ΔΙΕΘΝΕΣ
ΠΑΝΕΠΙΣΤΗΜΙΟ
ΤΗΣ ΕΛΛΑΔΟΣ

ΣΧΟΛΗ ΜΗΧΑΝΙΚΩΝ

ΤΜΗΜΑ ΜΗΧΑΝΙΚΩΝ ΠΛΗΡΟΦΟΡΙΚΗΣ
ΚΑΙ ΗΛΕΚΤΡΟΝΙΚΩΝ ΣΥΣΤΗΜΑΤΩΝ

ΔΙΠΛΩΜΑΤΙΚΗ ΕΡΓΑΣΙΑ

«ΕΠΙΛΥΣΗ ΤΟΥ ΠΡΟΒΛΗΜΑΤΟΣ ΤΗΣ ΔΙΑΔΟΣΗΣ ΚΑΙ ΔΙΑΣΠΟΡΑΣ ΦΗΜΩΝ ΣΤΑ ΜΕΣΑ ΚΟΙΝΩΝΙΚΗΣ ΔΙΚΤΥΩΣΗΣ ΜΕ ΤΗΝ ΧΡΗΣΗ ΝΕΥΡΩΝΙΚΩΝ ΔΙΚΤΥΩΝ ΒΑΘΙΑΣ ΜΑΘΗΣΗΣ- DEEP LEARNING»

Του φοιτητή
Πέτκαρη Άγγελου Παναγιώτη
Αρ. Μητρώου: 144377

Επιβλέπων
Γουλιάνας Κωνσταντίνος
Καθηγητής

Ημερομηνία 17-05-2025

Τίτλος Π.Ε. Επίλυση Του Προβλήματος Της Διάδοσης Και Διασποράς Φημών Στα Μέσα Κοινωνικής Δικτύωσης Με Τη Χρήση Νευρωνικών Δικτύων Βαθιάς Μάθησης Deep Learning

Κωδικός Π.Ε. 23144

Ονοματεπώνυμο φοιτητή Πέτκαρης Άγγελος Παναγιώτης

Ονοματεπώνυμο εισηγητή Γουλιάνας Κωνσταντίνος

Ημερομηνία ανάληψης Π.Ε. 23-03-2023

Ημερομηνία περάτωσης Π.Ε. 17-05-2025

Βεβαιώνω ότι είμαι ο συγγραφέας αυτής της εργασίας και ότι κάθε βοήθεια την οποία είχα για την προετοιμασία της είναι πλήρως αναγνωρισμένη και αναφέρεται στην εργασία. Επίσης, έχω καταγράψει τις όποιες πηγές από τις οποίες έκανα χρήση δεδομένων, ιδεών, εικόνων και κειμένου, είτε αυτές αναφέρονται ακριβώς είτε παραφρασμένες. Επιπλέον, βεβαιώνω ότι αυτή η εργασία προετοιμάστηκε από εμένα προσωπικά, ειδικά ως πτυχιακή εργασία, στο Τμήμα Μηχανικών Πληροφορικής και Ηλεκτρονικών Συστημάτων του ΔΙ.ΠΑ.Ε.

Πρόλογος

Στη σύγχρονη εποχή η διάδοση φημών στα μέσα κοινωνικής δικτύωσης είναι ένα συχνό φαινόμενο, το οποίο παρατηρείται σε πλήθος περιπτώσεων και μπορεί να δημιουργεί σύγχυση και να κοινωνική αναστάτωση. Η φύση των εν λόγω πλατφορμών ευνοεί τη γρήγορη διάδοση περιεχομένου, ενώ παράλληλα καθιστά δύσκολη τη διασταύρωση της ακρίβειας των διαφόρων πληροφοριών. Η ανάγκη για αντιμετώπιση του φαινομένου είναι κρίσιμη, δεδομένου του γεγονότος πως οι φήμες έχουν άμεσες συνέπειες στη ζωή των ατόμων αλλά και σε ολόκληρη την κοινωνία.

Κατά τη διάρκεια των τελευταίων ετών, η ανάπτυξη τόσο της μηχανικής μάθησης όσο και της βαθιάς μάθησης έχει φέρει στο προσκήνιο μεθόδους με σκοπό την αντιμετώπιση της διάδοσης των φημών. Μέσω αυτών των τεχνικών είναι εφικτή η αυτόματη ανάλυση πολλών δεδομένων και η αναγνώριση μοτίβων που υπάρχει πιθανότητα να αποτελούν φήμη. Με τη χρήση αλγορίθμων, όπως είναι τα νευρωνικά δίκτυα, οι ερευνητές πλέον μπορούν να εντοπίσουν και να περιορίσουν τη διάδοση φημών με μεγαλύτερη ακρίβεια από ποτέ.

Περίληψη

Η παρούσα πτυχιακή εργασία εξετάζει το πρόβλημα της διάδοσης και διασποράς φημών στα μέσα κοινωνικής δικτύωσης, φαινόμενο που μπορεί να έχει σοβαρές κοινωνικές επιπτώσεις. Στόχος της εργασίας είναι η διερεύνηση της εφαρμογής τεχνικών μηχανικής μάθησης και βαθιάς μάθησης, όπως τα νευρωνικά δίκτυα, για τον εντοπισμό, την ανάλυση και την αντιμετώπιση της παραπληροφόρησης στο ψηφιακό περιβάλλον.

Η εργασία αποτελείται από έξι κεφάλαια:

1. Το 1ο κεφάλαιο εισάγει την έννοια της φήμης, τους τρόπους διάδοσής της και τη σχέση της με τα μέσα κοινωνικής δικτύωσης και την τεχνητή νοημοσύνη.
2. Το 2ο κεφάλαιο παρουσιάζει τις βασικές αρχές της μηχανικής μάθησης, τους τύπους και τις βασικές αλγοριθμικές τεχνικές.
3. Το 3ο κεφάλαιο εστιάζει στη βαθιά μάθηση και αναλύει σχετικούς αλγορίθμους όπως RNN, CNN, και Autoencoders.
4. Στο 4ο κεφάλαιο, γίνεται ανασκόπηση σχετικών ερευνών που έχουν εφαρμοστεί για την ανίχνευση φημών με χρήση ML/DL.
5. Το 5ο κεφάλαιο περιλαμβάνει την πειραματική μεθοδολογία, την ανάλυση του συνόλου δεδομένων PHEME, τα εργαλεία και τη διαδικασία εκπαίδευσης του μοντέλου.
6. Τέλος, στο 6ο κεφάλαιο, παρουσιάζονται τα συμπεράσματα και προτείνονται κατευθύνσεις για μελλοντική έρευνα.

Η μελέτη καταλήγει ότι οι τεχνικές βαθιάς μάθησης προσφέρουν σημαντικά πλεονεκτήματα στην ανίχνευση φημών και μπορούν να συμβάλουν ουσιαστικά στην καταπολέμηση της παραπληροφόρησης στο ψηφιακό περιβάλλον.

Επίλυση Του Προβλήματος Της Διάδοσης Και Διασποράς Φημών Στα Μέσα Κοινωνικής Δικτύωσης Με Τη Χρήση Νευρωνικών Δικτύων Βαθιάς Μάθησης

PETKARIS ANGELOS PANAGIOTIS

Abstract

This undergraduate thesis investigates the problem of rumor propagation and diffusion on social media platforms, a phenomenon with significant societal implications. The main objective is to explore the use of machine learning and deep learning techniques, such as neural networks, for detecting, analyzing, and combating the spread of misinformation online.

The thesis is structured into six chapters:

1. Chapter 1 introduces the concept of rumors, their dissemination mechanisms, and the influence of social media and artificial intelligence.
2. Chapter 2 discusses machine learning fundamentals, including its types and key algorithmic approaches.
3. Chapter 3 focuses on deep learning, presenting algorithms such as RNN, CNN, and Autoencoders.
4. Chapter 4 offers a literature review of studies applying ML and DL to rumor detection.
5. Chapter 5 details the experimental methodology, including the PHEME dataset analysis, tools used, and model training process.
6. Chapter 6 presents conclusions and recommendations for future research.

The findings demonstrate that deep learning methods can effectively detect rumor-related patterns and offer promising solutions to the challenge of misinformation in the digital age.

Περιεχόμενα

Κατάλογος Σχημάτων

Κατάλογος Πινάκων

Κεφάλαιο 1ο: Διάδοση Φήμης στα Κοινωνικά Μέσα

1.1 Η έννοια της φήμης και τρόποι αντιμετώπισης

Ως «φήμη» ορίζονται οι μη επαληθευμένες δηλώσεις πληροφοριών, οι οποίες μεταδίδονται και προκύπτουν σε πλαίσια ασάφειας, κινδύνου καθώς επίσης και πιθανής απειλής. Μέσω της φήμης οι άνθρωποι προσπαθούν να κατανοήσουν γεγονότα. Επιπλέον, καθιστά μια συλλογική συναλλαγή, όπου τα μέλη της κοινότητας προσφέρουν, αξιολογούν και ερμηνεύουν πληροφορίες με σκοπό να κατανοήσουν αβέβαιες καταστάσεις, κατευνάζοντας την κοινωνική ένταση. Ακόμη, επιδιώκεται η επίλυση προβλημάτων συλλογικής κρίσης. Συνεπώς, οι φήμες αποτελούν μια μορφή συλλογικής κατανόησης εντός μιας κοινότητας και αποσκοπούν στην κατανόηση αβέβαιων καταστάσεων, στις περιπτώσεις όπου δε διατίθενται επίσημες πληροφορίες. Εντούτοις, οι φήμες ενδέχεται να επηρεάσουν αρνητικά τα εμπλεκόμενα άτομα ή ομάδες ατόμων, ανάλογα με το περιεχόμενο που αφορούν και τη βούληση όσων τις διαδίδουν (Goh et al., 2017)[1].

Σύμφωνα με τους DiFonzo και Bordia (2006)[2] η φήμη αφορά τη μετάδοση από άτομο σε άτομο, γεγονός που υποδηλώνει μια αλυσίδα επικοινωνίας μέσω ατόμων και κοινοποιούνται συχνά σε μια ομάδα ανθρώπων. Ορισμένες φορές, μια φήμη μπορεί να περιγραφεί καλύτερα ως μια δήλωση που μεταδίδεται μέσω μιας ομάδας, μπορεί επίσης να εκληφθεί ως μια υπόθεση που εκτίθεται προσωρινά κατά τη διάρκεια της ομαδικής συζήτησης στην οποία οι άνθρωποι προσπαθούν να κατανοήσουν μια διφορούμενη κατάσταση

Οι Chen και Wang (2020)[3] σημειώνουν, πως η διάδοση φημών συνιστά μια διαδικασία κοινωνικής μετάδοσης. Στο πλαίσιο αυτής της διαδικασίας, τόσο οι συμπεριφορές των ανθρώπων όσο και το κοινωνικό περιβάλλον είναι δυνατόν να επηρεάσουν τη διαδικασία διάδοσης φημών. Σημαντικό ρόλο στη διάδοση φημών έχει το ποσοστό λήθης, το οποίο αλλάζει με την πάροδο του χρόνου, η εμπιστοσύνη μεταξύ των ανθρώπων και η πειστικότητα των φημών.

Οι διαφορές που υπάρχουν μεταξύ των ανθρώπων καθώς επίσης και ορισμένα από τα χαρακτηριστικά τους, συμπεριλαμβανομένου του επιπέδου εκπαίδευσης αλλά και της ικανότητας αναγνώρισης πληροφοριών, ενδέχεται να επηρεάσουν τις αποφάσεις τους όταν έρχονται αντιμέτωποι με φήμες. Σε σχετική μελέτη, ο Afassinou (2014)[4]. εισήγαγε το ποσοστό εκπαίδευσης του πληθυσμού στο μοντέλο διάδοσης φημών κι εν συνεχεία χώρισε τους αδαείς σε δύο κατηγορίες και συγκεκριμένα στους μορφωμένους αδαείς και στους μη μορφωμένους αδαείς. Σύμφωνα με τα αποτελέσματα της προσομοίωσης του, το ποσοστό εκπαίδευσης επηρέασε τη διαδικασία διάδοσης φημών.

Σε μια άλλη έρευνα, οι Wang και Wang (2017)[5] χώρισαν τους αδαείς σε δύο ομάδες ανάλογα με την ικανότητά τους να αναγνωρίζουν πληροφορίες και σύμφωνα με τα αποτελέσματα της προσομοίωσης, επαληθεύθηκε ότι η διαίρεση είναι σημαντική. Σε μια άλλη έρευνα, οι Li και Ma (2017)[6] συζήτησαν την ευαισθησία των ατόμων στη διάδοση φημών, καταλήγοντας στο συμπέρασμα ότι η χαμηλότερη ευαισθησία μπορεί να εμποδίσει τη διάδοση φημών. Ενδιαφέρον παρουσιάζει και το εύρημα του Cheng και των συνεργατών του (2013)[7] ότι η ισχύς των συνδέσεων των ατόμων σε ένα κοινωνικό δίκτυο επηρεάζει σημαντικά τη διαδικασία διάδοσης φημών.

Προκειμένου να περιοριστούν οι φήμες και οι συνέπειές τους διατίθενται διάφορες μέθοδοι, συμπεριλαμβανομένης της αγνόησης, της επιβεβαίωσης της αλήθειας αλλά και της άρνησης. Σχετικά με την αγνόηση μιας φήμης θεωρείται η πιο αδύναμη μέθοδος και χρησιμοποιείται σε περιπτώσεις όπου η φήμη είναι απίθανη σε μεγάλο βαθμό. Σε αρκετές περιπτώσεις οι φήμες τείνουν να παίρνουν μεγάλη

διάσταση στα μέσα κοινωνικής δικτύωσης και εξαπλώνονται ανεξέλεγκτα. Υπό αυτό το πλαίσιο είναι αναγκαίο να εφαρμόζονται μηχανισμοί διόρθωσης, που καλούνται ως αντιφήμες. Σχετικά με την άρνηση καθιστά μια δημοφιλή τακτική αντιφήμης που χρησιμοποιείται προκειμένου να διαψεύσει τις φήμες, ενώ η αποτελεσματικότητά της έχει αμφισβητηθεί. Επιπλέον τακτικές αντιμετώπισης φημών είναι η παροχή των απαιτούμενων πληροφοριών όπως επίσης και η ενίσχυση της εμπιστοσύνης και της αξιοπιστίας μέσω της ενασχόλησης με τις δημόσιες σχέσεις, στις περιπτώσεις εταιρειών (Goh et al., 2017)[1].

Λαμβάνοντας υπόψη τα παραπάνω, η ανίχνευση φημών συμβάλλει στην επιλογή τακτικών αντιμετώπισης. Πιο συγκεκριμένα, η ανίχνευση φημών καθιστά τη διαδικασία που καθορίζει την αλήθεια οποιασδήποτε περίπτωσης. Υπό αυτό το πλαίσιο πρέπει να αντιμετωπιστούν διάφορες προκλήσεις, όπως είναι η συλλογή δεδομένων και η αναγνώριση προέλευσης μιας φήμης. Μια διαδικασία ελέγχου γεγονότων είναι μια διαδικασία που μπορεί να περιλαμβάνει τα εξής βήματα: α) ανάκτηση από δυνητικά σχετικά αποδεικτικά στοιχεία, β) πρόβλεψη της στάσης κάθε στοιχείου όσον αφορά τον ισχυρισμό, γ) εκτίμηση της αξιοπιστίας των στοιχείων δ) λήψη απόφασης με βάση τα προαναφερθέντα (Pathak et al., 2020)[8].

Οι φήμες μπορούν να εξαπλωθούν στα κοινωνικά μέσα όταν οι χρήστες δεν γνώριζαν προηγουμένως για την ύπαρξή τους και μάθουν για αυτήν μέσω ψηφιακού περιεχομένου που είναι διαθέσιμο σε εκείνους και στη συνέχεια μπορούν να αποφασίσουν να τις υποστηρίξουν και οι ίδιοι. Αυτό μπορεί να γίνει με διάφορους τρόπους, όπως η προσθήκη επιβεβαιωτικών σχολίων σε ψηφιακό περιεχόμενο το οποίο αναφέρεται σε μια φήμη και η αναδημοσίευσή του, ανάλογα με τις δυνατότητες που παρέχει η εκάστοτε πλατφόρμα. Συνεπώς, καθίσταται κατανοητό πως με αυτόν τον τρόπο, διαδίδουν μια φήμη και σε άλλα άτομα πέρα από το κοινό της αρχικής ανάρτησης και συμμετέχουν σε συζητήσεις που σχετίζονται με φήμες (Eismann, 2021)[9].

1.2 Κοινωνικά μέσα και διάδοση φήμης

Στο παρελθόν, οι άνθρωποι χρησιμοποιούσαν λεκτικά αλλά και μη λεκτικά μέσα με σκοπό να μοιραστούν πληροφορίες με άλλα άτομα. Ωστόσο, αυτά τα μέσα επικοινωνίας δεν ήταν επαρκή για να μεταφέρουν πληροφορίες. Προκειμένου να ξεπεράσουν αυτούς τους περιορισμούς, στράφηκαν στα έντυπα μέσα, όπως για παράδειγμα οι εφημερίδες και άλλα κυρίαρχα μέσα, όπως η τηλεόραση και το ραδιόφωνο. Σε αυτό το πλαίσιο υπάρχει ένα σαφές όριο μεταξύ των δημοσιογράφων και των απλών πολιτών, όπου περιορίζονται οι άνθρωποι στο να μοιραστούν τις εμπειρίες τους. Με την έλευση των μέσων κοινωνικής δικτύωσης καταρρίφθηκε αυτό το όριο, δίνοντας τη δυνατότητα στα άτομα να ενεργούν ως κοινό και ως «δημοσιογράφοι» ταυτόχρονα, διευρύνοντας την ποικιλία των πληροφοριών που μπορούν να λάβουν οι άνθρωποι από άλλα άτομα (Ahsan et al., 2019)[10].

