη διαδικασία εξόρυξης. Η προστασία της ιδιωτικής ζωής μπορεί να προστατευθεί με διάφορες τεχνικές, όπως η γενίκευση (generalization), η καταστολή (suppression) και η διαταραχή (perturbation). Αυτοί οι μηχανισμοί τροποποιούν τα δεδομένα με διάφορους τρόπους για να διασφαλίσουν τη μυστικότητα τους. Όμως, η υπερβολική τροποποίηση των δεδομένων μπορεί να καταστήσει τα δεδομένα που προκύπτουν ως άχρηστα. Αντίθετα, η ελάχιστη τροποποίηση των δεδομένων παρέχει καλή χρησιμότητα αλλά δεν μπορεί να εγγυηθεί την ιδιωτικότητα.

5.6 Τεχνολογία ανοικτού κώδικα - Open-Source technology

Σύμφωνα με το έγγραφο του Carrier (2002) η τεχνολογία ανοικτού κώδικα μπορεί πράγματι είναι πιο αξιόπιστη και πληροί τις κατευθυντήριες γραμμές πιο ολοκληρωμένα από την τεχνολογία κλειστού κώδικα. Ο Carrier εξετάζει την αξιοπιστία και τη συνάφεια των εργαλείων που χρησιμοποιούνται ως αποδεικτικά στοιχεία σε νομικές δίκες και δείχνει ότι τα εργαλεία ανοικτού κώδικα μπορεί να πληρούν με μεγαλύτερη σαφήνεια και πληρότητα τις κατευθυντήριες γραμμές. Προτείνει επίσης μια ισορροπημένη λύση για να επιτραπεί στις εμπορικές εταιρείες λογισμικού να παραμείνουν ανταγωνιστικές, διατηρώντας παράλληλα τον κώδικα εξαγωγής ως ανοικτού κώδικα και διαθέσιμο για δημοσίευση και αξιολόγηση από ομότιμους. Καταλήγει στο συμπέρασμα ότι η ποιότητα των ψηφιακών εγκληματολογικών εργαλείων μπορεί να βελτιωθεί επιτρέποντας την πρόσβαση στον πηγαίο κώδικα του εργαλείου και διεξάγοντας μια ολοκληρωμένη μεθοδολογία δοκιμών.

Τα παραδείγματα ανοικτού κώδικα αναφέρονται σε προγράμματα λογισμικού, έργα ή ακόμη εφαρμογές που αναπτύσσονται, διανέμονται και συντηρούνται με άδεια ανοικτού κώδικα, η οποία επιτρέπει στους χρήστες να έχουν πρόσβαση, να χρησιμοποιούν, να τροποποιούν και να διανέμουν ελεύθερα τον πηγαίο κώδικα του λογισμικού. Μερικά παραδείγματα ανοικτού κώδικα είναι τα εξής:

5.6.1 Apache Hadoop

Το Apache Hadoop είναι μια τεχνολογία ανοικτού κώδικα που παρέχει μια λύση για την αποθήκευση, τη διανομή και την εξόρυξη μεγάλων δεδομένων. Τα μεγάλα δεδομένα αναφέρονται σε μεγάλες και πολύπλοκες βάσεις δεδομένων που ενημερώνονται συχνά και μπορεί να είναι δομημένες, ημιδομημένες ή αδόμητες. Το Hadoop έχει σχεδιαστεί για να χειρίζεται τα τρία V των μεγάλων δεδομένων: όγκος, ταχύτητα και ποικιλία. Αποτελεί μια αποτελεσματική και δωρεάν λύση για τη διαχείριση μεγάλων δεδομένων σε κατανεμημένο περιβάλλον και τρέχει στο λειτουργικό σύστημα Linux. Το κύριο πλεονέκτημα της χρήσης του Apache Hadoop είναι η ικανότητά του να χειρίζεται μεγάλα και πολύπλοκα σύνολα δεδομένων, καθιστώντας το πολύτιμο εργαλείο για την ανάλυση δεδομένων (Nogueira, R. R., & Gueiber, E. (2014)).

5.6.2 Apache Spark

Ο αντίκτυπος των αναλυτικών στοιχείων μεγάλων δεδομένων στην αγορά του κλάδου είναι τεράστιος λόγω της ικανότητάς του να ανακαλύπτει κρυφά μοτίβα και άλλες πληροφορίες σε μεγάλα και διαφορετικά σύνολα δεδομένων. Το Apache Spark είναι μια ενοποιημένη μηχανή ανάλυσης μεγάλων δεδομένων που παρέχει απόλυτο παραλληλισμό δεδομένων και εξαιρετικές δυνατότητες επεξεργασίας παρτίδων και ροής. Παρέχεται μια τεχνική επισκόπηση της χρήσης του Apache Spark για ανάλυση μεγάλων δεδομένων, εστιάζοντας στον υπολογισμό στη μνήμη, καθιστώντας τον ταχύτερο από άλλα παρόμοια πλαίσια. Επίσης, αναφέρονται οι δυνατότητες πολλαπλών νημάτων και ταυτόχρονης χρήσης του Apache Spark, την αρχιτεκτονική του, τις απαιτήσεις υλικού, το οικοσύστημα και τις περιπτώσεις χρήσης, καθώς και την

εφαρμογή του σε αναδυόμενες τεχνολογίες. Συνοπτικά, εξετάζεται τη χρήση του Apache Spark ως ισχυρού εργαλείου για ανάλυση μεγάλων δεδομένων, εστιάζοντας στην ταχύτητα, την ευελιξία και τις προηγμένες δυνατότητές του (Shaikh, E., Mohiuddin, I., Alufaisan, Y., & Nahvi, I. (2019, November)).

