Πολυτεχνική Σχολή

Τμήμα Μηχανικών Η/Υ & Πληροφορικής

Διπλωματική Εργασία

Εφαρμογή τεχνικών Μηχανικής Μάθησης για αναγνώριση κακόβουλων χρηστών σε δίκτυα υπολογιστών

Αλεξόπουλος Άγγελος

ΑΜ: 235713 (παλαιός) 1041439 (νέος)

Επιβλέπων

Καθηγητής Χρήστος Μπούρας

Μέλος Επιτροπής Αξιολόγησης

Καθηγητής Ιωάννης Γαροφαλάκης

Αναπληρωτής Καθηγητής Εμμανουήλ Ψαράκης

Πάτρα, Ιανουάριος 2022

ΠΑΝΕΠΙΣΤΗΜΙΟ ΠΑΤΡΩΝ
UNIVERSITY OF PATRAS

Η έγκριση της διπλωματικής εργασίας από το Τμήμα Μηχανικών Ηλεκτρονικών Υπολογιστών & Πληροφορικής του Πανεπιστημίου Πατρών, τον Επιβλέποντα ή της επιτροπής που την ενέκρινε, δεν υποδηλώνει απαραιτήτως και αποδοχή των απόψεων του συγγραφέα εκ μέρους του Τμήματος.

Υπεύθυνη Δήλωση

Δηλώνω ότι είμαι συγγραφέας αυτής της πτυχιακής εργασίας, και ότι κάθε βοήθεια είναι πλήρως αναγνωρισμένη, όπως και κάθε πηγή από όπου αντλήθηκε κάποια ιδέα και δεδομένα, και γίνεται αναφορά των παραπάνω στην πτυχιακή εργασία

(Υπογραφή)

.....

Αλεξόπουλος Άγγελος

<Η επιτυχία δεν είναι τελική, η αποτυχία δεν είναι μοιραία: είναι το θάρρος να συνεχίσουμε αυτό που μετράει>

<Winston Churchill>

Περίληψη

Τα σύγχρονα δίκτυα, αποτελούνται από αρκετές διαφορετικές συσκευές, που ποικίλουν από smartphones έως και ψυγεία, με διαφορετικά χαρακτηριστικά και δυνατότητες με σκοπό την ικανοποίηση των αναγκών μας. Αυτό το επιτυγχάνουν μέσω της ομαλής συνεργασίας τους μέσα στο δίκτυο. Βέβαια αυτό δε σημαίνει ότι τα δίκτυα δεν μπορούν να πέσουν θύμα κάποιας κακόβουλης επίθεσης μέσα από τους χρήστες του δικτύου ή ακόμα και εξωτερικούς. Στόχος, πέρα από την ομαλή λειτουργία του δικτύου, είναι και η ασφάλεια του δικτύου αλλά και η ασφάλεια των ίδιων των χρηστών. Η Μηχανική Μάθηση αξιοποιείται σε πολλούς τομείς, τόσο στα δίκτυα υπολογιστών όσο και σε άλλους επιστημονικούς τομείς, και η χρήση της μέσα στα δίκτυα αναμένεται να ενισχύσει την απόδοση των δικτύων αλλά και τη ποιότητα εμπειρίας μέσα σε αυτά ενισχύοντας την ασφάλεια των δικτύων.

Στόχος αυτής της διπλωματικής εργασίας είναι η ενσωμάτωση της Μηχανικής Μάθησης στα δίκτυα, με σκοπό την εύρεση κακόβουλων χρηστών μέσα σε αυτά. Κάνοντας αυτήν την ενσωμάτωση αναμένουμε ένα ασφαλέστερο δίκτυο και κατ' επέκταση καλύτερη εμπειρία μέσα σε αυτό.

Λέξεις Κλειδιά

Μηχανική Μάθηση, Δίκτυα, Ασφάλεια

Abstract

Modern networks consist of different devices, ranging from smartphones even to refrigerators, with different features and capabilities in order to meet our needs. They achieve this through smooth collaboration within the network. Of course, this does not mean that networks cannot fall victim to a malicious attack from network users or even outsiders. The goal, in addition to the smooth operation of the network, is the security of the network and the security of the users themselves. Machine Learning is utilized in many fields, equally in computer networks as in other scientific fields, and its use within networks is expected to enhance the efficiency of said networks and the quality of experience in them by enhancing network security.

The aim of this diploma thesis is to integrate Machine Learning into networks, in order to find malicious users within them. By doing this integration we expect a more secure network and consequently a better experience within it.

Key Words

Machine Learning, Networks, Security

Ευχαριστίες

Θα ήθελα να ευχαριστήσω τον καθηγητή κ. Χρήστο Ι. Μπούρα και τον Βασίλειο Κόκκινο για την επίβλεψη αυτής της διπλωματικής εργασίας. Είμαι ιδιαίτερα ευγνώμων για την καθοδήγηση και την εξαιρετική συνεργασία που είχαμε. Τέλος θα ήθελα να ευχαριστήσω συγγενείς και φίλους που βρισκόντουσαν στο πλευρό μου όλα αυτά τα χρόνια και στις καλές και στις δύσκολες στιγμές, και με καθοδήγησαν στο να μπορώ να αποτυπώσω τις ιδέες μου όσο πιο επιστημονικώς ορθά γίνεται.

Περιεχόμενα

Κατάλογος Εικόνων

Κεφάλαιο 1

Εισαγωγή

1.1 Εισαγωγή Κεφαλαίου

Στο κεφάλαιο αυτό θα κάνουμε μία αναφορά στις ανάγκες των ανθρώπων για πρόοδο από την αρχαιότητα, που οδήγησε στη δημιουργία των δικτύων, και όχι απαραίτητα τα δίκτυα επικοινωνιών μόνο. Μαζί όμως με την ανάπτυξη των δικτύων προέκυψε και η ανάγκη για ασφάλεια μέσα σε αυτά. Επιπλέον θα κάνουμε και μία αναφορά και στη Μηχανική Μάθηση καθώς μπορεί να συνεισφέρει σημαντικά στην πιο αποδοτική λειτουργία των δικτύων.

Επιπρόσθετα θα μιλήσουμε και για τους στόχους αυτής της διπλωματικής εργασίας καθώς συνδυάζει δύο αρκετά σημαντικά θέματα, την Μηχανική Μάθηση και τα δίκτυα υπολογιστών, που αποτελούν το επίκεντρο πολλών επιστημονικών ερευνών.

Τέλος θα αναλύσουμε και τη διάρθρωση αυτής της διπλωματικής, δίνοντας μία πρόχειρη περιγραφή για το τι μπορεί να περιμένει κανείς σε κάθε ένα από τα κεφάλαια που ακολουθούν.

1.2 Δημιουργία Δικτύων και Μηχανική Μάθηση

Οι άνθρωποι μπορούν να χαρακτηριστούν ως ένα συλλογικό ον, και αυτό με τη σειρά του σημαίνει πως έχει την ανάγκη να επικοινωνήσει και να συναναστραφεί με άλλους ανθρώπους. Η ανάγκη αυτή οδήγησε στην ανάπτυξη των δικτύων. Με αυτό τον τρόπο έγινε εφικτή και η ανάπτυξη της κοινωνίας. Οι άνθρωποι, από τα παλιά τα χρόνια μέχρι και σήμερα δημιουργούν δίκτυα, που ποικίλουν από δίκτυα επικοινωνιών έως και δίκτυα εμπορίου, για να καλύψουν τις ανάγκες τους.

Τη σήμερον ημέρα όταν μιλάμε για δίκτυα η πρώτη σκέψη των ανθρώπων είναι τα δίκτυα επικοινωνιών. Ο λόγος είναι καθαρά στο ότι ο άνθρωπος ως συλλογικό ον, έδωσε μεγαλύτερη σημασία στα δίκτυα επικοινωνιών με σκοπό να μπορεί ο κάθε άνθρωπος να επικοινωνήσει με την οικογένειά του, τους φίλους του κ.ο.κ σε όποια γωνία του πλανήτη

και αν βρίσκεται, μία δυσκολία που υπήρξε αρκετά μεγάλο πρόβλημα στα παλαιότερα χρόνια σε αντίθεση με τη σήμερον ημέρα.

Πως όμως μπορούμε να ορίσουμε τα δίκτυα; Σύμφωνα με τον ορισμό του, ένα δίκτυο είναι ένα πολύπλοκο σύμπλεγμα από γραμμές ή αγωγούς που διασταυρώνονται με τρόπο που μοιάζει με δίχτυ, όπως για παράδειγμα το οδικό δίκτυο ή το ηλεκτρικό δίκτυο. Ο συγκεκριμένος ορισμός ισχύει απόλυτα και στα δίκτυα επικοινωνιών και των υπολογιστών. Όπως είπαμε και προηγουμένως, τα δίκτυα επικοινωνιών εξελίχθηκαν σε τέτοιο βαθμό που πλέον είναι και ασύρματα.

Τι διαφορά έχουν λοιπόν τα ασύρματα δίκτυα από τα κανονικά; Ουσιαστικά στα ασύρματα δίκτυα η μετάδοση των δεδομένων μεταξύ των κόμβων και του δικτύου, γίνεται, όπως είναι ξεκάθαρο από το όνομα, ασύρματα. Συγκεκριμένα ως φορείς πληροφορίας χρησιμοποιούνται τα ραδιοκύματα, και τα δεδομένα μεταφέρονται μέσω ηλεκτρομαγνητικών κυμάτων, με συχνότητα φέροντος εξαρτώμενη από τον ρυθμό μετάδοσης δεδομένων που καθορίζεται από το δίκτυο. Οι κόμβοι που αναφέραμε, χωρίζονται στους αποστολείς δεδομένων και στους παραλήπτες δεδομένων, και με αυτή τη λογική δημιουργήθηκε ο προκάτοχος των κινητών δικτύων, που αποτελούνταν από μία κεραία και υπήρχαν και οι χρήστες που εξυπηρετούνταν μέσω των διαθέσιμων καναλιών. Σαφώς με την αύξηση των χρηστών έπρεπε να γίνει και προσαρμογή του συστήματος για την εξυπηρέτηση των χρηστών.

Η προσαρμογή που καταλήξαμε είναι η κυψελωτή επικοινωνία, στην οποία οι περιοχές χωρίζονται και σε κάθε μία υπάρχει μία κεραία που εξυπηρετεί όλους τους χρήστες στην εκάστοτε περιοχή. Το όνομα κυψελωτή επικοινωνία προκύπτει από την ομοιότητα που έχει το δίκτυο με τις κυψέλες των μελισσών στη φύση. Όπως και με τον προκάτοχο τους, τα κυψελωτά συστήματα συνεχώς εξελίσσονται και προσαρμόζονται σε κάθε γενιά δικτύων που δημιουργείται, με στόχο την καλύτερη εμπειρία μέσα στο δίκτυο και τις ταχύτερες μεταφορές δεδομένων, που τη σήμερον ημέρα είναι συγκλονιστικές σε αντίθεση με αυτές του παρελθόντος.

Με αυτές όμως τις εξελίξεις, δημιουργείται και η ανάγκη της ασφάλειας του δικτύου, όπως και των χρηστών αυτού. Σε μερικούς ανθρώπους υπερισχύει η ανάγκη να κατέχουν όλο και περισσότερα ακόμα και αν με αυτόν τον τρόπο επηρεάζουν αρνητικά τους υπόλοιπους. Αντίστοιχα στα δίκτυα, ένας χρήστης, για παράδειγμα, μπορεί να θέλει να μάθει προσωπικές πληροφορίες για τους υπόλοιπους χρήστες της περιοχής του. Αυτό δεν μπορεί να είναι αποδεκτό, καθώς αποτελεί παραβίαση προσωπικών δεδομένων του κάθε χρήστη. Άρα πρέπει να αναπτύσσονται τεχνικές που να προστατεύουν τους χρήστες μέσα στο δίκτυο κάνοντας καλύτερη έτσι, την εμπειρία μέσα στο δίκτυο.

Η Μηχανική Μάθηση αξιοποιείται σε πάρα πολλά πεδία και επιστημονικούς κλάδους, και έχει χαρακτηριστεί ως μία από τις πιο αποδοτικές προσθήκες σε αυτά. Όσο αναφορά την ασφάλεια στα δίκτυα, η Μηχανική Μάθηση μπορεί να προσαρμοστεί με τέτοιο

τρόπο ώστε να μπορεί να αναγνωρίζει κακόβουλους χρήστες, όπως θα έκανε ένας άνθρωπος, μέσω εκπαιδευτικών δεδομένων έτσι ώστε να ξέρει ακριβώς τι να ψάξει, κάνοντας έτσι αποτελεσματικά τη δουλειά που απαιτείται, και πιο αποδοτικά από τους ανθρώπους, καθώς ο μεγάλος όγκος δεδομένων απαιτεί μεγάλο χρόνο για τους ίδιους αλλά όχι για τη Μηχανική Μάθηση.

1.3 Στόχοι της Διπλωματικής εργασίας

Λαμβάνοντας υπόψιν τα παραπάνω, κύριος στόχος αυτής της διπλωματικής εργασίας είναι η εφαρμογή της Μηχανικής Μάθησης στα Δίκτυα Υπολογιστών. Η Μηχανική Μάθηση θα λειτουργεί ως ένα συμπληρωματικό εργαλείο με σκοπό την δημιουργία ασφάλειας μέσα στα δίκτυα για τους χρήστες που τα χρησιμοποιούν. Με αυτό τον τρόπο θα ενισχύει την ομαλή λειτουργία του δικτύου αλλά και ταυτόχρονα την εμπειρία των χρηστών μέσα σε αυτό.

1.4 Συνεισφορά της Διπλωματικής Εργασίας

Στη διπλωματική εργασία θα μελετήσουμε την Μηχανική Μάθηση ως ένα εργαλείο εντοπισμού κακόβουλων χρηστών μέσα στα δίκτυα υπολογιστών. Πιο συγκεκριμένα, θα την εφαρμόσουμε σε ένα δικτυακό περιβάλλον, και θέλουμε μέσω της εκπαίδευσης να αναγνωρίζει τους κακόβουλους χρήστες έτσι ώστε να αυξάνεται η αίσθηση της ασφάλειας μέσα στο δίκτυο. Αυτό θα μπορέσουμε να το υλοποιήσουμε κάνοντας χρήση των δένδρων αποφάσεων, ως μία τεχνική Μηχανικής Μάθησης. Συνολικά λοιπόν θα έχουμε ένα σύστημα που θα παίρνει αποφάσεις για το δικτυακό περιβάλλον που θα εφαρμόζεται, κάνοντας το πιο ασφαλές.

1.5 Διάρθρωση της Διπλωματικής Εργασίας

Στο Κεφάλαιο 2, θα κάνουμε μία λεπτομερής αναφορά στη Μηχανική Μάθηση. Συγκεκριμένα θα μιλήσουμε για το τι είναι η Μηχανική Μάθηση, μερικές εφαρμογές της και γενικότερη ανάλυσή της. Στην ανάλυσή της θα δούμε και τις τεχνικές που χρησιμοποιούνται σε αυτήν αρκετά λεπτομερειακά.

Στο Κεφάλαιο 3, θα μιλήσουμε για τα δίκτυα κινητής επικοινωνίας. Θα μιλήσουμε για όλες τις γενιές δικτύων που έχουν υπάρξει και για κάθε μία θα κάνουμε μία ανάλυση για το πως λειτουργούσε αλλά θα κάνουμε και μία αναφορά για το κομμάτι της ασφάλειας στη κάθε γενιά. Στην πέμπτη γενιά, συγκεκριμένα θα αναφερθούμε και λίγο περισσότερο

καθώς είναι και η πιο σύγχρονη γενιά δικτύων με αποτέλεσμα να έχει και το περισσότερο ενδιαφέρον.

Στο Κεφάλαιο 4, θα γίνει εκτενής περιγραφή του μηχανισμού που προτείνουμε και των τεχνικών στους οποίους βασίστηκε η διπλωματική εργασία. Θα αναλύσουμε το μοντέλο μας και θα περιγράψουμε τον τρόπο λειτουργίας του αρκετά λεπτομερειακά έτσι ώστε να είναι εύκολη η κατανόησή του.

Στο Κεφάλαιο 5, θα παρουσιάσουμε τα αποτελέσματα του προγράμματος που αναπτύξαμε. Θα εξηγήσουμε συγκεκριμένα τι σημαίνουν τα αποτελέσματα αλλά και το πως αυτά παρήχθησαν, δείχνοντας τον κώδικα που αναπτύξαμε και παράλληλα δίνοντας μία εξήγηση αυτού για τη λειτουργία του.

Στο Κεφάλαιο 6 έχουμε τον επίλογο, όπου θα παρουσιάσουμε μία σύνοψη της διπλωματικής εργασίας και θα κάνουμε ιδέες για το πως αυτή μπορεί να εξελιχθεί στο μέλλον και να προσαρμοστεί σε περισσότερες εφαρμογές.

Κεφάλαιο 2

Μηχανική Μάθηση

2.1 Εισαγωγή Κεφαλαίου

Στο κεφάλαιο αυτό θα μιλήσουμε για την Μηχανική Μάθηση ή όπως είναι ευρέως γνωστή παγκοσμίως ως Machine Learning, δηλαδή με την αγγλική ορολογία της. Η Μηχανική Μάθηση διαθέτει τεράστια ευελιξία στη χρήση της καθώς προσφέρει άπειρες δυνατότητες τόσο σε τομείς της καθημερινής μας ζωής, όσο και σε πολλούς και διαφορετικούς επιστημονικούς κλάδους όπου και αξιοποιείται. Η ένταξή της λοιπόν ως ένα εργαλείο στα δίκτυα υπολογιστών είναι απολύτως λογική και αναμενόμενη καθώς θα αυξήσει την απόδοση αυτών και θα αλλάξει τον τρόπο λειτουργίας τους ως προς το καλύτερο.

Με την εξέλιξη της τεχνολογίας την σήμερον ημέρα αυξάνεται διαρκώς το πλήθος διαφορετικών συσκευών που έχουν την δυνατότητα σύνδεσης στο διαδίκτυο, με παραδείγματα όπως τα ευρέως γνωστά smartphones μέχρι και σε κλιματιστικά. Όλες αυτές οι συσκευές αλληλεπιδρούν μέσω ενός δικτύου με σκοπό να πετύχουν κοινούς στόχους. Η ορολογία αυτής της διαδικασίας είναι το Διαδίκτυο των Πραγμάτων ή όπως είναι γνωστό στην αγγλικής της ορολογία Internet of Things. Συγκεκριμένα το Internet of Things αναμένεται να απαιτεί πιο αποτελεσματικές και αποδοτικές ασύρματες επικοινωνίες από ποτέ. Για αυτό το λόγο, τεχνικές όπως η εξαγωγή ευφυΐας σήματος και η βελτιστοποιημένη δρομολόγηση θα γίνουν βασικά στοιχεία της ασύρματης επικοινωνίας των Internet of Things [1].

Όπως είναι λογικό οι συσκευές αυτές έχουν γίνει μέρος της καθημερινής μας ζωής σε πολλούς τομείς, αλλάζοντας την προς το καλύτερο, καθώς λύνουν αρκετά προβλήματα κάνοντας χρήση της, ουσιαστικά ατελείωτης, γνώσης που υπάρχει στο διαδίκτυο και όχι μόνο. Η ανάπτυξη εφαρμογών, που αξιοποιούν τις δυνατότητες αυτών των συσκευών, αυξάνεται διαρκώς και αυτό με τη σειρά του επιβαρύνει τα ασύρματα και ενσύρματα δίκτυα υπολογιστών με τη συνεχή αύξηση του όγκου δεδομένων που παράγουν. Παραδείγματα αυτών των εφαρμογών ξεκινάνε από ιατρικές εφαρμογές [2] έως και σε παροχές νερού [3].

Με την είσοδο αυτών των συσκευών είναι λογικό ότι και οι ανάγκες του δικτύου θα χρειαστούν ανάπτυξη μηχανισμών που θα είναι ικανοί να διαχειρίζονται αποδοτικά τους

δικτυακούς πόρους αλλά και την αποδοτική διαχείριση της κίνησης που αναπτύσσεται στο δίκτυο. Η ανάπτυξη αυτών των μηχανισμών έχει αποτελέσει το επίκεντρο ερευνών στην εισαγωγή της Μηχανικής Μάθησης στα δίκτυα υπολογιστών με σκοπό την βελτιστοποίηση των υπαρχόντων μηχανισμών που χρησιμοποιούνται ή ακόμα και στη παραγωγή καινούργιων μηχανισμών Μηχανικής Μάθησης πάνω σε αυτά τα προβλήματα.

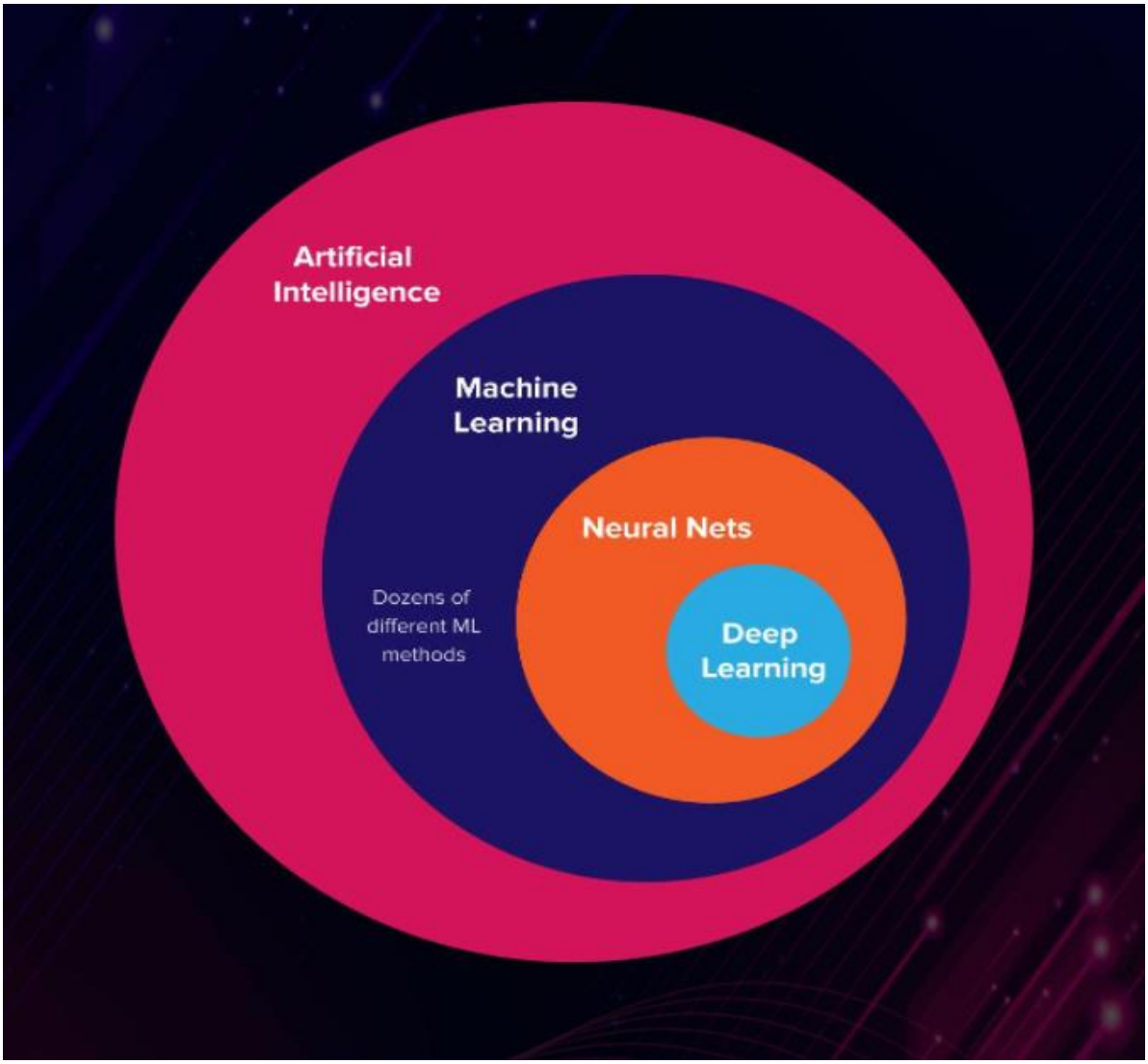
2.2 Εισαγωγή στη Μηχανική Μάθηση

Η Μηχανική Μάθηση διαθέτει διάφορες τεχνικές και αλγορίθμους που με χρήση αυτών από τους δικτυακούς μηχανισμούς μπορεί να βελτιωθεί η απόδοση του δικτύου. Σε αυτό το σημείο θα γίνει ανάλυση αυτών των τεχνικών και αλγορίθμων, και γενικότερα της Μηχανικής Μάθησης, έτσι ώστε να μπορούμε να είμαστε σε θέση να την εφαρμόζουμε στα δίκτυα υπολογιστών

Αρχικά ας μιλήσουμε λίγο για την Μηχανική Μάθηση, ας μιλήσουμε δηλαδή για το τι είναι ακριβώς. Η Μηχανική Μάθηση είναι ένας κλάδος υπολογιστικών αλγορίθμων, που συνεχώς εξελίσσεται, με σκοπό να μιμούνται την ανθρώπινη νοημοσύνη ή την ανθρώπινη λογική, όπως για παράδειγμα στη λήψη αποφάσεων, μαθαίνοντας από το περιβάλλον τους. Στην εποχή των big data (μεγάλου όγκου δεδομένων) που ζούμε η Μηχανική Μάθηση θεωρείται ως το workhorse (άλογο εργασίας) καθώς παράγει αρκετά μεγάλο έργο, που χωρίς αυτήν δε θα ήταν τόσο εύκολο [4]. Με τη διάδοση των πληροφοριών και τη δημιουργία πολλών βάσεων δεδομένων, ο τρόπος εξαγωγής δεδομένων από τις χρήσιμες πληροφορίες στα παραπάνω αποτελεί και το επείγον πρόβλημα που πρέπει να λυθεί [5]. Για αυτό άλλωστε θεωρείται ως το workhorse που αναφέραμε.

