



ΠΑΝΕΠΙΣΤΗΜΙΟ ΘΕΣΣΑΛΙΑΣ

ΠΟΛΥΤΕΧΝΙΚΗ ΣΧΟΛΗ

ΤΜΗΜΑ ΗΛΕΚΤΡΟΛΟΓΩΝ ΜΗΧΑΝΙΚΩΝ ΚΑΙ ΜΗΧΑΝΙΚΩΝ ΥΠΟΛΟΓΙΣΤΩΝ

“Εφαρμογές Κυβερνοασφάλειας & Ανωνυμοποίησης Δεδομένων”

Διπλωματική Εργασία

Ρήγας Νικόλαος

Επιβλέπων: Δασκαλοπούλου Ασπασία

Οκτώβριος 2023



ΠΑΝΕΠΙΣΤΗΜΙΟ ΘΕΣΣΑΛΙΑΣ

ΠΟΛΥΤΕΧΝΙΚΗ ΣΧΟΛΗ

ΤΜΗΜΑ ΗΛΕΚΤΡΟΛΟΓΩΝ ΜΗΧΑΝΙΚΩΝ ΚΑΙ ΜΗΧΑΝΙΚΩΝ ΥΠΟΛΟΓΙΣΤΩΝ

“Εφαρμογές Κυβερνοασφάλειας & Ανωνυμοποίησης Δεδομένων”

Διπλωματική Εργασία

Ρήγας Νικόλαος

Επιβλέπων: Δασκαλοπούλου Ασπασία

Οκτώβριος 2023



UNIVERSITY OF THESSALY

SCHOOL OF ENGINEERING

DEPARTMENT OF ELECTRICAL AND COMPUTER ENGINEERING

“Cyber Security & Data Anonymization Applications”

Diploma Thesis

Rigas Nikolaos

Supervisor: Daskalopulu Aspassia

October 2023

Εγκρίνεται από την Επιτροπή Εξέτασης:

Επιβλέπουσα

Δασκαλοπούλου Ασπασία

Αναπληρώτρια Καθηγήτρια, Τμήμα Ηλεκτρολόγων Μηχανικών και
Μηχανικών Υπολογιστών, Πανεπιστήμιο Θεσσαλίας

Μέλος

Βασιλακόπουλος Μιχαήλ

Καθηγητής, Τμήμα Ηλεκτρολόγων Μηχανικών και Μηχανικών Υπολογιστών,
Πανεπιστήμιο Θεσσαλίας

Μέλος

Χροναίος Αλέξανδρος

Καθηγητής, Τμήμα Ηλεκτρολόγων Μηχανικών και Μηχανικών Υπολογιστών,
Πανεπιστήμιο Θεσσαλίας

ΥΠΕΥΘΥΝΗ ΔΗΛΩΣΗ ΠΕΡΙ ΑΚΑΔΗΜΑΪΚΗΣ ΔΕΟΝΤΟΛΟΓΙΑΣ ΚΑΙ ΠΝΕΥΜΑΤΙΚΩΝ

ΔΙΚΑΙΩΜΑΤΩΝ

Με πλήρη επίγνωση των συνεπειών του νόμου περί πνευματικών δικαιωμάτων, δηλώνω ρητά ότι η παρούσα διπλωματική εργασία, καθώς και τα ηλεκτρονικά αρχεία και πηγαίοι κώδικες που αναπτύχθηκαν ή τροποποιήθηκαν στα πλαίσια αυτής της εργασίας, αποτελούν αποκλειστικά προϊόν προσωπικής μου εργασίας, δεν προσβάλλουν οποιασδήποτε μορφής δικαιώματα διανοητικής ιδιοκτησίας, προσωπικότητας και προσωπικών δεδομένων τρίτων, δεν περιέχουν έργα/εισφορές τρίτων για τα οποία απαιτείται άδεια των δημιουργών/δικαιούχων και δεν είναι προϊόν μερικής ή ολικής αντιγραφής, οι πηγές δε που χρησιμοποιήθηκαν περιορίζονται στις βιβλιογραφικές αναφορές και μόνον και πληρούν τους κανόνες της επιστημονικής παράθεσης. Τα σημεία όπου έχω χρησιμοποιήσει ιδέες, κείμενο, αρχεία ή/και πηγές άλλων συγγραφέων αναφέρονται ευδιάκριτα στο κείμενο με την κατάλληλη παραπομπή και η σχετική αναφορά περιλαμβάνεται στο τμήμα των βιβλιογραφικών αναφορών με πλήρη περιγραφή. Δηλώνω επίσης ότι τα αποτελέσματα της εργασίας δεν έχουν χρησιμοποιηθεί για την απόκτηση άλλου πτυχίου. Αναλαμβάνω πλήρως, ατομικά και προσωπικά, όλες τις νομικές και διοικητικές συνέπειες που δύναται να προκύψουν στην περίπτωση κατά την οποία αποδειχθεί, διαχρονικά, ότι η εργασία αυτή ή τμήμα της δεν μου ανήκει διότι είναι προϊόν λογοκλοπής.

Ο/Η Δηλών/ούσα

Ρήγας Νικόλαος

DISCLAIMER ON ACADEMIC ETHICS AND INTELLECTUAL PROPERTY RIGHTS

Being fully aware of the implications of copyright laws, I expressly state that this diploma thesis, as well as the electronic files and source codes developed or modified in the course of this thesis, are solely the product of my personal work and do not infringe any rights of intellectual property, personality and personal data of third parties, do not contain work / contributions of third parties for which the permission of the authors / beneficiaries is required and are not a product of partial or complete plagiarism, while the sources used are limited to the bibliographic references only and meet the rules of scientific citing. The points where I have used ideas, text, files and / or sources of other authors are clearly mentioned in the text with the appropriate citation and the relevant complete reference is included in the bibliographic references section. I also declare that the results of the work have not been used to obtain another degree. I fully, individually and personally undertake all legal and administrative consequences that may arise in the event that it is proven, in the course of time, that this thesis or part of it does not belong to me because it is a product of plagiarism.

The Declarant

Rigas Nikolaos

Ευχαριστίες

Στην οικογένειά μου και στους καθηγητές μου.

“Εφαρμογές Κυβερνοασφάλειας & Ανωνυμοποίησης Δεδομένων”

Ρήγας Νικόλαος

Περίληψη

Σε έναν συνεχώς εξελισσόμενο ψηφιακό κόσμο, οι έννοιες της κυβερνοασφάλειας, της τεχνητής νοημοσύνης (TN) της ανωνυμοποίησης των δεδομένων των τεχνολογιών blockchain αποτελούν κρίσιμης σημασίας παράγοντες. Η παρούσα διατριβή ξεκινά μια ολοκληρωμένη διερεύνηση αυτών των τομέων, ξετυλίγοντας τις περιπλοκές, τις προκλήσεις και τις δυνατότητές τους. Η έρευνα περιηγείται σε ένα ποικιλόμορφο τοπίο, καλύπτοντας την κυβερνοασφάλεια σε περιβάλλοντα πραγματικού κόσμου, τις ειδικές τεχνολογικές προκλήσεις και τις ηθικές διαστάσεις της υιοθέτησης της TN και της τεχνολογίας blockchain.

Η διατριβή ξεκινά με την εμβάθυνση στην κρίσιμη σημασία της κυβερνοασφάλειας. Ανακαλύπτει ένα δυναμικό και εξελισσόμενο τοπίο απειλών, αποκαλύπτοντας την πολυπλοκότητα των απειλών στον κυβερνοχώρο - από τις κοινές επιθέσεις phishing έως την κρατικά χρηματοδοτούμενη κυβερνοκατασκοπεία. Τα ευρήματα υπογραμμίζουν την ανάγκη για συνεχή καινοτομία, προσαρμοστικές στρατηγικές ασφαλείας και ισχυρή εκπαίδευση των χρηστών για την ενίσχυση της άμυνας στον κυβερνοχώρο. Στη συνέχεια, η έρευνα εξετάζει τις ειδικές τεχνολογικές προκλήσεις, συμπεριλαμβανομένης της ασφάλειας των Web servers και της κρυπτογράφησης. Φωτίζει τη σημασία των πρακτικών ασφαλούς ανάπτυξης, των τακτικών ελέγχων ασφαλείας και των αλγορίθμων κρυπτογράφησης ως βασικούς παράγοντες διασφάλισης της ομαλής λειτουργίας του ψηφιακού κόσμου εποχή. Επιπλέον, υπογραμμίζεται ιδιαίτερα το γεγονός ότι η υιοθέτηση της τεχνητής νοημοσύνης εισάγει τόσο υποσχέσεις όσο και κινδύνους. Η έρευνα διευκρινίζει τις μετασχηματιστικές δυνατότητες της TN, τις ποικίλες εφαρμογές της και τις ηθικές εκτιμήσεις για την υπεύθυνη ανάπτυξη της TN.

Επιπλέον, στην εποχή της ιδιωτικότητας των δεδομένων, η ανωνυμοποίηση και οι τεχνολογίες blockchain βρίσκονται στο επίκεντρο. Η διατριβή αναδεικνύει την ανωνυμοποίηση ως ένα ισχυρό εργαλείο για τη διατήρηση της ιδιωτικότητας με παράλληλη διατήρηση της χρησιμότητας των δεδομένων. Επιπλέον, διερευνά την τεχνολογία blockchain ως παράγοντα εμπιστοσύνης σε διάφορους τομείς, αναδεικνύοντας τις αρχές της αποκέντρωσης, της διαφάνειας και της κρυπτογραφικής ασφάλειας. Τέλος, η παρούσα διατριβή προσφέρει πολύτιμες γνώσεις για επαγγελματίες, υπεύθυνους χάραξης πολιτικής και ερευνητές τονίζοντας τον καθοριστικό ρόλο της καινοτομίας, της προσαρμοστικότητας αλλά και της διατήρησης της ηθικής για την εξασφάλιση ενός ασφαλούς και αξιόπιστου ψηφιακού μέλλοντος.

Λέξεις-κλειδιά:

Κυβερνοασφάλεια, Τεχνητή Νοημοσύνη (TN), Τεχνολογία Blockchain, Ψηφιακή Ασφάλεια, Κρυπτογράφηση, Ασφάλεια Διακομιστών Ιστού, Ανωνυμοποίηση, Αμυντική Τεχνητή Νοημοσύνη, Επιθετική Τεχνητή Νοημοσύνη, Προστασία Δεδομένων

Diploma Thesis

“Cyber Security & Data Anonymization Applications”

Rigas Nikolaos

Abstract

In an increasingly digital world, the paramount importance of cybersecurity, artificial intelligence (AI), and blockchain technologies cannot be overstated. This thesis embarks on a comprehensive exploration of these domains, unravelling their intricacies, challenges, and transformative potential. The research navigates through a diverse landscape, covering cybersecurity in real-world environments, technology-specific challenges, and the ethical dimensions of AI and blockchain adoption.

The thesis commences by delving into the critical importance of cybersecurity. It unearths a dynamic and evolving threat landscape, revealing the sophistication of cyber threats—from common phishing attacks to state-sponsored cyber-espionage. The findings underscore the necessity for continuous innovation, adaptive security strategies, and robust user education in bolstering cybersecurity defenses. Moving forward, the research examines technology-specific challenges, including web server security and encryption. It illuminates the significance of secure development practices, regular security audits, and quantum-resistant encryption algorithms as essential safeguards in the digital age. Moreover, the adoption of AI introduces both promise and perils. The research elucidates AI's transformative potential, its diverse applications, and the ethical considerations in responsible AI development.

In the era of data privacy, anonymization and blockchain technologies take center stage. The thesis showcases anonymization as a powerful tool for preserving privacy while maintaining data utility. Furthermore, it explores blockchain as a trust-enabler in various sectors, highlighting its principles of

decentralization, transparency, and cryptographic security. By unravelling the intricacies of these domains and presenting pragmatic solutions to challenges, this thesis offers valuable insights for practitioners, policy-makers, and researchers. It emphasizes the pivotal role of innovation, adaptability, and ethics in navigating the digital frontier, ensuring a secure and trustworthy digital future.

Keywords:

Cybersecurity, Artificial Intelligence (AI), Blockchain Technology, Digital Security, Encryption, Multi-factor Authentication (MFA), Web Server Security, Anonymization, Defensive AI, Offensive AI, Data Privacy

Πίνακας Περιεχομένων

Ευχαριστίες	VIII
Περίληψη	IX
Abstract	XI
Ευρετήριο Εικόνων	XV
Πρόλογος	1
1. Κυβερνοασφάλεια	2
1.1 Η σημασία της Κυβερνοασφάλειας	2
1.1.1 Web servers	3
1.1.2 Mobile Networks	4
1.1.3 Encryption	7
1.1.4 APT's και Στοχευμένες επιθέσεις	10
1.2 Ανάλυση της κυβερνοασφάλειας στο πραγματικό περιβάλλον	13
1.3 Χρήση ελέγχου ταυτότητας πολλαπλών παραγόντων	15
1.4 Απόρρητα Δεδομένα	17
2. Τεχνητή Νοημοσύνη (TN) στην Πληροφορική	21
2.1 Τεχνολογίες TN	21
2.1.1 Αλγόριθμοι Μηχανικής Μάθησης - Machine Learning (ML)	22
2.1.2 Deep Learning (DL)	42
2.1.3 Natural Language Generation	44
2.1.4 Speech Recognition	45
2.1.5 Biometrics	46
2.2 Εφαρμογή και Είδη TN	49
2.2.1 Είδη TN	49
2.2.2 Εφαρμογές TN	54
2.3 Πλεονεκτήματα της TN	56
2.4 Κίνδυνοι της TN	59
3. Defensive AI	61
3.1 Neural Nets	62
3.2 Intelligent Agents	65
3.3 Learning	66
3.4 Επίλυση περιορισμών	67

4. Offensive AI	69
4.1 Επιθέσεις με χρήση Τεχνητής Νοημοσύνης	69
4.2 Επιθέσεις κατά της Τεχνητής Νοημοσύνης – Μηχανική Μάθηση	71
4.2.1 Μεθοδολογίες επίθεσης	72
4.2.2 Κίνδυνοι και επιπτώσεις	72
4.2.3 Στρατηγικές μετριασμού	73
4.3 Επιθετική Τεχνητή Νοημοσύνη έναντι Οργανισμών	73
4.3.1 Το μοντέλο της επίθεσης	74
5. Anonymization	76
5.1 Εισαγωγή	76
5.2 Μεθοδολογίες/Τεχνικές	76
5.3 Blockchain Technique	78
6. Συζήτηση και συμπεράσματα	81
6.1. Κύρια ευρήματα	81
6.2. Προκλήσεις και λύσεις	82
Βιβλιογραφία-Παραπομπές	85

Ευρετήριο Εικόνων

Εικόνα 1 : Συστατικά στοιχεία των δικτύων κινητής τηλεφωνίας	6
Εικόνα 2 : Συμμετρική κρυπτογράφηση	8
Εικόνα 3 : Ασύμμετρη κρυπτογράφηση	9
Εικόνα 4 : Ο κύκλος ζωής των APTs [13]	12
Εικόνα 5 : Σχηματική αναπαράσταση λειτουργίας MFA	17
Εικόνα 6 : Αρχιτεκτονική ενός Τεχνητού Νευρωνικού Δικτύου [37]	31
Εικόνα 7 : Αρχιτεκτονική Δικτύου RBF [30]	33
Εικόνα 8 : Παράδειγμα Τεχνητού Νευρωνικού Δικτύου	62
Εικόνα 10 : Τομείς εφαρμογής Blockchain	80

Πρόλογος

Σε μια εποχή που ορίζεται από την ταχεία ψηφιοποίηση σχεδόν κάθε πτυχής της ανθρώπινης ζωής, οι έννοιες της κυβερνοασφάλειας, της τεχνητής νοημοσύνης και των τεχνολογιών blockchain έχουν αναδειχθεί ως βασικά συστατικά του ψηφιακού κόσμου. Η παρούσα διατριβή ξεκινά ένα ταξίδι εξερεύνησης, έρευνας και ανάλυσης, εμβαθύνοντας στην περίπλοκη αλληλεπίδραση αυτών των τομέων στο σύγχρονο ψηφιακό τοπίο. Η ψηφιακή εποχή φέρνει μαζί της απaráμιλλες ευκαιρίες για πρόοδο, συνδεσιμότητα και καινοτομία. Ωστόσο, εισάγει επίσης άνευ προηγουμένου προκλήσεις που απαιτούν βαθιά κατανόηση και προληπτική απάντηση. Σε αυτό το πλαίσιο, αυτή η διατριβή προσπαθεί να αποσαφηνίσει τις πολύπλευρες πτυχές της κυβερνοασφάλειας, τις μετασχηματιστικές δυνατότητες της τεχνητής νοημοσύνης και τα χαρακτηριστικά που επιτρέπουν την εμπιστοσύνη της τεχνολογίας blockchain. Επιπλέον, αναγνωρίζει τον κεντρικό ρόλο που παίζει η τεχνολογία στη διαμόρφωση της ζωής, των επιχειρήσεων και των κοινωνιών μας. Στα πλαίσια της παρούσας διατριβής επιχειρείται η εξέταση και η κατανόηση τις περίπλοκων αυτών εξελισσόμενων τομέων, η ανακάλυψη των κρυμμένων απειλών και η ανάδειξη των απαραίτητων διασφαλίσεων και της ηθικής ευθύνης.

1. Κυβερνοασφάλεια

1.1 Η σημασία της Κυβερνοασφάλειας

Σε μια εποχή που χαρακτηρίζεται από πρωτοφανείς τεχνολογικές εξελίξεις και ένα διαρκώς διευρυνόμενο ψηφιακό τοπίο, η έννοια της κυβερνοασφάλειας έχει αναδειχθεί σε ένα από τα πιο κρίσιμα και πιεστικά ζητήματα της εποχής μας. Καθώς ο κόσμος διασυνδέεται ολοένα και περισσότερο μέσω του διαδικτύου, της τεχνολογίας των πληροφοριών και του πολλαπλασιασμού των ψηφιακών συσκευών, η προστασία των ευαίσθητων δεδομένων, των υποδομών και της ιδιωτικής ζωής έχει καταστεί υψίστης σημασίας. Η κυβερνοασφάλεια, που συχνά περιγράφεται ως η επιστήμη της προστασίας των ψηφιακών συστημάτων, δικτύων και δεδομένων από κλοπή, ζημία ή μη εξουσιοδοτημένη πρόσβαση, διαδραματίζει καθοριστικό ρόλο στη διατήρηση της σταθερότητας και της ασφάλειας του σύγχρονου κόσμου [1].

Στον πυρήνα της, η κυβερνοασφάλεια αντιπροσωπεύει την πρακτική της προστασίας του ψηφιακού χώρου από μια πληθώρα απειλών, που περιλαμβάνει από κυβερνοεπιθέσεις και παραβιάσεις δεδομένων έως κλοπή πληροφοριών και ψηφιακή κατασκοπεία. Σε μια εποχή που χαρακτηρίζεται από τον πολλαπλασιασμό των πληροφοριών, την εξάρτηση από τις τεχνολογίες που βασίζονται στα δεδομένα και τα θολά φυσικά και ψηφιακά όρια, η κυβερνοασφάλεια έχει ξεπεράσει τη σφαίρα του τεχνικού όρου και έχει καταστεί θεμελιώδης πυλώνας της παγκόσμιας υποδομής της ασφαλείας μας. Ο ίδιος ο όρος "κυβερνοασφάλεια" περικλείει έναν ευρύ και διαρκώς εξελισσόμενο τομέα, που περιλαμβάνει ένα σύνθετο πλέγμα στρατηγικών, τεχνολογιών, πολιτικών και πρακτικών με στόχο την προστασία της ιδιωτικότητας, της ακεραιότητας και της διαθεσιμότητας των ψηφιακών στοιχείων. Η κυβερνοασφάλεια δεν είναι μια στατική έννοια - αντίθετα, είναι ένας εξελισσόμενος κλάδος που πρέπει να προσαρμόζεται συνεχώς στο δυναμικό τοπίο των ψηφιακών απειλών [1].

Σε αυτή την εποχή του ψηφιακού μετασχηματισμού, όπου τα όρια μεταξύ της φυσικής και της ψηφιακής σφαίρας θολώνουν, η κατανόηση του νοήματος και της σημασίας της κυβερνοασφάλειας δεν αποτελεί απλώς μια ακαδημαϊκή επιδίωξη, αλλά μια κοινωνική επιταγή. Η παρούσα διατριβή επιχειρεί να εμβαθύνει στον πολύπλευρο τομέα της κυβερνοασφάλειας, διερευνώντας τον εξελισσόμενο ορισμό της, τη σύνθετη σχέση της με την τεχνολογία και τις βαθιές επιπτώσεις της για άτομα, οργανισμούς και έθνη.

1.1.1 Web servers

Οι διακομιστές ιστού (Web servers) είναι ένα κρίσιμο στοιχείο της υποδομής του Διαδικτύου, υπεύθυνο για την εξυπηρέτηση ιστότοπων και εφαρμογών ιστού σε χρήστες σε όλο τον κόσμο. Ο ρόλος τους στην κυβερνοασφάλεια είναι πολύπλευρος και ουσιαστικός. Ακολουθεί μια επισκόπηση της σχέσης μεταξύ Web servers και κυβερνοασφάλειας [2], [3]:

- **Αποθήκευση και διαχείριση δεδομένων:** Οι Web servers αποθηκεύουν και διαχειρίζονται τα δεδομένα που συνθέτουν τους ιστότοπους και τις εφαρμογές ιστού. Τα δεδομένα αυτά μπορεί να περιλαμβάνουν ευαίσθητες πληροφορίες, όπως λογαριασμούς χρηστών, στοιχεία πληρωμών και προσωπικά δεδομένα. Η διασφάλιση αυτών των δεδομένων από μη εξουσιοδοτημένη πρόσβαση και παραβιάσεις αποτελεί πρωταρχικό μέλημα της κυβερνοασφάλειας.
- **Έλεγχος πρόσβασης:** Η πρόσβαση στους Web servers πρέπει να ελέγχεται αυστηρά, ώστε να αποτρέπεται η πρόσβαση μη εξουσιοδοτημένων χρηστών. Οι μηχανισμοί ελέγχου ταυτότητας, όπως τα ονόματα χρήστη και οι κωδικοί πρόσβασης, καθώς και πιο προηγμένες μέθοδοι, όπως ο έλεγχος ταυτότητας πολλαπλών παραγόντων (multi-factor authentication - MFA), είναι απαραίτητοι για την εξασφάλιση της πρόσβασης στον διακομιστή.
- **Ασφάλεια εφαρμογών ιστού:** Πολλές επιθέσεις στον κυβερνοχώρο έχουν ως στόχο εφαρμογές ιστού που φιλοξενούνται σε Web servers. Οι ευπάθειες των εφαρμογών ιστού, όπως η έγχυση SQL, το cross-site scripting (XSS) και το cross-site request forgery (CSRF), μπορούν να αξιοποιηθούν από επιτιθέμενους σε αυτές προκειμένου να αποκτήσουν μη εξουσιοδοτημένη πρόσβαση ή να αλλοιώσουν τα δεδομένα. Οι διαχειριστές των Web servers πρέπει να διασφαλίζουν ότι οι εφαρμογές ιστού είναι ασφαλείς εφαρμόζοντας επιδιορθώσεις ασφαλείας, χρησιμοποιώντας τείχη προστασίας εφαρμογών ιστού (web application firewalls - WAF) και ακολουθώντας πρακτικές ασφαλούς κωδικοποίησης.
- **Προστασία κατανεμημένης άρνησης υπηρεσίας (Distributed Denial of Service - DDoS):** Στόχος των επιθέσεων DDoS είναι να κατακλύσουν τους Web servers με πολλαπλές κινήσεις, προκαλώντας διακοπή των υπηρεσιών. Μέτρα κυβερνοασφάλειας, όπως συστήματα ανίχνευσης εισβολών (intrusion detection systems - IDS) και υπηρεσίες περιορισμού των DDoS, αναπτύσσονται συχνά για την προστασία των Web servers από αυτές τις επιθέσεις.
- **Κρυπτογράφηση SSL/TLS:** Για την ασφάλεια των δεδομένων κατά τη διαμετακίνηση μεταξύ χρηστών και διακομιστών ιστού, χρησιμοποιούνται πρωτόκολλα κρυπτογράφησης SSL/TLS. Η

σωστή διαμόρφωση των πιστοποιητικών SSL/TLS στους διακομιστές ιστού είναι ζωτικής σημασίας για την αποτροπή επιθέσεων man-in-the-middle και υποκλοπής δεδομένων.

- **Τακτικές ενημερώσεις και διαχείριση διορθώσεων:** Οι Web servers «τρέχουν» σε λογισμικά Apache, Nginx, Microsoft IIS ή άλλα. Αυτά τα λογισμικά χρειάζονται τακτικές ενημερώσεις και διαχείριση επιδιορθώσεων για την αντιμετώπιση των τρωτών σημείων ασφαλείας. Η μη ενημέρωσή τους μπορεί να εκθέσει τον διακομιστή σε γνωστά exploits.
- **Επικεφαλίδες ασφαλείας:** Η εφαρμογή επικεφαλίδων ασφαλείας στην απόκριση του Web server μπορεί να βοηθήσει στην προστασία από κοινές επιθέσεις στο διαδίκτυο. Για παράδειγμα, το HTTP Strict Transport Security (HSTS) εξασφαλίζει ασφαλείς συνδέσεις και η Πολιτική Ασφάλειας Περιεχομένου (Content Security Policy - CSP) βοηθά στην αποτροπή επιθέσεων XSS.
- **Καταγραφή και παρακολούθηση:** Οι αξιόπιστες πρακτικές καταγραφής και παρακολούθησης είναι ζωτικής σημασίας για τον εντοπισμό και την αντιμετώπιση περιστατικών ασφαλείας. Τα εργαλεία παρακολούθησης μπορούν να βοηθήσουν στον εντοπισμό ασυνήθιστων ή κακόβουλων δραστηριοτήτων σε Web servers.
- **Δημιουργία αντιγράφων ασφαλείας και δυνατότητα ανάκτησης:** Η τακτική δημιουργία αντιγράφων ασφαλείας των δεδομένων των Web servers είναι απαραίτητη για τον μετριασμό των επιπτώσεων της απώλειας δεδομένων λόγω κυβερνοεπιθέσεων ή βλαβών του συστήματος. Μια αποτελεσματική στρατηγική δημιουργίας αντιγράφων ασφαλείας και ανάκτησης αποτελεί μέρος μιας ολοκληρωμένης προσέγγισης για την ασφάλεια στον κυβερνοχώρο.

1.1.2 Mobile Networks

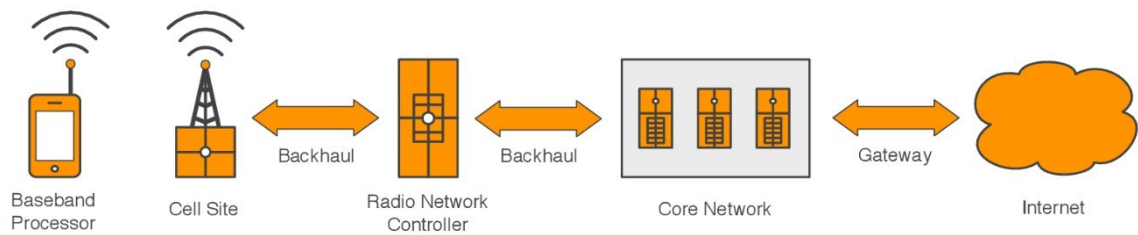
Τα δίκτυα κινητής τηλεφωνίας (Mobile Networks), γνωστά και ως κυψελοειδή δίκτυα ή ασύρματα δίκτυα, αποτελούν ζωτικό μέρος της σύγχρονης τηλεπικοινωνιακής υποδομής. Παρέχουν τη δυνατότητα σε κινητές συσκευές όπως smartphones, tablets και φορητούς υπολογιστές να συνδέονται στο διαδίκτυο και να επικοινωνούν μεταξύ τους ασύρματα. Τα δίκτυα κινητής τηλεφωνίας οργανώνονται σε κυψέλες, οι οποίες είναι γεωγραφικές περιοχές που καλύπτονται από έναν σταθμό βάσης ή πύργο κυψέλης. Κάθε κυψέλη αποτελεί μέρος ενός μεγαλύτερου δικτύου και έχει περιορισμένη εμβέλεια. Όταν μια κινητή συσκευή κινείται εκτός της εμβέλειας μιας κυψέλης, συνδέεται με τον πλησιέστερο πύργο κυψέλης με ισχυρότερο σήμα [4].

Γενιές Mobile Networks:

- ✚ 1G (πρώτη γενιά): Τα δίκτυα 1G ήταν αναλογικά και παρείχαν βασική φωνητική επικοινωνία.
- ✚ 2G (Δεύτερη γενιά): Εισήχθησαν τη δεκαετία του 1990, τα δίκτυα 2G έφεραν ψηφιακή φωνή και περιορισμένες υπηρεσίες δεδομένων.
- ✚ 3G (Τρίτη γενιά): Εισήχθησαν στις αρχές της δεκαετίας του 2000, τα δίκτυα 3G παρείχαν ταχύτερους ρυθμούς μεταφοράς δεδομένων, επιτρέποντας υπηρεσίες όπως η περιήγηση στο διαδίκτυο μέσω κινητού τηλεφώνου και οι βιντεοκλήσεις.
- ✚ 4G (τέταρτη γενιά): Εισήχθησαν γύρω στο 2010, τα δίκτυα 4G αύξησαν σημαντικά τις ταχύτητες δεδομένων, καθιστώντας εφικτή τη ροή βίντεο υψηλής ποιότητας και τα διαδικτυακά παιχνίδια.
- ✚ 5G (πέμπτη γενιά): Τα δίκτυα 5G είναι τα πιο δημοφιλή δίκτυα που έχουν αναπτυχθεί σε όλη την Ευρώπη: Τα δίκτυα 5G που αναπτύσσονται επί του παρόντος παγκοσμίως, υπόσχονται ακόμη υψηλότερες ταχύτητες δεδομένων, χαμηλότερη καθυστέρηση και μεγαλύτερη χωρητικότητα. Αναμένεται να επιτρέψουν νέες εφαρμογές όπως η επαυξημένη πραγματικότητα (AR), η εικονική πραγματικότητα (VR) και το Διαδίκτυο των πραγμάτων (IoT).

Συστατικά στοιχεία των δικτύων κινητής τηλεφωνίας:

- ✚ Σταθμοί βάσης (Cell Towers): Είναι οι φυσικοί πύργοι ή κεραίες που μεταδίδουν και λαμβάνουν σήματα από κινητές συσκευές εντός της περιοχής κάλυψής τους.
- ✚ Κινητές συσκευές: Έξυπνα τηλέφωνα, ταμπλέτες και άλλες φορητές συσκευές που είναι εξοπλισμένες με κυψελοειδείς ασυρμάτους και κάρτες SIM.
- ✚ Δίκτυο πυρήνα: Αυτό το τμήμα του δικτύου διαχειρίζεται τη δρομολόγηση κλήσεων, την κυκλοφορία δεδομένων και τις λειτουργίες διαχείρισης του δικτύου. Περιλαμβάνει στοιχεία όπως τα κέντρα μεταγωγής κινητής τηλεφωνίας (Mobile Switching Centers - MSC) και τους κόμβους εξυπηρέτησης δεδομένων πακέτων (Packet Data Serving Nodes - PDSN).
- ✚ Δίκτυο οπισθοπορείας: Αυτό το δίκτυο συνδέει το δίκτυο πυρήνα με μεμονωμένους πύργους κυψέλης και είναι υπεύθυνο για τη μεταφορά δεδομένων μεταξύ τους.
- ✚ Κάρτες SIM: Οι κάρτες SIM (Subscriber Identity Module) πιστοποιούν και ταυτοποιούν τις κινητές συσκευές στο δίκτυο.
- ✚ Ζώνες συχνотήτων: Τα δίκτυα κινητής τηλεφωνίας λειτουργούν σε συγκεκριμένες ζώνες ραδιοσυχνотήτων που κατανέμονται από τις ρυθμιστικές αρχές.



Εικόνα 1: Συστατικά στοιχεία των δικτύων κινητής τηλεφωνίας¹

Υπηρεσίες δικτύου κινητής τηλεφωνίας [5]:

- ✚ Φωνητικές κλήσεις: Οι παραδοσιακές φωνητικές κλήσεις εξακολουθούν να αποτελούν πρωταρχική υπηρεσία των δικτύων κινητής τηλεφωνίας.
- ✚ Μηνύματα: Τα SMS (Short Message Service) και MMS (Multimedia Messaging Service) επιτρέπουν την αποστολή μηνυμάτων κειμένου και πολυμέσων.
- ✚ Κινητό Διαδίκτυο: Οι χρήστες μπορούν να έχουν πρόσβαση στο διαδίκτυο, να περιηγηθούν σε ιστότοπους, να χρησιμοποιήσουν εφαρμογές και να ελέγξουν μηνύματα ηλεκτρονικού ταχυδρομείου μέσω δικτύων κινητής τηλεφωνίας.
- ✚ Βιντεοκλήσεις: Υπηρεσίες όπως το FaceTime και το Skype επιτρέπουν βιντεοκλήσεις μέσω δικτύων κινητής τηλεφωνίας.
- ✚ Υπηρεσίες εφαρμογών: Οι εφαρμογές κινητής τηλεφωνίας, όπως τα μέσα κοινωνικής δικτύωσης, η πλοήγηση και οι εφαρμογές ροής, βασίζονται στα δίκτυα κινητής τηλεφωνίας για να λειτουργήσουν.

Τα δίκτυα κινητής τηλεφωνίας εξελίσσονται συνεχώς για να καλύψουν την αυξανόμενη ζήτηση για δεδομένα και νέες υπηρεσίες. Αυτό περιλαμβάνει την ανάπτυξη νέων γενεών (π.χ. 5G), την αναβάθμιση των υποδομών και την ανάπτυξη δικτύων IoT βελτιστοποιημένων για επικοινωνία χαμηλής ισχύος και ευρείας περιοχής. Τα δίκτυα κινητής τηλεφωνίας έχουν γίνει αναπόσπαστο μέρος της σύγχρονης ζωής, επιτρέποντας την παγκόσμια επικοινωνία, την πρόσβαση σε πληροφορίες και ένα ευρύ φάσμα υπηρεσιών. Θα συνεχίσουν να εξελίσσονται και να διαμορφώνουν τον τρόπο με τον οποίο συνδεόμαστε και αλληλοεπιδρούμε με τον ψηφιακό κόσμο. Με τα δίκτυα κινητής τηλεφωνίας 5ης γενιάς (5G) να αναπτύσσονται σταδιακά από το 2019, η έρευνα για το δίκτυο κινητής τηλεφωνίας επόμενης γενιάς (6G) έχει ξεκινήσει και στοχεύει στο 2030 [6].

¹ <https://aosabook.org/en/posa/secrets-of-mobile-network-performance.html>

Τα δίκτυα κινητής τηλεφωνίας χρησιμοποιούν διάφορα μέτρα ασφαλείας για την προστασία των δεδομένων και της ιδιωτικής ζωής των χρηστών, συμπεριλαμβανομένης της κρυπτογράφησης των δεδομένων κατά τη μεταφορά και της αυθεντικοποίησης του χρήστη. Ωστόσο, αντιμετωπίζουν επίσης προκλήσεις ασφαλείας, όπως η υποκλοπή, η κλωνοποίηση καρτών SIM και το κακόβουλο λογισμικό που στοχεύει τις κινητές συσκευές. Η εκτεταμένη χρήση κινητών συσκευών, καθώς και η αυξανόμενη δημοτικότητα των κινητών υπηρεσιών έχει δημιουργήσει σοβαρές προκλήσεις στον τομέα της κυβερνοασφάλειας. Τα τελευταία χρόνια, ο αριθμός των κυβερνοεπιθέσεων έχει αυξηθεί δραματικά, καθώς και η πολυπλοκότητά τους. Τα παραδοσιακά συστήματα κυβερνοασφάλειας έχουν αποτύχει να ανιχνεύσουν πολύπλοκες επιθέσεις, άγνωστο κακόβουλο λογισμικό και δεν εγγυώνται τη διατήρηση της ιδιωτικής ζωής των χρηστών. Κατά συνέπεια, τα συστήματα κυβερνοασφάλειας έχουν αγκαλιάσει τα μοντέλα βαθιάς μάθησης (Deep Learning - DL), καθώς παρέχουν αποτελεσματική ανίχνευση νέων επιθέσεων και καλύτερη ακρίβεια [7].

1.1.3 Encryption

Η κρυπτογράφηση είναι μια θεμελιώδης τεχνική ασφαλείας που χρησιμοποιείται για την προστασία ευαίσθητων πληροφοριών με τη μετατροπή τους σε μη αναγνώσιμη μορφή. Η διαδικασία αυτή διασφαλίζει ότι ακόμη και αν μη εξουσιοδοτημένα άτομα ή οντότητες αποκτήσουν πρόσβαση στα κρυπτογραφημένα δεδομένα, δεν μπορούν να τα κατανοήσουν ή να τα χρησιμοποιήσουν χωρίς το κλειδί αποκρυπτογράφησης. Η κρυπτογράφηση διαδραματίζει κρίσιμο ρόλο στη διασφάλιση της ιδιωτικότητας και της ασφάλειας των δεδομένων σε διάφορα πλαίσια, όπως η επικοινωνία, η αποθήκευση δεδομένων και οι διαδικτυακές συναλλαγές. Ακολουθεί μια επισκόπηση της κρυπτογράφησης [8], [9]:

Βασικές αρχές κρυπτογράφησης:

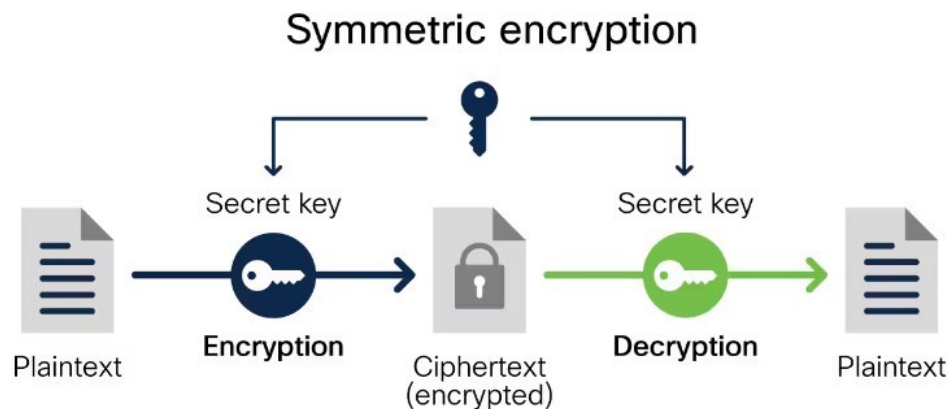
- ✓ Απλό κείμενο και κρυπτογραφημένο κείμενο: Η κρυπτογράφηση ξεκινά με το απλό κείμενο, το οποίο είναι τα αρχικά, αναγνώσιμα δεδομένα που χρήζουν προστασίας. Χρησιμοποιώντας έναν αλγόριθμο κρυπτογράφησης και ένα μυστικό κλειδί, αυτό το απλό κείμενο μετατρέπεται σε κρυπτοκείμενο, το οποίο είναι η κρυπτογραφημένη, μη αναγνώσιμη έκδοση των δεδομένων.
- ✓ Αλγόριθμος κρυπτογράφησης: Ένας αλγόριθμος κρυπτογράφησης είναι μια μαθηματική διαδικασία που καθορίζει τον τρόπο μετατροπής του απλού κειμένου σε κρυπτογράφημα και

αντίστροφα. Οι συνήθεις αλγόριθμοι κρυπτογράφησης περιλαμβάνουν τους AES (Advanced Encryption Standard), RSA (Rivest–Shamir–Adleman) και DES (Data Encryption Standard).