Τα μέσα κοινωνικής δικτύωσης σημειώνουν μεγάλη δημοτικότητα, καθώς επιτρέπουν στους χρήστες να διατηρούν επαφή με άλλα άτομα, όπως η οικογένεια και οι φίλοι τους. Επίσης, μέσω των κοινωνικών δικτύων ενημερώνονται για διάφορα γεγονότα και έκτακτες ειδήσεις. Ένα από τα κύρια χαρακτηριστικά των κοινωνικών μέσων είναι η δυνατότητα γρήγορης διάδοσης πληροφοριών μέσω μιας μεγάλης κοινότητας χρηστών. Αναλυτικότερα, τα μέσα που είναι ανοιχτά σε όλους, επιτρέπουν τόσο σε ειδησεογραφικούς οργανισμούς όσο και σε απλούς πολίτες να αναφέρουν τις δικές τους απόψεις και εμπειρίες. Κατ' αυτόν τον τρόπο διευρύνεται το εύρος των πληροφοριών που μπορεί να λάβουν οι

χρήστες, οδηγώντας στην πρόσβαση σε πιο ολοκληρωμένες πληροφορίες είτε σε παραπληροφόρηση (Zubiaga et al., 2016)[11].

Τα διαδικτυακά μέσα συμπεριλαμβανομένων των κοινωνικών δικτύων, των διαδικτυακών κοινοτήτων, των άμεσων μηνυμάτων και των e-mail, αποτελούν δημοφιλείς φορείς για τη διάδοση ειδήσεων, περιεχομένου, ενημερωτικών εκστρατειών και διαφημίσεων προϊόντων. Λόγω της φύσης τους διαδίδουν πληροφορίες ευρέως με ταχείς ρυθμούς και γι' αυτόν το λόγο έχουν γίνει προσπάθειες μέσω της χρήσης τους για διάδοση παραπληροφόρησης, ψευδών ειδήσεων και φημών. Τέτοιες φήμες συνήθως επηρεάζουν τους ανθρώπους, την κοινωνία και γενικά την οικονομία (Choi et al., 2020)[12].

Σχετικά με τα μέσα κοινωνικής δικτύωσης έχουν διατυπωθεί διάφοροι ορισμοί εντός του κλάδου της επικοινωνίας αλλά και σε άλλους κλάδους όπως είναι για παράδειγμα οι δημόσιες σχέσεις και τα μέσα μαζικής ενημέρωσης. Οι περισσότεροι ορισμοί συγκλίνουν στο ότι τα κοινωνικά δίκτυα βασίζονται σε ψηφιακές τεχνολογίες που εστιάζουν στο περιεχόμενο και την αλληλεπίδραση που δημιουργείται μεταξύ των χρηστών (Carr & Hayes, 2015)[13].

Σύμφωνα με τους Kaplan και Haenlein (2010, 61)[14] τα μέσα κοινωνικής δικτύωσης καθιστούν «*μια ομάδα εφαρμογών στο Διαδίκτυο που βασίζονται στα ιδεολογικά και τεχνολογικά θεμέλια του Web 2.0 και που επιτρέπουν τη δημιουργία και την ανταλλαγή περιεχομένου που δημιουργείται από τους χρήστες*».

Δημοφιλείς εφαρμογές είναι το Twitter, το Instagram το Facebook και το YouTube. Είναι αξιοσημείωτο, πως τα μέσα κοινωνικής δικτύωσης εκτός από μέσο διασκέδασής, επηρεάζουν σε μεγάλο βαθμό τους καταναλωτές και τους οργανισμούς (Kaplan, 2015)[15]. Ακόμη, επιτρέπουν σε χρήστες με διαφορετικό υπόβαθρο, όπως διαφορετικό φύλο και πολιτικές πεποιθήσεις, να μοιράζονται ειδήσεις, πληροφορίες αλλά και προσωπικές απόψεις (Choi et al., 2020)[12].

Τα μέσα κοινωνικής δικτύωσης έχουν κληθεί ως «μύλος φημών» σε ότι αφορά στη διάδοση ψευδών φημών και παραπληροφόρησης κατά τη διάρκεια καταστάσεων κρίσης, η οποία έχει τη δυνατότητα να προάγει μεγάλης κλίμακας πανικό και οικονομικές απώλειες. Οι χρήστες των κοινωνικών μέσων αξιολογούν την ακρίβεια των πληροφοριών από μόνοι τους κι εν συνεχεία μπορεί να προβούν σε ενέργεια για τη διάδοση, την παράβλεψη ή την απομυθοποίηση των πληροφοριών που κυκλοφορούν. Επίσης, στους ανθρώπους αρέσει να διαδίδουν φήμες λόγω σπουδαιότητας, κοινωνικής ευθύνης και επίγνωσης των δυσμενών συνεπειών (Agarwal et al., 2022)[16].

Επιπροσθέτως, λόγω χαρακτήρων με προκατειλημμένη αλήθεια, οι άνθρωποι είναι επιρρεπείς να πιστεύουν τις ψευδείς φήμες και να τις διαδίδουν ως αληθινές πληροφορίες. Υπό αυτό το πλαίσιο, οι πλατφόρμες μέσων κοινωνικής δικτύωσης καταχρώνται όλο και περισσότερο παραπλάνηση και χειραγωγούν τους χρήστες διαδίδοντας φήμες, παραπληροφορώντας με ανεπιθύμητη αλληλογραφία και κακόβουλο λογισμικό (Agarwal et al., 2022)[16].

Σε αυτό το σημείο είναι σημαντικό να αναφερθεί, πως η στάση των χρηστών απέναντι στο ψηφιακό περιεχόμενο δεν συνιστά απλώς μια έκφραση των ανεξάρτητων πεποιθήσεών τους, αλλά μπορεί να επηρεαστεί από τις απόψεις άλλων χρηστών. Συνεπώς, η κοινωνική επιρροή που αφορά την επιρροή του ατόμου από τις πεποιθήσεις, τις στάσεις αλλά και τις συμπεριφορές άλλων ατόμων, μπορεί να είναι ιδιαίτερα εμφανής κατά τη χρήση των κοινωνικών μέσων στο διαδίκτυο, όταν οι πληροφορίες σχετικά με τις δραστηριότητες ενός χρήστη γίνονται διαθέσιμες σε άλλους. Ακόμη, οι χρήστες του διαδικτύου μπορούν να παρακολουθούν το ψηφιακό περιεχόμενο, τις ομάδες που ακολουθούν και τα σχόλια άλλων χρηστών οι οποίοι υποστηρίζουν μια φήμη (Kane et al. 2014)[17].

Επιπροσθέτως, άλλες λειτουργίες όπως είναι για παράδειγμα οι ενημερώσεις κατάστασης και η προσθήκη ετικετών, είναι εφικτό να ενισχύσουν την ευαισθητοποίηση των χρηστών σχετικά με τις

δραστηριότητες άλλων και να διευκολύνουν τις αλληλεπιδράσεις. Οι αλληλεπιδράσεις κατά μήκος των σχέσεων των χρηστών συχνά αντιμετωπίζονται ως ο κύριος μηχανισμός διάχυσης πληροφοριών στα μέσα κοινωνικής δικτύωσης, όπου το περιεχόμενο, η κατευθυντικότητα και η ισχύς της ροής πληροφοριών μπορούν να διαμορφώσουν τη διάχυση.

Βάσει σχετικών στοιχείων, οι δυαδικές αλληλεπιδράσεις των χρηστών αντιπροσωπεύουν μόνο ένα μέρος των αναμεταδόσεων φημών στα μέσα κοινωνικά στο διαδίκτυο (Kwon et al. 2016)[18]. Ενώ, οι χρήστες που διαδίδουν ψευδείς φήμες πολλές φορές μαθαίνουν γι' αυτές μέσω αποκλειστικών λειτουργιών πρόσβασης περιεχομένου. Συνεπώς, η δυνατότητα αναζήτησης είναι πιθανό να συμβάλλει στη διάδοση μιας φήμης με βάση τα χαρακτηριστικά της εκάστοτε πλατφόρμας, όπως είναι οι αναζητήσεις λέξεων-κλειδιών, διευκολύνοντας τους χρήστες να έχουν πρόσβαση σε ψηφιακό περιεχόμενο εκτός από τις άμεσες επαφές τους (Eismann, 2021)[9].

1.3 Τεχνητή Νοημοσύνη (AI)

Οι άνθρωποι θεωρούνται το πιο ευφυές είδος στη γη λόγω της ικανότητας σκέψης, της εφαρμογής της λογικής, της κατανόησης της πολυπλοκότητας και της ικανότητας να λαμβάνουν αποφάσεις μόνοι τους. Επιπροσθέτως, μπορούν να κάνουν προγραμματισμό, να καινοτομούν και να λύνουν προβλήματα. Ήδη από την εποχή ανακάλυψης της φωτιάς μέχρι την προσεδάφιση στον Άρη, ο άνθρωπος έχει εφεύρει πολλά πράγματα προς όφελος του είδους. Μια από τις σημαντικότερες εφευρέσεις είναι ο υπολογιστής, ο οποίος έχει διαδραματίσει καίριο ρόλο στη μείωση του φόρτου εργασίας και την επίλυση πολύπλοκων μαθηματικών προβλημάτων. Υπό αυτό το πλαίσιο έγιναν προσπάθειες για δημιουργία ενός είδους «ανθρωπογενούς homosapien», το οποίο μπορεί να συσχετιστεί με τον κόσμο των υπολογιστών με τη μορφή τεχνητής νοημοσύνης (Artificial Intelligence -AI). Στην περίπτωση που ένα σύστημα μπορεί να έχει τις βασικές δεξιότητες όπως η μάθηση, η λογική, η κατανόηση της γλώσσας και η επίλυση προβλημάτων, τότε μπορεί να υποτεθεί ότι υπάρχει AI (Ghosh & Thirugnanam, 2021)[19].

Το έτος 1956, στο πλαίσιο ενός συνεδρίου στο Πανεπιστήμιο Dartmouth, από ομάδα μελετητών προτάθηκε για πρώτη φορά επίσημα ο όρος «Τεχνητή Νοημοσύνη». Έτσι έγινε το πρώτο βήμα για τη μελέτη του τρόπου που οι μηχανές προσομοιώνουν τις ανθρώπινες ευφυείς δραστηριότητες. Αρκετά χρόνια αργότερα, το ενδιαφέρον στράφηκε στην Τεχνητή Νοημοσύνη (AI) όταν το 2016, η AlphaGo νίκησε τον παγκόσμιο πρωταθλητή στο σκάκι. Μέσω της ανάπτυξης της Τεχνητής Νοημοσύνης έχουν επέλθει σημαντικά οικονομικά οφέλη σε πολλές πτυχές της ανθρωπότητας, προωθώντας σε μεγάλο βαθμό την κοινωνική ανάπτυξη.

Σύμφωνα με τους Ghosh και Thirugnanam (2021, 23)[19], η Τεχνητή Νοημοσύνη (AI) «*είναι ένας τομέας της επιστήμης των υπολογιστών που ασχολείται με την ανάπτυξη ευφυών συστημάτων υπολογιστών, τα οποία είναι ικανά να αντιλαμβάνονται, να αναλύουν και να αντιδρούν ανάλογα στις εισροές*».

Ο όρος AI χρησιμοποιείται ευρέως για την επιστήμη της Τεχνητής Νοημοσύνης και προέρχεται από τον αγγλικό όρο Artificial Intelligence. Ειδικότερα, χρησιμοποιεί υπολογιστές τόσο για να προσομοιώνει ανθρώπινες ευφυείς συμπεριφορές όσο και για να εκπαιδεύει τους υπολογιστές να μαθαίνουν ανθρώπινες συμπεριφορές, συμπεριλαμβανομένης της μάθησης και της λήψης αποφάσεων (Zhang & Lu, 2021)[20].

Πιο συγκεκριμένα, το AI συνιστά ένα έργο γνώσης που λαμβάνει τη γνώση ως αντικείμενο, αποκτά γνώση και μπορεί να αναλύει και να μελετά τις μεθόδους έκφρασης της γνώσης με σκοπό να επιτύχει το αποτέλεσμα προσομοίωσης ανθρώπινων πνευματικών δραστηριοτήτων. Αποτελεί έναν συνδυασμό επιστήμης υπολογιστών, βιολογίας, ψυχολογίας, φιλοσοφίας καθώς επίσης και άλλων επιστημονικών κλάδων, επιτυγχάνοντας αξιοσημείωτα αποτελέσματα σε εύρος εφαρμογών όπως είναι για παράδειγμα η αναγνώριση ομιλίας, η επεξεργασία εικόνας και φυσικής γλώσσας και η απόδειξη αυτόματων θεωρημάτων (Zhang & Lu, 2021)[20].

Σήμερα, η τεχνητή νοημοσύνη διαδραματίζει καίριο ρόλο στην κοινωνική ανάπτυξη, επιφέροντας επαναστατικά αποτελέσματα μεταξύ άλλων για την αποδοτικότητα της εργασίας, τη βελτιστοποίηση της δομής των ανθρώπινων πόρων και την ανάπτυξη νέων απαιτήσεων εργασίας (Zhang & Lu, 2021)[20].

Η Τεχνητή Νοημοσύνη αξιοποιείται μεταξύ άλλων τομέων και στα παρακάτω (Padmaja et al., 2024)[21]:

- Υγειονομική περίθαλψη. Η συνεισφορά της Τεχνητής Νοημοσύνη είναι ιδιαίτερα σημαντική στην υγειονομική περίθαλψη, δίνοντας τη δυνατότητα για βελτιωμένες διαγνώσεις, εξατομικευμένη θεραπεία και αποτελεσματική περίθαλψη ασθενών. Επίσης, έχει αξιοποιηθεί για την ανακάλυψη φαρμάκων.
- Οικονομικά. Η συνεισφορά της Τεχνητής Νοημοσύνης στη χρηματοοικονομική βιομηχανία δε μπορεί να υποτιμηθεί. Ειδικότερα, επιτρέπει τον εντοπισμό απάτης, την αξιολόγηση κινδύνου καθώς επίσης και τις εξατομικευμένες χρηματοοικονομικές υπηρεσίες. Ακόμη, μέσω της χρήσης τής εντοπίζεται η απάτη.
- Παραγωγή. Η χρήση της Τεχνητής Νοημοσύνης στην παραγωγή προϊόντων συντελεί στην αναδιαμόρφωση των διαδικασιών παραγωγής, έχοντας ως αποτέλεσμα τη βελτιστοποίηση των γραμμών παραγωγής και της ποιότητας των προϊόντας αλλά και τη μείωση του κόστους.
- Ψυχαγωγία: Η Τεχνητή Νοημοσύνη έχει συνεισφέρει και στη βιομηχανία της ψυχαγωγίας, παρέχοντας βελτιωμένες εμπειρίες στους χρήστες και ενισχύοντας τη δημιουργία περιεχομένου. Επιπλέον, χρησιμοποιούνται αλγόριθμοι παραγωγής περιεχομένου που βασίζονται σε AI για δημιουργία ρεαλιστικών γραφικών, μουσικής καθώς επίσης και εικονικών χαρακτήρων.

Είναι αξιοσημείωτο, πως οι πληροφορίες που παράγονται από άτομα σε ιστότοπους κοινωνικής δικτύωσης (SNSs) μπορεί να έχουν ηθικές συνέπειες για τις εταιρείες, όπως είναι για παράδειγμα η διάδοση παραπληροφόρησης και η διακύβευση της ασφάλειας και της εμπιστοσύνης. Ωστόσο, με τις εξελίξεις της τεχνητής νοημοσύνης και κυρίως με την ανάδειξη της βαθιάς μηχανικής μάθησης δίνεται η δυνατότητα για αντιμετώπιση αυτών των προκλήσεων (Ross et al., 2019; Wang et al., 2019)[22][23].

Πιο συγκεκριμένα, τα κοινωνικά ρομπότ εντός των ιστότοπων κοινωνικής δικτύωσης αποτελούν αυτόνομους φορείς οι οποίοι οδηγούνται από αλγόριθμους και λογισμικό που δημοσιεύουν περιεχόμενο στις εν λόγω πλατφόρμες. Διατίθενται και κακόβουλα ρομπότ, τα οποία αναπτύσσονται με σκοπό να βλάψουν, εξαπατώντας και χειραγωγώντας διαλόγους των μέσων κοινωνικής δικτύωσης, διαδίδοντας ψεύτικες ειδήσεις και έχοντας ως σκοπό την παραπληροφόρηση (Kudugunta & Ferrara, 2018)[24]. Ωστόσο, η τεχνολογία συμβάλλει και στον εντοπισμό φημών δίνοντας έμφαση στη μηχανική μάθηση και τη βαθιά μάθηση, που αναλύονται στα επόμενα κεφάλαια (Zhao et al., 2023)[25].