5.6.3 Elastic Stack

Σύμφωνα με τους Vaibhav Aggarwal και B Thangaraju (Feb 02, 2020) το Elastic Stack είναι ένα εργαλείο που βοηθά στην καλύτερη ανάλυση δεδομένων. Τα μεγάλα δεδομένα ορίζονται ως μεγάλα σύνολα δεδομένων που συνδυάζουν δομημένα, ημιδομημένα και μη δομημένα δεδομένα από διάφορες πηγές. Η ανάλυση μεγάλων δεδομένων μπορεί να παρέχει πληροφορίες σχετικά με τα πρότυπα συμπεριφοράς των χρηστών, τις μεταβαλλόμενες τάσεις, την υγεία του συστήματος ή τις ανωμαλίες και μπορεί να βοηθήσει στη βελτίωση της ενσωμάτωσης των πελατών, στον εντοπισμό ανωμαλιών και στον μετριασμό της απάτης.

5.6.4 H2O.ai

Το H2O.ai είναι ένα εργαλείο που μπορεί να χρησιμοποιηθεί για την ανάλυση μεγάλων δεδομένων. Προσφέρει μια σειρά αλγορίθμων μηχανικής μάθησης και αναλυτικών εργαλείων που μπορούν να βοηθήσουν τους χρήστες να εξάγουν πληροφορίες από μεγάλα και πολύπλοκα σύνολα δεδομένων. Με τις δυνατότητες κατανεμημένης επεξεργασίας του, το H2O.ai μπορεί να χειριστεί τεράστιες ποσότητες δεδομένων και να επιτρέψει στους χρήστες να δημιουργήσουν μοντέλα πρόβλεψης που μπορούν να αναπτυχθούν σε πραγματικό χρόνο. Το H2O.ai παρέχει επίσης μια διαισθητική διεπαφή που επιτρέπει στους χρήστες να δημιουργούν, να δοκιμάζουν και να βελτιώνουν μοντέλα μηχανικής μάθησης γρήγορα και εύκολα, ακόμη και αν έχουν μικρή ή καθόλου εμπειρία στον προγραμματισμό. Η ικανότητά του να συνεργάζεται με δημοφιλείς γλώσσες προγραμματισμού και να ενσωματώνεται με άλλες πλατφόρμες μεγάλων δεδομένων το καθιστά ένα εύελικτο εργαλείο για την ανάλυση και την άντληση αξίας από μεγάλα δεδομένα (2021, May 6). H2O.ai)).

Κεφάλαιο 6

Πίνακας Σύνοψης Ευρημάτων

Σε αυτή την ενότητα παρουσιάζεται ένας συνοπτικός πίνακας όλων των ευρημάτων που εντοπίστηκαν για την παρούσα διατριβή. Ο συνοπτικός πίνακας συμπερασμάτων αποτελεί βασικό στοιχείο, καθώς παρέχει μια συνοπτική επισκόπηση των βασικών αποτελεσμάτων και συμπερασμάτων της έρευνας. Αποτελεί κρίσιμο εργαλείο για την κοινοποίηση των κύριων συμπερασμάτων σε ένα ευρύτερο ακροατήριο, συμπεριλαμβανομένων άλλων ερευνητών.

Πίνακας Σύνοψης Ευρημάτων						
Τίτλος Άρθρου	Συγγραφείς	Έτος	Ερευνητικός Σχεδιασμός	Τίτλος Περιοδικού/ Συνεδρίου	Λέξεις Κλειδιά	Κύρια Ευρήματα
1	Guarino, A.	2013	Βιβλιογραφική ανασκόπηση και εννοιολογική ανάλυση	ISSE 2013 Securing Electronic Business Processes: Highlights of the Information Security Solutions Europe 2013 Conference (pp. 197-203)	Ψηφιακή δικανική Μεγάλα Δεδομένα	Ο αφάνητος όγκος, η ποικιλία και η ταχύτητα των ψηφιακών δεδομένων αποτελούν σημαντική πρόκληση για τις ψηφιακές δικανικές έρευνες. Η προσέγγιση των Μεγάλων Δεδομένων μπορεί να συμβάλει στην αντιμετώπιση των προκλήσεων της ψηφιακής δικανικής.
2	Zawoad, S., Hasan, R.	2015	Βιβλιογραφική ανασκόπηση και εννοιολογική ανάλυση	2015 IEEE 17th International Conference on High Performance Computing and Communications, 2015 IEEE 7th International Symposium on Cyberspace Safety and Security, and 2015 IEEE 12th International Conference on Embedded Software and Systems (pp. 1320-1325)	Ψηφιακή δικανική Μεγάλα Δεδομένα, προκλήσεις, προσεγγίσεις, ευκαιρίες	Το άρθρο υπογραμμίζει τις ευκαιρίες και τις προκλήσεις της ψηφιακής δικανικής σε περιβάλλοντα Μεγάλων Δεδομένων και υποδεικνύει την ανάγκη για περισσότερη έρευνα και ανάπτυξη εξειδικευμένων εργαλείων και τεχνικών για τη μερτιστοποίηση των δυνατοτήτων της σε αυτόν τον τομέα.
3	Adedayo, O. M	2016	Ανασκόπηση	2016 IEEE International Conference on Cybercrime and Computer Forensics (ICCCF) (pp. 1-7). IEEE.	Μεγάλα Δεδομένα, ψηφιακή δικανική	Τα Μεγάλα Δεδομένα μπορούν να χρησιμοποιηθούν για την ενίσχυση των ερευνών ψηφιακής δικανικής, αλλά πρέπει να αντιμετωπιστούν οι προκλήσεις που σχετίζονται με την προστασία της ιδιωτικής ζωής, την ασφάλεια και τη δεοντολογία.
4	Banica, L., Burtescu, E., & Enescu, F.	2017	Έρευνα	Scientific Bulletin-Economic Sciences, 16(1), 53-59.	Internet-of-Things, Τριτοβάθμια Εκπαίδευση	Το IoT μπορεί να αποτελέσει ένα πολύτιμο εργαλείο για την τριτοβάθμια εκπαίδευση, αλλά τα ιδρύματα πρέπει να εξετάσουν προσεκτικά τους κινδύνους και τις προκλήσεις που συνδέονται με την εφαρμογή και τη χρήση του. Οι συσκευές IoT μπορούν να χρησιμοποιηθούν για τη βελτίωση της ποιότητας και της αποτελεσματικότητας των εκπαιδευτικών υπηρεσιών, αλλά υπάρχουν επίσης ανησυχίες σχετικά με την ασφάλεια, την ιδιωτικότητα και την προστασία των δεδομένων.