Η Μηχανική Μάθηση αποτελεί ένα υποσύνολο της Τεχνητής Νοημοσύνης (Artificial Intelligence). Αυτές οι δύο τεχνολογίες είναι οι πιο σύγχρονες τεχνολογίες που χρησιμοποιούνται για τη δημιουργία ευφυών συστημάτων. Βέβαια πρόκειται για δύο σχετικές τεχνολογίες και μερικές φορές οι άνθρωποι τις χρησιμοποιούν ως συνώνυμα μεταξύ τους, ωστόσο όμως και οι δύο είναι διαφορετικοί όροι. Σε ένα ευρύ επίπεδο, μπορούμε να διαφοροποιήσουμε τόσο την Τεχνητή Νοημοσύνη όσο και τη Μηχανική Μάθηση λέγοντας το εξής: Η Τεχνητή Νοημοσύνη είναι μία μεγαλύτερη ιδέα για τη δημιουργία έξυπνων μηχανών που μπορούν να προσομοιώσουν την ανθρώπινη ικανότητα σκέψης και συμπεριφοράς, ενώ η Μηχανική Μάθηση είναι μια εφαρμογή ή, όπως αναφέραμε προηγουμένως, ένα υποσύνολο της τεχνητής νοημοσύνης που επιτρέπει στις μηχανές να μαθαίνουν από δεδομένα χωρίς να προγραμματίζονται ρητά [6].

Επίσης η Μηχανική Μάθηση αποτελεί ένα υπερσύνολο της Βαθειάς Μάθησης (Deep Learning). Η Βαθειά Μάθηση κάνει χρήση των νευρωνικών δικτύων με σκοπό την ανάλυση διαφορετικών παραγόντων με δομή που είναι παρόμοια με το ανθρώπινο νευρικό δίκτυο [7]. Στην Εικόνα 1 φαίνονται τα σύνολα αυτά μεταξύ τους:

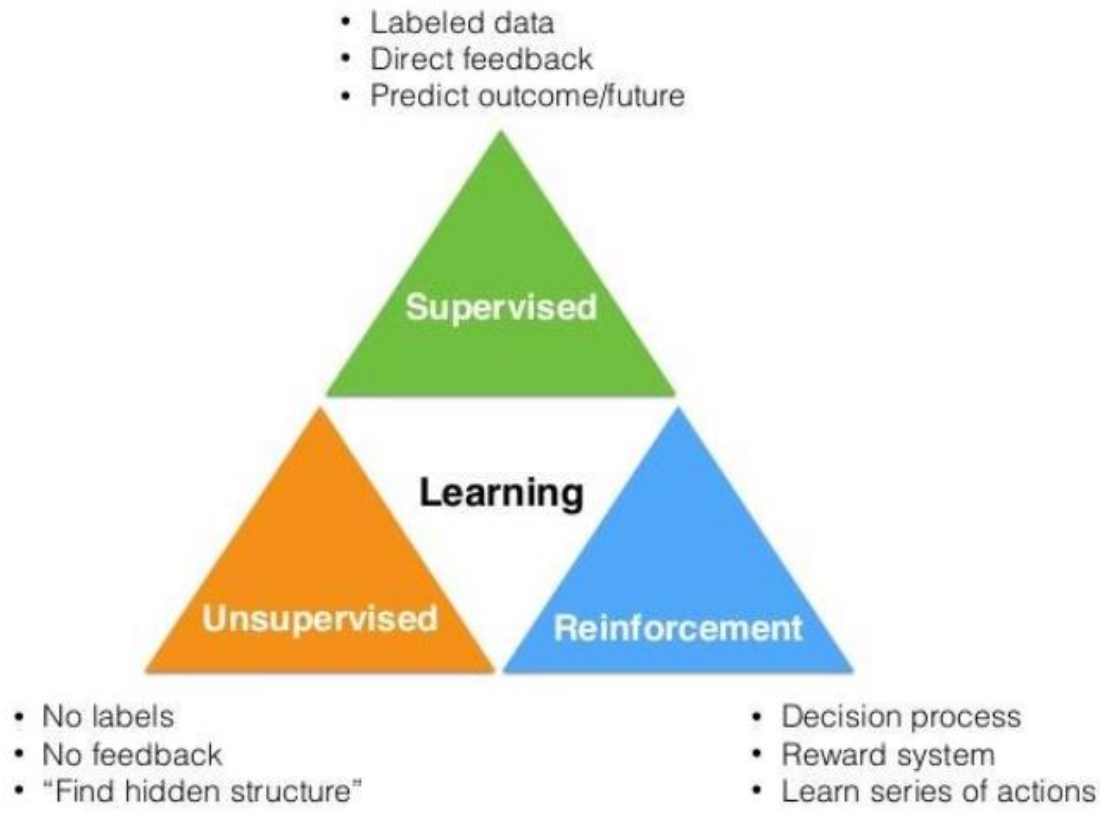


Εικόνα 1: Διάγραμμα Euler των συνόλων Τεχνητής Νοημοσύνης, Μηχανικής Μάθησης και Βαθειάς Μάθησης [7]

Η Μηχανική Μάθηση, όπως είδαμε, αποτελεί μέρος ενός μεγαλύτερο συνόλου, της Τεχνητής Νοημοσύνης, που περιέχει την ίδια καθώς και τη Βαθειά Μάθηση ως υποσύνολά της. Όμως και η ίδια χωρίζεται σε υποκατηγορίες , οι οποίες είναι οι εξής [8]:

1. **Εποπτευόμενη Μάθηση (Supervised Learning)**

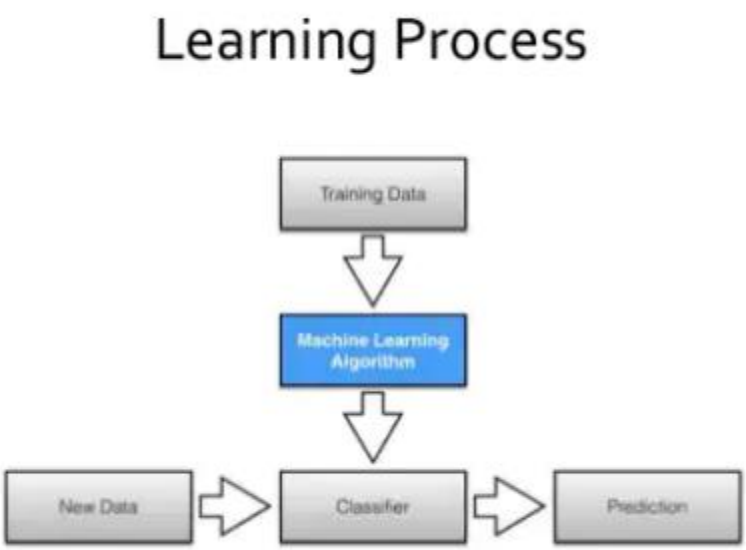
2. **Μη εποπτευόμενη Μάθηση (Unsupervised Learning)**
3. **Ημι-εποπτευόμενη Μάθηση (Semi-supervised Learning)**
4. **Ενισχυμένη Μάθηση (Reinforcement Learning)**



Εικόνα 2: Υποκατηγορίες Μηχανικής Μάθησης [9]

2.2.1 Εποπτευόμενη Μάθηση

Η διαδικασία εκμάθησης σε ένα απλό μοντέλο μηχανικής μάθησης χωρίζεται σε δύο βήματα, την εκπαίδευση και στις δοκιμές. Στη διαδικασία εκπαίδευσης, τα δείγματα στα δεδομένα εκπαίδευσης λαμβάνονται ως είσοδος, και με αυτόν τον τρόπο ο αλγόριθμος μάθησης χτίζει το μοντέλο μάθησης [10]. Στη διαδικασία δοκιμής, το μοντέλο εκμάθησης χρησιμοποιεί τη μηχανή εκτέλεσης για να κάνει την πρόβλεψη για τα δεδομένα δοκιμής. Τα δεδομένα με ετικέτα είναι το αποτέλεσμα του μοντέλου που δίνει τη τελική πρόβλεψη. Παρακάτω φαίνεται και αυτή η διαδικασία:



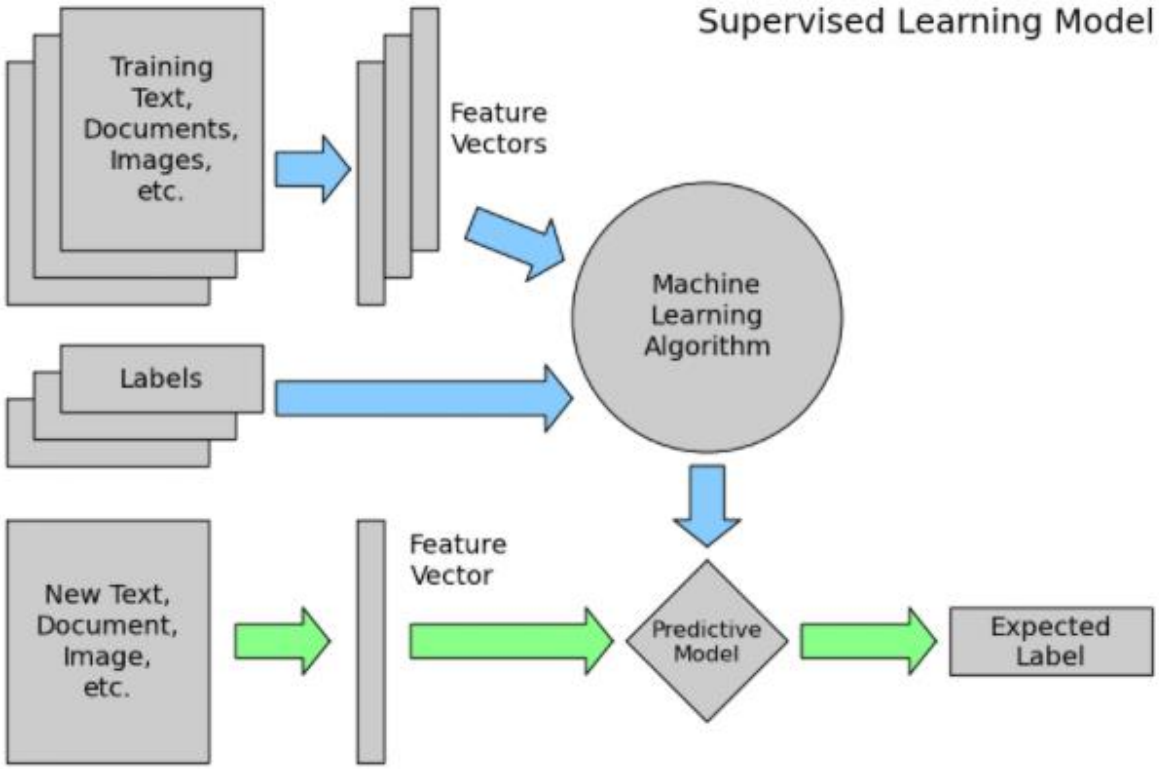
Εικόνα 3: Εποπτευόμενη μαθησιακή διαδικασία [11]

Η εποπτευόμενη μάθηση είναι η πιο κοινή τεχνική στα προβλήματα ταξινόμησης, καθώς ο στόχος συχνά είναι να κάνουμε το μηχάνημα να μάθει ένα σύστημα ταξινόμησης που δημιουργήσαμε. Πιο συχνά η εποπτευόμενη μάθηση αφήνει απροσδιόριστη την πιθανότητα εισαγωγής, όπως μία είσοδος όπου η αναμενόμενη έξοδος είναι γνωστή. Αυτή η διαδικασία παρέχει ένα σύνολο δεδομένων (dataset) που αποτελείται από χαρακτηριστικά και ετικέτες.

Ο κύριος σκοπός είναι η κατασκευή ενός εκτιμητή ή μοντέλου, που θα είναι ικανό να προβλέψει την ετικέτα ενός αντικειμένου, από το σύνολο των χαρακτηριστικών. Στη συνέχεια, ο αλγόριθμος εκμάθησης λαμβάνει ένα σύνολο χαρακτηριστικών ως εισόδων μαζί με τις σωστές εξόδους, και μαθαίνει συγκρίνοντας την πραγματική του έξοδο με διορθωμένα αποτελέσματα, έτσι ώστε να βρει πιθανά σφάλματα και τροποποιεί το μοντέλο αναλόγως.

Η εποπτευόμενη μάθηση είναι η πιο κοινή τεχνική εκπαίδευσης για ουδέτερα δίκτυα και τα δέντρα αποφάσεων. Και τα δύο εξαρτώνται από τις πληροφορίες που δίνονται από την προκαθορισμένη ταξινόμηση.

Στην Εικόνα 4 φαίνεται το μοντέλο που δημιουργείται από αυτή τη διαδικασία:



Εικόνα 4: Μοντέλο εποπτευόμενης μάθησης [12]

Στην εποπτευόμενη μάθηση υπάρχουν αλγόριθμοι που υλοποιούν τα παραπάνω που συζητήσαμε. Οι κυριότεροι από αυτούς είναι οι εξής [13]:

1. **Δέντρα αποφάσεων (Decision Trees)**
2. **Γραμμική παλινδρόμηση (Linear Regression)**
3. **Αφελής Bayes (Naive Bayes)**
4. **Λογιστική παλινδρόμηση (Logistic Regression)**

Στα παρακάτω υποκεφάλαια ακολουθεί μία ανάλυση αυτών των αλγορίθμων.

2.2.1.1 Decision Trees

Το δένδρο απόφασης [14] αντιπροσωπεύει έναν ταξινομητή που εκφράζεται ως αναδρομική κατάτμηση του χώρου παραδείγματος. Το δέντρο απόφασης αποτελείται από κόμβους που σχηματίζουν το λεγόμενο δέντρο ρίζας, το οποίο σημαίνει ότι είναι ένα

κατανεμημένο δέντρο με έναν βασικό κόμβο που ονομάζεται ρίζα χωρίς εισερχόμενα άκρα.

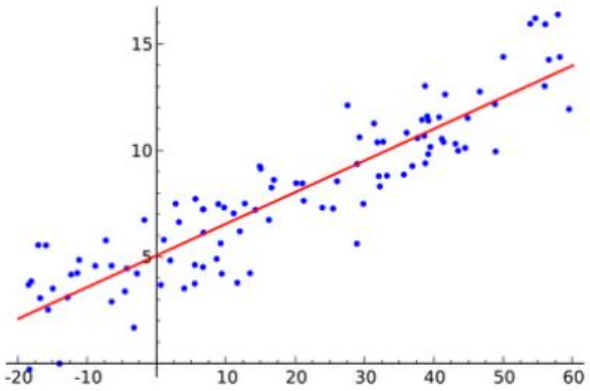
Όλοι οι άλλοι κόμβοι έχουν ακριβώς ένα εισερχόμενο άκρο. Ο κόμβος που έχει εξερχόμενες ακμές ονομάζεται εσωτερικός κόμβος ή κόμβος δοκιμής. Οι υπόλοιποι κόμβοι ονομάζονται φύλλα. Σε ένα δέντρο απόφασης, κάθε δοκιμαστικός κόμβος χωρίζει τον χώρο του στιγμιότυπου σε δύο ή περισσότερους υποχώρους σύμφωνα με μία ορισμένη διακριτή συνάρτηση των τιμών εισόδου. Στην πιο απλή περίπτωση, κάθε δοκιμή εξετάζει ένα μεμονωμένο χαρακτηριστικό, έτσι ώστε ο χώρος του στιγμιότυπου να τμηματοποιείται σύμφωνα με την τιμή του χαρακτηριστικού. Στην περίπτωση αριθμητικών χαρακτηριστικών, η συνθήκη αναφέρεται σε ένα εύρος.

2.2.1.2 Linear Regression

Ο στόχος της Linear Regression, ως μέρος της οικογενείας αλγορίθμων παλινδρόμησης, είναι η εύρεση σχέσεων και εξαρτήσεων μεταξύ μεταβλητών. Αντιπροσωπεύει μία μοντελοποίηση μεταξύ μιας συνεχούς βαθμωτής εξαρτώμενης μεταβλητής y (επίσης ετικέτα ή στόχος στην ορολογία της μηχανικής μάθησης) και ενός ή περισσότερων (ένα διάνυσμα D-διαστάσεων) μεταβλητών επεξήγησης (επίσης ανεξάρτητες μεταβλητές, μεταβλητές εισόδου, παρατηρούμενα δεδομένα κ.ο.κ) συμβολίζεται ως X χρησιμοποιώντας μία γραμμική συνάρτηση. Στην ανάλυση παλινδρόμησης ο στόχος είναι η πρόβλεψη μιας συνεχούς μεταβλητής στόχου, ενώ μία άλλη περιοχή που ονομάζεται ταξινόμηση είναι η πρόβλεψη μιας ετικέτας από ένα πεπερασμένο σύνολο. Το μοντέλο για πολλαπλή παλινδρόμηση που περιλαμβάνει τον γραμμικό συνδυασμό μεταβλητών εισόδου παίρνει τη μορφή:

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \ldots + e$$

Η Linear Regression, όπως είπαμε, ανήκει στους αλγορίθμους της εποπτευόμενης μάθησης [15]. Αυτό σημαίνει ότι εκπαιδεύουμε το μοντέλο σε ένα σύνολο δεδομένων με ετικέτα και στη συνέχεια χρησιμοποιούμε το μοντέλο για να κάνουμε προβλέψεις ετικετών σε δεδομένα χωρίς ετικέτα.



Εικόνα 5: Αναπαράσταση της Linear Regression [16]

Όπως φαίνεται στην Εικόνα 5 το μοντέλο, που εδώ είναι η κόκκινη γραμμή, υπολογίζεται χρησιμοποιώντας δεδομένα εκπαίδευσης, τα μπλε σημεία, όπου κάθε σημείο έχει μία γνωστή ετικέτα (άξονα y) για να ταιριάζει τα σημεία όσο το δυνατόν με περισσότερη ακρίβεια, με την ελαχιστοποίηση της τιμής μιας επιλεγμένης συνάρτησης απώλειας. Στην συνέχεια μπορούμε να χρησιμοποιήσουμε το μοντέλο για να προβλέψουμε άγνωστες ετικέτες, δηλαδή γνωρίζουμε την τιμή x και θέλουμε να προβλέψουμε την τιμή y.

2.2.1.3 Naive Bayes

Η ταξινόμηση Bayes [17] είναι και αυτή μέθοδος της εποπτευόμενης μάθησης καθώς και στατιστική μέθοδος ταξινόμησης. Προϋποθέτει ένα υποκειμενικό πιθανοτικό μοντέλο και επιτρέπει την αβεβαιότητα για το μοντέλο με έναν τρόπο αρχής με τον προσδιορισμό των πιθανοτήτων του αποτελέσματος. Ο βασικός σκοπός της ταξινόμησης Bayes είναι ότι μπορεί να λύσει προβλήματα πρόβλεψης. Αυτή η ταξινόμηση παρέχει πρακτικούς αλγόριθμους μάθησης και μπορεί να συνδυάσει παρατηρούμενα δεδομένα. Η ταξινόμηση Bayes παρέχει χρήσιμη προοπτική για καλύτερη κατανόηση και αξιολόγηση αλγορίθμων μάθησης. Υπολογίζει ρητές πιθανότητες για υπόθεση και ενισχύει το θόρυβο στα δεδομένα εισόδου. Ας εξετάσουμε τη γενική κατανομή δύο τιμών $P(x_1, x_2)$. Χρησιμοποιώντας τον κανόνα του Bayes, χωρίς απώλεια γενικότητας παίρνουμε την εξίσωση:

$$P(x_1, x_2) = P(x_1|x_2)P(x_2)$$

Αντίστοιχα, αν υπάρχει και άλλη μεταβλητή κλάσης c, παίρνουμε την εξίσωση:

$$P(x_1, x_2|c) = P(x_1|x_2, c)P(x_2|c)$$

Εάν η κατάσταση γενικεύεται με δύο μεταβλητές σε μία υπό όρους υπόθεση ανεξαρτησίας για ένα σύνολο μεταβλητών $x_1,x_2,\ldots,x_N$, υπό τον όρο μιας άλλης μεταβλητής c, παίρνουμε το ακόλουθο:

$$P(x|c) = \prod_{i=i}^{N} P(x_i|c)$$

2.2.1.4 Logistic Regression

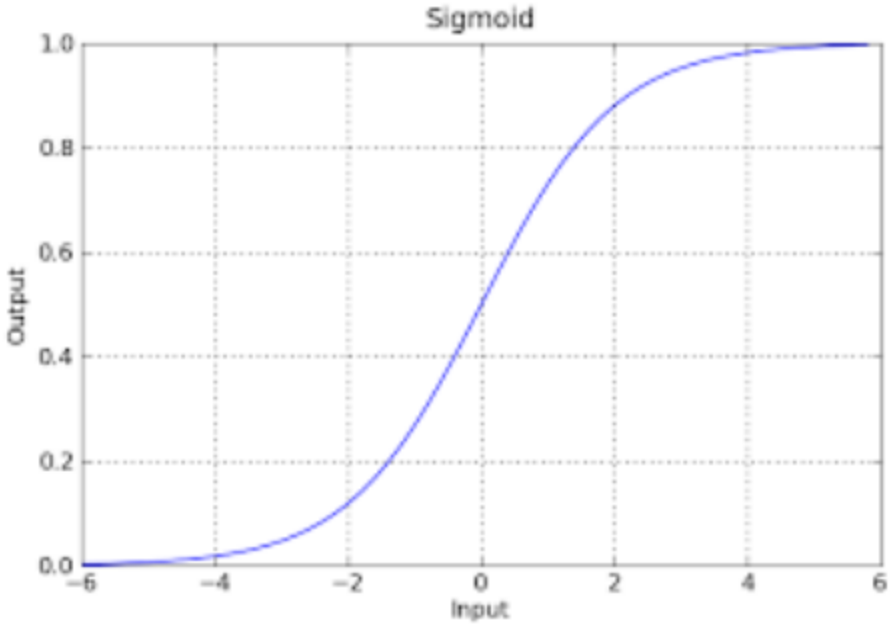
Όπως η naive Bayes, η logistic regression [18] λειτουργεί εξάγοντας κάποιο σύνολο χαρακτηριστικών από την είσοδο, λαμβάνοντας αρχεία καταγραφής και συνδυάζοντάς τα γραμμικά, το οποίο σημαίνει ότι κάθε χαρακτηριστικό πολλαπλασιάζεται με ένα βάρος και στη συνέχεια αθροίζεται. Η πιο σημαντική διαφορά μεταξύ του naive Bayes και της logistic regression είναι ότι η logistic regression είναι ένας διακριτικός ταξινομητής ενώ ο naive Bayes είναι ένας γενεσιουργός ταξινομητής. Η logistic regression [18] είναι ένας τύπος παλινδρόμησης που προβλέπει την πιθανότητα εμφάνισης ενός συμβάντος με την προσαρμογή δεδομένων σε μία λογιστική συνάρτηση. Όπως πολλές μορφές ανάλυσης παλινδρόμησης, η λογιστική κάνει χρήση πολλών μεταβλητών πρόβλεψης που μπορεί να είναι αριθμητικές ή κατηγορικές. Η υπόθεση της logistic regression ορίζεται ως εξής:

$$h_\theta(x) = g(\theta^T x)$$

Όπου η συνάρτηση g είναι σιγμοειδής συνάρτηση και ορίζεται ως εξής:

$$g(z) = \frac{1}{1 + e^{-z}}$$

Η σιγμοειδής συνάρτηση έχει ειδικές ιδιότητες που έχουν ως αποτέλεσμα τις τιμές στην περιοχή [0,1], όπως φαίνεται στην Εικόνα 6



Εικόνα 6 Αναπαράσταση της λογιστικής συνάρτησης [19]

Η συνάρτηση κόστους για την λογιστική παλινδρόμηση δίνεται ως εξής:

$$J(\theta) = \frac{1}{m}\sum_{i=1}^{m}[-y^{(i)}\log\left(h_\theta\left(x^{(i)}\right)\right) - (1-y^{(i)})\log(1-h_\theta(x^{(i)}))]$$

Για να βρούμε το ελάχιστο αυτής της συνάρτησης κόστους, στη μηχανική μάθηση χρησιμοποιείται μια ενσωματωμένη συνάρτηση που ονομάζεται fmin_bfgs2, που βρίσκει τις καλύτερες παραμέτρους θ για τη συνάρτηση κόστους της logistic regression δεδομένου ενός σταθερού συνόλου δεδομένων (τιμών x και y). Οι παράμετροι είναι οι αρχικές τιμές των παραμέτρων που πρέπει να βελτιστοποιηθούν και μία συνάρτηση που όταν της δίνεται το σύνολο εκπαίδευσης και ενός συγκεκριμένου θ, υπολογίζει το κόστος της λογιστικής παλινδρόμησης και κλίση σε σχέση με το σύνολο θ για το σύνολο δεδομένων με τιμές x και y. Η τελική τιμή της θ χρησιμοποιείται για τη γραφική παράσταση του ορίου απόφασης των δεδομένων εκπαίδευσης.