- ✓ Κλειδί κρυπτογράφησης: Το κλειδί κρυπτογράφησης είναι μια μυστική πληροφορία που χρησιμοποιείται από τον αλγόριθμο κρυπτογράφησης για την εκτέλεση των διαδικασιών κρυπτογράφησης και αποκρυπτογράφησης. Η ασφάλεια των κρυπτογραφημένων δεδομένων εξαρτάται σε μεγάλο βαθμό από την ισχύ του κλειδιού κρυπτογράφησης.

Τύποι κρυπτογράφησης:

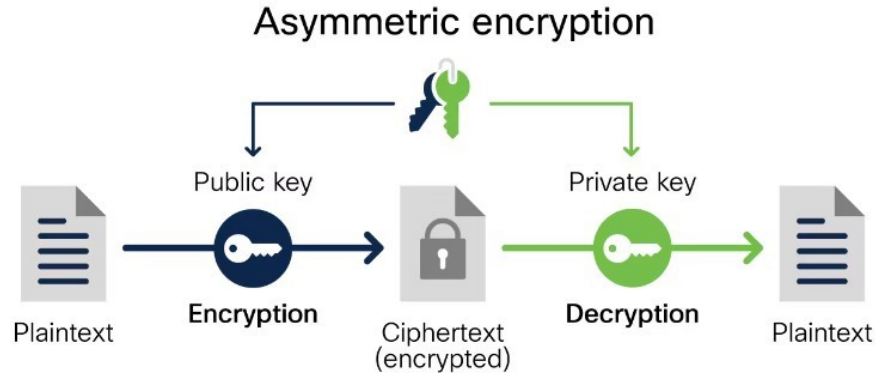
- ✓ Συμμετρική κρυπτογράφηση: Στην συμμετρική κρυπτογράφηση, το ίδιο κλειδί χρησιμοποιείται τόσο για την κρυπτογράφηση όσο και για την αποκρυπτογράφηση. Αυτό σημαίνει ότι τόσο ο αποστολέας όσο και ο παραλήπτης πρέπει να έχουν πρόσβαση στο ίδιο κλειδί. Ο AES είναι ένας ευρέως χρησιμοποιούμενος αλγόριθμος συμμετρικής κρυπτογράφησης.



Εικόνα 2: Συμμετρική κρυπτογράφηση²

- ✓ Ασύμμετρη κρυπτογράφηση: Η ασύμμετρη κρυπτογράφηση χρησιμοποιεί ένα ζεύγος κλειδίων: ένα δημόσιο κλειδί για την κρυπτογράφηση και ένα ιδιωτικό κλειδί για την αποκρυπτογράφηση. Το δημόσιο κλειδί μπορεί να διαμοιραστεί ελεύθερα, ενώ το ιδιωτικό κλειδί παραμένει μυστικό. Ο RSA είναι ένας κοινός αλγόριθμος ασύμμετρης κρυπτογράφησης.

² <https://www.cisco.com/c/en/us/products/security/encryption-explained.html#~encryption-algorithms>



Εικόνα 3³: Ασύμμετρη κρυπτογράφηση

Περιπτώσεις χρήσης της κρυπτογράφησης [10]:

- ✓ Η κρυπτογράφηση χρησιμοποιείται για την προστασία δεδομένων που είναι αποθηκευμένα σε συσκευές, διακομιστές ή στο cloud. Ακόμη και αν κάποιος αποκτήσει φυσική ή ψηφιακή πρόσβαση στο μέσο αποθήκευσης, δεν μπορεί να διαβάσει τα κρυπτογραφημένα δεδομένα χωρίς το κλειδί κρυπτογράφησης.
- ✓ Ασφαλής επικοινωνία: Η κρυπτογράφηση είναι ζωτικής σημασίας για την εξασφάλιση της επικοινωνίας μέσω του διαδικτύου. Το HTTPS (Hypertext Transfer Protocol Secure) κρυπτογραφεί τα δεδομένα που ανταλλάσσονται μεταξύ ενός προγράμματος περιήγησης στον ιστό και ενός διακομιστή ιστού, διασφαλίζοντας ότι ευαίσθητες πληροφορίες όπως τα στοιχεία σύνδεσης και οι αριθμοί πιστωτικών καρτών παραμένουν εμπιστευτικές.
- ✓ Κρυπτογράφηση ηλεκτρονικού ταχυδρομείου: Τα εργαλεία και οι υπηρεσίες κρυπτογράφησης ηλεκτρονικού ταχυδρομείου κρυπτογραφούν το περιεχόμενο των μηνυμάτων ηλεκτρονικού ταχυδρομείου για την αποτροπή μη εξουσιοδοτημένης πρόσβασης. Τα PGP (Pretty Good Privacy) και S/MIME (Secure/Multipurpose Internet Mail Extensions) χρησιμοποιούνται συνήθως για την κρυπτογράφηση ηλεκτρονικού ταχυδρομείου.
- ✓ Κρυπτογράφηση αρχείων και δίσκων: Η κρυπτογράφηση πλήρους δίσκου και η κρυπτογράφηση σε επίπεδο αρχείου προστατεύουν το περιεχόμενο ολόκληρων δίσκων ή

³ <https://www.cisco.com/c/en/us/products/security/encryption-explained.html#~encryption-algorithms>

μεμονωμένων αρχείων και φακέλων. Αυτό χρησιμοποιείται συνήθως για την ασφάλεια φορητών υπολογιστών και κινητών συσκευών.

Ισχύς κρυπτογράφησης και διαχείριση κλειδιών [8], [9]:

- ✓ Μήκος κλειδιού: Το μήκος των κλειδιών κρυπτογράφησης καθορίζει την ισχύ της κρυπτογράφησης. Τα μακρύτερα κλειδιά είναι πιο ασφαλή, αλλά απαιτούν περισσότερη υπολογιστική ισχύ για την επεξεργασία τους.
- ✓ Διαχείριση κλειδιών: Οι κατάλληλες πρακτικές διαχείρισης κλειδιών, όπως η δημιουργία ισχυρών κλειδιών, η ασφαλής αποθήκευση και κοινή χρήση κλειδιών και η τακτική εναλλαγή κλειδιών, είναι απαραίτητες για τη διατήρηση της ασφάλειας των κρυπτογραφημένων δεδομένων.

Περιορισμοί της κρυπτογράφησης[8]:

- ✓ Επιπτώσεις στις επιδόσεις: Οι διαδικασίες κρυπτογράφησης και αποκρυπτογράφησης μπορεί να εισάγουν επιβάρυνση επιδόσεων, ιδίως σε περιβάλλοντα με περιορισμένους πόρους.
- ✓ Προκλήσεις διαχείρισης κλειδιών: Η ασφαλής διαχείριση των κλειδιών κρυπτογράφησης μπορεί να είναι πολύπλοκη και προκλητική, ιδίως σε συστήματα μεγάλης κλίμακας.
- ✓ Backdoors και νομικές εκτιμήσεις: Στην κυβερνοασφάλεια, μια Backdoor είναι ένα μέσο παράκαμψης των υφιστάμενων συστημάτων ασφαλείας ενός οργανισμού. Ενώ μια εταιρεία μπορεί να διαθέτει διάφορες λύσεις ασφαλείας, μπορεί να υπάρχουν μηχανισμοί που επιτρέπουν σε έναν νόμιμο χρήστη ή επιτιθέμενο να τους παρακάμψει. Έχουν υπάρξει συζητήσεις σχετικά με τη χρήση της κρυπτογράφησης σε περιπτώσεις όπου οι υπηρεσίες επιβολής του νόμου απαιτούν πρόσβαση σε κρυπτογραφημένα δεδομένα για ποινικές έρευνες.

1.1.4 APT's και Στοχευμένες επιθέσεις

APTs

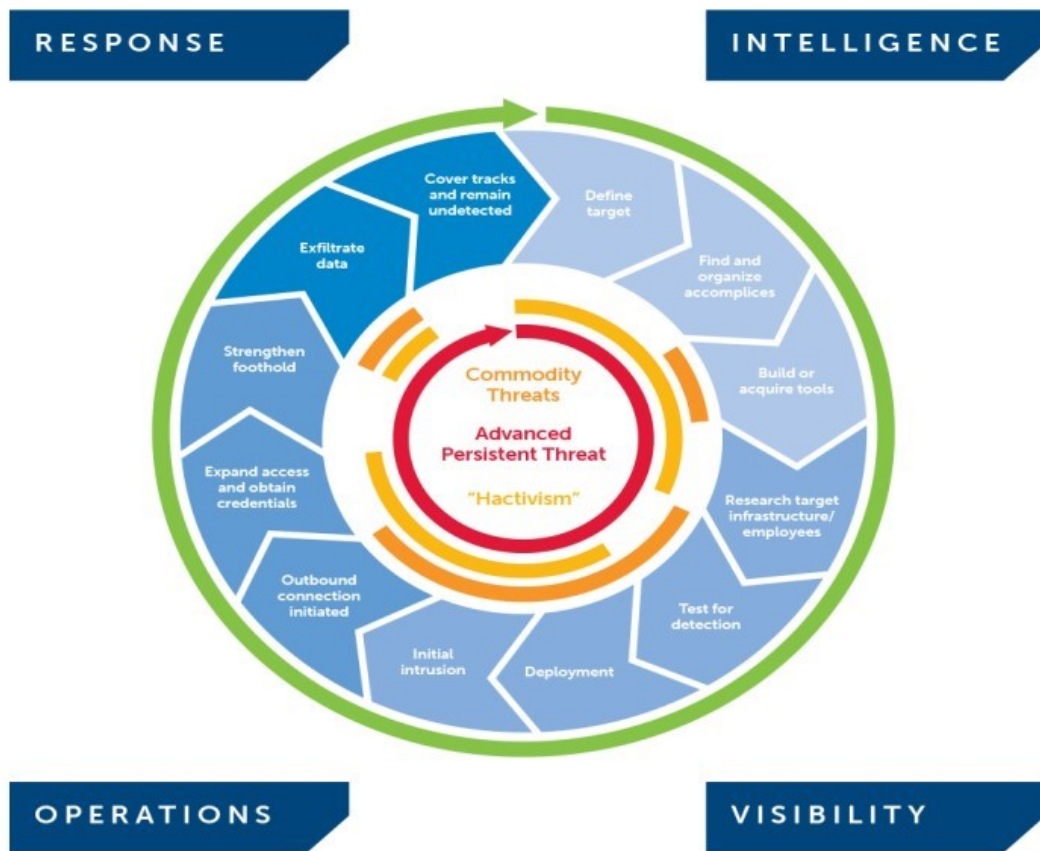
Οι APTs (Advanced Persistent Threats) αναφέρονται σε μια κατηγορία κυβερνοεπιθέσεων που διεξάγονται από καλά χρηματοδοτημένους, ειδικευμένους φορείς με συγκεκριμένο στόχο. Οι επιτιθέμενοι είναι συνήθως έθνη-κράτη, εγκληματικές οργανώσεις ή εξαιρετικά οργανωμένες ομάδες χάκερ. Η "επίμονη" (Persistent) πτυχή των APT σημαίνει ότι οι επιτιθέμενοι διατηρούν

μακροχρόνια παρουσία στο δίκτυο του στόχου, παραμένοντας συχνά απαρατήρητοι για μεγάλα χρονικά διαστήματα.

Οι APTs έχουν σχεδιαστεί για να είναι αόρατες, καθιστώντας δύσκολο τον εντοπισμό τους. Οι επιτιθέμενοι συχνά χρησιμοποιούν εξελιγμένες τεχνικές για να αποφύγουν την ανίχνευση, όπως προσαρμοσμένο κακόβουλο λογισμικό, ευπάθειες μηδενικής ημέρας (zero-day vulnerabilities) και τακτικές κοινωνικής μηχανικής (social engineering tactics). Οι δράστες των APTs έχουν έναν συγκεκριμένο στόχο, όπως η κλοπή ευαίσθητων δεδομένων, η κατασκοπεία ή η δολιοφθορά κρίσιμων υποδομών. Είναι επίμονοι και προσαρμόζουν τις τακτικές τους με την πάροδο του χρόνου για να επιτύχουν τους στόχους τους [11].

Οι APTs συχνά αναπτύσσουν ή χρησιμοποιούν προσαρμοσμένα εργαλεία και κακόβουλο λογισμικό προσαρμοσμένα στο στόχο τους. Αυτό καθιστά δύσκολο για τις παραδοσιακές λύσεις ασφαλείας να ανιχνεύσουν και να αποτρέψουν τις δραστηριότητές τους. Οι επιτιθέμενοι χρησιμοποιούν προηγμένες τεχνικές, συμπεριλαμβανομένης της πλευρικής μετακίνησης σε ένα δίκτυο (lateral movement), της κλιμάκωσης προνομίων (privilege escalation) και της διαρροής δεδομένων, ενώ παράλληλα διατηρούν την επιμονή τους.

Οι APTs έχουν διάφορα κίνητρα, αυτά περιλαμβάνουν την πολιτική, οικονομική και στρατιωτική κατασκοπεία. Οι επιτιθέμενοι μπορεί να στοχεύουν κυβερνητικές υπηρεσίες, αμυντικά συστήματα, κρίσιμες υποδομές ή εταιρείες με πολύτιμη πνευματική ιδιοκτησία. Συμπερασματικά, η άμυνα κατά των APTs απαιτεί μια πολυεπίπεδη προσέγγιση, που περιλαμβάνει τμηματοποίηση του δικτύου, συστήματα ανίχνευσης εισβολών, ασφάλεια τελικών σημείων, ανταλλαγή πληροφοριών που σχετίζονται με απειλές και εκπαίδευση των εργαζομένων. Οι τακτικοί έλεγχοι ασφαλείας και τα σχέδια αντιμετώπισης περιστατικών είναι επίσης απαραίτητα [12], [13].



Εικόνα 4: Ο κύκλος ζωής των APTs [13]

Στοχευμένες επιθέσεις:

Οι στοχευμένες επιθέσεις, όπως υποδηλώνει το όνομα, επικεντρώνονται σε ένα συγκεκριμένο άτομο, οργανισμό ή οντότητα. Αυτές οι επιθέσεις μπορεί να είναι ή να μην είναι τόσο εξελιγμένες όσο οι APTs, αλλά είναι προσαρμοσμένες στο στόχο (θύμα), αξιοποιώντας συχνά προσωπικές πληροφορίες για να αυξήσουν τις πιθανότητες επιτυχίας τους. Συγκεκριμένα, οι στοχευμένες επιθέσεις περιλαμβάνουν την προσαρμογή του κακόβουλου περιεχομένου ή των τεχνικών στο στόχο (θύμα). Οι επιτιθέμενοι μπορούν να χρησιμοποιήσουν πληροφορίες που συλλέγονται από τα μέσα κοινωνικής δικτύωσης ή άλλες πηγές προκειμένου να δημιουργήσουν πειστικά δέλεαρ.

Τα ηλεκτρονικά μηνύματα ηλεκτρονικού ψαρέματος (Spear Phishing) αποτελούν κοινή μέθοδο των στοχευμένων επιθέσεων. Αυτά τα μηνύματα ηλεκτρονικού ταχυδρομείου έχουν σχεδιαστεί για να

εξαπατήσουν τον παραλήπτη ώστε να προβεί σε συγκεκριμένες ενέργειες, όπως να κάνει κλικ σε έναν κακόβουλο σύνδεσμο ή να ανοίξει ένα κακόβουλο συνημμένο αρχείο.

Σε αντίθεση με τις APTs, οι οποίες μπορεί να στοχεύουν μεγάλους οργανισμούς ή κυβερνητικές οντότητες, οι στοχευμένες επιθέσεις συχνά επικεντρώνονται σε μικρότερους και συγκεκριμένους στόχους. Αυτό μπορεί να περιλαμβάνει άτομα, μικρές επιχειρήσεις ή συγκεκριμένα τμήματα σε μεγαλύτερους οργανισμούς. Οι στοχευμένες επιθέσεις μπορεί να έχουν ποικίλα κίνητρα, όπως οικονομικό κέρδος, κλοπή δεδομένων, εταιρική κατασκοπεία ή απλώς πρόκληση αναστάτωσης.

Η άμυνα κατά των στοχευμένων επιθέσεων περιλαμβάνει εκπαίδευση και ευαισθητοποίηση των χρηστών, λύσεις φιλτραρίσματος ηλεκτρονικού ταχυδρομείου και προστασίας από το «ψάρεμα», καθώς και ασφάλεια τελικών σημείων. Η τακτική ενημέρωση των συστημάτων και των εφαρμογών μπορεί επίσης να συμβάλει στον μετριασμό των ευπαθειών που μπορεί να εκμεταλλευτούν οι επιτιθέμενοι.

Συνοπτικά, οι APTs και οι στοχευμένες επιθέσεις αντιπροσωπεύουν δύο διαφορετικές κατηγορίες απειλών στον κυβερνοχώρο. Οι APTs χαρακτηρίζονται συνήθως από υψηλό επίπεδο πολυπλοκότητας, επιμονή και συμμετοχή καλά χρηματοδοτημένων απειλητικών φορέων. Οι στοχευμένες επιθέσεις, αν και επίσης απευθύνονται σε συγκεκριμένα θύματα, μπορεί να μην είναι πάντα τόσο προηγμένες, αλλά είναι ιδιαίτερα εξατομικευμένες για να αυξήσουν τις πιθανότητες επιτυχίας τους. Και οι δύο απαιτούν προληπτικά μέτρα κυβερνοασφάλειας για τον εντοπισμό, τον μετριασμό και την πρόληψη των επιπτώσεών τους σε άτομα και οργανισμούς [13].

1.2 Ανάλυση της κυβερνοασφάλειας στο πραγματικό περιβάλλον

Στον σημερινό διασυνδεδεμένο κόσμο, οι οργανισμοί αντιμετωπίζουν πλήθος απειλών στον κυβερνοχώρο, από κοινές επιθέσεις κακόβουλου λογισμικού και phishing μέχρι APTs που εξαπολύονται από εθνικά κράτη και καλά χρηματοδοτούμενες ομάδες χάκερ. Αυτό το δυναμικό περιβάλλον απαιτεί συνεχή παρακολούθηση, συλλογή πληροφοριών σχετικά με τις απειλές και προσαρμοστικά μέτρα ασφαλείας για την προστασία ευαίσθητων δεδομένων και κρίσιμων υποδομών. Η ανάλυση της ασφάλειας στον κυβερνοχώρο σε σενάρια του πραγματικού κόσμου περιλαμβάνει την κατανόηση των εξελισσόμενων τακτικών των κυβερνοεγκληματιών και των τρωτών σημείων που εκμεταλλεύονται [14].

Μια κρίσιμη πτυχή της ανάλυσης της ασφάλειας στον κυβερνοχώρο σε ένα πραγματικό περιβάλλον είναι η διενέργεια εκτιμήσεων κινδύνου για τον εντοπισμό πιθανών απειλών και τρωτών σημείων. Αυτό περιλαμβάνει την αξιολόγηση του αντίκτυπου μιας επιτυχημένης κυβερνοεπίθεσης και της πιθανότητας να συμβεί. Οι οργανισμοί πρέπει να ιεραρχούν τους κινδύνους με βάση τη δυνητική τους βλάβη και να εφαρμόζουν ανάλογες στρατηγικές μετριασμού. Αυτές οι στρατηγικές μπορεί να περιλαμβάνουν την ενίσχυση των ελέγχων πρόσβασης, την εφαρμογή συστημάτων ανίχνευσης εισβολών, την ενημέρωση του λογισμικού και των συστημάτων και την εκπαίδευση των εργαζομένων σχετικά με τις βέλτιστες πρακτικές ασφάλειας. Η αποτελεσματική αξιολόγηση και ο μετριασμός των κινδύνων είναι ουσιώδεις για την ελαχιστοποίηση των επιπτώσεων των περιστατικών στον κυβερνοχώρο.

Η ανάλυση της ασφάλειας στον κυβερνοχώρο σε πραγματικά περιβάλλοντα συνεπάγεται επίσης την ύπαρξη ενός ισχυρού σχεδίου αντιμετώπισης περιστατικών. Το σχέδιο αυτό περιγράφει τα βήματα που πρέπει να ακολουθηθούν όταν εντοπιστεί παραβίαση της ασφάλειας, συμπεριλαμβανομένου του περιορισμού, της διερεύνησης, της εξάλειψης και της αποκατάστασης. Η έγκαιρη και αποτελεσματική αντιμετώπιση περιστατικών είναι ζωτικής σημασίας για τον περιορισμό των ζημιών που προκαλούνται από μια παραβίαση και την αποκατάσταση των κανονικών λειτουργιών. Η ανάλυση των προηγούμενων περιστατικών και των διδαγμάτων που αντλήθηκαν ενημερώνει για τη βελτίωση των σχεδίων αντιμετώπισης περιστατικών με στόχο τη μελλοντική ετοιμότητα.

Σε πολλούς κλάδους, η συμμόρφωση με τους κανονισμούς και τα πρότυπα κυβερνοασφάλειας είναι υποχρεωτική. Η ανάλυση της ασφάλειας στον κυβερνοχώρο σε πραγματικά περιβάλλοντα περιλαμβάνει τη διασφάλιση της συμμόρφωσης των οργανισμών με αυτές τις απαιτήσεις, οι οποίες μπορεί να περιλαμβάνουν τον Γενικό Κανονισμό Προστασίας Δεδομένων (General Data Protection Regulation GDPR⁴), τον Νόμο περί Φορητότητας και Ευθύνης Ασφαλιστικών Συμβάσεων Υγείας (Health Insurance Portability and Accountability Act - HIPAA⁵) ή ειδικά βιομηχανικά πρότυπα, όπως το Πρότυπο Ασφάλειας Δεδομένων της Βιομηχανίας Καρτών Πληρωμών (Payment Card Industry Data Security Standard - PCI DSS⁶). Οι αξιολογήσεις και οι έλεγχοι συμμόρφωσης αποτελούν μέρος της διαδικασίας ανάλυσης, καθώς βοηθούν τους οργανισμούς να ευθυγραμμίσουν τις πρακτικές

⁴ <https://eur-lex.europa.eu/legal-content/EL/TXT/?uri=celex%3A32016R0679>

⁵ <https://www.cdc.gov/php/publications/topic/hipaa.html>

⁶ <https://www.enisa.europa.eu/topics/risk-management/current-risk/laws-regulation/corporate-governance/pci-dss>

κυβερνοασφάλειας με τις νομικές και κανονιστικές εντολές, μειώνοντας τους νομικούς και οικονομικούς κινδύνους.

Συμπερασματικά, η ανάλυση της ασφάλειας στον κυβερνοχώρο σε ένα πραγματικό περιβάλλον είναι μια πολύπλευρη προσπάθεια που περιλαμβάνει την κατανόηση του δυναμικού τοπίου απειλών, τη διενέργεια αξιολογήσεων κινδύνου, την εφαρμογή στρατηγικών μετριασμού, την προετοιμασία για την αντιμετώπιση περιστατικών και τη διασφάλιση της συμμόρφωσης με τα κανονιστικά πρότυπα. Σε μια εποχή όπου οι απειλές στον κυβερνοχώρο είναι πανταχού παρούσες, οι οργανισμοί πρέπει να προσαρμόζουν διαρκώς τις στρατηγικές τους για την ασφάλεια στον κυβερνοχώρο ώστε να προστατεύουν τα περιουσιακά τους στοιχεία και να διατηρούν την ακεραιότητα, την εμπιστευτικότητα και τη διαθεσιμότητα των κρίσιμων δεδομένων και συστημάτων [14].

1.3 Χρήση ελέγχου ταυτότητας πολλαπλών παραγόντων

Οι παραδοσιακές μέθοδοι ελέγχου ταυτότητας, όπως η χρήση ονόματος χρήστη και κωδικού πρόσβασης, έχουν αποδειχθεί ευάλωτες σε διάφορες μορφές επιθέσεων στον κυβερνοχώρο. Για την ενίσχυση της ασφάλειας, οι οργανισμοί και τα άτομα στρέφονται όλο και περισσότερο στον έλεγχο ταυτότητας πολλαπλών παραγόντων (Multi-Factor Authentication - MFA). Ο MFA είναι μια ισχυρή μέθοδος ελέγχου ταυτότητας που απαιτεί από τους χρήστες να παρέχουν πολλαπλές μορφές επαλήθευσης πριν από τη χορήγηση πρόσβασης σε συστήματα, λογαριασμούς ή δεδομένα [15].

Ο MFA βασίζεται στην αρχή της πολυεπίπεδης ασφάλειας, όπου πολλαπλοί παράγοντες ελέγχου ταυτότητας συνδυάζονται για να επιβεβαιώσουν την ταυτότητα ενός χρήστη. Αυτοί οι παράγοντες εμπίπτουν σε τρεις κύριες κατηγορίες:

- ✓ Κάτι που γνωρίζετε: Αυτό περιλαμβάνει παραδοσιακά διαπιστευτήρια όπως κωδικούς πρόσβασης ή προσωπικούς αριθμούς αναγνώρισης (PIN). Ωστόσο, ο MFA ενισχύει αυτόν τον παράγοντα απαιτώντας σύνθετους και μοναδικούς κωδικούς πρόσβασης.
- ✓ Κάτι που έχετε: Οι χρήστες πρέπει να κατέχουν ένα φυσικό αντικείμενο για να πιστοποιήσουν τον εαυτό τους, όπως μια έξυπνη κάρτα, ένα κουπόνι ασφαλείας ή μια κινητή συσκευή. Αυτά τα αντικείμενα δημιουργούν κωδικούς πρόσβασης μίας χρήσης (OTP) ή κωδικούς, προσθέτοντας ένα επιπλέον επίπεδο ασφάλειας.
- ✓ Κάτι που είστε: Αυτός ο παράγοντας βασίζεται σε βιομετρικά χαρακτηριστικά, όπως δακτυλικά αποτυπώματα, αναγνώριση προσώπου ή σαρώσεις ίριδας. Τα βιομετρικά

δεδομένα είναι δύσκολο να αναπαραχθούν, παρέχοντας υψηλό επίπεδο εμπιστοσύνης στην ταυτότητα του χρήστη.

Η εφαρμογή του MFA προσφέρει πολλά βασικά πλεονεκτήματα [15], [16]:

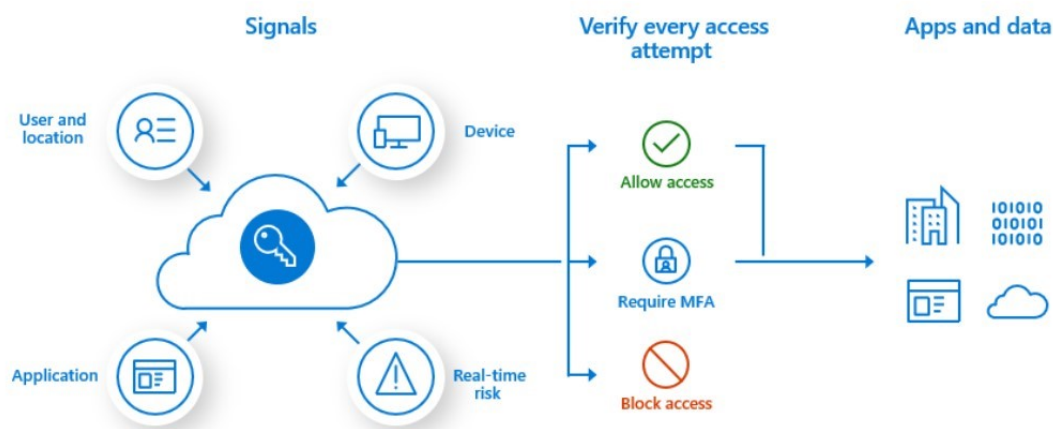
- ✓ *Ενισχυμένη ασφάλεια:* Ο MFA μειώνει σημαντικά τον κίνδυνο μη εξουσιοδοτημένης πρόσβασης. Ακόμη και αν ένας κακόβουλος φορέας αποκτήσει έναν παράγοντα ελέγχου ταυτότητας, θα χρειαστεί και τους υπόλοιπους για να αποκτήσει πλήρη πρόσβαση.
- ✓ *Προστασία κατά του phishing:* Ο MFA συμβάλλει στην αποτροπή των επιθέσεων phishing. Ακόμα και αν ένας χρήστης αποκαλύψει άθελά του τον κωδικό πρόσβασής του, ο επιτιθέμενος θα χρειαστεί και πάλι τον δεύτερο παράγοντα για να αποκτήσει πρόσβαση στον λογαριασμό.
- ✓ *Βελτιωμένη συμμόρφωση:* Πολλά ρυθμιστικά πρότυπα, όπως το GDPR και το HIPAA, απαιτούν τη χρήση του MFA για τη διασφάλιση ευαίσθητων δεδομένων. Η συμμόρφωση με αυτά τα πρότυπα είναι ζωτικής σημασίας για την αποφυγή νομικών κυρώσεων.
- ✓ *Ευκολία για τους χρήστες:* Οι σύγχρονες λύσεις MFA έχουν σχεδιαστεί ώστε να είναι φιλικές προς τον χρήστη. Με επιλογές όπως οι εφαρμογές ελέγχου ταυτότητας που βασίζονται σε smartphone, οι χρήστες μπορούν να επαληθεύουν με ευκολία την ταυτότητά τους.

Η εφαρμογή του MFA περιλαμβάνει τα εξής βήματα [15], [16]:

- ✚ *Επιλογή των σωστών παραγόντων:* Επιλογή των παραγόντων ελέγχου ταυτότητας που ευθυγραμμίζονται με τις απαιτήσεις ασφαλείας του εκάστοτε φορέα/προσώπου. Οι περισσότερες λύσεις MFA επιτρέπουν ευελιξία στον συνδυασμό παραγόντων.
- ✚ *Επιλογή της σωστής λύσης MFA:* Διατίθενται πολυάριθμες λύσεις MFA, που κυμαίνονται από hardware tokens έως εφαρμογές για κινητά και βιομετρικούς σαρωτές. Ένα τυπικό hardware token υλικού είναι μια μικρή συσκευή, συνήθως στο σχήμα μιας πιστωτικής κάρτας ή ενός μπρελόκ. Τα απλούστερα hardware tokens μοιάζουν με μια μονάδα USB και περιέχουν μια μικρή ποσότητα αποθηκευτικού χώρου που περιέχει ένα πιστοποιητικό ή ένα μοναδικό αναγνωριστικό και συχνά ονομάζονται dongles.
- ✚ *Εκπαίδευση χρηστών:* Απαραίτητη είναι η εκπαίδευση των χρηστών σχετικά με τη σημασία του MFA και η κατάλληλη καθοδήγησή τους με σαφείς οδηγίες κατά τη διαδικασία εγκατάστασης.

✚ **Παρακολούθηση και συντήρηση:** Κρίνεται εξαιρετικά σημαντική η τακτική παρακολούθηση των αρχείων καταγραφής του MFA για τυχόν ύποπτη δραστηριότητα. Είναι απαραίτητη η διαρκής ενημέρωση και συντήρηση της λύσης MFA για την αντιμετώπιση των τρωτών σημείων και τη διασφάλιση της συνεχούς αποτελεσματικότητας.

Συμπερασματικά, ο έλεγχος ταυτότητας πολλαπλών παραγόντων αποτελεί κρίσιμο στοιχείο των σύγχρονων στρατηγικών κυβερνοασφάλειας. Απαιτώντας από τους χρήστες να παρέχουν πολλαπλές μορφές επαλήθευσης, ο MFA ανεβάζει σημαντικά τον πήχη για τους επιτιθέμενους που επιχειρούν μη εξουσιοδοτημένη πρόσβαση. Καθώς οι απειλές στον κυβερνοχώρο συνεχίζουν να εξελίσσονται, η υιοθέτηση του MFA καθίσταται όχι μόνο βέλτιστη πρακτική αλλά και αναγκαιότητα για την προστασία ευαίσθητων δεδομένων και ψηφιακών περιουσιακών στοιχείων. Είτε στον εταιρικό κόσμο είτε για προσωπικούς διαδικτυακούς λογαριασμούς, η χρήση του MFA είναι ένα αποτελεσματικό μέσο για την ενίσχυση της ασφάλειας και της άμυνας έναντι των απειλών στον κυβερνοχώρο [15].



Εικόνα 5: Σχηματική αναπαράσταση λειτουργίας MFA⁷

1.4 Απόρρητα Δεδομένα

Σε μια εποχή που χαρακτηρίζεται από ταχεία ψηφιοποίηση, ο όγκος των απόρρητων ή αλλιώς ευαίσθητων δεδομένων που παράγονται, αποθηκεύονται και μεταδίδονται έχει αυξηθεί εκθετικά. Τα δεδομένα αυτά περιλαμβάνουν ένα ευρύ φάσμα πληροφοριών, από προσωπικά στοιχεία ταυτοποίησης και οικονομικά αρχεία έως επιχειρηματικά μυστικά και απόρρητες κυβερνητικές πληροφορίες. Η προστασία και η διαχείριση των απόρρητων δεδομένων έχουν καταστεί πρωταρχικό

⁷ <https://monosign.com/en/blog/what-is-multi-factor-authentication>

μέλημα για τους επαγγελματίες της ασφάλειας, τους οργανισμούς και τις κυβερνήσεις. Το παρόν υποκεφάλαιο εμβαθύνει στη φύση, τη σημασία και τις προκλήσεις που σχετίζονται με τα απόρρητα δεδομένα στο σημερινό ψηφιακό τοπίο. Τα απόρρητα δεδομένα μπορούν να οριστούν ευρέως ως οποιαδήποτε πληροφορία η οποία, εάν αποκαλυφθεί σε μη εξουσιοδοτημένα άτομα ή οντότητες, θα μπορούσε να οδηγήσει σε βλάβη, οικονομική απώλεια, ζημία φήμης ή ακόμα και νομικές συνέπειες [17].

Τα δεδομένα αυτά συνήθως κατηγοριοποιούνται σε διάφορους βασικούς τύπους [18]:

- ✓ *Προσωπικές πληροφορίες / δεδομένα:* Τα προσωπικά δεδομένα περιλαμβάνουν ονόματα, διευθύνσεις, αριθμούς κοινωνικής ασφάλισης, ιατρικά αρχεία και οικονομικά στοιχεία. Διέπονται από νόμους περί προστασίας δεδομένων, όπως ο προαναφερθέν GDPR και HIPAA.
- ✓ *Οικονομικά δεδομένα:* Αυτή η κατηγορία περιλαμβάνει αριθμούς πιστωτικών καρτών, πληροφορίες τραπεζικών λογαριασμών, φορολογικά αρχεία και οικονομικές συναλλαγές. Αποτελούν πρωταρχικό στόχο για τους εγκληματίες του κυβερνοχώρου που επιδιώκουν οικονομικό κέρδος.
- ✓ *Πνευματική ιδιοκτησία:* Οι οργανισμοί θεωρούν την πνευματική τους ιδιοκτησία (Intellectual Property - IP) μεταξύ των πιο πολύτιμων μορφών εμπιστευτικών δεδομένων. Η πνευματική ιδιοκτησία περιλαμβάνει εμπορικά μυστικά, διπλώματα ευρεσιτεχνίας, δεδομένα έρευνας και ανάπτυξης και υλικό που προστατεύεται από πνευματικά δικαιώματα.
- ✓ *Κυβερνητικά και αμυντικά δεδομένα:* Οι κυβερνήσεις κατέχουν ευαίσθητες πληροφορίες που σχετίζονται με την εθνική ασφάλεια, τη διπλωματία και την επιβολή των νόμων. Η μη εξουσιοδοτημένη πρόσβαση σε αυτά τα δεδομένα μπορεί να έχει εκτεταμένες συνέπειες με εθνικές συνέπειες.
- ✓ *Δεδομένα υγειονομικής περίθαλψης:* Τα απόρρητα δεδομένα ασθενών, συμπεριλαμβανομένων των ιατρικών ιστορικών και των σχεδίων θεραπείας, είναι ζωτικής σημασίας για τους παρόχους υγειονομικής περίθαλψης. Η προστασία αυτών των πληροφοριών δεν αποτελεί μόνο νομική απαίτηση, αλλά είναι επίσης απαραίτητη για την εμπιστοσύνη των ασθενών απέναντι στο σύστημα υγειονομικής περίθαλψης.

Τα απόρρητα δεδομένα είναι κρίσιμης σημασίας για διάφορους λόγους [17], [18]:

- ✚ *Προστασία της ιδιωτικής ζωής:* Τα εμπιστευτικά δεδομένα συχνά περιέχουν πληροφορίες που μπορούν να ταυτοποιηθούν προσωπικά (Personally Identifiable Information - PII). Η

διασφάλιση της εμπιστευτικότητάς τους είναι απαραίτητη για την προστασία της ιδιωτικής ζωής των ατόμων και την πρόληψη της κλοπής ταυτότητας.

-  **Οικονομικός αντίκτυπος:** Η απώλεια ή η έκθεση εμπιστευτικών δεδομένων μπορεί να οδηγήσει σε σημαντικές οικονομικές απώλειες για άτομα και οργανισμούς. Οι παραβιάσεις δεδομένων μπορεί να οδηγήσουν σε αγωγές και κανονιστικά πρόστιμα.
-  **Εθνική ασφάλεια:** Στον τομέα της κυβέρνησης και της άμυνας, η εμπιστευτικότητα των διαβαθμισμένων πληροφοριών είναι ζωτικής σημασίας για την εθνική ασφάλεια. Οι παραβιάσεις μπορούν να θέσουν σε κίνδυνο τις διπλωματικές σχέσεις και τις στρατιωτικές επιχειρήσεις.
-  **Ανταγωνιστικό πλεονέκτημα:** Η πνευματική ιδιοκτησία και τα εμπορικά μυστικά παρέχουν στους οργανισμούς ανταγωνιστικό πλεονέκτημα. Η διαφύλαξη αυτών των δεδομένων είναι απαραίτητη για τη διατήρηση της καινοτομίας και της ηγετικής θέσης στην αγορά.