Πίνακας Σύνοψης Ευρημάτων						
Τίτλος Άρθρου	Συγγραφείς	Έτος	Ερευνητικός Σχεδιασμός	Τίτλος Περιοδικού/ Συνεδρίου	Λέξεις Κλειδιά	Κύρια Ευρήματα
5	Comti, M., Debigantha, A., Franke, K., & Watson, S.	2018	Ανασκόπηση της βιβλιογραφίας	Future Generation Computer Systems, 78, 544-546.	Διαδίκτυο των πραγμάτων, ασφάλεια, Δικανική	Επιπτώσεις Οι επηρεασμένες της ασφάλειας και της δικανικής πρέπει να γνωρίζουν τις μοναδικές προκλήσεις που θέτει το IoT και να αναπτύξουν κατάλληλες στρατηγικές και τεχνικές για τον μετριασμό των σχετικών κινδύνων.
6	Du, X., Hargreaves, C., Sheppard, J., Anda, F., Sayakkara, A., Le-Khac, N. A., & Scamlon, M.	2020	Ανασκόπηση της βιβλιογραφίας	Proceedings of the 15th International Conference on Availability, Reliability and Security (pp. 1-10)	Τεχνητή νοημοσύνη, ψηφιακή δικανική	Το άρθρο υπογραμμίζει τα πιθανά οφέλη και τις προκλήσεις της χρήσης της Τεχνητής Νοημοσύνης στην ψηφιακή δικανική έρευνα και παρέχει οδηγίες για τη μελλοντική έρευνα στον τομέα αυτό.
7	Riabi, Y., & Riabi, S.	2018	Ανασκόπηση της βιβλιογραφίας	International Journal of Research and Engineering, 5(9), 524-528	Μεγάλα Δεδομένα, Ανάλυση Μεγάλων Δεδομένων	Το άρθρο παρέχει μια εισαγωγή στα Μεγάλα Δεδομένα και την ανάλυση Μεγάλων Δεδομένων, η οποία μπορεί να είναι χρήσιμη για ερευνητές και επαγγελματίες που ενδιαφέρονται να εφαρμόσουν αυτές τις έννοιες σε διάφορους τομείς, συμπεριλαμβανομένης της ψηφιακής δικανικής έρευνας.
8	Patterson and C. J. Hargreaves	2012	Μελέτη περίπτωσης	Digital Investigation	Ανάλυση διαστυρωμένων δίσκων, ψηφιακά δικανικά χρονοδιαγράμματα	Προτείνεται ότι τα αυτοματοποιημένα ψηφιακά δικανικά χρονοδιαγράμματα θα μπορούσαν να αποτελέσουν χρήσιμο εργαλείο στην ανάλυση πολλαπλών δίσκων, βελτιώνοντας ενδεχομένως την αποδοτικότητα και την αποτελεσματικότητα των δικανικών ερευνών.
9	Y. Al Balushi, H. Shaker, and B. Kumar	2023	Ανασκόπηση της βιβλιογραφίας	1st International Conference on Innovation in Information Technology and Business	Μηχανική μάθηση, ψηφιακή δικανική	Επισημαίνει τις δυνατότητες της μηχανικής μάθησης στην ψηφιακή δικανική υπογραμμίζοντας παράλληλα την ανάγκη προσεκτικής εξέτασης των ρίσκων και νομικών ζητημάτων και τη σημασία της αντιμετώπισης των περιορισμών που σχετίζονται με την ποιότητα των δεδομένων και την προκατάληψη.

Πίνακας Σύνοψης Ευρημάτων						
Τίτλος Άρθρου	Συγγραφείς	Έτος	Ερευνητικός Σχεδιασμός	Τίτλος Περιοδικού/ Συνεδρίου	Λέξεις Κλειδιά	Κύρια Ευρήματα
10	Akinbowale, O. E., Mashigo, P., & Zerihun, M. F.	2023	Μέλητη περίπτωσης	Cogent Business & Management	Δικαστική λογιστική, τεχνολογία Μεγάλων Δεδομένων, εσωτερική απάτη	Η ενσωμάτωση της εγκληματολογικής λογιστικής και των Μεγάλων Δεδομένων μπορεί να συμβάλει στον μετριασμό της εσωτερικής απάτης στον τραπεζικό κλάδο.
11	Thapaliya, S., & Sharma, P. K.	2023	Πειραματική έρευνα	International Journal of Information Security and Privacy (IJISP)	Βαθύ Νευρο-Ασφαρές Δίκτυο, Δικαστική έρευνα στον κυβερνοχώρο, Μεγάλα δεδομένα, IoT	Το προτεινόμενο βελτιστοποιημένο βαθύ νευρωνικό ασφαλές δίκτυο μπορεί να επιτύχει μεγαλύτερη ακρίβεια στην δικαστική έρευνα στον κυβερνοχώρο σε υποδομές IoT με βάση Μεγάλα Δεδομένα.
12	Casey, E.	2011	Ανασκόπηση της βιβλιογραφίας	Academic Press	Ψηφιακά αποδεικτικά στοιχεία, Έγκλημα ηλεκτρονικών υπολογιστών, Δικαστική επιστήμη	Παρέχει γνώσεις σχετικά με τις θεμελιώδεις έννοιες και αρχές της ψηφιακής δικαστικής, οι οποίες μπορούν να χρησιμοποιηθούν ως βάση για περαιτέρω έρευνα και πρακτική.
13	Nelson, B., Phillips, A., & Stewart, C.	2014	Ανασκόπηση της βιβλιογραφίας	Cengage Learning	Δικαστική υπολογιστών, Έρευνες	Αποτελεί πολύτιμο βοήθημα για επαγγελματίες και ερευνητές στον τομέα της ψηφιακής δικαστικής, παρέχοντας πρακτική καθοδήγηση και μελέτες περιπτώσεων.
14	Garfinkel, S. L.	2010	Ανασκόπηση της βιβλιογραφίας	Digital Investigation	Έρευνα ψηφιακής δικαστικής	Το άρθρο αναφέρεται τη σημασία της διεπιστημονικής έρευνας και συνεργασίας για την πρόοδο του τομέα της ψηφιακής δικαστικής και παρέχει έναν οδικό χάρτη για τη μελλοντική έρευνα.