Κεφάλαιο 2ο: Μηχανική Μάθηση

2.1 Ορισμός

Η Μηχανική Μάθηση (Machine Learning - ML) αποτελεί την επιστημονική μελέτη αλγορίθμων καθώς επίσης και στατιστικών μοντέλων που χρησιμοποιούν τα συστήματα υπολογιστών με σκοπό να υλοποιήσουν μια συγκεκριμένη εργασία χωρίς να έχει προγραμματιστεί ρητά. Αναλυτικότερα, από εκατομμύρια χρόνια πριν οι άνθρωποι χρησιμοποιούν διάφορα είδη εργαλείων προκειμένου να ολοκληρώσουν εργασίες με απλό τρόπο. Αργότερα, η δημιουργικότητα του ανθρώπινου εγκεφάλου συνέβαλε στην εφεύρεση μηχανών, καθιστώντας την ανθρώπινη ζωή ευκολότερη. Ειδικότερα, μέσω αυτών τους δόθηκε η δυνατότητα να ανταποκρίνονται σε διάφορες ανάγκες τους, όπως τα ταξίδια και η βιομηχανία, συμπεριλαμβανομένης της μηχανικής μάθησης.

Η μηχανική μάθηση ορίζεται ως το πεδίο σπουδών που δίνει στους υπολογιστές τη δυνατότητα να μαθαίνουν χωρίς να είναι ρητά προγραμματισμένοι. Χρησιμοποιείται για να διδάξει τις μηχανές πώς να χειρίζονται τα δεδομένα πιο αποτελεσματικά. Σε ορισμένες περιπτώσεις, έπειτα από την προβολή των δεδομένων, δεν είναι εφικτή η ερμηνεία των πληροφοριών που εξάγονται και εφαρμόζεται η μηχανική εκμάθηση. Καθίσταται αξιοσημείωτο, πως λόγω της αφθονίας των διαθέσιμων συνόλων δεδομένων, υπάρχει αυξημένη ζήτηση για μηχανική μάθηση. Πολλές βιομηχανίες εφαρμόζουν μηχανική εκμάθηση για την εξαγωγή σχετικών δεδομένων και ο απώτερος σκοπός της είναι η μάθηση μέσω των δεδομένων (Mahesh, 2018)[26].

Σύμφωνα με τους Mahammad και Bakirova (2021)[27] η μηχανική μάθηση συνιστά μια σημαντική λειτουργικότητα και κλάδο της Τεχνητής Νοημοσύνης και οι εφαρμογές αυτής της τεχνολογίας είναι ευρείες. Ειδικότερα, περιλαμβάνει μεταξύ άλλων την αναγνώριση ομιλίας, την αναγνώριση εικόνων, τα spam email και το φιλτράρισμα κακόβουλου λογισμικού, την ανίχνευση απάτης στο διαδίκτυο, τα αυτό-οδηγούμενα αυτοκίνητα και την ιατρική.

2.2 Βασικές έννοιες και εφαρμογές μηχανικής μάθησης

Η μηχανική μάθηση ασχολείται με δεδομένα και με σύνολα δεδομένων. Ένα σύνολο δεδομένων περιλαμβάνει πολλαπλά σημεία δεδομένων, τα οποία καλούνται και ως δείγματα. Κάθε σημείο δεδομένων αντιπροσωπεύει μια οντότητα που πρόκειται να αναλυθεί (Badillo et al., 2020)[28].

Πολλοί αλγόριθμοι μηχανικής μάθησης έχουν αναπτυχθεί προκειμένου να καλύψουν την ποικιλία δεδομένων και τους τύπους προβλημάτων που προκύπτουν. Εννοιολογικά, οι αλγόριθμοι μηχανικής μάθησης σύμφωνα με τους Jordan και Mitchell (2015, 255)[29] «*μπορούν να θεωρηθούν ως αναζήτηση σε ένα μεγάλο χώρο υποψήφιων προγραμμάτων, με γνώμονα την εκπαιδευτική εμπειρία, για την εύρεση ενός προγράμματος που βελτιστοποιεί τη μέτρηση απόδοσης*».

Επίσης, οι αλγόριθμοι μηχανικής μάθησης ποικίλλουν σημαντικά από τον τρόπο με τον οποίο αντιπροσωπεύουν τα υποψήφια προγράμματα. Ακόμη, πολλοί αλγόριθμοι εστιάζουν σε προβλήματα προσέγγισης συναρτήσεων, όπου η εργασία ενσωματώνεται σε μια συνάρτηση. Για παράδειγμα μπορεί να αναφερθεί η περίπτωση όπου δίνεται μια συναλλαγή εισόδου, εξάγεται μια ετικέτα "απάτη" είτε "όχι απάτη". Το πρόβλημα μάθησης είναι να επιτευχθεί η βελτίωση της ακρίβειας της συνάρτησης. με

εμπειρία που αποτελείται από ένα δείγμα γνωστών ζευγών εισόδου-εξόδου της συνάρτησης (Jordan & Mitchell, 2015)[29].

Μια ακόμη έννοια που παρουσιάζει ενδιαφέρον είναι τα νευρωνικά δίκτυα. Η ευέλικτη δομή των νευρωνικών δικτύων επιτρέπει την τροποποίηση για πολλά πλαίσια και τύπους μηχανικής μάθησης. Εμπνευσμένα από την αρχή της επεξεργασίας πληροφοριών σε βιολογικά συστήματα, τα τεχνητά νευρωνικά δίκτυα αποτελούνται από μαθηματικές αναπαραστάσεις συνδεδεμένων μονάδων επεξεργασίας, που καλούνται ως τεχνητοί νευρώνες. Επιπλέον, όπως στην περίπτωση των συνάψεων του εγκεφάλου, η σύνδεση μεταξύ νευρώνων μεταδίδει σήματα. Η ισχύς των σημάτων μπορεί να ενισχυθεί είτε να εξασθενήσει από ένα βάρος που προσαρμόζεται μέσω της μαθησιακής διαδικασίας (Janiesch et al., 2021)[30].

Αναφορικά με τα σήματα επεξεργάζονται από επόμενους νευρώνες εάν ξεπεραστεί ένα ορισμένο όριο όπως καθορίζεται από μια συνάρτηση ενεργοποίησης. Συνήθως, οι νευρώνες οργανώνονται σε δίκτυα με διαφορετικά επίπεδα. Ένα επίπεδο εισόδου λαμβάνει τα δεδομένα εισαγωγής, όπως για παράδειγμα εικόνες προϊόντων σε ένα ηλεκτρονικό κατάστημα και ένα επίπεδο εξόδου παράγει το τελικό αποτέλεσμα, όπως για παράδειγμα κατηγοριοποίηση προϊόντων. Σε αυτό το σημείο είναι σημαντικό να αναφερθεί, πως ενδιάμεσα υπάρχουν μηδέν ή περισσότερα κρυφά επίπεδα που είναι υπεύθυνα για την εκμάθηση μιας μη γραμμικής αντιστοίχισης μεταξύ εισόδου και εξόδου, ενώ ο αριθμός των επιπέδων και των νευρώνων δεν μπορεί να μαθευτεί από τον αλγόριθμο εκμάθησης (Janiesch et al., 2021)[30].

Η μηχανική μάθηση εφαρμόζεται σε πολλά πεδία παρέχοντας σημαντικά οφέλη. Ειδικότερα, παρέχει οφέλη στην υγειονομική περίθαλψη, στον στρατό, στα συστήματα αυτοκινήτων ή στη δημιουργία φωνητικών διεπαφών και φωνητικών βοηθών στην καθημερινή ζωή, καθώς συμβάλλει στη βελτίωση της προσβασιμότητας. Ένας από τους τομείς οφέλους στη μηχανική μάθηση είναι ο τραπεζικός και ο χρηματοοικονομικός τομέας, όπου οι πιθανότητες ανίχνευσης απάτης είναι υψηλές σε περίπτωση που οι χρηματικές συναλλαγές γίνονται ηλεκτρονικά. Τόσο ο εντοπισμός όσο και η πρόληψη της απάτης επιτυγχάνονται βάσει του εντοπισμού προτύπων στις συναλλαγές πελατών, τον εντοπισμό περίεργης συμπεριφοράς και τα πιστωτικά όρια (Sharma et al., 2021)[31].

Η συμβολή της μηχανικής μάθησης είναι φανερή και στην υγειονομική περίθαλψη, καθώς βοηθά μεταξύ άλλων στη διαλογή και τις διαγνώσεις περιστατικών, στη βελτίωση της σάρωσης και την τμηματοποίηση εικόνας, τη λήψη αποφάσεων, την πρόβλεψη κινδύνου ασθένειας και τη νευροαπεικόνιση (Habehh & Gohel, 2021)[32].

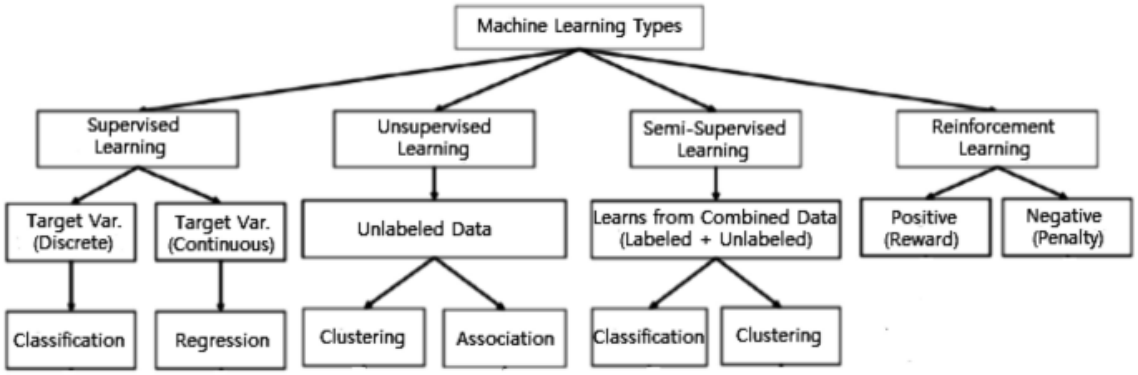
2.3 Τύποι μηχανικής μάθησης

Έχει διεξαχθεί πλήθος μελετών σχετικά με τον τρόπο που μπορούν οι μηχανές να μαθαίνουν μόνες τους χωρίς να είναι ρητά προγραμματισμένες. Μαθηματικοί αλλά και προγραμματιστές εφαρμόζουν διάφορες προσεγγίσεις προκειμένου να βρουν τη λύση (Mahesh, 2018)[26].

Οι αλγόριθμοι Μηχανικής Μάθησης διακρίνονται σε τέσσερις κατηγορίες (Sarker, 2021)[33], όπως φαίνεται στην Εικόνα 1:

1) Εποπτευόμενη μάθηση
2) Μη εποπτευόμενη μάθηση
3) Ημι-εποπτευόμενη μάθηση
4) Ενισχυτική μάθηση.

Εικόνα 1. Τύποι τεχνικών μηχανικής μάθησης

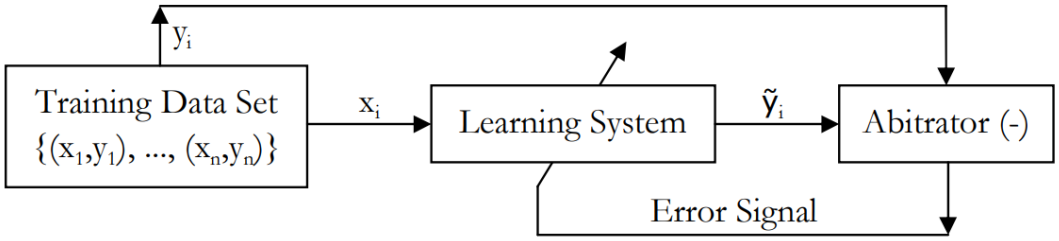


Πηγή: https://link.springer.com/article/10.1007/s42979-021-00592-x

Εποπτευόμενη μάθηση. Σχετικά με την εποπτευόμενη μάθηση καθιστά έναν τύπο μηχανικής μάθησης που έχει ως σκοπό την απόκτηση των πληροφοριών για τη σχέση εισόδου-εξόδου ενός συστήματος το οποίο είναι βασισμένο σε ένα σύνολο ζευγών δειγμάτων εκπαίδευσης εισόδου-εξόδου. Η έξοδος θεωρείται πως είναι η ετικέτα των δεδομένων εισόδου είτε της επίβλεψης, ενώ ένα δείγμα εκπαίδευσης εισόδου-εξόδου καλείται ως ετικετοποιημένα δεδομένα εκπαίδευσης είτε ως εποπτευόμενα δεδομένα. Η εποπτευόμενη μάθηση αποσκοπεί στην κατασκευή ενός τεχνητού συστήματος το οποίο είναι εφικτό να μάθει τη χαρτογράφηση μεταξύ της εισόδου και της εξόδου και έχει τη δυνατότητα να προβλέψει την έξοδο του εν λόγω συστήματος με νέες εισόδους. Στην περίπτωση που η έξοδος λάβει ένα πεπερασμένο σύνολο τιμών που υποδεικνύουν τις ετικέτες κλάσεων της εισόδου, τότε η εκμάθηση αντιστοίχισης συντελεί στην ταξινόμηση των δεδομένων εισόδου. Ενώ, στην περίπτωση που η έξοδος λάβει συνεχείς τιμές, έχει ως αποτέλεσμα την παλινδρόμηση της εισόδου. Σχετικά με τις πληροφορίες της σχέσης εισόδου-εξόδου αναπαρίστανται συνήθως με παραμέτρους μοντέλου μάθησης. Αν οι εν λόγω παράμετροι δεν καθίστανται άμεσα διαθέσιμες από δείγματα εκπαίδευσης, τότε ένα σύστημα εκμάθησης είναι σημαντικό να περάσει από μια διαδικασία εκτίμησης με σκοπό να αποκτήσει τις συγκεκριμένες παραμέτρους (Liu & Wu, 2012)[34].

Στην παρακάτω Εικόνα (Εικόνα 2) απεικονίζεται ένα διάγραμμα που δείχνει τη μορφή της Εποπτευόμενης Μάθησης. Σε αυτό το διάγραμμα, (x_i,y_i) φαίνεται ένα εποπτευόμενο δείγμα εκπαίδευσης, όπου το «x» αντιπροσωπεύει την είσοδο του συστήματος, ενώ το «y» αντιπροσωπεύει την έξοδο του συστήματος, που πρόκειται είτε για την επίβλεψη είτε για την επισήμανση του είσοδος x, όπως επίσης και το «i» που αποτελεί το δείκτης του δείγματος εκπαίδευσης. Στο πλαίσιο της διαδικασίας εποπτευόμενης μάθησης, μια εισαγωγή εκπαίδευσης x_i τροφοδοτείται στο σύστημα εκμάθησης και το σύστημα εκμάθησης αναπτύσσει μια έξοδο $\tilde{y}_i$. Έπειτα, η έξοδος του συστήματος εκμάθησης $\tilde{y}_i$ συγκρίνεται με τη σήμανση βασικής αλήθειας y_i και υπολογίζεται η μεταξύ τους διαφορά, η οποία καλείται ως Σήμα Σφάλματος. Ακολούθως, αποστέλλεται στο σύστημα εκμάθησης για προσαρμογή των παραμέτρων του εκπαιδευόμενου. Αυτή η μαθησιακή διαδικασία έχει ως σκοπό την απόκτηση βέλτιστων παραμέτρων του συστήματος εκμάθησης που ενδέχεται να ελαχιστοποιήσουν τις διαφορές μεταξύ $\tilde{y}i$ και yi για όλα i, ελαχιστοποιώντας το συνολικό σφάλμα στο σύνολο των δεδομένων εκπαίδευσης (Liu & Wu, 2012)[34].

Εικόνα 2. Διάγραμμα μορφής εποπτευόμενης μάθησης

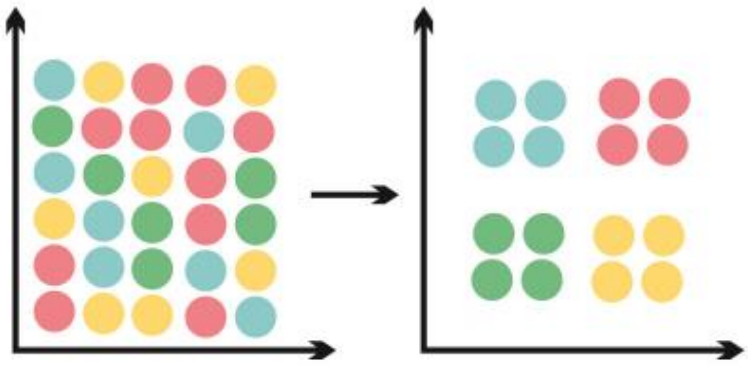


Πηγή: https://www.researchgate.net/publication/229031588_Supervised_Learning

Μη εποπτευόμενη μάθηση. Στη μη εποπτευόμενη μάθηση χωρίς επίβλεψη, ένας επιστήμονας δεδομένων δίνει φωτογραφίες και εναπόκειται στο σύστημα να εξετάσει τα δεδομένα και να καθορίσει για τι εικόνες πρόκειται. Είναι αξιοσημείωτο, πως απαιτούνται μεγάλες ποσότητες δεδομένων στην περίπτωση της μη εποπτευόμενης μηχανικής μάθησης. Όταν οι επιστήμονες δεδομένων χρησιμοποιούν σύνολα δεδομένων για την εκπαίδευση αλγορίθμων, ξεκινά η διαδικασία μη εποπτευόμενης μάθησης. Τα σύνολα δεδομένων δεν περιλαμβάνουν σημεία δεδομένων με ετικέτα ή ταξινόμηση, ενώ ο σκοπός της εκμάθησης του αλγορίθμου καθίσταται η εύρεση μοτίβων στο σύνολο δεδομένων καθώς επίσης και η βαθμολόγηση των σημείων δεδομένων βάσει των εν λόγω μοτίβων. Τα ζητήματα ομαδοποίησης, συσχέτισης, ανίχνευσης ανωμαλιών και αυτόματου κωδικοποιητή είναι τέσσερις τύποι προκλήσεων μάθησης χωρίς επίβλεψη, από όπου προκύπτουν οι τύποι μη εποπτευόμενης μάθησης, που παρατίθενται παρακάτω (Naeem et al., 2023)[35]:

1) Ομαδοποίηση. Πρόκειται για μια πρακτική ταξινόμησης στοιχείων σε ομάδες που ονομάζεται ομαδοποίηση ή ανάλυση συστάδων. Η ομαδοποίηση είναι δυνατόν να χωριστεί σε διάφορες μορφές, όπως είναι η κατάτμηση, η ιεράρχηση, αλλά και η επικαλυπτόμενη και η πιθανολογική μορφή. Τα δεδομένα χωρίζονται με τέτοιο τρόπο ώστε καθεμία πληροφορία να ανήκει αποκλειστικά σε ένα σύμπλεγμα. Χρησιμοποιείται για την οργάνωση δεδομένων σε επικαλυπτόμενα ασαφή σύνολα (Εικόνα 3).