Προκλήσεις στη διαχείριση απόρρητων δεδομένων [17], [18]:

-  **Απειλές κυβερνοασφάλειας:** Η ψηφιακή εποχή έχει επιφέρει πολλαπλασιασμό των απειλών στον κυβερνοχώρο, συμπεριλαμβανομένων των επιθέσεων hacking, κακόβουλου λογισμικού, ransomware και κοινωνικής μηχανικής. Το ransomware αποτελεί μια παραλλαγή κακόβουλου λογισμικού (malware) που χρησιμοποιείται για να διευκολύνει περιπτώσεις εκβιασμών. Μόλις διεισδύσει επιτυχώς σε μια συσκευή, είτε ακινητοποιεί την οθόνη είτε κρυπτογραφεί τα δεδομένα στο δίσκο. Στη συνέχεια, παρουσιάζει ένα αίτημα λύτρων που περιέχει αναλυτικές οδηγίες πληρωμής. Όπως καταλαβαίνουμε οι απειλές αυτές εξελίσσονται συνεχώς, καθιστώντας την προστασία των δεδομένων μια συνεχή πρόκληση.
-  **Κανονισμοί συμμόρφωσης:** Οι οργανισμοί πρέπει να περιηγηθούν σε ένα πολύπλοκο πλέγμα νόμων και κανονισμών για την προστασία των δεδομένων, ο καθένας με τις δικές του απαιτήσεις. Οι αποτυχίες συμμόρφωσης μπορεί να οδηγήσουν σε αυστηρές κυρώσεις.
-  **Διαχείριση του κύκλου ζωής των δεδομένων:** Η κατανόηση του πλήρους κύκλου ζωής των δεδομένων, από τη δημιουργία και την αποθήκευση έως την καταστροφή, είναι ζωτικής σημασίας. Τα δεδομένα πρέπει να διασφαλίζονται κατάλληλα σε κάθε στάδιο.
-  **Ανθρώπινοι παράγοντες:** Οι εργαζόμενοι και οι χρήστες διαδραματίζουν σημαντικό ρόλο στην προστασία των δεδομένων. Οι τυχαίες διαρροές δεδομένων και οι εσωτερικές απειλές αποτελούν κοινές προκλήσεις.

 **Τεχνολογικές εξελίξεις:** Καθώς εξελίσσεται η τεχνολογία, εξελίσσονται και οι δυνατότητες δράσης των εγκληματιών στον κυβερνοχώρο για αυτό το λόγο απαιτείται συνεχής επαγρύπνηση και επενδύσεις σε μέτρα κυβερνοασφάλειας.

Είναι γεγονός λοιπόν ότι τα απόρρητα δεδομένα έχουν τεράστια αξία στην ψηφιακή εποχή, η οποία περιλαμβάνει την προστασία της ιδιωτικής ζωής, την οικονομική σταθερότητα και την εθνική ασφάλεια. Η προστασία αυτών των δεδομένων είναι μια πολύπλευρη πρόκληση που απαιτεί μια ολιστική προσέγγιση, η οποία περιλαμβάνει ισχυρά μέτρα κυβερνοασφάλειας, τήρηση της συμμόρφωσης και εκπαίδευση των χρηστών. Η κατανόηση της σημασίας και της πολυπλοκότητας των εμπιστευτικών δεδομένων είναι ζωτικής σημασίας για τα άτομα, τους οργανισμούς και τους υπεύθυνους χάραξης πολιτικής, καθώς περιηγούνται στο διαρκώς εξελισσόμενο τοπίο της προστασίας των δεδομένων και της ιδιωτικής ζωής [17].

2. Τεχνητή Νοημοσύνη (TN) στην Πληροφορική

2.1 Τεχνολογίες TN

Η τεχνητή νοημοσύνη (TN) είναι μια ακμάζουσα τεχνολογία για πολλές έξυπνες εφαρμογές σε διάφορους τομείς. Μερικά παραδείγματα TN είναι τα αυτόνομα οχήματα στην εμπορική αυτοκινητοβιομηχανία, η ιατρική διάγνωση και η τηλεϊατρική στον υγειονομικό τομέα, τα συστήματα ασφάλειας στον κυβερνοχώρο (π.χ. ανίχνευση κακόβουλου λογισμικού και botnet). Μεταξύ πολλών κλάδων της TN, υπάρχουν δύο σημαντικές προσεγγίσεις: η μηχανική μάθηση (MM – Machine Learning - ML) και η Βαθιά Μάθηση (BM - Deep Learning - DL). Γενικά, η MM αναφέρεται στην ικανότητα εκμάθησης και εξαγωγής σημαντικών προτύπων από τα δεδομένα. Η απόδοση αλγορίθμων και συστημάτων που βασίζονται σε MM εξαρτώνται σε μεγάλο βαθμό από τα αντιπροσωπευτικά χαρακτηριστικά των δεδομένων εισόδου. Η βαθιά μάθηση είναι σε θέση να λύσει πολύπλοκα συστήματα μαθαίνοντας από απλές αναπαραστάσεις. Η BM έχει δύο κύρια χαρακτηριστικά: 1) Η ικανότητα εκμάθησης των σωστών αναπαραστάσεων αποτελεί το ένα χαρακτηριστικό της BM και 2) η BM επιτρέπει στο σύστημα να μαθαίνει τα δεδομένα με πιο πολύπλοκο τρόπο, και έτσι χρησιμοποιούνται πολλαπλά επίπεδα νευρώνων διαδοχικά για να υποστηριχθεί η μάθηση όλο και πιο πολύπλοκων αναπαραστάσεων [19][20].

Η τεχνητή νοημοσύνη (Artificial Intelligence - AI) συνδέεται με την πληροφορική και με υπολογιστικά συστήματα τα οποία μιμούνται εν μέρει την ανθρώπινη συμπεριφορά παρουσιάζοντας ευφυΐα. Η τεχνητή νοημοσύνη χωρίζεται σε δυο μέρη, την συμβολική και την υπο-συμβολική νοημοσύνη. Η πρώτη σχετίζεται με την προσπάθεια εξομίωσης της ανθρώπινης συμπεριφοράς με χρήση αλγορίθμων, με ένα σύνολο συμβόλων και λογικών κανόνων υψηλού επιπέδου, ενώ η δεύτερη αποσκοπεί στην αναπαραγωγή της ανθρώπινης ευφυΐας μέσω αριθμητικών μοντέλων που συνθέτουν συμπεριφορές ευφυΐας.

Οι στόχοι της τεχνητής νοημοσύνης χωρίζονται σε ένα μεγαλύτερο σύνολο από τομείς, όπως στην μηχανική εκμάθηση, την επίλυση προβλημάτων, τα συστήματα γνώσης(κ.α.). Σε πλήθος περιπτώσεων παρουσιάζεται επικάλυψη με άλλα επιστημονικά πεδία όπως η υπολογιστική όραση, η σύνθεση και αναγνώριση φυσικής γλώσσας και ρομποτική. Σήμερα, η τεχνητή νοημοσύνη αποτελεί ένα από τα γρηγορότερα αναπτυσσόμενα πεδία της επιστήμης. Με τον όρο μηχανική μάθηση (machine learning) θεωρούμε ένα εκ των βασικότερων τμημάτων της τεχνητής νοημοσύνης και

αφορά την ανάπτυξη αλγορίθμων που θα δώσουν τη δυνατότητα “εκμάθησης” στους υπολογιστές. Στη μηχανική μάθηση υπάρχουν αλγόριθμοι αναγνώρισης μοτίβων στα δεδομένα και λήψη αποφάσεων που βασίζονται στην στατιστική, τη θεωρία των πιθανοτήτων και την βελτιστοποίηση. Στις περιπτώσεις, στις οποίες η ψηφιακή επεξεργασία εικόνας συνδυάζεται με τεχνικές μηχανικής μάθησης τα αποτελέσματα για τη διάγνωση του καρκίνου του μαστού αποκτούν μεγαλύτερη ακρίβεια, επομένως η χρήση τέτοιων τεχνικών έχει αρχίσει να χρησιμοποιείται στην υποβοήθηση στη λήψη ιατρικών αποφάσεων σχετιζόμενες με αυτή τη νόσο [19][20].

2.1.1 Αλγόριθμοι Μηχανικής Μάθησης - Machine Learning (ML)

Υπάρχουν τρεις βασικές κατηγορίες στις οποίες κατατάσσονται οι αλγόριθμοι ML ανάλογα με την διαδικασία μάθησης:

- Μάθηση με Επίβλεψη (ή Μάθηση με Εκπαιδευτή) – Supervised Learning
- Μάθηση χωρίς Εκπαιδευτή
 - Μάθηση χωρίς Επίβλεψη - Unsupervised Learning
 - Μάθηση με Ενίσχυση – Reinforcement Learning
- Μάθηση με Ημι-επίβλεψη – Semi-supervised Learning

Μάθηση με Επίβλεψη (Supervised Learning)

Όταν υπάρχει εκ των προτέρων (a priori) πληροφορία μέσω ενός συνόλου δεδομένων με γνωστή κλάση, αναφερόμαστε σε μάθηση με επίβλεψη. Απαιτείται να υπάρχει μια επιθυμητή απόκριση, προκειμένου να υλοποιηθεί η κατάλληλη αντιστοίχιση μεταξύ εισόδου και εξόδου, ελαχιστοποιώντας μια συνάρτηση κόστους. Η επιθυμητή απόκριση αντιπροσωπεύει την βέλτιστη ενέργεια που πρέπει να εκτελείται από το νευρωνικό δίκτυο. Μέσω, λοιπόν, της συνάρτησης κόστους εκτελείται μια προσαρμογή των παραμέτρων του συστήματος με επαναληπτικό τρόπο, με τελικό στόχο το νευρωνικό δίκτυο να συμπεριφέρεται με τρόπο που προσομοιάζει τον εκπαιδευτή, δηλαδή να παράγει τη επιθυμητή απόκριση που έχει ορίσει αυτός [21].

Μάθηση χωρίς Επίβλεψη (Unsupervised Learning)

Η περίπτωση της μάθησης χωρίς επίβλεψη, αφορά σύνολα δεδομένων των οποίων η κλάση δεν είναι γνωστή. Το σύστημα με αυτό-οργανούμενο τρόπο πρέπει να μάθει τον συσχετισμό μεταξύ εισόδου και εξόδου. Τα ερεθίσματα υπάρχουν και σε αφθονία και στόχος του αλγόριθμου είναι να

μετατρέπει ένα σήμα εισόδου σε ένα άλλο σήμα που μπορεί να χρησιμοποιηθεί για την επίλυση ενός προβλήματος. Ένας τρόπος είναι η εύρεση των ομοιοτήτων μεταξύ των εισόδων και η ομαδοποίηση τους [21], [22].

Μάθηση με Ενίσχυση (Reinforcement Learning)

Στην μάθηση με ενίσχυση, η εκπαίδευση μιας αντιστοίχισης εισόδου-εξόδου γίνεται με την συνεχή αλληλεπίδραση συστήματος και περιβάλλοντος. Στόχος είναι η μεγιστοποίηση ενός δείκτη απόδοσης. Ένα σύστημα ενισχυτικής μάθησης διαθέτει έναν μηχανισμό που λειτουργεί ως κριτής και παράγει ένα σήμα βασισμένο στα ερεθίσματα που δέχεται από το περιβάλλον του. Μέσω του σήματος αυτού και της απόκρισης από το περιβάλλον, το σύστημα μαθαίνει και λειτουργεί με στόχο να ανακαλύπτει τις βέλτιστες ενέργειες για την συμπεριφορά του.

Μάθηση με Ημι-επίβλεψη (Semi-supervised Learning)

Ο σχεδιασμός ενός αλγόριθμου μηχανικής μάθησης με ημιεπίβλεψη έχει τους ίδιους στόχους με αυτούς της μάθησης με επίβλεψη. Η διαφορά έγκειται στο ότι υπάρχουν σύνολα δεδομένων με άγνωστη την κλάση, δηλαδή *χωρίς ετικέτα* (unlabeled) και σύνολα εκπαίδευσης με γνωστή την κλάση τους, δηλαδή *με ετικέτα* (labeled). Συνήθως, ο αριθμός των δεδομένων χωρίς ετικέτα είναι μεγαλύτερων από αυτών με ετικέτα. Η μάθηση με ημιεπίβλεψη είναι χρήσιμη όταν υπάρχει, λοιπόν, πρόσβαση σε περιορισμένο αριθμό δεδομένων με ετικέτα. Στις περιπτώσεις αυτές, η επεξεργασία των χωρίς ετικέτα δεδομένων και η ανάκτηση πληροφοριών από αυτά μπορεί να οδηγήσει σε βελτιστοποίηση του συστήματος ταξινόμησης. Παράλληλα, αυτό το είδος μάθησης βρίσκει έδαφος στα προβλήματα ομαδοποίησης δεδομένων, υποχρεώνοντας το σύστημα σε καταχώρηση δεδομένων σε μία κλάση με παράλληλη απαγόρευση καταχώρησης άλλων σημείων στην ίδια ομάδα. Η μάθηση με ημιεπίβλεψη, επομένως, παρέχει εκ των προτέρων γνώση την οποία πρέπει να σεβαστεί ο αλγόριθμος ομαδοποίησης [22].

Βασικοί Αλγόριθμοι Μηχανικής Μάθησης

Στην ενότητα αυτή θα αναφερθούν οι σημαντικότεροι αλγόριθμοι μηχανικής μάθησης, και ιδιαίτερα εκείνοι που χρησιμοποιούνται συστηματικά στον τομέα την ιατρικής και στις έρευνες ταξινόμησης για διαγνωστικό σκοπό και θα συναντήσουμε αργότερα στην εργασία αυτή.

Ταξινομητής k-Nearest Neighbors (kNN)

Ο ταξινομητής kNN είναι μια απλή αλλά ισχυρή μέθοδος μηχανικής μάθησης που χρησιμοποιείται στην ταξινόμηση και ομαδοποίηση παρατηρήσεων/δεδομένων χωρίς ετικέτα, με το να τις αναθέτει στην κλάση με την οποία έχουν τα πιο πολλά κοινά χαρακτηριστικά. Οι kNN μέθοδοι δουλεύουν πολύ καλά σε περιπτώσεις μεγάλων συνόλων δεδομένων και μικρών διαστάσεων. Βασίζονται στην ιδέα ότι τα πιο κοντινά μοτίβα σε ένα μοτίβο x' , για το οποίο αναζητούμε την ετικέτα ταξινόμησης, μας δίνουν χρήσιμες πληροφορίες για την ταξινόμηση αυτή. Η kNN, λοιπόν, εφαρμόζει στο μοτίβο x' την ετικέτα της κλάσης με την οποία έχει τους περισσότερους κοντινούς γείτονες στον χώρο των δεδομένων. Κατά κύριο λόγο, η kNN μέθοδος εφαρμόζει την Ευκλείδεια απόσταση ως μέτρο εξαγωγής συμπερασμάτων, η οποία προκύπτει από την παρακάτω εξίσωση για $p = 2$:

$$\|x' - x_j\|^p = \left(\sum_{i=1}^q |(x_i)' - (x_i)_j|^p \right)^{1/p}$$

Πρόκειται για την νόρμα του Minkowski, όπου p είναι ο αριθμός των κλάσεων και x_i είναι οι παρατηρήσεις με q χαρακτηριστικά (διάσταση χώρου).

Επιπλέον, βασικό στοιχείο για την διαμόρφωση του ταξινομητή είναι η παράμετρος k που θα επιλεγεί. Η κατάλληλη επιλογή του k επηρεάζει την διαγνωστική απόδοση του αλγόριθμου διότι εκφράζει την «τοπικότητα» του ταξινομητή. Ένα μεγάλο k μειώνει την πιθανότητα τυχαίου σφάλματος, όμως εμπεριέχει τον κίνδυνο της αγνόησης μικρών αλλά σημαντικών προτύπων. Το κλειδί για την σωστή επιλογή της παραμέτρου αυτής είναι η επίτευξη ισορροπίας μεταξύ υπέρ-εκπαίδευσης και υπό-εκπαίδευσης του ταξινομητή [23], [24].

Ο αλγόριθμος K-Nearest Neighbours είναι μια πολύ απλή και λειτουργική μη παραμετρική μέθοδος που χρησιμοποιείται για την ταξινόμηση και την παλινδρόμηση. Με την εφαρμογή του, παράγονται αποτελέσματα από το εκπαιδευτικό δείγμα, που σχετίζονται με μια συνάρτηση απόστασης και τη συνάρτηση επιλογής της τάξης. Πριν ταξινομηθεί ένα νέο στοιχείο, το συγκρίνει με άλλα στοιχεία χρησιμοποιώντας ένα μέτρο ομοιότητας. Στη συνέχεια λαμβάνονται υπόψη οι πλησιέστεροι γείτονες των στοιχείων και η κλάση που είναι πιο συνηθισμένη μεταξύ των γειτόνων, αντιστοιχεί στο στοιχείο που πρέπει να ταξινομηθεί. Χρησιμοποιώντας την απόσταση, οι γείτονες σταθμίζονται. Το KNN

μπορεί να απαιτήσει πολύ μνήμη αλλά και χώρο για να αποθηκεύσει όλη τη βάση δεδομένων με τις εκτελέσεις των υπολογισμών για την πρόβλεψη [25].

Support Vector Machine (SVM) Αλγόριθμοι

Οι μηχανές διανυσματικής στήριξης (SVM) αναπτύχθηκαν αρχικά με σκοπό την χρήση τους σε προβλήματα δυαδικής ταξινόμησης. Ο βασικός στόχος των μεθόδων αυτών είναι να βρεθεί το βέλτιστο υπερεπίπεδο (hyperplane) $f(w, x) = w \cdot x + b$ που διαχωρίζει δύο κλάσεις σε ένα δεδομένο σύνολο, έχοντας χαρακτηριστικά εισόδου $x \in \mathbb{R}^p$ και ετικέτες $y \in \{-1, +1\}$. Ο SVM λαμβάνει υπόψη τα σημεία που είναι πιο κοντά στο υπερεπίπεδο και τα ονομάζει διανύσματα υποστήριξης (Support Vectors). Η απόσταση μεταξύ του υπερεπιπέδου και του σημείου που βρίσκεται πλησιέστερα σε κάποια από τις κλάσεις ορίζεται ως περιθώριο (margin).

Ο SVM εκπαιδεύεται με την επίλυση του ακόλουθου προβλήματος βελτιστοποίησης με περιορισμό:

$$\begin{aligned} \min & \frac{1}{p} \mathbf{w}^T \mathbf{w} + C \sum_{i=1}^p \xi_i \\ \text{s.t.} & y_i'(\mathbf{w} \cdot \mathbf{x} + b) \geq 1 - \xi_i \\ & \xi_i \geq 0, i = 1, \dots, p \end{aligned}$$

Όπου $\mathbf{w}^T \mathbf{w}$ είναι η νόρμα Manhattan, ξ μία συνάρτηση κόστους και C παράμετρος ποινής. Το αντίστοιχο πρόβλημα βελτιστοποίησης χωρίς περιορισμό είναι το εξής:

$$\min \frac{1}{p} \mathbf{w}^T \mathbf{w} + C \sum_{i=1}^p \max(0, 1 - y_i'(w_i x_i + b))$$

Όπου $\mathbf{w} \mathbf{x} + b$ είναι η συνάρτηση πρόβλεψης. Μία άλλη εκδοχή της παραπάνω εξίσωσης που προσφέρει την ευκολία του ότι είναι διαφοροποιήσιμη είναι η ακόλουθη:

$$\min \frac{1}{p} \|\mathbf{w}\|_2^2 + C \sum_{i=1}^p \max(0, 1 - y_i'(w_i x_i + b))^2$$

Το γεγονός αυτό κάνει την τελευταία να παράγει πιο σταθερά και αξιόπιστα αποτελέσματα. Τελικά, ο βέλτιστος ταξινομητής (υπερεπίπεδο) είναι γνωστός ως μηχανή διανυσματικής υποστήριξης (SVM) [22], [26], [27].

Πιο αναλυτικά οι Μηχανές Διανυσμάτων Υποστήριξης αποτελούν έναν αλγόριθμο ταξινόμησης προτύπων ο οποίος βασίζεται στη στατιστική θεωρία μάθησης. Αποτελούν ένα είδος συγκερασμού γραμμικών μοντέλων και της εύρεσης κοντινότερων γειτόνων. Για την ταξινόμηση, οι μηχανές διανυσμάτων υποστήριξης αναζητούν ένα υπερεπίπεδο (hyperplane) στο χώρο των πιθανών εισόδων. Αυτό το υπερεπίπεδο δημιουργείται ώστε να διαχωρίσει τα θετικά παραδείγματα από τα αρνητικά επιδιώκοντας την μέγιστη απόσταση από το κοντινότερο θετικό και αρνητικό παράδειγμα. Προφανώς υπάρχουν πολλά υπερεπίπεδα που διαχωρίζουν τα θετικά και αρνητικά παραδείγματα όμως αυτό που μας ενδιαφέρει περισσότερο είναι το βέλτιστο υπερεπίπεδο διαχωρισμού το οποίο είναι και μοναδικό [28].

Δέντρα Απόφασης (Decision Trees – DT)

Τα Δέντρα Απόφασης (DT) είναι μία μεγάλη κατηγορία μη-γραμμικών ταξινομητών. Πρόκειται για συστήματα απόφασης πολλών σταδίων, στα οποία οι κλάσεις απορρίπτονται ακολουθιακά έως ότου φτάσουμε σε μία αποδεκτή κλάση ταξινόμησης. Προκειμένου να γίνει αυτό, ο χώρος χαρακτηριστικών διαιρείται σε μοναδικές περιοχές που αντιστοιχίζονται, επίσης με ακολουθιακό τρόπο, με μία μοναδική κλάση. Ένα κατάλληλα κατασκευασμένο δέντρο αποτελείται από ένα μονοπάτι στο οποίο υπάρχουν κόμβοι διακλάδωσης (nodes). Η ταξινόμηση γίνεται τελικά μετά από μία σειρά αποφάσεων, οι οποίες εφαρμόζονται με ερωτήσεις πάνω στα χαρακτηριστικά του διανύσματος προς ταξινόμηση. Κατά την φάση της εκπαίδευσης, επομένως, σε έναν κόμβο πρέπει να έχει αποφασιστεί το σύνολο των χαρακτηριστικών που θα χρησιμοποιηθούν για την εφαρμογή των ερωτήσεων διαχωρισμού. Προκειμένου να υπάρξει διαχωρισμός σε κάθε κόμβο πρέπει επίσης να υιοθετηθεί ένα κριτήριο διαίρεσης. Επιπλέον, απαιτείται ένας κανόνας διακοπής διαιρέσεων, ώστε να μην αναπτύσσεται το δέντρο επ' άπειρον και τέλος, ένας κανόνας που θα αναθέτει κάθε φύλλο σε μία συγκεκριμένη κλάση. Το πλεονέκτημα αυτών των μεθόδων διαφαίνεται στις περιπτώσεις όπου το πλήθος των πιθανών κλάσεων είναι μεγάλο [29].

J48 Decision Tree

Το J48 είναι ένας πιο συγκεκριμένος αλγόριθμος παραγωγής ενός δέντρου απόφασης. Δημιουργείται με αναδρομικό τρόπο διαχωρίζοντας τα σύνολα εκπαίδευσης σε όλο και πιο ξεκάθαρα υποσύνολα. Χρησιμοποιείται ευρέως για την εξόρυξη δεδομένων. Η γενική διαδικασία που ακολουθείται είναι η εξής:

Έστω D_t ένα σύνολο δεδομένων εκπαίδευσης που φτάνουν σε έναν κόμβο t . Τότε:

1. Αν το D_t περιέχει δεδομένα που ανήκουν στην ίδια κλάση y_t . Τότε ο t είναι ένα φύλλο που λαμβάνει την ετικέτα y_t .
2. Αν το D_t είναι ένα κενό σύνολο, τότε ο t ονομάζεται από την προκαθορισμένη κλάση y_d .
3. Αν το D_t περιέχει δεδομένα που ανήκουν σε παραπάνω από μία κλάσεις, τότε χρησιμοποιήσει ένα χαρακτηριστικό ώστε να διαχωρίσεις το σύνολο σε μικρότερα υποσύνολα.

Ο αλγόριθμος εφαρμόζεται αναδρομικά σε κάθε υποσύνολο έως ότου όλα τα δεδομένα του υποσυνόλου να ανήκουν στην ίδια κλάση. Επιπλέον, υπάρχει η υπόθεση ότι κάθε συνδυασμός χαρακτηριστικών διαθέτει διαφορετική ετικέτα κλάσης. Στην περίπτωση που όλα τα δεδομένα του D_t έχουν τις ίδιες τιμές χαρακτηριστικών, τότε ο διαχωρισμός δεν μπορεί να συνεχιστεί και ο κόμβος καθορίζεται ως φύλλο και αποκτά την ίδια κλάση με αυτή που έχει η πλειοψηφία των δεδομένων που σχετίζονται με τον συγκεκριμένο κόμβο [30].

Ο αλγόριθμος Random Forest (RF)

Το Random Forest είναι ένας αλγόριθμος ομαδοποίησης που στοχεύει με επιτυχία στο σημείο κανονικοποίησης, όπου η ποιότητα του μοντέλου είναι όσο το δυνατόν υψηλότερη και τα προβλήματα διακύμανσης και κλίσης έχουν ρίσκο αποτυχίας. Για να ξεπεραστεί το πρόβλημα υπερπροσαρμογής που αντιμετωπίζουν τα δέντρα απόφασης, το Random Forests κατασκευάζει εκατοντάδες ή μπορεί χιλιάδες από αυτά. Προκειμένου τα δέντρα να είναι διαφορετικά μεταξύ τους, το Random Forest χρησιμοποιεί τυχαία δείγματα με αντικατάσταση. Κάθε δέντρο ταξινομεί τις παρατηρήσεις του και η τελική απόφαση ταξινόμησης προκύπτει από τον κανόνα της πλειοψηφίας. Το Random Forest μπορεί επίσης να χρησιμοποιηθεί σε εφαρμογές χωρίς επίβλεψη για την αξιολόγηση γειτνίασης μεταξύ σημείων.

Τα τυχαία δάση αποτελούν μια ειδική κατηγορία των συνδυαστικών μεθόδων ταξινόμησης η οποία χρησιμοποιεί για ταξινομητές δέντρα απόφασης. Εφαρμόζεται η τεχνική bagging για την παροχή διαφορετικών υποομάδων εκπαίδευσης σε δέντρα αποφάσεων με στόχο τη μείωση της διακύμανσης και ταυτόχρονα ο αλγόριθμος εκμάθησης δέντρων τροποποιείται για να επιλέξει ένα τυχαίο υποσύνολο χαρακτηριστικών για κάθε διαχωρισμό [29].

Ο απλοϊκός ταξινομητής κατά Bayes (Naive Bayes Classifier)

Ο απλοϊκός (Naive) Bayes είναι ένας πιθανοτικός μη-γραμμικός ταξινομητής που εφαρμόζει το θεώρημα του Bayes με ισχυρές υποθέσεις ανεξαρτησίας. Αυτό σημαίνει ότι όσο πιο πολλά είναι τα δεδομένα, τόσο περισσότερη εμπειρία συσσωρεύεται και όλο και πιο ακριβής θα είναι η απόφαση που θα ληφθεί. Σε αυτό το μοντέλο, όλες οι ιδιότητες εξετάζονται χωριστά για να ανιχνευθεί οποιαδήποτε υπάρχουσα σχέση μεταξύ τους. Υποθέτει ότι τα χαρακτηριστικά πρόβλεψης είναι υπό όρους ανεξάρτητα δεδομένης μιας κλάσης. Ωστόσο, πρέπει να λάβουμε υπόψιν ότι σε πραγματικές συνθήκες είναι δύσκολο να βρεθούν ανεξάρτητες ιδιότητες, εξ 'ου και το όνομα «απλοϊκός». Επιπλέον, οι τιμές των αριθμητικών χαρακτηριστικών κατανέμονται σε κάθε κλάση. Στον ταξινομητή αυτόν διακρίνονται δύο τύποι πιθανοτήτων, η εκ των υστέρων πιθανότητα δεδομένης μιας πρόβλεψης και η εκ των προτέρων πιθανότητα, η οποία είναι απλούστερη να υπολογιστεί. Το NB είναι γρήγορο και αποδίδει καλά ακόμη και με ένα μικρό σύνολο δεδομένων. Επιπλέον, ο ταξινομητής κατά Bayes είναι βέλτιστος ως προς την ελαχιστοποίηση της πιθανότητας του σφάλματος ταξινόμησης [22], [26], [30], [31].

Βασίζεται στον Κανόνα του Bayes και μπορεί να εκφραστεί ως εξής:

$$\text{Αν } P(\omega_1|x) > P(\omega_2|x), \text{ το } x \text{ ταξινομείται στην } \omega_1$$

$$\text{Αν } P(\omega_1|x) < P(\omega_2|x), \text{ το } x \text{ ταξινομείται στην } \omega_2$$

Ο αλγόριθμος Naïve Bayes είναι ένας από τους κορυφαίους αλγόριθμους στη μηχανική μάθηση για δυαδική ταξινόμηση. Χρησιμοποιεί για την ταξινόμηση κανόνες πιθανότητας και συγκεκριμένα του θεωρήματος του Bayes με τη λανθασμένη υπόθεση ότι η επίδραση της τιμής ενός χαρακτηριστικού σε μια τάξη είναι ανεξάρτητη από τις άλλες τιμές των χαρακτηριστικών. Ο τύπος του θεωρήματος του Bayes είναι: $P(A | B) = P(B | A) P(A) P(B)$ Όπου το $P(A | B)$ είναι μεταγενέστερη πιθανότητα, το $P(B | A)$ είναι πιθανότητα, $P(A)$ είναι πιθανότητα προηγούμενης κατηγορίας και $P(B)$ είναι πιθανότητα προηγούμενης πρόβλεψης. Ένα πλεονέκτημα του ταξινομητή Naive Bayes είναι ότι δεν εξαρτάται

από την ποσότητα των δεδομένων εκπαίδευσης. Μπορεί να χρησιμοποιήσει μια μικρή ποσότητα δεδομένων εκπαίδευσης για να εκτιμήσει τις παραμέτρους που απαιτούνται για την ταξινόμηση. Αποδίδει καλύτερα σε πολύπλοκες πλατφόρμες όπως η ταξινόμηση ανεπιθύμητων μηνυμάτων, η ιατρική διάγνωση και η πρόγνωση καιρού. Κατά τον χειρισμό συνεχών δεδομένων θεωρείται ότι οι συνεχείς τιμές που αντιστοιχούν σε κάθε κατηγορία κατανέμονται σύμφωνα με μια κατανομή Gauss [32]. Στα πλεονεκτήματα του NB συμπεριλαμβάνονται και η απλότητά του (άρα και η ευκολότερη κατανόησή του), καθώς και η ακρίβειά του ακόμα και σε εφαρμογές όπου δεν υφίσταται η ανεξαρτησία των χαρακτηριστικών [33].

Τεχνητά Νευρωνικά Δίκτυα (Artificial Neural Networks – ANN)

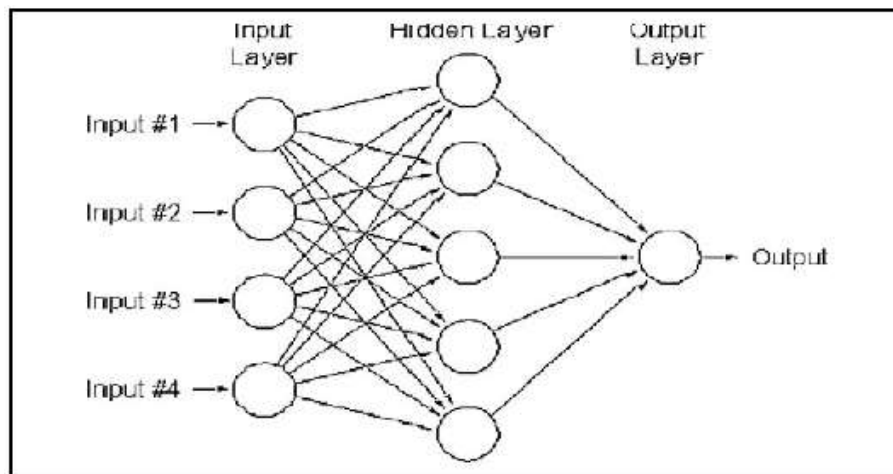
Τα τεχνητά νευρωνικά δίκτυα είναι υπολογιστικές μέθοδοι που εκτελούν πολυπαραγοντική ανάλυση, εμπνευσμένη από τα δίκτυα των βιολογικών νευρώνων. Στη βιολογία τα νευρικά δίκτυα αποτελούνται από ένα μεγάλο βαθμό νευρωνικών κυττάρων που είναι ευρέως διασυνδεδεμένα μεταξύ τους. Μέσα από ένα πολύπλοκο ιστό κλάδων το νευρωνικό κύτταρο μπορεί να αλληλοεπιδράσει με άλλους νευρώνες είτε κοντινούς είτε μακρινούς [34]. Ως νευρωνικά δίκτυα, τα τεχνητά νευρωνικά δίκτυα παρέχουν επίσης επίπεδα απλών σημείων (κόμβων) των δεδομένων, που αλληλοεπιδρούν μέσω προσεχτικά σταθμισμένων γραμμών σύνδεσης έτσι ώστε να μπορούν να επεξεργάζονται ένα αποτέλεσμα. Το ισοζύγιο των βαρών των γραμμών αυτών επιτυγχάνεται από μια διαδικασία εκπαίδευσης των δεδομένων εισόδου. Χαρακτηριστικά υπάρχουν τρεις τύποι νευρώνων-κόμβων:

1. Οι νευρώνες εισόδου. Αυτοί δεν επιτελούν κανένα υπολογισμό.
2. Οι νευρώνες εξόδου οι οποίοι διοχετεύουν στο περιβάλλον τις τελικές αριθμητικές εξόδους του δικτύου.
3. Οι υπολογιστικοί ή κρυφοί νευρώνες. Οι υπολογιστικοί νευρώνες πολλαπλασιάζουν κάθε είσοδο τους με το αντίστοιχο συναπτικό βάρος και υπολογίζουν το ολικό άθροισμα των γινομένων. Το άθροισμα αυτό τροφοδοτείται ως όρισμα στη συνάρτηση ενεργοποίησης, την οποία υλοποιεί εσωτερικά κάθε κόμβος. Η τιμή της συνάρτησης για το εκάστοτε όρισμα αποτελεί και την έξοδο του νευρώνα για τις τρέχουσες εισόδους και βάρη.

Είναι κατάλληλα για τους σκοπούς της δυαδικής ταξινόμησης και της πρόβλεψης επιβίωσης [35].

Ο αλγόριθμος αποτελείται από έναν μεγάλο αριθμό διασυνδεδεμένων κόμβων οργανωμένων σε στρώματα, κάνοντας την αρχιτεκτονική για την επεξεργασία πληροφοριών ικανή να εντοπίσει πολύπλοκα και μη γραμμικά μοτίβα μεταξύ δεδομένων εισόδου και μεταβλητών εξόδου. Τα νευρωνικά δίκτυα πραγματοποιούν τη λειτουργία της ταξινόμησης χρησιμοποιώντας τεχνικές εκπαίδευσης είτε με επίβλεψη είτε χωρίς επίβλεψη. Η αναδρομή των νευρωνικών δικτύων είναι μαύρα κουτιά, στα οποία ο χρήστης τροφοδοτεί δεδομένα και λαμβάνει απαντήσεις. Οι απαντήσεις μπορούν να ρυθμιστούν σωστά, αλλά δεν υπάρχει πρόσβαση στην ακριβή διαδικασία λήψης αποφάσεων, εξ ου και ο τίτλος του μαύρου κουτιού. Μία από τις μεγαλύτερες προκλήσεις είναι ο χρόνος που απαιτείται για την εκπαίδευση δικτύων, το οποίο μπορεί επίσης να απαιτεί σημαντικό υπολογιστικό δυναμικό [36].

Ένα ANN είναι ένα μοντέλο που μιμείται την λειτουργία του νευρικού συστήματος του εγκεφάλου. Αποτελείται από πολυάριθμα στοιχεία επεξεργασίας που συνδέονται σε μεγάλο βαθμό και επικοινωνούν όπως οι συνάψεις του εγκεφάλου. Αυτά τα στοιχεία ονομάζονται τεχνητοί νευρώνες και είναι μονάδες επεξεργασίας πληροφορίας. Το νευρωνικό δίκτυο περιέχει πολλαπλά στρώματα μεταξύ του επιπέδου εισόδου και εξόδου του, τα οποία χαρακτηρίζονται ως κρυφά στρώματα. Σε κάθε στρώμα υπάρχουν κόμβοι που μπορεί να βρίσκονται σε μία από τις εξής δύο καταστάσεις: 0 που σημαίνει ανενεργός και 1 που σημαίνει ενεργός. Επιπλέον, διαθέτει βάρη που καθορίζουν την ισχύ του και την δυνατότητα αποθήκευσης γνώσης (πληροφορίας) και ελέγχουν την κατάσταση του. Με αυτόν τον τρόπο, η δομή που προκύπτει είναι γνωστή ως πολυστρωματικό νευρωνικό δίκτυο, ενώ οι κόμβοι σε κάθε στρώμα συνδέονται μόνο με τους κόμβους του επόμενου επιπέδου. Ουσιαστικά, ένα νευρωνικό δίκτυο είναι ένας παράλληλος επεξεργαστής, ο οποίος διαθέτει πολλές απλές μονάδες επεξεργασίας και έχει την δυνατότητα να αποθηκεύει πληροφορία και εμπειρική γνώση με αποτέλεσμα να καθίσταται δυνατή η εκπαίδευσή του [21], [37].



Εικόνα 6: Αρχιτεκτονική ενός Τεχνητού Νευρωνικού Δικτύου [37]

Στην παραπάνω εικόνα εμφανίζεται ένα πολυστρωματικό νευρωνικό δίκτυο εμπρόσθιας τροφοδοσίας. Διαθέτει κρυμμένα στρώματα, το στρώμα εισόδου και το στρώμα εξόδου. Οι εισοδοί που εισέρχονται στο στρώμα εισόδου θα φτάσουν στο στρώμα εξόδου, αφού υποστούν επεξεργασία στα κρυφά στρώματα. Η εγγενής αυτή δομή των νευρωνικών δικτύων, τα καθιστά πολύ ισχυρά εργαλεία για την επεξεργασία πολύπλοκων δεδομένων, όπως αυτά σε μελέτες καρκίνου του μαστού που χαρακτηρίζονται από έντονα μη γραμμικές αλληλεπιδράσεις μεταξύ δεδομένων εισόδου και πρόβλεψη στόχων (target prediction) [37].

Ο αλγόριθμος Perceptron

Το Perceptron είναι η απλούστερη μορφή νευρωνικού δικτύου, το οποίο χρησιμοποιείται για την ταξινόμηση ενός ειδικού τύπου προτύπων, που είναι γραμμικά διαχωριζόμενα. Προτάθηκε αρχικά από τον Rosenblatt, στο τέλος της δεκαετίας του 50. Αναπτύχθηκε, συγκεκριμένα, προκειμένου να εκπαιδευτεί το perceptron, δηλαδή η βασική μονάδα του χρησιμοποιείται για την μοντελοποίηση των νευρώνων του εγκεφάλου που θεωρήθηκε κρίσιμο για την ανάπτυξη ισχυρών μοντέλων μηχανικής μάθησης. Αν διαθέτουν την κατάλληλη δομή, οι perceptrons κάνουν δυνατή την επίλυση οποιασδήποτε γραμμικής συνάρτησης. Ο πολυεπίπεδος Perceptron (Multilayer Perceptron – MLP), συνίσταται από κρυμμένα επίπεδα, που περιλαμβάνουν έναν αριθμό από perceptrons, κάνοντας δυνατή την επίλυση προβλημάτων με μη γραμμικά διαχωρίσιμες τιμές εξόδου. Κάθε νευρώνας του MLP λαμβάνει το άθροισμα των τιμών εισόδων αφού πολλαπλασιαστούν με τα βάρη τους. Είναι, με

λίγα λόγια ένας απλός γραμμικός ταξινομητής, που σε συνδυασμό με τους υπόλοιπους νευρώνες χτίζουν ένα πολύπλοκο δίκτυο που καταλήγει σε ένα μη γραμμικό ταξινομητή.