Πίνακας Σύντομης Ευρημάτων						
Τίτλος Άρθρου	Συγγραφείς	Έτος	Ερευνητικός Σχεδιασμός	Τίτλος Περιοδικού/ Συνεδρίου	Λέξεις Κλειδιά	Κύρια Ευρήματα
15	Atzori, L., Iera, A., & Morabito, G.	2010	Ανασκόπηση της βιβλιογραφίας	Computer Networks	Internet of Things	Παρέχει μια επισκόπηση του Διαδικτύου των Πραγμάτων (IoT) και των εφαρμογών του, συμπεριλαμβανομένων των προκλήσεων και των ευκαιριών του IoT σε διάφορους τομείς.
16	Hou, Y.	2018	Μελέτη περίπτωσης	2018 International Conference on Transportation & Logistics, Information & Communication, Smart City (TLICSC 2018)	Αλγόριθμος δέντρου αποφάσεων, Μεγάλα Δεδομένα	Το δέντρο αποφάσεων είναι μια αποτελεσματική και ακριβής προσέγγιση για την ανάλυση Μεγάλων Δεδομένων. Είναι κατάλληλο τόσο για εργασίες ταξινόμησης όσο και για εργασίες πρόβλεψης και μπορεί να χειριστεί δεδομένα που λείπουν.
17	Liu, Y.	2014	Ανασκόπηση της βιβλιογραφίας	Computer modelling & new technologies, 18(12A)	Αλγόριθμος Random Forest, Μεγάλα Δεδομένα	Ο αλγόριθμος Random Forest είναι ένα ισχυρό εργαλείο μηχανικής μάθησης που μπορεί να χειριστεί πολύπλοκα και πολύδιάστατα δεδομένα, καθώς επίσης και να βελτιώσει την ακρίβεια πρόβλεψης.
18	Sarker, I. H.	2021	Ανασκόπηση της βιβλιογραφίας	SN computer science, 2(3)	Μηχανική Μάθηση, Αλγόριθμοι	Παρουσιάζεται μια ολοκληρωμένη κατανόηση των αλγορίθμων μηχανικής μάθησης, η οποία μπορεί να καθοδηγήσει τους ερευνητές να επλέξουν τον κατάλληλο αλγόριθμο για το πρόβλημά τους. Μπορεί να βοηθήσει τους επαγγελματίες να εφαρμόσουν αλγορίθμους μηχανικής μάθησης στις εφαρμογές τους.
19	Shrivastav, L. K., & Kumar, R.	2022	Μελέτη περίπτωσης	Journal of Information Technology Research (JITR), 15(1)	Gradient Boosting Machine, Βαθιά μάθηση, Ανάλυση Μεγάλων Δεδομένων	Επιδείκνουν τον αλγόριθμο GBM και DL, στη χρηματοπιστωτική αγορά, οι οποίες μπορούν να βοηθήσουν τους επενδυτές να λάβουν τεκμηριωμένες αποφάσεις. Καταδεικνύει τις δυνατότητες της DL στην ανάλυση Μεγάλων Δεδομένων, γεγονός που μπορεί να εμπνεύσει περαιτέρω έρευνα στον τομέα αυτό.

Πίνακας Σύνοψης Ευρημάτων						
Τίτλος Άρθρου	Συγγραφείς	Έτος	Ερευνητικός Σχεδιασμός	Τίτλος Περιοδικού/ Συνεδρίου	Λέξεις Κλειδιά	Κύρια Ευρήματα
20	VanderPlas, J.	2016	Ανασκόπηση της βιβλιογραφίας	O'Reilly Media, Inc.	Python, Επιστήμη δεδομένων	<p>Το βιβλίο αυτό παρέχει έναν ολοκληρωμένο οδηγό για τα εργαλεία και τις τεχνικές της επιστήμης των δεδομένων με τη χρήση της Python.</p> <p>Καλύπτει ένα ευρύ φάσμα θεμάτων, όπως η οπτικοποίηση δεδομένων, η μηχανική μάθηση και η επεξεργασία Μεγάλων Δεδομένων.</p>
21	Ferrari, R.	2015	Ανασκόπηση της βιβλιογραφίας	Medical writing, 24(4)	Κριτικές βιβλιογραφίας	<p>Επιστημονικά οφέλη της συγγραφής μιας αφηγηματικής βιβλιογραφικής ανασκόπησης</p> <p>Παρέχει κατευθυντήριες γραμμές για τη συγγραφή μιας αφηγηματικής βιβλιογραφικής ανασκόπησης</p>
22	Saxena, I., Usha, G., et al.	2023	Ανασκόπηση της βιβλιογραφίας	Ανασκόπηση της βιβλιογραφίας	Τεχνητή νοημοσύνη, ψηφιακή δικαστική, επαναστατική προσέγγιση	<p>Διερεύνησε τις δυνατότητες της τεχνητής νοημοσύνης στην ψηφιακή δικαστική και τα οφέλη της.</p>
23	Chen, M., Mao, & Liu, Y.	2014	Έρευνα	Mobile Networks and Applications	Μεγάλα Δεδομένα, Έρευνα	<p>Εξερεύνησε τα Μεγάλα Δεδομένα σε διάφορους τομείς, όπως η κυβερνήση, η υγειονομική περίθαλψη, η εκπαίδευση και η χρηματοδότηση.</p>
						<p>Επιπτώσεις</p> <p>Χρησιμοποιεί και ως εργαλείο για φοιτητές που θέλουν να μάθουν την επιστήμη των δεδομένων.</p> <p>Το άρθρο παρέχει ένα χρήσιμο πλαίσιο για τους ερευνητές προκειμένου να συντάξουν μια βιβλιογραφική ανασκόπηση που συνθέτει την υπάρχουσα γνώση και εντοπίζει τα ερευνητικά κενά.</p> <p>Παρέχει εικόνα των δυνατοτήτων της Τεχνητής Νοημοσύνης στην ψηφιακή δικαστική.</p> <p>Παρέχει μια επισκόπηση των εφαρμογών για Μεγάλα Δεδομένα.</p>