Εικόνα 3. Παράδειγμα ομαδοποίησης μη εποπτευόμενης μάθησης



Πηγή: https://journals.uob.edu.bh/handle/123456789/4777

2) Συσχέτιση. Ο συγκεκριμένος τύπος χρησιμοποιείται για την αποκάλυψη συσχετίσεων μεταξύ μεταβλητών σε περιπτώσεις μεγάλων συνόλων δεδομένων. Αυτός ο τύπος μπορεί να δέχεται μη αριθμητικά σημεία δεδομένων, σε αντίθεση με άλλες μεθόδους μηχανικής μάθησης. Πιο συγκεκριμένα, ασχολείται με το πώς συνδέονται συγκεκριμένες μεταβλητές.
3) Ανίχνευση ανωμαλιών. Κάθε διαδικασία ανακαλύπτει ακραίες τιμές στο πλαίσιο ενός συνόλου δεδομένων και καλείται ως ανίχνευση ανωμαλιών. Οι εν λόγω ανωμαλίες μπορεί να υποδηλώνουν ασυνήθιστη δραστηριότητα δικτύου, ελαττωματικό αισθητήρα καθώς επίσης και δεδομένα τα οποία είναι αναγκαίο να καθαριστούν πριν γίνει η ανάλυση. Στην περίπτωση που τα μοντέλα δεδομένων είτε υπερβαίνουν είτε αποκλίνουν από τα συνηθισμένα μοντέλα, τότε θεωρείται πως υπάρχει μια ανωμαλία. Ως παράδειγμα μπορεί να αναφερθεί η περίπτωση όπου ένα ασυνήθιστο μοτίβο κίνησης δικτύου, φανερώνει ότι το παραβιασμένο σύστημα μεταφέρει ευαίσθητα δεδομένα σε μη εξουσιοδοτημένο διακομιστή. Σε αυτό το σημείο είναι σημαντικό να αναφερθεί, πως οι ανωμαλίες μπορεί να εντοπίζονται είτε να προβλέπονται με την εύρεση και την πρόβλεψη σημείων δεδομένων τα οποία διαφέρουν σε σχέση με ένα το τυπικό μοντέλο. Ορισμένες τεχνικές για τον εντοπισμό ανωμαλιών είναι η ανίχνευση εισβολής, η ασφάλιση και η ανίχνευση απάτης (Prasad & Balakrishnan, 2022)[36].
4) Αυτό-κωδικοποιητές. Οι αυτό-κωδικοποιητές αποτελούν μια προσέγγιση μη εποπτευόμενης μάθησης, η οποία χρησιμοποιεί νευρωνικά δίκτυα με σκοπό να πραγματοποιηθεί εκμάθηση αναπαράστασης. Ειδικότερα, δημιουργείται μια αρχιτεκτονική νευρωνικού δικτύου με ένα σημείο συμφόρησης και αναγκάζει το δίκτυο να προβεί σε χρήση μιας συμπιεσμένης αναπαράστασης γνώσης της αρχικής εισόδου. Η συγκεκριμένη συμπίεση αλλά και η επακόλουθη ανακατασκευή καθίστανται περίπλοκες στην περίπτωση που οι ιδιότητες εισόδου δεν σχετίζονται. Στην περίπτωση που τα δεδομένα έχουν κάποια δομή (για παράδειγμα, συσχετίσεις μεταξύ των χαρακτηριστικών εισόδου), τότε αυτή ενδέχεται να μαθευτεί και να αξιοποιηθεί οδηγώντας την είσοδο μέσω του σημείου συμφόρησης του δικτύου. Ακόμη, η παρουσία ενός σημείου συμφόρησης πληροφοριών αποτελεί κύριο χαρακτηριστικό του σχεδιασμού του δικτύου, διότι αν απουσιάζει μπορεί να μάθει γρήγορα να αποθηκεύει δεδομένα εισόδου μέσω του ιστού (Kottmann et al., 2020)[37]. Στην παρακάτω Εικόνα (Εικόνα 4) απεικονίζονται διαγραμματικά τα στοιχεία του αλγόριθμου του αυτόματου κωδικοποιητή.

Εικόνα 4. Στοιχεία αλγόριθμου αυτόματου κωδικοποιητή

Πηγή: https://journals.uob.edu.bh/handle/123456789/4777

Ημι- εποπτευόμενη μάθηση. Σχετικά με την ημι-εποπτευόμενη μάθηση (SSL) αποτελεί έναν τύπο μηχανικής μάθησης, ο οποίος εντάσσεται μεταξύ εποπτευόμενης και μη εποπτευόμενης μάθησης και έχει ως στόχο να ξεπεράσει τα μειονεκτήματα αυτών των τύπων μάθησης. Ειδικότερα, η εποπτευόμενη μάθηση απαιτεί τεράστιο όγκο δεδομένων εκπαίδευσης για την ταξινόμηση των δεδομένων δοκιμής, καθιστώντας την αρκετά κοστοβόρα και χρονοβόρα διαδικασία. Από την άλλη πλευρά, η μη εποπτευόμενη μάθηση δεν απαιτεί δεδομένα με ετικέτα κι έτσι δεν είναι εφικτό να ομαδοποιήσει με ακρίβεια άγνωστα δεδομένα. Προκειμένου να ξεπεραστούν τα προαναφερθέντα ζητήματα, ο συγκεκριμένος τύπος έχει προταθεί από την ερευνητική κοινότητα, καθώς μπορεί να επιτευχθεί η μάθηση με μικρό όγκο δεδομένων εκπαίδευσης και να ονομάσει τα άγνωστα δεδομένα ή τα δεδομένα δοκιμής. Αναλυτικότερα, αναπτύσσει ένα μοντέλο με λίγα μοτίβα με ετικέτα ως δεδομένα εκπαίδευσης και αντιμετωπίζει τα υπόλοιπα μοτίβα ως δεδομένα δοκιμής. Ακόμη, η ημι-εποπτευόμενη μάθηση χωρίζεται στους δύο παρακάτω τύπους, την ημι-εποπτευόμενη ταξινόμηση και την ημι-εποπτευόμενη ομαδοποίηση (Reddy et al., 2018)[38].

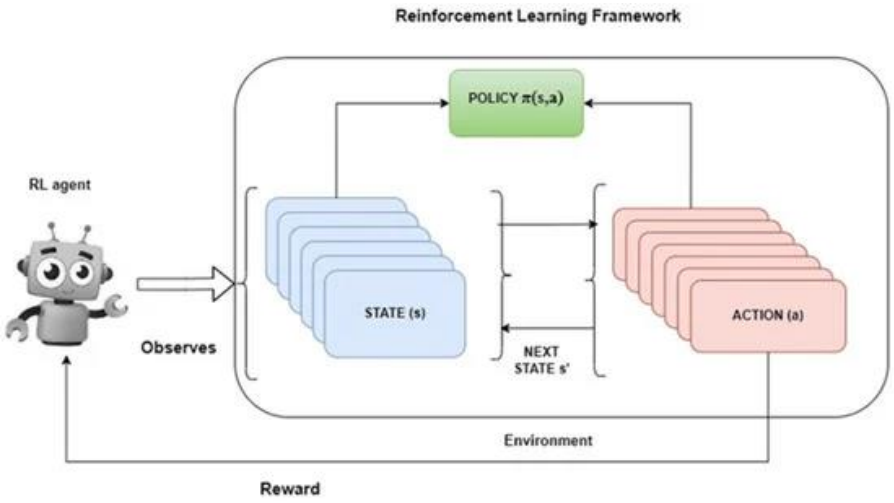
1. Ημι-εποπτευόμενη ταξινόμηση. Η ημι-εποπτευόμενη ταξινόμηση (Semi-Supervised Classification - SSC) ομοιάζει σε μεγάλο βαθμό με την εποπτευόμενη προσέγγιση. Ωστόσο, χρησιμοποιεί λιγότερα δεδομένα για την ταξινόμηση μεγάλου όγκου δεδομένων δοκιμής. Χρησιμοποιώντας την μειώνεται η χρήση των δεδομένων εκπαίδευσης (Reddy et al., 2018)[38].

2. Ημι-εποπτευόμενη ομαδοποίηση (Semi-supervised Clustering). Στην προκειμένη περίπτωση, εκτός από τις πληροφορίες ομοιότητας οι οποίες χρησιμοποιούνται από τη μη εποπτευόμενη ομαδοποίηση, μπορεί να είναι διαθέσιμη μια μικρή ποσότητα γνώσης σχετικά με περιορισμούς είτε κατά ζεύγη (must-link ή not not-link) μεταξύ στοιχείων δεδομένων είτε ετικετών κλάσεων για ορισμένα στοιχεία. Αντί να γίνεται χρήση αυτής της γνώσης για την εξωτερική επικύρωση των αποτελεσμάτων της ομαδοποίησης, μπορεί να «καθοδηγήσει» είτε να «προσαρμόσει» τη διαδικασία ομαδοποίησης, με άλλα λόγια να παρέχει μια περιορισμένη μορφή εποπτείας. Η προσέγγιση η οποία προκύπτει καλείται ως ημι-εποπτευόμενη ομαδοποίηση (Grira et al., 2004)[39].

Ενισχυτική μάθηση. Ο τέταρτος τύπος μηχανικής μάθησης είναι η ενισχυτική μάθηση, η οποία είναι κατάλληλη για διαδοχικές διαδικασίες λήψης αποφάσεων και η μάθηση επιτυγχάνεται μέσα από την άμεση αλληλεπίδραση με το περιβάλλον προκειμένου να επιτευχθούν μακροπρόθεσμοι στόχοι χωρίς εξωτερικό κίνητρο είτε πλήρης γνώση του περιβάλλοντος. Στην παρακάτω εικόνα (Εικόνα 5) ο πράκτορας ενισχυτικής μάθησης εξετάζει την κατάσταση του περιβάλλοντος και εν συνεχεία επιλέγει την κατάλληλη ενέργεια. Εφόσον προβεί στη σωστή ενέργεια, δέχεται θετική ανταμοιβή, ενώ στην περίπτωση που γίνει λάθος κίνηση, λαμβάνεται αρνητική ανταμοιβή. Η ενισχυτική μάθηση πρέπει να επιτύχει την εξισορρόπηση μεταξύ της εξερεύνησης και της εκμετάλλευσης. Αναφορικά με την

εκμετάλλευση συμβαίνει αν ένας πράκτορας επιχειρεί να μεγιστοποιήσει την ανταμοιβή με βάση μια προηγουμένως καθορισμένη διαδρομή. Ενώ, στην περίπτωση που προσπαθεί να εξερευνήσει έναν νέο τρόπο ώστε να φτάσει στον προορισμό, τότε αυτή η διαδικασία ονομάζεται εξερεύνηση. Αυτός ο τύπος μηχανικής μάθησης δε χρειάζεται μεγάλο σύνολο δεδομένων και μαθαίνει με δοκιμή και σφάλμα (Sivamayil et al., 2023)[40].

Εικόνα 5. Πλαίσιο ενισχυτικής μάθησης



Πηγή: https://www.mdpi.com/1996-1073/16/3/1512

2.4 Αλγοριθμικές τεχνικές μηχανικής μάθησης

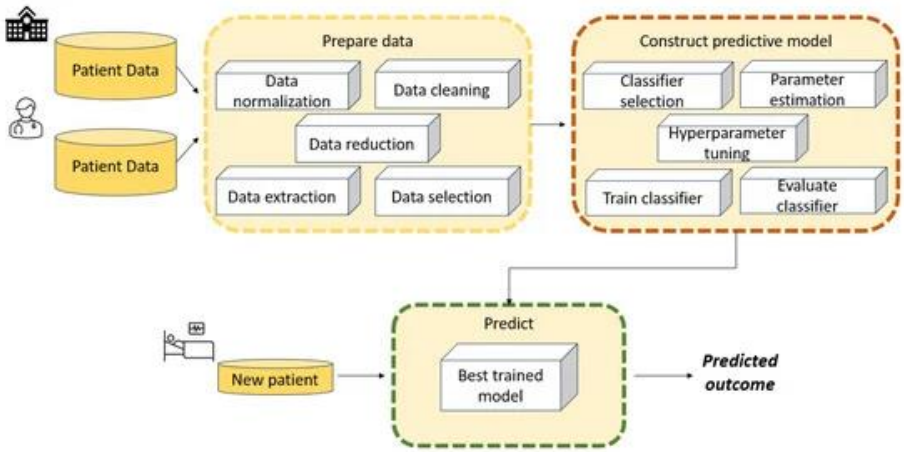
- Μηχανή Διανυσμάτων Υποστήριξης (Support Vector Machine – SVM)

Η μηχανή διανυσμάτων υποστήριξης (SVM) αποτελεί ένα σύνολο εποπτευόμενων μεθόδων εκμάθησης, οι οποίες αξιοποιούνται με σκοπό την ταξινόμηση, την παλινδρόμηση αλλά και ανίχνευση ακραίων στοιχείων. Ειδικότερα, το SVM αναπτύχθηκε από τους Cortes και Vapnik στο AT&T Bell Laboratories το 1995 και λειτουργούσε βάσει στατιστικών πλαισίων μάθησης και συγκεκριμένα με βάση τη θεωρία Vapnik–Chervonenkis (1989)[41]. Είναι εφικτό να εφαρμοστεί σε γραμμικές όπως επίσης και σε μη γραμμικές ταξινομήσεις. Αναφορικά με το γραμμικό SVM χρησιμοποιεί ένα υπερεπίπεδο μέγιστου περιθωρίου (hard-margin ή soft-margin), ενώ το μη γραμμικό SVM χρησιμοποιεί πυρήνες για ταξινομήσεις. Επιπροσθέτως, το SVM χρησιμοποιείται για μάθηση χωρίς επίβλεψη, η οποία ονομάζεται Ομαδοποίηση Διανυσμάτων Υποστήριξης (Support Vector Clustering - SVC). Ακόμη, οι αλγόριθμοι SVM χρησιμοποιούν διαφορετικούς τύπους συναρτήσεων πυρήνα, όπου οι πιο κοινές συναρτήσεις είναι γραμμικές, μη γραμμικές, πολυωνυμικές καθώς επίσης και σιγμοειδές (Cemiloglu et al., 2023)[42].

Στην Εικόνα 6 παρουσιάζονται τα κύρια στάδια, τα οποία εμπλέκονται στην ανάπτυξη ενός μοντέλου ML, όπως ένα SVM με σκοπό την επίλυση ενός διαγνωστικού προβλήματος είτε προβλήματος παλινδρόμησης. Αρχικά, συλλέγονται δεδομένα από διάφορες πηγές, όπως βάσεις δεδομένων. Έπεται η φάση επεξεργασίας δεδομένων, η οποία στοχεύει στο χειρισμό τιμών που λείπουν, ακραίων τιμών αλλά και ασυνέπειας. Στις τεχνικές προεπεξεργασίας δεδομένων μπορεί να συγκαταλέγεται μεταξύ

άλλων, ο καθαρισμός δεδομένων και η κωδικοποίηση κατηγορικών μεταβλητών. Στο επόμενο στάδιο, η μηχανική χαρακτηριστικών περιλαμβάνει την επιλογή, τη μετατροπή κι εν συνεχεία την ανάπτυξη νέων χαρακτηριστικών από ακατέργαστα δεδομένα με σκοπό τη βελτίωση της απόδοσης των μοντέλων SVM. Ακολουθεί η μοντελοποίηση δεδομένων που είναι ένα βασικό συστατικό της αρχιτεκτονικής, και απαιτεί προσεκτική εξέταση της προεπεξεργασίας δεδομένων και των τεχνικών αξιολόγησης με σκοπό την δημιουργία ακριβών συστημάτων μηχανικής εκμάθησης. Τέλος, πραγματοποιείται η αξιολόγηση του μοντέλου (Guido et al., 2024)[43].