2.2.2 Μη εποπτευόμενη Μάθηση

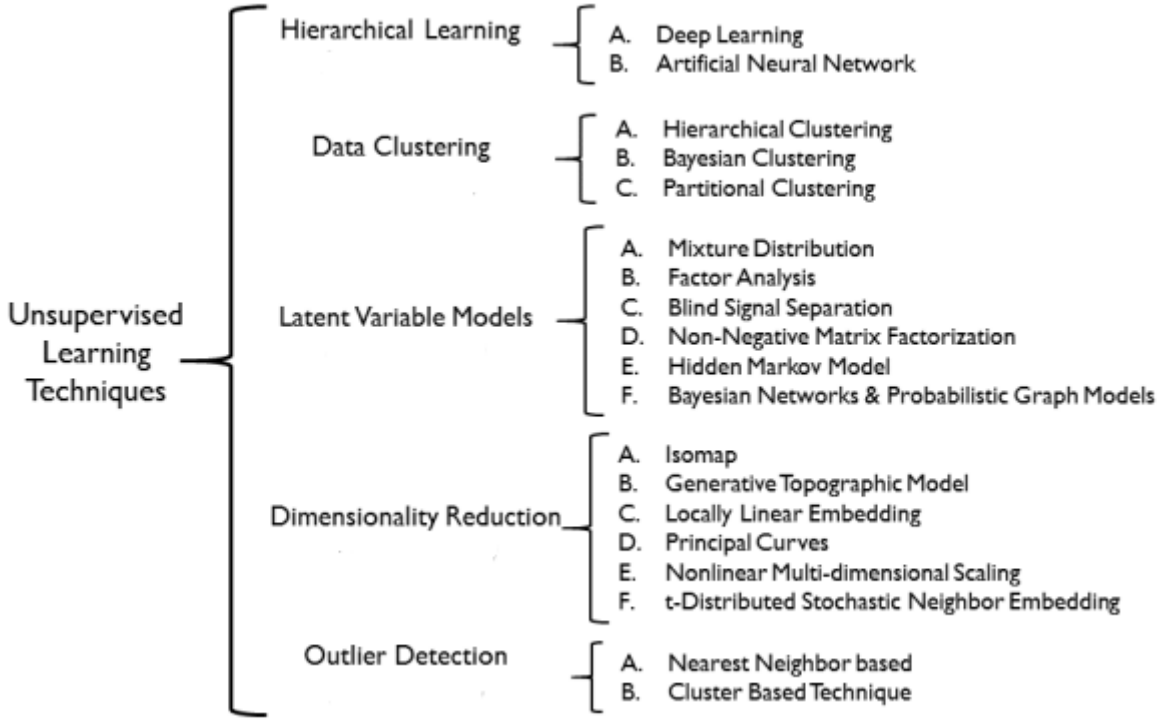
Οι αλγόριθμοι της μη εποπτευόμενης μάθησης, σε αντίθεση με αυτών της εποπτευόμενης, δε έχουν κάποιο δείγμα για εκπαίδευση αλλά έχουν μόνο δείγμα

δεδομένων και πρέπει από μόνοι τους να βγάλουν συμπεράσματα και να κάνουν ταξινόμηση [20].

Σε αυτό το υποκεφάλαιο, θα παρουσιάσουμε ορισμένες ευρέως χρησιμοποιούμενες τεχνικές μη εποπτευόμενης μάθησης. Έχουμε διαχωρίσει αυτή τη μάθηση σε πέντε μεγάλες κατηγορίες οι οποίες είναι οι εξής [21]:

- **Ιεραρχική Μάθηση (Hierarchical Learning)**
- **Ομαδοποίηση Δεδομένων (Data clustering)**
- **Μοντέλα Λανθάνουσας Μεταβλητής (Latent Variable Models)**
- **Μείωση Διαστάσεων (Dimensionality Reduction)**
- **Ανίχνευση Ακραίων Τιμών (Outlier Detection)**

Στην Εικόνα 7 φαίνονται επίσης αυτές οι κατηγορίες:



Εικόνα 7: Κατηγορίες Μη εποπτευόμενης Μάθησης [21]

Στα παρακάτω υποκεφάλαια ακολουθεί η ανάλυση αυτών των κατηγοριών

2.2.2.1 Ιεραρχική Μάθηση

Η ιεραρχική μάθηση ορίζεται ως η μάθηση απλών και σύνθετων χαρακτηριστικών από μία ιεραρχία πολλαπλών γραμμικών και μη γραμμικών ενεργοποιήσεων. Στα μοντέλα εκμάθησης ένα χαρακτηριστικό είναι μία μετρήσιμη ιδιότητα των δεδομένων εισόδου, όπου τα επιθυμητά χαρακτηριστικά είναι ιδανικά ενημερωτικά, διακριτικά και ανεξάρτητα. Στα στατιστικά στοιχεία τα χαρακτηριστικά είναι επίσης γνωστά ως επεξηγηματικές μεταβλητές [22].

Η εκμάθηση χαρακτηριστικών, που είναι γνωστή και ως εκμάθηση αναπαράστασης δεδομένων, είναι ένα σύνολο τεχνικών που μπορούν να μάθουν ένα ή περισσότερα χαρακτηριστικά από τα δεδομένα εισόδου [23]. Περιλαμβάνει τη μεταμόρφωση ακατέργαστων δεδομένων σε ποσοτικοποιήσιμη και συγκρίσιμη αναπαράσταση, που είναι ειδική για την ιδιότητα της εισόδου, αλλά γενική αρκετά για σύγκριση με παρόμοιες εισόδους.

Συμβατικά, τα χαρακτηριστικά είναι φτιαγμένα ειδικά για την τρέχουσα εφαρμογή. Βασίζεται στο πεδίο γνώσης, αλλά ακόμα και τότε δεν γενικεύουν καλά τη διακύμανση των δεδομένων του πραγματικού κόσμου, η οποία οδηγεί σε αυτοματοποιημένη εκμάθηση γενικευμένων χαρακτηριστικών από την υποκείμενη δομή των δεδομένων εισόδου.

2.2.2.2 Ομαδοποίηση Δεδομένων

Η ομαδοποίηση δεδομένων ανήκει και αυτή στη μη εποπτευόμενη μάθηση, και στοχεύει στην εύρεση κρυφών μοτίβων σε δεδομένα εισόδου χωρίς ετικέτα με τη μορφή συστάδων [24]. Με απλά λόγια περιλαμβάνει τη διάταξη των δεδομένων σε ουσιαστικές φυσικές ομαδοποιήσεις, με βάση την ομοιότητα μεταξύ διαφορετικών χαρακτηριστικών για να μάθει σχετικά με τη δομή του.

Η ομαδοποίηση περιλαμβάνει την οργάνωση των δεδομένων με τέτοιο τρόπο ώστε να υπάρχει υψηλή εντός (high intra-cluster) και χαμηλή (low inter-cluster) ομοιότητα μεταξύ των συστάδων. Τα δομημένα δεδομένα που προκύπτουν ονομάζονται ως έννοια δεδομένων [25].

Η ομαδοποίηση χρησιμοποιείται σε πολλές εφαρμογές από τους τομείς της Μηχανικής Μάθησης, της εξόρυξης δεδομένων, της ανάλυσης δικτύου. Στη δικτύωση, οι τεχνικές ομαδοποίησης έχουν αναπτυχθεί ευρέως για εφαρμογές όπως η ανάλυση

κυκλοφορίας και η ανίχνευση ανωμαλιών σε όλα τα είδη δικτύων, όπως για παράδειγμα ασύρματα δίκτυα αισθητήρων, με ανίχνευση ανωμαλίας [26].

Η ομαδοποίηση βελτιώνει την απόδοση σε διάφορα είδη εφαρμογών. Για παράδειγμα οι McGregor et al [27] προτείνουν μία αποτελεσματική προσέγγιση ανίχνευσης πακέτων που χρησιμοποιεί τον αλγόριθμο πιθανολογικής ομαδοποίησης Προσδοκιών-Μεγιστοποίησης (Expectation-Maximization) , ο οποίος ομαδοποιεί ροές σε ένα μικρό αριθμό συστάδων, με στόχο την ανάλυση κυκλοφορίας δικτύου χρησιμοποιώντας απλά ένα σύνολο αντιπροσωπευτικών συμπλεγμάτων.

2.2.2.3 Μοντέλα Λανθάνουσας Μεταβλητής

Ένα μοντέλο λανθάνουσας μεταβλητής είναι ένα στατιστικό μοντέλο που σχετίζει τις φανερές μεταβλητές με ένα σύνολο κρυφών μεταβλητών. Το μοντέλο αυτό μας επιτρέπει να εκφράσουμε σχετικά σύνθετες κατανομές όσον αφορά τις ελκόμενες κοινές κατανομές σε έναν διευρυμένο μεταβλητό χώρο [28].

Οι υποκείμενες μεταβλητές μιας διεργασίας αναπαρίστανται σε χώρο υψηλότερων διαστάσεων χρησιμοποιώντας σταθερό μετασχηματισμό και οι στοχαστικές παραλλαγές είναι γνωστές ως μοντέλα λανθάνουσας μεταβλητής όπου η κατανομή σε μεγαλύτερη διάσταση οφείλεται στον μικρό αριθμό κρυφών μεταβλητών που ενεργούν συνδυαστικά [29].

Αυτά τα μοντέλα χρησιμοποιούνται για:

- Οπτικοποίηση δεδομένων (Data visualization)
- Μείωση διαστάσεων (Dimensionality reduction)
- Βελτιστοποίηση (Optimization)
- Μάθηση διανομής (Distribution learning)
- Τυφλό διαχωρισμό σημάτων (Blind signal separation)
- Ανάλυση παραγόντων (Factor Analysis)

2.2.2.4 Μείωση Διαστάσεων

Η αναπαράσταση δεδομένων σε λιγότερες διαστάσεις είναι ένα άλλο καθιερωμένο καθήκον της μη εποπτευόμενης μάθησης. Πραγματικά δεδομένα έχουν συχνά υψηλές διαστάσεις, δηλαδή σε πολλά σύνολα δεδομένων οι διαστάσεις μπορεί να ανέλθουν σε

χιλιάδες, ακόμη και εκατομμύρια, δυνητικά συσχετιζόμενες διαστάσεις [30]. Ωστόσο, παρατηρείται ότι η εγγενής διάσταση των δεδομένων είναι μικρότερη από τον συνολικό αριθμό των διαστάσεων.

Για να βρεθεί το ουσιαστικό πρότυπο των υποκείμενων δεδομένων με εξαγωγή εγγενών διαστάσεων, είναι απαραίτητο να μην χαθεί η πραγματική ουσία. Για παράδειγμα, μπορεί να συμβαίνει ότι ένα φαινόμενο είναι παρατηρήσιμο σε δεδομένα υψηλότερης διάστασης και να καταστέλλεται στις χαμηλότερες. Τα φαινόμενα αυτά λέγεται ότι πάσχουν από τη κατάρα της διάστασης [31]. Ενώ η μείωση διαστάσεων χρησιμοποιείται μερικές φορές εναλλακτικά με την επιλογή χαρακτηριστικών [32], [33], και υπάρχει μία λεπτή διαφορά μεταξύ των δύο [34].

Η επιλογή χαρακτηριστικών εκτελείται παραδοσιακά ως εποπτευόμενη με έναν ειδικό τομέα που βοηθά στη χειροποίητη δημιουργία ενός συνόλου κρίσιμων χαρακτηριστικών των δεδομένων. Μία τέτοια απόδοση, μπορεί να αποδώσει καλά, αλλά δεν είναι κλιμακωτή και επιρρεπής στην κρίση προκατάληψης. Η μείωση των διαστάσεων, από την άλλη πλευρά, είναι μία μη εποπτευόμενη εργασία, όπου αντί να επιλεγεί ένα υποσύνολο χαρακτηριστικών, δημιουργεί νέα χαρακτηριστικά (διαστάσεις) ως συνάρτηση όλων των χαρακτηριστικών. Για να το πούμε διαφορετικά, η επιλογή χαρακτηριστικών λαμβάνει υπόψη εποπτευόμενες ετικέτες δεδομένων, ενώ η μείωση διαστάσεων εστιάζει στα σημεία δεδομένων και τις κατανομές τους σε ένα χώρο Ν διαστάσεων.

2.2.2.5 Ανίχνευση Ακραίων Τιμών

Η ανίχνευση ακραίων τιμών είναι μία σημαντική εφαρμογή μη εποπτευόμενης μάθησης. Ένα σημείο δείγματος που απέχει από άλλα δείγματα ονομάζεται ως ακραίο. Μία ακραία τιμή μπορεί να προκύψει για τους εξής λόγους:

- Θόρυβος
- Σφάλμα μέτρησης
- Μεγάλων κατανομών ουράς
- Μείγμα δύο κατανομών

Υπάρχουν δύο δημοφιλείς υποκείμενες τεχνικές, για μη εποπτευόμενη ανίχνευση ακραίων στοιχείων, βάσει των οποίων έχουν σχεδιαστεί πολλοί αλγόριθμοι. Αυτές είναι, η τεχνική που βασίζεται στον πλησιέστερο γείτονα και η μέθοδος που βασίζεται στην ομαδοποίηση [21].

2.2.3 Ημι-εποπτευόμενη Μάθηση

Η ημι-εποπτευόμενη μάθηση είναι ένας τύπος τεχνικής Μηχανικής Μάθησης. Βρίσκεται στα μισά του δρόμου μεταξύ της εποπτευόμενης και μη εποπτευόμενης μάθησης. Ο κύριος στόχος της ημι-εποπτευόμενης μάθησης είναι να ξεπεράσει τα μειονεκτήματα τόσο της εποπτευόμενης όσο και της μη εποπτευόμενης μάθησης.

Η εποπτευόμενη μάθηση απαιτεί τεράστιο όγκο δεδομένων εκπαίδευσης για την ταξινόμηση των δεδομένων δοκιμής, κάτι που είναι μία οικονομικά αποδοτική (cost effective) και χρονοβόρα διαδικασία. Από την άλλη πλευρά, η μη εποπτευόμενη μάθηση δεν απαιτεί δεδομένα με ετικέτα, τα οποία ομαδοποιούν τα δεδομένα με βάση την ομοιότητα στα σημεία δεδομένων, χρησιμοποιώντας είτε ομαδοποίηση είτε προσέγγιση μέγιστης πιθανότητας. Το κύριο μειονέκτημα αυτής της προσέγγισης είναι ότι δε μπορεί να ομαδοποιήσει με ακρίβεια άγνωστα δεδομένα.

Για να ξεπεραστούν αυτά τα ζητήματα, η ημι-εποπτευόμενη μάθηση έχει προταθεί από την ερευνητική κοινότητα, διότι μπορεί να μάθει με μικρή ποσότητα δεδομένων εκπαίδευσης και να επισημάνει τα άγνωστα δεδομένα δοκιμής. Η ημι-εποπτευόμενη μάθηση δημιουργεί ένα μοντέλο με λίγα μοτίβα με ετικέτα ως δεδομένα εκπαίδευσης και αντιμετωπίζει τα υπόλοιπα ως δεδομένα δοκιμής [35].

2.2.4 Ενισχυτική Μάθηση

Ένας από τους πρωταρχικούς στόχους του τομέα της τεχνητής νοημοσύνης είναι η παραγωγή πλήρως αυτόνομων παραγόντων που αλληλεπιδρούν με τα περιβάλλοντά τους για να μάθουν βέλτιστες συμπεριφορές, βελτιώνοντας ξανά στο πέρασμα του χρόνου μέσω δοκιμής και λάθους (trial and error). Η δημιουργία συστημάτων τεχνητής νοημοσύνης που είναι ανταποκριστικά και που μπορούν να μάθουν αποτελεσματικά, αποτελεί μία μακροχρόνια πρόκληση, που ποικίλλει από ρομπότ, που μπορούν να αισθανθούν και να αντιδράσουν στον κόσμο γύρω τους, μέχρι σε πράκτορες που βασίζονται αποκλειστικά σε λογισμικό, που μπορεί να αλληλεπιδράσει με τη φυσική γλώσσα και τα πολυμέσα [36].

Ένα θεμελιώδες μαθηματικό πλαίσιο για την αυτόνομη μάθηση μέσω της εμπειρίας, είναι η ενισχυτική μάθηση [37]. Αν και η ενισχυμένη μάθηση είχε κάποιες επιτυχίες στο παρελθόν [38], [39], [40], [41], οι προηγούμενες προσεγγίσεις δεν είχαν επεκτασιμότητα

και ήταν περιορισμένες σε προβλήματα αρκετά χαμηλών διαστάσεων. Αυτοί οι περιορισμοί υπάρχουν επειδή οι αλγόριθμοι ενισχυμένης μάθησης μοιράζονται τα ίδια ζητήματα πολυπλοκότητας με τους αλγορίθμους πολυπλοκότητας μνήμης, της υπολογιστικής πολυπλοκότητας και στην περίπτωση αλγορίθμων μηχανικής μάθησης, την πολυπλοκότητα δείγματος [42].

Η ουσία της ενισχυμένης μάθησης είναι η μάθηση μέσω της αλληλεπίδρασης. Ένας πράκτορας της ενισχυμένης μάθησης αλληλεπιδρά με το περιβάλλον του και παρατηρώντας τις συνέπειες των πράξεων του μπορεί να μάθει να αλλάζει τη συμπεριφορά του σε ανταπόκριση με τις ανταμοιβές που λαμβάνει. Αυτό το παράδειγμα της μάθησης δοκιμής και λάθους έχει τις ρίζες του στη συμπεριφοριστική ψυχολογία και είναι ένα από τα κύρια θεμέλια της ενισχυμένης μάθησης [37]. Η άλλη βασική επιρροή στην ενισχυμένη μάθηση είναι ο βέλτιστος έλεγχος, ο οποίος έχει δανείσει τους μαθηματικούς φορμαλισμούς (κυρίως ο δυναμικός προγραμματισμός [43]) που υποστηρίζουν το πεδίο.

Βέβαια στην ενισχυμένη μάθηση υπάρχουν και οι εξής προκλήσεις:

- Η βέλτιστη πολιτική πρέπει να συνάγεται από αλληλεπίδραση δοκιμής και σφάλματος με το περιβάλλον. Το μόνο σήμα εκμάθησης που λαμβάνει ο πράκτορας είναι η ανταμοιβή
- Οι παρατηρήσεις του πράκτορα εξαρτώνται από τις ενέργειές του και αυτό μπορεί να περιέχει ισχυρούς χρονικούς συσχετισμούς.
- Οι πράκτορες πρέπει να αντιμετωπίζουν μακροχρόνιες εξαρτήσεις. Συχνά οι συνέπειες μιας πράξης πραγματοποιούνται μόνο μετά από πολλές μεταβάσεις του περιβάλλοντος. Αυτό είναι γνωστό ως το πρόβλημα εκχώρησης πίστωσης [37].

Κεφάλαιο 3

Δίκτυα

3.1 Εισαγωγή Κεφαλαίου

Σε αυτό το κεφάλαιο θα κάνουμε μία ανάλυση στις γενιές δικτύων. Συγκεκριμένα θα κάνουμε αναφορά σε όλες τις γενιές δικτύων που έχουν χρησιμοποιηθεί έως σήμερα. Αυτές είναι οι:

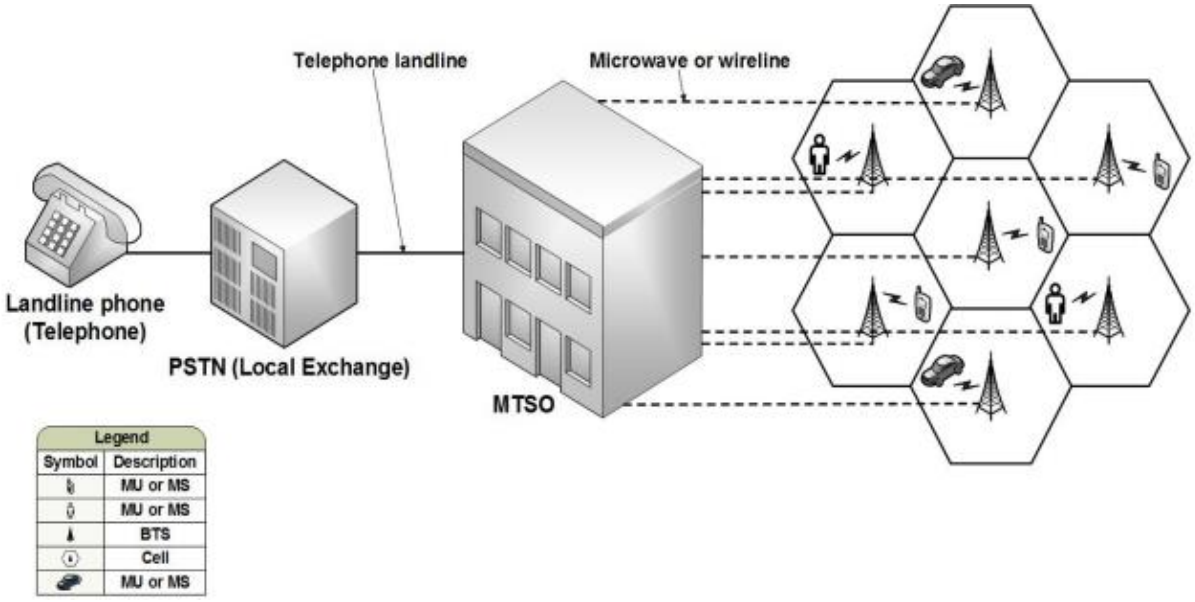
- **Πρώτη γενιά δικτύων (1G)**
- **Δεύτερη γενιά δικτύων (2G)**
- **Τρίτη γενιά δικτύων (3G)**
- **Τέταρτη γενιά δικτύων (4G)**
- **Πέμπτη γενιά δικτύων (5G)**

Πέρα από τα δίκτυα που έχουν χρησιμοποιηθεί, θα κάνουμε και αναφορά στην ασφάλεια των δικτύων. Η ασφάλεια δικτύων αποτελεί ένα πολύ σημαντικό ζήτημα για αρκετούς λόγους οι οποίοι ποικίλουν από προσωπικά δεδομένα έως και ομαλή λειτουργία του δικτύου.