Κεφάλαιο 7

Συμπεράσματα

Στο κεφάλαιο αυτό παρουσιάζονται τα συμπεράσματα της παρούσας μεταπτυχιακής διατριβής, καταγράφεται η συνεισφορά της όπως επίσης οι περιορισμοί και κατευθύνσεις για μελλοντική έρευνα. Αρχικά ανατρέχοντας στα ερευνητικά ερωτήματά μας και ξεκινώντας με τον πρώτο ερευνητικό ερώτημα: **Τι προβλήματα δημιουργούνται από τα Μεγάλα Δεδομένα στην ψηφιακή δικανική ανάλυση;** Συμπερασματικά, η χρήση Μεγάλων Δεδομένων στην ψηφιακή εγκληματολογία παρουσιάζει πολλές προκλήσεις και πολυπλοκότητες που απαιτούν προσεκτική εξέταση και εξειδικευμένα εργαλεία. Ο τεράστιος όγκος, η ποικιλία και η ταχύτητα των δεδομένων που μπορούν να συλλεχθούν από διάφορες πηγές, συμπεριλαμβανομένων των περιβαλλόντων νέφους, δημιουργούν σημαντικές προκλήσεις για τους ερευνητές ψηφιακής εγκληματολογίας. Τα ζητήματα που σχετίζονται με την ασφάλεια των δεδομένων, την προστασία της ιδιωτικής ζωής και τη νομική συμμόρφωση προσθέτουν περαιτέρω επίπεδα πολυπλοκότητας. Επιπλέον, η έλλειψη τυποποίησης στην τεχνολογία και την ανάλυση Μεγάλων Δεδομένων περιπλέκει την ανάπτυξη εξειδικευμένων εργαλείων και τεχνικών για ψηφιακές εγκληματολογικές έρευνες. Η δικανική στο νέφος αποτελεί επίσης πρόκληση λόγω της εξαιρετικά κατανεμημένης και πολύπλοκης αρχιτεκτονικής του νέφους, του μοντέλου χρήσης πολλαπλών μισθωτών, της εικονικοποίησης και της ευμετάβλητης φύσης των ίδιων των δεδομένων. Παράλληλα, η δικανική λογιστική θεωρείται πρόκληση λόγω του ότι η διερεύνηση και η ανίχνευση της απάτης στην ψηφιακή εποχή είναι πολύπλοκη λόγω της συνωμοσίας, της εξαπάτησης και της απόκρυψης, γεγονός που καθιστά δύσκολο τον εντοπισμό της αιτίας και των δραστών.

Προχωρώντας στο επόμενο ερευνητικό ερώτημα: **Ποιες τάσεις έχουν δημιουργηθεί για την “state-of-the-art” διαχείριση των Μεγάλων Δεδομένων στην ψηφιακή δικανική ανάλυση;** Οι σύγχρονες τάσεις διαχείρισης Μεγάλων

Δεδομένων στην ψηφιακή εγκληματολογία εξελίσσονται συνεχώς καθώς αυξάνονται ο όγκος, η ταχύτητα και η ποικιλία των ψηφιακών αποδεικτικών στοιχείων. Μια τάση είναι η ανάπτυξη κατανεμημένων συστημάτων επεξεργασίας, τα οποία μπορούν να διαχειριστούν μεγάλους όγκους δεδομένων κατανέμοντας την επεξεργασία σε πολλαπλούς κόμβους. Μια άλλη τάση είναι η χρήση αλγορίθμων μηχανικής μάθησης και τεχνητής νοημοσύνης για την αυτοματοποίηση του εντοπισμού των σχετικών δεδομένων. Επιπλέον, το υπολογιστικό νέφος χρησιμοποιείται όλο και περισσότερο για την αποθήκευση και την επεξεργασία ψηφιακών αποδεικτικών στοιχείων. Τέλος, αυξάνεται η υιοθέτηση λογισμικού και εργαλείων ανοικτού κώδικα (open source), καθώς παρέχουν οικονομικά αποδοτικές και προσαρμόσιμες λύσεις για τους επαγγελματίες της ψηφιακής δικανικής. Συνολικά, η τάση είναι προς κλιμακούμενες, αποτελεσματικές και αυτοματοποιημένες διαδικασίες ψηφιακής δικανικής που μπορούν να χειριστούν τον αυξανόμενο όγκο και την πολυπλοκότητα των ψηφιακών αποδεικτικών στοιχείων.