Ο αλγόριθμος perceptron μπορεί να αποτυπωθεί με την ακόλουθη σχέση:

$$\mathbf{w}(t+1) = \mathbf{w}(t) - \rho_t \sum_{x \in Y} \delta_x \mathbf{x}$$

και αρχικοποιείται με ένα αυθαίρετο διάνυσμα βαρών $\mathbf{w}(0)$, και το διάνυσμα διόρθωσης $\sum_{x \in Y} \delta_x \mathbf{x}$ σχηματίζεται χρησιμοποιώντας τα λάθος ταξινομημένα διανύσματα. Αυτό επαναλαμβάνεται μέχρι να συγκλίνει ο αλγόριθμος σε μία λύση. Ο ψευδοκώδικας για τον αλγόριθμο perceptron είναι ο εξής:

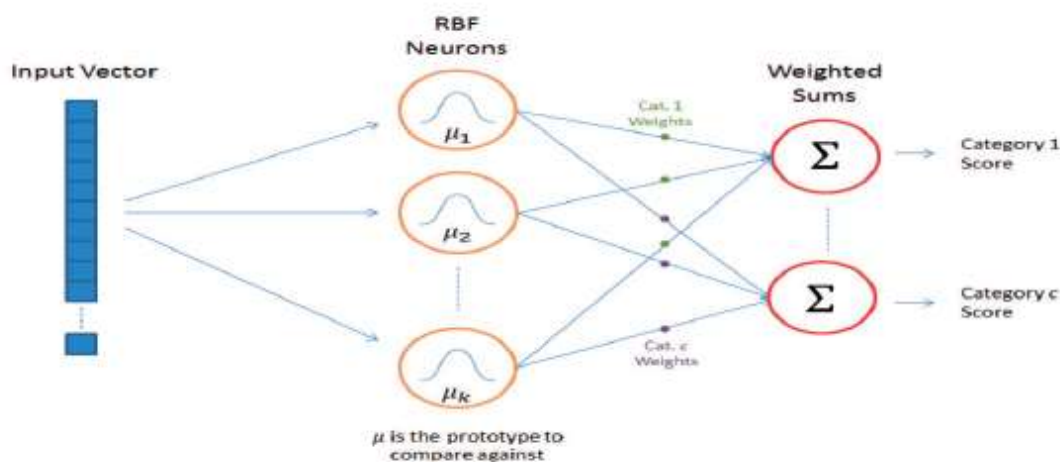
- Επίλεξε τυχαία το $\mathbf{w}(0)$
- Επίλεξε το ρ_0
- $t = 0$
- Επανάλαβε:
 - $Y = \emptyset$
 - Για $i = 1$ έως N
 - Αν $\delta_{x_i} \mathbf{w}(t)^T \mathbf{x}_i \geq 0$ τότε $Y = Y \cup \{\mathbf{x}_i\}$
 - Τέλος {Για}
 - $\mathbf{w}(t+1) = \mathbf{w}(t) - \rho_t \sum_{x \in Y} \delta_x \mathbf{x}$
 - Προσάρμοσε το ρ_t
 - $t = t + 1$
- Μέχρις ότου $Y = \emptyset$

Όπου Y είναι το υποσύνολο των διανυσμάτων εκπαίδευσης που ταξινομήθηκαν εσφαλμένα από το υπερεπίπεδο που ορίζεται από το αυθαίρετο διάνυσμα βαρών $\mathbf{w}$. Η μεταβλητή δ_x ορίζεται ώστε $\delta_x = \begin{cases} -1 & \text{αν } x \in \omega_1 \\ +1 & \text{αν } x \in \omega_2 \end{cases}$, με (ω_1, ω_2) τις κλάσεις ταξινόμησης [22], [27].

Δίκτυο Radial basis function (RBF Network)

Το RBF είναι ένας πιο εξειδικευμένος τύπος νευρωνικού δικτύου. Συχνά με την αναφορά σε ένα ANN, γίνεται αυτόματος συσχετισμός με τον MLP. Η προσέγγιση. Όμως του RBF είναι πιο ενστικτώδης και εκτελεί την ταξινόμηση με το να υπολογίζει την ομοιότητα της εισόδου με τα παραδείγματα του συνόλου εκπαίδευσης. Κάθε νευρώνας του RBF αποθηκεύει ένα «πρωτότυπο», δηλαδή ένα

μοναδικό από τα δεδομένα εκπαίδευσης. Έτσι λοιπόν, όταν επιδιώκεται η ταξινόμηση μιας νέας εισόδου, κάθε νευρώνας υπολογίζει την Ευκλείδεια απόσταση μεταξύ της εισόδου και του «πρωτότυπου». Αν η τιμή της εισόδου είναι πιο κοντά στην κλάση A από την κλάση B, τότε ταξινομείται στην τάξη A.



Εικόνα 7: Αρχιτεκτονική Δικτύου RBF [30]

Ένα RBF δίκτυο διακρίνεται για την ταχύτητα μάθησης του και τον εύκολο σχεδιασμό. Επιπλέον, μπορεί να γενικευθεί εύκολα και να δημιουργήσει ευέλικτους ταξινομητές, ενώ έχει ισχυρή αντίσταση στον θόρυβο εισόδου [30].

Αλγόριθμος Γραμμικής Παλινδρόμησης - Linear Regression

Παρότι πρόκειται για αλγόριθμο προβλήματος παλινδρόμησης, ο Linear Regression θα δούμε στα επόμενα κεφάλαια να χρησιμοποιείται ως ταξινομητής. Ο αλγόριθμος αυτός «εκπαιδεύει» το μοντέλο, το οποίο είναι ένας γραμμικός συνδυασμός των χαρακτηριστικών των παραδειγμάτων εισόδου. Προκειμένου να επιτευχθεί αυτό, εφαρμόστηκε ένα όριο για την έξοδο, ταξινομώντας έτσι τα αποτελέσματα σε δύο κλάσεις [27].

Logistic Regression

Το Logistic Regression θεωρείται μία κλασσική προσέγγιση για προβλήματα δυαδικής ταξινόμησης. Οι logistic regression αλγόριθμοι στην ουσία είναι αλγόριθμοι μάθησης μέσω ταξινόμησης. Ο όρος regression οφείλεται καθαρά στο γεγονός πως βασίζεται σε παρόμοια μαθηματική φόρμουλα με αυτή της linear regression. Στη μέθοδο αυτή υπολογίζεται η σχέση μεταξύ ενός απόλυτα εξαρτώμενου χαρακτηριστικού και ενός ή περισσότερων ανεξάρτητων χαρακτηριστικών, χρησιμοποιώντας τις τιμές πιθανότητας ως τις τιμές πρόβλεψης του εξαρτημένου χαρακτηριστικού. Είναι αρκετά δημοφιλές για τον καθορισμό των παραγόντων κινδύνου στον ιατρικό χώρο [38].

Softmax Regression (SR)

Το μοντέλο αυτό αποτελεί μια γενίκευση του logistic regression σε πολυωνυμικά προβλήματα. Σε αντίθεση με το Linear regression που παράγει ξεκάθαρα αποτελέσματα για τις κλάσεις, το SR παράγει μια πιθανότητα διανομής για αυτές. Η εξίσωση που περιγράφει την μέθοδο αυτή είναι η εξής:

$$P(\hat{y}|x) = \frac{e^{\hat{y}^t}}{\sum_{i=0}^n e^{\hat{y}^t}}$$

Το σφάλμα υπολογίζεται με την χρήση της συνάρτησης εντροπίας:

$$L(y, \hat{y}) = - \sum_{i=0}^n y_i \cdot \log(\hat{y}_i)$$

Όπου y είναι η πραγματική κλάση και $\hat{y}$ η κλάση που προβλέφθηκε [27].

Τεχνικές Boosting Συνδυασμού Ταξινομητών

Με στόχο την βελτίωση της απόδοσης ενός δεδομένου ταξινομητή, αναπτύχθηκε η μέθοδος boosting, που είναι μια ιδιαίτερη τεχνική συνδυασμού ταξινομητών. Αρχικά, υπάρχει ο λεγόμενος ταξινομητής βάσης, ο οποίος είναι ασθενής. Χρησιμοποιώντας επαναληπτικά τον ταξινομητή βάσης και ένα διαφορετικό υποσύνολο του συνόλου εκπαίδευσης ή μια διαφορετική στάθμιση πάνω σε αυτά, σχεδιάζεται μια σειρά ταξινομητών. Σε κάθε επανάληψη δίνεται έμφαση στα λάθος

ταξινομημένα δείγματα. Ο τελικός ταξινομητής προκύπτει από τον σταθμισμένο μέσο όρο των προηγούμενων ιεραρχικά σχεδιασμένων ταξινομητών. Μάλιστα, το σφάλμα ταξινόμησης του τελικού συνδυασμού, για έναν αρκετά ικανό αριθμό επαναλήψεων, γίνεται αυθαίρετα μικρό. Μία από τις βασικές ιδιότητες της μεθόδου boosting είναι η σχετική ανοσία της στο πρόβλημα την υπερβολικής εκπαίδευσης.

Ένας από τους πιο δημοφιλείς αλγόριθμους που εφαρμόζει την μέθοδο boosting είναι ο AdaBoost (adaptive boosting – προσαρμοστική ενίσχυση). Ο AdaBoost χρησιμοποιείται σε προβλήματα ταξινόμησης και παλινδρόμησης. Μετατρέπει ασθενείς «μαθητές» σε ισχυρούς με τον συνδυασμό πολλών ασθενών προκειμένου να δημιουργηθεί ένας μοναδικός ισχυρός κανόνας ταξινόμησης.

Η τεχνική αυτή μπορεί να εφαρμοστεί σε διαφορετικά ήδη ταξινομητών, όπως τα δέντρα απόφασης, παράγοντας πιο ισχυρούς ταξινομητές [22].

Τεχνικές επεξεργασίας δεδομένων και επιλογής χαρακτηριστικών

Ένα ευρέως γνωστό πρόβλημα στην ταξινόμηση και εν γένει στην αναγνώριση προτύπων προκειμένου να επιτευχθεί η εκπαίδευση ενός συστήματος είναι η «κατάρτα της διαστατικότητας», δηλαδή η υπολογιστική πολυπλοκότητα. Ο αριθμός των συλλεγμένων στοιχείων και δεδομένων είναι συνήθως πολύ μεγάλος, με αποτέλεσμα ο ταξινομητής να φέρει και την ευθύνη της μείωσης και επιλογής των επαρκώς ελάχιστου αριθμού χαρακτηριστικών (Feature Selection – FS).

Εξίσου σημαντικός λόγος για την μείωση είναι και το κέρδος που λαμβάνεται από έναν πιθανό συνδυασμό χαρακτηριστικών. Παρότι μεμονωμένα χαρακτηριστικά μπορεί να φέρουν ικανή πληροφορία ταξινόμησης, ο συνδυασμός τους σε περίπτωση αμοιβαίας συσχέτισης δεν αποδίδει κέρδος που να δικαιολογεί την αύξηση της πολυπλοκότητας. Επιπλέον, όσο λιγότερα είναι τα δεδομένα εκπαίδευσης ενός ταξινομητή, τόσο πιο εύκολη είναι η γενίκευση στην χρήση του [22].

Correlation Feature Selection (CFS)

Η επιλογή χαρακτηριστικών βάση συσχέτισης, είναι μια FS τεχνική που χρησιμοποιεί την προσέγγιση του φίλτρου (Filter Approach). Αυτό σημαίνει ότι δεν βασίζεται σε κάποιον αλγόριθμο ML που εφαρμόζεται σε αυτά, αλλά το βάρος των χαρακτηριστικών αξιολογείται μελετώντας τις εγγενείς

ιδιότητες των δεδομένων. Όπως, προαναφέρθηκε, συχνά τα χαρακτηριστικά ενός συνόλου δεδομένων παρατηρείται αν έχουν υψηλή συσχέτιση μεταξύ τους, με αποτέλεσμα να δίνουν περιττή πληροφορία. Η CFS, λοιπόν, βρίσκει τους συσχετισμούς μεταξύ των χαρακτηριστικών. Αυτά που εμφανίζουν μεγάλη συσχέτιση μεταξύ τους αποκλείονται από το σύνολο δεδομένων που χρησιμοποιούνται για την εκπαίδευση του ταξινομητή. Παρομοίως, εκείνα που είναι αλληλένδετα με την κλάση ταξινόμησης διατηρούνται και επιλέγονται [37].

Recursive Feature Elimination (RFE)

Η αναδρομική επιλογή χαρακτηριστικών (RFE) είναι μια τεχνική που χρησιμοποιεί την προσέγγιση Wrapper. Ένας περιορισμός της μεθόδου αυτής είναι η ελλιπής υπάρχουσα βιβλιογραφία σχετικά με τον τρόπο που λειτουργεί, ωστόσο είναι αρκετά δημοφιλής χάρη στην ευκολία στην διαμόρφωση και την χρήση του. Η RFE περιλαμβάνει την δημιουργία ενός ML μοντέλου με όλα τα αρχικά χαρακτηριστικά του συνόλου δεδομένων, τα οποία ταξινομούνται βάση της ποσοτικής σημασίας τους στην μείωση του σφάλματος. Με αυτόν τον τρόπο επιτυγχάνεται η εξάλειψη των πιο αδύναμων χαρακτηριστικών, έως ότου επιτευχθεί ο επιθυμητός αριθμός αυτών. Δεδομένου ότι ο βέλτιστος αριθμός χαρακτηριστικών δεν είναι εκ των προτέρων γνωστός, συχνά χρησιμοποιείται cross-validation με την RFE για τον υπολογισμό της επίδοσης διαφορετικών υποσυνόλων χαρακτηριστικών και την τελική επιλογή των χαρακτηριστικών με την καλύτερη συμπεριφορά. Η RFE αποτελεί, επομένως, μία αποτελεσματική μέθοδο για την επιλογή εκείνων των χαρακτηριστικών σε ένα σύνολο δεδομένων εκπαίδευσης, που είναι περισσότερο σχετικές με την πρόβλεψη μιας της επιθυμητής μεταβλητής [37].

Principal Component Analysis (PCA)

Η ανάλυση κύριων συνιστωσών (PCA) είναι μια γραμμική στατιστική διαδικασία για την εξαγωγή χαρακτηριστικών που μετατρέπει το αρχικό σύνολο δεδομένων σε μειωμένο αριθμό παραγόμενων μεταβλητών που δεν συσχετίζονται και ονομάζονται κύριες συνιστώσες (PC). Η συνολική ποσότητα των PC είναι ίση με την ποσότητα των αρχικών μεταβλητών και παρουσιάζει τις ίδιες πληροφορίες στατιστικής. Κατά την εκτέλεση PCA, η αθροιστική διακύμανση χρησιμοποιείται ως εμπειρικός κανόνας για τη μείωση της διάστασης χαρακτηριστικών του συνόλου δεδομένων. Αυτός ο εμπειρικός κανόνας επιλέγει τον αριθμό των κύριων συνιστωσών σύμφωνα με τα μεγέθη της τιμής Eigen ή την αναλογία διακύμανσης που εξηγεί κάθε μεμονωμένο κύριο συστατικό. Έτσι, λοιπόν, η μέθοδος αυτή

επιτρέπει την μείωση του συνόλου των χαρακτηριστικών διατηρώντας έως και 90% της χρήσιμης πληροφορίας [37].

Linear Discriminant Analysis (LDA)

Η ανάλυση γραμμικής διάκρισης (LDA) είναι μια από τις μεθόδους εξαγωγής στατιστικών χαρακτηριστικών που χρησιμοποιείται, επίσης, για τη μείωση της υψηλής διαστατικότητας των δεδομένων. Χρησιμοποιείται κυρίως για εποπτευόμενη μείωση διαστάσεως. Αυτό σημαίνει ότι λαμβάνει υπόψη τις διαφορετικές ετικέτες κλάσεων. Το LDA είναι μια τεχνική εξαγωγής χαρακτηριστικών που υπολογίζει τον μετασχηματισμό μεγιστοποιώντας τη διασπορά μεταξύ κλάσεων και ελαχιστοποιώντας τη διασπορά εντός κλάσης. Αυτό γίνεται ταυτόχρονα και επιτυγχάνεται η υψηλότερης κατηγορίας διάκριση. Για τον υπολογισμό του βέλτιστου μετασχηματισμού στο LDA, χρησιμοποιείται η αποσύνθεση Eigen στους πίνακες συνδιακύμανσης. Το LDA προσδιορίζει χαρακτηριστικά που ευθύνονται για τη μεγαλύτερη απόκλιση μεταξύ των κλάσεων. Το LDA δύναται να παράγει καλύτερα αποτελέσματα από την PCA, αφού δημιουργούνται σύνολα δεδομένων εκπαίδευσης και επικύρωσης για το LDA [37].

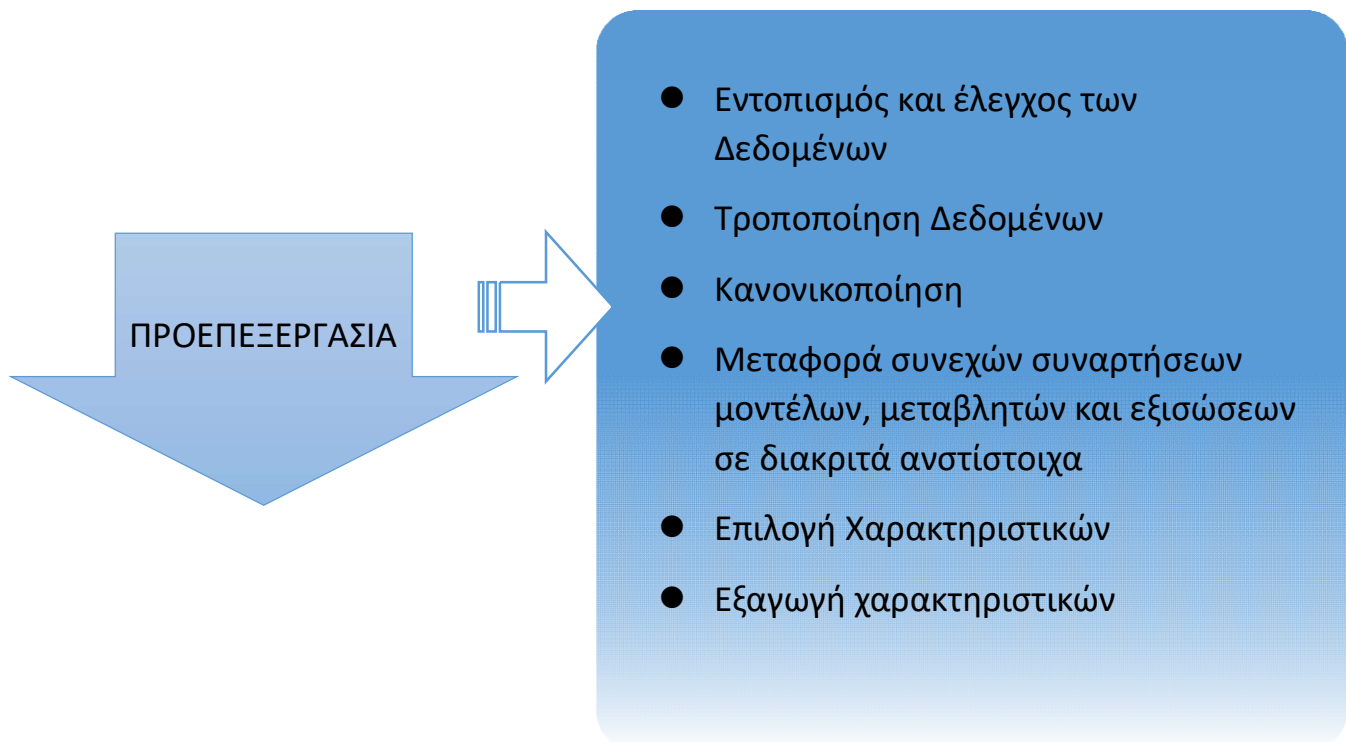
Λογιστική Παλινδρόμηση (LR)

Η λογιστική παλινδρόμηση είναι μια άλλη εποπτευόμενη τεχνική μηχανικής μάθησης που δανείζεται χαρακτηριστικά από μοντέλα στατιστικής και χρησιμοποιείται όταν επιθυμούμε να προβλέψουμε την απουσία ή την παρουσία ενός χαρακτηριστικού, ή ενός συμβάντος. Στο Logistic Regression η μεταβλητή εξόδου ή στόχου είναι μεταβλητή, σε αντίθεση με τη Γραμμική παλινδρόμηση, και ως εκ τούτου είναι ένας δυαδικός αλγόριθμος ταξινόμησης που κατατάσσει ένα σημείο δεδομένων σε μία από τις κατηγορίες των δεδομένων. Η γενική εξίσωση του LogisticRegression είναι: $\log(p(X) / 1 - p(X)) = \beta_0 + \beta_1 X$ Όπου, $p(X)$ είναι η εξαρτημένη μεταβλητή, το X είναι η ανεξάρτητη μεταβλητή, β_0 είναι η αναχαίτιση και β_1 είναι η συντελεστής κλίσης. Οι τιμές εισόδου (X) συνδυάζονται γραμμικά χρησιμοποιώντας τιμές συντελεστή για τον προσδιορισμό μιας εξόδου (Y). Διαφέρει σημαντικά από τη γραμμική παλινδρόμηση επειδή η τιμή εξόδου που διαμορφώνεται είναι δυαδική τιμή (0 ή 1) και όχι αριθμητική τιμή [29]. Η τιμή του σκορ βρίσκεται μεταξύ θετικού και αρνητικού άπειρου, αλλά πρέπει να είναι μεταξύ [0, 1]. Ως εκ τούτου, η λογιστική συνάρτηση, επίσης γνωστή ως η σιγμοειδής συνάρτηση χρησιμοποιείται για τη μετατροπή. Είναι μια καμπύλη που έχει σχήμα S, το οποίο παίρνει οποιοδήποτε πραγματικό αριθμό και το χαρτογραφεί σε τιμή μεταξύ 0 και 1. Χρησιμοποιείται ευρέως για προβλήματα δυαδικής ταξινόμησης και λειτουργεί καλύτερα όταν τα χαρακτηριστικά που

μειώνονται συσχετίζονται μεταξύ τους. Επομένως, αποτελεί ένα γρήγορο μοντέλο και είναι ιδιαίτερα αποτελεσματική σε δυαδικά προβλήματα ταξινόμησης [39].

Τεχνικές Μηχανικής Μάθησης

Η μηχανική μάθηση, ως κλάδος της τεχνητής νοημοσύνης, αντιπροσωπεύει το σύνολο των μεθόδων που έχουν τη δυνατότητα να μάθουν από δεδομένα, δηλαδή, να ανιχνεύουν κρυμμένα μοτίβα σε βάσεις δεδομένων με σκοπό τη χρήση παραγόμενης γνώσης για την πρόβλεψη νέων εξόδων του συστήματος [40]. Διάφορα μαθηματικά μοντέλα εφαρμόζονται για την προσέγγιση πολύπλοκων σχέσεων σε δεδομένα για περιγραφικούς ή προγνωστικούς σκοπούς. Επομένως, η μηχανική μάθηση (Machine Learning) είναι μια πολύ χρήσιμη προσέγγιση για διάφορους τομείς όπου οι συλλογές σχετικών πραγματικών δεδομένων δεν μπορούν να διερευνηθούν με ντετερμινιστικό τρόπο λόγω της παρουσίας αβεβαιότητας και μη γραμμικών εξαρτήσεων μεταξύ των χαρακτηριστικών. Υπάρχει μια τυποποιημένη διαδικασία για την εκτέλεση της ταξινόμησης. Ο κύριος σκοπός είναι να οργανωθούν επαρκώς τα δεδομένα έτσι ώστε οι μέθοδοι Μηχανικής Μάθησης να μπορούν να εφαρμοστούν με τον κατάλληλο τρόπο και να αξιολογηθεί το ανεπτυγμένο μοντέλο που θα μπορούσε να οδηγήσει στην περαιτέρω φάση ανάπτυξης του. Τα πιο σημαντικά βήματα για την ανάπτυξη μοντέλων μηχανικής μάθησης παρουσιάζονται παρακάτω [41]:



ΕΚΜΑΘΗΣΗ



- Κατασκευή Μοντέλου
- Επιλογή κατάλληλου Αλγορίθμου Μηχανικής Μάθησης
- Χειρισμός άνισης κατανομής των πληροφοριών στις βάσεις δεδομένων

ΜΕΤΕΠΕΞΕΡΓΑΣΙΑ



- Αποκτημένες πληροφορίες από τα αποτελέσματα του Μοντέλου
- Σύγκριση Μοντέλων
- Αξιοπιστία Πρόγνωσης
- Εξαγωγή Γνώσης

Αξιολόγηση Αλγορίθμων

Αξιολογώντας τους αλγόριθμους μηχανικής μάθησης με βάση την ακρίβεια, την ευαισθησία και την καμπύλη ROC (Receiver Operating Characteristic) προέκυψε ότι η Trees Random Forest (TRF) έχει τη μεγαλύτερη ακρίβεια και ως εκ τούτου δύναται να υποβοηθήσει στη λήψη ιατρικών αποφάσεων σχετικά με τον καρκίνο του μαστού[42]. Παράλληλα ο υπολογισμός του πιθανού κινδύνου που διατρέχει ο ασθενής εκφράζεται ακριβέστερα με τη μέθοδο Support Vector Machine (SVM), η οποία εμφανίζει τα μικρότερα ποσοστά εσφαλμένων αποτελεσμάτων [43][44][45]. Μια τεχνική μηχανικής μάθησης σχετική με τη μέθοδο Support Vector Machine (SVM) είναι η Least Squares Support Vector Machines (LSSVM), η οποία προσκομίζει υποσχόμενα αποτελέσματα στην ταξινόμηση του καρκίνου του μαστού δίνοντας τη δυνατότητα να εξετάζονται τα δεδομένα από τους γιατρούς ταχύτερα και πιο λεπτομερώς. Συνεπώς, φαίνεται ότι η διάγνωση αυτής της νόσου δύναται να υποβοηθηθεί από τη μέθοδο LSSVM [46]. Επιπλέον, οι αλγόριθμοι μηχανικής μάθησης πιθανόν να είναι το ίδιο ακριβείς σε μια διάγνωση όπως θα ήταν μια ομάδα ιατρών, στην περίπτωση όμως των αλγορίθμων οι χρονικές καθυστερήσεις στη διάγνωση εκμηδενίζονται. Η χρήση τους στην υποβοήθηση των ιατρικών αποφάσεων πρόκειται να κριθεί [47].

Σε ποίο βαθμό όμως μπορούμε να εμπιστευτούμε ένα μοντέλο πρόγνωση: Η αξιολόγηση της ακρίβειας του μοντέλου ή άλλων μέτρων του μοντέλου μας παρέχει τις εκτιμήσεις των μέσων επιδόσεων του μοντέλου με βάση το σύνολο δεδομένων εκπαίδευσης / επικύρωσης. Αλλά πόσο μπορεί να μας πείσει η μέση ακρίβεια ενός μοντέλου για την ακρίβεια των υπολογιστικών προβλέψεων για δεδομένα στα οποία δεν έχουμε οπτική επαφή; Συχνά, οι χρήστες μοντέλων Μηχανικής Μάθησης χρειάζονται κάποια επιπλέον παράμετρο που μπορεί να παρέχει τις πληροφορίες αξιοπιστίας της ακρίβειας πρόβλεψης. Η εκτίμηση της αξιοπιστίας για τις προβλέψεις ταξινόμησης θα μπορούσε να είναι πολύ χρήσιμη σε τομείς με ευαίσθητες αποφάσεις όπως η ιατρική και πιο ειδικά την περίπτωση όπου ο ασθενής είναι υγιής ή πάσχει από καρκίνο του μαστού. Με βάση αυτήν την επιπλέον παράμετρο, οι κλινικοί γιατροί μπορούν να αποφασίσουν σε ποιο βαθμό να εμπιστευτούν μια συγκεκριμένη πρόβλεψη του συστήματος υποστήριξης αποφάσεων. Έτσι, η επέκταση του πλαισίου πρόβλεψης με τις εκτιμήσεις εμπιστοσύνης ενισχύει το μοντέλο πρόβλεψης με κλινική χρηστικότητα.

Τρεις εκτιμήσεις αξιοπιστίας έχουν προταθεί για εκτίμηση της αξιοπιστίας του αποτελέσματος [48]:

1. Η προσέγγιση CNK (εκτίμηση αξιοπιστίας με χρήση τοπικού μοντέλου σφάλματος πρόβλεψης) – που βασίζεται στις ετικέτες των πλησιέστερων γειτόνων. Αν το μοντέλο πρόβλεψης που αναπτύχθηκε προβλέπει το K ως κατανομή πιθανότητας κλάσης για το μη επισημασμένο παράδειγμα X , η εκτίμηση CNK (CNeighbours - K), για παράδειγμα, X , ορίζεται ως η μέση απόσταση μεταξύ των προβλέψεων που βασίζονται σε k -πλησιέστερους γείτονες σύμφωνα με την παρακάτω εξίσωση:

$$CNK = 1 - \frac{\sum_{i=1}^n \|C_i, K\|}{k}$$

Όπου το C_i είναι η πραγματική ετικέτα του πλησιέστερου γείτονα, $i = 1, \dots, k$.

2. Η προσέγγιση LCV (εκτίμηση της αξιοπιστίας μέσω διασταυρούμενης επικύρωσης) βασίζεται στις ετικέτες των πλησιέστερων γειτόνων και στη διασταυρούμενη επικύρωση LOOCV (Leave One Out Cross Validation). Για το μη επισημασμένο παράδειγμα X , στο πρώτο βήμα, ο αλγόριθμος σχηματίζει τον υποχώρο των k -πλησιέστερων γειτόνων. Στη συνέχεια, ακολουθώντας τις αρχές του LOOCV, δημιουργούνται k τοπικά μοντέλα, καθένα από αυτά αποκλείει έναν από τους k -πλησιέστερους γείτονες, και υπολογίζονται τα διαφορετικά K_i , $i = 1, \dots, k$ για καθένα από τους πλησιέστερους γείτονες. Επίσης, για καθέναν από τους πλησιέστερους γείτονες, είναι γνωστές οι αληθινές ετικέτες C_i , $i = 1, \dots, k$, και μπορούν να υπολογιστούν τα σφάλματα πρόβλεψης $E_i = \|(C_i, K_i)\|$. Στη συνέχεια, η εκτίμηση LCV υπολογίζεται χρησιμοποιώντας την παρακάτω εξίσωση:

$$LCV = 1 - \frac{\sum_{i=1}^k E_i}{k}$$

3. Η προσέγγιση DENS (εκτίμηση αξιοπιστίας βάσει πυκνότητας) βασίζεται στην πυκνότητα του προβληματικού χώρου γύρω από το μη επισημασμένο γεγονός για το οποίο θέλουμε να γνωρίζουμε την πρόβλεψη. Η εκτίμηση πυκνότητας $p(X)$ για το μη επισημασμένο γεγονός X υπολογίζεται χρησιμοποιώντας Parzen windows (Τα παράθυρα Parzen είναι μια μη παραμετρική μέθοδος για τον υπολογισμό της συνεχούς πυκνότητας των δεδομένων) και Gaussian kernel (Γκαουσιανο Πυρήνα). Ο συνδυασμός και των δύο εφαρμόζονται στην απόσταση μεταξύ ενός γεγονότος, για το οποίο το μοντέλο μας δεν έχει εκπαιδευτεί ακόμα,

και όλων των γεγονότων από τη βάση δεδομένων.

Έχουμε την παρακάτω εξίσωση:

$$p(X) = \frac{\sum_{i=1}^n K(\|X, X_i\|)}{n}$$

Όπου $\|\cdot, \cdot\|$ δηλώνει μια συνάρτηση Ευκλείδειας απόστασης μεταξύ δύο δειγμάτων, το K υποδηλώνει συνάρτηση Gaussian kernel και $X_i, i = 1, \dots, n$ δηλώνει τα γεγονότα στη βάση δεδομένων. Με αυτόν τον τρόπο, το πρόβλημα του υπολογισμού του πολυδιάστατου πυρήνα Gauss περιορίζεται σε υπολογισμό του δισδιάστατου πυρήνα χρησιμοποιώντας μια συνάρτηση απόστασης.

Η κύρια παραδοχή βασίζεται πως το σφάλμα είναι χαμηλότερο για προβλέψεις που γίνονται σε πυκνότερους υποχώρους και υψηλότερο για προβλέψεις που γίνονται σε αραιότερους υποχώρους.

Επομένως, η εκτίμηση αξιοπιστίας δίνεται από την παρακάτω εξίσωση:

$$DENS = \max_{x_i, i=1, \dots, n} (p(X_i) - p(X))$$

2.1.2 Deep Learning (DL)

Η βαθιά μάθηση (Deep Learning - DL) έχει αναδειχθεί σε αναπόσπαστο κομμάτι δύναμη του τομέα της τεχνητής νοημοσύνης (AI) και της τεχνολογίας γενικότερα. Στη σημερινή ψηφιακή εποχή, δημιουργούνται τεράστιες ποσότητες δεδομένων κάθε δευτερόλεπτο. Η βαθιά μάθηση υπερέχει στην επεξεργασία και την εξαγωγή πολύτιμων πληροφοριών από τεράστια σύνολα δεδομένων, καθιστώντας την απαραίτητη για εργασίες όπως η ανάλυση δεδομένων, η αναγνώριση προτύπων και η προγνωστική μοντελοποίηση [49].

Οι αλγόριθμοι βαθιάς μάθησης είναι ικανοί να χειρίζονται σύνθετα και πολύπλευρα προβλήματα που προηγουμένως θεωρούνταν ανυπέρβλητα για τις παραδοσιακές τεχνικές μηχανικής μάθησης. Η ικανότητα αυτή είναι ιδιαίτερα σημαντική σε τομείς όπως η αναγνώριση εικόνας και ομιλίας, η επεξεργασία φυσικής γλώσσας και τα αυτόνομα συστήματα. Τα μοντέλα βαθιάς μάθησης, αφού εκπαιδευτούν, μπορούν να αυτοματοποιήσουν πολύπλοκες εργασίες, βελτιώνοντας σημαντικά την

αποδοτικότητα σε διάφορους κλάδους. Αυτή η αυτοματοποίηση είναι ιδιαίτερα πολύτιμη σε τομείς όπως είναι η υγειονομική περίθαλψη, ο οικονομικός τομέας και η λογιστική. Η βαθιά μάθηση επιτρέπει την ανάπτυξη ιδιαίτερα εξατομικευμένων προϊόντων και υπηρεσιών. Η παραγωγή συγκεκριμένων συστάσεων από ένα σύστημα ή μια εφαρμογή, για παράδειγμα, αξιοποιούν τη βαθιά μάθηση για την παροχή προσαρμοσμένου περιεχομένου και εμπειριών στους χρήστες, ενισχύοντας την ικανοποίησή τους. Επιπλέον, Η βαθιά μάθηση έχει επιταχύνει την επιστημονική έρευνα σε τομείς όπως η γονιδιωματική, η ανακάλυψη φαρμάκων και η επιστήμη των υλικών. Η ικανότητά της να αναλύει πολύπλοκα δεδομένα έχει οδηγήσει στην παραγωγή νέας πολύτιμης γνώσης σε αυτούς τους τομείς [50], [51].

Η σημασία της βαθιάς μάθησης στο σημερινό τεχνολογικό τοπίο είναι εμφανής από διάφορες απόψεις [52]:

- ✓ Η εξέλιξη της τεχνητής νοημοσύνης: Η βαθιά μάθηση βρίσκεται στην πρώτη γραμμή της προόδου της τεχνητής νοημοσύνης. Έχει τροφοδοτήσει εφαρμογές τεχνητής νοημοσύνης όπως οι φωνητικοί βοηθοί, τα αυτόνομα οχήματα και η γλωσσική μετάφραση, βελτιώνοντας την καθημερινή μας ζωή.
- ✓ Αλλαγές στη βιομηχανία: Η βαθιά μάθηση έχει διαταράξει παραδοσιακούς κλάδους, δημιουργώντας νέα επιχειρηματικά μοντέλα και ευκαιρίες. Για παράδειγμα, οι υπηρεσίες και εφαρμογές όπως η Uber και οι πλατφόρμες παράδοσης φαγητού όπως η DoorDash βασίζονται στη βαθιά μάθηση για τη δρομολόγηση, την τιμολόγηση και την πρόβλεψη της ζήτησης.
- ✓ Επανάσταση στην υγειονομική περίθαλψη: Στην υγειονομική περίθαλψη, η βαθιά μάθηση ενισχύει την έγκαιρη και έγκυρη διάγνωση, την ανακάλυψη νέων αποτελεσματικών φαρμάκων και τα εξατομικευμένα θεραπευτικά σχήματα.
- ✓ Αυτόνομα συστήματα: Η σημασία της βαθιάς μάθησης είναι πιο εμφανής στα αυτόνομα συστήματα. Τα αυτοκινούμενα αυτοκίνητα, τα μη επανδρωμένα αεροσκάφη και η ρομποτική βασίζονται σε μεγάλο βαθμό στη βαθιά μάθηση για την αντίληψη, τη λήψη αποφάσεων και την πλοήγηση. Αυτές οι τεχνολογίες είναι έτοιμες να φέρουν επανάσταση στις μεταφορές, την εφοδιαστική αλυσίδα και τη οικονομία γενικότερα.
- ✓ Κατανόηση φυσικής γλώσσας: Η βαθιά μάθηση έχει βελτιώσει σημαντικά την κατανόηση της φυσικής γλώσσας, επιτρέποντας chatbots, εικονικούς βοηθούς και εργαλεία ανάλυσης

συναισθήματος. Αυτό βελτιώνει την υποστήριξη πελατών, τη δημιουργία περιεχομένου και την επικοινωνία.

- ✓ Επιστημονικές ανακαλύψεις: Η βαθιά μάθηση έχει επιτρέψει επαναστατικές ανακαλύψεις στην έρευνα, συμπεριλαμβανομένης της πρόβλεψης της αναδίπλωσης πρωτεϊνών, της μοντελοποίησης του κλίματος και της ανακάλυψης νέων φαρμάκων. Αυτές οι εξελίξεις έχουν επιπτώσεις στην αντιμετώπιση παγκόσμιων προκλήσεων.