3.2 Δίκτυα Πρώτης Γενιάς

Τα δίκτυα πρώτης γενιάς αντιπροσωπεύουν μία τεχνολογία αναλογικής μετάδοσης που έχει σχεδιαστεί για την παροχή βασικής υπηρεσίας μετάδοσης φωνής. Τα χρησιμοποιούμενα ραδιοφωνικά σήματα είναι αναλογικής φύσεως χωρίς δυνατότητες δεδομένων, αν και η ψηφιακή σηματοδότηση χρησιμοποιείται για την σύνδεση των ραδιοπύργων με τα υπόλοιπα τηλεφωνικά συστήματα με διαμόρφωση των τηλεφωνικών κλήσεων, σε υψηλότερη συχνότητα των περίπου 150 MHz. Με άλλα λόγια, η

πολυπλεξία διαίρεσης συχνότητας χρησιμοποιείται για τη διαίρεση του εύρους ζώνης σε συγκεκριμένες συχνότητες που αντιστοιχίζονται σε μεμονωμένες κλήσεις. Το μέγεθος κελιού για ένα τυπικό, πρώτης γενιάς δίκτυο, κυμαίνεται στα 2 έως 20 χιλιόμετρα [44], [45], [46]. Στην Εικόνα 8 φαίνεται ένα παράδειγμα ενός δικτύου πρώτης γενιάς.



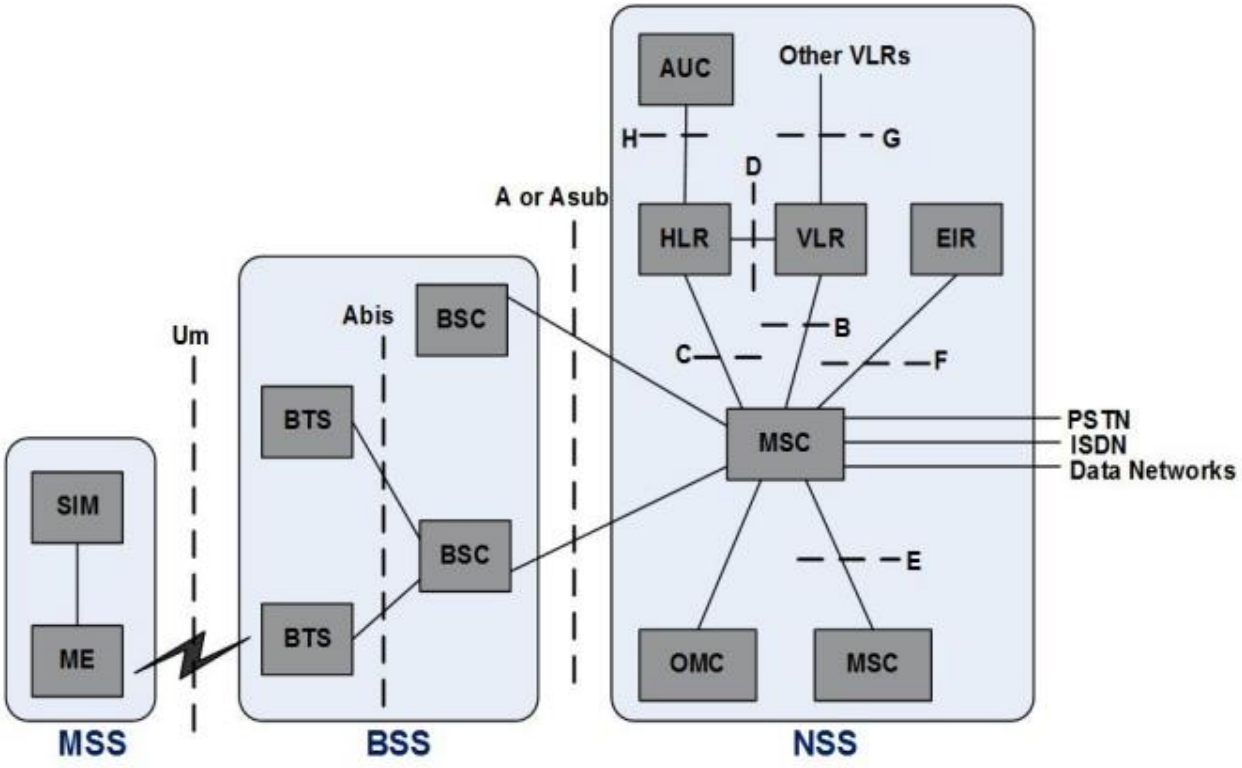
Εικόνα 8:Τυπική κυψελοειδής αρχιτεκτονική 1G AMPS(Advanced Mobile Phone System) [47]

Το γεγονός ότι τα δίκτυα πρώτης γενιάς βασίζονται σε αναλογικά σήματα, σημαίνει ότι ένα κοινό πρόβλημα είναι η ευαισθησία σε παρεμβολές, η οποία έχει ως συνέπεια την μείωση της ποιότητας της κλήσης. Επιπλέον υπάρχει και η ουσιαστική έλλειψη ασφάλειας, καθώς τα αναλογικά σήματα δεν επιτρέπουν την εφαρμογή προηγμένων μεθόδων κρυπτογράφησης. Η τεχνολογία επικοινωνίας των δικτύων πρώτης γενιάς είναι παγιδευμένη από τα εξής:

- Περιορισμένη χωρητικότητα (περιορισμένος αριθμός χρηστών)
- Μεγάλο μέγεθος τηλεφώνου
- Κακή ποιότητα φωνής
- Διάρκεια μπαταρίας
- Αξιοπιστία παράδοσης [45]

3.3 Δίκτυα Δεύτερης Γενιάς

Η τεχνολογία δικτύων δεύτερης γενιάς έφερε την ψηφιοποίηση στη δικτύωση κινητής τηλεφωνίας καθώς παρείχε τα πρώτα ψηφιακά συστήματα ως επικαλύψεις σε αναλογικά συστήματα. Τα δίκτυα αυτά μπόρεσαν να παρέχουν σημαντική βελτίωση της ποιότητας της φωνής και γέννησε τη πρώτη προσφορά δεδομένων, αν και περιορισμένη, στην εξέλιξη των κυψελωτών δικτύων (cellular network). Σε μεγάλο βαθμό, μέσω της χρήσης μιας πιο αποτελεσματικής κατανομής εύρους ζώνης/φάσματος μέσω πολλαπλών συστημάτων πρόσβασης, όπως πολλαπλή πρόσβαση διαίρεσης συχνότητας, ή πολλαπλή πρόσβαση διαίρεσης χρόνου, η επικοινωνία των δικτύων δεύτερης γενιάς ήταν πολύ επιτυχημένη και εξαιρετική για εφαρμογές μετάδοσης φωνής [48]. Το τελευταίο δε θα μπορούσε να είναι μόνο ψηφιακά κρυπτογραφημένο, αλλά και ικανό να παρέχει μία ασφαλή υπηρεσία σύντομου μηνύματος (short message service SMS) καθώς και ασφαλή υπηρεσία μηνυμάτων πολυμέσων (multimedia messaging service MMS), υπηρεσίες ικανές για να ξεπεραστούν ορισμένοι από τους περιορισμούς των δικτύων πρώτης γενιάς. Επιπλέον τα δίκτυα αυτά ήταν ικανά να παρέχουν ένα ημι-παγκόσμιο σύστημα περιαγωγής για την προώθηση της συνδεσιμότητας σε όλο τον κόσμο, ένα κατόρθωμα που δεν μπόρεσε να επιτευχθεί με τα δίκτυα πρώτης γενιάς. Στην πράξη, το παγκόσμιο σύστημα δικτύων δεύτερης γενιάς υποστηρίζει προδιαγραφές για κινητές επικοινωνίες (Global System for Mobile Communications GSM) με μεγέθη κελιών έως και 35 χιλιομέτρων χρησιμοποιώντας macro, micro, pico ή femto κελιά [46]. Στην Εικόνα 9 φαίνεται μία τυπική αρχιτεκτονική GSM δεύτερης γενιάς.



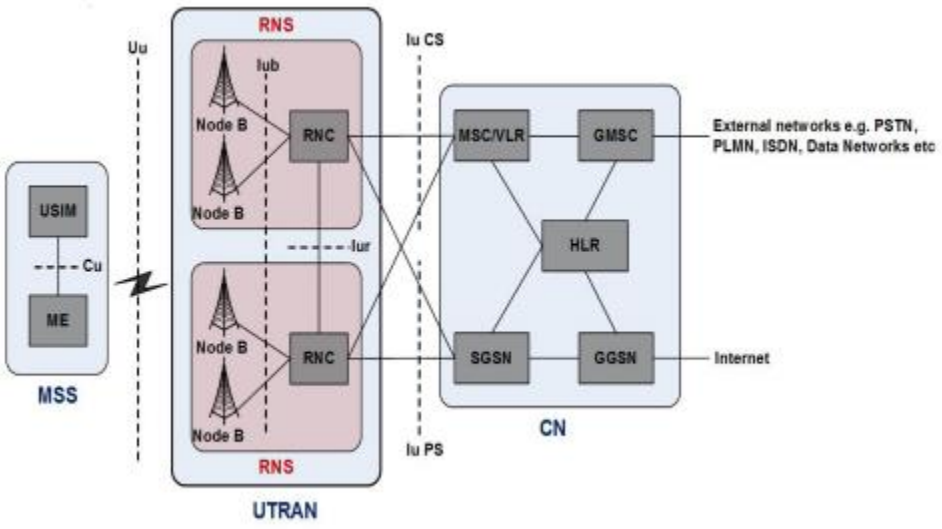
Εικόνα 9: GSM αρχιτεκτονική δικτυών δεύτερης γενιάς [49]

Η εποχή των δικτύων δεύτερης γενιάς εξελίχθηκε δραστικά από την GSM σε γενική υπηρεσία πακέτων ασύρματης διεπαφής (General Packet Radio Service GPRS) και βελτιωμένους ρυθμούς δεδομένων για την εξέλιξη του GSM (EDGE) το 1999, εν μέρει λόγω της αχόρταγης φύσης των χρηστών που πάντα ήθελαν περισσότερα από την άποψη των υπηρεσιών δεδομένων, της ποιότητας των υπηρεσιών (quality of service QoS) και τις ταχύτητες διεκπερέωσης [48], [50]. Παρόλα αυτά όμως τα δίκτυα δεύτερης γενιάς είχαν και αυτά τα μειονεκτήματά τους, όπως ζητήματα παρεμβολών (λόγω επαναχρησιμοποίησης συχνότητας [48]), ή διαλείπουσα διακοπή κλήσης ή ακόμα και πλήρη αποτυχία λόγω της παλμικής φύσης της πολλαπλής πρόσβασης με διαίρεση χρόνου και της καμπύλης γωνιακής αποσύνθεσης κάτω από δυσμενές έδαφος, τοπογραφικές ή ηλεκτρομαγνητικές συνθήκες.

Σε επίπεδο ασφάλειας, ο έλεγχος ταυτότητας, η κρυπτογράφηση και η ανωνυμία αποτελούν τις βασικές πτυχές της παροχής ασφάλειας στα GSM. Παρόλο όμως που παρέχεται έλεγχος ταυτότητας και κρυπτογράφηση μέσω αλγοριθμικών μηχανισμών, υπάρχει μια σειρά από μειονεκτήματα όπως ελαττώματα κρυπτογράφησης, επίθεση υποκλοπής, ψευδής ή ψεύτικος σταθμός βάσης, άρνησης υπηρεσίας κ.ο.κ. Αυτές είναι γνωστό ότι είναι ελλείψεις των ρυθμίσεων ασφαλείας στα GSM [49], [51], [52].

3.4 Δίκτυα Τρίτης Γενιάς

Τα δίκτυα τρίτης γενιάς παρέχουν αποκλειστικά ψηφιακά δίκτυα που χρησιμοποιούνται για την παράδοση ευρυζωνικών υπηρεσιών όπως και υπηρεσίες πολυμέσων. Στην Εικόνα 10 φαίνεται η αρχιτεκτονική της καθολικής υπηρεσίας τηλεπικοινωνιών (Universal Mobile Telecommunication Service UMTS)



Εικόνα 10: Αρχιτεκτονική UMTS δικτύων τρίτης γενιάς [53]

Καθοδηγούμενη, εν μέρη από την πρόοδο στην διαδικτυακή τεχνολογία καθώς και στη τεχνολογία δικτύου IP, η αρχιτεκτονική των δικτύων τρίτης γενιάς παρέχει υποστήριξη για βελτιωμένο ρυθμό μετάδοσης δεδομένων και ποιότητα υπηρεσιών. Υπηρεσίες όπως η παγκόσμια περιαγωγή (global roaming) και η βελτιωμένη ποιότητα φωνής αποτελούν σημαντικά επιτεύγματα της τεχνολογίας δικτύων τρίτης γενιάς. Ένα μειονέκτημα της τεχνολογίας αυτής είναι ότι στον τομέα της αποδοτικότητας ενέργειας, ο εξοπλισμός του χρήστη καταναλώνει σημαντικά περισσότερη ενέργεια σε σύγκριση με τα περισσότερα μοντέλα των δικτύων δεύτερης γενιάς. Ένα άλλο μειονέκτημα είναι ότι είναι λιγότερο οικονομική η εγκατάστασή του, όπως και η συντήρηση του δικτύου τρίτης γενιάς, σε αντίθεση φυσικά με τη προηγούμενη γενιά δικτύων [50], [53], [54], [55].

Επιπλέον το UMTS τρίτης γενιάς δικτύων, είναι συμβατό με τις προηγούμενες γενιές ασυρμάτων κυψελοειδών τεχνολογιών μέσω της ικανότητας ύπαρξης σε ετερογένεια, με τη τεχνολογία GSM ή τη τεχνολογία AMPS. Η εξέλιξη από το UMTS μέσω της πρόσβασης πακέτων υψηλής ταχύτητας (high speed packet access HSPA) και του

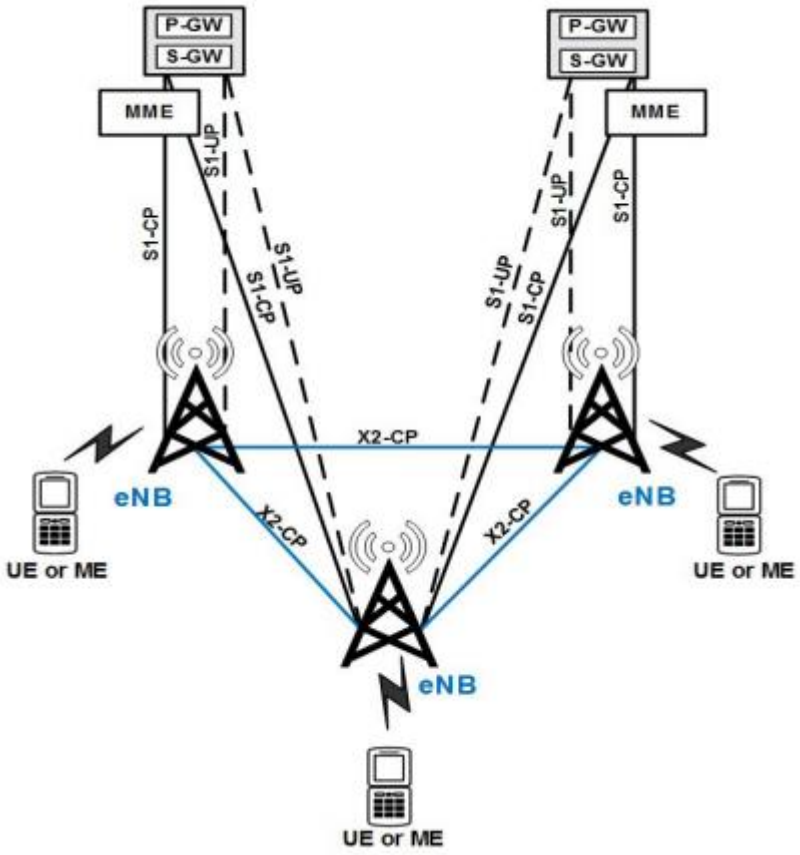
εξελιγμένου HSPA (HSPA+) παρείχε περαιτέρω σημαντικά ενισχυμένη απόδοση σε από άκρο σε άκρο δίκτυο (end-to-end network) και τελικά οδήγησε στην ανάπτυξη της επόμενης γενιάς δικτύων, η οποία είναι η τέταρτη γενιά δικτύων.

Από άποψη ασφαλείας στα δίκτυα τρίτης γενιάς γίνεται χρήση των εξής χαρακτηριστικών του GSM που αναφέρονται ενδεικτικά σύμφωνα με το [56]:

- **Έλεγχος ταυτότητας Βάσει SIM**
- **Εμπιστευτικότητα των κινήσεων του χρήστη στην διεπαφή αέρα (air interface)**
- **Εμπιστευτικότητα της ταυτότητας χρήστη στη διεπαφή αέρα**

3.5 Δίκτυα Τέταρτης Γενιάς

Τα δίκτυα τέταρτης γενιάς αντιπροσωπεύουν τη γενιά της τεχνολογίας κινητής τηλεφωνίας που αναμένεται να ανταποκρίνεται παραγωγικά στις απαιτήσεις για ευρυζωνική μετάδοση δεδομένων [54]. Από αρχιτεκτονική άποψη, στα δίκτυα τέταρτης γενιάς η μακροπρόθεσμη εξέλιξη δικτύου (long term evolution LTE) έχει σχεδιαστεί με στόχο την παροχή υποστήριξης για κίνηση με μεταγωγή πακέτων (packet- switched traffic) με απρόσκοπτη κινητικότητα, ποιότητα υπηρεσιών και ελάχιστη καθυστέρηση. Αυτή η προσέγγιση επιτρέπει την υποστήριξη όλων των υπηρεσιών (δεδομένα, φωνή, πολυμέσα) μέσω πακέτων σύνδεσης (packet connections) [57]. Χρησιμοποιώντας τον ενισχυμένο κόμβο B (enchanced node B eNB), την οντότητα διαχείρισης κινητικότητας (mobility management entity MME) και την system architecture evolution gateway (SAE GW) βγαίνει μία εξαιρετικά βελτιωμένη αρχιτεκτονική που μπορεί να οριστεί για το LTE όπως απεικονίζεται στην Εικόνα 11 [57], [58], [59].



Εικόνα 11:Αρχιτεκτονική LTE τέταρτης γενιάς δικτύων [57], [58], [60]

Στην Εικόνα 11, ο εξοπλισμός χρήστη ή το κινητό τηλέφωνο είναι συνδεδεμένο ασύρματα στο eNB. Όλα τα πρωτόκολλα της ασύρματης διεπαφής, διαχείριση κινητικότητας, συμπίεση κεφαλίδων, κρυπτογράφηση, αξιόπιστη παράδοση πακέτων και όλες οι αναμεταδόσεις πακέτων ενορχηστρώνονται από το eNB, ως ελεγκτής ασύρματης διεπαφής δικτύου (radio network controller RNC) [57], [58].

Η δυνατότητα επίτευξης ελάχιστης καθυστέρησης, προόδους σε τεχνικές πολλαπλής εισόδου, πολλαπλής εξόδου (multiple input multiple output MIMO) μέσω διάφορων τεχνολογιών πρόσβασης ασύρματης διεπαφής (radio access technologies RAT) όπως για παράδειγμα πολλαπλή πρόσβαση με ορθογώνια διαίρεση συχνότητας (orthogonal frequency division multiple access OFDMA) και η πολλαπλή πρόσβαση με διαίρεση συχνότητας ενός φορέα (single carrier frequency division multiple access SCFDMA) αποτελούν μέρος των βασικών απαιτήσεων και των κατορθωμάτων της τεχνολογίας των δικτύων τέταρτης γενιάς. Πολυάριθμες καινοτόμες ιδέες όπως η συνάθροιση φορέα (carrier aggregation), η αναμετάδοση και συντονισμένη, πολλαπλών σημείων, μετάδοση ή/και λήψη σχεδιάστηκε και υλοποιήθηκε τελικά για να παρέχει σημαντικά

βελτιωμένους ρυθμούς αιχμής δεδομένων, να υποστηρίζει ετερογενή ανάπτυξη δικτύου και ευελιξία φάσματος μεταξύ άλλων δυνατοτήτων. Αυτά αποτελούν τα υπέροχα επιτεύγματα που παρέχονται από την τεχνολογία δικτύων τέταρτης γενιάς [58], [61].

Άλλες σημαντικές βελτιώσεις που έφερε η εποχή των δικτύων τέταρτης γενιάς βρίσκονται στους τομείς της πολλαπλής μετάδοσης και του μετριασμού των παρεμβολών. Το σχέδιο συνεργασίας 3ης γενιάς (The 3rd Generation Partnership Project 3GPP) ορίζει πολλές βασικές δυνατότητες και απαιτήσεις για ενεργειακή απόδοση, LTE για υπηρεσίες δημόσιας ασφάλειας, υπηρεσίες έκτακτης ανάγκης και τοποθεσίας και πολλαπλές εκπομπές και υπηρεσίες multicasting, δηλαδή με λίγα λόγια εξελιγμένη μετάδοση πολυμέσων και υπηρεσία πολλαπλής εκπομπής [62], [63], [64].

Το CCI (Communication Components Inc) ήταν ένα σημαντικό εμπόδιο για την επίτευξη υψηλότερης χωρητικότητας στα δίκτυα κινητής τηλεφωνίας. Ωστόσο, η εξέλιξη από τα δίκτυα πρώτης γενιάς έως της τέταρτης έχει προκαλέσει διάφορες παρεμβολές σε συστήματα συντονισμού και δέκτες με επίγνωση παρεμβολών, με στόχο τον μετριασμό του CCI. Αυτά έχουν δώσει πολλά υποσχόμενες βελτιώσεις απόδοσης σε σύγκριση με την προβολή δεκτών CCI ως προσθετικός λευκός Gaussian θόρυβος. Συγκεκριμένα το 3GPP εφαρμόζει μία τεχνική γνωστή ως ακύρωση παρεμβολών υποβοηθούμενη από το δίκτυο και καταστολή (network assisted interference cancellation and suppression NAICS), μέσω σημαντικών βελτιώσεων στον μετριασμό παρεμβολών στη πλευρά του δέκτη. Αυτό επιτυγχάνεται με τη χρήση προηγμένων δεκτών για την αύξηση του βαθμού ευαισθητοποίησης σχετικά με παρεμβολές μετάδοσης με πιθανή βοήθεια στο δίκτυο [62], [63], [64].

Αξίζει να αναφέρουμε ότι άλλες ασύρματες τεχνολογίες που εξελίχθηκαν γύρω από την εποχή των δικτύων δεύτερης, τρίτης και τέταρτης γενιάς περιλαμβάνουν τα Bluetooth, Wi-Fi, WiMAX και ZigBee μεταξύ άλλων. Από τη δεύτερη γενιά και της οικογένειας προτύπων 3GPP υπάρχει διαλειτουργικότητα με προηγούμενες γενιές τεχνολογιών κινητής, κυψελοειδούς ασύρματης σύνδεσης. Αυτές αποτελούν θεμελιώδη αρχή στο σχεδιασμό, στην ανάπτυξη και στην εξάπλωση. Αυτό αντικατοπτρίζεται από το γεγονός ότι το 4G είναι συμβατό προς τα πίσω και ενσωματώνεται με διάφορες τεχνολογίες ασύρματων επικοινωνιών που κυμαίνονται από GSM δικτύων δεύτερης γενιάς, σε Wi-Fi έως και ZigBee.