Το επόμενο ερευνητικό ερώτημα: **Ποια τα πλαίσια - μοντέλα - τεχνικές αντιμετώπισης των Μεγάλων Δεδομένων (MapReduce , Machine Learning, κ.λπ.) από τη σκοπιά της ψηφιακής δικανικής;** Η άνοδος των Μεγάλων Δεδομένων στην ψηφιακή δικανική οδήγησε στην ανάπτυξη πλαισίων, μοντέλων και τεχνικών για τη διαχείριση και ανάλυση των τεράστιων όγκων δεδομένων που παράγονται. Τα MapReduce, Hadoop είναι μερικά από τα πιο συχνά χρησιμοποιούμενα πλαίσια για την επεξεργασία Μεγάλων Δεδομένων, παρέχοντας δυνατότητες κατανεμημένου υπολογισμού που επιτρέπουν την παράλληλη επεξεργασία τους. Η μηχανική μάθηση είναι μια άλλη δημοφιλής τεχνική που χρησιμοποιείται για την ανάλυση Μεγάλων Δεδομένων στην ψηφιακή δικανική, με αλγορίθμους όπως οι Random Forest, Naïve Bayes, Gradient Boosting Machines και Decision trees. Άλλες τεχνικές, όπως η ομαδοποίηση, η οπτικοποίηση και η εξόρυξη κανόνων συσχέτισης, χρησιμοποιούνται επίσης για την απόκτηση γνώσεων σχετικά με τα Μεγάλα Δεδομένα. Ωστόσο, οι προκλήσεις που θέτουν τα Μεγάλα Δεδομένα στην ψηφιακή εγκληματολογία απαιτούν πιο εξειδικευμένα εργαλεία και τεχνικές, για την αντιμετώπιση της πολυπλοκότητας των κατανεμημένων αρχιτεκτονικών που βασίζονται στο νέφος, του απορρήτου των δεδομένων και των νομικών περιορισμών.

Επιπρόσθετα, ένα άλλο ερευνητικό ερώτημα: **Ποιος διαφαίνεται ως ο οδικός χάρτης για τον αποτελεσματικό χειρισμό των Μεγάλων Δεδομένων στην ψηφιακή δικανική;** Η ψηφιακή δικανική αποκτά ολοένα και μεγαλύτερη σημασία, καθώς όλο και περισσότεροι οργανισμοί βασίζονται στην ψηφιακή τεχνολογία για την αποθήκευση και την επεξεργασία των δεδομένων τους. Η εκθετική αύξηση του όγκου, της ποικιλίας και της ταχύτητας των δεδομένων απαιτεί έναν οδικό χάρτη για τον αποτελεσματικό χειρισμό των Μεγάλων Δεδομένων. Ο οδικός χάρτης θα πρέπει να περιλαμβάνει διάφορα βασικά στοιχεία, όπως η απόκτηση, η αποθήκευση, η ανάλυση και η οπτικοποίηση δεδομένων. Για τον αποτελεσματικό χειρισμό των Μεγάλων Δεδομένων στην ψηφιακή εγκληματολογία, θα πρέπει να αξιοποιηθούν προηγμένα εργαλεία και τεχνικές, όπως MapReduce, μηχανική μάθηση και τεχνητή νοημοσύνη. Επιπλέον, μπορούν να χρησιμοποιηθούν πλαίσια και μοντέλα όπως το μοντέλο ροής εργασιών ανάλυσης της εγκληματολογίας στον κυβερνοχώρο και το πλαίσιο διερεύνησης της ψηφιακής δικανικής για την καθοδήγηση της όλης διαδικασίας. Καθώς αναδύονται νέες τεχνολογίες και τεχνικές, ο οδικός χάρτης πρέπει να είναι ευέλικτος και προσαρμόσιμος, ώστε να διασφαλίζεται ότι οι ειδικοί της ψηφιακής δικανικής είναι εξοπλισμένοι για να χειρίζονται τον συνεχώς αυξανόμενο όγκο, την ποικιλία και την ταχύτητα των δεδομένων.

Σημαντικό ερευνητικό ερώτημα αποτελεί: **Ποιες είναι οι ευκαιρίες των Μεγάλων Δεδομένων στη ψηφιακή δικανική;** Τα Μεγάλα Δεδομένα μπορούν να παρέχουν ευκαιρίες για ψηφιακή εγκληματολογία, αποκαλύπτοντας προηγουμένως άγνωστες γνώσεις μέσω της ανάλυσης μεγάλων συνόλων δεδομένων. Με τα κατάλληλα εργαλεία και τεχνικές, η ανίχνευση, η συλλογή και η ανάλυση μεγάλου όγκου δεδομένων μπορεί να δώσει νέες γνώσεις για εγκληματικά περιστατικά. Επιπλέον, προτείνεται ένα εννοιολογικό μοντέλο εγκληματολογίας μεγάλων δεδομένων, το οποίο μπορεί να προσφέρει νέες γνώσεις σχετικά με τη διαπίστωση εγκληματικών γεγονότων. Συνολικά, η διαθεσιμότητα Μεγάλων Δεδομένων στην ψηφιακή εγκληματολογία μπορεί να προσφέρει στους ερευνητές την ευκαιρία να εξάγουν σημαντικές πληροφορίες από τα συσσωρευμένα δεδομένα.

Ανακεφαλαιώνοντας, αποτελεί κρίσιμο το εξής ερώτημα: **Πώς επηρεάζουν νομικά και ηθικά ζητήματα τη συλλογή και ανάλυση δεδομένων μεγάλου όγκου**