Συνοπτικά, η συνάφεια και η σημασία της βαθιάς μάθησης έγκειται στην ικανότητά της να αντιμετωπίζει πολύπλοκα προβλήματα, να αξιοποιεί μεγάλα δεδομένα και να οδηγεί σε εξελίξεις στην ΤΝ και την τεχνολογία σε διάφορους τομείς. Ο αντίκτυπός της στην αυτοματοποίηση, την εξατομίκευση των τεχνολογιών, την υγειονομική περίθαλψη και την επιστημονική ανακάλυψη υπογραμμίζει τον ρόλο της ως μια τεχνολογία-κλειδί στο σημερινό ταχέως εξελισσόμενο τεχνολογικό τοπίο. Καθώς η βαθιά μάθηση συνεχίζει να εξελίσσεται, αναμένεται να διαμορφώσει καθοριστικά το μέλλον των καινοτομιών και λύσεων που βασίζονται στην ΤΝ.

2.1.3 Natural Language Generation

Η παραγωγή φυσικής γλώσσας (Natural Language Generation - NLG) είναι ένα υποπεδίο της τεχνητής νοημοσύνης και της επεξεργασίας φυσικής γλώσσας (Natural Language Processing - NLP) που επικεντρώνεται στην παραγωγή κειμένου ή προφορικού λόγου που μοιάζει με ανθρώπινο κείμενο από δομημένα δεδομένα. Τα συστήματα NLG χρησιμοποιούν αλγορίθμους και τεχνικές μηχανικής μάθησης για να μετατρέψουν δεδομένα, όπως αριθμητικές πληροφορίες, σε συνεκτικό και συναφές με το περιβάλλον κείμενο φυσικής γλώσσας. Ο πρωταρχικός στόχος της NLG είναι η αυτοματοποίηση της διαδικασίας παραγωγής περιεχομένου αναγνώσιμου από τον άνθρωπο, η οποία έχει ευρείες εφαρμογές σε διάφορους κλάδους, όπως η δημοσιογραφία, το ηλεκτρονικό εμπόριο, η εξυπηρέτηση πελατών και η αναφορά δεδομένων [53].

Η NLG έχει εκτεταμένες εφαρμογές σε διάφορους τομείς. Στον τομέα της δημοσιογραφίας, η NLG χρησιμοποιείται για την αυτόματη παραγωγή ειδησεογραφικών άρθρων και αναφορών με βάση δεδομένα και γεγονότα. Στο ηλεκτρονικό εμπόριο, μπορεί να δημιουργήσει περιγραφές προϊόντων, κριτικές και εξατομικευμένο περιεχόμενο μάρκετινγκ. Η NLG διαδραματίζει επίσης κρίσιμο ρόλο στην επιχειρηματική ευφυΐα, μετατρέποντας πολύπλοκες αναλύσεις δεδομένων και πληροφορίες σε εύκολα κατανοητές αφηγήσεις και εκθέσεις. Επιπλέον, στην υγειονομική περίθαλψη, η NLG χρησιμοποιείται για τη δημιουργία αναφορών ασθενών, ιατρικών σημειωμάτων και συστάσεων

θεραπείας. Τα chatbots εξυπηρέτησης πελατών αξιοποιούν τη NLG για να παρέχουν απαντήσεις που μοιάζουν με ανθρώπινες απαντήσεις και να βοηθούν τους χρήστες στις ερωτήσεις τους [53].

Η σημασία της NLG έγκειται στην ικανότητά της να εξορθολογίζει τις διαδικασίες δημιουργίας περιεχομένου, να βελτιώνει την επικοινωνία δεδομένων και να βελτιώνει τις εμπειρίες των χρηστών. Τα συστήματα NLG μπορούν να παράγουν χρήσιμο περιεχόμενο, μειώνοντας το χρόνο και το κόστος που σχετίζονται με τη φυσική παραγωγή περιεχομένου. Επιπλέον, το σύστημα NLG ενισχύει την προσβασιμότητα με τη μετατροπή πληροφοριών πλούσιων σε δεδομένα σε μορφές που είναι εύκολα κατανοητές από ένα ευρύτερο κοινό. Αυτή η τεχνολογία είναι ιδιαίτερα πολύτιμη στην εποχή των μεγάλων δεδομένων, όπου οι επιχειρήσεις και οι οργανισμοί χρειάζονται αποτελεσματικούς τρόπους για τη μετάδοση πληροφοριών και γνώσεων. Καθώς η NLG συνεχίζει να εξελίσσεται, είναι έτοιμη να φέρει επανάσταση στη δημιουργία περιεχομένου και στην επικοινωνία σε πολυάριθμους κλάδους, καθιστώντας τελικά τις πληροφορίες πιο προσιτές και αξιοποιήσιμες τόσο για τις επιχειρήσεις όσο και για τους ιδιώτες [54].

2.1.4 Speech Recognition

Η αναγνώριση ομιλίας, επίσης γνωστή ως αυτόματη αναγνώριση ομιλίας (Automatic Speech Recognition - ASR) ή μετατροπή ομιλίας σε κείμενο, είναι μια τεχνολογία που επιτρέπει σε ένα υπολογιστικό σύστημα να μετατρέπει τον προφορικό λόγο σε γραπτό κείμενο. Αποτελεί κρίσιμο συστατικό της επεξεργασίας φυσικής γλώσσας (NLP) και της αλληλεπίδρασης ανθρώπου-υπολογιστή. Η τεχνολογία αναγνώρισης ομιλίας έχει εξελιχθεί σημαντικά τα τελευταία χρόνια και έχει βρει εφαρμογές σε διάφορους τομείς [55].

Βασικά στοιχεία της αναγνώρισης ομιλίας [55]:

- ✓ *Ακουστική μοντελοποίηση:* Αυτή η συνιστώσα ασχολείται με την αποτύπωση των ακουστικών χαρακτηριστικών της ομιλίας, όπως τα φωνήματα, τα δίφωνα και τα τρίφωνα. Τα ακουστικά μοντέλα εκπαιδεύονται για να αναγνωρίζουν αυτές τις μονάδες ομιλίας και να τις αντιστοιχούν σε κείμενο.
- ✓ *Γλωσσική μοντελοποίηση:* Τα γλωσσικά μοντέλα βοηθούν το σύστημα να κατανοήσει τη δομή και τη γραμματική μιας γλώσσας. Χρησιμοποιούν στατιστικές τεχνικές για να προβλέψουν την πιθανότητα ακολουθιών λέξεων, βοηθώντας στην επιλογή της πιο πιθανής μεταγραφής των προφορικών λέξεων.

- ✓ *Λεξικό και μοντελοποίηση προφοράς:* Ένα λεξικό περιέχει έναν κατάλογο λέξεων και τις φωνητικές τους αναπαραστάσεις. Τα μοντέλα προφοράς βοηθούν το σύστημα να προσδιορίσει τον τρόπο προφοράς των λέξεων, λαμβάνοντας υπόψη τις παραλλαγές στην ομιλία.

Πώς λειτουργεί η αναγνώριση ομιλίας [55]:

-  *Εισαγωγή ήχου:* Η διαδικασία ξεκινά με μια ηχητική είσοδο, η οποία μπορεί να προέρχεται από διάφορες πηγές, όπως μικρόφωνα, τηλεφωνικές γραμμές ή ηχογραφημένο ήχο.
-  *Εξαγωγή χαρακτηριστικών:* Το ηχητικό σήμα υποβάλλεται σε επεξεργασία για την εξαγωγή σχετικών χαρακτηριστικών, όπως φασματογραφήματα (spectrograms) ή συντελεστές συχνότητας Mel-frequency (Mel-frequency cepstral coefficients - MFCC). Αυτά τα χαρακτηριστικά αντιπροσωπεύουν τα ακουστικά χαρακτηριστικά της ομιλίας.
-  *Ακουστική μοντελοποίηση:* Τα εξαγόμενα χαρακτηριστικά συγκρίνονται με προ-εκπαιδευμένα ακουστικά μοντέλα για να προσδιοριστεί η πιθανότητα διαφόρων φωνημάτων ή μονάδων ομιλίας.
-  *Μοντελοποίηση γλώσσας:* Τα γλωσσικά μοντέλα χρησιμοποιούνται για τον υπολογισμό της πιθανότητας διαφόρων ακολουθιών λέξεων. Το σύστημα επιλέγει την ακολουθία λέξεων με την υψηλότερη πιθανότητα ως αναγνωρισμένη μεταγραφή.
-  *Μεταγενέστερη επεξεργασία:* Μπορούν να εφαρμοστούν τεχνικές μετα-επεξεργασίας για τη βελτίωση της ακρίβειας αναγνώρισης, συμπεριλαμβανομένης της μείωσης του θορύβου, της ανάλυσης περιεχομένου και της προσαρμογής του ομιλητή.

2.1.5 Biometrics

Η βιομετρία (Biometrics) είναι μια τεχνολογία που περιλαμβάνει τη μέτρηση και τη στατιστική ανάλυση των μοναδικών φυσικών ή συμπεριφορικών χαρακτηριστικών ενός ατόμου για σκοπούς ταυτοποίησης και πιστοποίησης. Σε αντίθεση με τις παραδοσιακές μορφές ταυτοποίησης, όπως οι κωδικοί πρόσβασης ή οι κάρτες ταυτότητας, τα βιομετρικά συστήματα χρησιμοποιούν εγγενή, διακριτικά χαρακτηριστικά για την επαλήθευση της ταυτότητας ενός ατόμου [56].

Τύποι βιομετρικών χαρακτηριστικών [57]:

- ✓ *Αναγνώριση δακτυλικών αποτυπωμάτων:* Πρόκειται για έναν από τους πιο ευρέως χρησιμοποιούμενους βιομετρικούς τρόπους, όπου τα μοτίβα των κορυφογραμμών και κοιλάδων στα ακροδάχτυλα ενός ατόμου σαρώνονται και αντιστοιχίζονται για την ταυτοποίηση.
- ✓ *Αναγνώριση προσώπου:* Τα βιομετρικά χαρακτηριστικά προσώπου χρησιμοποιούν χαρακτηριστικά του προσώπου, όπως η απόσταση μεταξύ των ματιών, το σχήμα της μύτης και η γραμμή του σαγονιού, για την αναγνώριση ατόμων.
- ✓ *Αναγνώριση ίριδας και αμφιβληστροειδούς:* Αυτές οι τεχνικές αναλύουν τα μοναδικά μοτίβα στο έγχρωμο τμήμα του ματιού (ίριδα) ή στο λεπτό στρώμα των αιμοφόρων αγγείων στο πίσω μέρος του ματιού (αμφιβληστροειδής).
- ✓ *Αναγνώριση φωνής:* Τα βιομετρικά στοιχεία φωνής αναλύουν τα φωνητικά χαρακτηριστικά, συμπεριλαμβανομένου του ύψους, του τόνου και των μοτίβων ομιλίας, για να επαληθεύσουν την ταυτότητα ενός ατόμου.
- ✓ *Αναγνώριση αποτυπωμάτων παλάμης:* Παρόμοια με τα δακτυλικά αποτυπώματα, τα αποτυπώματα παλάμης καταγράφουν τα μοναδικά μοτίβα στην επιφάνεια της παλάμης για την ταυτοποίηση.
- ✓ *Αναγνώριση γεωμετρίας χεριού:* Αυτή η μέθοδος μετρά το μέγεθος και το σχήμα του χεριού ενός ατόμου, συμπεριλαμβανομένου του μήκους και του πλάτους των δακτύλων.
- ✓ *Βιομετρικά χαρακτηριστικά συμπεριφοράς:* Τα βιομετρικά χαρακτηριστικά συμπεριφοράς περιλαμβάνουν την ανάλυση μοναδικών μοτίβων συμπεριφοράς, όπως η δυναμική πληκτρολόγησης (ρυθμός πληκτρολόγησης), η δυναμική υπογραφής και η ανάλυση βάδισης (στυλ βαδίσματος).

Πώς λειτουργεί η βιομετρία [57]:

-  **Εγγραφή:** Κατά τη διαδικασία εγγραφής, τα βιομετρικά δεδομένα ενός ατόμου καταγράφονται και αποθηκεύονται σε μια βάση δεδομένων. Αυτή η αρχική καταγραφή δημιουργεί ένα πρότυπο αναφοράς που θα χρησιμοποιείται για μελλοντικές συγκρίσεις.
-  **Επαλήθευση:** Για την επαλήθευση της ταυτότητας, το σύστημα καταγράφει εκ νέου τα βιομετρικά δεδομένα και τα συγκρίνει με το αποθηκευμένο πρότυπο. Εάν υπάρχει ταύτιση εντός ενός αποδεκτού ορίου, η ταυτότητα επιβεβαιώνεται.
-  **Ταυτοποίηση:** Στη λειτουργία ταυτοποίησης, το σύστημα αναζητά μια βάση δεδομένων με αποθηκευμένα πρότυπα για να βρει μια αντιστοιχία για τα παρουσιαζόμενα βιομετρικά

δεδομένα. Αυτό χρησιμοποιείται συχνά σε σενάρια όπου η ταυτότητα του ατόμου είναι άγνωστη.

Εφαρμογές της βιομετρίας [57]:

- ✚ Έλεγχος πρόσβασης: Βιομετρικά συστήματα χρησιμοποιούνται συνήθως για τον έλεγχο της φυσικής πρόσβασης σε ασφαλείς χώρους, όπως κτίρια γραφείων, αεροδρόμια και κυβερνητικές εγκαταστάσεις.
- ✚ Έξυπνα τηλέφωνα και συσκευές: Πολλά smartphones και άλλες συσκευές διαθέτουν αναγνώριση δακτυλικών αποτυπωμάτων και προσώπου για το ξεκλείδωμα και την ασφάλεια της συσκευής.
- ✚ Ασφάλεια συνόρων: Τα βιομετρικά στοιχεία χρησιμοποιούνται στις συνοριακές διαβάσεις για την επαλήθευση της ταυτότητας των ταξιδιωτών και την ενίσχυση της ασφάλειας.
- ✚ Χρηματοοικονομικές υπηρεσίες: Ο βιομετρικός έλεγχος ταυτότητας χρησιμοποιείται όλο και περισσότερο στον χρηματοπιστωτικό κλάδο για ασφαλή πρόσβαση σε λογαριασμούς και συναλλαγές.
- ✚ Υγειονομική περίθαλψη: Η βιομετρία μπορεί να χρησιμοποιηθεί για την ταυτοποίηση ασθενών, την επαλήθευση συνταγών και την εξασφάλιση ηλεκτρονικών αρχείων υγείας.
- ✚ Επιβολή του νόμου: Οι αστυνομικές υπηρεσίες χρησιμοποιούν τα βιομετρικά στοιχεία για την ταυτοποίηση εγκληματιών και τις εγκληματολογικές έρευνες.
- ✚ Εκλογές: Ορισμένες χώρες χρησιμοποιούν βιομετρικά στοιχεία για την πρόληψη της απάτης, επαληθεύοντας την ταυτότητα των ψηφοφόρων.

Προκλήσεις και ανησυχίες [58]:

- ✚ Απόρρητο και ασφάλεια: Τα βιομετρικά δεδομένα είναι ιδιαίτερα ευαίσθητα και υπάρχουν ανησυχίες σχετικά με την αποθήκευση, τη διαβίβαση και την πιθανή κατάχρησή τους.
- ✚ Ακρίβεια: Παράγοντες όπως οι συνθήκες φωτισμού και η ποιότητα της εικόνας μπορούν να επηρεάσουν την ακρίβεια των βιομετρικών χαρακτηριστικών.
- ✚ Απομίμηση: Τα βιομετρικά συστήματα μπορεί να είναι ευάλωτα σε απόπειρες πλαστογράφησης με τη χρήση πλαστών δακτυλικών αποτυπωμάτων, φωτογραφιών ή ηχογραφήσεων φωνής.



Νομοθεσία και κανονισμοί: Η χρήση των βιομετρικών συστημάτων υπόκειται σε διάφορους νόμους και κανονισμούς που διαφέρουν ανάλογα με τη δεδομένη δικαιοδοσία.

Τα βιομετρικά χαρακτηριστικά προσφέρουν έναν βολικό και ασφαλή τρόπο για την πιστοποίηση της ταυτότητας των ατόμων σε διάφορες πτυχές, ωστόσο η εφαρμογή τους χρήζει ιδιαίτερης προσοχής και πρέπει να λαμβάνει υπόψη τις ανησυχίες για την προστασία της ιδιωτικής ζωής και την ασφάλεια. Καθώς η τεχνολογία εξελίσσεται, τα βιομετρικά συνεχίζουν να διαδραματίζουν σημαντικό ρόλο στην επαλήθευση της ταυτότητας και στον έλεγχο πρόσβασης.

2.2 Εφαρμογή και Είδη TN

2.2.1 Είδη TN

Τύποι τεχνητής νοημοσύνης βασισμένοι στις ικανότητες μάθησης

Ο διαχωρισμός αυτός γίνεται με βάση τον τρόπο με τον οποίο μαθαίνουν και σε ποιο βαθμό μπορούν να εφαρμόσουν τις γνώσεις τους τα συστήματα TN, όλη η τεχνητή νοημοσύνη μπορεί να χωριστεί σε τρεις τύπους ικανοτήτων: Τεχνητή «στενή» νοημοσύνη, τεχνητή «γενική» νοημοσύνη και την τεχνητή υπερνοημοσύνη [59]:

Artificial narrow intelligence

Η τεχνητή «στενή» νοημοσύνη (Artificial narrow intelligence - ANI), επίσης γνωστή ως στενή TN ή αδύναμη TN, περιγράφει εργαλεία TN σχεδιασμένα να εκτελούν πολύ συγκεκριμένες ενέργειες ή εντολές. Οι τεχνολογίες ANI είναι κατασκευασμένες για να εξυπηρετούν και να υπερέχουν σε μια γνωστική ικανότητα και δεν μπορούν να μάθουν δεξιότητες ανεξάρτητα από τον σχεδιασμό τους. Συχνά χρησιμοποιούν αλγορίθμους μηχανικής μάθησης και νευρωνικών δικτύων για να ολοκληρώσουν αυτές τις συγκεκριμένες εργασίες. Για παράδειγμα, η επεξεργασία φυσικής γλώσσας με TN είναι ένας τύπος στενής νοημοσύνης, επειδή μπορεί να αναγνωρίζει και να ανταποκρίνεται σε φωνητικές εντολές, αλλά δεν μπορεί να εκτελέσει άλλες εργασίες πέραν αυτού. Ορισμένα παραδείγματα τεχνητής στενής νοημοσύνης περιλαμβάνουν το λογισμικό αναγνώρισης εικόνων, τα αυτοκινούμενα αυτοκίνητα και τους εικονικούς βοηθούς TN όπως η Siri [60].

Artificial general intelligence

Η τεχνητή «γενική» νοημοσύνη (Artificial general intelligence – AGI), που ονομάζεται επίσης γενική ΤΝ ή ισχυρή ΤΝ, περιγράφει την ΤΝ που μπορεί να μαθαίνει, να σκέφτεται και να εκτελεί ένα ευρύ φάσμα ενεργειών παρόμοια με τον άνθρωπο. Ο στόχος του σχεδιασμού της τεχνητής γενικής νοημοσύνης είναι να μπορέσουμε να δημιουργήσουμε μηχανές που θα είναι ικανές να εκτελούν πολυλειτουργικές εργασίες και θα λειτουργούν ως αληθοφανείς, εξίσου ευφυείς βοηθοί των ανθρώπων στην καθημερινή ζωή. Παρόλο που είναι ακόμη σε εξέλιξη, τα θεμέλια της τεχνητής γενικής νοημοσύνης θα μπορούσαν να οικοδομηθούν από τεχνολογίες όπως οι υπερυπολογιστές, το κβαντικό υλικό και τα παραγωγικά μοντέλα ΤΝ όπως το ChatGPT [61].

Artificial superintelligence (ASI)

Η τεχνητή υπερ-νοημοσύνη (Artificial superintelligence - ASI), ή υπερ-ΤΝ, είναι υλικό επιστημονικής φαντασίας. Θεωρείται ότι όταν η τεχνητή νοημοσύνη φτάσει στο επίπεδο της γενικής νοημοσύνης, σύντομα θα μαθαίνει με τόσο γρήγορο ρυθμό που οι γνώσεις και οι ικανότητές της θα γίνουν ισχυρότερες ακόμη και από αυτές της ανθρωπότητας.

Η Τεχνητή Υπερνοημοσύνη αντιπροσωπεύει μια υποθετική μορφή ΤΝ που ξεπερνά την ανθρώπινη νοημοσύνη και ικανότητες σε κάθε πτυχή. Η ASI θα έχει τη δυνατότητα να διαθέτει ένα πρωτοφανές επίπεδο γνώσεων, ικανότητας επίλυσης προβλημάτων και δημιουργικότητας. Πρόκειται για μια έννοια που εγείρει θέματα συζήτησης και προβληματισμού όσον αφορά στην ηθική και τη φιλοσοφία της ΤΝ. Η ASI θα λειτουργήσει ως η ραχοκοκαλιά της τεχνολογίας της πλήρως αυτοσυνείδητης τεχνητής νοημοσύνης και των αυτόνομων ρομπότ [62].

«Η τεχνητή υπερ-νοημοσύνη θα γίνει μακράν η πιο ικανή μορφή νοημοσύνης στη γη», δήλωσε ο David Rogenmoser, διευθύνων σύμβουλος της εταιρείας συγγραφής AI Jasper «Θα έχει τη νοημοσύνη των ανθρώπων και θα είναι εξαιρετικά καλύτερη σε όλα όσα κάνουμε εμείς»⁸.

Τύποι τεχνητής νοημοσύνης βασισμένοι στη λειτουργικότητα

Η λειτουργικότητα αφορά τον τρόπο με τον οποίο μια τεχνητή νοημοσύνη εφαρμόζει τις μαθησιακές της ικανότητες για να επεξεργάζεται δεδομένα, να ανταποκρίνεται σε ερεθίσματα και να

⁸ <https://builtin.com/artificial-intelligence/types-of-artificial-intelligence>

αλληλεπιδρά με το περιβάλλον της. Ως εκ τούτου, η τεχνητή νοημοσύνη μπορεί να ταξινομηθεί με βάση τέσσερις τύπους λειτουργικότητας:

Reactive Machines (Αντιδραστικές Μηχανές)

Η γένεση της τεχνητής νοημοσύνης ξεκίνησε με την ανάπτυξη των αντιδραστικών μηχανών, του πιο θεμελιώδους τύπου τεχνητής νοημοσύνης. Οι αντιδραστικές μηχανές μπορούν να ανταποκριθούν σε άμεσα αιτήματα και εργασίες, αλλά δεν είναι ικανές να αποθηκεύουν μνήμη ή να μαθαίνουν από προηγούμενες εμπειρίες. Δηλαδή, δεν μπορούν να βελτιώσουν τη λειτουργικότητά τους μέσω της εμπειρίας και μπορούν να ανταποκριθούν μόνο σε έναν περιορισμένο συνδυασμό εργασιών.

Στην πράξη, οι αντιδραστικές μηχανές μπορούν να διαβάζουν και να ανταποκρίνονται σε εξωτερικά ερεθίσματα σε πραγματικό χρόνο. Αυτό τις καθιστά χρήσιμες για την εκτέλεση βασικών αυτόνομων λειτουργιών, όπως το φιλτράρισμα των ανεπιθύμητων μηνυμάτων από τα εισερχόμενα email ή η σύσταση ταινιών με βάση τις πιο πρόσφατες αναζητήσεις των χρηστών.

Η πιο διάσημη, η αντιδραστική μηχανή τεχνητής νοημοσύνης Deep Blue της IBM μπόρεσε να διαβάσει ενδείξεις σε πραγματικό χρόνο για να νικήσει τον Ρώσο γκραν-μετρ του σκακιού Γκάρι Κασπάροφ σε έναν αγώνα σκακιού το 1997⁹. Αλλά πέρα από αυτό, η αντιδραστική τεχνητή νοημοσύνη δεν μπορεί να βασιστεί σε προηγούμενες γνώσεις ή να εκτελέσει πιο σύνθετες εργασίες. Προκειμένου να εφαρμοστεί η TN σε πιο προηγμένα σενάρια, έπρεπε να υπάρξουν εξελίξεις στην αποθήκευση δεδομένων και στη διαχείριση της μνήμης [63].

Limited Memory (Μηχανές Περιορισμένης Μνήμης)

Το επόμενο βήμα στην εξέλιξη της τεχνητής νοημοσύνης είναι η ανάπτυξη ικανότητας αποθήκευσης γνώσεων. Τα συστήματα τεχνητής νοημοσύνης περιορισμένη μνήμης μπορούν να αποθηκεύουν και να ανακτούν γνώσεις από προηγούμενες εμπειρίες για να βελτιώνουν την απόδοσή τους και να προσαρμόζονται σε διαφορετικές καταστάσεις. Τα συστήματα αυτά μπορούν να μαθαίνουν και να εκπαιδεύονται για μελλοντικές εργασίες αξιοποιώντας τις αποθηκευμένες πληροφορίες. Η TN με περιορισμένη μνήμη είναι ζωτικής σημασίας σε εφαρμογές όπως τα αυτόνομα οχήματα και τα παιχνίδια TN. Όλα τα σημερινά συστήματα τεχνητής νοημοσύνης εκπαιδεύονται από μεγάλους

⁹ <https://builtin.com/artificial-intelligence/types-of-artificial-intelligence>

όγκους δεδομένων εκπαίδευσης που αποθηκεύουν στη μνήμη τους για να σχηματίσουν ένα μοντέλο αναφοράς για την επίλυση μελλοντικών προβλημάτων.

Το 2012, ο τομέας της τεχνητής νοημοσύνης σημείωσε σημαντική πρόοδο. Νέες καινοτομίες από την Google και την Image Net κατέστησαν δυνατή την αποθήκευση δεδομένων του παρελθόντος από την τεχνητή νοημοσύνη και την πραγματοποίηση προβλέψεων με τη χρήση τους. Αυτός ο τύπος τεχνητής νοημοσύνης αναφέρεται ως τεχνητή νοημοσύνη περιορισμένης μνήμης, επειδή μπορεί να δημιουργήσει τη δική της περιορισμένη βάση γνώσεων και να χρησιμοποιήσει αυτή τη γνώση για να βελτιωθεί με την πάροδο του χρόνου. Σήμερα, το μοντέλο περιορισμένης μνήμης αντιπροσωπεύει την πλειονότητα των εφαρμογών τεχνητής νοημοσύνης [64].

Theory of Mind

Όσον αφορά την πρόοδο της Τεχνητής Νοημοσύνης, η τεχνολογία περιορισμένης μνήμης είναι το πιο μακρινό σημείο στο οποίο έχουμε φτάσει - αλλά δεν είναι ο τελικός προορισμός. Οι μηχανές περιορισμένης μνήμης μπορούν να μαθαίνουν από προηγούμενες εμπειρίες και να αποθηκεύουν γνώσεις, αλλά δεν μπορούν να αντιληφθούν λεπτές περιβαλλοντικές αλλαγές, συναισθηματικές ενδείξεις ή να φτάσουν στο ίδιο επίπεδο με την ανθρώπινη νοημοσύνη.

Η τεχνητή νοημοσύνη του τύπου Θεωρίας του Νου (Theory of Mind) αντιπροσωπεύει συστήματα ικανά να αντιλαμβάνονται και να ανταποκρίνονται στα ανθρώπινα συναισθήματα και προθέσεις. Εκτός από τις ικανότητες της Τεχνητής Νοημοσύνης Περιορισμένης Μνήμης, τα συστήματα αυτά μπορούν να κατανοούν και να συναισθάνονται τα ανθρώπινα συναισθήματα. Μπορούν να συμπεραίνουν τις νοητικές καταστάσεις των χρηστών και να προσαρμόζουν τη συμπεριφορά τους ανάλογα. Το συγκεκριμένο είδος ΤΝ έχει εφαρμογές στην κοινωνική ρομποτική και στην αλληλεπίδραση ανθρώπου-υπολογιστή.

«Τα τρέχοντα μοντέλα έχουν μια μονόδρομη σχέση», δήλωσε ο Rogenmoser. «Τα εργαλεία τεχνητής νοημοσύνης όπως η Alexa και η Siri δεν αντιδρούν με συναισθηματική υποστήριξη όταν τους φωνάζετε»¹⁰.

Η έννοια της τεχνητής νοημοσύνης που μπορεί να αντιλαμβάνεται τα ανθρώπινα συναισθήματα δεν έχει ακόμη υλοποιηθεί πλήρως. Αυτή η έννοια αναφέρεται ως «θεωρία του νου», ένας όρος

¹⁰ <https://builtin.com/artificial-intelligence/types-of-artificial-intelligence>

δανεισμένος από την ψυχολογία που περιγράφει την ικανότητα των ανθρώπων να διαβάζουν τα συναισθήματα των άλλων και να προβλέπουν μελλοντικές ενέργειες με βάση αυτές τις πληροφορίες.

«Οι μηχανές μπορεί να λειτουργούν καλύτερα από εμάς στο 90 τοις εκατό του χρόνου, αλλά αυτό το τελευταίο δέκα τοις εκατό, αυτό που θα περιγράφαμε ως κοινή λογική, είναι πραγματικά δύσκολο να το πετύχουμε» σύμφωνα με τους επιστήμονες¹¹. Για παράδειγμα, ένα αυτοκινούμενο αυτοκίνητο μπορεί να αποδίδει καλύτερα από έναν ανθρώπινο οδηγό την πλειονότητα του χρόνου, επειδή δεν θα κάνει τα ίδια ανθρώπινα λάθη. Αλλά αν ο οδηγός, γνωρίζει ότι το παιδί του γείτονά του έχει την τάση να παίζει κοντά στο δρόμο μετά το σχολείο, θα ξέρει ενστικτωδώς ότι πρέπει να επιβραδύνει περνώντας από το δρομάκι του γείτονα - κάτι που ένα όχημα τεχνητής νοημοσύνης εξοπλισμένο με βασική περιορισμένη μνήμη δεν θα ήταν σε θέση να κάνει.

Η θεωρία του νου θα μπορούσε να φέρει πολλές θετικές αλλαγές στον κόσμο της τεχνολογίας, αλλά ενέχει και τους δικούς της κινδύνους. Δεδομένου ότι οι συναισθηματικές ενδείξεις είναι τόσο διαφοροποιημένες, θα χρειαστεί πολύς χρόνος για να τελειοποιήσουν οι μηχανές τεχνητής νοημοσύνης την ανάγνωσή τους, και θα μπορούσαν ενδεχομένως να κάνουν μεγάλα λάθη ενώ βρίσκονται στο στάδιο της εκμάθησης. Ορισμένοι φοβούνται επίσης ότι μόλις οι τεχνολογίες είναι σε θέση να ανταποκρίνονται στα συναισθηματικά σήματα, το αποτέλεσμα θα μπορούσε να σημαίνει την πλήρη αυτοματοποίηση ορισμένων θέσεων εργασίας [65].

Self - Aware

Η ΤΝ με αυτογνωσία (Self – Aware) είναι ένα θεωρητικό είδος όπου τα συστήματα ΤΝ όχι μόνο διαθέτουν τις δυνατότητες της περιορισμένης μνήμης και της θεωρίας του νου, αλλά και αυτογνωσία και νοημοσύνη ανθρώπινου επιπέδου. Αυτό το επίπεδο ΤΝ θα έχει βαθιά κατανόηση της ίδιας της ύπαρξής του, των συναισθημάτων και της συνείδησής του. Η επίτευξη της αυτογνωσίας της ΤΝ παραμένει μια εξαιρετικά κερδοσκοπική και φιλοσοφική ιδέα.

Στο στάδιο αυτό όταν η τεχνητή νοημοσύνη αναπτύξει αυτογνωσία, θα αναφέρεται ως το σημείο μοναδικότητας της ΤΝ. Πιστεύεται ότι όταν επιτευχθεί αυτό το σημείο, οι μηχανές ΤΝ θα είναι πέρα

¹¹ <https://builtin.com/artificial-intelligence/types-of-artificial-intelligence>

από τον έλεγχό μας, επειδή θα μπορούν όχι μόνο να αισθάνονται τα συναισθήματα των άλλων, αλλά θα έχουν και αίσθηση του εαυτού τους.

«Οι άνθρωποι τόσο προσπαθούν να δημιουργήσουν αυτού του είδους την τεχνητή νοημοσύνη όσο και φοβούνται τις συνέπειες της δημιουργίας της, ανησυχώντας ότι αυτού του είδους η τεχνητή νοημοσύνη θα μπορούσε να κλέψει τις δουλειές μας ή να καταλάβει τον κόσμο μας», «Αν δημιουργηθεί επιτυχώς αυτός ο τύπος ΤΝ, κανείς δεν ξέρει ποιες θα είναι οι επιπτώσεις» δήλωσε ο Rogenmoser¹².

Οι ερευνητές και οι μηχανικοί κάνουν βήματα για να αναπτύξουν στοιχειώδεις εκδόσεις αυτοσυνείδητης τεχνητής νοημοσύνης. Ίσως ένα από τα πιο διάσημα εργαλεία αυτού του είδους ΤΝ είναι η Sophia, ένα ρομπότ που αναπτύχθηκε από την εταιρεία ρομποτικής Hanson Robotics¹³. Αν και τεχνικά δεν έχει αυτογνωσία, η προηγμένη εφαρμογή των σημερινών τεχνολογιών ΤΝ στη Sophia παρέχει μια ματιά στο δυνητικά αυτοσυνείδητο μέλλον της ΤΝ. Πρόκειται για ένα μέλλον που υπόσχεται πολλά αλλά και εγκυμονεί κινδύνους - και υπάρχει συζήτηση για το αν είναι ηθικό να δημιουργούμε καθόλα νοήμονα ΤΝ εργαλεία [66].

2.2.2 Εφαρμογές ΤΝ

Κάποιες από βασικές εφαρμογές της Τεχνητής Νοημοσύνης συναντώνται στους κάτωθι τομείς [67]:

Υγειονομική περίθαλψη

-  **Ιατρική διάγνωση:** Τα διαγνωστικά εργαλεία με τεχνητή νοημοσύνη χρησιμοποιούν την αναγνώριση εικόνων και την ανάλυση προτύπων για τον εντοπισμό ασθενειών όπως ο καρκίνος, τα κατάγματα και οι παθήσεις των ματιών. Τα μοντέλα βαθιάς μάθησης μπορούν να αναλύουν ιατρικές εικόνες με μεγάλη ακρίβεια και να βοηθούν τους γιατρούς να λαμβάνουν τεκμηριωμένες αποφάσεις με βάση οπτικά δεδομένα.
-  **Ανακάλυψη φαρμάκων:** Η τεχνητή νοημοσύνη επιταχύνει την ανακάλυψη φαρμάκων αναλύοντας τεράστια σύνολα βιολογικών δεδομένων. Τα μοντέλα μηχανικής μάθησης προβλέπουν πιθανά υποψήφια φάρμακα, τις αλληλεπιδράσεις τους με τα μόρια-στόχους και τις παρενέργειες, μειώνοντας σημαντικά το χρόνο και το κόστος της έρευνας.

¹² <https://builtin.com/artificial-intelligence/types-of-artificial-intelligence>

¹³ <https://www.hansonrobotics.com/sophia/>

- ✚ *Φροντίδα ασθενών:* Οι προγνωστικές αναλύσεις στην υγειονομική περίθαλψη χρησιμοποιούν την TN για την ανάλυση δεδομένων ασθενών, τον εντοπισμό πρώιμων προειδοποιητικών σημείων και την πρόβλεψη της εξέλιξης της νόσου. Αυτό επιτρέπει προληπτικές παρεμβάσεις, μειώνοντας τις επανεισαγωγές στα νοσοκομεία και βελτιώνοντας τα αποτελέσματα των ασθενών [67].

Οικονομία

- ✚ *Algorithmic trading:* Οι αλγόριθμοι τεχνητής νοημοσύνης αναλύουν ιστορικά δεδομένα της αγοράς και ροές ειδήσεων σε πραγματικό χρόνο για τη λήψη αποφάσεων που αφορούν σε οικονομικές διαπραγματεύσεις. Αυτοί οι αλγόριθμοι χρησιμοποιούν προγνωστικές αναλύσεις για τον εντοπισμό ευκαιριών διαπραγμάτευσης, τη διαχείριση του κινδύνου και τη βελτιστοποίηση των αποδόσεων του εκάστοτε χαρτοφυλακίου.
- ✚ *Credit Scoring:* Τα μοντέλα TN αναλύουν το πιστωτικό ιστορικό, το εισόδημα και τη συμπεριφορά ενός δανειολήπτη για την αξιολόγηση του πιστωτικού κινδύνου. Η προγνωστική ανάλυση βοηθά τους δανειστές να λαμβάνουν τεκμηριωμένες αποφάσεις δανεισμού και να καθορίζουν τα κατάλληλα επιτόκια.
- ✚ *Ανίχνευση απάτης:* Η TN εφαρμόζει αλγορίθμους ανίχνευσης ανωμαλιών και μηχανικής μάθησης για να εντοπίζει παρατυπίες στις χρηματοοικονομικές συναλλαγές. Αυτά τα μοντέλα μαθαίνουν από ιστορικά μοτίβα απάτης και προσαρμόζονται σε νέους τύπους απάτης, μειώνοντας τα ψευδώς θετικά αποτελέσματα και τις ενδεχόμενες απώλειες [67].

Αυτόνομα οχήματα

- ✚ *Αυτοκινούμενα αυτοκίνητα:* Τα αυτόνομα οχήματα χρησιμοποιούν δεδομένα αισθητήρων (LiDAR, κάμερες, ραντάρ) σε πραγματικό χρόνο, για να αναλύουν το περιβάλλον τους. LiDAR, κάμερες και ραντάρ, για να αναλύσουν το περιβάλλον τους. Τα μοντέλα μηχανικής μάθησης ερμηνεύουν αυτά τα δεδομένα για να λαμβάνουν αποφάσεις οδήγησης, όπως η ανίχνευση πεζών, άλλων οχημάτων και σηματοδοτών.
- ✚ *Μη επανδρωμένα αεροσκάφη:* Τα μη επανδρωμένα αεροσκάφη με τεχνητή νοημοσύνη χρησιμοποιούν την όραση υπολογιστή και τη συγχώνευση αισθητήρων για την πλοήγηση και την εκτέλεση εργασιών. Αναλυτικοί αλγόριθμοι επεξεργάζονται δεδομένα από αισθητήρες που βρίσκονται επί του σκάφους για να επιτύχουν αυτόνομη πτήση και να ολοκληρώσουν αποστολές, όπως η αεροφωτογράφιση και η επιτήρηση [67].

Επεξεργασία φυσικής γλώσσας (NLP)

- ✚ *Chatbots*: Τα chatbots χρησιμοποιούν την κατανόηση της φυσικής γλώσσας και την ανάλυση συναισθήματος για να συνομιλούν με τους χρήστες σε συνομιλίες που μοιάζουν με τις ανθρώπινες. Τα αναλυτικά μοντέλα NLP εξάγουν πολύτιμες πληροφορίες από τις αλληλεπιδράσεις των πελατών για τη βελτίωση της ποιότητας των υπηρεσιών.
- ✚ *Αυτόματη Μετάφραση*: Στατιστικά και νευρωνικά μοντέλα αυτόματης μετάφρασης αναλύουν γλωσσικά πρότυπα και περιβάλλοντα για τη μετάφραση κειμένου μεταξύ γλωσσών. Τα αναλυτικά μοντέλα βελτιώνουν συνεχώς την ακρίβεια της μετάφρασης με βάση την ανατροφοδότηση των χρηστών και τα δεδομένα παράλληλων κειμένων [67].