Εικόνα 6. Στάδια ανάπτυξης SVM



https://www.mdpi.com/2078-2489/15/4/235

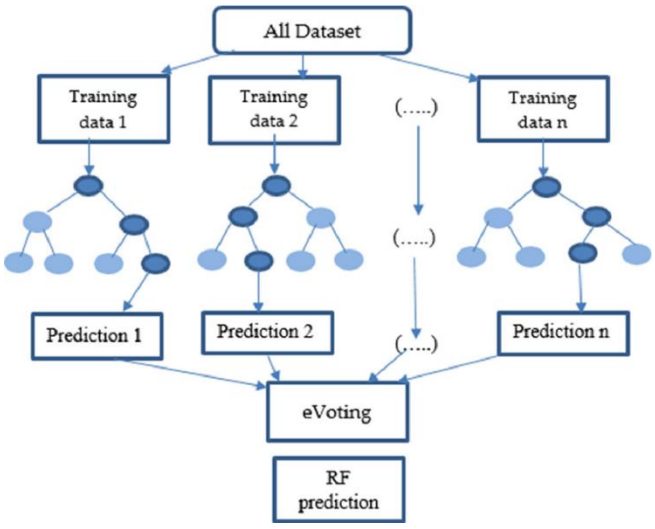
- *k*-Nearest Neighbor (KNN)

Σχετικά με τον αλγόριθμο k- Nearest Neighbor (kNN) αποτελεί έναν δημοφιλή αλγόριθμο μηχανικής μάθησης. Πιο συγκεκριμένα, η κύρια ιδέα τού είναι η πρόβλεψη της ετικέτας ενός στιγμιότυπου ερωτήματος με βάση τις ετικέτες των k πλησιέστερων στιγμιότυπων στα αποθηκευμένα δεδομένα, κάνοντας την υπόθεση ότι η ετικέτα ενός στιγμιότυπου παρομοιάζει εκείνη των στιγμών του k NN. Ο εν λόγω αλγόριθμος είναι εύκολος σε ότι αφορά στην εφαρμογή και αποτελεσματικός σχετικά με την απόδοση πρόβλεψης. Λαμβάνοντας υπόψη αυτά τα πλεονεκτήματα από τους ερευνητές, έχει αξιοποιηθεί σε διάφορες εργασίες εποπτευόμενης μάθησης, συμπεριλαμβανομένων των εργασιών ταξινόμησης αλλά και παλινδρόμησης (Kang, 2021)[44].

- Random Forest (RF)

Αναφορικά με το RF συνιστά μια εποπτευόμενη τεχνική μηχανικής εκμάθησης, η οποία είναι δυνατόν να χρησιμοποιηθεί για προβλήματα ταξινόμησης αλλά και για προβλήματα παλινδρόμησης. Αναλυτικότερα, αυτή η τεχνική βασίζεται σε μεγάλο αριθμό δέντρων αποφάσεων που λειτουργούν ως σύνολο, ενώ η τελική απόφαση βασίζεται στην πλειοψηφία των ψήφων για ταξινόμηση και η μέση πρόβλεψη θεωρείται η λύση στα προβλήματα παλινδρόμησης (Εικόνα 7) (Vergni & Todisco, 2023)[45].

Εικόνα 7. Αλγόριθμος Random Forest (RF)



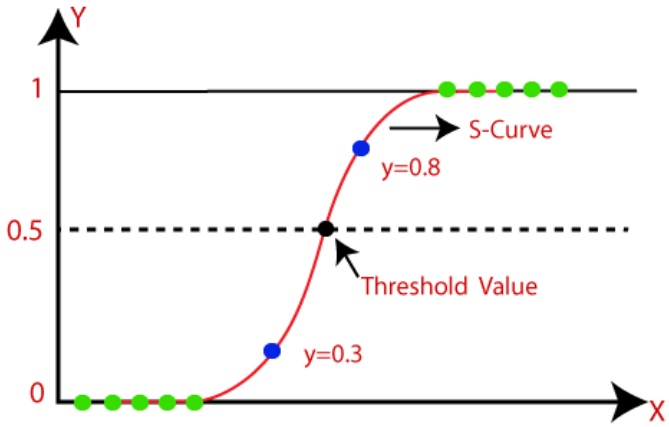
Πηγή: https://www.researchgate.net/figure/The-Processes-of-Random-Forest-RF-Algorithm_fig3_354076063

Τις περισσότερες φορές, το αρχικό σύνολο δεδομένων, το οποίο αποτελείται από μια μεταβλητή απόκρισης και μία είτε περισσότερες μεταβλητές πρόβλεψης (χαρακτηριστικά), συνιστά το υποσύνολο που σχηματίζει σύνολα δεδομένων εκπαίδευσης και επικύρωσης. Έπειτα, κάθε δέντρο απόφασης του δάσους λαμβάνεται από ένα δείγμα bootstrap του συνόλου δεδομένων εκπαίδευσης, κάνοντας χρήση μερικών τυχαία επιλεγμένων χαρακτηριστικών κατά την ανάπτυξη του δέντρου. Ακόμη, το σύνολο out of bag (OOB) περιλαμβάνει τα δεδομένα τα οποία δεν επιλέχθηκαν στη διαδικασία δειγματοληψίας ενός δέντρου. Σε ότι αφορά στα επόμενα βήματα, αυτά διαφέρουν ανάλογα με το στόχο (Vergni & Todisco, 2023)[45].

- Λογιστική Παλινδρόμηση (Logistic Regression - LR)

Η λογιστική παλινδρόμηση καθιστά ένα εποπτευόμενο αλγόριθμο μηχανικής μάθησης, ο οποίος αναπτύχθηκε με σκοπό την επίλυση προβλημάτων μάθησης ταξινόμησης. Ειδικότερα, ένα πρόβλημα μάθησης ταξινόμησης συναντάται, στην περίπτωση που η μεταβλητή στόχος είναι κατηγορηματική. Η λογιστική παλινδρόμηση έχει ως στόχο να χαρτογραφήσει μια συνάρτηση από τα χαρακτηριστικά του συνόλου δεδομένων στους στόχους προκειμένου να καταστεί εφικτή η πρόβλεψη της πιθανότητας ότι ένα νέο παράδειγμα ανήκει σε μία από τις κατηγορίες-στόχους (Bisong, 2019)[46].

Εικόνα 8. Αλγόριθμος λογιστικής παλινδρόμησης



Πηγή: https://www.researchgate.net/figure/Logistic-regression-algorithm-9_fig1_374387841

- Ακραία Ενίσχυση Κλάσης (Extreme Gradient Boosting – XGBoost)\

Η Ακραία Ενίσχυση Κλάσης αποτελεί μια κλιμακούμενη τεχνολογία μηχανικής μάθησης βελτιστοποίησης δέντρων. Η εν λόγω τεχνική προτάθηκε ως ένα εφαρμοσμένο μηχάνημα ενίσχυσης κλίσης, κυρίως σε περιπτώσεις δέντρων παλινδρόμησης και ταξινόμησης (Alshboul et al., 2022)[47].

Κεφάλαιο 3ο: Βαθιά Μάθηση

3.1 Ορισμός

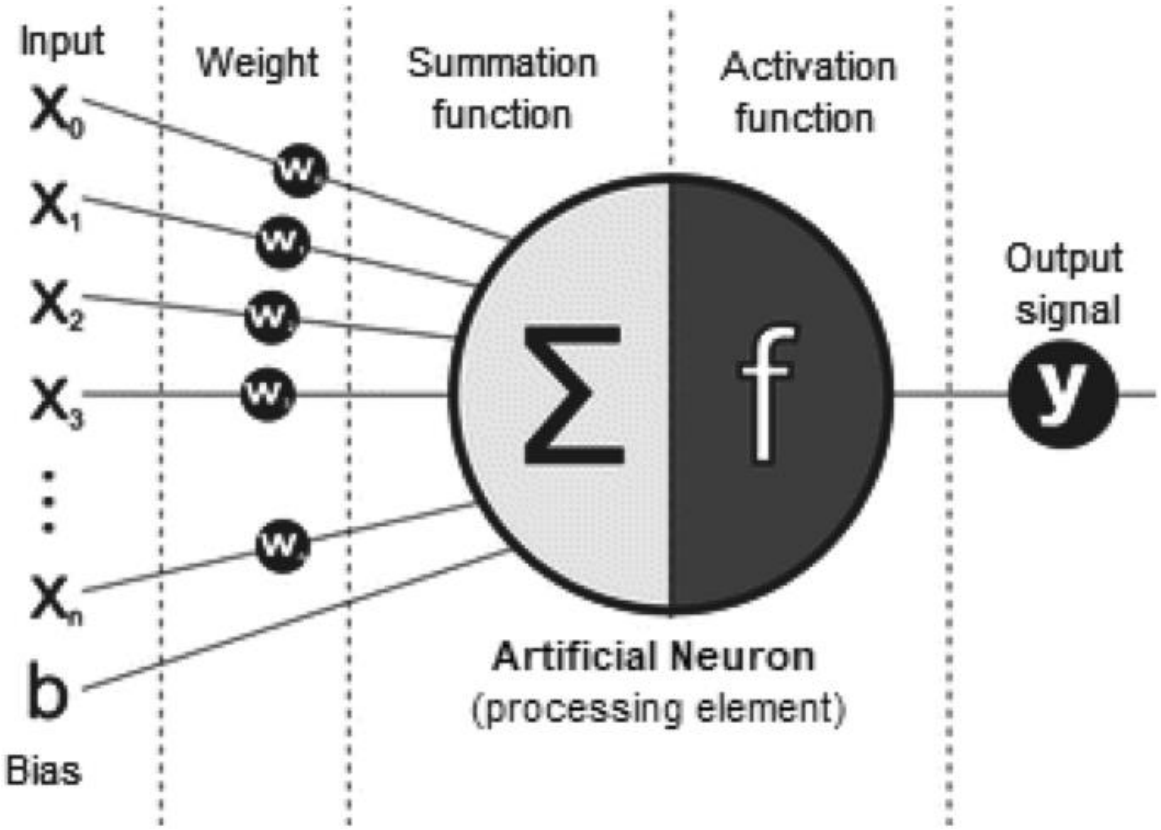
Η βαθιά μάθηση αποτελεί κλάδο της μηχανικής μάθησης και της τεχνητής νοημοσύνης. Πολλές εταιρείες, συμπεριλαμβανομένης της Google και της Microsoft μελετούν ενεργά τη βαθιά μάθηση καθώς μπορεί να παρέχει σημαντικά αποτελέσματα σε διαφορετικά προβλήματα ταξινόμησης, παλινδρόμησης και σύνολα δεδομένων. Αναλυτικότερα, η βαθιά μάθηση θεωρείται ως ένα υποσύνολο της μηχανικής μάθησης και της τεχνητής νοημοσύνης, και υπό αυτό το πλαίσιο μπορεί να θεωρηθεί ως μια λειτουργία τεχνητής νοημοσύνης που μιμείται την επεξεργασία δεδομένων του ανθρώπινου εγκεφάλου (Sarker, 2021)[48].

Η βαθιά μάθηση διαφέρει από την τυπική μηχανική μάθηση σχετικά με την αποτελεσματικότητα καθώς αυξάνεται ο όγκος των δεδομένων. Η βαθιά μάθηση χρησιμοποιεί πολλαπλά επίπεδα με σκοπό να αναπαραστήσει τις αφαιρέσεις δεδομένων για τη ανάπτυξη υπολογιστικών μοντέλων. Παρόλο που χρειάζεται αρκετό χρόνο για την εκπαίδευση ενός μοντέλου εξαιτίας μεγάλου αριθμού παραμέτρων, χρειάζεται μικρό χρονικό διάστημα εκτέλεσης κατά τη δοκιμή συγκριτικά με άλλους αλγόριθμους μηχανικής μάθησης.

Η βαθιά μάθηση θεωρείται η Τέταρτη Βιομηχανική Επανάσταση (4IR ή Industry 4.0) και επικεντρώνεται στον αυτοματισμό καθώς επίσης και τα ευφυή συστήματα, που βασίζονται στην τεχνολογία που προέρχεται από τα τεχνητά νευρωνικά δίκτυα (Sarker, 2021)[48].

Στην παρακάτω εικόνα (Εικόνα 9) φαίνεται ένα μαθηματικό μοντέλο ενός τεχνητού νευρώνα, που περιλαμβάνει στοιχείο επεξεργασίας, επισήμανση εισόδου\(X_i\)), βάρος (w), προκατάληψη (b), συνάρτηση άθροισης (\(\sum\)), λειτουργία ενεργοποίησης (f) και αντίστοιχο σήμα εξόδου (y). Η τεχνολογία βαθιάς μάθησης που βασίζεται σε νευρωνικά δίκτυα εφαρμόζεται πλέον ευρέως σε πολλούς τομείς όπως η υγειονομική περίθαλψη, η ανάλυση συναισθημάτων, η επεξεργασία φυσικής γλώσσας, η οπτική αναγνώριση, η επιχειρηματική ευφυΐα και η ασφάλεια στον κυβερνοχώρο (Sarker, 2021)[48].

Εικόνα 9. Μαθηματικό μοντέλο τεχνητού νευρώνα



Πηγή: https://link.springer.com/article/10.1007/s42979-021-00815-1

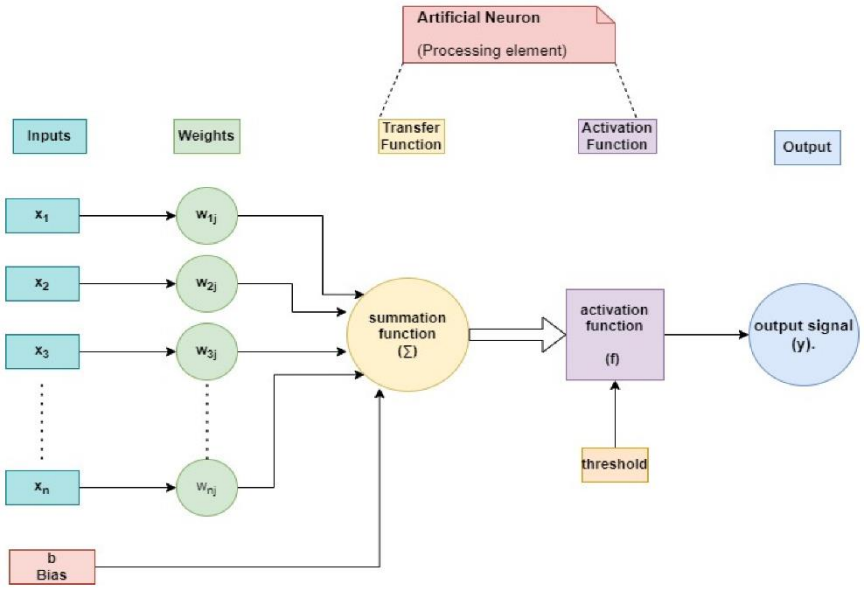
Η πρώτη κίνηση προς τα νευρωνικά δίκτυα χρονολογείται το 1943. Ειδικότερα, ο Warren McCulloch, νευροφυσιολόγος, και ένας μαθηματικός, ο Walter Pitts, προέβησαν στη δημιουργία μιας εργασίας για τον τρόπο λειτουργίας των νευρώνων. Εντός αυτού του πλαισίου πρότειναν ένα βασικό νευρωνικό δίκτυο με ηλεκτρικά κυκλώματα. Λίγα χρόνια αργότερα, το 1949 ο Donald Hebb θεώρησε ότι τα νευρικά μονοπάτια ενισχύονται κάθε φορά που χρησιμοποιούνται, ενώ τη δεκαετία του 1950, ο Nathanial Rochester πραγματοποίησε έρευνα για την προσομοίωση αφηρημένου νευρωνικού δικτύου σε υπολογιστές IBM (Dong et al., 2021)[49]. Σημαντικός παράγοντας σε ένα νευρωνικό δίκτυο καθίστανται οι λειτουργίες ενεργοποίησης που εμπνέονται από την ανθρώπινη νευρική πυροδότηση είτε τη μη πυροδότηση (Dong et al., 2021)[49].

3.2 Βασικές έννοιες βαθιάς μάθησης

Η εφαρμογή της τεχνολογίας βαθιάς μάθησης η οποία βασίζεται σε νευρωνικά δίκτυα καθίσταται ευρέως διαδεδομένη σε πολλούς κλάδους αλλά και τομείς μελέτης, όπως είναι για παράδειγμα η υγειονομική περίθαλψη και η ασφάλεια στον κυβερνοχώρο. Τόσο η δυναμική φύση όσο και η ποικιλία των πραγματικών συνθηκών αλλά και των δεδομένων καθιστούν δύσκολο τον σχεδιασμό ενός αποδεκτού μοντέλου βαθιάς μάθησης. Επιπροσθέτως, επικρατεί η πεποίθηση ότι τα μοντέλα βαθιάς μάθησης είναι εγγενώς μυστηριώδη και υπό αυτό το πλαίσιο περιορίζουν την ανάπτυξη του πεδίου της βαθιάς μάθησης (Taye, 2023)[50].

Πολλές μικρές συνδεδεμένες μονάδες επεξεργασίας, που είναι οι νευρώνες, αποτελούν τη ραχοκοκαλιά ενός τυπικού νευρωνικού δικτύου. Οι εν λόγω νευρώνες είναι υπεύθυνοι για την παραγωγή μιας ακολουθίας ενεργειών με πραγματική αξία, παρέχοντας συνδυαστικά το επιθυμητό αποτέλεσμα. Το μαθηματικό μοντέλο ενός τεχνητού νευρώνα είτε στοιχείου επεξεργασίας όπως φαίνεται στην Εικόνα 10 αποτελεί απλοποιημένη σχηματική μορφή. Η έξοδος σήματος που συμβολίζεται με (s) επισημαίνεται μαζί με την είσοδό του (Xi), το βάρος με (w), η μεροληψία με (b), η συνάρτηση άθροισης με (Σ), η συνάρτηση ενεργοποίησης με (f) καθώς επίσης και η σχετική είσοδο με (Xi) (y). Σχετικά με την τεχνολογία βαθιάς μάθησης υποστηρίζει ευφυή συστήματα αυτοματισμού που βασίζονται στην τεχνολογία (Schmidhuber, 2015; Mathew et al., 2021)[51][52].