για εγκληματολογικούς σκοπούς και ποια μέτρα μπορούν να ληφθούν για να εξασφαλιστεί η συμμόρφωση με τους σχετικούς νόμους και κανονισμούς; Η ανάπτυξη και εφαρμογή ψηφιακών συστημάτων δικαιοσύνης εγείρει διάφορα ηθικά ζητήματα όπως το απόρρητο και το απόρρητο, η άρνηση και η σύγχυση, η ακρίβεια και η αποτελεσματικότητα, η χειραγώγηση, η παραποίηση, η υποκατάσταση και η μείωση. Οι ηθικές ανησυχίες επιδεινώνονται από τα μοναδικά χαρακτηριστικά των ψηφιακών συστημάτων δικαιοσύνης που υπερβαίνουν τα τυπικά συστήματα πληροφοριών. Επίσης, εντοπίστηκαν άλλα ηθικά ζητήματα που σχετίζονται με τα ψηφιακά συστήματα δικαιοσύνης, συμπεριλαμβανομένης της χρήσης εργαλείων, ανεπαρκών νομικών μέτρων για την αντιμετώπιση σύνθετων ζητημάτων, την ανάγκη για συνεχή αναθεώρηση των νομοθετικών πλαισίων και την εφαρμογή τεχνικών μέτρων όπως η εκπαίδευση, η αναδιάρθρωση και η ρύθμιση. Θεσμική μεταρρύθμιση. Για την αντιμετώπιση ηθικών ζητημάτων που προκύπτουν από την ανάπτυξη και την εφαρμογή ψηφιακών συστημάτων δικαιοσύνης, συνιστώνται προσωρινά μέτρα που αντιμετωπίζουν άμεσα τις ανησυχίες σχετικά με το απόρρητο των ενδιαφερομένων που προκύπτουν από τεχνικά μέτρα όπως η τεχνολογική πρόοδος, η νομική παρέμβαση και εκπαίδευση και ο επανασχεδιασμός και η αναδιάρθρωση συστημάτων. Είναι σημαντικό να σημειωθεί ότι τα μέτρα αυτά θα πρέπει να εφαρμοστούν εγκαίρως και αποτελεσματικά, ώστε να μετριαστούν οι πιθανές ηθικές παραβιάσεις.

Επίλογος

Τα τελευταία χρόνια, ο πολλαπλασιασμός των ψηφιακών συσκευών και ο τεράστιος όγκος δεδομένων που παράγουν έχουν δημιουργήσει σημαντικές προκλήσεις για την ψηφιακή δικανική. Ο τεράστιος όγκος, η ποικιλία και η ταχύτητα των δεδομένων έχουν καταστήσει όλο και πιο δύσκολο για τους ερευνητές να αναλύσουν και να εξάγουν σχετικές πληροφορίες. Ωστόσο, καθώς η τεχνολογία συνεχίζει να εξελίσσεται, το ίδιο συμβαίνει και με τις τεχνικές και τα εργαλεία που είναι διαθέσιμα για τη διαχείριση και την ανάλυση Μεγάλων Δεδομένων στην ψηφιακή δικανική.

Όπως περιγράφεται στην παρούσα έρευνα, οι σύγχρονες τάσεις στη διαχείριση Μεγάλων Δεδομένων, όπως η χρήση καταναμημένων συστημάτων αρχείων, έχουν επιτρέψει στους ερευνητές να αποθηκεύουν και να επεξεργάζονται αποτελεσματικά τεράστιες ποσότητες δεδομένων. Επιπλέον, αναφέρθηκαν πλαίσια και μοντέλα όπως το MapReduce, την τεχνητή νοημοσύνη, και τη μηχανική μάθηση για την ανάλυση και την εξαγωγή πολύτιμων πληροφοριών από τα Μεγάλα Δεδομένα στην ψηφιακή δικανική.

Ωστόσο, η χρήση των Μεγάλων Δεδομένων στην ψηφιακή δικανική εγείρει νομικά και ηθικά ζητήματα που πρέπει να εξεταστούν προσεκτικά. Αυτό περιλαμβάνει τη λήψη της κατάλληλης νομικής άδειας, την προστασία της ιδιωτικής ζωής και του απορρήτου των ατόμων και τη διασφάλιση της αξιοπιστίας και της ακρίβειας των δεδομένων.

Καθώς η χρήση ψηφιακών συσκευών συνεχίζει να αυξάνεται, οι προκλήσεις που θέτουν τα Μεγάλα Δεδομένα στην ψηφιακή δικανική θα συνεχίσουν να εξελίσσονται. Επομένως, με την υιοθέτηση νέων τεχνικών και εργαλείων και την προσεκτική εξέταση νομικών και δεοντολογικών προβληματισμών, οι ερευνητές μπορούν να διαχειρίζονται και να αναλύουν αποτελεσματικά τα Μεγάλα Δεδομένα, εξάγοντας πολύτιμες πληροφορίες για την υποστήριξη των ερευνών τους.

Βιβλιογραφία

Adedayo, O. M. (2016, June). Big data and digital forensics. In 2016 IEEE International Conference on Cybercrime and Computer Forensic (ICCCF) (pp. 1- 7). IEEE.

Akinbowale, O. E., Mashigo, P., & Zerihun, M. F. (2023). The integration of forensic accounting and big data technology frameworks for internal fraud mitigation in the banking industry. *Cogent Business & Management*, 10(1), 2163560.

Al Balushi, Y., Shaker, H., & Kumar, B. (2023, January). The Use of Machine Learning in Digital Forensics. In 1st International Conference on Innovation in Information Technology and Business (ICIITB 2022) (pp. 96-113). Atlantis Press.

Arshad, H., Jantan, A. B., & Abiodun, O. I. (2018). Digital forensics: review of issues in scientific validation of digital evidence. *Journal of Information Processing Systems*, 14(2), 346-376.

Atzori, L., Iera, A., & Morabito, G. (2010). The internet of things: A survey. *Computer networks*, 54(15), 2787-2805.

Balogun, A. M., & Zuva, T. (2017). Open ethical issues in digital forensic systems. *International Journal of eBusiness and eGovernment Studies*, 9(1), 55-69.

Banica, L., Burtescu, E., & Enescu, F. (2017). The impact of internet-of-things in higher education. *Scientific Bulletin-Economic Sciences*, 16(1), 53-59.

Carrier, B. (2002). Open source digital forensics tools: The legal argument.

Casey, E. (2011). *Digital evidence and computer crime: Forensic science, computers, and the internet*. Academic press.