Ηλεκτρονικό εμπόριο

- ✚ *Συστήματα συστάσεων*: Τα συστήματα συστάσεων συνεργατικού φιλτραρίσματος και τα συστήματα συστάσεων βάσει περιεχομένου αναλύουν τη συμπεριφορά, τις προτιμήσεις και τα ιστορικά δεδομένα των χρηστών για να παρέχουν εξατομικευμένες συστάσεις προϊόντων. Τα αναλυτικά μοντέλα βελτιστοποιούν τα ποσοστά κλικ και οδηγούν σε αύξηση των πωλήσεων.
- ✚ *Βελτιστοποίηση της αλυσίδας εφοδιασμού*: Τα συστήματα πρόβλεψης ζήτησης και διαχείρισης αποθεμάτων με βάση την τεχνητή νοημοσύνη χρησιμοποιούν αναλυτικές μεθόδους για την ελαχιστοποίηση του κόστους της αλυσίδας εφοδιασμού, τη μείωση των πλεοναζόντων αποθεμάτων και την αποτελεσματικότερη ικανοποίηση της ζήτησης των πελατών [67].

2.3 Πλεονεκτήματα της ΤΝ

Η Τεχνητή Νοημοσύνη (ΤΝ) προσφέρει ένα ευρύ φάσμα πλεονεκτημάτων σε διάφορους κλάδους και εφαρμογές, προωθώντας την καινοτομία και την αποτελεσματικότητα. Ακολουθούν ορισμένα από τα βασικά πλεονεκτήματα της Τεχνητής Νοημοσύνης [68],[69]:

- ✓ *Αυτοματοποίηση επαναλαμβανόμενων εργασιών*: Η ΤΝ μπορεί να αυτοματοποιήσει τις συνήθεις και επαναλαμβανόμενες εργασίες, επιτρέποντας στους ανθρώπους να επικεντρωθούν σε πιο δημιουργικές και σύνθετες πτυχές της εργασίας τους. Αυτό οδηγεί σε αυξημένη παραγωγικότητα και αποδοτικότητα.

- ✓ *Ανάλυση δεδομένων και αναγνώριση προτύπων:* Τα συστήματα TN υπερέχουν στην επεξεργασία και ανάλυση τεράστιων ποσοτήτων δεδομένων γρήγορα και με ακρίβεια. Μπορούν να εντοπίζουν μοτίβα, τάσεις και ιδέες που μπορεί να είναι δύσκολο ή αδύνατο να διακρίνουν οι άνθρωποι, επιτρέποντας τη λήψη αποφάσεων βάσει δεδομένων.
- ✓ *Βελτιωμένη ακρίβεια και συνέπεια:* Τα συστήματα τεχνητής νοημοσύνης εκτελούν με συνέπεια εργασίες με μεγάλη ακρίβεια και χωρίς κόπωση, μειώνοντας τα σφάλματα που μπορεί να προκύψουν λόγω της ανθρώπινης κόπωσης ή ασυνέπειας.
- ✓ *Μείωση του κόστους:* Η αυτοματοποίηση μέσω της TN μπορεί να οδηγήσει σε σημαντική εξοικονόμηση κόστους στον εργασιακό τομέα το οποίο προκύπτει με τη μείωση του εργατικού κόστους, την ελαχιστοποίηση των ανθρώπινων σφαλμάτων και τη βελτιστοποίηση της κατανομής των πόρων.
- ✓ *Βελτιωμένη εξυπηρέτηση πελατών:* Τα chatbots και οι εικονικοί βοηθοί που λειτουργούν με TN παρέχουν υποστήριξη πελατών 24 ώρες το 24ωρο, απαντούν σε ερωτήματα και επιλύουν ζητήματα άμεσα, ανταποκρίνονται στις ανάγκες και βελτιώνουν την ικανοποίηση των πελατών.
- ✓ *Εξατομίκευση περιεχομένου:* Τα συστήματα συστάσεων με βάση την τεχνητή νοημοσύνη αναλύουν τα δεδομένα των χρηστών για να παρέχουν εξατομικευμένο περιεχόμενο, συστάσεις προϊόντων και υπηρεσιών, βελτιώνοντας τις εμπειρίες των χρηστών και ενισχύοντας την αφοσίωση των πελατών.
- ✓ *Ταχύτερη λήψη αποφάσεων:* Οι αλγόριθμοι TN μπορούν να επεξεργάζονται και να αναλύουν δεδομένα σε πραγματικό χρόνο, επιτρέποντας στις επιχειρήσεις να λαμβάνουν ταχύτερες και πιο τεκμηριωμένες αποφάσεις, ιδίως σε κρίσιμες καταστάσεις που χαρακτηρίζονται από περιορισμένο χρόνο.
- ✓ *Ανάλυση προβλέψεων:* Η TN μπορεί να προβλέψει μελλοντικά αποτελέσματα και τάσεις αναλύοντας προηγούμενα δεδομένα. Αυτή η ικανότητα είναι πολύτιμη σε τομείς όπως η πρόβλεψη της ζήτησης, η ανίχνευση απάτης, η προληπτική συντήρηση μηχανημάτων και φυσικά η υγειονομική περίθαλψη.
- ✓ *Βελτιωμένα διαγνωστικά αποτελέσματα στον τομέα της υγειονομικής περίθαλψης:* Η τεχνητή νοημοσύνη βοηθά τους επαγγελματίες υγείας στη διάγνωση ασθενειών, στην ερμηνεία ιατρικών εικόνων και στον εντοπισμό θεραπευτικών επιλογών, οδηγώντας σε έγκαιρες και ακριβέστερες διαγνώσεις.

- ✓ *Αυτόνομα συστήματα:* Η τεχνητή νοημοσύνη επιτρέπει την ανάπτυξη αυτόνομων οχημάτων, μη επανδρωμένων αεροσκαφών και ρομπότ που μπορούν να εκτελούν εργασίες σε επικίνδυνα ή δύσκολα περιβάλλοντα, βελτιώνοντας την ασφάλεια και την αποτελεσματικότητα.
- ✓ *Ενισχυμένη ασφάλεια:* Τα συστήματα τεχνητής νοημοσύνης μπορούν να εντοπίζουν ανωμαλίες και πιθανές απειλές σε πραγματικό χρόνο, ενισχύοντας τα μέτρα κυβερνοασφάλειας και ανίχνευσης απάτης.
- ✓ *Επεξεργασία φυσικής γλώσσας (NLP):* Τα συστήματα TN που βασίζονται σε NLP κατανοούν και παράγουν την ανθρώπινη γλώσσα, βελτιώνοντας την επικοινωνία και επιτρέποντας εφαρμογές όπως οι φωνητικοί βοηθοί, η γλωσσική μετάφραση και η ανάλυση συναισθήματος.
- ✓ *Περιβαλλοντικές επιπτώσεις:* Η τεχνητή νοημοσύνη μπορεί να βελτιστοποιήσει την κατανάλωση ενέργειας, τη διαχείριση των πόρων και την παρακολούθηση κρίσιμων περιβαλλοντικών παραγόντων συμβάλλοντας στις προσπάθειες βιωσιμότητας και μειώνοντας τις επιβλαβείς επιπτώσεις της δράσης του ανθρώπινου παράγοντα στο περιβάλλον.
- ✓ *Έρευνα και ανακάλυψη:* Η τεχνητή νοημοσύνη βοηθά τους επιστήμονες και τους ερευνητές στην ανάλυση σύνθετων συνόλων δεδομένων, στην προσομοίωση πειραμάτων και στην ανακάλυψη νέων προτύπων ή λύσεων σε τομείς όπως η φυσική, η βιολογία και η χημεία.
- ✓ *Προσβασιμότητα:* Οι τεχνολογίες TN βελτιώνουν την προσβασιμότητα για τα άτομα με αναπηρίες παρέχοντας βοηθητικές συσκευές και προσαρμοσμένες στις ανάγκες τους εφαρμογές, καθιστώντας την τεχνολογία ένα χρήσιμο εργαλείο βελτίωσης της ποιότητας ζωής τους.

Αυτά τα πλεονεκτήματα αναδεικνύουν τη δυνατότητα της TN να ενισχύσει την παραγωγικότητα, την αποδοτικότητα και την καινοτομία σε διάφορους τομείς, καθιστώντας την μετασχηματιστική δύναμη στον σύγχρονο κόσμο. Ωστόσο, είναι σημαντικό να σημειωθεί ότι η υπεύθυνη και ηθική ανάπτυξη και ανάπτυξη της TN είναι κρίσιμης σημασίας για τη μεγιστοποίηση των πλεονεκτημάτων της, ενώ παράλληλα αντιμετωπίζονται οι πιθανές προκλήσεις και κίνδυνοι [68],[69].

2.4 Κίνδυνοι της ΤΝ

Ενώ η Τεχνητή Νοημοσύνη προσφέρει πολυάριθμα οφέλη, ενέχει επίσης ορισμένους κινδύνους και προκλήσεις που πρέπει να διαχειριστούν προσεκτικά. Ακολουθούν ορισμένοι από τους βασικούς κινδύνους που σχετίζονται με την ΤΝ [69]–[71]:

- ❖ *Μεροληψία και δικαιοσύνη:* Τα συστήματα τεχνητής νοημοσύνης μπορεί να «κληρονομήσουν» προκαταλήψεις που υπάρχουν στα δεδομένα με τα οποία εκπαιδεύονται, οδηγώντας σε άδικα ή μεροληπτικά αποτελέσματα. Είναι ζωτικής σημασίας να αντιμετωπιστεί η προκατάληψη στους αλγόριθμους ΤΝ για να διασφαλιστεί η δικαιοσύνη και η ισότητα, ιδίως σε εφαρμογές όπως οι προσλήψεις, ο δανεισμός και η ποινική δικαιοσύνη.
- ❖ *Ανησυχίες σχετικά με την προστασία της ιδιωτικής ζωής:* Η ικανότητα της τεχνητής νοημοσύνης να επεξεργάζεται και να αναλύει τεράστιες ποσότητες προσωπικών δεδομένων εγείρει ανησυχίες για την προστασία της ιδιωτικής ζωής. Η μη εξουσιοδοτημένη πρόσβαση, οι παραβιάσεις δεδομένων και η κατάχρηση προσωπικών πληροφοριών μπορεί να οδηγήσουν σε σημαντικές παραβιάσεις της ιδιωτικής ζωής.
- ❖ *Τρωτά σημεία ασφαλείας:* Τα συστήματα τεχνητής νοημοσύνης μπορεί να είναι ευάλωτα σε επιθέσεις, συμπεριλαμβανομένων των επιθέσεων όπου κακόβουλοι φορείς χειραγωγούν δεδομένα εισόδου για να εξαπατήσουν αλγόριθμους τεχνητής νοημοσύνης. Επιπλέον, οι απειλές κυβερνοασφάλειας που βασίζονται στην ΤΝ μπορούν να γίνουν πιο εξελιγμένες όπως αναλύθηκαν στο προηγούμενο κεφάλαιο.
- ❖ *Αλλαγές στο εργασιακό πεδίο:* Η αυτοματοποίηση των εργασιακών καθηκόντων μέσω της ΤΝ και της ρομποτικής μπορεί να οδηγήσει σε εκτόπιση θέσεων εργασίας σε ορισμένους κλάδους, επηρεάζοντας δυνητικά τις ευκαιρίες απασχόλησης για ορισμένους εργαζομένους.
- ❖ *Ηθικά διλήμματα:* Η τεχνητή νοημοσύνη μπορεί να θέσει ηθικά διλήμματα στις διαδικασίες λήψης αποφάσεων. Για παράδειγμα, τα αυτόνομα οχήματα κατά τη λειτουργία του μπορεί να χρειαστεί να αντιμετωπίσουν καταστάσεις όπου πρέπει να κάνουν ηθικές επιλογές, όπως η προτεραιότητα της ασφάλειας των επιβατών έναντι της ασφάλειας των πεζών.
- ❖ *Έλλειψη διαφάνειας:* Πολλοί αλγόριθμοι τεχνητής νοημοσύνης, ιδίως τα μοντέλα βαθιάς μάθησης, θεωρούνται "μαύρα κουτιά" επειδή οι διαδικασίες λήψης αποφάσεων δεν είναι εύκολα ερμηνεύσιμες. Η έλλειψη διαφάνειας μπορεί να οδηγήσει σε προκλήσεις όσον αφορά την κατανόηση και την εξήγηση των αποφάσεων της ΤΝ.

- ❖ *Λογοδοσία και απόδοση ευθυνών:* Ο καθορισμός της ευθύνης σε περιπτώσεις σφαλμάτων ή ζημιών της λειτουργίας ενός συστήματος ΤΝ μπορεί να είναι πολύπλοκος. Χρειάζονται σαφείς κατευθυντήριες γραμμές και πολύ καλά καθορισμένα νομικά πλαίσια για την αντιμετώπιση των ζητημάτων λογοδοσίας.
- ❖ *Κατάχρηση της τεχνητής νοημοσύνης σε κυβερνοεπιθέσεις:* Οι κακόβουλοι φορείς μπορούν να αξιοποιήσουν την ΤΝ για να εξαπολύσουν πιο εξελιγμένες κυβερνοεπιθέσεις, καθιστώντας δυσκολότερο τον εντοπισμό και την άμυνα κατά των απειλών.
- ❖ *Απώλεια ελέγχου:* Καθώς τα συστήματα ΤΝ γίνονται όλο και πιο αυτόνομα και ικανά, υπάρχει ανησυχία για την απώλεια του ελέγχου αυτών των συστημάτων, ιδίως σε κρίσιμες εφαρμογές όπως είναι τα αυτόνομα όπλα.
- ❖ *Κανονισμοί για το απόρρητο των δεδομένων:* Οι κανονισμοί, όπως ο Γενικός Κανονισμός για την Προστασία Δεδομένων (GDPR), απαιτούν από τους οργανισμούς να χειρίζονται τα δεδομένα με υπευθυνότητα. Τα συστήματα τεχνητής νοημοσύνης πρέπει να συμμορφώνονται με αυτούς τους κανονισμούς, γεγονός που οδηγεί σε πιθανές νομικές προκλήσεις και προκλήσεις συμμόρφωσης.
- ❖ *Εξάρτηση από την τεχνητή νοημοσύνη:* Η υπερβολική εξάρτηση από τα συστήματα ΤΝ, ιδίως σε διαδικασίες λήψης αποφάσεων, μπορεί να οδηγήσει τους ανθρώπους να ακολουθούν τυφλά τις συστάσεις της ΤΝ χωρίς κριτική αξιολόγηση.
- ❖ *Μετατόπιση θέσεων εργασίας και δεξιοτήτων:* Ενώ η ΤΝ μπορεί να εκτοπίσει ορισμένες θέσεις εργασίας, δημιουργεί επίσης νέους ρόλους και απαιτήσεις για δεξιότητες στην ανάπτυξη, τη συντήρηση και την εποπτεία της ΤΝ. Ενδέχεται να υπάρξει ανάγκη για αναβάθμιση και επανεκπαίδευση του εργατικού δυναμικού.

Για την αντιμετώπιση αυτών των κινδύνων, είναι σημαντικό να υιοθετηθούν υπεύθυνες πρακτικές διαχείρισης της ΤΝ, συμπεριλαμβανομένης της διασφάλισης της ποιότητας των δεδομένων, του μετριασμού των προκαταλήψεων, της εξασφάλισης της διαφάνειας και την αντιμετώπιση των ηθικών διλημάτων που εγείρονται στην ανάπτυξη της ΤΝ. Κυβερνήσεις, οργανισμοί και ερευνητές εργάζονται ενεργά πάνω σε κατευθυντήριες γραμμές και κανονισμούς για να διασφαλίσουν την υπεύθυνη και ηθική χρήση της τεχνολογίας ΤΝ. Η εξισορρόπηση των πλεονεκτημάτων της ΤΝ με τη διαχείριση των συναφών κινδύνων είναι ζωτικής σημασίας για την αξιοποίηση του πλήρους δυναμικού της ΤΝ με ασφαλή και ηθικό τρόπο [69]–[71].

3. Defensive AI

Στο διαρκώς εξελισσόμενο ψηφιακό τοπίο του 21ου αιώνα, η διάχυτη ανάπτυξη της τεχνολογίας έχει επιφέρει ασύγκριτες ευκαιρίες και ευκολίες. Ωστόσο, αυτή η ψηφιακή επανάσταση δημιουργήσε επίσης έναν τρομερό αντίπαλο: τις απειλές στον κυβερνοχώρο που παραμονεύουν στις σκιές του διασυνδεδεμένου κόσμου. Η ταχεία επέκταση των διαδικτυακών πλατφορμών, ο πολλαπλασιασμός των πολύτιμων δεδομένων και η αυξανόμενη πολυπλοκότητα των κυβερνοεπιθέσεων έχουν υπογραμμίσει την κρίσιμη σημασία της κυβερνοασφάλειας για τη διαφύλαξη των ψηφιακών μας περιουσιακών στοιχείων και της ιδιωτικής μας ζωής [72], [73].

Η ασφάλεια στον κυβερνοχώρο, που κάποτε θεωρούνταν δευτερεύουσα ανησυχία, έχει πλέον ανέλθει στην πρωταρχική θέση στον παγκόσμιο ψηφιακό κόσμο. Η προστασία των ευαίσθητων δεδομένων, των κρίσιμων υποδομών και της ιδιωτικής ζωής των ατόμων έχει καταστεί επιτακτική ανάγκη. Σε αυτή τη μάχη κατά των αντιπάλων στον κυβερνοχώρο, τόσο οι οργανισμοί όσο και τα άτομα πρέπει να χρησιμοποιούν καινοτόμους και ισχυρούς αμυντικούς μηχανισμούς για να αποκρούσουν τις πανταχού παρούσες απειλές.

Η αμυντική τεχνητή νοημοσύνη (Defensive AI) είναι μια πρωτοποριακή και μετασχηματιστική προσέγγιση για την ενίσχυση της ψηφιακής μας άμυνας. Η Defensive AI, μια διασταύρωση της τεχνητής νοημοσύνης και της ασφάλειας στον κυβερνοχώρο, αντιπροσωπεύει μια αλλαγή παραδείγματος στην ικανότητά μας να προβλέπουμε, να εντοπίζουμε και να μετριάζουμε τις απειλές στον κυβερνοχώρο με απaráμιλλη ακρίβεια και ταχύτητα. Αυτό το κεφάλαιο εμβαθύνει στον πολύπλευρο κόσμο της Defensive AI διερευνώντας τις βασικές αρχές, τις μεθοδολογίες και τον καθοριστικό ρόλο που διαδραματίζει στην ενίσχυση της κυβερνοασφάλειας [72], [73].

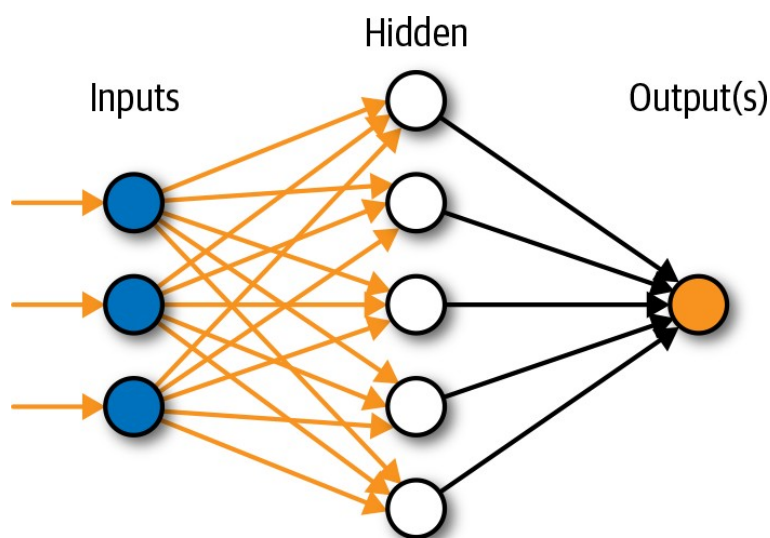
Η Defensive AI αποτελεί απόδειξη της δύναμης των ευφυών συστημάτων στην προληπτική προστασία από μια σειρά αντιπάλων στον κυβερνοχώρο, από μεμονωμένους χάκερ έως κρατικούς φορείς. Αξιοποιώντας τις δυνατότητες της μηχανικής μάθησης, της ανίχνευσης ανωμαλιών και των μηχανισμών απόκρισης σε πραγματικό χρόνο, της Defensive AI εξοπλίζει τους οργανισμούς με τα μέσα για να προβλέπουν τις απειλές, να προσαρμόζονται στους εξελισσόμενους φορείς επιθέσεων και να ανταποκρίνονται γρήγορα σε πιθανές παραβιάσεις. Καθώς επιχειρούμε να εισέλθουμε βαθύτερα στην ψηφιακή εποχή, η κατανόηση των αποχρώσεων της Defensive AI δεν είναι απλώς μια

επιλογή, αλλά μια αναγκαιότητα για τη διαφύλαξη των ψηφιακών μας περιουσιακών στοιχείων και την τήρηση των αρχών της ιδιωτικότητας, της ακεραιότητας και της αξιοπιστίας [72], [73].

3.1 Neural Nets

Στο πεδίο της Defensive AI, η χρήση τεχνητών νευρωνικών δικτύων (Neural Nets) έχει αναδειχθεί σε καθοριστική δύναμη. Τα τεχνητά νευρωνικά δίκτυα, που συχνά αναφέρονται ως νευρωνικά δίκτυα, αποτελούν θεμελιώδες συστατικό των συστημάτων της Defensive AI, στηρίζοντας την ικανότητά τους να ανιχνεύουν, να αναλύουν και να ανταποκρίνονται σε απειλές στον κυβερνοχώρο με αξιοσημείωτη ακρίβεια. Σε αυτό το υποκεφάλαιο, εμβαθύνουμε στον κρίσιμο ρόλο που διαδραματίζουν τα νευρωνικά δίκτυα στην ενίσχυση της ασφάλειας στον κυβερνοχώρο.

Στον πυρήνα του, ένα νευρωνικό δίκτυο είναι ένα διασυνδεδεμένο σύστημα τεχνητών νευρώνων, εμπνευσμένο από τα βιολογικά νευρωνικά δίκτυα στον ανθρώπινο εγκέφαλο. Αυτοί οι διασυνδεδεμένοι νευρώνες σχηματίζουν επίπεδα, με ένα επίπεδο εισόδου, ένα ή περισσότερα κρυφά επίπεδα και ένα επίπεδο εξόδου. Τα νευρωνικά δίκτυα φημίζονται για την ικανότητά τους να εκτελούν πολύπλοκες εργασίες, ιδίως στους τομείς της αναγνώρισης προτύπων, της ταξινόμησης και της λήψης αποφάσεων.



Εικόνα 8: Παράδειγμα Τεχνητού Νευρωνικού Δικτύου¹⁴

¹⁴ <https://www.innoarchitech.com/blog/artificial-intelligence-deep-learning-neural-networks-explained>

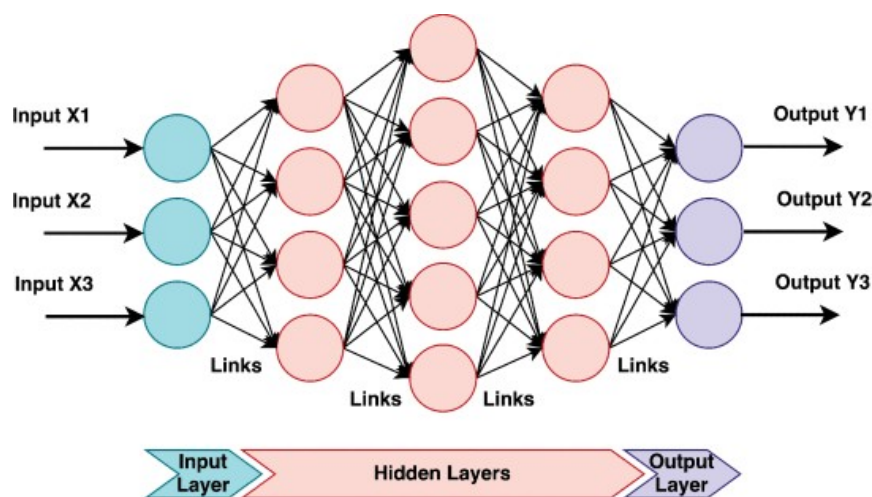
Η εφαρμογή των νευρωνικών δικτύων στην Defensive AI είναι πολύπλευρη, με κύρια έμφαση στην αναγνώριση προτύπων και την ανίχνευση ανωμαλιών. Τα νευρωνικά δίκτυα είναι ικανά να μαθαίνουν από μεγάλα σύνολα δεδομένων, επιτρέποντάς τους να διακρίνουν λεπτά μοτίβα και αποκλίσεις από την κανονική συμπεριφορά εντός του ψηφιακού οικοσυστήματος. Ακολουθούν ορισμένες βασικές εφαρμογές των νευρωνικών δικτύων στην κυβερνοασφάλεια [74]:

-  Συστήματα ανίχνευσης εισβολών (Intrusion Detection Systems - IDS): Τα νευρωνικά δίκτυα παίζουν καθοριστικό ρόλο στα IDS, συγκεκριμένα αναλύουν την κυκλοφορία του δικτύου και τα αρχεία καταγραφής του συστήματος για τον εντοπισμό ύποπτων ή κακόβουλων δραστηριοτήτων. Με την κατάλληλη εκπαίδευση σε προηγούμενα δεδομένα, τα νευρωνικά δίκτυα μπορούν να εντοπίζουν προηγουμένως αθέατα μοτίβα επιθέσεων και να εγείρουν συναγερμούς / ειδοποιήσεις σε πραγματικό χρόνο.
-  Ανίχνευση κακόβουλου λογισμικού: Στο αέναο παιχνίδι «γάτας – ποντικιού» που συντελείται κατά την προσπάθεια της ανίχνευσης κακόβουλου λογισμικού, τα νευρωνικά δίκτυα έχουν καθοριστικής σημασίας συμβολή στην αντιμετώπιση των απειλών καθώς μπορούν να εντοπίσουν κακόβουλο λογισμικό αναλύοντας τον κώδικα, τη συμπεριφορά και τους φορείς επίθεσης, ακόμη και όταν το κακόβουλο λογισμικό προσπαθεί να αποκρύψει την παρουσία του.
-  Ανίχνευση phishing: Τα νευρωνικά δίκτυα υπερέχουν στον εντοπισμό επιθέσεων phishing εξετάζοντας διεξοδικά το περιεχόμενο του ηλεκτρονικού ταχυδρομείου, τους συνδέσμους και τη συμπεριφορά του αποστολέα. Μπορούν να εντοπίσουν μηνύματα ηλεκτρονικού ταχυδρομείου phishing που μπορεί να αποφύγουν τα παραδοσιακά φίλτρα που βασίζονται σε κανόνες.
-  Ανάλυση της συμπεριφοράς των χρηστών: Παρακολουθώντας τη συμπεριφορά των χρηστών, τα νευρωνικά δίκτυα μπορούν να ανιχνεύσουν ανωμαλίες που μπορεί να υποδεικνύουν παραβιασμένους λογαριασμούς ή εσωτερικές απειλές. Αναλύουν τα μοτίβα της δραστηριότητας των χρηστών και προκαλούν συναγερμούς/ειδοποιήσεις όταν εμφανίζονται αποκλίσεις από τον κανόνα.

Ενώ τα νευρωνικά δίκτυα υπόσχονται πολλά για την ενίσχυση της ασφάλειας στον κυβερνοχώρο, η χρήση τους ενέχει και προκλήσεις. Η φύση του «μαύρου κουτιού» των βαθιών νευρωνικών δικτύων μπορεί να εμποδίσει την ερμηνευσιμότητα, καθιστώντας δύσκολη την εξήγηση των αποφάσεών τους.

Επιπλέον, οι αντίπαλες επιθέσεις μπορούν να εκμεταλλευτούν τα τρωτά σημεία των μοντέλων των νευρωνικών δικτύων.

Ως βαθιά νευρωνικά δίκτυα (Deep Neural Networks - DNN) μπορούν να οριστούν τα δίκτυα με περισσότερα από ένα κρυμμένα επίπεδα. Με αυτό τον ορισμό καλύπτονται όλες οι βαθιές δομές, όπως τα συνελκτικά, τα Μακράς Βραχυπρόθεσμης Μνήμης, και τα παραγωγικά αντιπαραθετικά δίκτυα [2]. Ένα βαθύ νευρωνικό δίκτυο είναι ένα νευρωνικό δίκτυο με πολλαπλά κρυφά επίπεδα. Παρακάτω απεικονίζεται η βασική του δομή [75]:



Εικόνα 9: Απλό παράδειγμα DNN [75]

Οι πρόσφατες εξελίξεις στην έρευνα για τα νευρωνικά δίκτυα, όπως οι τεχνικές εξηγήσιμης τεχνητής νοημοσύνης (explainable AI - XAI) και η εκπαίδευση αντιπάλων, στοχεύουν στην αντιμετώπιση αυτών των προκλήσεων. Οι μέθοδοι XAI παρέχουν πληροφορίες για το πώς τα νευρωνικά δίκτυα λαμβάνουν αποφάσεις, ενισχύοντας τη διαφάνεια και τη λογοδοσία τους. Η αντίπαλη εκπαίδευση βοηθά τα νευρωνικά δίκτυα να γίνουν πιο ανθεκτικά σε αντίπαλες επιθέσεις. Η ικανότητά τους για αναγνώριση προτύπων, προσαρμοστικότητα και ανάλυση σε πραγματικό χρόνο εξοπλίζει τους οργανισμούς με ένα ισχυρό εργαλείο στη μάχη κατά των αντιπάλων στον κυβερνοχώρο.

3.2 Intelligent Agents

Οι ευφυείς πράκτορες (Intelligent Agents), στις διάφορες μορφές τους, χαρακτηρίζονται από την ικανότητά τους να αντιλαμβάνονται το περιβάλλον τους, να σκέφτονται γι' αυτό, να λαμβάνουν αποφάσεις και να εκτελούν ενέργειες για την επίτευξη συγκεκριμένων στόχων. Αυτοί οι πράκτορες μπορεί να κυμαίνονται από απλά σενάρια και συστήματα βασισμένα σε κανόνες έως πολύπλοκα μοντέλα που βασίζονται στη μάθηση. Στο πλαίσιο της ασφάλειας στον κυβερνοχώρο, οι ευφυείς πράκτορες λειτουργούν ως ψηφιακοί φρουροί, παρακολουθώντας, αναλύοντας και αντιδρώντας συνεχώς σε πιθανές απειλές. Οι ευφυείς πράκτορες στην Defensive AI εμφανίζονται σε διάφορες μορφές, η καθεμία προσαρμοσμένη σε συγκεκριμένα καθήκοντα και τοπία απειλών. Ακολουθούν ορισμένοι βασικοί τύποι ευφυών πρακτόρων που χρησιμοποιούνται συνήθως στην κυβερνοασφάλεια [76], [77]:

-  Πράκτορες ανίχνευσης εισβολών: Αυτοί οι πράκτορες λειτουργούν μέσα σε περιβάλλοντα δικτύου και συστήματα υποδοχής για να ανιχνεύουν ανώμαλες δραστηριότητες, απόπειρες μη εξουσιοδοτημένης πρόσβασης και πιθανές εισβολές. Χρησιμοποιούν αλγορίθμους μηχανικής μάθησης, ανάλυση συμπεριφοράς και αναγνώριση προτύπων για τον εντοπισμό απειλών.
-  Πράκτορες αυτόνομης αντίδρασης: Οι πράκτορες αυτόνομης αντίδρασης έχουν τη δυνατότητα να αναλαμβάνουν προκαθορισμένες ενέργειες ως απάντηση σε εντοπισμένες απειλές. Μπορούν να απομονώσουν συστήματα που έχουν παραβιαστεί, να μπλοκάρουν την ύποπτη κυκλοφορία δικτύου ή να δρομολογήσουν διαδικασίες αντιμετώπισης περιστατικών, όλα αυτά χωρίς ανθρώπινη παρέμβαση.
-  Πράκτορες πληροφοριών απειλών: Αυτοί οι πράκτορες έχουν σχεδιαστεί για να συλλέγουν, να αναλύουν και να διαδίδουν δεδομένα πληροφοριών για απειλές. Συλλέγουν συνεχώς πληροφορίες σχετικά με τις αναδυόμενες απειλές, τα τρωτά σημεία και τις τεχνικές επίθεσης, δίνοντας στους οργανισμούς τη δυνατότητα να βρίσκονται μπροστά από τους πιθανούς επιτιθέμενους.
-  Πράκτορες ανάλυσης της συμπεριφοράς των χρηστών: Οι πράκτορες ανάλυσης της συμπεριφοράς των χρηστών επικεντρώνονται στην παρακολούθηση και ανάλυση της συμπεριφοράς των χρηστών στο ψηφιακό οικοσύστημα ενός οργανισμού. Μπορούν να

εντοπίσουν αποκλίσεις από τα συνήθη πρότυπα συμπεριφοράς, βοηθώντας στον εντοπισμό εσωτερικών απειλών και παραβιασμένων λογαριασμών.

Οι ευφυείς πράκτορες στην Defensive AI συχνά αξιοποιούν τεχνικές μηχανικής μάθησης και τεχνητής νοημοσύνης για να ενισχύσουν τις δυνατότητές τους. Οι αλγόριθμοι μηχανικής μάθησης επιτρέπουν στους πράκτορες να προσαρμόζονται στα εξελισσόμενα τοπία απειλών, να αναγνωρίζουν προηγουμένως αθέατα μοτίβα επιθέσεων και να λαμβάνουν τεκμηριωμένες αποφάσεις σε πραγματικό χρόνο. Παρά το γεγονός ότι οι ευφυείς πράκτορες προσφέρουν σημαντικά πλεονεκτήματα στην ενίσχυση της άμυνας της κυβερνοασφάλειας, θα πρέπει να εξεταστούν προσεκτικά οι ηθικοί προβληματισμοί σχετικά με την αυτονομία αυτών των πρακτόρων, το ενδεχόμενο ψευδώς θετικών αποτελεσμάτων και την ανάγκη ανθρώπινης επίβλεψης. Επιπλέον, η προσαρμοστικότητα των ευφύων πρακτόρων μπορεί να τους καταστήσει ευάλωτους σε αντίπαλες επιθέσεις. Οι μελλοντικές ερευνητικές κατευθύνσεις μπορεί να περιλαμβάνουν την ανάπτυξη πιο επεξηγήσιμων και ερμηνεύσιμων πρακτόρων, καθώς και τεχνικές για τον μετριασμό των αντιπολιτευτικών επιθέσεων σε αυτούς τους πράκτορες.

3.3 Learning

Η μηχανική μάθηση (Learning), ένα υποσύνολο της τεχνητής νοημοσύνης, βρίσκεται στην πρώτη γραμμή της Defensive AI, επιτρέποντας στους οργανισμούς να αξιοποιήσουν τη δύναμη των γνώσεων που βασίζονται σε δεδομένα για να προβλέψουν, να ανιχνεύσουν και να ανταποκριθούν στις απειλές στον κυβερνοχώρο. Αυτό το υπόκεφάλαιο εμβαθύνει στον καθοριστικό ρόλο που διαδραματίζει η μηχανική μάθηση στην ενίσχυση της άμυνας της κυβερνοασφάλειας [73].

Στον πυρήνα της, η μηχανική μάθηση είναι μια προσέγγιση που βασίζεται σε δεδομένα και δίνει τη δυνατότητα στους υπολογιστές να μαθαίνουν από αυτά τα δεδομένα, να κάνουν προβλέψεις ή να λαμβάνουν αποφάσεις με βάση αυτά. Περιλαμβάνει ένα ποικίλο σύνολο αλγορίθμων και τεχνικών, καθένα από τα οποία έχει σχεδιαστεί για συγκεκριμένες εργασίες, όπως η ταξινόμηση, η παλινδρόμηση, η ομαδοποίηση και η ανίχνευση ανωμαλιών.

Η εφαρμογή της μηχανικής μάθησης στην Defensive AI είναι πολύπλευρη, φέρνοντας επανάσταση στην ικανότητά μας να εντοπίζουμε και να μετριάζουμε προληπτικά τις απειλές στον κυβερνοχώρο. Οι τεχνικές μηχανικής μάθησης συμβάλλουν καθοριστικά στην ανάλυση μεγάλου όγκου δεδομένων, στην αναγνώριση μοτίβων και στην προσαρμογή σε εξελισσόμενους φορείς επιθέσεων. Ακολουθούν βασικές εφαρμογές της μηχανικής μάθησης στην κυβερνοασφάλεια [78]:

- ✚ Ανίχνευση ανωμαλιών: Τα μοντέλα μηχανικής μάθησης υπερέχουν στην ανίχνευση ανωμαλιών μαθαίνοντας τη φυσιολογική συμπεριφορά των συστημάτων και των χρηστών. Μπορούν να εντοπίσουν αποκλίσεις από τις καθιερωμένες βασικές γραμμές, που ενδεχομένως σηματοδοτούν παραβιάσεις της ασφάλειας ή ασυνήθιστες δραστηριότητες.
- ✚ Ανίχνευση κακόβουλου λογισμικού: Οι αλγόριθμοι μηχανικής μάθησης αναλύουν τα χαρακτηριστικά των αρχείων, τη συμπεριφορά του κώδικα και την κυκλοφορία του δικτύου για τον εντοπισμό υπογραφών και συμπεριφορών κακόβουλου λογισμικού. Μπορούν να ταξινομήσουν και να ανιχνεύσουν γνωστές απειλές κακόβουλου λογισμικού και απειλές μηδενικής ημέρας.
- ✚ Ανίχνευση phishing: Τα μοντέλα μηχανικής μάθησης εξετάζουν διεξοδικά το περιεχόμενο του ηλεκτρονικού ταχυδρομείου, τη συμπεριφορά του αποστολέα και τα μοτίβα διευθύνσεων URL για να ανιχνεύσουν απόπειρες phishing. Ενισχύουν την ασφάλεια ηλεκτρονικού ταχυδρομείου με τον εντοπισμό κακόβουλων μηνυμάτων ηλεκτρονικού ταχυδρομείου.
- ✚ Ανάλυση της συμπεριφοράς χρηστών και οντοτήτων (User and Entity Behavior Analytics - UEBA): Η UEBA αξιοποιεί τη μηχανική μάθηση για την παρακολούθηση και την ανάλυση της συμπεριφοράς χρηστών και οντοτήτων σε έναν οργανισμό. Μπορεί να εντοπίσει ασυνήθιστα μοτίβα πρόσβασης, που ενδεχομένως υποδεικνύουν εσωτερικές απειλές.