Εικόνα 10. Μαθηματικό μοντέλο τεχνητού νευρώνα



Πηγή: https://www.mdpi.com/2073-431X/12/5/91

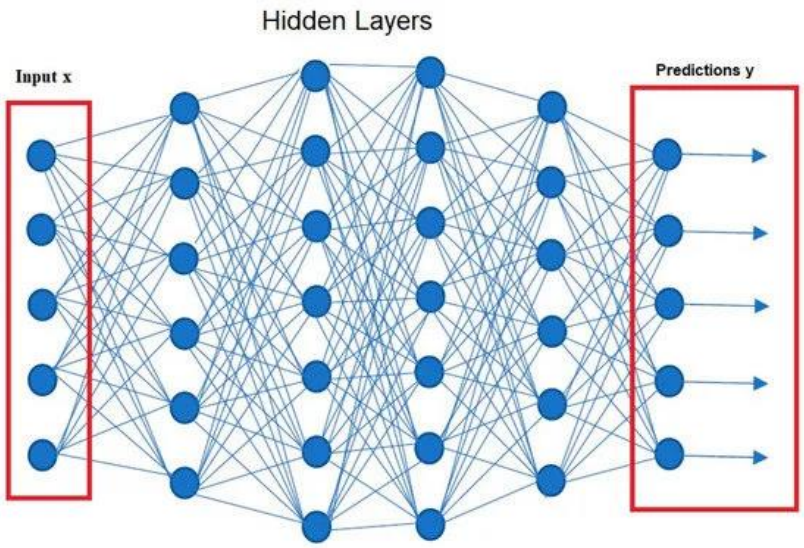
Καθίσταται αξιοσημείωτο, πως το θεμέλιο της βαθιάς μάθησης είναι οι τεχνητοί νευρώνες που μιμούνται τους νευρώνες του ανθρώπινου εγκεφάλου. Ένας τεχνητός νευρώνας μιμείται τη συμπεριφορά ενός πραγματικού νευρώνα λαμβάνοντας πληροφορίες από μια συλλογή εισόδων, σε καθεμία από τις οποίες δίνεται ένα συγκεκριμένο βάρος. Εντός αυτού του πλαισίου ο νευρώνας χρησιμοποιεί τις εν λόγω σταθμισμένες εισόδους με σκοπό να υπολογίσει μια συνάρτηση και να παράσχει μια έξοδο, ενώ Ν είσοδοι αποστέλλονται στον νευρώνα, δηλαδή μία για κάθε χαρακτηριστικό. Έπειτα, αθροίζονται οι είσοδοι, εκτελεί κάποιο είδος λειτουργίας σε αυτές και παράγει την έξοδο. Η σημασία μιας εισόδου μετριέται από το βάρος της, ενώ το νευρωνικό δίκτυο θα δίνει έμφαση σε εισόδους που έχουν μεγαλύτερο βάρος. Εξασφαλίζεται η καλύτερη δυνατή εφαρμογή μοντέλου-δεδομένων. Μια συνάρτηση ενεργοποίησης είναι ένας μετασχηματισμός μεταξύ εισόδων και αποτελεσμάτων, ενώ εφαρμογή ενός κατωφλίου οδηγεί σε έξοδο. Δεδομένου του γεγονότος πως ένας

μεμονωμένος νευρώνας δεν είναι σε θέση να επεξεργαστεί πολλές εισόδους, χρησιμοποιούνται πολλαπλοί νευρώνες για να καταλήξουμε σε ένα συμπέρασμα (Taye, 2023)[50].

Ένα νευρωνικό δίκτυο αποτελείται από perceptrons (Αντίληπτρα) που συνδέονται με διαφορετικούς τρόπους και λειτουργούν με διακριτές συναρτήσεις ενεργοποίησης. Κάθε νευρωνικό δίκτυο με περισσότερα από δύο επίπεδα θεωρείται μοντέλο βαθιάς μάθησης. Στην επεξεργασία δεδομένων, τα «κρυφά επίπεδα» αναφέρονται στα ενδιάμεσα επίπεδα μεταξύ εισόδου και εξόδου. Τα συγκεκριμένα στρώματα μπορούν να ενισχύσουν την ακρίβεια (Taye, 2023)[50].

Παρόλο που τα νευρωνικά δίκτυα μοιάζουν με τον εγκέφαλο, η επεξεργαστική τους ισχύς δεν προσομοιάζει αυτή του ανθρώπινου εγκεφάλου. Ακόμη, τα νευρωνικά δίκτυα επωφελούνται από μεγάλα σύνολα δεδομένων για εκπαίδευση. Ως ένα από τα πιο ταχέως αναπτυσσόμενα υπο-πεδία στην υπολογιστική επιστήμη, η βαθιά μάθηση αξιοποιεί μεγάλα δίκτυα πολλαπλών επιπέδων με σκοπό τη μοντελοποίηση προτύπων ανώτατου επιπέδου στα δεδομένα (Εικόνα 11) (Taye, 2023)[50].

Εικόνα 11. Μοντελοποίηση προτύπου στη βαθιά μάθηση



Πηγή: https://www.mdpi.com/2073-431X/12/5/91

Από την άλλη πλευρά η μηχανική εκμάθηση μαθαίνει να χαρτογραφεί την είσοδο στην έξοδο, δεδομένης μιας συγκεκριμένης παγκόσμιας αναπαράστασης χαρακτηριστικών, που έχουν σχεδιαστεί χειροκίνητα για κάθε εργασία. Δομημένα αλλά και μη δομημένα δεδομένα χρησιμοποιούνται για τη συλλογή και την εξαγωγή των απαραίτητων πληροφοριών αναφορικά με την εκάστοτε εργασία (Taye, 2023)[50].

Σε ότι αφορά στα μοντέλα βαθιάς μάθησης έχουν ένα πλεονέκτημα συγκριτικά με τα παραδοσιακά μοντέλα μηχανικής μάθησης εξαιτίας του αυξημένου αριθμού επιπέδων μάθησης αλλά και του υψηλότερου επιπέδου αφαίρεσης. Επίσης, υπάρχει άμεση μάθηση βάσει δεδομένων για όλα τα στοιχεία

του μοντέλου. Εξαιτίας του αλγορίθμου στα παραδοσιακά μοντέλα μηχανικής εκμάθησης υπάρχουν περιορισμούς καθώς το μέγεθος των δεδομένων όπως επίσης και η ζήτηση για πληροφορίες από τα δεδομένα αυξάνονται. Η επέκταση των δεδομένων έχει οδηγήσει σε ανάπτυξη ταχύτερων και ακριβέστερων αλγορίθμων μάθησης. Προκειμένου να διατηρηθεί το ανταγωνιστικό πλεονέκτημα, κάθε οργανισμός θα χρησιμοποιήσει ένα μοντέλο που παράγει τις πιο ακριβείς προβλέψεις (Taye, 2023)[50].

Επιπροσθέτως, η βαθιά μάθηση δεν απαιτεί κανόνες σχεδιασμένους από ανθρώπους με σκοπό να λειτουργήσει. Αντ' αυτού αξιοποιεί μεγάλο όγκο δεδομένων ώστε να αντιστοιχίσει την παρεχόμενη είσοδο σε συγκεκριμένες ετικέτες. Ακόμη, δημιουργείται χρησιμοποιώντας πολλαπλά επίπεδα αλγορίθμων, εκ των οποίων καθένα δίνει μια μοναδική ερμηνεία των δεδομένων για τα δεδομένα που του παρέχονται (LeCun et al., 2015)[53].

Αναφορικά με τις συμβατικές τεχνικές μηχανικής μάθησης περιλαμβάνουν πολλά διαδοχικά βήματα με σκοπό την ολοκλήρωση της εργασίας ταξινόμησης, όπως είναι η προεπεξεργασία, η έξυπνη επιλογή και η ταξινόμηση. Σε αυτό το σημείο είναι σημαντικό να αναφερθεί, πως η επιλογή χαρακτηριστικών επιδρά σε μεγάλο βαθμό στην απόδοση των αλγορίθμων μηχανικής εκμάθησης, ενώ η μεροληπτική επιλογή χαρακτηριστικών ενδέχεται να οδηγήσει σε ανακρίβεια στη διάκριση κατηγορίας. Από την άλλη πλευρά, η βαθιά μάθηση μπορεί να αυτοματοποιήσει την εκμάθηση συνόλων χαρακτηριστικών για εύρος εργασιών, σε αντίθεση με τους τυπικούς αλγόριθμους μηχανικής μάθησης. Μέσω της βαθιάς μάθησης καθίσταται εφικτή η ταυτόχρονη μάθηση και ταξινόμηση (LeCun et al., 2015; Shrestha & Mahmood, 2019)[53][54].

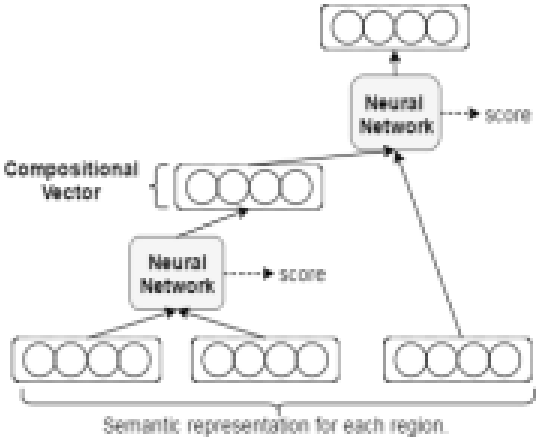
Εξαιτίας της πολύπλοκης δομής των πολλαπλών επιπέδων, ένα σύστημα βαθιάς μάθησης χρειάζεται μεγάλο όγκο δεδομένων με σκοπό τον περιορισμό του θορύβου και την ακρίβεια των ερμηνειών. Συνεπώς, η βαθιά εκμάθηση απαιτεί περισσότερα δεδομένα συγκριτικά με τους παραδοσιακούς αλγόριθμους μηχανικής εκμάθησης, καθώς στη μηχανική μάθηση μπορεί να χρησιμοποιηθούν 1000 σημεία δεδομένων, ενώ στη βαθιά εκμάθηση εκατομμύρια σημεία δεδομένων (Bengio, 2009; LeCun et al., 2015)[55][53].

3.3 Αλγοριθμικές τεχνικές βαθιάς μάθησης

- Αναδρομικό νευρωνικό δίκτυο (Recursive Neural Network - RvNN)

Σχετικά με το RvNN έχει τη δυνατότητα να πραγματοποιεί προβλέψεις στο πλαίσιο μιας ιεραρχικής δομής. Επίσης, μπορεί να ταξινομεί τις εξόδους κάνοντας χρήση διανυσμάτων σύνθεσης. Καθίσταται αξιοσημείωτο, πως η ανάπτυξη του RvNN εμπνεύστηκε από την Recursive Autoassociative Memory (RAAM), η οποία αποτελεί αρχιτεκτονική που αναπτύχθηκε με σκοπό για την επεξεργασία αντικειμένων δομημένων σε κάποιο σχήμα, όπως για παράδειγμα σε γραφήματα. Η εν λόγω προσέγγιση ήταν να ληφθεί μια αναδρομική δομή δεδομένων μεταβλητού μεγέθους καθώς επίσης και να αναπτυχθεί μια κατανεμημένη αναπαράσταση με σταθερό πλάτος. Ακόμη, το πρόγραμμα εκμάθησης Backpropagation Through Structure (BTS) εφαρμόστηκε με σκοπό την εκπαίδευση του δικτύου, ακολουθώντας μια προσέγγιση η οποία είναι παρόμοια με τον τυπικό αλγόριθμο backpropagation και είναι εφικτό να υποστηρίξει μια τέτοια δομή (Εικόνα 12). Ακόμη, το δίκτυο εκπαιδεύεται με αυτόματη συσχέτιση ώστε να προβεί σε αναπαραγωγή του μοτίβου του επιπέδου εισόδου στο επίπεδο εξόδου (Pouyanfar et al., 2018)[56].

Εικόνα 12. RvNN



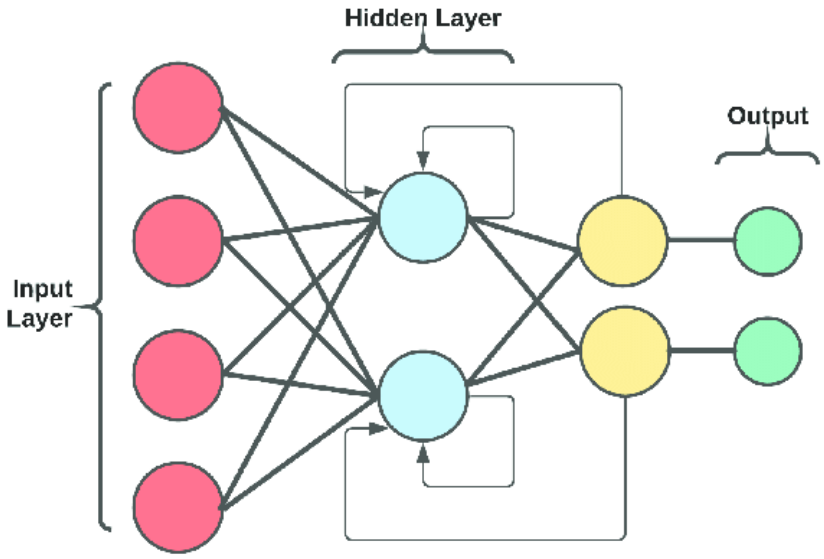
Πηγή: https://courses.cs.duke.edu/spring20/compsci527/papers/Pouyanfar.pdf

Το RvNN ήταν ιδιαίτερα επιτυχημένο στο Νευρογλωσσικό Προγραμματισμό (NLP). Προτάθηκε από τον Socher και τους συνεργάτες του (2011)[57] μια αρχιτεκτονική RvNN που είναι εφικτό να χειριστεί τις εισόδους διαφορετικών τρόπων. Διατίθενται δύο παραδείγματα χρήσης του RvNN για την ταξινόμηση φυσικών εικόνων αλλά και προτάσεων φυσικής γλώσσας. Παρόλο που μια εικόνα μπορεί να διαιρεθεί σε διάφορες περιοχές ενδιαφέροντος, μια πρόταση διασπάται σε λέξεις, και το RvNN αξιολογεί τη βαθμολογία ενός πιθανού ζεύγους με στόχο να τα συγχωνεύσει και να κατασκευάσει ένα συντακτικό δέντρο. Για κάθε πιθανό ζεύγος, το RvNN υπολογίζει μια βαθμολογία που εκτιμά την καταλληλότητα της συγχώνευσης. Το ζεύγος που συγκεντρώνει την υψηλότερη βαθμολογία συνδυάζεται τελικά σε ένα διάνυσμα σύνθεσης. Στη συνέχεια, μετά από κάθε συγχώνευση, το RvNN δημιουργεί (1) μια μεγαλύτερη περιοχή που αποτελείται από πολλαπλές μονάδες, (2) ένα διάνυσμα σύνθεσης το οποίο περιγράφει τη νέα περιοχή, καθώς και (3) την κατάλληλη ετικέτα κλάσης. Στη ρίζα του δέντρου RvNN βρίσκεται η αναπαράσταση του διανύσματος σύνθεσης της συνολικής περιοχής (Pouyanfar et al., 2018)[56].

- Επαναλαμβανόμενο Νευρωνικό Δίκτυο (RNN)

Μεταξύ των πιο γνωστών και διαδεδομένων αλγορίθμων της βαθιάς μάθησης, ιδιαίτερα σε τομείς όπως η Επεξεργασία Φυσικής Γλώσσας (NLP) και η ανάλυση ομιλίας, συγκαταλέγεται το RNN. Σε αντίθεση με τα κλασικά νευρωνικά δίκτυα, τα RNN εκμεταλλεύονται τη σειριακή πληροφορία μέσα στο δίκτυο. Η εν λόγω ιδιότητα είναι κρίσιμη για πολλές εφαρμογές όπου η διαδοχική δομή των δεδομένων μεταφέρει ουσιώδη πληροφορία, όπως για παράδειγμα, η κατανόηση μιας λέξης σε μια πρόταση απαιτεί γνώση του συμφραζομένου. Συνεπώς, το RNN μπορεί να παρομοιαστεί με ένα σύνολο μονάδων που διαθέτουν βραχυπρόθεσμη μνήμη, αποτελούμενο από το επίπεδο εισόδου x, το κρυφό επίπεδο (κατάσταση) s και το επίπεδο εξόδου y. Μπορεί να υπάρχουν τρεις διαφορετικές προσεγγίσεις βαθιών RNN, οι οποίες περιλαμβάνουν τις στρατηγικές "Είσοδος σε Κρυφό", "Κρυφό σε Έξοδο" και "Κρυφό σε Κρυφό". Αξιοποιώντας αυτές τις λύσεις, προτείνεται ένα βαθύ RNN που επωφελείται από τη χρήση βαθύτερων δομών αλλά και διευκολύνει τη μάθηση στα βαθιά δίκτυα (Pouyanfar et al., 2018)[56].