- Chen, M., Mao, S., & Liu, Y. (2014). Big data: A survey. *Mobile networks and applications*, 19(2), 171-209.
- Conti, M., Dehghantanha, A., Franke, K., & Watson, S. (2018). Internet of Things security and forensics: Challenges and opportunities. *Future Generation Computer Systems*, 78, 544-546.
- Du, X., Hargreaves, C., Sheppard, J., Anda, F., Sayakkara, A., Le-Khac, N. A., & Scanlon, M. (2020, August). SoK: Exploring the state of the art and the future potential of artificial intelligence in digital forensic investigation. In *Proceedings of the 15th International Conference on Availability, Reliability and Security* (pp. 1-10).
- Ferrari, R. (2015). Writing narrative style literature reviews. *Medical writing*, 24(4), 230-235.
- Garfinkel, S. L. (2010). Digital forensics research: The next 10 years. *digital investigation*, 7, S64-S73., Chicago
- George, G., Haas, M. R., & Pentland, A. (2014). Big data and management. *Academy of management Journal*, 57(2), 321-326.
- Ghandan Gaur (Oct 24, 2010). What is Big Data: Characteristics, Challenges, Tools & Use Cases. XENONSTACK. <https://www.xenonstack.com/blog/what-is-big-data>
- Guarino, A. (2013). Digital forensics as a big data challenge. In *ISSE 2013 securing electronic business processes* (pp. 197-203). Springer Vieweg, Wiesbaden.
- Gubbi, J., Buyya, R., Marusic, S., & Palaniswami, M. (2013). Internet of Things (IoT): A vision, architectural elements, and future directions. *Future generation computer systems*, 29(7), 1645-1660.
- Hassani, H., & Silva, E. S. (2015). Forecasting with big data: A review. *Annals of Data Science*, 2, 5-19.

Hou, Y. (2018, December). Decision tree algorithm for big data analysis. In 2018 International Conference on Transportation & Logistics, Information & Communication, Smart City (TLICSC 2018) (pp. 270-274). Atlantis Press.

<https://h2o.ai/> (2021, May 6). H2O.ai

Liu, Y. (2014). Random forest algorithm in big data environment. *Computer modelling & new technologies*, 18(12A), 147-151.

Neilson, A., Daniel, B., & Tjandra, S. (2019). Systematic review of the literature on big data in the transportation domain: Concepts and applications. *Big Data Research*, 17, 35-44.

Nelson, B., Phillips, A., & Steuart, C. (2014). *Guide to computer forensics and investigations*. Cengage Learning.

Nogueira, R. R., & Gueiber, E. (2014). Big data: An approach on Apache Hadoop. *Iberoamerican Journal of Applied Computing*, 4(3).

Olaronke, I., & Oluwaseun, O. (2016, December). Big data in healthcare: Prospects, challenges and resolutions. In 2016 Future technologies conference (FTC) (pp. 1152-1157). IEEE.

Patterson, J., & Hargreaves, C. J. (2012). The potential for cross-drive analysis using automated digital forensic timelines.

Raghavan, S. (2013). Digital forensic research: current state of the art. *Csi Transactions on ICT*, 1, 91-114.

Riahi, Y., & Riahi, S. (2018). Big data and big data analytics: concepts, types and technologies. *International Journal of Research and Engineering*, 5(9), 524-528.

https://www.researchgate.net/publication/328783489_Big_Data_and_Big_Data_Analytics_Concepts_Types_and_Technologies/

Ritesh Pathak (Jan 23, 2021). What is Big Data Analytics? Definition, Advantages, and

Types. AnalyticsSteps. <https://www.analyticssteps.com/blogs/what-big-data-analytics-definition-advantages-and-types/>

Sachin P Bappalige. (Aug 26, 2014). An introduction to Apache Hadoop for big data. Opensource.com. <https://opensource.com/life/14/8/intro-apache-hadoop-big-data>

Sarker, I. H. (2021). Machine learning: Algorithms, real-world applications and research directions. SN computer science, 2(3), 160.

Saxena, I., Usha, G., Vinoth, N. A. S., Veena, S., & Nancy, M. (2023). The Future of Artificial Intelligence in Digital Forensics: A Revolutionary Approach. In Artificial Intelligence and Blockchain in Digital Forensics (pp. 133-151). River Publishers.

Shaikh, E., Mohiuddin, I., Alufaisan, Y., & Nahvi, I. (2019, November). Apache spark: A big data processing engine. In 2019 2nd IEEE Middle East and North Africa COMMUNICATIONS Conference (MENACOMM) (pp. 1-6). IEEE.

Shrivastav, L. K., & Kumar, R. (2022). Gradient Boosting Machine and Deep Learning Approach in Big Data Analysis: A Case Study of the Stock Market. Journal of Information Technology Research (JITR), 15(1), 1-20.

Sruthi E R (17, June, 2021). Understand Random Forest Algorithms With Examples. Analytics Vidhya. <https://www.analyticsvidhya.com/blog/2021/06/understanding-random-forest/>

Thapaliya, S., & Sharma, P. K. (2023). Optimized Deep Neuro Fuzzy Network for Cyber Forensic Investigation in Big Data-Based IoT Infrastructures. International Journal of Information Security and Privacy (IJISP), 17(1), 1-22.

Vaibhav Aggarwal – B Thangaraju. (Feb 02, 2020). How to Analyse Big Data Using the ELK Stack. OpenSourceForU.com. <https://www.opensourceforu.com/2020/02/how-to-analyse-big-data-using-the-elk-stack/>

VanderPlas, J. (2016). Python data science handbook: Essential tools for working with data. " O'Reilly Media, Inc."

Zawoad, S., & Hasan, R. (2015, August). Digital forensics in the age of big data: Challenges, approaches, and opportunities. In 2015 IEEE 17th International Conference on High Performance Computing and Communications, 2015 IEEE 7th International Symposium on Cyberspace Safety and Security, and 2015 IEEE 12th International Conference on Embedded Software and Systems (pp. 1320-1325). IEEE.

Σούμπλης, Α. (2021). Προκλήσεις στην Ανάλυση Μεγάλων Δεδομένων. [Original source: <https://studycrumb.com/alphabetizer>]