Ζητήματα όπως η ποιότητα των δεδομένων, οι αντίπαλες επιθέσεις και η ερμηνευσιμότητα των μοντέλων πρέπει να αντιμετωπιστούν για να μεγιστοποιηθεί η αποτελεσματικότητα της μηχανικής μάθησης στην Defensive AI. Οι πρόσφατες εξελίξεις στην έρευνα της μηχανικής μάθησης, συμπεριλαμβανομένης της εξηγήσιμης ΤΝ (ΧΑΙ) και της ομοσπονδιακής μάθησης (Federated Learning), στοχεύουν στην αντιμετώπιση αυτών των προκλήσεων. Οι μέθοδοι ΧΑΙ παρέχουν διαφάνεια στις αποφάσεις των μοντέλων μηχανικής μάθησης, ενισχύοντας την εμπιστοσύνη και την ερμηνευσιμότητα. Η ομοσπονδιακή μάθηση επιτρέπει την εκπαίδευση μοντέλων σε αποκεντρωμένες πηγές δεδομένων, διατηρώντας την ιδιωτικότητα και την ασφάλεια.

3.4 Επίλυση περιορισμών

Η επίλυση περιορισμών, που χαρακτηρίζεται από την ικανότητά της Defensive AI να επιβάλλει πολιτικές και περιορισμούς ασφαλείας, διαδραματίζει καθοριστικό ρόλο στην ενίσχυση της ψηφιακής άμυνας και στη διατήρηση της ακεραιότητας των ψηφιακών οικοσυστημάτων. Σε αυτό το

υποκεφάλαιο, διερευνούμε τη σημασία των τεχνικών επίλυσης περιορισμών στο πλαίσιο της Defensive AI.

Η επίλυση περιορισμών, που συχνά αναφέρεται ως έλεγχος πρόσβασης, αντιπροσωπεύει ένα σύνολο μηχανισμών και πρωτοκόλλων που διέπουν την πρόσβαση σε ψηφιακούς πόρους και συστήματα. Αυτοί οι μηχανισμοί επιβάλλουν πολιτικές, περιορισμούς και δικαιώματα για να διασφαλίσουν ότι μόνο εξουσιοδοτημένες οντότητες μπορούν να έχουν πρόσβαση σε συγκεκριμένους πόρους ή να εκτελούν ορισμένες ενέργειες.

Οι τεχνικές επίλυσης περιορισμών βρίσκουν εφαρμογή σε διάφορες πτυχές της ασφάλειας στον κυβερνοχώρο, διασφαλίζοντας ότι τα κρίσιμα περιουσιακά στοιχεία παραμένουν προστατευμένα και οι μη εξουσιοδοτημένες δραστηριότητες αποτρέπονται. Ακολουθούν βασικές εφαρμογές της επίλυσης περιορισμών στην Defensive AI [79]:

-  Πολιτικές ελέγχου πρόσβασης: Οι πολιτικές ελέγχου πρόσβασης, που επιβάλλονται μέσω της επίλυσης περιορισμών, υπαγορεύουν ποιος μπορεί να έχει πρόσβαση σε συγκεκριμένους πόρους εντός ενός ψηφιακού οικοσυστήματος. Αυτές οι πολιτικές είναι απαραίτητες για τη διασφάλιση ευαίσθητων δεδομένων, δικτυακών πόρων και κρίσιμων υποδομών.
-  Διαχείριση ταυτότητας και πιστοποίησης ταυτότητας: Οι τεχνικές επίλυσης περιορισμών διαδραματίζουν κρίσιμο ρόλο στη διαχείριση ταυτότητας και αυθεντικοποίησης. Επαληθεύουν την ταυτότητα των χρηστών, των συσκευών ή των οντοτήτων που επιδιώκουν πρόσβαση σε συστήματα και διασφαλίζουν ότι χορηγείται πρόσβαση μόνο σε πιστοποιημένες οντότητες.
-  Έλεγχος πρόσβασης βάσει ρόλων (Role-Based Access Control - RBAC): Τα συστήματα RBAC χρησιμοποιούν την επίλυση περιορισμών για να κατανέμουν δικαιώματα με βάση προκαθορισμένους ρόλους σε έναν οργανισμό. Η προσέγγιση αυτή βελτιώνει τη διαχείριση της πρόσβασης και μειώνει τον κίνδυνο μη εξουσιοδοτημένης πρόσβασης.

Ενώ οι τεχνικές επίλυσης περιορισμών ενισχύουν τις άμυνες στον χώρο της κυβερνοασφάλειας, δεν είναι χωρίς προκλήσεις και ηθικούς προβληματισμούς. Οι προκλήσεις περιλαμβάνουν την πολυπλοκότητα της διαχείρισης των πολιτικών ελέγχου πρόσβασης σε μεγάλους οργανισμούς, τη διασφάλιση της επεκτασιμότητας και την αντιμετώπιση του κινδύνου κλιμάκωσης των προνομιών. Οι ηθικοί προβληματισμοί περιστρέφονται γύρω από τη δικαιοσύνη και τη διαφάνεια των πολιτικών ελέγχου πρόσβασης. Η διασφάλιση ότι οι πολιτικές δεν εισάγουν διακρίσεις και ότι τα άτομα έχουν

ορατότητα και έλεγχο των δικαιωμάτων πρόσβασής τους είναι υψίστης σημασίας. Οι τεχνικές επίλυσης περιορισμών αποτελούν έναν κρίσιμο πυλώνα στο οικοδόμημα της Defensive AI, παρέχοντας τις απαραίτητες προστατευτικές μπάρες για την προστασία των ψηφιακών περιουσιακών στοιχείων και τη διατήρηση της ακεραιότητας των ψηφιακών οικοσυστημάτων. Επιβάλλοντας πολιτικές ελέγχου πρόσβασης, επαληθεύοντας ταυτότητες και διαχειριζόμενες άδειες, η επίλυση περιορισμών διασφαλίζει ότι μόνο εξουσιοδοτημένες οντότητες περιηγούνται στο ψηφιακό τοπίο [80].

4. Offensive AI

Στο διαρκώς εξελισσόμενο πεδίο της κυβερνοασφάλειας, η δυαδικότητα της τεχνολογίας γίνεται όλο και πιο εμφανής. Ενώ η επιδίωξη των αμυντικών συστημάτων TN (Defensive AI) αποσκοπεί στην ενίσχυση των ψηφιακών μας οχυρώσεων και στην προστασία από τις απειλές στον κυβερνοχώρο, ταυτόχρονα αναδύεται μια αντίθετη δύναμη: Η επιθετική τεχνητή νοημοσύνη (Offensive AI). Αυτό το κεφάλαιο ξεκινά ένα ταξίδι στον ιδιαίτερα ενδιαφέροντα τομέα της Offensive AI, ρίχνοντας φως στις προκλήσεις και τις απειλές που παρουσιάζει στο τοπίο της κυβερνοασφάλειας.

Καθώς εξελίσσεται η τεχνολογία, εξελίσσονται και οι στρατηγικές που χρησιμοποιούν οι κακόβουλοι φορείς στο ψηφιακό πεδίο. Η Offensive AI, στον πυρήνα της, αντιπροσωπεύει την ενσωμάτωση της τεχνητής νοημοσύνης και της μηχανικής μάθησης στο οπλοστάσιο των επιτιθέμενων στον κυβερνοχώρο. Ενσαρκώνει μια νέα γενιά απειλών στον κυβερνοχώρο που είναι προσαρμοστικές, εξελιγμένες και ικανές να παρακάμπτουν τις παραδοσιακές άμυνες. Η Offensive AI παρουσιάζει μια έντονη αντίθεση με την Defensive AI, καθώς επιδιώκει να εκμεταλλευτεί τα τρωτά σημεία, να παρακάμψει τα μέτρα ασφαλείας και να σπείρει την απειλή μέσα στα ψηφιακά οικοσυστήματα [81].

4.1 Επιθέσεις με χρήση Τεχνητής Νοημοσύνης

Στο δυναμικό τοπίο της κυβερνοασφάλειας, η σύγκλιση της τεχνητής νοημοσύνης και των επιθετικών τακτικών σηματοδοτεί μια αλλαγή παραδείγματος στις δυνατότητες των κακόβουλων φορέων. Αυτό το υποκεφάλαιο εμβαθύνει στο πεδίο των επιθέσεων με χρήση τεχνητής νοημοσύνης, διερευνώντας πώς οι τεχνικές που βασίζονται στην TN δίνουν τη δυνατότητα στους αντιπάλους του κυβερνοχώρου να εξαπολύουν εξελιγμένες επιθέσεις σε ψηφιακά συστήματα και οργανισμούς [81].

Η εξάπλωση της τεχνολογίας ΤΝ έχει μεταμορφώσει τη σύγχρονη βιομηχανία και έχει ενισχύσει σε μεγάλο βαθμό την παραγωγικότητα, ωστόσο έχει ενδυναμώσει τους εγκληματίες του κυβερνοχώρου παρέχοντάς τους νέα εργαλεία και μεθοδολογίες. Οι επιθέσεις με χρήση ΤΝ (AI – powered) αποτελούν μια κατηγορία απειλών όπου οι κακόβουλοι φορείς αξιοποιούν τεχνικές ΤΝ για να σχεδιάζουν, να εκτελούν και να αποφεύγουν την ανίχνευση κατά τη διάρκεια κυβερνοεπιθέσεων. Αυτές οι επιθέσεις αξιοποιούν την ικανότητα της ΤΝ να αναλύει τεράστια σύνολα δεδομένων, να προσαρμόζεται στις μεταβαλλόμενες συνθήκες και να λαμβάνει αποφάσεις σε πραγματικό χρόνο. Η εφαρμογή της τεχνητής νοημοσύνης σε επιθετικές επιχειρήσεις στον κυβερνοχώρο έχει οδηγήσει στην ανάπτυξη νέων φορέων και τεχνικών επίθεσης. Ακολουθούν βασικοί τομείς στους οποίους η ΤΝ αξιοποιείται από τους εγκληματίες του κυβερνοχώρου [81]:

- ❖ Αντικρουόμενες επιθέσεις: Οι επιθέσεις κατά του αντιπάλου αφορούν τη χειραγώγηση των μοντέλων μηχανικής μάθησης για την παραγωγή εσφαλμένων ή μεροληπτικών αποτελεσμάτων. Οι αντίπαλοι στον κυβερνοχώρο μπορούν να δημιουργήσουν κακόβουλες εισόδους που παραπλανούν τους αλγορίθμους τεχνητής νοημοσύνης, οδηγώντας σε λανθασμένες αποφάσεις στα συστήματα ασφαλείας.
- ❖ Αυτοματοποιημένο ψάρεμα: Τα bots που λειτουργούν με τεχνητή νοημοσύνη (AI – powered bots) μπορούν να δημιουργήσουν εξαιρετικά πειστικά μηνύματα ηλεκτρονικού «ψαρέματος», μιμούμενα το στυλ γραφής και τη συμπεριφορά των νόμιμων αποστολέων. Αυτά τα μηνύματα ηλεκτρονικού ταχυδρομείου έχουν σχεδιαστεί για να εξαπατούν τους παραλήπτες και να αυξάνουν την πιθανότητα επιτυχών επιθέσεων phishing.
- ❖ Σπάσιμο κωδικών πρόσβασης και γέμισμα διαπιστευτηρίων (Credential Stuffing): Οι αλγόριθμοι τεχνητής νοημοσύνης επιταχύνουν τις προσπάθειες παραβίασης κωδικών πρόσβασης, δημιουργώντας και δοκιμάζοντας με επαναληπτικό έξυπνο τρόπο κωδικούς πρόσβασης. Στις επιθέσεις πλήρωσης διαπιστευτηρίων, καθοδηγούμενα από την ΤΝ bots χρησιμοποιούν κλεμμένα διαπιστευτήρια για να αποκτήσουν μη εξουσιοδοτημένη πρόσβαση σε λογαριασμούς χρηστών.
- ❖ Αποφυγή ανίχνευσης ανωμαλιών: Οι κακόβουλοι φορείς αξιοποιούν την Τεχνητή Νοημοσύνη για να αποφεύγουν τα συστήματα ανίχνευσης ανωμαλιών. Το κακόβουλο λογισμικό ενισχυμένο με ΑΙ μπορεί να μιμηθεί την κανονική συμπεριφορά του συστήματος, καθιστώντας δύσκολο για τους παραδοσιακούς ανιχνευτές ανωμαλιών να εντοπίσουν παρατυπίες.

Προκλήσεις για την άμυνα κατά των επιθέσεων με τεχνητή νοημοσύνη

Η άμυνα κατά των επιθέσεων που χρησιμοποιούν τεχνητή νοημοσύνη παρουσιάζει σημαντικές προκλήσεις για τους επαγγελματίες της κυβερνοασφάλειας. Ορισμένες από αυτές τις προκλήσεις περιλαμβάνουν [82]:

-  *Ταχέως εξελισσόμενες τακτικές:* Οι κακόβουλοι φορείς προσαρμόζουν συνεχώς τις τακτικές που βασίζονται στην τεχνητή νοημοσύνη, απαιτώντας από τους αμυντικούς να συμβαδίζουν με τις αναδυόμενες απειλές.
-  *Αδιαφάνεια μοντέλου:* Η πολυπλοκότητα και η αδιαφάνεια των μοντέλων τεχνητής νοημοσύνης που χρησιμοποιούνται στις επιθέσεις μπορεί να καταστήσει δύσκολη την κατανόηση και την πρόβλεψη της συμπεριφοράς τους.
-  *Εντοπισμός επιθέσεων αντιπάλων:* Ο εντοπισμός αντίπαλων επιθέσεων απαιτεί ισχυρά συστήματα ανίχνευσης ανωμαλιών και την ικανότητα διάκρισης μεταξύ νόμιμων και κακόβουλων εισροών.
-  *Εκτιμήσεις δεοντολογίας και μεροληψίας:* Η χρήση της τεχνητής νοημοσύνης σε επιθέσεις εγείρει ηθικούς προβληματισμούς, συμπεριλαμβανομένων ζητημάτων που σχετίζονται με την ιδιωτικότητα και τη δικαιοσύνη.

4.2 Επιθέσεις κατά της Τεχνητής Νοημοσύνης – Μηχανική Μάθηση

Αυτό το υποκεφάλαιο εξετάζει το τοπίο των επιθέσεων κατά της τεχνητής νοημοσύνης, με ιδιαίτερη έμφαση στα μοντέλα μηχανικής μάθησης, διερευνώντας τις μεθοδολογίες, τους κινδύνους και τις στρατηγικές μετριασμού που σχετίζονται με αυτές τις εξελισσόμενες απειλές. Τα μοντέλα μηχανικής μάθησης, γνωστά για την ικανότητά τους να μαθαίνουν από δεδομένα και να κάνουν προβλέψεις ή αποφάσεις, έχουν γίνει αναπόσπαστα συστατικά πολλών συστημάτων κυβερνοασφάλειας. Αυτά τα μοντέλα χρησιμοποιούνται για εργασίες όπως η ανίχνευση εισβολών, η ταξινόμηση κακόβουλου λογισμικού και η ανίχνευση ανωμαλιών. Ωστόσο, η αυξανόμενη υιοθέτησή τους τα έχει καταστήσει επίσης ευάλωτα σε επιθέσεις από αντιπάλους του κυβερνοχώρου [82], [83].

4.2.1 Μεθοδολογίες επίθεσης

Οι επιθέσεις σε μοντέλα μηχανικής μάθησης περιλαμβάνουν ένα φάσμα μεθοδολογιών που αποσκοπούν στην υπονόμευση της ακεραιότητάς τους ή στην αποφυγή της λειτουργικότητάς τους. Οι βασικοί φορείς επίθεσης περιλαμβάνουν [82]:

-  **Επιθέσεις δηλητηρίασης:** Στις επιθέσεις δηλητηρίασης, οι αντίπαλοι χειραγωγούν τα δεδομένα εκπαίδευσης που χρησιμοποιούνται για την ανάπτυξη μοντέλων μηχανικής μάθησης. Εισάγοντας κακόβουλα δεδομένα ή τροποποιώντας νόμιμα δείγματα, οι επιτιθέμενοι μπορούν να θέσουν σε κίνδυνο την απόδοση του μοντέλου ή να εισάγουν ευπάθειες.
-  **Επιθέσεις αποφυγής:** Οι επιθέσεις αποφυγής αποσκοπούν στην εξαπάτηση των μοντέλων μηχανικής μάθησης κατά τη διάρκεια της εξαγωγής συμπερασμάτων. Οι επιτιθέμενοι δημιουργούν εισόδους σχεδιασμένες να προκαλέσουν εσφαλμένη ταξινόμηση ή να εκμεταλλευτούν ευπάθειες του μοντέλου, συχνά για να παρακάμψουν τα μέτρα ασφαλείας.
-  **Αντιστροφή μοντέλου:** Οι επιθέσεις αντιστροφής μοντέλου επικεντρώνονται στην εξαγωγή ευαίσθητων πληροφοριών από μοντέλα μηχανικής μάθησης. Οι αντίπαλοι αντιστρέφουν τα μοντέλα για να αποκαλύψουν εμπιστευτικά δεδομένα που χρησιμοποιήθηκαν κατά την εκπαίδευση, θέτοντας σημαντικούς κινδύνους για την προστασία της ιδιωτικής ζωής.
-  **Αντίπαλες επιθέσεις (adversarial):** Οι επιθέσεις adversarial χειραγωγούν τις εισόδους για να προκαλέσουν εσφαλμένη ταξινόμηση από τα μοντέλα μηχανικής μάθησης. Τα αντίπαλα παραδείγματα είναι προσεκτικά διαμορφωμένα ώστε να φαίνονται σωστά, αλλά οδηγούν το μοντέλο σε λανθασμένες αποφάσεις.

4.2.2 Κίνδυνοι και επιπτώσεις

Οι επιθέσεις σε μοντέλα μηχανικής μάθησης ενέχουν βαθύτατους κινδύνους και συνέπειες για την ασφάλεια στον κυβερνοχώρο [84]:

- ❖ **Παράκαμψη ασφάλειας:** Οι επιτυχείς επιθέσεις μπορούν να υπονομεύσουν τους μηχανισμούς ασφαλείας που βασίζονται στη μηχανική μάθηση, επιτρέποντας ενδεχομένως σε κακόβουλους φορείς να παρακάμψουν την ανίχνευση εισβολών, την ανίχνευση κακόβουλου λογισμικού ή άλλα μέτρα προστασίας.
- ❖ **Παραβιάσεις της ιδιωτικής ζωής:** Οι επιθέσεις αντιστροφής μοντέλου μπορούν να αποκαλύψουν ευαίσθητες πληροφορίες, παραβιάζοντας την ιδιωτικότητα και την εμπιστευτικότητα των χρηστών.

- ❖ Μειωμένη εμπιστοσύνη: Οι επιθέσεις στα μοντέλα μηχανικής μάθησης διαβρώνουν την εμπιστοσύνη στα συστήματα κυβερνοασφάλειας που βασίζονται στην τεχνητή νοημοσύνη, καθιστώντας δύσκολη τη στήριξη σε αυτά τα συστήματα για την ανίχνευση και την πρόληψη απειλών.
- ❖ Λειτουργική διαταραχή: Η διατάραξη της λειτουργικότητας των μοντέλων μηχανικής μάθησης μπορεί να παρεμποδίσει τις συνήθεις επιχειρηματικές λειτουργίες και να οδηγήσει σε οικονομικές απώλειες.

4.2.3 Στρατηγικές μετριασμού

Η άμυνα έναντι επιθέσεων σε μοντέλα μηχανικής μάθησης απαιτεί μια πολύπλευρη προσέγγιση [82]:

- 🛡️ Η διασφάλιση της ποιότητας και της ακεραιότητας των δεδομένων εκπαίδευσης είναι απαραίτητη για τον μετριασμό των επιθέσεων δηλητηρίασης.
- 🛡️ Αντίπαλη (Adversarial) εκπαίδευση: Τεχνικές όπως η αντίπαλη εκπαίδευση και οι στιβαρές (robust) αρχιτεκτονικές μοντέλων μπορούν να ενισχύσουν την ανθεκτικότητα των μοντέλων στις προαναφερθείσες επιθέσεις.
- 🛡️ Παρακολούθηση του μοντέλου: Η συνεχής παρακολούθηση της συμπεριφοράς του μοντέλου μπορεί να βοηθήσει στην ανίχνευση και την αντιμετώπιση των αντίπαλων εισροών και της παρέκκλισης των εννοιών.
- 🛡️ Δεοντολογικά ζητήματα: Οι ηθικοί προβληματισμοί πρέπει να καθοδηγούν την ανάπτυξη και την ανάπτυξη μοντέλων μηχανικής μάθησης, ιδίως σε ευαίσθητους τομείς.

4.3 Επιθετική Τεχνητή Νοημοσύνη έναντι Οργανισμών

Οι τεχνολογίες τεχνητής νοημοσύνης έχουν δώσει τη δυνατότητα σε επιχειρήσεις, κυβερνήσεις και ιδρύματα να βελτιστοποιήσουν τις λειτουργίες τους, να βελτιώσουν τη λήψη αποφάσεων και να αποκτήσουν ανταγωνιστικά πλεονεκτήματα. Ωστόσο, αυτό το κεφάλαιο εξετάζει μια ξεχωριστή πτυχή της επιρροής της ΤΝ - τη δυνατότητά της να χρησιμοποιηθεί επιθετικά εναντίον των οργανισμών. Διερευνούμε πώς η ΤΝ, όταν χρησιμοποιείται με εχθρική πρόθεση, μπορεί να αποτελέσει σημαντική απειλή για τις επιχειρησιακές, οικονομικές πτυχές και τη φήμη των σύγχρονων οργανισμών [85].

4.3.1 Το μοντέλο της επίθεσης

Η ταχεία εξελισσόμενη ανάπτυξη της ΤΝ έχει εξοπλίσει τους κακόβουλους φορείς με μια σειρά εργαλείων και τακτικών σχεδιασμένων να θέτουν σε κίνδυνο την ακεραιότητα και τη σταθερότητα των οργανισμών. Από τις αυτοματοποιημένες κυβερνοεπιθέσεις έως τις εκστρατείες παραπληροφόρησης που καθοδηγούνται από την ΤΝ, η επιθετική ΤΝ εναντίον οργανισμών ενσαρκώνει τη σκοτεινή πλευρά της τεχνολογικής προόδου. Αυτές οι απειλές που τροφοδοτούνται από την ΤΝ συχνά υπερβαίνουν τα παραδοσιακά μέτρα ασφαλείας, απαιτώντας προσαρμοστικές και εξελιγμένες απαντήσεις.

Το υποκεφάλαιο θα εμβαθύνει σε διάφορες μορφές επιθετικής ΤΝ που στοχεύουν οργανισμούς [85]:

-  Κυβερνοεπιθέσεις και εισβολές: Η τεχνητή νοημοσύνη χρησιμοποιείται ολοένα και περισσότερο για την εκτέλεση κυβερνοεπιθέσεων με ταχύτητα και ακρίβεια, παραβιάζοντας την άμυνα των οργανισμών, εξαφανίζοντας ευαίσθητα δεδομένα και προκαλώντας λειτουργικές διαταραχές.
-  Χειραγώγηση από τους αντιπάλους: Οι αντίπαλες τεχνικές μηχανικής μάθησης επιτρέπουν στους επιτιθέμενους να χειραγωγούν συστήματα ΤΝ, όπως αυτοματοποιημένους αλγόριθμους συναλλαγών ή μηχανές συστάσεων, για οικονομικό κέρδος ή ζημία στη φήμη.
-  Κοινωνική μηχανική ενισχυμένη με ΤΝ (AI-Enhanced Social Engineering): Τα κακόβουλα συστήματα ΤΝ μπορούν να διεξάγουν εξαιρετικά στοχευμένες και πειστικές εκστρατείες κοινωνικής μηχανικής, εκμεταλλευόμενα την ανθρώπινη ψυχολογία για να αποκτήσουν μη εξουσιοδοτημένη πρόσβαση ή να θέσουν σε κίνδυνο άτομα εντός οργανισμών.
-  Παραπληροφόρηση με βάση την ΤΝ: Οι οργανισμοί μπορεί να αντιμετωπίσουν εκστρατείες παραπληροφόρησης που δημιουργούνται από ΤΝ και έχουν σχεδιαστεί για να υπονομεύσουν την εμπιστοσύνη του κοινού, να διαταράξουν τις επιχειρηματικές λειτουργίες ή να επηρεάσουν τη λήψη αποφάσεων.

Καθώς οι οργανισμοί βασίζονται όλο και περισσότερο στην ΤΝ για κρίσιμες λειτουργίες, το τοπίο των απειλών συνεχίζει να εξελίσσεται. Οι απειλές που βασίζονται στην ΤΝ μπορεί να εκμεταλλεύονται τρωτά σημεία στα ίδια τα συστήματα ΤΝ ή να χειραγωγούν την ανθρώπινη συμπεριφορά μέσω περιεχομένου που παράγεται από την ΤΝ. Η αντιμετώπιση των κινδύνων που ενέχει η επιθετική ΤΝ απαιτεί μια πολύπλευρη προσέγγιση. Οι οργανισμοί πρέπει να επενδύσουν σε μέτρα ασφαλείας που βασίζονται στην ΤΝ, να καλλιεργήσουν μια κουλτούρα ευαισθητοποίησης στον τομέα της

κυβερνοασφάλειας και να συνεργαστούν με εμπειρογνώμονες του κλάδου για να παραμείνουν ένα βήμα μπροστά από τις αναδυόμενες απειλές.

Η δράση της επιθετικής τεχνητής νοημοσύνης κατά των οργανισμών υπογραμμίζει την ανάγκη για συνεχή επαγρύπνηση και καινοτομία απέναντι στους εξελισσόμενους τεχνολογικούς κινδύνους. Καθώς οι οργανισμοί περιηγούνται στην πολύπλοκη αλληλεπίδραση μεταξύ των πλεονεκτημάτων της ΤΝ και των πιθανών απειλών, πρέπει να παραμείνουν ανθεκτικοί και προσαρμοστικοί για να διαφυλάξουν τα συμφέροντά τους σε έναν ψηφιακό κόσμο όπου η ΤΝ μπορεί να αποτελέσει ταυτόχρονα ασπίδα αλλά και όπλο [85].

5. Anonymization

5.1 Εισαγωγή

Σε μια εποχή που χαρακτηρίζεται από την αμείλικτη αύξηση των ψηφιακών δεδομένων, η διατήρηση της ιδιωτικής ζωής έχει γίνει ένα ολοένα και πιο πιεστικό ζήτημα. Η ανωνυμοποίηση, η πρακτική του μετασχηματισμού ή της αφαίρεσης των προσωπικά αναγνωρίσιμων πληροφοριών (personally identifiable information - PII) από σύνολα δεδομένων, βρίσκεται στο μεταίχμιο της προστασίας της ιδιωτικής ζωής και της χρησιμότητας των δεδομένων. Το παρόν κεφάλαιο ξεκινά μια εξερεύνηση του περίπλοκου τοπίου της ανωνυμοποίησης, εξετάζοντας τον ρόλο της στη διαφύλαξη της ιδιωτικής ζωής των ατόμων, διατηρώντας παράλληλα την αξία και τη χρησιμότητα των δεδομένων στον σύγχρονο κόσμο.

Η εποχή της πληροφορίας έχει εγκαινιάσει μια άνευ προηγουμένου πρόσβαση σε τεράστιες ποσότητες δεδομένα που στηρίζουν τη λήψη αποφάσεων, την καινοτομία και την πρόοδο. Ωστόσο, αυτό το τοπίο με γνώμονα τα δεδομένα έχει επίσης εγείρει βαθιά ερωτήματα σχετικά με την προστασία της ιδιωτικής ζωής και την ασφάλεια των προσωπικών πληροφοριών των ατόμων. Καθώς οι οργανισμοί και τα ιδρύματα συλλέγουν, αναλύουν και μοιράζονται δεδομένα, η ανάγκη για ισχυρές εγγυήσεις προστασίας της ιδιωτικής ζωής δεν ήταν ποτέ μεγαλύτερη.

Η ανωνυμοποίηση αναδεικνύεται ως μια καίρια τεχνική στην προσπάθεια εξισορρόπησης της χρησιμότητας των δεδομένων με τη διατήρηση της ιδιωτικής ζωής. Με τη μετατροπή των δεδομένων κατά τρόπο που καθιστά τα άτομα μη αναγνωρίσιμα, η ανωνυμοποίηση επιδιώκει να προστατεύσει τις ευαίσθητες πληροφορίες από μη εξουσιοδοτημένη πρόσβαση, ενώ παράλληλα επιτρέπει τη χρήση των δεδομένων για νόμιμους σκοπούς, όπως η έρευνα, η ανάλυση και η χάραξη πολιτικής [86].

5.2 Μεθοδολογίες/Τεχνικές

Η ανωνυμοποίηση, ως θεμελιώδης τεχνική για τη διασφάλιση της ιδιωτικής ζωής στην ψηφιακή εποχή, βασίζεται σε μια ποικιλία μεθοδολογιών και τεχνικών για τη μετατροπή των ακατέργαστων δεδομένων σε μια μορφή που αποκρύπτει την ταυτότητα των ατόμων, διατηρώντας παράλληλα τη χρησιμότητα των δεδομένων. Το παρόν υποκεφάλαιο εμβαθύνει στο ποικίλο φάσμα των τεχνικών ανωνυμοποίησης, διερευνώντας τα πλεονεκτήματα, τους περιορισμούς και τις πολύπλοκες εκτιμήσεις που διέπουν την εφαρμογή τους [87]:

Τεχνικές αλλοίωσης δεδομένων

- ✓ Τυχαιοποίηση: Η τυχαιοποίηση εισάγει θόρυβο ή τυχαιότητα στα δεδομένα, καθιστώντας δύσκολο τον εντοπισμό ατομικών ταυτοτήτων. Τεχνικές όπως η προσθήκη τυχαίου θορύβου σε αριθμητικές τιμές ή η ανταλλαγή εγγραφών δεδομένων εμπίπτουν σε αυτή την κατηγορία.
- ✓ Διαφορικό απόρρητο: Η διαφορική ιδιωτικότητα προσφέρει ένα αυστηρό μαθηματικό πλαίσιο για την ποσοτικοποίηση και τον έλεγχο των κινδύνων ιδιωτικότητας. Περιλαμβάνει την προσθήκη προσεκτικά βαθμονομημένου θορύβου στα αποτελέσματα των ερωτημάτων για την προστασία της ατομικής ιδιωτικότητας με παράλληλη διατήρηση των συνολικών στατιστικών στοιχείων [87].

Τεχνικές μετασχηματισμού δεδομένων

- ✓ Γενίκευση: Η γενίκευση περιλαμβάνει την αντικατάσταση συγκεκριμένων χαρακτηριστικών με πιο γενικευμένες κατηγορίες. Για παράδειγμα, ο μετασχηματισμός ακριβών δεδομένων ηλικίας σε ηλικιακές ομάδες (π.χ. 20-30, 31-40) προστατεύει την ιδιωτικότητα, διατηρώντας παράλληλα τα πρότυπα δεδομένων.
- ✓ Απόκρυψη: Η απόκρυψη περιλαμβάνει την παράλειψη ή την απόκρυψη συγκεκριμένων στοιχείων δεδομένων που θα μπορούσαν να χρησιμοποιηθούν για την ταυτοποίηση ατόμων. Η τεχνική αυτή διασφαλίζει ότι δεν αποκαλύπτονται ευαίσθητες πληροφορίες.
- ✓ Ομαδοποίηση (Bucketization ή Binning): Η ομαδοποίηση με κάδους ομαδοποιεί τα δεδομένα σε προκαθορισμένους «κουβάδες» - buckets. Για παράδειγμα, η μετατροπή των συγκεκριμένων τιμών εισοδήματος σε εύρος εισοδήματος βοηθά στην ανωνυμοποίηση των δεδομένων, διατηρώντας παράλληλα χρήσιμες πληροφορίες [87].

Τεχνικές ανταλλαγής δεδομένων

- ✓ k-ανωνυμία: Η k-ανωνυμία διασφαλίζει ότι κάθε εγγραφή σε ένα σύνολο δεδομένων είναι δυσδιάκριτη από τουλάχιστον k-1 άλλες εγγραφές όσον αφορά ένα σύνολο χαρακτηριστικών. Η τεχνική αυτή μετριάζει τον κίνδυνο επαναπροσδιορισμού της ταυτότητας.
- ✓ l-ποικιλομορφία: Η l-ποικιλομορφία επεκτείνει την k-ανωνυμία εξασφαλίζοντας ότι σε κάθε k-ανώνυμη ομάδα, τα ευαίσθητα χαρακτηριστικά έχουν τουλάχιστον l διαφορετικές τιμές. Αυτό αποτρέπει την αποκάλυψη των χαρακτηριστικών.

- ✓ t-Αποκλειστικότητα (t-Closeness): Η t-Closeness επικεντρώνεται στην κατανομή των τιμών των ευαίσθητων χαρακτηριστικών εντός κάθε k-ανώνυμης ομάδας, με στόχο τη μείωση της διαρροής πληροφοριών σχετικά με την κατανομή [87].

5.3 Blockchain Technique

Η τεχνική Blockchain είναι μια επαναστατική τεχνολογία που έχει κερδίσει ευρεία προσοχή λόγω της δυνατότητάς της να εφαρμόζεται σε διάφορους κλάδους, συμπεριλαμβανομένων των χρηματοοικονομικών, της εφοδιαστικής αλυσίδας, της υγειονομικής περίθαλψης και άλλων. Πρόκειται ουσιαστικά για ένα κατανεμημένο σύστημα λογιστικών βιβλίων που καταγράφει με ασφάλεια τις συναλλαγές σε ένα δίκτυο υπολογιστών. Ακολουθεί μια επισκόπηση της τεχνολογίας blockchain [88]:

-  Αποκεντρωση (Decentralization): Ένα από τα βασικά χαρακτηριστικά του blockchain είναι η αποκεντρωση. Σε αντίθεση με τα παραδοσιακά συγκεντρωτικά συστήματα (π.χ. τράπεζες ή κεντρικές βάσεις δεδομένων), το blockchain λειτουργεί σε ένα ομότιμο δίκτυο. Κάθε συμμετέχων (κόμβος) στο δίκτυο διαθέτει ένα αντίγραφο ολόκληρης της αλυσίδας μπλοκ, διασφαλίζοντας ότι δεν υπάρχει ένα μόνο σημείο ελέγχου ή αποτυχίας. Αυτός ο αποκεντρωμένος χαρακτήρας ενισχύει την ασφάλεια και την ανθεκτικότητα.
-  Κατανεμημένος λογιστικός κατάλογος (Distributed Ledger): Η τεχνολογία Blockchain είναι ένα κατανεμημένο βιβλίο/κατάλογος που καταγράφει τις συναλλαγές με χρονολογική σειρά. Αυτές οι συναλλαγές ομαδοποιούνται σε «μπλοκ» και κάθε μπλοκ συνδέεται με το προηγούμενο, σχηματίζοντας μια αλυσίδα. Μόλις ένα μπλοκ προστεθεί στην αλυσίδα μπλοκ, δεν μπορεί να τροποποιηθεί, διασφαλίζοντας την ακεραιότητα των δεδομένων.
-  Διαφάνεια και αμεταβλητότητα: Οι συναλλαγές στην αλυσίδα μπλοκ είναι διαφανείς και αμετάβλητες. Οποιοσδήποτε στο δίκτυο μπορεί να δει ολόκληρο το ιστορικό των συναλλαγών, αλλά από τη στιγμή που μια συναλλαγή καταγράφεται, δεν μπορεί να τροποποιηθεί ή να διαγραφεί. Αυτή η διαφάνεια και η αμεταβλητότητα καθιστούν το blockchain ένα ασφαλές και αξιόπιστο σύστημα.
-  Κρυπτογραφική ασφάλεια: Το Blockchain βασίζεται σε προηγμένες κρυπτογραφικές τεχνικές για την ασφάλεια των δεδομένων. Κάθε συμμετέχων διαθέτει ένα ζεύγος κρυπτογραφικών κλειδιών: ένα δημόσιο κλειδί για την ταυτοποίηση και ένα ιδιωτικό κλειδί για τον έλεγχο

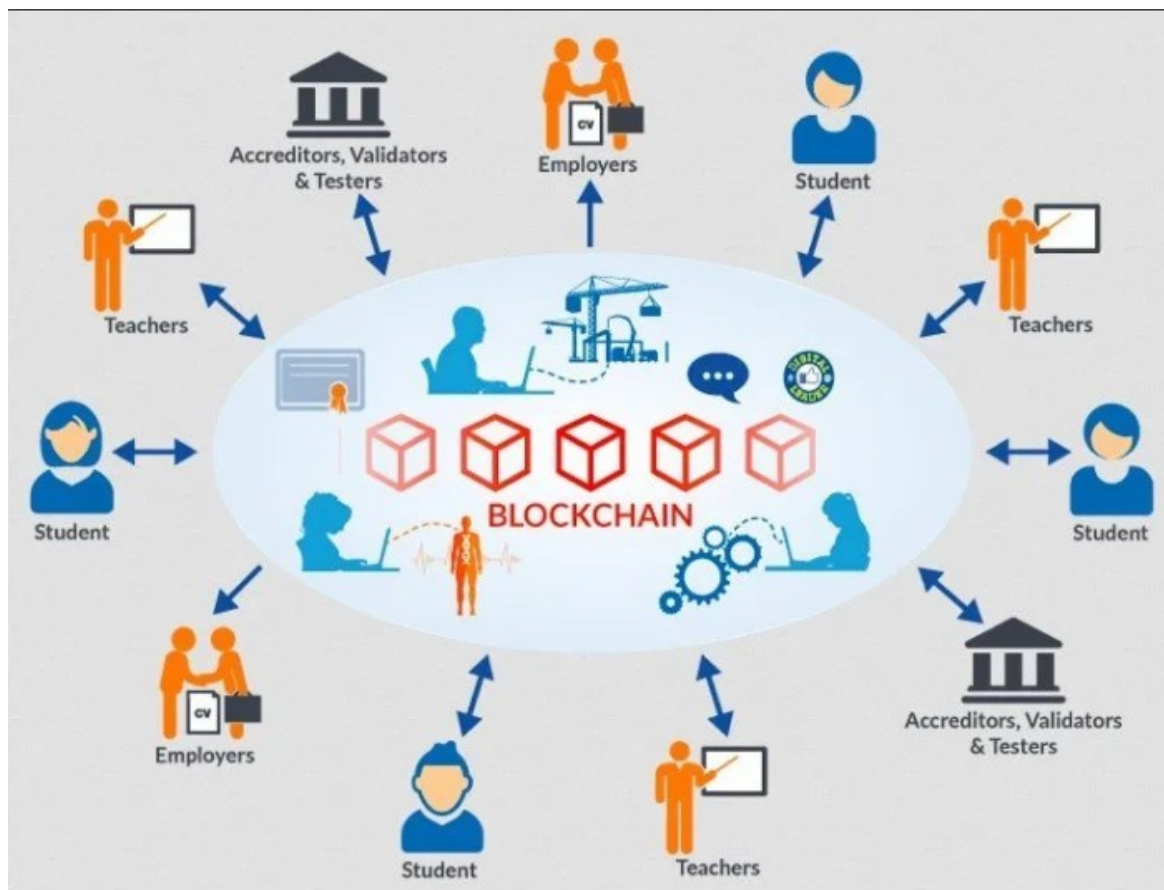
ταυτότητας και πρόσβασης. Αυτή η κρυπτογραφική ασφάλεια διασφαλίζει την ιδιωτικότητα και την ακεραιότητα των συναλλαγών.