Όσο αναφορά την ασφάλεια στα δίκτυα τέταρτης γενιάς, σε αυτά γίνεται χρήση των εξής στοιχείων που αναφέρονται ενδεικτικά σύμφωνα με το [65]:

- **Έλεγχος ταυτότητας και συμφωνία κλειδιού**

- **Εμπιστευτικότητα και ακεραιότητα της σηματοδότησης**
- **Εμπιστευτικότητα του πεδίου χρήστη**

3.6 Δίκτυα Πέμπτης Γενιάς

Τα δίκτυα τέταρτης γενιάς κάποτε επισημάνθηκαν ως τα δίκτυα της επόμενης γενιάς, ενώ τα δίκτυα πέμπτης γενιάς έχουν πρόσφατα αναφερθεί ως το δίκτυο του μέλλοντος, σύμφωνα με ορισμένους μελετητές του χώρου [66]. Σύμφωνα με το [67] στα δίκτυα πρώτης γενιάς θεμελιώθηκε η κινητή τηλεφωνία, ενώ στα δίκτυα δεύτερης γενιάς η κινητή τηλεφωνία έγινε διαθέσιμη για όλους. Αντίστοιχα, στα δίκτυα τρίτης γενιάς πραγματοποιήθηκε η θεμελίωση της κινητής ευρυζωνικότητας και η εξέλιξη αυτού έγινε η ημερήσια διάταξη στα δίκτυα τέταρτης γενιάς.

Τα δίκτυα πέμπτης γενιάς είναι ικανά να συνδέουν τα πάντα, παρέχοντας απρόσκοπτη συνδεσιμότητα, συγχωνεύοντας το ύφασμα συνδεσιμότητας για τουλάχιστον την επόμενη δεκαετία και πιθανώς ακόμα περισσότερο [68], [69], [70]. Με άλλα λόγια, τα δίκτυα πέμπτης γενιάς παρέχουν απεριόριστη πρόσβαση οπουδήποτε, ανά πάσα στιγμή, για οποιονδήποτε και για οτιδήποτε [71]. Αυτό είναι επειδή αυτή η τεχνολογία, δημιουργεί μία ενοποιημένη διεπαφή αέρα (air interface) στην ίδρυση μίας σύνδεσης από άκρο σε άκρο (end-to-end) μεταξύ πραγμάτων της καθημερινότητας όπως το smartphone, ψυγείο, καταψύκτη, βοηθητικούς μετρητές και πολλά άλλα [72], [73], [74].

Για να το θέσουμε συνοπτικά με τεχνικούς όρους, τα δίκτυα πέμπτης γενιάς φέρνουν ένα κόσμο με αισθητά βελτιωμένη ευρυζωνική σύνδεση δεδομένων κινητής τηλεφωνίας, υπεραπόκριση, εξαιρετική αξιοπιστία, εξαιρετικά χαμηλή καθυστέρηση, εξαιρετικά γρήγορους ρυθμούς δεδομένων και τεράστιες δυνατότητες IoT. Στα δίκτυα πέμπτης γενιάς, τα βασικά συστατικά της κατανομής πόρων ασύρματης διεπαφής, συμπεριλαμβανομένου της καθυστέρησης, της απόδοσης, της αξιοπιστίας, της ποιότητας των υπηρεσιών και της εμπειρίας χρήστη είναι σημαντικά βελτιστοποιημένα σε εντελώς πρωτόγνωρα επίπεδα.

Πιο συγκεκριμένα τα δίκτυα πέμπτης γενιάς είναι ικανά να [75]:

- Παρέχουν ελάχιστη έως μηδαμινή καθυστέρηση και καθηλωτική εμπειρία πολυμέσων.
- Μπορούν να ανταπεξέρχονται σε αιτήματα εξαιρετικά υψηλής χωρητικότητας και απόδοσης.

- Παρέχουν συνοχή υπηρεσιών (υψηλής διαθεσιμότητας) σε τραίνα και στις αραιές αλλά και στις πυκνές περιοχές
- Υποστήριξη συνδεσιμότητας για συσκευές χρήστη και IoT συσκευές που ξεπερνούν τα είκοσι εκατομμύρια και ένα τρισεκατομμύρια αντίστοιχα, σε ή πολύ κοντά, εκατό τις εκατό αξιοπιστία.

Μερικά από τα βασικά χαρακτηριστικά για τα δίκτυα πέμπτης γενιάς θα συζητηθούν παρακάτω:

1. **Εξαιρετικά γρήγορος ρυθμός μετάδοσης δεδομένων:** Ο μέγιστος ρυθμός δεδομένων αναφέρεται στον μέγιστο δυνατό ρυθμό δεδομένων υπό ιδανικές συνθήκες, δηλαδή χωρίς σφάλματα στα ληφθέντα bit δεδομένων που μπορούν να μεταφερθούν υπό τη μέγιστη χρήση πόρων ασύρματης διεπαφής. Μαθηματικά ο μέγιστος ρυθμός δεδομένων χρήστη ορίζεται ως το προϊόν του καναλιού BW και του PSE σύμφωνα με το [76]. Ο μέγιστος ρυθμός δεδομένων που μπορεί να επιτευχθεί στα δίκτυα πέμπτης γενιάς είναι θεωρητικά έως 1 Tbps, αν και αυτό αναμένεται να γίνει πραγματικότητα γύρω στο 2030 [77].
2. **Εξαιρετικά χαμηλή καθυστέρηση:** Εξ ορισμού, η καθυστέρηση αναφέρεται στον χρόνο, για τη μετάδοση δεδομένων από μία συσκευή, και την χωρίς σφάλματα λήψη των δεδομένων από άλλη συσκευή. Ευρέως η καθυστέρηση έχει τέσσερα συσχετισμένα στοιχεία, την καθυστέρηση μετάδοσης, την καθυστέρηση ουράς, την καθυστέρηση διάδοσης και την καθυστέρηση επεξεργασίας [78].
3. **Εξαιρετική Αξιοπιστία:** Σύμφωνα με το [79], η αξιοπιστία μπορεί να θεωρηθεί ως τη διαθεσιμότητα ή την παροχή συγκεκριμένου επιπέδου υπηρεσιών περίπου το 100% του χρόνου. Στο πλαίσιο της κυτταρικής ασύρματης δικτύωσης, το [76] ορίζει την αξιοπιστία ως την ικανότητα της μετάδοσης μιας συγκεκριμένης ποσότητας κίνησης μέσα σε μια προκαθορισμένη χρονική περίοδο με μεγάλη πιθανότητα επιτυχίας. Αυτό είναι παρόμοιο με τον ορισμό που παρέχεται στο [67], ο οποίος ορίζει την αξιοπιστία ως ικανή να εγγυηθεί μια επιτυχημένη παράδοση μηνυμάτων εντός ενός καθορισμένου χρόνου.
4. **Υπερσυνδεσιμότητα:** Μεγάλη γκάμα νέων εφαρμογών, συσκευών όπως φορητές συσκευές, έξυπνες πόλεις, σπίτια καταστήματα, αυτοκίνητα και πολλά άλλα, θα απολαμβάνουν απρόσκοπτα, πανταχού παρούσα ασύρματη συνδεσιμότητα μέσω των δικτύων πέμπτης γενιάς [69].
5. **Εξαιρετική απόκριση:** Η ανταπόκριση αναφέρεται σε μία χρονική μέτρηση της ικανότητας ενός στοιχείου, συστήματος ή μίας ολόκληρης μονάδας να ολοκληρώσει μία καθορισμένη εργασία. Για να φέρουμε μία εντελώς

καινούργια διάσταση στην επικοινωνία ανθρώπου με μηχανή σε κυψελοειδή δικτύωση πέμπτης γενιάς, προκύπτει η ανάγκη για μια εξαιρετικά αποκριτική συνδεσιμότητα [75].

6. **Υπερπύκνωση:** Η πυκνότητα δικτύου είναι μία πολλά υποσχόμενη κυτταρική τεχνική που αξιοποιεί τη χωρική επαναχρησιμοποίηση για την ενίσχυση της κάλυψης και διεκπεραίωση για το κυψεολοειδές δίκτυο πέμπτης γενιάς. Η πύκνωση δικτύου έχει να κάνει με την προσθήκη σημείων πρόσβασης και αξιοποιώντας τη χωρική επαναχρησιμοποίηση του φάσματος , βελτιώνοντας τη χωρητικότητα του δικτύου [80].

Οι επιθέσεις ασφαλείας στα δίκτυα πέμπτης γενιάς, μπορούν να ταξινομηθούν σε δύο τύπους οι οποίοι είναι οι παθητικές και οι ενεργητικές επιθέσεις [81]. Σε μία παθητική επίθεση οι εισβολείς προσπαθούν να μάθουν ή να χρησιμοποιήσουν πληροφορίες από τους νόμιμους χρήστες αλλά δε σκοπεύουν να επιτεθούν στην επικοινωνία. Σε αντίθεση με τις παθητικές επιθέσεις, οι ενεργητικές επιθέσεις μπορούν να τροποποιούν τα δεδομένα ή και ακόμα να διακόπτουν τις νόμιμες διαβιβάσεις. Αντίστοιχα οι μηχανισμοί που χρησιμοποιούνται για την αντιμετώπιση επιθέσεων ασφαλείας χωρίζονται και αυτοί σε δύο κατηγορίες, τις κρυπτογραφικές προσεγγίσεις με νέα πρωτόκολλα δικτύωσης και τις προσεγγίσεις ασφάλειας φυσικού επιπέδου σύμφωνα με το [82].

Κεφάλαιο 4

Περιγραφή Μηχανισμού

4.1 Εισαγωγή Κεφαλαίου

Στο παρόν κεφάλαιο θα παρουσιάσουμε μία προσομοίωση ενός δικτύου που χρησιμοποιεί την device to device (συσκευή σε συσκευή) επικοινωνία. Παράλληλα σε αυτή τη προσομοίωση με χρήση τεχνικών μηχανικής μάθησης θα εντοπίζουμε πιθανούς κακόβουλους χρήστες μέσα σε αυτό το δίκτυο.

Από την μηχανική μάθηση γίνεται χρήση της τεχνικής των Decision Trees (Δέντρων Αποφάσεων) καθώς θέλουμε να μας γίνεται μία πρόβλεψη για τον κάθε χρήστη, εάν δηλαδή είναι κακόβουλος ή όχι, μέσα στο δίκτυό μας.

Στην device to device επικοινωνία υπάρχουν αρκετές απειλές ασφαλείας, από τις οποίες οι ακόλουθες, αποτελούν τις κυριότερες αλλά υπάρχουν και στην προσομοίωσή μας [83]:

- **Επίθεση υποκλοπής (Eavesdropping attack)**
- **Επίθεση πλαστοπροσωπίας (Impersonate attack)**
- **Επίθεση πλαστογραφίας (Forge attack)**
- **Επίθεση ελεύθερης Πρόσβασης (Free-riding attack)**
- **Ενεργή επίθεση σε δεδομένα ελέγχου (Active attack on control data)**
- **Παραβίαση απορρήτου (Privacy violation)**
- **Άρνηση παροχής υπηρεσιών (Denial of Service)**

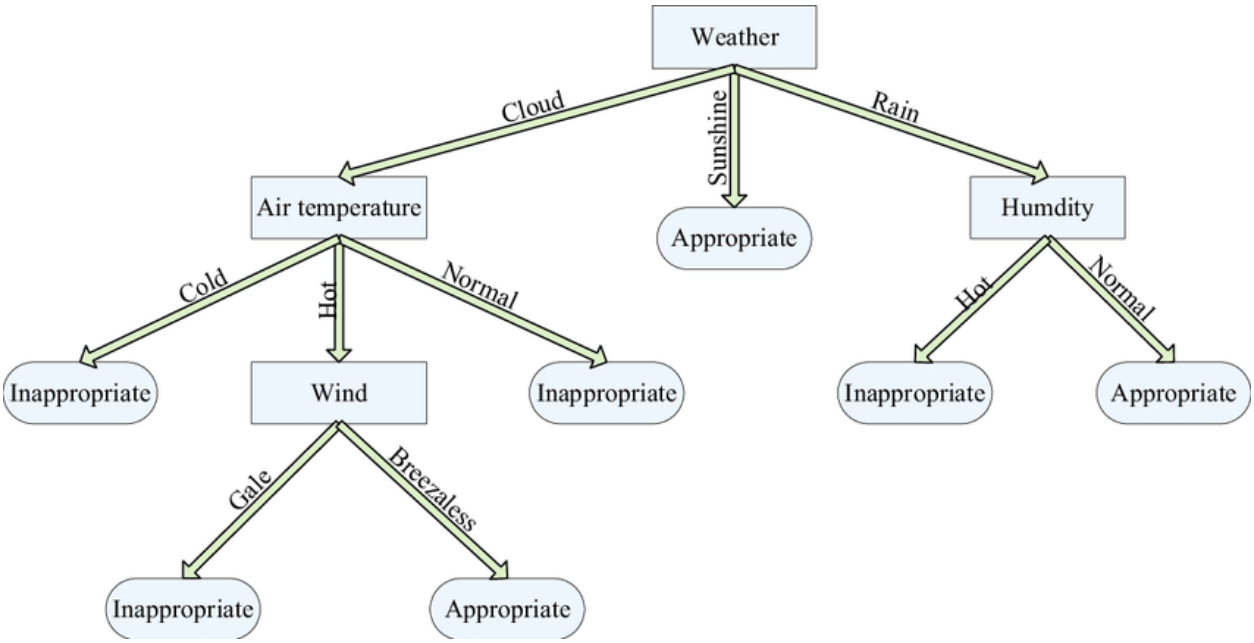
Παράλληλα η προσομοίωση μας, που θα περιγραφεί σε αυτό το κεφάλαιο, συνδυάζει την τεχνική των Decision Trees αλλά και τις απειλές ασφαλείας στην επικοινωνία device to device.

4.2 Μοντέλα Μηχανικής Μάθησης

Η Μηχανική Μάθηση αποκτά όλο και πιο σημαντικό ρόλο καθώς έχει μεγάλη ευελιξία μιας και μπορεί να εισαχθεί τόσο σε κάθε επιστημονικό κλάδο, όσο και σε όλους τους τομείς στην καθημερινότητά μας, όπως για παράδειγμα στην βελτίωση της διαδικτυακής αναζήτησης ή ακόμα και στην εξοικονόμηση ενέργειας εικονικής μηχανής [84]. Στη διπλωματική αυτή έγινε χρήση της τεχνικής των Decision Trees οπότε παρακάτω ακολουθεί η ανάλυση της τεχνικής αυτής.

4.2.1 Δέντρα Αποφάσεων

Το πιο διαδεδομένο και πιο πολυχρησιμοποιημένο εργαλείο για πραγματοποίηση πρόβλεψης είναι αυτό των Decision trees. Αποτελεί ένα εργαλείο που για να πάρει μία απόφαση δημιουργεί μία απεικόνιση που παραπέμπει σε δένδρο, άλλωστε αυτός είναι και ο λόγος της ονομασίας αυτής της τεχνικής. Στην παρακάτω εικόνα, Εικόνα 12, μπορούμε να δούμε και ένα παράδειγμα ενός Decision Tree.



Εικόνα 12 : Decision Tree [84]

Στο παραπάνω δένδρο βλέπουμε τη διαδικασία λήψης απόφασης, για το αν ο καιρός είναι κατάλληλος, που ακολουθεί η τεχνική των Decision Tree. Μπορούμε να διακρίνουμε ότι είναι ένα rooted tree, καθώς στην κορυφή του δένδρου βλέπουμε ότι ο κόμβος έχει μόνο εξερχόμενες ακμές και καμία εισερχόμενη ακμή. Οι υπόλοιποι κόμβοι διαθέτουν από μία εισερχόμενη και από τουλάχιστον μία εξερχόμενη ακμή. Βέβαια υπάρχουν και οι κόμβοι που διαθέτουν μόνο μία εισερχόμενη και καμία εξερχόμενη ακμή. Αυτοί ονομάζονται φύλλα του δένδρου και στη συγκεκριμένη περίπτωση είναι οι κόμβοι απόφασης του Decision Tree μας. Αυτό φαίνεται και στο σχήμα κιόλας καθώς αυτοί οι κόμβοι διαθέτουν την απόφαση, το αν δηλαδή ο καιρός είναι κατάλληλος (appropriate) ή ακατάλληλος (inappropriate).

4.2.2 Περιορισμοί των Decision Trees

Πέρα όμως από την απλότητα που παρέχουν, υπάρχουν και περιορισμοί στα Decision trees [84].

- Τα δένδρα αυτά μπορεί να μην είναι πολύ ανθεκτικά, καθώς μία μικρή αλλαγή στα δεδομένα εκπαίδευσης μπορεί να οδηγήσει σε μεγάλη αλλαγή στο δένδρο και, όπως είναι λογικό, στις τελικές προβλέψεις.
- Ένας άλλος περιορισμός είναι ότι οι Decision-tree learners (μαθητές) μπορούν να δημιουργήσουν υπερβολικά πολύπλοκα δένδρα που δεν μπορούν να γενικευτούν καλά από τα δεδομένα εκπαίδευσης (αυτό είναι γνωστό ως υπερπροσαρμογή). Μηχανισμοί όπως το κλάδεμα (pruning) είναι απαραίτητοι για την αποφυγή αυτού του προβλήματος.
- Το μέσο βάθος του δένδρου που ορίζεται από τον αριθμό των κόμβων ή ακόμα και των δοκιμών μέχρι την ταξινόμηση δεν είναι εγγυημένο ότι είναι ελάχιστο ή μικρό σύμφωνα με διάφορα κριτήρια διαχωρισμού.
- Για δεδομένα που περιλαμβάνουν κατηγορικές μεταβλητές με διαφορετικούς αριθμούς επιπέδων, το κέρδος πληροφοριών στα δένδρα αποφάσεων είναι προκατειλημμένο υπέρ των χαρακτηριστικών με τα περισσότερα επίπεδα.

4.3 Device To Device Επικοινωνία

Η επικοινωνία Device to Device αποτελεί μία πολλά υποσχόμενη τεχνολογία για την επόμενη γενιά κινητών δικτύων επικοινωνίας. Αναμένεται να επιτρέπει υψηλή απόδοση, μείωση των καθυστερήσεων επικοινωνίας αλλά και μείωση της κατανάλωσης ενέργειας και του κυκλοφοριακού φόρτου. Η τεχνολογία device to device θα ενισχύσει την ικανότητα και την απόδοση των παραδοσιακών κυτταρικών δικτύων. Παρόλα αυτά

όμως, τα ζητήματα ασφαλείας πρέπει να λαμβάνονται υπόψη σε όλους τους τύπους επικοινωνίας , και ειδικά όταν αυτή πρόκειται για ασύρματη επικοινωνία [83].

4.3.1 Ζητήματα Ασφαλείας: Απειλές Ασφαλείας

Η τεχνολογία device to device εισάγει διάφορες απειλές ασφαλείας, από τις οποίες οι κυριότερες αλλά και αυτές που χρησιμοποιούμε στη διπλωματική αυτή είναι [83]:

- ***Eavesdropping attack***: Η επίθεση υποκλοπής είναι όταν ένας εισβολέας ακούει παθητικά μεταξύ συσκευών εξοπλισμού χρηστών (οποιαδήποτε μέσο επικοινωνίας όπως για παράδειγμα κινητά, υπολογιστές κ.ο.κ) για να λάβει ευαίσθητα δεδομένα. Η εμπιστευτικότητα δεδομένων στη κρυπτογραφική προσέγγιση μπορεί να αντιμετωπίσει αυτήν την απειλή βέβαια.
- ***Impersonate attack***: Στην επίθεση μίμησης ο εισβολέας μπορεί να προσποιηθεί ότι είναι μία νόμιμη συσκευή εξοπλισμού χρηστών, για να αποκτήσει πρόσβαση στα δεδομένα κυκλοφορίας. Ο έλεγχος ταυτότητας στην κρυπτογραφική προσέγγιση μπορεί να αντιμετωπίσει αυτήν την απειλή.
- ***Forge attack***: Στην επίθεση πλαστογραφίας ο εισβολέας μπορεί να παραποιήσει το περιεχόμενο και να στείλει τα πλαστά δεδομένα στους υπόλοιπους χρήστες, προδικάζοντας έτσι το σύστημα. Η ακεραιότητα των δεδομένων (με λίγα λόγια η ψηφιακή υπογραφή) στην κρυπτογραφική προσέγγιση μπορεί να αντιμετωπίσει αυτήν την απειλή.
- ***Free riding attack***: Προκειμένου να μειωθεί το σύστημα διαθεσιμότητας στις device to device επικοινωνίες, ο εισβολέας μπορεί να ενθαρρύνει την εγωιστική συμπεριφορά ορισμένων χρηστών για τη διατήρηση της κατανάλωσης ενέργειας, ώστε να μην είναι πρόθυμοι να στείλουν τα περιεχόμενά τους σε άλλους ενώ αυτοί να λαμβάνουν τα δεδομένα από τους υπόλοιπους. Μια τέτοια αδυναμία είναι πιθανό να επηρεάσει την ποιότητα εμπειρίας του χρήστη. Η ανάπτυξη της συνεργασίας είναι απαραίτητη για να μειωθεί η εμφάνιση μιας τέτοιας επίθεσης.
- ***Active attack on control data***: Στην ενεργή επίθεση σε δεδομένα ελέγχου ο εισβολέας προσπαθεί να αλλάξει τα δεδομένα ελέγχου του συστήματος. Η αυθεντικοποίηση, η εμπιστευτικότητα και ακεραιότητα στην κρυπτογραφική προσέγγιση μπορεί να αντιμετωπίσει αυτήν την απειλή.
- ***Privacy violation***: Ορισμένα ευαίσθητα, ως προς το απόρρητο, δεδομένα όπως είναι η ταυτότητα ή και η τοποθεσία του χρήστη, είναι πιο σημαντικά στις device to device communications από τις λειτουργίες υπηρεσιών που παρέχουν. Οπότε αυτές οι προσωπικές πληροφορίες πρέπει να παραμένουν αποκρυμμένες σε μη εξουσιοδοτημένα άτομα.
- ***Denial of Service (DoS)***: Η επίθεση αυτή γίνεται με το να καθιστά, ο εισβολέας, μη διαθέσιμη μία από τις υπηρεσίες του device to device communication. Έχει δειχθεί ακόμα σε πειραματική μελέτη, ότι κακόβουλες συσκευές μπορούν να

βλάψουν κρυφά ή και ακόμα να μπλοκάρουν εντελώς τη σύνδεση νόμιμων συσκευών στο υποκείμενο δίκτυο [85].