Εικόνα 13. RNN



Πηγή: https://www.researchgate.net/figure/Recurrent-neural-network-architecture-diagram_fig2_361681838

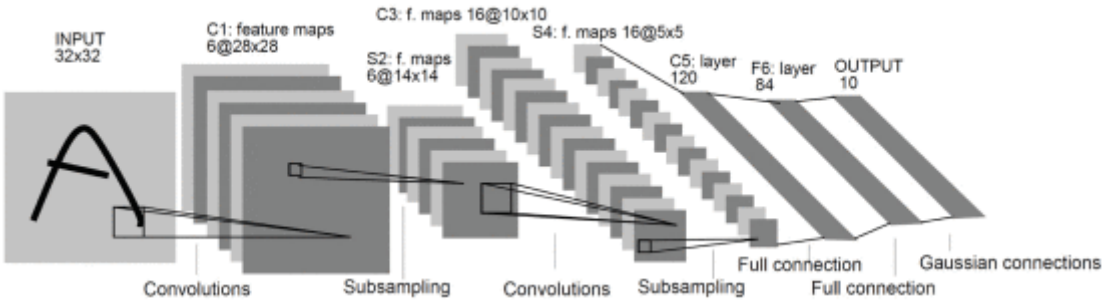
Ένα από τα κύρια προβλήματα που αντιμετωπίζει ένα RNN είναι η ευαισθησία του στις κλίσεις που είτε εξαφανίζονται είτε εκρήγνυνται. Δηλαδή, οι κλίσεις ενδέχεται να μειώνονται ή να αυξάνονται εκθετικά λόγω του πολλαπλασιασμού πολλών πολύ μικρών ή πολύ μεγάλων παραγώγων κατά τη διαδικασία εκπαίδευσης. Με την πάροδο του χρόνου, αυτή η ευαισθησία οδηγεί στη μείωση της απομνημόνευσης των αρχικών εισόδων του δικτύου, καθώς εισάγονται νέες πληροφορίες. Για την αντιμετώπιση αυτού του ζητήματος, χρησιμοποιείται η Μακροπρόθεσμη Μνήμη (LSTM), η οποία προσφέρει μπλοκ μνήμης στις επαναλαμβανόμενες συνδέσεις. Τα συγκεκριμένα μπλοκ μνήμης περιέχουν κελιά μνήμης που διατηρούν χρονικές καταστάσεις του δικτύου. Επιπλέον, ενσωματώνουν περιφραγμένες μονάδες που ρυθμίζουν τη ροή των πληροφοριών. Οι υπολειπόμενες συνδέσεις σε εξαιρετικά βαθιά δίκτυα μπορούν επίσης να μειώσουν σημαντικά το πρόβλημα της εξαφάνισης των κλίσεων (Pouyanfar et al., 2018)[56].

- Νευρωνικό Δίκτυο Συνέλιξης (Convolution Neural Network - CNN)

Σχετικά με το CNN βασίζεται στον ανθρώπινο οπτικό φλοιό και συνιστά το νευρωνικό δίκτυο επιλογής σε ότι αφορά στην αναγνώριση εικόνας και βίντεο. Ακόμη, αξιοποιείται σε διάφορους τομείς όπως για παράδειγμα στην ανακάλυψη φαρμάκων. Όπως απεικονίζεται στην Εικόνα 14. το CNN περιλαμβάνει μια σειρά επιπέδων συνέλιξης και δευτερεύουσας δειγματοληψίας, κι έπεται ένα πλήρως συνδεδεμένο στρώμα και ένα κανονικοποιητικό στρώμα, όπως για παράδειγμα συνάρτηση softmax. Ειδικότερα, απεικονίζεται η αρχιτεκτονική LeNet-5 CNN 7 επιπέδων για αναγνώριση ψηφίων. Ακόμη, η σειρά πολλαπλών επιπέδων συνέλιξης πραγματοποιεί προοδευτικά πιο εκλεπτυσμένη εξαγωγή χαρακτηριστικών από τα επίπεδα εισόδου στα επίπεδα εξόδου. Τα πλήρως συνδεδεμένα επίπεδα που εκτελούν ταξινόμηση ακολουθούν τα επίπεδα συνέλιξης. Σχετικά με τα στρώματα υποδειγματοληψίας ή συγκέντρωσης, πολλές φορές παρεμβάλλονται ανάμεσα στα στρώματα συνέλιξης. Το CNN παίρνει ένα 2Dn × n εικόνα με pixel ως είσοδο, ενώ τα στρώματα αποτελούνται από ομάδες δισδιάστατων νευρώνων, οι οποίες καλούνται ως φίλτρα είτε ως πυρήνες. Οι νευρώνες στα στρώματα εξαγωγής

χαρακτηριστικών του CNN δεν συνδέονται με όλους τους νευρώνες στα κοντίνα στρώματα (Shrestha & Mahmood, 2019)[52].

Εικόνα 14. Νευρωνικό δίκτυο συνέλιξης



Πηγή: https://ieeexplore.ieee.org/abstract/document/8694781/references#references

Ωστόσο, συνδέονται μόνο με τους χωρικά αντιστοιχισμένους νευρώνες που έχουν σταθερό μέγεθος και είναι μερικώς επικαλυπτόμενοι στην εικόνα εισόδου είτε στον χάρτη χαρακτηριστικών του προηγούμενου στρώματος.

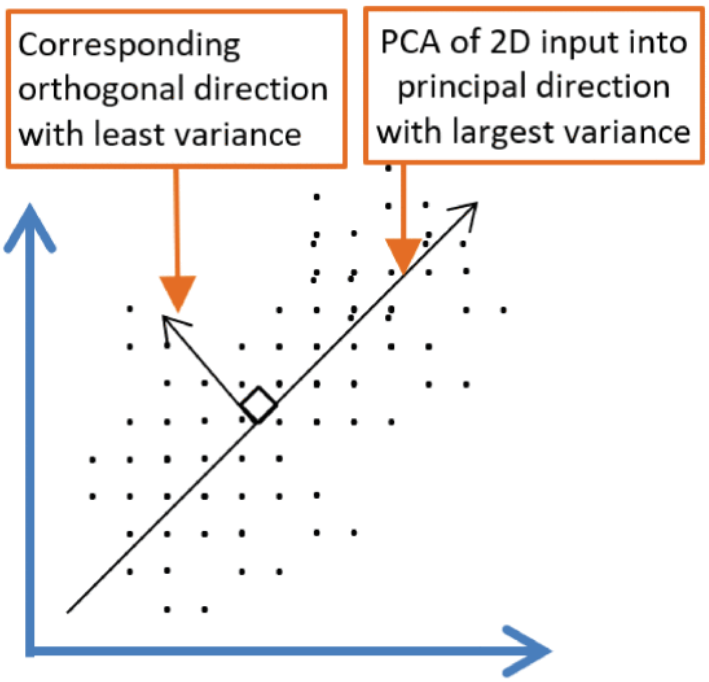
Η εν λόγω περιοχή στην είσοδο αναφέρεται ως τοπικό δεκτικό πεδίο. Ο ελαχιστοποιημένος αριθμός συνδέσεων συμβάλλει στη μείωση του χρόνου εκπαίδευσης και περιορίζει τον κίνδυνο υπερπροσαρμογής. Κάθε νευρώνας σε ένα φίλτρο συνδέεται με συγκεκριμένο αριθμό νευρώνων στο προηγούμενο επίπεδο εισόδου είτε στο χάρτη χαρακτηριστικών και αναφέρεται στην ίδια σειρά βαρών καθώς επίσης και προκαταλήψεων. Αυτά τα χαρακτηριστικά επιταχύνουν τη διαδικασία εκμάθησης, ενώ παράλληλα μειώνουν τις απαιτήσεις σχετικά με τη μνήμη του δικτύου. Κατ' αυτόν τον τρόπο οι νευρώνες ενός φίλτρου εντοπίζουν το ίδιο μοτίβο σε διαφορετικά σημεία της εισερχόμενης εικόνας. Σε αυτό το σημείο αξίζει να αναφερθεί, πως τα επίπεδα υποδειγματοληψίας ελαχιστοποιούν το μέγεθος του δικτύου, ενώ, σε συνδυασμό με τα τοπικά δεκτικά πεδία και τη χρήση κοινών βαρών εντός του ίδιου φίλτρου, ελαχιστοποιούν την ευαισθησία του δικτύου σε μετατοπίσεις, αλλαγές κλίμακας και παραμορφώσεις των εικόνων. Για την υποδομή συχνά αξιοποιούνται φίλτρα μέγιστης ή μέσης συγκέντρωσης, καθώς και τοπικού μέσου όρου. Στα τελικά επίπεδα του CNN, που ασχολούνται με την ταξινόμηση, οι νευρώνες μεταξύ των στρωμάτων καθίστανται απόλυτα συνδεδεμένοι (Shrestha & Mahmood, 2019)[52].

Καθίσταται αξιοσημείωτο, πως τα CNN μπορούν να χρησιμοποιηθούν με πολλαπλές ακολουθίες επιπέδων συνέλιξης που κάνουν χρήση τόσο κοινών βαρών όσο και επιπέδων υποδειγματοληψίας. Η πολυεπίπεδη φύση του CNN εξασφαλίζει αναπαραστάσεις υψηλής ποιότητας, ενώ παράλληλα διατηρεί την εντοπιότητα, μειώνει τον αριθμό των παραμέτρων και παραμένει ανθεκτική σε μικρές μεταβολές της εικόνας εισόδου.

- Αυτόματος κωδικοποιητής

Σχετικά με τον Αυτόματο κωδικοποιητή αποτελεί ένα νευρωνικό δίκτυο, το οποίο χρησιμοποιεί αλγόριθμο χωρίς να υπάρχει επίβλεψη και μαθαίνει την αναπαράσταση στο σύνολο δεδομένων εισόδου με σκοπό να μειωθούν οι διαστάσεις και να αναδημιουργηθούν τα αρχικά δεδομένα. Σε αυτήν την περίπτωση οι αυτοκωδικοποιητές επεκτείνουν την ιδέα της ανάλυσης κύριου συστατικού (PCA). Όπως απεικονίζεται στην Εικόνα 15, ένα PCA μετατρέπει πολυδιάστατα δεδομένα σε γραμμική αναπαράσταση. Ακόμη, διαπιστώνεται πώς μια δισδιάστατη είσοδος δεδομένων είναι δυνατόν να αναχθεί σε γραμμικό διάνυσμα χρησιμοποιώντας PCA. Επιπροσθέτως, οι αυτοκωδικοποιητές είναι δυνατόν να προχωρήσουν περισσότερο, παράγοντας μη γραμμική αναπαράσταση. Ο ρόλος του PCA είναι καθοριστικός σε ότι αφορά στο σύνολο γραμμικών μεταβλητών στις κατευθύνσεις με τη μεγαλύτερη διακύμανση. Τα σημεία δεδομένων εισόδου διαστάσεων αντιπροσωπεύονται ως m ορθογώνιες κατευθύνσεις, έτσι ώστε $m \leq p$ και αποτελεί χαμηλότερο (μικρότερο από m) διαστασιακό χώρο. Αναφορικά με τα αρχικά σημεία δεδομένων προβάλλονται στις κύριες κατευθύνσεις παραλείποντας με αυτόν τον τρόπο πληροφορίες στις αντίστοιχες ορθογώνιες κατευθύνσεις. Ακόμη, είναι σημαντικό να αναφερθεί πως το PCA εστιάζει κυρίως στις διακυμάνσεις και όχι στις συνδιακυμάνσεις και τις συσχετίσεις, ενώ αναζητά τη γραμμική συνάρτηση με τη μεγαλύτερη διακύμανση. Εδώ ο στόχος είναι να προσδιοριστεί η κατεύθυνση με το ελάχιστο μέσο τετραγωνικό σφάλμα, παρουσιάζοντας το μικρότερο σφάλμα ανακατασκευής.

Εικόνα 15. Αυτόματος κωδικοποιητής

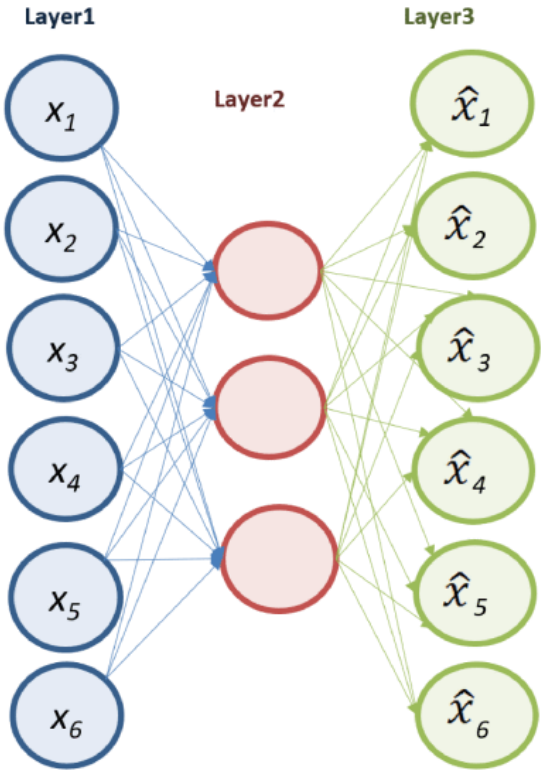


Πηγή: https://ieeexplore.ieee.org/abstract/document/8694781/references#references

Επιπροσθέτως, οι αυτόματοι κωδικοποιητές αξιοποιούν μπλοκ κωδικοποιητών αλλά και αποκωδικοποιητών μη γραμμικών κρυφών επιπέδων προκειμένου να γενικεύσουν το PCA για να μειώσουν τις διαστάσεις και την τελική ανακατασκευή των αρχικών δεδομένων. Κατά την εκτέλεση μείωσης διαστάσεων, οι αυτόματοι κωδικοποιητές παρουσιάζουν ενδιαφέρουσες αναπαραστάσεις του

διανύσματος εισόδου στο κρυφό στρώμα. Αυτό μπορεί να αποδοθεί στον μικρότερο αριθμό κόμβων είτε στο κρυφό στρώμα είτε σε κάθε δεύτερο επίπεδο των μπλοκ δύο επιπέδων. Ωστόσο, ακόμη και στην περίπτωση που υπάρχει μεγαλύτερος αριθμός κόμβων στο κρυφό επίπεδο, ένας περιορισμός πυκνότητας μπορεί να επιβληθεί στις κρυφές μονάδες ώστε να διατηρηθούν ενδιαφέρουσες αναπαραστάσεις χαμηλότερης διάστασης των εισόδων. Για την επίτευξη αραιότητας η έξοδος ορίζεται σε τιμή κοντά στο μηδέν. Στην Εικόνα 16 φαίνεται το μπλοκ ανιχνευτών μονής στρώσης χαρακτηριστικών RBM που χρησιμοποιούνται στην προεκπαίδευση. Αναλυτικότερα, στη συγκεκριμένη εικόνα απεικονίζεται μια απλοποιημένη αναπαράσταση του τρόπου με τον οποίο οι αυτόματοι κωδικοποιητές μπορούν να μειώσουν τη διάσταση των δεδομένων εισόδου και να μάθουν να τα αναδημιουργούν στο επίπεδο εξόδου (Shrestha & Mahmood, 2019)[52].

Εικόνα 16. Μπλοκ ανιχνευτών μονής στρώσης χαρακτηριστικών RBM



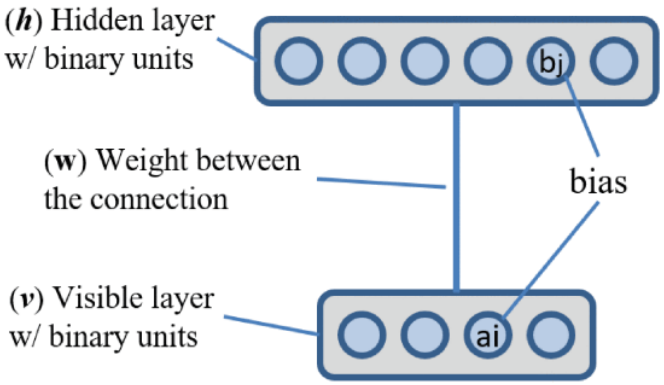
Πηγή: https://ieeexplore.ieee.org/abstract/document/8694781/references#references

- Περιορισμένη μηχανή Boltzmann (Restricted Boltzmann Machine - RBM)

Η RBM αποτελεί ένα τεχνητό νευρωνικό δίκτυο στο οποίο είναι εφικτό να εφαρμοστεί αλγόριθμος μάθησης χωρίς επίβλεψη με σκοπό να δημιουργηθούν μη γραμμικά παραγωγικά μοντέλα από δεδομένα χωρίς ετικέτα. Στην προκειμένη περίπτωση ο στόχος είναι να εκπαιδευτεί το δίκτυο και να αυξήσει μια συνάρτηση, όπως για παράδειγμα ένα προϊόν, της πιθανότητας του διανύσματος στις ορατές μονάδες, για να καταστεί εφικτή η πιθανολογική ανακατασκευή στην είσοδο. Όπως απεικονίζεται στη Εικόνα 17, η RBM αποτελείται από δίκτυο δύο επιπέδων, περιλαμβάνοντας το ορατό στρώμα και το κρυφό στρώμα. Αναλυτικότερα, η κάθε μονάδα στο ορατό επίπεδο συνδέεται με όλες τις μονάδες στο κρυφό

στρώμα, ενώ δεν υφίστανται συνδέσεις μεταξύ των μονάδων του ίδιου επιπέδου Shrestha & Mahmood, 2019)[52].