✚ Μηχανισμοί συναίνεσης: Για να επικυρώσουν και να προσθέσουν νέες συναλλαγές στην αλυσίδα μπλοκ, οι συμμετέχοντες πρέπει να καταλήξουν σε συναίνεση. Διάφοροι μηχανισμοί συναίνεσης, όπως η απόδειξη εργασίας (Proof of Work - PoW) και η απόδειξη συμμετοχής (Proof of Stake - PoS), χρησιμοποιούνται για να διασφαλιστεί ότι οι συναλλαγές είναι νόμιμες και να αποτραπεί η χειραγώγηση της αλυσίδας μπλοκ από κακόβουλους φορείς.

✚ Έξυπνες συμβάσεις: Οι έξυπνες συμβάσεις είναι αυτοεκτελούμενες συμβάσεις με προκαθορισμένους κανόνες και προϋποθέσεις. Εκτελούνται αυτόματα όταν πληρούνται συγκεκριμένες προϋποθέσεις. Το Ethereum, μια πλατφόρμα blockchain, διέδωσε την έννοια των έξυπνων συμβολαίων, επιτρέποντας προγραμματιζόμενες και αποκεντρωμένες εφαρμογές (decentralized applications - DApps) [88], [89].

Η τεχνολογία Blockchain έχει βρει εφαρμογές σε πολλούς τομείς, όπως [88], [89]:

- ❖ Κρυπτονομίσματα: Υπάρχουν χιλιάδες κρυπτονομίσματα που έχουν κατασκευαστεί σε διάφορες πλατφόρμες blockchain.
- ❖ Διαχείριση της αλυσίδας εφοδιασμού: Το blockchain μπορεί να χρησιμοποιηθεί για την παρακολούθηση της προέλευσης των αγαθών, εξασφαλίζοντας διαφάνεια και αυθεντικότητα.
- ❖ Χρηματοοικονομικά: Το blockchain μπορεί να βελτιώσει τις διασυνοριακές πληρωμές, να μειώσει την απάτη και να επιτρέψει τον ταχύτερο διακανονισμό των χρηματοοικονομικών συναλλαγών.
- ❖ Υγειονομική περίθαλψη: Μπορεί να διασφαλίσει τα ιατρικά αρχεία, να ενισχύσει τη διαλειτουργικότητα και να επιτρέψει την ελεγχόμενη από τον ασθενή κοινή χρήση δεδομένων.
- ❖ Συστήματα ψηφοφορίας: Το blockchain μπορεί να παρέχει ασφαλή και διαφανή συστήματα ψηφοφορίας, μειώνοντας τον κίνδυνο απάτης.
- ❖ Επαλήθευση ταυτότητας: Το blockchain μπορεί να βελτιώσει την επαλήθευση της ταυτότητας και να μειώσει την κλοπή ταυτότητας.



Εικόνα 10: Τομείς εφαρμογής Blockchain¹⁵

Η τεχνολογία blockchain συνεχίζει να εξελίσσεται. Αναδύονται νέοι μηχανισμοί συναίνεσης, blockchain με επίκεντρο την προστασία της ιδιωτικής ζωής και επιχειρηματικές λύσεις με χρήση blockchain για την αντιμετώπιση συγκεκριμένων περιπτώσεων χρήσης και προκλήσεων. Το blockchain έχει τη δυνατότητα να φέρει επανάσταση στον τρόπο με τον οποίο πραγματοποιούμε συναλλαγές, αποθηκεύουμε δεδομένα και οικοδομούμε εμπιστοσύνη στα ψηφιακά οικοσυστήματα. Η αποκεντρωμένη και ασφαλής φύση του το καθιστά ανατρεπτική δύναμη σε διάφορους κλάδους και η εξέλιξή του συνεχίζει να διαμορφώνει το μέλλον της τεχνολογίας και των επιχειρήσεων [88], [89].

¹⁵ <https://www.indiamart.com/proddetail/blockchain-software-development-services-21571340412.html>

6. Συζήτηση και συμπεράσματα

6.1. Κύρια ευρήματα

Αρχικά όσον αφορά στην κυβερνοασφάλεια, η διερεύνηση της φώτισε το συνεχώς διευρυνόμενο φάσμα απειλών που αντιμετωπίζουν οι οργανισμοί και τα άτομα στο ψηφιακό πεδίο. Από τις διαδεδομένες αλλά επικύνδινες επιθέσεις phishing έως τις εξελιγμένες εκστρατείες κυβερνοκατασκοπείας που χρηματοδοτούνται από κράτη, τα ευρήματά μας υπογραμμίζουν τις ποικίλες τακτικές που χρησιμοποιούν οι κακόβουλοι φορείς. Ειδικότερα, η συνεχής εξέλιξη των απειλών στον κυβερνοχώρο απαιτεί μια προληπτική προσέγγιση της ασφάλειας στον κυβερνοχώρο. Η έρευνά μας επιβεβαίωσε ότι τα αποτελεσματικά μέτρα ασφαλείας πρέπει να επεκτείνονται πέρα από τα παραδοσιακά τείχη προστασίας και το λογισμικό προστασίας από ιούς. Οι οργανισμοί πρέπει να επενδύουν σε πληροφορίες σχετικά με τις απειλές, ανάλυση βάσει συμπεριφοράς και προσαρμοστικές στρατηγικές ασφάλειας για τον εντοπισμό και τον μετριασμό των αναδυόμενων απειλών σε πραγματικό χρόνο.

Επιπλέον, ο ανθρώπινος παράγοντας αναδεικνύεται σε καθοριστικό παράγοντα για την ανθεκτικότητα της ασφάλειας στον κυβερνοχώρο. Παρά τις πιο προηγμένες τεχνολογίες, η ανθρώπινη ευπάθεια παραμένει μια σημαντική πρόκληση. Σύμφωνα με τα ευρήματά μας ότι η ευαισθητοποίηση και η εκπαίδευση των χρηστών διαδραματίζουν κρίσιμο ρόλο στην ενίσχυση της άμυνας ενός οργανισμού. Επιπλέον στα πλαίσια της παρούσας διατριβής εξετάστηκε η αποτελεσματικότητα του ελέγχου ταυτότητας πολλαπλών παραγόντων (MFA) ως κρίσιμου επιπέδου ασφάλειας. Το MFA ενισχύει σημαντικά την ασφάλεια απαιτώντας πολλαπλές μορφές ελέγχου ταυτότητας, μειώνοντας τον κίνδυνο μη εξουσιοδοτημένης πρόσβασης. Το κύριο εύρημα σε αυτήν την ενότητα είναι η αποτελεσματικότητα του ελέγχου ταυτότητας πολλαπλών παραγόντων (MFA) ως κρίσιμου επιπέδου ασφάλειας. Το MFA ενισχύει σημαντικά τη στάση ασφαλείας απαιτώντας πολλαπλές μορφές ελέγχου ταυτότητας, μειώνοντας τον κίνδυνο μη εξουσιοδοτημένης πρόσβασης. Επιπρόσθετα, η έρευνά μας υπογραμμίζει τη σημασία της διαφύλαξης εμπιστευτικών δεδομένων καθώς οι παραβιάσεις δεδομένων μπορεί να έχουν σοβαρές συνέπειες τόσο για τα άτομα όσο και τους οργανισμούς. Η κρυπτογράφηση, οι έλεγχοι πρόσβασης και η ταξινόμηση δεδομένων είναι βασικά στοιχεία για τη διατήρηση του απορρήτου των δεδομένων. Η κύρια διαπίστωση μας είναι ότι η κρυπτογράφηση χρησιμεύει ως ισχυρή εγγύηση για τα δεδομένα, εξασφαλίζοντας την εμπιστευτικότητα και την ακεραιότητά τους κατά τη μετάδοση και την αποθήκευση. Διερευνήσαμε

τις αρχές της κρυπτογράφησης, συμπεριλαμβανομένης της συμμετρικής και της ασύμμετρης κρυπτογράφησης, ρίχνοντας φως στον τρόπο με τον οποίο λειτουργούν οι κρυπτογραφικοί αλγόριθμοι για την ασφάλεια των δεδομένων

Στον τομέα της τεχνητής νοημοσύνης, τα κύρια ευρήματα περιλαμβάνουν τη μετασχηματιστική δύναμη των τεχνολογιών τεχνητής νοημοσύνης, την ποικιλομορφία των εφαρμογών της, τα πλεονεκτήματα και την καθοριστική συμβολή της στη βελτίωση της αποτελεσματικότητας και της αυτοματοποίησης του σύγχρονου ψηφιακού κόσμου χωρίς να παραβλέπονται ωστόσο οι σχετικοί κίνδυνοι, ιδίως όσον αφορά την ιδιωτικότητα των δεδομένων και τις ηθικές ανησυχίες που εγείρονται γύρω από τον συγκεκριμένο τομέα.

Τα ευρήματα της έρευνας μας για την Defensive AI υπογραμμίζουν τον ύψιστης σημασίας ρόλο των νευρωνικών δικτύων, των Intelligent Agents και εν γένει της μηχανικής μάθησης στην ενίσχυση της άμυνας και της διασφάλισης της κυβερνοασφάλειας. Αυτές οι προσεγγίσεις που βασίζονται στην τεχνητή νοημοσύνη επιτρέπουν την προληπτική ανίχνευση και αντιμετώπιση των απειλών. Στον τομέα της Offensive AI, η έρευνά μας αναλύει το ρόλο της τεχνητής νοημοσύνης σε κυβερνοεπιθέσεις, τα τρωτά σημεία των συστημάτων τεχνητής νοημοσύνης αλλά και τις αναδυόμενες απειλές και επιθέσεις κατά της τεχνητής νοημοσύνης. Επιπλέον, διερευνήσαμε τη δράση της Offensive AI εναντίον οργανισμών, αποκαλύπτοντας πιθανά μοντέλα επίθεσης. Τέλος, όσον αφορά στην ανάγκη ανωνυμοποίησης των δεδομένων και στην τεχνολογία Blockchain, η παρούσα διατριβή τονίζει τη σημασία της ανωνυμοποίησης για τη διατήρηση του απορρήτου κατά την εξισορρόπηση της χρησιμότητας δεδομένων αλλά και τη σημασία της τεχνολογίας blockchain για την ενίσχυση της εμπιστοσύνης και της ασφάλειας σε διάφορους κλάδους. Η διαφύλαξη των απόρρητων δεδομένων κρίνεται ζωτικής σημασίας καθώς οι σχετικές παραβιάσεις μπορεί να έχουν σοβαρές συνέπειες τόσο για τα άτομα όσο και τους οργανισμούς ή ακόμα και τα κράτη. Η κρυπτογράφηση, οι έλεγχοι πρόσβασης και η ταξινόμηση δεδομένων είναι βασικά στοιχεία για τη διατήρηση του απορρήτου των δεδομένων.

6.2. Προκλήσεις και λύσεις

Η έννοια της κυβερνοασφάλειας αποτελεί μια δυναμική και συνεχώς εξελισσόμενη πρόκληση. Μία από τις κύριες προκλήσεις είναι το συνεχώς εξελισσόμενο τοπίο απειλών, με τους αντιπάλους να χρησιμοποιούν όλο και πιο περίπλοκες τακτικές στις κυβερνοεπιθέσεις τους. Μια λύση βρίσκεται στην προληπτική αντιμετώπιση των απειλών, όπου οι οργανισμοί συλλέγουν και αναλύουν δεδομένα

σε πραγματικό χρόνο για να προβλέψουν τις επερχόμενες απειλές. Τα προσαρμοστικά μέτρα ασφαλείας, όπως ο εντοπισμός ανωμαλιών με τη βοήθεια της μηχανικής μάθησης, μπορούν να παρέχουν την ευελιξία που απαιτείται για την αποτελεσματική αντιμετώπιση των αναδυόμενων απειλών. Επιπλέον, μια ανθρωποκεντρική προσέγγιση είναι απαραίτητη. Ισχυρά προγράμματα εκπαίδευσης χρηστών και ευαισθητοποίησης για την ασφάλεια δίνουν τη δυνατότητα στα άτομα να αναγνωρίζουν και να ανταποκρίνονται σε πιθανές απειλές, διαμορφώνοντας μια κρίσιμη γραμμή άμυνας.

Στους τομείς της τεχνολογίας, οι προκλήσεις και οι λύσεις είναι μοναδικές. Η ασφάλεια των διακομιστών Ιστού, για παράδειγμα, αντιμετωπίζει την επίμονη πρόκληση των ευάλωτων εφαρμογών Ιστού. Η λύση βρίσκεται σε ασφαλείς πρακτικές ανάπτυξης, τακτικούς ελέγχους ασφαλείας και αυστηρή διαχείριση ενημερώσεων κώδικα. Ομοίως, η υιοθέτηση της τεχνητής νοημοσύνης εισάγει ανησυχίες για την ιδιωτικότητα. Οι τεχνικές της τεχνητής νοημοσύνης που διατηρούν το απόρρητο, όπως είναι η τεχνική του διαφορικού απόρρητου και η ομοσπονδιακή μάθηση, προσφέρουν λύσεις προστατεύοντας ευαίσθητα δεδομένα κατά την εκπαίδευση μοντέλων τεχνητής νοημοσύνης.

Επιπρόσθετα, η εφαρμογή της τεχνητής νοημοσύνης εισάγει ηθικές και υπεύθυνες προκλήσεις ανάπτυξης. Οι επιθέσεις που βασίζονται στην τεχνητή νοημοσύνη εγείρουν ανησυχίες. Οι ισχυρές λύσεις συνεπάγονται την ενίσχυση των μηχανισμών ανίχνευσης του αντιπάλου με την παράλληλη τήρηση των πλαισίων ηθικής της τεχνητής νοημοσύνης. Οι υπεύθυνες πρακτικές ανάπτυξης τεχνητής νοημοσύνης είναι κρίσιμες για τον μετριασμό των κινδύνων. Η διαφανής ανάπτυξη των μοντέλων τεχνητής νοημοσύνης και η επεξήγηση αλγορίθμων είναι ζωτικής σημασίας πτυχές της ηθικής τεχνητής νοημοσύνης. Στον τομέα του blockchain, προκλήσεις όπως η επεκτασιμότητα και η κανονιστική συμμόρφωση απαιτούν προσοχή. Οι λύσεις επεκτασιμότητας όπως τα πρωτόκολλα διαμοιρασμού μπορούν να αντιμετωπίσουν την πρόκληση της επεκτασιμότητας, ενώ η δέσμευση με ρυθμιστικές αρχές, βιομηχανικά πρότυπα και πλαίσια συμμόρφωσης προσφέρει λύσεις για την κανονιστική συμμόρφωση.

Συμπερασματικά, οι προαναφερθείσες προκλήσεις και λύσεις υπογραμμίζουν την ανάγκη για μια ολοκληρωμένη και προσαρμοστική προσέγγιση στον τομέα της κυβερνοασφάλειας και της τεχνολογίας. Αντιμετωπίζοντας αυτές τις προκλήσεις προληπτικά και αξιοποιώντας τις καινοτόμες λύσεις, οι οργανισμοί και τα ενδιαφερόμενα μέρη της τεχνολογίας μπορούν να πλοηγηθούν πιο

αποτελεσματικά στην πολυπλοκότητα του ψηφιακού τοπίου, ενισχύοντας τόσο την ασφάλεια όσο και την εμπιστοσύνη.

Βιβλιογραφία-Παραπομπές

- [1] S. AlDaajeh, H. Saleous, S. Alrabaee, E. Barka, F. Breitingner, and K. K. Raymond Choo, "The role of national cybersecurity strategies on the improvement of cybersecurity education," *Comput. Secur.*, vol. 119, p. 102754, 2022, doi: 10.1016/j.cose.2022.102754.
- [2] S. Vyamajala, T. K. Mohd, and A. Javaid, "A Real-World Implementation of SQL Injection Attack Using Open Source Tools for Enhanced Cybersecurity Learning," *IEEE Int. Conf. Electro Inf. Technol.*, vol. 2018-May, pp. 198–202, Oct. 2018, doi: 10.1109/EIT.2018.8500136.
- [3] N. Howard and S. Kanareykin, "Intention awareness in cyber security," *Proc. 2012 Int. Conf. Cyber Secur. Cyber Warf. Digit. Forensic, CyberSec 2012*, pp. 6–11, 2012, doi: 10.1109/CYBERSEC.2012.6246079.
- [4] M. Ratana Bhalla and A. Vardhan Bhalla, "Generations of Mobile Wireless Technology: A Survey," *Int. J. Comput. Appl.*, vol. 5, no. 4, pp. 975–8887, 2010.
- [5] J. P. Shim, U. Varshney, S. Dekleva, and G. Knoerzer, "Mobile and wireless networks: Services, evolution and issues," *Int. J. Mob. Commun.*, vol. 4, no. 4, pp. 405–417, 2006, doi: 10.1504/IJMC.2006.008949.
- [6] G. Liu *et al.*, "Vision, requirements and network architecture of 6G mobile network beyond 2030," *China Commun.*, vol. 17, no. 9, pp. 92–104, Sep. 2020, doi: 10.23919/JCC.2020.09.008.
- [7] E. Rodriguez, B. Otero, N. Gutierrez, and R. Canal, "A survey of deep learning techniques for cybersecurity in mobile networks," *IEEE Commun. Surv. Tutorials*, vol. 23, no. 3, pp. 1920–1955, Jul. 2021, doi: 10.1109/COMST.2021.3086296.
- [8] M. Y. Shabir, A. Iqbal, Z. Mahmood, and A. Ghafoor, "Analysis of classical encryption techniques in cloud computing," *Tsinghua Sci. Technol.*, vol. 21, no. 1, pp. 102–113, Feb. 2016, doi: 10.1109/TST.2016.7399287.
- [9] A. Biryukov, C. De Cannière, A. Braeken, and B. Preneel, "A toolbox for cryptanalysis: Linear and affine equivalence algorithms," *Lect. Notes Comput. Sci. (including Subser. Lect. Notes Artif.*

- Intell. Lect. Notes Bioinformatics*), vol. 2656, pp. 33–50, 2003, doi: 10.1007/3-540-39200-9_3/COVER.
- [10] “Data Encryption Standard (DES) and Advanced Encryption Standard (AES),” *Encycl. Multimed.*, pp. 143–144, Jun. 2006, doi: 10.1007/0-387-30038-4_46.
 - [11] A. Alshamrani, S. Myneni, A. Chowdhary, and D. Huang, “A Survey on Advanced Persistent Threats: Techniques, Solutions, Challenges, and Research Opportunities,” *IEEE Commun. Surv. Tutorials*, vol. 21, no. 2, pp. 1851–1877, Apr. 2019, doi: 10.1109/COMST.2019.2891891.
 - [12] F. Mohammed Al-Matarneh, “ADVANCED PERSISTENT THREATS AND ITS ROLE IN NETWORK SECURITY VULNERABILITIES,” *Int. J. Adv. Res. Comput. Sci.*, vol. 11, no. 1, doi: 10.26483/ijarcs.v11i1.6502.
 - [13] I. Ghafir and V. Prenosil, “Proposed approach for targeted attacks detection,” *Lect. Notes Electr. Eng.*, vol. 362, pp. 73–80, 2016, doi: 10.1007/978-3-319-24584-3_7/FIGURES/2.
 - [14] M. Droppa and M. Harakal, “Analysis of Cybersecurity in the Real Environment,” *2021 Commun. Inf. Technol. Conf. Proceedings, KIT 2021 - 11th Int. Sci. Conf.*, Oct. 2021, doi: 10.1109/KIT52904.2021.9583748.
 - [15] A. Ometov, S. Bezzateev, N. Mäkitalo, S. Andreev, T. Mikkonen, and Y. Koucheryavy, “Multi-Factor Authentication: A Survey,” *Cryptogr. 2018, Vol. 2, Page 1*, vol. 2, no. 1, p. 1, Jan. 2018, doi: 10.3390/CRYPTOGRAPHY2010001.
 - [16] A. Ometov, V. Petrov, S. Bezzateev, S. Andreev, Y. Koucheryavy, and M. Gerla, “Challenges of Multi-Factor Authentication for Securing Advanced IoT Applications,” *IEEE Netw.*, vol. 33, no. 2, pp. 82–88, Mar. 2019, doi: 10.1109/MNET.2019.1800240.
 - [17] S. M. Diesburg and A. I. A. Wang, “A survey of confidential data storage and deletion methods,” *ACM Comput. Surv.*, vol. 43, no. 1, Dec. 2010, doi: 10.1145/1824795.1824797.
 - [18] D. Clark, S. Hunt, and P. Malacaria, “Quantitative Analysis of the Leakage of Confidential Data,” *Electron. Notes Theor. Comput. Sci.*, vol. 59, no. 3, pp. 238–251, Nov. 2002, doi: 10.1016/S1571-0661(04)00290-7.

- [19] W. H. Wolberg, W. N. Street, and O. L. Mangasarian, "Machine learning techniques to diagnose breast cancer from image-processed nuclear features of fine needle aspirates," *Cancer Lett.*, 1994, doi: 10.1016/0304-3835(94)90099-X.
- [20] W. H. Wolberg, W. N. Street, and O. L. Mangasarian, "Image analysis and machine learning applied to breast cancer diagnosis and prognosis," *Anal. Quant. Cytol. Histol.*, 1995.
- [21] S. Haykin, *Neural Networks and Learning Machines*, 3rd ed. Pearson, 2009.
- [22] S. Theodoridis and K. Koutroumbas, *Αναγνώριση Προτύπων*, 4th ed. ΕΚΔΟΣΕΙΣ Π.Χ. ΠΑΣΧΑΛΙΔΗΣ - BROKEN HILL PUBLISHERS LTD, 2009.
- [23] Z. Zhang, "Introduction to machine learning: K-nearest neighbors," *Ann. Transl. Med.*, vol. 4, no. 11, 2016, doi: 10.21037/atm.2016.03.37.
- [24] O. Kramer, "K-Nearest Neighbors," Springer, Berlin, Heidelberg, 2013, pp. 13–23.
- [25] N. S. Altman, "An introduction to kernel and nearest-neighbor nonparametric regression," *Am. Stat.*, 1992, doi: 10.1080/00031305.1992.10475879.
- [26] M. N. Elgedawy, "Prediction of Breast Cancer using Random Forest, Support Vector Machines and Naïve Bayes," *Int. J. Eng. Comput. Sci.*, Jan. 2017, doi: 10.18535/ijecs/v6i1.07.
- [27] A. F. Agarap, "On Breast Cancer Detection: An Application of Machine Learning Algorithms on the Wisconsin Diagnostic Dataset," *ACM Int. Conf. Proceeding Ser.*, pp. 5–9, Nov. 2017, doi: 10.1145/3184066.3184080.
- [28] Q. Pan, Y. Zhang, D. Chen, and G. Xu, "Character-based convolutional grid neural network for breast cancer classification," 2017, doi: 10.1109/ICGI.2017.31.
- [29] B. R. Andjelkovic Cirkovic, "Machine learning approach for breast cancer prognosis prediction," in *Computational Modeling in Bioengineering and Bioinformatics*, 2020.
- [30] V. Chaurasia, S. Pal, and B. Tiwari, "Prediction of benign and malignant breast cancer using data mining techniques," *J. Algorithm. Comput. Technol.*, vol. 12, no. 2, pp. 119–126, Jun. 2018, doi: 10.1177/1748301818756225.

- [31] M. Tahmooresi, A. Afshar, B. B. Rad, K. B. Nowshath, and M. A. Bamiah, "Early Detection of Breast Cancer Using Machine Learning Techniques," no. September, 2018.
- [32] J. D. M. Rennie, L. Shih, J. Teevan, and D. Karger, "Tackling the Poor Assumptions of Naive Bayes Text Classifiers," 2003.
- [33] H. Zhang, "The Optimality of Naive Bayes," Accessed: Oct. 18, 2021. [Online]. Available: www.aaai.org.
- [34] T. Anagnostoua, M. Remzi, M. Lykourinas, and B. Djavan, "Artificial neural networks for decision-making in urologic oncology," *European Urology*. 2003, doi: 10.1016/s0302-2838(03)00133-7.
- [35] G. P. Zhang, "Neural networks for classification: a survey," *IEEE Trans. Syst. Man, Cybern. Part C Appl. Rev.*, 2000.
- [36] R. R. Janghel, A. Shukla, R. Tiwari, and R. Kala, "Breast Cancer diagnosis using Artificial Neural Network models," 2010, doi: 10.1109/ICICIS.2010.5534716.
- [37] D. A. Omondiagbe, S. Veeramani, and A. S. Sidhu, "Machine Learning Classification Techniques for Breast Cancer Diagnosis," *IOP Conf. Ser. Mater. Sci. Eng.*, vol. 495, no. 1, p. 012033, Apr. 2019, doi: 10.1088/1757-899X/495/1/012033.
- [38] Y.-J. Tseng *et al.*, "Predicting breast cancer metastasis by using serum biomarkers and clinicopathological data with machine learning technologies," *Int. J. Med. Inform.*, vol. 128, pp. 79–86, Aug. 2019, doi: 10.1016/j.ijmedinf.2019.05.003.
- [39] D. W. Hosmer, S. Lemeshow, and R. X. Sturdivant, *Applied Logistic Regression: Third Edition*. 2013.
- [40] I. Kononenko and M. Kukar, *Machine learning and data mining*. 2007.
- [41] L. Rokach and O. Maimon, "Chapter 15— Clustering methods," *Data Min. Knowl. Discov. Handb.*, 2010, doi: 10.1007/0-387-25465-X_15.
- [42] M. Montazeri, M. Montazeri, M. Montazeri, and A. Beigzadeh, "Machine learning models in breast cancer survival prediction," *Technol. Heal. Care*, vol. 24, no. 1, pp. 31–42, Jan. 2016, doi:

10.3233/THC-151071.

- [43] H. Asri, H. Mousannif, H. Al Moatassime, and T. Noel, "Using Machine Learning Algorithms for Breast Cancer Risk Prediction and Diagnosis," 2016, doi: 10.1016/j.procs.2016.04.224.
- [44] R. D. Nindrea, T. Aryandono, L. Lazuardi, and I. Dwiprahasto, "Diagnostic accuracy of different machine learning algorithms for breast cancer risk calculation: A meta-analysis," *Asian Pacific Journal of Cancer Prevention*. 2018, doi: 10.22034/APJCP.2018.19.7.1747.
- [45] A. LG and E. AT, "Using Three Machine Learning Techniques for Predicting Breast Cancer Recurrence," *J. Heal. Med. Informatics*, 2013, doi: 10.4172/2157-7420.1000124.
- [46] K. Polat and S. Güneş, "Breast cancer diagnosis using least square support vector machine," *Digit. Signal Process. A Rev. J.*, 2007, doi: 10.1016/j.dsp.2006.10.008.
- [47] B. E. Bejnordi *et al.*, "Diagnostic assessment of deep learning algorithms for detection of lymph node metastases in women with breast cancer," *JAMA - J. Am. Med. Assoc.*, vol. 318, no. 22, pp. 2199–2210, Dec. 2017, doi: 10.1001/jama.2017.14585.
- [48] D. Pevec, E. Štrumbelj, and I. Kononenko, "Evaluating reliability of single classifications of neural networks," 2011, doi: 10.1007/978-3-642-20282-7_3.
- [49] A. Krizhevsky, I. Sutskever, and G. E. Hinton, "ImageNet Classification with Deep Convolutional Neural Networks," Accessed: Sep. 28, 2023. [Online]. Available: <http://code.google.com/p/cuda-convnet/>.
- [50] J. Devlin, M. W. Chang, K. Lee, and K. Toutanova, "BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding," *NAACL HLT 2019 - 2019 Conf. North Am. Chapter Assoc. Comput. Linguist. Hum. Lang. Technol. - Proc. Conf.*, vol. 1, pp. 4171–4186, Oct. 2018, Accessed: Sep. 28, 2023. [Online]. Available: <https://arxiv.org/abs/1810.04805v2>.
- [51] K. He, X. Zhang, S. Ren, and J. Sun, "Deep Residual Learning for Image Recognition," *Proc. IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognit.*, vol. 2016-December, pp. 770–778, Dec. 2015, doi: 10.1109/CVPR.2016.90.
- [52] M. M. Mijwil, I. E. Salem, and M. M. Ismaeel, "The Significance of Machine Learning and Deep

Learning Techniques in Cybersecurity: A Comprehensive Review,” doi:
10.52866/ijcsm.2023.01.01.008.

- [53] R. Dale, “Natural language generation: The commercial state of the art in 2020,” *Nat. Lang. Eng.*, vol. 26, no. 4, pp. 481–487, Jul. 2020, doi: 10.1017/S135132492000025X.
- [54] R. Evans and E. Grefenstette, “Survey of the State of the Art in Natural Language Generation: Core tasks, applications and evaluation,” *J. Artif. Intell. Res.*, vol. 61, pp. 65–170, Jan. 2018, doi: 10.1613/JAIR.5477.
- [55] D. Wang, X. Wang, and S. Lv, “An Overview of End-to-End Automatic Speech Recognition,” *Symmetry 2019, Vol. 11, Page 1018*, vol. 11, no. 8, p. 1018, Aug. 2019, doi: 10.3390/SYM11081018.
- [56] “Guide to Biometrics - Ruud M. Bolle, Jonathan H. Connell, Sharath Pankanti, Nalini K. Ratha, Andrew W. Senior - Βιβλία Google.”
https://books.google.gr/books?hl=el&lr=&id=DLLbBwAAQBAJ&oi=fnd&pg=PR13&dq=biometrics&ots=K6wWdPSKcP&sig=UiY-EegbLkPZD2WMp0WvK1DH4r8&redir_esc=y#v=onepage&q=biometrics&f=false (accessed Sep. 28, 2023).
- [57] A. K. Jain and A. Kumar, “Biometrics of Next Generation: An Overview,” 2010.
- [58] J. Wickins, “The ethics of biometrics: The risk of social exclusion from the widespread use of electronic identification,” *Sci. Eng. Ethics*, vol. 13, no. 1, pp. 45–54, Mar. 2007, doi: 10.1007/S11948-007-9003-Z/METRICS.
- [59] M. M. Mariani, I. Machado, and S. Nambisan, “Types of innovation and artificial intelligence: A systematic quantitative literature review and research agenda,” *J. Bus. Res.*, vol. 155, p. 113364, Jan. 2023, doi: 10.1016/J.JBUSRES.2022.113364.
- [60] A. Kaplan and M. Haenlein, “Siri, Siri, in my hand: Who’s the fairest in the land? On the interpretations, illustrations, and implications of artificial intelligence,” *Bus. Horiz.*, vol. 62, no. 1, pp. 15–25, Jan. 2019, doi: 10.1016/J.BUSHOR.2018.08.004.
- [61] B. Goertzel, “Artificial General Intelligence: Concept, State of the Art, and Future Prospects,” *J.*

Artif. Gen. Intell., vol. 5, no. 1, pp. 2013–2015, 2014, doi: 10.2478/jagi-2014-0001.

- [62] S. Pueyo, “Growth, degrowth, and the challenge of artificial superintelligence,” *J. Clean. Prod.*, vol. 197, pp. 1731–1736, Oct. 2018, doi: 10.1016/J.JCLEPRO.2016.12.138.
- [63] A. Kar, A. Subash, and V. U. S. Rao, “Reactive Artificial Intelligence Using Big Data in the Era of Precision Medicine,” *JAMA Surg.*, vol. 155, no. 7, pp. 671–671, Jul. 2020, doi: 10.1001/JAMASURG.2020.0839.
- [64] G. Molas and E. Nowak, “Advances in Emerging Memory Technologies: From Data Storage to Artificial Intelligence,” *Appl. Sci. 2021, Vol. 11, Page 11254*, vol. 11, no. 23, p. 11254, Nov. 2021, doi: 10.3390/APP112311254.
- [65] F. Cuzzolin, A. Morelli, B. Cîrstea, and B. J. Sahakian, “Knowing me, knowing you: theory of mind in AI,” *Psychol. Med.*, vol. 50, no. 7, pp. 1057–1061, May 2020, doi: 10.1017/S0033291720000835.
- [66] J. Cámara *et al.*, “Self-aware computing systems: Related concepts and research areas,” *Self-Aware Comput. Syst.*, pp. 17–49, Jan. 2017, doi: 10.1007/978-3-319-47474-8_2/COVER.
- [67] J. Borana, “Applications of Artificial Intelligence & Associated Technologies,” 2016.
- [68] K. Chhaya, A. Khanzode, and R. D. Sarode, “ADVANTAGES AND DISADVANTAGES OF ARTIFICIAL INTELLIGENCE AND MACHINE LEARNING: A LITERATURE REVIEW,” pp. 9–10, Accessed: Sep. 29, 2023. [Online]. Available: <http://www.iaeme.com/IJLIS/index.asp30http://www.iaeme.com/IJLIS/issues.asp?JType=IJLIS&VType=9&IType=1JournalImpactFactor>.
- [69] M. Nadimpalli, “Artificial Intelligence Risks and Benefits Big Data & Analytics View project Artificial Intelligence View project Artificial Intelligence Risks and Benefits,” *Int. J. Innov. Res. Sci. Eng. Technol. (An ISO)*, vol. 3297, 2007, Accessed: Sep. 29, 2023. [Online]. Available: <https://www.researchgate.net/publication/319321806>.
- [70] K. Manheim and L. Kaplan, “Artificial Intelligence: Risks to Privacy and Democracy,” *Yale J. Law Technol.*, vol. 21, 2019, Accessed: Sep. 29, 2023. [Online]. Available: <https://heinonline.org/HOL/Page?handle=hein.journals/yjolt21&id=106&div=&collection=>.

- [71] B. Cheatham, K. Javanmardian, and H. Samandari, "Confronting the risks of artificial intelligence With great power comes great responsibility. Organizations can mitigate the risks of applying artificial intelligence and advanced analytics by embracing three principles," 2019.
- [72] K. Kaushik, S. Dahiya, K. Kaushik, and · S Dahiya, "An Artificial Intelligence Assisted Defensive Framework for Securing Cyberspace," *Algorithms Intell. Syst.*, pp. 323–333, 2022, doi: 10.1007/978-981-19-1657-1_28.
- [73] J. B. Michael and T. C. Wingfield, "Defensive AI: The Future Is Yesterday," *Computer (Long. Beach. Calif.)*, vol. 54, no. 9, pp. 90–96, Sep. 2021, doi: 10.1109/MC.2021.3092480.
- [74] X. Zhang, X. Zheng, and W. Mao, "Adversarial Perturbation Defense on Deep Neural Networks," *ACM Comput. Surv.*, vol. 54, no. 8, Oct. 2021, doi: 10.1145/3465397.
- [75] R. A. Khalil, E. Jones, M. I. Babar, T. Jan, M. H. Zafar, and T. Alhussain, "Speech Emotion Recognition Using Deep Learning Techniques: A Review," *IEEE Access*, vol. 7, pp. 117327–117345, 2019, doi: 10.1109/ACCESS.2019.2936124.
- [76] C. H. Heinl, "Artificial (intelligent) agents and active cyber defence: Policy implications," *Int. Conf. Cyber Conflict, CYCON*, vol. 2014, pp. 53–66, Oct. 2014, doi: 10.1109/CYCON.2014.6916395.
- [77] J. Gabirondo-Lopez, J. Egana, J. Miguel-Alonso, and R. Orduna Urrutia, "Towards Autonomous Defense of SDN Networks Using MuZero Based Intelligent Agents," *IEEE Access*, vol. 9, pp. 107184–107199, 2021, doi: 10.1109/ACCESS.2021.3100706.
- [78] S. Qiu, Q. Liu, S. Zhou, and C. Wu, "Review of Artificial Intelligence Adversarial Attack and Defense Technologies," *Appl. Sci. 2019, Vol. 9, Page 909*, vol. 9, no. 5, p. 909, Mar. 2019, doi: 10.3390/APP9050909.
- [79] W. Jiang, Z. He, J. Zhan, and W. Pan, "Attack-Aware Detection and Defense to Resist Adversarial Examples," *IEEE Trans. Comput. Des. Integr. Circuits Syst.*, vol. 40, no. 10, pp. 2194–2198, Oct. 2021, doi: 10.1109/TCAD.2020.3033746.
- [80] L. Chongrui *et al.*, "Artificial Intelligence in Cyber Security," *J. Phys. Conf. Ser.*, vol. 1964, no. 4, p. 042072, Jul. 2021, doi: 10.1088/1742-6596/1964/4/042072.

- [81] M. François, P. E. Arduin, and M. Merad, "Artificial Intelligence & Cybersecurity: A Preliminary Study of Automated Pentesting with Offensive Artificial Intelligence," *Lect. Notes Bus. Inf. Process.*, vol. 425 LNBIP, pp. 131–138, 2021, doi: 10.1007/978-3-030-85977-0_10/COVER.
- [82] S. Kaviani, K. J. Han, and I. Sohn, "Adversarial attacks and defenses on AI in medical imaging informatics: A survey," *Expert Syst. Appl.*, vol. 198, p. 116815, Jul. 2022, doi: 10.1016/J.ESWA.2022.116815.
- [83] M. Oreyomi and H. Jahankhani, "Challenges and Opportunities of Autonomous Cyber Defence (ACyD) Against Cyber Attacks," *Adv. Sci. Technol. Secur. Appl.*, pp. 239–269, 2022, doi: 10.1007/978-3-030-98225-6_9/FIGURES/8.
- [84] M. M. Rahman, A. Siddika Arshi, M. M. Hasan, S. Farzana Mishu, H. Shahriar, and F. Wu, "Security Risk and Attacks in AI: A Survey of Security and Privacy," *Proc. - Int. Comput. Softw. Appl. Conf.*, vol. 2023-June, pp. 1834–1839, 2023, doi: 10.1109/COMPSAC57700.2023.00284.
- [85] Y. Mirsky *et al.*, "The Threat of Offensive AI to Organizations," *Comput. Secur.*, vol. 124, p. 103006, Jan. 2023, doi: 10.1016/J.COSE.2022.103006.
- [86] "The Complete Book of Data Anonymization: From Planning to Implementation - Balaji Raghunathan - Βιβλία Google."
https://books.google.gr/books?hl=el&lr=&id=RF8WQD6m530C&oi=fnd&pg=PP1&dq=data+anonymization&ots=0tFZ6-K0FI&sig=_bSa6d1RHe46CUPhzZpf-LFiKXM&redir_esc=y#v=onepage&q=data+anonymization&f=false (accessed Oct. 01, 2023).
- [87] A. Majeed and S. Lee, "Anonymization Techniques for Privacy Preserving Data Publishing: A Comprehensive Survey," *IEEE Access*, vol. 9, pp. 8512–8545, 2021, doi: 10.1109/ACCESS.2020.3045700.
- [88] T. Huynh-The *et al.*, "Blockchain for the Metaverse: A Review," *Futur. Gener. Comput. Syst.*, vol. 143, pp. 401–419, Mar. 2022, doi: 10.1016/j.future.2023.02.008.
- [89] M. Javaid, A. Haleem, R. Pratap Singh, S. Khan, and R. Suman, "Blockchain technology applications for Industry 4.0: A literature-based review," *Blockchain Res. Appl.*, vol. 2, no. 4, p. 100027, Dec. 2021, doi: 10.1016/J.BCRA.2021.100027.