4.3.2 Ζητήματα Ασφαλείας: Απαιτήσεις Ασφαλείας

Λόγω των παραπάνω απειλών ασφαλείας, ένα ασφαλές σύστημα device to device θα πρέπει να διαθέτει απαιτήσεις ασφαλείας, είτε αυτές είναι ελεγχόμενες είτε αυτόνομες. Αυτές ακολουθούν παρακάτω [83]:

- ***Αυθεντικοποίηση (Authentication)***: Η ταυτοποίηση των χρηστών που επικοινωνούν στο δίκτυο πρέπει να ελέγχεται για πιθανές κακόβουλες προσπάθειες.
- ***Εμπιστευτικότητα Δεδομένων (Data confidentiality)***: Τα μεταδιδόμενα δεδομένα μεταξύ των χρηστών θα πρέπει να παραμένουν κρυφά χρησιμοποιώντας τους κατάλληλους μηχανισμούς κρυπτογράφησης.
- ***Ακεραιότητα Δεδομένων (Data integrity)***: Όλα τα μεταδιδόμενα δεδομένα που μεταφέρονται από εξουσιοδοτημένους χρήστες θα πρέπει να ελέγχονται με σκοπό την επαλήθευση ότι δεν έχουν τροποποιηθεί με οποιονδήποτε τρόπο.
- ***Ιδιωτικότητα (Privacy)***: Όλα τα προσωπικά δεδομένα όπως για παράδειγμα η ταυτότητα η τοποθεσία κ.ο.κ θα πρέπει να διατηρούνται κρυφά.
- ***Ιχνηλασιμότητα (Traceability)*****:** Είναι απαραίτητο να είναι δυνατή η αναγνώριση της ταυτότητας μίας πηγής που μεταδίδει ψευδή μηνύματα.
- ***Ανωνυμία (Anonymity)***: Οι επιβεβαιωμένοι ή εξουσιοδοτημένοι χρήστες που επικοινωνούν, να μπορούν να παραμένουν ανώνυμοι μεταξύ τους.
- ***Μη αποκήρυξη (Non repudiation)***: Ουσιαστικά είναι στην ικανότητα αποτροπής από την άρνηση μετάδοσης ή λήψης μηνύματος από κάποιον χρήστη. Στην κρυπτογραφική προσέγγιση, ένα αποτελεσματικό εργαλείο για την αποτροπή της μη απόρριψης μετάδοσης, ενώ παράλληλα απαιτείται πρόσθετος μηχανισμός για την διασφάλιση της μη άρνησης λήψης, είναι αυτός της ψηφιακής υπογραφής.
- ***Διαθεσιμότητα (Availability)***: Οι υπηρεσίες των device to device θα πρέπει να είναι προσβάσιμες ανά πάσα στιγμή, ακόμα και κατά τη διάρκεια επιθέσεων του τύπου DoS ή free riding, έτσι ώστε να μην αποθαρρύνονται οι χρήστες από το να χρησιμοποιούν την τεχνολογία αυτή.
- ***Δυνατότητα ανάκλησης (Revocability)***: Είναι η ικανότητα αναστολής ενός χρήστη από κάποιο προνόμιο που διαθέτει σε κάποια υπηρεσία της device to device, εάν εντοπιστεί ως κακόβουλος.
- ***Λεπτόκοκκος Έλεγχος Πρόσβασης (Fine grained Access Control)***: Λαμβάνει υπόψη μικρή ευαισθησία λογαριασμού ενός καθορισμένου κανόνα πρόσβασης ενός χρήστη κατά την πρόσβασή του στην υπηρεσία. Θεωρείται ως μία αποτελεσματική λύση για την υπέρβαση του απορρήτου και των δεδομένων.

4.4 Περιγραφή του Κώδικα

Έχοντας λάβει τα παραπάνω υπόψιν, στο κομμάτι που ακολουθεί θα παρουσιαστεί ο κώδικας, ο οποίος είναι υλοποιημένος στη γλώσσα προγραμματισμού Python, που ουσιαστικά κάνει τη προσομοίωση ενός δικτύου όπου υπάρχουν χρήστες, κάποιοι είναι κακόβουλοι, και με χρήση μηχανικής μάθησης τους εντοπίζουμε όσο το δυνατόν καλύτερα μπορούμε.

Οι χρήστες μας βρίσκονται σε ένα dataset (users.csv), από όπου αντλούμε τις πληροφορίες που χρειαζόμαστε και τις τροφοδοτούμε στο decision tree μας για να τους αξιολογήσει έναν έναν.

Εκπαιδεύουμε το μοντέλο μας κάνοντας χρήση ενός δεύτερου dataset (trainusers.csv) της ίδιας μορφής με αυτό των χρηστών, απλά με την προσθήκη ότι είναι ήδη αξιολογημένοι, έτσι ώστε να μπορεί να πάρει απόφαση ο μηχανισμός.

Θέλουμε ο μηχανισμός μας να μπορεί να αναγνωρίζει έναν χρήστη ως κακόβουλο, όταν έχει διαπράξει μία από τις απειλές που περιγράψαμε προηγουμένως (για παράδειγμα να έχει κάνει forge attack, ή free ride attack κ.ο.κ).

4.4.1 Dataset Χρηστών προς αξιολόγηση

Το dataset μας, users.csv, το δημιουργούμε μέσα στον κώδικα με χρήση της συνάρτησης makeusers(x) η οποία δέχεται έναν ακέραιο αριθμό σαν είσοδο, που ουσιαστικά θα είναι ο συνολικός αριθμός των χρηστών μας. Στην Εικόνα 13 φαίνεται ο κώδικας αυτής της συνάρτησης:

```
7
8  def makeusers(x):
9      namelist = ['Zion','Kai',
10                 'Maeve','Luca','Nova','Mia','Aurora','Quinn','Ezra','Eliana','Ivy','Jayden','Amara',
11                 'Kayden','Lilibet','Isabella','Alina','Elliot','River','Xavier','Zoey','Isla','Lyla',
12                 'Alex','Molly','Andrea','Remi','Rowan','Elias','Alice','Hayden','Rohan','Ophelia','Kyle',
13                 'Jude','Mya','Shia','Cecilia','Milo','Finn','Leilani','Aria','Evan','Millie','Axel','Urban',
14                 'Amaya','Kayla','Jesse','Ian','Riley','Bailey','Julia','Blake','Ari','Savannah','Freya']
15     chance = [1,2,3,4,5,6,7,8,9,10]
16     table = ['0','1'] #No = 0 Yes = 1
17     filename = "users.csv"
18     fields = ['ID','Name','Eavesdropping attack','Impersonate attack','Forge attack',
19               'Free-riding attack','Active attack on control data','Privacy violation','Denial-of-Service']
20     with open(filename, 'w') as csvfile:
21         csvwriter = csv.writer(csvfile)
22         csvwriter.writerow(fields)
23         for i in range(x):
24             if(random.choice(chance)>1):
25                 row = [i+1,random.choice(namelist),'0','0','0','0','0','0','0']
26                 csvwriter.writerow(row)
27             else:
28                 row = [i+1,random.choice(namelist),random.choice(table),
29                        random.choice(table),random.choice(table),random.choice(table),
30                        random.choice(table),random.choice(table),random.choice(table)]
31                 csvwriter.writerow(row)
```

Εικόνα 13 : Κώδικας συνάρτησης makeusers(x)

Θέλουμε ο κάθε χρήστης μας να έχει ένα μοναδικό id, το οποίο δίνεται κατά την εισαγωγή του στο dataset ξεκινώντας από το ένα και φτάνει μέχρι και τον συνολικό αριθμό των χρηστών, για παράδειγμα εάν έχουμε συνολικά ‘250’ χρήστες ο πρώτος θα έχει το id ένα, δεύτερος το δύο και ο τελευταίος το ‘250’. Πέρα από το id, ο κάθε χρήστης έχει και ένα όνομα, το οποίο δεν είναι μοναδικό, αλλά με τυχαίο τρόπο δίνεται στον καθένα από μία λίστα ‘57’ διαφορετικών ονομάτων (namelist) . Κάνοντας χρήση του ονόματος και του id ο κάθε χρήστης αποκτάει μία μοναδικότητα. Ο κάθε χρήστης θα μπορεί να είναι είτε κακόβουλος, είτε καλοκάγαθος, που ουσιαστικά αυτό θέλουμε να εξακριβώσουμε, αλλά έχουμε σαν αρχή ότι ο χρήστης θεωρείται κακόβουλος εάν έχει διαπράξει τουλάχιστον μία από τις απειλές ασφάλειας στο σύστημα που περιγράψαμε προηγουμένως. Άρα στα συγκεκριμένα πεδία που δημιουργούμε (Eavesdropping attack, Impersonate attack κ.ο.κ), εάν δεν είναι κακόβουλος, τοποθετούμε σε όλα την τιμή ‘0’. Στην αντίθετη περίπτωση, που συμβαίνει με πιθανότητα 10%, με τυχαίο τρόπο επιλέγουμε κάποιο από αυτά τα πεδία να πάρει την τιμή ‘1’, που σημαίνει ότι το έχει διαπράξει το πεδίο που βρίσκεται. Ουσιαστικά αυτό το dataset αναπαριστά ένα στιγμιότυπο του δικτύου μας, όπου έχουν αντληθεί αυτά τα δεδομένα με αυτή τη μορφή με σκοπό την αξιολόγηση από το πρόγραμμά μας. Άρα το dataset μας έχει αυτήν την μορφή, που φαίνεται στην Εικόνα 14:

ID	Name	Eavesdropping attack	Impersonate attack	Forge attack	Free-riding attack	Active attack on control data	Privacy violation	Denial-of-Service
1	Ophelia	0	0	0	0	0	0	0
2	Freya	0	0	0	0	0	0	0
3	Mya	0	0	0	0	0	0	0
4	Nova	0	0	0	0	0	0	0
5	Aria	0	0	0	0	0	0	0
6	Alex	0	0	0	0	0	0	0
7	Eliana	0	0	0	0	0	0	0
8	Quinn	0	0	0	0	0	0	0
9	Ophelia	0	0	0	0	0	0	0
10	Cecilia	1	1	0	1	0	0	0
11	River	0	0	0	0	0	0	0
12	Ivy	0	0	0	0	0	0	0
13	Finn	0	0	0	0	0	0	0
14	Mia	0	0	0	0	0	0	0
15	Ari	0	0	0	0	0	0	0
16	Alice	0	0	0	0	0	0	0
17	Shia	0	0	0	0	0	0	0
18	Ari	0	0	0	0	0	0	0
19	Jude	0	0	0	0	0	0	0
20	Isabella	0	0	0	0	0	0	0

Εικόνα 14: Dataset users.csv

Όπως βλέπουμε ο χρήστης με το id ‘10’ και το όνομα ‘Cecilia’διαθέτει σε μερικά πεδία την τιμή ‘1’ άρα όταν αξιολογηθεί από το μοντέλο μας θα πρέπει να αναγνωριστεί ως ‘Malicious’ (κακόβουλος), ενώ οι υπόλοιποι ως ‘Benevolent’ (καλοκάγαθος).

4.4.2 Dataset Χρηστών για εκπαίδευση

Ο τρόπος που δημιουργείται αυτό το dataset είναι σχεδόν ίδιος με αυτόν του dataset των χρηστών με μικρές διαφορές όμως. Αυτές είναι ότι προστίθεται ένα ακόμα πεδίο το ‘Status’ το οποίο παίρνει τις τιμές ‘Malicious’ και ‘Benevolent’ ανάλογα με τον χρήστη σαφώς, και η δεύτερη αλλαγή είναι ότι η πιθανότητα εμφάνισης κακόβουλου χρήστη έχει αυξηθεί κατά 50% διότι θέλουμε η εκπαίδευση να γίνει με μεγαλύτερο ποσοστό κακόβουλων χρηστών από ότι η αξιολόγηση. Στην Εικόνα 15 φαίνεται ο κώδικας της συνάρτησης maketrainingusers(x) που υλοποιεί αυτή τη διαδικασία:

```
24
25  def maketrainingusers(x):
26      namelist = ['Zion','Kai','Maeve','Luca','Nova','Mia','Aurora',
27                  'Quinn','Ezra','Eliana','Ivy','Jayden','Amara', 'Kayden',
28                  'Lilibet','Isabella','Alina','Elliot','River','Xavier',
29                  'Zoey','Isla','Lyla','Alex','Molly','Andrea','Remi','Rowan',
30                  'Elias','Alice','Hayden','Rohan','Ophelia','Kyle','Jude','Mya',
31                  'Shia','Cecilia','Milo','Finn','Leilani','Aria','Evan','Millie',
32                  'Axel','Urban','Amaya','Kayla','Jesse','Ian','Riley','Bailey','Julia',
33                  'Blake','Ari','Savannah','Freya']
34      chance = [1,2,3,4,5,6,7,8,9,10]
35      table = ['0','1'] #No = 0 Yes = 1
36      filename = "trainusers.csv"
37      fields = ['ID','Name','Eavesdropping attack','Impersonate attack','Forge attack',
38                'Free-riding attack','Active attack on control data','Privacy violation',
39                'Denial-of-Service','Status']
40      with open(filename, 'w') as csvfile:
41          csvwriter = csv.writer(csvfile)
42          csvwriter.writerow(fields)
43          for i in range(x):
44              if(random.choice(chance)>=5):
45                  row = [i+1,random.choice(namelist),'0','0','0','0','0','0','0','Benevolent']
46                  csvwriter.writerow(row)
47              else:
48                  row = [i+1,random.choice(namelist),random.choice(table),
49                         random.choice(table),random.choice(table),random.choice(table),
50                         random.choice(table),random.choice(table),random.choice(table),'Malicious']
51                  csvwriter.writerow(row)
52
```

Εικόνα 15: Κώδικας συνάρτησης maketrainingusers(x)

Και κατά την εκτέλεσή του δημιουργείται το αρχείο trainusers.csv το οποίο φαίνεται στην Εικόνα 16:

	A	B	C	D	E	F	G	H	I	J
1	ID	Name	Eavesdropping attack	Impersonate attack	Forge attack	Free-riding attack	Active attack on control data	Privacy violation	Denial-of-Service	Status
2										
3	1	Urban	1	0	1	0	0	0	1	Malicious
4										
5	2	Jesse	0	0	0	0	0	0	0	Benevolent
6										
7	3	Evan	0	0	0	0	0	0	0	Benevolent
8										
9	4	Aria	0	0	0	0	0	0	0	Benevolent
10										
11	5	Ophelia	0	0	0	0	0	0	0	Benevolent
12										
13	6	Savannah	0	0	0	0	0	0	0	Benevolent
14										
15	7	Amaya	0	0	0	0	0	0	0	Benevolent
16										
17	8	Nova	0	0	0	0	0	0	0	Benevolent
18										
19	9	Maeve	1	1	1	1	1	0	0	Malicious
20										
21	10	Andrea	1	1	0	0	0	0	0	Malicious
22										
23	11	Riley	0	0	0	0	0	0	0	Benevolent
24										
25	12	Jesse	0	0	0	0	0	0	0	Benevolent
26										
27	13	Zion	0	1	1	1	1	1	0	Malicious
28										
29	14	Julia	0	0	1	0	1	1	1	Malicious
30										
31	15	Hayden	0	0	0	0	0	0	0	Benevolent
32										
33	16	Xavier	0	0	0	0	0	0	0	Benevolent
34										
35	17	Cecilia	0	0	0	0	0	0	0	Benevolent
36										
37	18	Kayla	0	0	0	0	0	0	0	Benevolent
38										
39	19	Rohan	0	0	0	0	0	0	0	Benevolent
40										
41	20	Jude	1	1	0	0	0	1	1	Malicious

Εικόνα 16: Dataset trainusers.csv

Που εδώ είναι ξεκάθαρο το πως διαχωρίζονται οι κακόβουλοι (Malicious) από τους καλοκάγαθους (Benevolent) χρήστες καθώς το πεδίο Status το κάνει ξεκάθαρο.

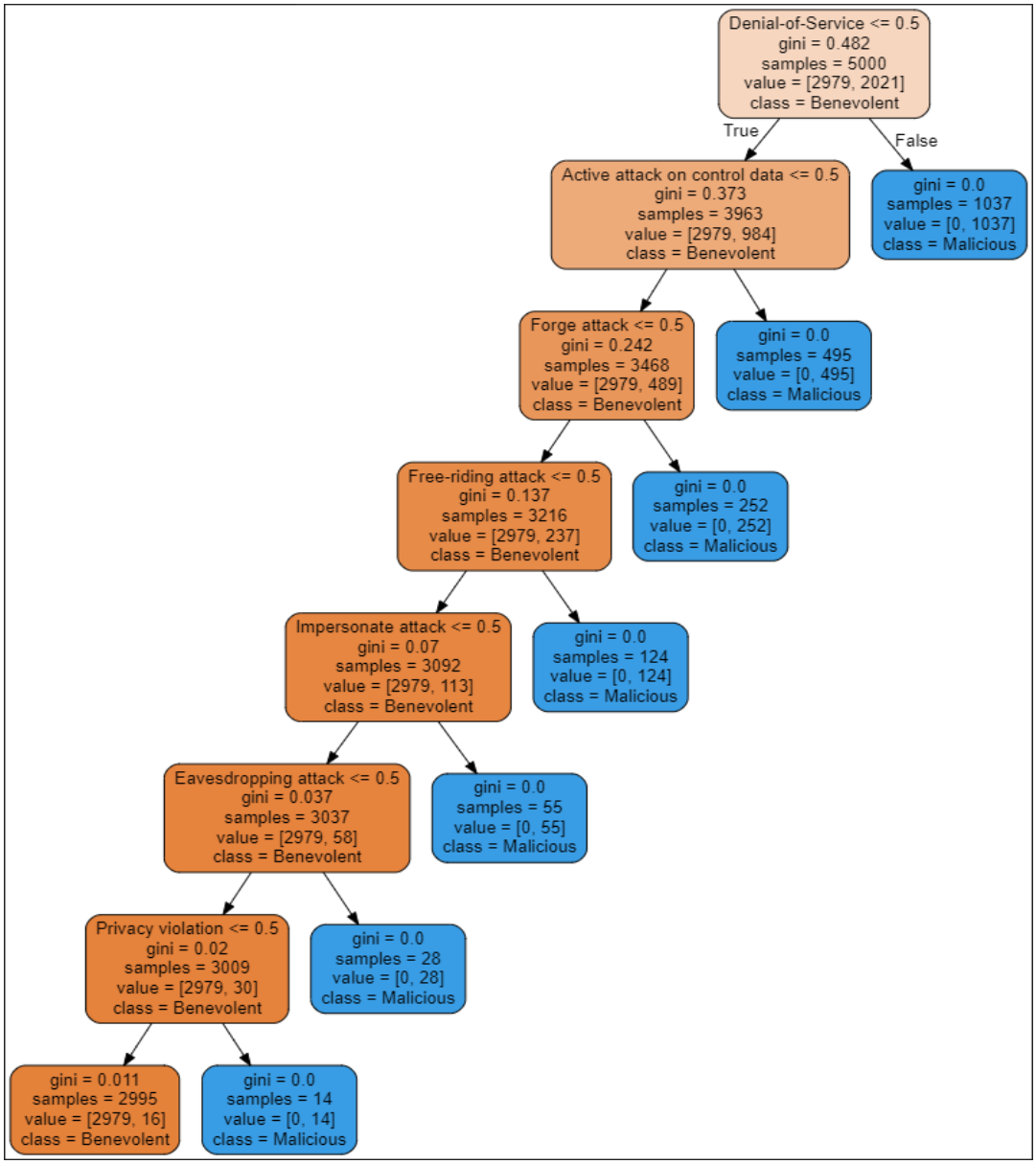
4.4.3 Εκπαίδευση και Πρόβλεψη του Μηχανισμού

Στόχος μας είναι να εκπαιδεύσουμε το μοντέλο μας, αξιοποιώντας τα στοιχεία του αρχείου trainusers.csv, έτσι ώστε να μπορεί στη συνέχεια να αξιολογήσει μόνο του τον κάθε χρήστη από το αρχείο users.csv, με όσο το δυνατόν μεγαλύτερη ακρίβεια γίνεται. Παρακάτω φαίνεται ο κώδικας που υλοποιεί την εκπαίδευση:

```
61      if trainflag:
62          #Training my model. Visualing my model
63          trainingdata = pd.read_csv("trainusers.csv")
64          X = trainingdata.drop(columns = ['Status','ID','Name'])
65          y = trainingdata['Status']
66          model = DecisionTreeClassifier()
67          model.fit(X, y)
```

Εικόνα 17: Εκπαίδευση του μοντέλου μας

Αρχικά φορτώνουμε τα δεδομένα εκπαίδευσης μας, χρησιμοποιώντας τη βιβλιοθήκη pandas. Μετά χωρίζουμε τα δεδομένα μας με τη λογική, ότι θέλουμε τις πιθανές απαντήσεις (στον y), που εδώ είναι 'Benevolent' και 'Malicious', αλλά και όλα τα υπόλοιπα δεδομένα που παίζουν ρόλο στη λήψη της απόφασης (στον X) , άρα όπως είναι λογικό τα περιεχόμενα των πεδίων 'ID', 'Name' και 'Status' δεν είναι απαραίτητα. Με αυτόν τον τρόπο το μοντέλο μας ψάχνει να βρει το μοτίβο στο γιατί ένας χρήστης είναι κακόβουλος ή καλοκάγαθος, και, όπως θα φανεί και παρακάτω, με μεγάλο δείγμα εκπαίδευσης το κάνει αρκετά αποτελεσματικά. Μόλις γίνει η εκπαίδευση, το decision tree μας έχει δημιουργηθεί, και για το δείγμα εκπαίδευσης των '5000' χρηστών είναι το εξής:



Εικόνα 18: Το Δέντρο Αποφάσεών μας

Το συγκεκριμένο γράφημα μας δείχνει, πως το Δέντρο Αποφάσεών μας, παίρνει τις αποφάσεις του. Για κάθε ένα από τα πεδία που έχουμε, ελέγχει εάν είναι μεγαλύτερα ή ίσα του '0.5', που οι τιμές είναι ή το '0' ή το '1', και αν ισχύει τότε ο χρήστης είναι κακόβουλος, διαφορετικά ελέγχει το αμέσως επόμενο πεδίο και με ίδιο τρόπο συνεχίζει στα επόμενα. Αυτή η 'λογική' που φτιάχνει η μηχανική μάθηση είναι και η επιθυμητή που θέλαμε να βρει καθώς και εμείς με ίδιο τρόπο αποφασίζουμε εαν ένας χρήστης είναι κακόβουλος ή όχι. Στον κώδικα που ακολουθεί, στην Εικόνα 19, φαίνεται ο τρόπος που δημιουργείται το συγκεκριμένο γράφημα, και με τη βοήθεια του VSCode έχουμε την Εικόνα 18:

```
68          #tree.export_graphviz(model,out_file = 'Malicious-Finder.dot',
69          #                      feature_names = ['Eavesdropping attack','Impersonate attack','Forge attack',
70                  #'Free-riding attack','Active attack on control data','Privacy violation','Denial-of-Service'],
71          #                      class_names = ['Benevolent','Malicious'],
72          #                      label = 'all',
73          #                      rounded = True,
74          #                      filled = True) #Creating Graph
75          #Training and visualizing finished
```

Εικόνα 19: Κώδικας για γράφημα δένδρου

Ο κώδικας είναι σε σχόλια διότι δεν είναι απαραίτητο να τον τρέχουμε σε κάθε εκτέλεση και αυτό επειδή σκοπός του ήταν να δούμε το γράφημα του δένδρου, για να καταλάβουμε αν η μηχανική μάθηση βρίσκει την λογική που θέλαμε μέσω των δειγμάτων εκπαίδευσης.

Στη λήψη απόφασης θα πρέπει να δώσουμε στο μοντέλο μας είσοδο κατάλληλα διαμορφωμένη με αυτή του δείγματος εκπαίδευσης, έτσι ώστε να ξέρει πως να πάρει απόφαση. Στη συγκεκριμένη περίπτωση θα πρέπει να είναι οι τιμές των πεδίων με τις απειλές ασφαλείας του κάθε χρήστη. Αυτό φαίνεται παρακάτω:

```
82              myid = data['ID'][p]
83              myname = str(data['Name'][p])
84              myeav = int(data['Eavesdropping attack'][p])
85              myimpers = int(data['Impersonate attack'][p])
86              myforge = int(data['Forge attack'][p])
87              myfre = int(data['Free-riding attack'][p])
88              myattd = int(data['Active attack on control data'][p])
89              mypv = int(data['Privacy violation'][p])
90              mydos = int(data['Denial-of-Service'][p])
91              mystatus = model.predict([[myeav,myimpers,myforge,myfre,myattd,mypv,mydos]])
92              myusers.append(User(myid,myname,myeav,myimpers,myforge,myfre,myattd,mypv,mydos,mystatus))
```

Εικόνα 20: Λήψη απόφασης

Όπως φαίνεται και στην Εικόνα 20, δίνουμε στο μοντέλο μας όλες τις τιμές από τα πεδία του κάθε χρήστη, τα οποία έχουν ρόλο στη λήψη απόφασης, και με χρήση της συνάρτησης predict της βιβλιοθήκης sklearn, αποφασίζει αν είναι κακόβουλος ή όχι.

Πιο συγκεκριμένα για τον κάθε χρήστη δημιουργούμε ένα αντικείμενο από την κλάση User, που ο κώδικάς της φαίνεται στην Εικόνα 21, με σκοπό την αποθήκευσή τόσο των δεδομένων από το αρχείο μας όσο και την απόφαση του μοντέλου μας.

```
42
43 class User:
44     def __init__(self,ID,Name,EDrop,Imper,Forge,FreRi,Adata,PViol,DOServ,Status):
45         self.ID = ID
46         self.Name = Name
47         self.EDrop = EDrop
48         self.Imper = Imper
49         self.Forge = Forge
50         self.FreRi = FreRi
51         self.Adata = Adata
52         self.PViol = PViol
53         self.DOServ = DOServ
54         self.Status = Status
55
```

Εικόνα 21: Κλάση User

4.4.4 Λογική Λειτουργίας Μηχανισμού

Το πρόγραμμά μας είναι ένας βρόγχος επανάληψης που δημιουργεί το εξής μενού:

```
Training the model was successful
Loading users was successful


Give 1 to exit
Give 2 to show the number of malicious users found
Give 3 to show the malicious users with info
Give 4 to retrain the model
Give 5 to reload the users
Give 6 to write the evaluated users in a csv
Give 7 to show the accuracy
Give 8 to show a graph chart with multiple trainings and their accuracy
```

Εικόνα 22: Μενού

Μπαίνοντας σε αυτόν τον βρόγχο γίνεται η εκπαίδευση του μοντέλου μας και αμέσως μετά η φόρτωση των χρηστών καθώς και η αξιολόγησή ακριβώς όπως περιγράψαμε προηγουμένως. Ύστερα δίνεται η επιλογή στον χρήστη να επιλέξει μία από τις ‘8’ επιλογές που φαίνονται στην Εικόνα 22 έως ότου επιλεγεί η επιλογή της εξόδου όπου και θα τερματίσει το πρόγραμμα. Τα αποτελέσματα της κάθε επιλογής θα συζητηθεί στο επόμενο κεφάλαιο.

Κεφάλαιο 5

Αποτελέσματα Προγράμματος

5.1 Εισαγωγή Κεφαλαίου

Όπως αναφέραμε στο προηγούμενο κεφάλαιο το πρόγραμμά μας έχει 8 επιλογές οι οποίες είναι οι ακόλουθες:

- **Επιλογή 1:** Με αυτήν την επιλογή γίνεται έξοδος και τερματισμός του προγράμματος
- **Επιλογή 2:** Με αυτήν την επιλογή γίνεται εμφάνιση του αριθμού των κακόβουλων χρηστών που εντοπίστηκαν
- **Επιλογή 3:** Με αυτήν την επιλογή παρουσιάζεται ο κάθε κακόβουλος χρήστης με πληροφορίες
- **Επιλογή 4:** Με αυτήν την επιλογή γίνεται επανεκπαίδευση του μοντέλου μας με νέο αρχείο (ξαναδημιουργείται) εκπαίδευσης χρηστών
- **Επιλογή 5:** Με αυτήν την επιλογή γίνεται επαναφόρτωση νέων χρηστών
- **Επιλογή 6:** Με αυτήν την επιλογή γίνεται αποθήκευση των αξιολογημένων χρηστών σε ένα αρχείο csv
- **Επιλογή 7:** Με αυτήν την επιλογή γίνεται εμφάνιση της ακρίβειας του μοντέλου μας
- **Επιλογή 8:** Με αυτήν την επιλογή εμφανίζεται ένα γράφημα που δείχνει την ακρίβεια σε συνάρτηση πολλαπλών εκπαιδεύσεων διαφορετικού πλήθους χρηστών εκπαίδευσης.

5.2 Επιλογές Χρήστη Προγράμματος

Παρακάτω ακολουθεί η περιγραφή της κάθε επιλογής, που μπορεί να έχει ο χρήστης του προγράμματος.

5.2.1 Επιλογή 1 'Έξοδος'

Με αυτήν την επιλογή ο χρήστης τερματίζει το πρόγραμμα, όπως είναι και προφανές από το όνομα της επιλογής άλλωστε, και την εκτύπωση σχετικού μηνύματος που φαίνεται παρακάτω:

```
Training the model was successful
Loading users was successful

Give 1 to exit
Give 2 to show the number of malicious users found
Give 3 to show the malicious users with info
Give 4 to retrain the model
Give 5 to reload the users
Give 6 to write the evaluated users in a csv
Give 7 to show the accuracy
Give 8 to show a graph chart with multiple trainings and their accuracy
1

Exiting..
```

Εικόνα 23: Μενού Επιλογής 1

Αυτό γίνεται αλλάζοντας τη συνθήκη του βρόγχου επανάληψης σε ψευδής όπως φαίνεται στην Εικόνα 24.

```
107     if (option == 1): #Exit
108         flag = False
109         print("\nExiting..")
```

Εικόνα 24: Κώδικας επιλογής 1

5.2.2 Επιλογή 2 'Εμφάνιση Αριθμού Κακόβουλων Χρηστών'

Με αυτήν την επιλογή δείχνουμε στην έξοδο τον αριθμό των κακόβουλων χρηστών. Αυτό το καταφέρνουμε αρχικά στο σημείο που φορτώνουμε τους χρήστες και τους αξιολογούμε, όπου με μία προσπέλαση σε αυτούς κρατάμε όλους τους κακόβουλους ξεχωριστά στον πίνακα malusers. Αυτό φαίνεται στην Εικόνα 25:

```
95          malusers = []
96          for p in range(len(myusers)):
97              if (str(myusers[p].Status) == str(['Malicious'])):
98                  malusers.append(myusers[p])
```

Εικόνα 25: Αποθήκευση κακόβουλων χρηστών

Και μόλις επιλεγεί η επιλογή 2 τότε εκτελείται ο εξής κώδικας και εκτυπώνεται το σχετικό μήνυμα όπως φαίνεται παρακάτω:

```
110     if (option == 2): #Show how many malicious users have been found
111         print('Number of malicious users found:',len(malusers))
```

Εικόνα 26: Κώδικας επιλογής 2

```
Training the model was successful
Loading users was successful

Give 1 to exit
Give 2 to show the number of malicious users found
Give 3 to show the malicious users with info
Give 4 to retrain the model
Give 5 to reload the users
Give 6 to write the evaluated users in a csv
Give 7 to show the accuracy
Give 8 to show a graph chart with multiple trainings and their accuracy
2
Number of malicious users found: 28
```

Εικόνα 27: Μενού Επιλογής 2

Το αποτέλεσμα αυτό, όπως θα φανεί και παρακάτω, μπορεί να αλλάξει ανάλογα με την ακρίβεια του μοντέλου.

5.2.3 Επιλογή 3 'Εμφάνιση των Κακόβουλων Χρηστών'

Με αυτήν την επιλογή γίνεται εμφάνιση όλων των κακόβουλων χρηστών που εντοπίστηκαν, μαζί με το όνομά τους, το id τους αλλά και τους λόγους που αξιολογήθηκαν ως κακόβουλοι χρήστες. Παρακάτω φαίνεται ο κώδικας που το υλοποιεί καθώς και τα αποτελέσματα, Εικόνα 28 και Εικόνα 29 αντίστοιχα:

```
    if (option == 3): #Show the users with info
        if mflag:
            print('No malicious user was found')
        else:
            for p in range(len(malusers)):
                if (p==0):
                    print('-------------------------------------------------------------------------------------------------------')
                save = ""
                if (malusers[p].EDrop == 1):
                    save =  save + " Eavesdropping"
                if (malusers[p].Imper == 1):
                    save =  save + " Impersonating"
                if (malusers[p].Forge == 1):
                    save =  save + " Forgery"
                if (malusers[p].FreRi == 1):
                    save =  save + " Free-Riding"
                if (malusers[p].Adata == 1):
                    save =  save + " Active-attack-on-control-data"
                if (malusers[p].PViol == 1):
                    save =  save + " Privacy-violation"
                if (malusers[p].DOServ == 1):
                    save =  save + " Denial-of-Service"
                print('User named',malusers[p].Name, 'with id', malusers[p].ID,'is',malusers[p].Status,'For:',save)
                print('-------------------------------------------------------------------------------------------------------')
```

Εικόνα 28: Κώδικας επιλογής 3

```
Give 1 to exit
Give 2 to show the number of malicious users found
Give 3 to show the malicious users with info
Give 4 to retrain the model
Give 5 to reload the users
Give 6 to write the evaluated users in a csv
Give 7 to show the accuracy
Give 8 to show a graph chart with multiple trainings and their accuracy
3
-------------------------------------------------------------------------------------------------------
User named Cecilia with id 10 is ['Malicious'] For:  Eavesdropping Impersonating Free-Riding
-------------------------------------------------------------------------------------------------------
User named Julia with id 55 is ['Malicious'] For:  Eavesdropping Forgery Free-Riding
-------------------------------------------------------------------------------------------------------
User named Jesse with id 65 is ['Malicious'] For:  Eavesdropping Impersonating Privacy-violation
-------------------------------------------------------------------------------------------------------
User named Kayden with id 68 is ['Malicious'] For:  Privacy-violation
-------------------------------------------------------------------------------------------------------
User named Isabella with id 83 is ['Malicious'] For:  Forgery Privacy-violation
-------------------------------------------------------------------------------------------------------
User named Lyla with id 91 is ['Malicious'] For:  Active-attack-on-control-data
-------------------------------------------------------------------------------------------------------
User named Alina with id 99 is ['Malicious'] For:  Forgery Active-attack-on-control-data Privacy-violation
-------------------------------------------------------------------------------------------------------
User named Lyla with id 131 is ['Malicious'] For:  Eavesdropping Forgery Privacy-violation Denial-of-Service
-------------------------------------------------------------------------------------------------------
User named Blake with id 134 is ['Malicious'] For:  Impersonating Privacy-violation
-------------------------------------------------------------------------------------------------------
User named Bailey with id 135 is ['Malicious'] For:  Eavesdropping Free-Riding Active-attack-on-control-data Privacy-
violation Denial-of-Service
-------------------------------------------------------------------------------------------------------
```

Εικόνα 29: Μενού Επιλογής 3

5.2.4 Επιλογή 4 ‘Επανεκπαίδευση Μοντέλου’

Σε αυτήν την επιλογή ζητάμε από τον χρήστη να δώσει τον αριθμό των χρηστών εκπαίδευσης, έτσι ώστε να επανεκπαιδευτεί το μοντέλο μας με αυτόν τον αριθμό χρηστών. Ο κώδικας που υλοποιεί αυτή τη διαδικασία φαίνεται στην Εικόνα 30:

```
136     if (option == 4): #Retrain the model with differnet set of users
137         x = int(input('Please give the number of the randomised users for retraining\n'))
138         maketrainingusers(x)
139         trainflag = True
140         loadflag = True
```

Εικόνα 30: Κώδικας επιλογής 4

Συγκεκριμένα σε αυτήν την επιλογή, αλλάζει το αρχείο εκπαίδευσης καλώντας την συνάρτηση maketrainingusers με τον αριθμό που έδωσε ο χρήστης του προγράμματος, και ανανεώνονται οι σημαίες της εκπαίδευσης, που επιτρέπει την λειτουργία που περιγράψαμε στο προηγούμενο κεφάλαιο για την εκπαίδευση του μοντέλου, αλλά και της αξιολόγησης των χρηστών με το νέο μοντέλο μας.

```
Give 1 to exit
Give 2 to show the number of malicious users found
Give 3 to show the malicious users with info
Give 4 to retrain the model
Give 5 to reload the users
Give 6 to write the evaluated users in a csv
Give 7 to show the accuracy
Give 8 to show a graph chart with multiple trainings and their accuracy
4
Please give the number of the randomised users for retraining
5000
Training the model was successful
Loading users was successful
```

Εικόνα 31: Μενού Επιλογής 4

5.2.5 Επιλογή 5 ‘Επαναφόρτωση Χρηστών’

Αντίστοιχα με την επιλογή 4, ακολουθείται η ίδια διαδικασία αλλά αλλάζουν μόνο οι χρήστες προς αξιολόγηση αυτή τη φορά και τίποτα άλλο. Ο κώδικας που το υλοποιεί φαίνεται στην Εικόνα 32:

```
141     if (option == 5): #Load different set of users
142         x = int(input('Please give the number of users for reloading\n'))
143         makeusers(x)
144         loadflag = True
```

Εικόνα 32: Κώδικας επιλογής 5

Δίνεται ξανά από τον χρήστη ο αριθμός των νέων χρηστών προς αξιολόγηση και αμέσως μετά καλείται η συνάρτηση που ανανεώνει το αρχείο χρηστών μας. Ύστερα ενεργοποιείται η σημαία φόρτωσης και αξιολογούνται οι νέοι χρήστες μας. Στην Εικόνα 33 φαίνεται και το μενού εκτέλεσης αυτής της εντολής:

```
Give 1 to exit
Give 2 to show the number of malicious users found
Give 3 to show the malicious users with info
Give 4 to retrain the model
Give 5 to reload the users
Give 6 to write the evaluated users in a csv
Give 7 to show the accuracy
Give 8 to show a graph chart with multiple trainings and their accuracy
5
Please give the number of users for reloading
300
Loading users was successful
```

Εικόνα 33: Μενού Επιλογής 5

5.2.6 Επιλογή 6 'Αποθήκευση Χρηστών σε αρχείο Csv'

Με αυτήν την επιλογή αποθηκεύουμε όλους τους αξιολογημένους χρήστες σε ένα αρχείο csv, ακριβώς όπως κάναμε και με τη δημιουργία των χρηστών μας ή των χρηστών εκπαίδευσης. Ακολουθεί ο κώδικας:

```
145     if (option == 6): #Write the users in a csv
146         filename = "evaluatedusers.csv"
147         fields = ['ID','Name','Eavesdropping attack','Impersonate attack','Forge attack',
148                   'Free-riding attack','Active attack on control data','Privacy violation',
149                   'Denial-of-Service','Status']
150         with open(filename, 'w') as csvfile:
151             csvwriter = csv.writer(csvfile)
152             csvwriter.writerow(fields)
153             for i in range(len(myusers)):
154                 row = [i+1,myusers[i].Name,myusers[i].EDrop,myusers[i].Imper,myusers[i].Forge,
155                        myusers[i].FreRi,myusers[i].Adata,myusers[i].PViol,myusers[i].DOServ,myusers[i].Status]
156                 csvwriter.writerow(row)
157         print('Writing users in file evaluatedusers.csv was successful')
```

Εικόνα 34: Κώδικας επιλογής 6

Αφού ακολουθήσει η εκτέλεσή του που δημιουργεί, πηγαίνοντας στο αρχείο evaluatedusers.csv δηλαδή, βλέπουμε ότι όλοι οι χρήστες έχουν αξιολογηθεί, όπως φαίνεται παρακάτω στην Εικόνα 35:

	A	B	C	D	E	F	G	H	I	J
1	ID	Name	Eavesdropping attack	Impersonate attack	Forge attack	Free-riding attack	Active attack on control data	Privacy violation	Denial-of-Service	Status
2										
3	1	Ophelia	0	0	0	0	0	0	0	['Benevolent']
4										
5	2	Freya	0	0	0	0	0	0	0	['Benevolent']
6										
7	3	Mya	0	0	0	0	0	0	0	['Benevolent']
8										
9	4	Nova	0	0	0	0	0	0	0	['Benevolent']
10										
11	5	Aria	0	0	0	0	0	0	0	['Benevolent']
12										
13	6	Alex	0	0	0	0	0	0	0	['Benevolent']
14										
15	7	Eliana	0	0	0	0	0	0	0	['Benevolent']
16										
17	8	Quinn	0	0	0	0	0	0	0	['Benevolent']
18										
19	9	Ophelia	0	0	0	0	0	0	0	['Benevolent']
20										
21	10	Cecilia	1	1	0	1	0	0	0	['Malicious']
22										
23	11	River	0	0	0	0	0	0	0	['Benevolent']
24										
25	12	Ivy	0	0	0	0	0	0	0	['Benevolent']
26										
27	13	Finn	0	0	0	0	0	0	0	['Benevolent']
28										
29	14	Mia	0	0	0	0	0	0	0	['Benevolent']
30										
31	15	Ari	0	0	0	0	0	0	0	['Benevolent']
32										
33	16	Alice	0	0	0	0	0	0	0	['Benevolent']
34										
35	17	Shia	0	0	0	0	0	0	0	['Benevolent']
36										
37	18	Ari	0	0	0	0	0	0	0	['Benevolent']
38										
39	19	Jude	0	0	0	0	0	0	0	['Benevolent']
40										
41	20	Isabella	0	0	0	0	0	0	0	['Benevolent']
42										

Εικόνα 35: Dataset evaluatedusers

Σε σύγκριση με το dataset των user που είδαμε προηγουμένως μπορούμε να παρατηρήσουμε ότι ο χρήστης ‘Cecilia’ με Id 10 έχει αξιολογηθεί ως κακόβουλος (Malicious), ενώ οι υπόλοιποι καλοκάγαθοι (Benevolent), όπως ακριβώς και θα έπρεπε να γίνει. Άρα το μοντέλο μας προβλέπει σωστά, όπως και θέλουμε.

5.2.7 Επιλογή 7 ‘Εμφάνιση Ακρίβειας’

Σε αυτήν την επιλογή εμφανίζουμε την ακρίβεια του μοντέλου μας. Αυτό το καταφέρνουμε κάνοντας χρήση των συναρτήσεων train_test_split και accuracy_score της βιβλιοθήκης sklearn. Η πρώτη συνάρτηση χωρίζει το δείγμα εκπαίδευσης στους πίνακες X_train, X_test, y_train και y_test. Οι πίνακες X_train (περιέχει τα πεδία forge attack, denial of service κ.ο.κ) και y_train (περιέχει τα Malicious Benevolent των χρηστών σε αντιστοιχία με τον X_train), χρησιμοποιούνται για εκπαίδευση, όμως επειδή έχει ήδη εκπαιδευτεί το μοντέλο μας μένουν αχρησιμοποίητοι. Οι X_test και y_test, είναι οι πίνακες με τους οποίους ελέγχουμε την ακρίβεια του μοντέλου μας. Αρχικά ο X_test είναι όπως το αρχείο με τους χρήστες προς αξιολόγηση, χωρίς βέβαια τα ονόματα και τα id, απλά τα πεδία με τις παραβάσεις, και ο y_test έχει τις απαντήσεις, σε πλήρη αντιστοιχία με τον X_test, το αν είναι Malicious ή Benevolent. Οπότε με χρήση της

predict το πρόγραμμα μας αξιολογεί τα δεδομένα του X_test και με χρήση της accuracy_score τα συγκρίνει με τον y_test και επιστρέφει την ακρίβεια. Αυτό φαίνεται στην Εικόνα 36:

```
157     if (option == 7): #Accuracy
158         X_train, X_test, y_train, y_test = train_test_split(X, y, test_size = 0.2)
159         predictions = model.predict(X_test)
160         acc = accuracy_score(y_test, predictions)
```

Εικόνα 36: Κώδικας επιλογής 7

```
Give 1 to exit
Give 2 to show the number of malicious users found
Give 3 to show the malicious users with info
Give 4 to retrain the model
Give 5 to reload the users
Give 6 to write the evaluated users in a csv
Give 7 to show the accuracy
Give 8 to show a graph chart with multiple trainings and their accuracy
7
The Accuracy of our model is 1.0
```

Εικόνα 37: Μενού Επιλογής 7

Σε αυτή τη περίπτωση βλέπουμε πως το μοντέλο μας έχει ακρίβεια ‘1’, που σημαίνει ότι όλες οι προβλέψεις που έκανε ήταν σωστές. Αξίζει να σημειωθεί ότι στη συγκεκριμένη εκτέλεση το μοντέλο εκπαιδεύτηκε με έναν αξιοσημείωτο αριθμό χρηστών εκπαίδευσης, συγκεκριμένα ‘5000’. Παρακάτω θα δούμε το ρόλο που παίζει η εκπαίδευση στην πρόβλεψη του μοντέλου.

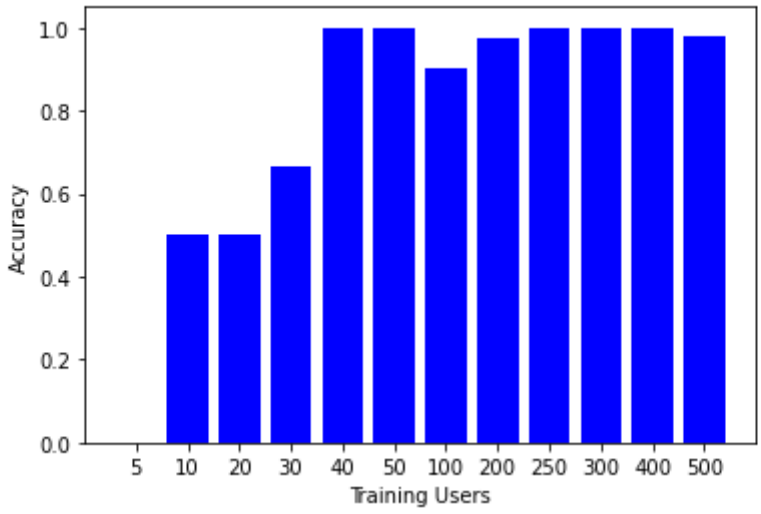
5.2.8 Επιλογή 8 ‘Εμφάνιση Γραφημάτων Ακρίβειας’

Σε αυτή την επιλογή δοκιμάζουμε το μοντέλο μας για να δούμε πως ανταποκρίνεται σε διαφορετικούς αριθμούς χρηστών εκπαίδευσης, συγκεκριμένα βλέπουμε πως ανταποκρίνεται στους εξής αριθμούς χρηστών 5, 10, 20, 30, 40 ,50, 100, 200, 250 , 300, 400, 500. Παρακάτω φαίνεται ο κώδικας καθώς και τα αποτελέσματα, Εικόνα 38 και Εικόνα 39:

```
    if (option == 8): #Graph Accuracy
        us = [5,10,20,30,40,50,100,200,250,300,400,500]
        accurracy = [0,0,0,0,0,0,0,0,0,0,0,0]
        for i in range(len(us)):
            maketrainingusers(us[i])
            trainingdata = pd.read_csv("trainusers.csv")
            X = trainingdata.drop(columns = ['Status','ID','Name'])
            y = trainingdata['Status']
            X_train, X_test, y_train, y_test = train_test_split(X, y, test_size = 0.2)
            model = DecisionTreeClassifier()
            model.fit(X_train, y_train)
            predictions = model.predict(X_test)
            accurracy[i] = accuracy_score(y_test,predictions)
        pos = list(range(12))
        plt.ylabel('Accuracy')
        plt.xlabel('Training Users')
        plt.bar(pos, accurracy, color='blue')
        plt.xticks(ticks=pos, labels=us)
        plt.show()
        print(accurracy)
```

Εικόνα 38: Κώδικας επιλογής 8

Εδώ πρέπει να πούμε πως γίνεται ένας βρόγχος 12 επαναλήψεων, μία για κάθε περίπτωση διαφορετικού αριθμού χρηστών, και ακολουθείται η εξής διαδικασία. Αρχικά καλείται η maketrainingusers που δημιουργεί ξανά το αρχείο χρηστών προς εκπαίδευση με το αντίστοιχο αριθμό χρηστών και ύστερα ακολουθεί η εκπαίδευση του μοντέλου μας με αυτό. Αμέσως μετά υπολογίζεται η ακρίβεια με τον ίδιο τρόπο που έγινε στην προηγούμενη επιλογή. Και αφού έχουμε αποθηκεύσει τα αποτελέσματά μας, δημιουργούμε και το παρακάτω γράφημα



Εικόνα 39: Μενού Επιλογής 8

Σε αυτό το σημείο μπορούμε να παρατηρήσουμε πως στα μεγαλύτερα δείγματα χρηστών εκπαίδευσης η ακρίβεια πλησιάζει όλο και περισσότερο το ‘1’, που είναι και το ιδανικό. Αυτό συμβαίνει διότι όσο περισσότερα δείγματα έχει η μηχανική μάθηση κατά

την εκπαίδευση, τόσο καλύτερα μπορεί να επεξεργαστεί και να οδηγηθεί στη σωστή απάντηση. Αυτό όμως δεν είναι και απόλυτο όπως βλέπουμε και στο παράδειγμα των ‘100’ χρηστών όπου η ακρίβεια είναι μικρότερη των προηγουμένων της, δηλαδή των ‘40’ και ‘50’. Στη συγκεκριμένη περίπτωση του δικού μας προγράμματος υπάρχει μία τυχαιότητα στη δημιουργία των χρηστών εκπαίδευσης, όπως περιγράψαμε και παραπάνω, που σημαίνει ότι στην περίπτωση των ‘100’ χρηστών μπορεί να υπάρχουν λιγότεροι ‘κακόβουλοι χρήστες’ από ότι των προηγούμενων περιπτώσεων που αυτό με τη σειρά του οδηγεί σε μικρότερη ακρίβεια. Το γενικό συμπέρασμα όμως είναι ότι στα μεγάλα δείγματα η ακρίβεια είναι σε πλήρως ικανοποιητικά επίπεδα σε αντίθεση με αυτά που υπάρχει ελλιπής αριθμός χρηστών εκπαίδευσης.

5.2.9 Συμπεράσματα

Σε αυτό το κεφάλαιο παρουσιάσαμε ένα μοντέλο το οποίο κάνοντας χρήση απλών δεδομένων είναι ικανό να εκπαιδευτεί κατάλληλα, με τη προϋπόθεση ότι έχει σεβαστό δείγμα εκπαίδευσης αφού όπως δείξαμε όσο μεγαλύτερο δείγμα έχουμε τόσο πιο ακριβή αποτελέσματα, και να μπορεί να εντοπίσει κακόβουλους χρήστες μέσα σε ένα δίκτυο. Όπως δείξαμε και μέσω του προγράμματος μας, η ακρίβεια έχει το πιο σημαντικό ρόλο καθώς, όσο μειώνεται αυτή τα αποτελέσματα που βγάζουμε είναι λανθασμένα και αυτό με τη σειρά του δημιουργεί προβλήματα. Πιο συγκεκριμένα εάν σε ένα δίκτυο υπολογιστών έχουμε ένα αντίστοιχο πρόγραμμα εντοπισμού κακόβουλων χρηστών, που όπως είναι λογικό ίσως να έχει διαφορετικό τρόπο εύρεσης ή ίσως καλύτερα άλλη λογική, θα πρέπει να έχει όσο το δυνατόν τέλεια ακρίβεια διότι μέσα στο δίκτυο δεν θα πρέπει κάποιος χρήστης να βγει κακόβουλος, ενώ ο ίδιος δεν είναι. Η μηχανική μάθηση, και συγκεκριμένα τα Decision Tree, λύνουν αρκετά αποτελεσματικά αυτό το πρόβλημα με μόνο παράγοντα το δείγμα εκπαίδευσης.

Κεφάλαιο 6

Επίλογος

6.1 Εισαγωγή Κεφαλαίου

Σε αυτό το κεφάλαιο θα μιλήσουμε για τα συμπεράσματα που καταλήγουμε για τα δίκτυα, την ασφάλεια σε αυτά αλλά και τη Μηχανική Μάθηση, διότι με την συνεχή εξέλιξη της τεχνολογίας, απαιτείται και συνεχής προσαρμογή στα νέα δεδομένα.

Επιπλέον θα κάνουμε και μία αναφορά στις πιθανές μελλοντικές επεκτάσεις που θα μπορεί να έχει ο μηχανισμός που παρουσιάσαμε.

6.2 Συμπεράσματα

Τα δίκτυα υπολογιστών εξελίσσονται διαρκώς, και με αρκετά γρήγορους ρυθμούς, με κάθε γενιά να φέρνει νέες δυνατότητες. Μαζί με αυτές τις δυνατότητες όμως προκύπτουν και νέες προκλήσεις, τόσο στην ανάπτυξη μηχανισμών ικανών να ανταπεξέλθουν σε αυτές όσο και στην ασφαλή χρήση αυτών.

Με την έλευση της πέμπτης γενιάς δικτύων έχουμε φτάσει στην εποχή όπου διαθέτουμε πλήρως ασύρματα δίκτυα μέσα στα οποία υπάρχουν συσκευές, που ποικίλουν από αυτοκίνητα έως κινητά, που αλληλεπιδρούν μεταξύ τους με σκοπό να ικανοποιήσουν την όποια μας ανάγκη.

Με αυτόν τον τρόπο όμως εντάσσονται και απειλές μέσα στα δίκτυα, καθώς όσο εξελίσσεται η τεχνολογία μαζί της εξελίσσονται και οι τρόποι που κάποιος χρήστης μπορεί να τα εκμεταλλευθεί και να προκαλέσει ζημιές τόσο στο ίδιο το δίκτυο όσο και

στους υπόλοιπους χρήστες. Παράλληλα όμως εξελίσσονται και τεχνικές με τις οποίες καταπολεμούνται τέτοιες απειλές.

Με την εφαρμογή της Μηχανικής Μάθησης, ως εργαλείο εντοπισμού κακόβουλων χρηστών, είδαμε αρκετά μεγάλη επιτυχία στον εντοπισμό των κακόβουλων χρηστών, αλλά με την προϋπόθεση να είναι διαθέσιμο ένα ικανοποιητικό δείγμα εκπαίδευσης. Μπορούμε να συμπεράνουμε λοιπόν, ότι η αξιοποίηση της Μηχανικής Μάθησης για την ασφάλεια των δικτύων, είναι μία βιώσιμη και αποδοτική επιλογή. Τα αποτελέσματα ήταν αρκετά αισιόδοξα καθώς το ίδιο το πρόγραμμα έπαιρνε σωστά αποφάσεις, όπως θα έκανε και ένας άνθρωπος, αλλά σε πολύ μικρότερο χρονικό διάστημα ακόμα και σε μεγάλο όγκο δεδομένων.

6.3 Μελλοντικές Επεκτάσεις

Ο μηχανισμός που αναπτύχθηκε στο πλαίσιο αυτής της διπλωματικής εργασίας θα μπορούσε να επεκταθεί και να αναπτυχθεί, ή ακόμα και να τροποποιηθεί, περαιτέρω ως εξής:

- Η Μηχανική Μάθηση, χάρη στη ευελιξία που διαθέτει, δεν περιορίζεται σε δεδομένα της μορφής που παρουσιάσαμε, αλλά λίγο πολύ σε οτιδήποτε. Αυτό σημαίνει ότι ο εξής μηχανισμός μπορεί να χρησιμοποιηθεί, για παράδειγμα σε σελίδες κοινωνικής δικτύωσης για να εκτελέσει την ίδια λειτουργία αλλά προφανώς με διαφορετικό δείγμα εκπαίδευσης.
- Ένας άλλος τρόπος επέκτασης, θα μπορούσε να είναι, να γίνει μέρος ενός μεγαλύτερου συστήματος που λαμβάνει χρήση σε δίκτυα πραγματικής χρήσης. Δηλαδή με μικρές αλλαγές θα παρακολουθεί τους χρήστες, και αναλόγως πως στο δίκτυο μπορεί να θεωρηθεί κακόβουλος, να αναλαμβάνει τον εντοπισμό τους με στόχο την ασφάλεια του δικτύου.
- Επιπλέον θα μπορούσαμε να χρησιμοποιούμε αυτόν τον μηχανισμό ως ένα εργαλείο, που θα αξιολογεί το κάθε δίκτυο, στο οποίο εφαρμόζεται, με βάση τον αριθμό των κακόβουλων χρηστών, σε σύγκριση με τον συνολικό αριθμό των χρηστών. Δηλαδή για το κάθε δίκτυο, θα βγάζει ένα ποσοστό ασφάλειας, με στόχο τη σύγκριση μεταξύ των διάφορων δικτύων που υπάρχουν, αλλά και με στόχο να μπορεί ο κάθε χρήστης, να βλέπει τα αποτελέσματα αυτά για το δίκτυο που θέλει να χρησιμοποιήσει.

- Τέλος, αυτός ο μηχανισμός θα μπορούσε να χρησιμοποιηθεί, με μία μικρή τροποποίηση φυσικά, στο να εντοπίζει παραβιάσεις που πάνε να γίνουν στο δίκτυο. Πιο συγκεκριμένα, αντί να εκπαιδευτεί και να εντοπίζει τους κακόβουλους χρήστες μέσα στο δίκτυο, θα εκπαιδευτεί με τέτοιο τρόπο, που θα παρατηρεί όλες τις κινήσεις που πραγματοποιούνται μέσα σε αυτό. Έτσι όταν θα εντοπίζει μία κίνηση που δεν αναγνωρίζει θα καταλαβαίνει ότι, πιθανότατα, αυτή πρόκειται για μία απόπειρα παραβίασης του δικτύου και θα λαμβάνει αντίστοιχα μέτρα.

Βιβλιογραφία

[1] Jithin Jagannath, Nicholas Polosky, Anu Jagannath, Francesco Restuccia kai Tommaso Melodia. Machine learning for wireless communications in the internet of things:A comprehensive survey, 2019

[2] Muhammad Ahmad. Selection and Ranking of Fog Computing-Based IoT Monitoring of Health Using the Analytic Network Approach. 2021

[3] Kazeem B. Adedeji, Nnamdi I. Nwulu, Aigbavboa Clinton. IoT-Based Smart Water Network Management: Challenges and Future Trend IEE AFRICON 2019

[4] El Naqa I., Murphy M.J. (2015) What Is Machine Learning?. In: El Naqa I., Li R., Murphy M. (eds) Machine Learning in Radiation Oncology. Springer, Cham. https://doi.org/10.1007/978-3-319-18305-3_1

[5] Zehui Meng, Hao Sun, Xiaotong Shen, Ziyue Chen, Marcelo H. Ang, "System integration: Application towards autonomous navigation in cluttered environments", System Integration (SII) 2016 IEEE/SICE International Symposium on, pp. 786-791, 2016

[6] Author: Sieuwert van Otterloo AI, Machine Learning and neural networks explained 27 Ιουλίου 2020

[7] Yulia Gavrilova. Artificial Intelligence vs. Machine Learning vs. Deep Learning: Essentials April 8th 2020

[8] Ivens Portugal, Paulo Alencar kai Donald Cowan. The use of machine learning algorithms in recommender systems: A systematic review. Expert Systems with Applications, 97, 2015

[9] Théotime Gros Can Artificial Intelligence Create Art. Thesis for: Master in International ManagementAdvisor: Alain Busson June 2019

[10] Sandhya N. dhage, Charanjeet Kaur Raina. (2016) A review on Machine Learning Techniques. In International Journal on Recent and Innovation Trends in Computing and Communication, Volume 4 Issue 3

[11] Girish Khanzode Machine Learning. https:// www. slideshare. net/ GirishKhanzode/ supervised-learning-52218215 August 2015

[12] Radim Rehurek. Practical Data Science in Python. University of Economics Prague. December 2014

[13] Vladimir Nasteski An overview of the supervised machine learning methods. Faculty of Information and Communication Technologies, Partizanska bb, 7000 Bitola, Macedonia

[14] Lior Rokach, Oded Z. Maimon. (2008) Data Mining with Decision Trees: Theory and Applications. In World Scientific.
[15] Daniel Jurafsky & James H. Martin, (2016) Speech and Language Processing Tom Mitchell, McGraw Hill (2015) Machine Learning
[16] Machine Learning – linear Regression Model https:// datacadamia . com / data_mining /linear_regression
[17] C.-J. Lin, R. C. Weng, S. S. Keerthi. (2008) Trust region Newton method for large-scale logistic regression. In Journal of Machine Learning Research, vol. 9
[18] Andrew Ng. (2012) CS229 Lecture notes Machine Learning - Supervised learning.
[19] Machine Learning with Python – Logistic Regression November 2011
[20] Memoona Khanam, Tahira Mahboob, Warda Imtiaz A Survey on Unsupervised Machine Learning Algorithms for Automation, Classification and Maintenance. International Journal of Computer Applications June 2015
[21] M. Usama et al., "Unsupervised Machine Learning for Networking: Techniques, Applications and Research Challenges," in IEEE Access, vol. 7, pp. 65579-65615, 2019, doi: 10.1109/ACCESS.2019.2916648.
[22] Feature Extraction: Foundations and Applications, Heidelberg, Germany:Springer, vol. 207, 2008.
[23] A. Coates, A. Y. Ng and H. Lee, "An analysis of single-layer networks in unsupervised feature learning", Proc. Int. Conf. Artif. Intell. Statist., pp. 215-223, 2011.
[24] N. Grira, M. Crucianu and N. Boujemaa, "Unsupervised and semi-supervised clustering: A brief survey", Rev. Mach. Learn. Techn. Process. Multimedia Content, vol. 1, pp. 9-16, Jul. 2004.
[25] P. Berkhin, "A survey of clustering data mining techniques" in Grouping Multidimensional Data, Berlin, Germany:Springer, pp. 25-71, 2006.
[26] M. H. Bhuyan, D. K. Bhattacharyya and J. K. Kalita, "Network anomaly detection: Methods systems and tools", IEEE Commun. Surveys Tuts., vol. 16, no. 1, pp. 303-336, 1st Quart. 2014.
[27] A. McGregor et al., "Flow clustering using machine learning techniques", Proc. Int. Workshop Passive Act. Netw. Meas, pp. 205-214, 2004.
[28] C. M. Bishop, "Latent variable models" in Learning in Graphical Models, Dordrecht, The Netherlands:Springer, pp. 371-403, 1998.
[29] A. Skrondal and S. Rabe-Hesketh, "Latent variable modelling: A survey", Scandin. J. Statist., vol. 34, no. 4, pp. 712-745, 2007.
[30] S. T. Roweis and L. K. Saul, "Nonlinear dimensionality reduction by locally linear embedding", Science, vol. 290, no. 5500, pp. 2323-2326, Dec. 2000.
[31] E. Keogh and A. Mueen, "Curse of dimensionality" in Encyclopedia of Machine Learning, Boston, MA, USA:Springer, pp. 257-258, 2010.
[32] P. Pudil and J. Novovičová, "Novel methods for feature subset selection with respect to problem knowledge" in Feature Extraction Construction Selection, Boston, MA, USA:Springer, pp. 101-116, 1998.

[33] L. Yu and H. Liu, "Feature selection for high-dimensional data: A fast correlation-based filter solution", Proc. Int. Conf. ICML, vol. 3, pp. 856-863, 2003.
[34] W. M. Hartmann, "Dimension reduction vs. variable selection", Proc. Int. Workshop Appl. Parallel Comput, pp. 931-938, 2004.
[35] Y C A Padmanabha Reddy, Viswanath Pulabaigari, Eswara Reddy B. Semi-supervised learning: a brief review. Ferbouary 2018
[36] Kai Arukumaran, Marc Peter Deisenroth, Miles Brundage, Anil Anthony Bharath. A Brief Survey of Deep Reinforcement Learning. IEEE SIGNAL PROCESSING MAGAZINE September 2017
[37] Richard S Sutton and Andrew G Barto. Reinforcement Learning: An Introduction. MIT Press, 1998.
[38] Gerald Tesauro. Temporal Difference Learning and TD-Gammon. Communications of the ACM, 38(3):58–68, 1995.
[39] Satinder Singh, Diane Litman, Michael Kearns, and Marilyn Walker. Optimizing Dialogue Management with Reinforcement Learning: Experiments with the NJFun System. JAIR, 16:105–133, 2002.
[40] Andrew Y Ng, Adam Coates, Mark Diel, Varun Ganapathi, Jamie Schulte, Ben Tse, Eric Berger, and Eric Liang. Autonomous Inverted Helicopter Flight via Reinforcement Learning. Experimental Robotics, pages 363–372, 2006.
[41] Nate Kohl and Peter Stone. Policy Gradient Reinforcement Learning for Fast Quadrupedal Locomotion. In ICRA, volume 3, 2004.
[42] Alexander L Strehl, Lihong Li, Eric Wiewiora, John Langford, and Michael L Littman. PAC Model-Free Reinforcement Learning. In ICML, 2006.
[43] Richard Bellman. On the Theory of Dynamic Programming. PNAS, 38(8):716–719, 1952.
[44] J. H. Schiller, “Mobile Communications,” 2nd Edition, Pearson Education Limited, 2003.
[45] W. C. Y. Lee, “Wireless and Cellular Telecommunications,” 3rd Edition, McGraw-Hill Companies, Inc., 2006.
[46] I. Stojmenovic, “Handbook of Wireless Networks and Mobile Computing,” John Wiley & Sons Inc., 2002.
[47] L. Goleniewski and K. W. Jarrett, “Telecommunications Essentials, Second Edition: The Complete Global Source,” 2nd Edition, Pearson Education Inc., October 2006
[48] J. Eberspächer, H-J Vögel, C. Bettstetter and C. Hartmann, “GSM – Architecture, Protocols and Services,” 3rd Edition, John Wiley & Sons Ltd, 2009
[49] T. S. Rappaport, “Wireless Communications - Principles and Practice,” 2nd Edition, Prentice Hall Inc., 2002.
[50] J. Bannister, P. Mather and S. Coope, “Convergence Technologies for 3G Networks - IP, UMTS, EGPRS and ATM,” John Wiley & Sons Ltd, 2004.
[51] P. Chandra, “Bulletproof Wireless Security - GSM, UMTS, 802.11 and Ad Hoc Security,” Elsevier Inc., 2005

[52] M. Stamp, “Information Security – Principles and Practice,” John Wiley & Sons, Inc., 2006
[53] H. Holma and A. Toskala, “WCDMA for UMTS, Radio Access for Third Generation Mobile Communications,” Third Edition, John Wiley & Sons Ltd, 2004.
[54] D. T. C. Wong, P. Kong, Y. Liang, K.C. Chua, and J.W. Mark, “Wireless Broadband Networks,” John Wiley & Sons, April 2009.
[55] T. Halonen, J. Romero and J. Melero, “GSM, GPRS and EDGE Performance: Evolution Towards 3G/UMTS,” 2nd Edition, John Wiley & Sons Ltd, 2003.
[56] Colin Blanchard. Security for the Third Generation (3G) Mobile System. Network Systems & Security Technologies
[57] F. Khan, “LTE for 4G Mobile Broadband - Air Interface Technologies and Performance,” Cambridge University Press, 2009.
[58] M. Rumney (Editor-in-Chief), “LTE and the Evolution to 4G Wireless – Design and Measurement Challenges,” Agilent Technologies Inc., 2009.
[59] H. Holma and A. Toskala, “LTE for UMTS, OFDMA and SC-FDMA Based Radio Access,” John Wiley & Sons, 2010.
[60] 3GPP TR 36.913, V8.0.0, Technical Specification Group RAN, “Requirements for Further Advancements for Evolved Universal Terrestrial Radio Access (E-UTRA) (LTE-Advanced) (Release 8),” June 2008.
[61] O. T. Eluwole and M. Lohi, “Coordinated Multipoint Power Consumption Modeling for Energy Efficiency Assessment in LTE/LTE-advanced Cellular Networks,” IEEE 19th International Conference on Telecommunications (ICT), Jounieh, Lebanon, April, 2012.
[62] Nokia Networks White Paper, “LTE Release 12 and Beyond,” 2014. [Online]. Available: https://resources.ext.nokia.com/asset/200174.
[63] 4G Americas, “Understanding 3GPP Release 12 Standards for HSPA+ and LTE-Advanced Enhancements”, February 2015. [Online]. Available: http://www.5gamericas.org/files/6614/2359/0457/4G_Americas_-_3GPP_Release_12_Executive_Summary_-_February_2015.pdf
[64] T. Nakamura, “Proposal for Candidate Radio Interface Technologies for IMT-Advanced Based on LTE Release 10 and Beyond (LTEAdvanced), 3GPP ITU-R WP 5D 3rd Workshop on IMT-Advanced, 15 October 2009.
[65] Nour Moustafa, Jiankun Hu. Security and privacy in 4G/LTE Network. Augus 2018t
[66] L. Gupta, R. Jain, and H. A. Chan, “Mobile Edge Computing – An Important Ingredient of 5G Networks,” IEEE Software Defined Networks Technical Community, Newsletter, March 2016.
[67] O. N. C. Yilmaz, “Ultra-Reliable and Low-Latency (URLLC) 5G Communication,” European Conference on Networks and Communications (EuCNC), June 2016.
[68] Ericsson, “5G - key Component of the Networked Society (RWS150009),” 3GPP RAN Workshop on 5G, Phoenix, AZ, USA, 17 – 18 September 2015.

[69] Samsung Electronics White Paper, “5G Vision,” DMC R&D Center, August 2015. [Online]. Available: http:// www .samsung.com /global / businessimages /insights /2015/Samsung-5G-Vision-2.pdf

[70] Ericsson White Paper, “5G systems – Enabling the Transformation of Industry and Society,” UEN 284 23-3251 rev B, January 2017. [Online]. Available: https://www.ericsson.com/assets/local/publications/white-papers/wp5g-systems.pdf

[71] A. Gupta and R. K. Jha, “A Survey of 5G Network: Architecture and Emerging Technologies,” IEEE Access Special Section on Recent Advances in Software Defined Networking for 5G Networks, Volume 3, pp. 1206 – 1232, 2015.

[72] Ericsson White Paper, “5G Radio Access – Capabilities and Technologies,” Uen 284 23-3204 Rev C, April 2016. [Online]. Available: https:// www. ericsson. com/assets /local/publications/whitepapers/wp-5g.pdf

[73] Nokia Networks, “Ten key rules of 5G deployment, Enabling 1 Tbit/s/km2 in 2030,” White Paper, May 2015. [Online]. Available: https://networks.nokia.com/file/39891/ten-key-rules-of-5gdeployment

[74] P. Sorrells, “5G: Advancing the world of connectivity,” CommScope White Paper, September 2016. [Online]. Available: http://www.commscope.com/5g/wp-advancing-the-world-ofconnectivity/

[75] M. Maier, M. Chowdhury, B. P. Rimal, and D. P. Van, “The Tactile Internet: Vision, Recent Progress, and Open Challenges,” IEEE Communications, Volume 54, Issue 5, pp. 138 – 145, May 2016.

[76] ITU-R SG05 Contribution 40, “Draft new Report ITU-R M.[IMT2020.TECH PERF REQ] - Minimum requirements related to technical performance for IMT-2020 radio interface(s),” February 2017. [Online]. Available: https://www.itu.int/md/R15-SG05-C0040/en

[77] BBC Technology News, “5G researchers manage record connection speed,” 25 February 2015. [Online]. Available: http://www.bbc.com/news/technology-31622297

[78] B. A. Forouzan, “Data Communications and Networking,” Fourth Edition, McGraw-Hill, 2007

[79] A. Aijaz, M. Dohler, A. H. Aghvami, V. Friderikos, and M. Frodigh, “Realizing the Tactile Internet: Haptic Communications over Next Generation 5G Cellular Networks,” IEEE Wireless Communications, Volume 24, Issue 2, pp. 82 – 89, April 2017.

[80] N. Bhushan, et al. "Network densification: the dominant theme for wireless evolution into 5G." IEEE Communications Magazine 52.2 (2014): 82-89.

[81] W. Stallings, Cryptography and Network Security: Principles and Practice, 6th ed. London, U.K.: Pearson, 2014.

[82] D. Fang, Y. Qian and R. Q. Hu, "Security for 5G Mobile Wireless Networks," in IEEE Access, vol. 6, pp. 4850-4874, 2018, doi: 10.1109/ACCESS.2017.2779146.

[83] A survey Othmane Nait Hamoud, Tayeb Kenaza, Yacine Challal. Security in Device-to-Device communications (D2D): a survey. IET Networks, IET, 2017, ff10.1049/iet-net.2017.0119ff. ffhal-01627604

[84] S. Nowozin, C. Rother, S. Bagon, T. Sharp, Bangpeng Yao and P. Kohli, "Decision tree fields," 2011 International Conference on Computer Vision, 2011, pp. 1668-1675, doi: 10.1109/ICCV.2011.6126429.Applied Research of Decision Tree Method. A paper by Jinhui Liu: Issue MATEC Web of Conferences Volume 25, 2015. 2015 International Conference on Energy, Materials and Manufacturing Engineering (EMME 2015)

[85] A. Hadiks, Y. Chen, F. Li, and B. Liu, “A study of stealthy denial-ofservice attacks in Wi-Fi direct device-to-device networks,” IEEE 11th Consumer Communications and Networking Conference (CCNC 2014), pp.507-508, January 2014.