Εικόνα 17. Περιορισμένη μηχανή Boltzmann

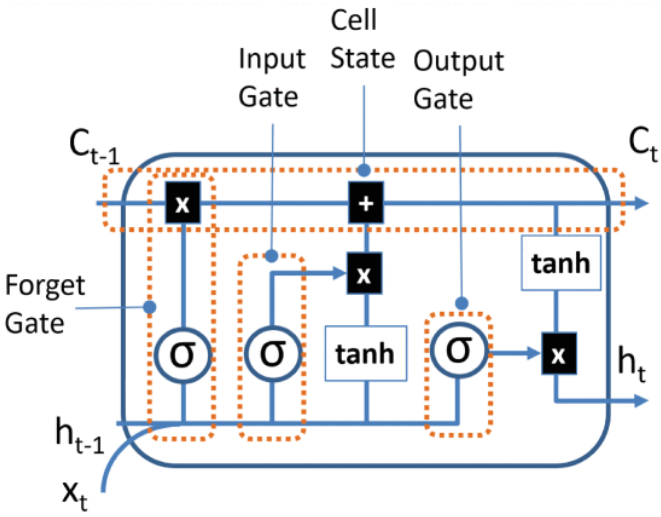


Πηγή: https://ieeexplore.ieee.org/abstract/document/8694781/references#references

- Μακροπρόθεσμη Μνήμη (LSTM)

Σχετικά με την LSTM αναφέρεται στην υλοποίηση του Recurrent Neural Network και προτάθηκε κατά το έτος 1997 από τον Hochreiter και τους συνεργάτες του. Σημαντικό χαρακτηριστικό του LSTM είναι πως μπορεί να διατηρήσει τη γνώση των προηγούμενων καταστάσεων και να εκπαιδευτεί για εργασία που απαιτεί μνήμη είτε επίγνωση κατάστασης. Επιπλέον, η LSTM αντιμετωπίζει εν μέρει έναν περιορισμό του RNN και ειδικότερα το πρόβλημα της εξαφάνισης των κλίσεων, επιτρέποντας στις κλίσεις να περνούν αναλλοίωτες. Από την Εικόνα 18 φαίνεται, πως το LSTM αποτελείται από μπλοκ κατάστασης κυψέλης μνήμης μέσω των οποίων ρέει το σήμα και ρυθμίζεται από πύλες εισόδου, λήθης αλλά και από πύλες εξόδου. Οι εν λόγω πύλες ελέγχουν οτιδήποτε αποθηκεύεται, διαβάζεται και γράφεται στο κελί. Τέλος η LSTM χρησιμοποιείται μεταξύ άλλων από την Google και την Apple στις πλατφόρμες αναγνώρισης φωνής που διαθέτουν (Shrestha & Mahmood, 2019)[52].

Εικόνα 18. Μακροπρόθεσμη μνήμη

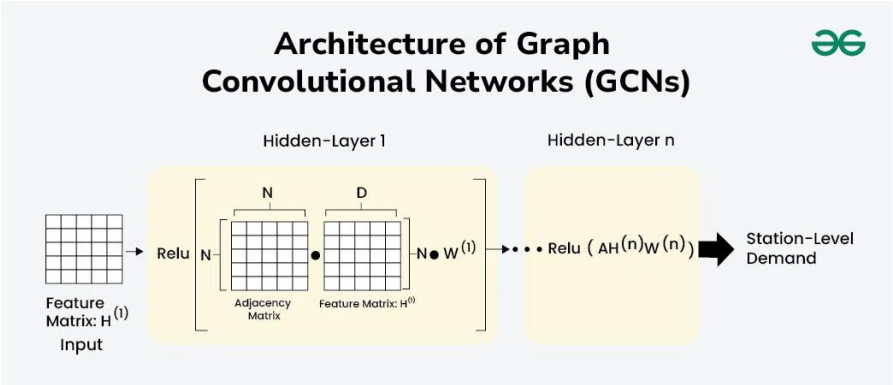


Πηγή: https://ieeexplore.ieee.org/abstract/document/8694781/references#references

- Συνελικτικό δίκτυο γραφημάτων (GCNs)

Τα συνελικτικά δίκτυα γραφημάτων (Graph Convolutional Networks - GCNs) αποτελούν τύπο νευρωνικού δικτύου που έχει σχεδιαστεί με σκοπό να λειτουργεί άμεσα με γραφήματα. Ειδικότερα, ένα γράφημα περιλαμβάνει κόμβους και ακμές. Σε ένα τέτοιο δίκτυο, καθένας από τους κόμβους αντιπροσωπεύει μια οντότητα, ενώ οι ακμές αντιπροσωπεύουν τις σχέσεις μεταξύ των εν λόγω οντοτήτων. Βασικός στόχος των GCN είναι η εκμάθηση ενσωματώσεων κόμβων, που αποτελούν διανυσματικές αναπαραστάσεις κόμβων, καταγράφοντας δομικές πληροφορίες όπως επίσης και πληροφορίες χαρακτηριστικών του γραφήματος. Συχνά, τα GCN αποτελούνται από πολλαπλά επίπεδα, που είναι υπεύθυνα για τη βελτίωση των ενσωματώσεων κόμβων μέσω της συλλογής πληροφοριών από τους γείτονες κόμβους σε αυξανόμενες αποστάσεις (Sejan et al., 2023; Gao et al., 2024)[58][59].

Εικόνα 19. Συνελικτικό δίκτυο γραφημάτων



Πηγή: https://www.geeksforgeeks.org/graph-convolutional-networks-gcns-architectural-insights-and-applications/

Κεφάλαιο 4ο: Ερευνητικά δεδομένα

4.1 Έρευνες για μηχανική μάθηση

Οι Bingöl και Alatas (2019)[60] εστιάζοντας στη σημαντικότητα της αποτελεσματικής ανίχνευσης φημών στα μέσα κοινωνικής δικτύωσης, προέβησαν σε μελέτη με τη χρήση εποπτευόμενων μεθόδων μηχανικής μάθησης σε πραγματικά δεδομένα. Για την ανίχνευση των φημών αξιοποίησαν τους αλγόριθμους OneR (One Rule), Naive Bayes, ZeroR, JRip, Random Forest, Sequential Minimal Optimization και Hoeffding Tree. Σύμφωνα με τα αποτελέσματα, το υψηλότερο ποσοστό ακρίβειας λήφθηκε από τον αλγόριθμο ταξινόμησης Random Forest και τον Naive Bayes.

Ο Saha και οι συνεργάτες (2022)[61] με σκοπό τον εντοπισμό ψευδών ειδήσεων, πρότειναν μια μέθοδο μηχανικής μάθησης που χρησιμοποιεί τον ταξινομητή Naïve Bayes. Ειδικότερα, το σύστημα μπορεί να συνδεθεί με οποιοδήποτε μέσο κοινωνικής δικτύωσης, καθώς προβλέπει την πιθανότητα οι ειδήσεις να είναι ψεύτικες ή η φήμη να μην εξαπλωθεί περαιτέρω με τον έλεγχο λογισμικού στα μέσα κοινωνικής δικτύωσης. Τα αποτελέσματα έδειξαν επιτυχία υψηλής ακρίβειας.

Ο Wang και οι συνεργάτες του (2019)[62] στο πλαίσιο μελέτης τους, συνέλεξαν πληροφορίες χρήστη από το πιο πρόσφατο δημοσιευμένο περιεχόμενο microblog 3793 χρηστών του Sina Weibo και χρησιμοποιήθηκε η επεξεργασία φυσικής γλώσσας (NLP) και τεχνικές μηχανικής μάθησης. Ειδικότερα, χρησιμοποιήθηκε η λογιστική παλινδρόμηση (LR), οι μηχανές υποστήριξης διανυσμάτων (SVM), ο αλγόριθμος τυχαίου δάσους (RF) καθώς επίσης και η ενίσχυση ακραίας κλίσης (XGBoost) για να προβλεφθεί η διάδοση φήμης μέσω retweet. Σύμφωνα με τα αποτελέσματα και σε σύγκριση με τις παραδοσιακές μεταβλητές πρόβλεψης που έχουν πρόσβαση μόνο σε πληροφορίες χρήστη, οι αναλύσεις ομοιότητας και συναισθήματος των πιο πρόσφατων περιεχομένων διαπιστώθηκε, πως βελτιώνουν σημαντικά την ακρίβεια της πρόβλεψης.

Σε σχετική μελέτη, οι Al-Alshaqi και συνεργάτες (2024) πρότειναν μια μεθοδολογία διπλής φάσης για την ανίχνευση ψευδών ειδήσεων, αξιοποιώντας αρχικά διάφορους ταξινομητές για την επεξεργασία και ανάλυση δεδομένων κειμένου. Η πρώτη φάση περιλάμβανε τη χρήση τεχνικών μηχανικής μάθησης για την κατηγοριοποίηση των ειδήσεων, ενώ στη δεύτερη φάση αναπτύχθηκε μια πολυτροπική προσέγγιση, η οποία συνδύαζε μοντέλο BERT για την κατανόηση του κειμένου με συνελικτικά νευρωνικά δίκτυα (CNN) για την ανάλυση των οπτικών δεδομένων. Η συνδυαστική χρήση γλωσσικών και οπτικών χαρακτηριστικών οδήγησε σε σημαντικά υψηλά ποσοστά ακρίβειας, καθιστώντας το προτεινόμενο σύστημα ιδιαίτερα αποτελεσματικό στην καταπολέμηση της παραπληροφόρησης. Η μελέτη αποδεικνύει τη σημασία των πολυτροπικών μοντέλων στην αναγνώριση ψεύτικων ειδήσεων, καθώς μπορούν να επεξεργαστούν και να ερμηνεύσουν πληροφορίες από διαφορετικές μορφές μέσων, ενισχύοντας τη συνολική ακρίβεια και αξιοπιστία των αποτελεσμάτων. (Al-Alshaqi et al., 2024)[63].

Σε μια πρόσφατη μελέτη, οι Arowolo και συνεργάτες (2023) πρότειναν μια τεχνική μηχανικής μάθησης με στόχο την ανίχνευση ψευδών ειδήσεων μέσω φιλτραρίσματος και ταξινόμησης περιεχομένου από τα μέσα κοινωνικής δικτύωσης. Η προσέγγιση βασίστηκε στην προεπεξεργασία των δεδομένων και στη χρήση κλασικών αλγορίθμων μηχανικής μάθησης για την κατηγοριοποίησή τους, δημιουργώντας έτσι ένα σύστημα το οποίο μπορεί να διακρίνει την αληθή από την παραπλανητική πληροφορία. Η μελέτη αξιολόγησε διάφορους ταξινομητές και κατέληξε στο συμπέρασμα ότι οι αλγόριθμοι SVM (Support Vector Machine) και KNN (K-Nearest Neighbors) παρουσιάζουν ικανοποιητική απόδοση στον εντοπισμό ψευδών ειδήσεων στα κοινωνικά δίκτυα. Οι ερευνητές υποστηρίζουν ότι, παρά την απλότητα αυτών των μεθόδων σε σύγκριση με πιο σύγχρονες προσεγγίσεις βαθιάς μάθησης, μπορούν να

λειτουργήσουν αποτελεσματικά όταν εφαρμοστούν σε καλά φιλτραρισμένα και ποιοτικά επεξεργασμένα δεδομένα (Arowolo et al., 2023)[64].

Ο Chang και οι συνεργάτες (2024)[65] πρότειναν σε μελέτη τους μια προσέγγιση, η οποία καλείται ως μετασχηματιστής επαυξημένης μνήμης με συνελικτικά δίκτυα γραφημάτων (GCNs-MT), με σκοπό ανίχνευση φημών σε πλατφόρμες κοινωνικής δικτύωσης. Ειδικότερα, το μοντέλο τους ενσωματώνει κελιά βραχυπρόθεσμης μνήμης αλλά και μηχανισμό προσοχής πολλαπλών κεφαλών για να καταγράψει τις τοπικές εξαρτήσεις αλλά και τις παγκόσμιες εξαρτήσεις στη διάδοση φημών. Με την ενσωμάτωση των GCN επιχείρησαν να αξιοποιήσουν τις δομικές πληροφορίες της διάδοσης φημών με σκοπό τη βελτιωμένη απόδοση ανίχνευσης. Ακόμη, ανέπτυξαν ένα σύνολο δεδομένων που κωδικοποιείται και ενσωματώνεται από προεκπαιδευμένες ενσωματώσεις λέξεων με βάση τα tweets και ακολούθησαν εκτενείς αξιολογήσεις. Σύμφωνα με τα αποτελέσματα της μελέτης, το προτεινόμενο πλαίσιο GCNs-MT συνιστά αποτελεσματική λύση σχετικά με την ανίχνευση φημών στα μέσα κοινωνικής δικτύωσης.

4.2 Έρευνες για βαθιά μάθηση

Ο Zhang και οι συνεργάτες (2018)[66] προτείνουν μια προσέγγιση η οποία καλείται ως DRI-RCNN (Deceptive Review Identification by Recurrent Convolutional Neural Network) με σκοπό τον εντοπισμό παραπλανητικών κριτικών αξιοποιώντας περιβάλλοντα λέξεων καθώς επίσης και βαθιά μάθηση. Το μοντέλο αυτό συνδυάζει επαναλαμβανόμενα και συνελικτικά νευρωνικά δίκτυα για να διακρίνει μεταξύ παραπλανητικών και αληθινών κριτικών, βασιζόμενο στην υπόθεση ότι οι συγγραφείς των παραπλανητικών κριτικών δεν έχουν πραγματική εμπειρία με το αντικείμενο της κριτικής, σε αντίθεση με τους συγγραφείς των αληθινών κριτικών. Σε αυτό το πλαίσιο, προκειμένου να διαφοροποιήσουν την παραπλανητική και αληθινή γνώση των συμφραζομένων που περιλαμβάνονται στις διαδικτυακές κριτικές, περιλαμβάνονται έξι στοιχεία ως επαναλαμβανόμενο συνελικτικό διάνυσμα. Ειδικότερα, το πρώτο και το δεύτερο στοιχείο καθιστούν δύο διανύσματα αριθμητικών λέξεων που προέρχονται από την εκπαίδευση τόσο παραπλανητικών όσο και αληθινών κριτικών, αντίστοιχα. Σχετικά με το τρίτο και το τέταρτο στοιχείο είναι γειτονικά παραπλανητικά και αληθινά διανύσματα περιβάλλοντος τα οποία προέρχονται από την εκπαίδευση ενός επαναλαμβανόμενου συνελικτικού νευρωνικού δικτύου σε διανύσματα περιβάλλοντος και διανύσματα αριστερών λέξεων. Τέλος, σχετικά με το πέμπτο και το έκτο στοιχείο, είναι γειτονικά παραπλανητικά και αληθή διανύσματα πλαισίου των σωστών λέξεων. Σύμφωνα με τα αποτελέσματα πειράματος, η προσέγγιση DRI-RCNN είναι μια χρήσιμη τεχνική για τον εντοπισμό παραπλανητικών φημών. Στη συνέχεια, το μοντέλο εφαρμόζει φίλτρα ReLU και μέγιστη συγκέντρωση (max-pooling) για να εξάγει τα πιο σημαντικά χαρακτηριστικά από τα επαναλαμβανόμενα συνελικτικά διανύσματα των λέξεων, δημιουργώντας έτσι ένα συνολικό διάνυσμα αναπαράστασης της κριτικής. Τα πειραματικά αποτελέσματα έδειξαν ότι το DRI-RCNN υπερέχει σε ακρίβεια ανίχνευσης παραπλανητικών κριτικών σε σύγκριση με προηγούμενες μεθόδους, όπως τα μοντέλα RCNN και GRNN-CNN, επιτυγχάνοντας ακρίβεια 85,24% και 87,24% στα σύνολα δεδομένων Deception και OpSpam, αντίστοιχα. Αυτά τα αποτελέσματα υποδηλώνουν ότι η ενσωμάτωση των συμφραζομένων και η διαφοροποίηση μεταξύ παραπλανητικών και αληθινών κριτικών βελτιώνουν σημαντικά την απόδοση του μοντέλου στην ανίχνευση παραπλανητικών κριτικών.

Στο πλαίσιο ενός πειράματος, ερευνητές παρουσίασαν ένα νέο μοντέλο βαθιάς μάθησης με αναδρομικά νευρωνικά δίκτυα (RNN) για την ανίχνευση φημών στην πλατφόρμα κοινωνικών μέσων Sina Weibo. Το μοντέλο αυτό, γνωστό ως Deep Recurrent Neural Network (DRNN), διαθέτει συμμετρική αρχιτεκτονική και αποτελείται από οκτώ επίπεδα. Περιλαμβάνει ένα επίπεδο εισόδου που δέχεται τη

ροή των αναρτήσεων, ακολουθούμενο από ένα επίπεδο κανονικοποίησης και δύο πλήρως συνδεδεμένα επίπεδα για την καλύτερη αναπαράσταση των χαρακτηριστικών. Στη συνέχεια, δύο επίπεδα RNN χρησιμοποιούνται για την καταγραφή των δυναμικών χρονικών σημάτων που είναι χαρακτηριστικά στη ροή των αναρτήσεων. Τέλος, ένα πλήρως συνδεδεμένο επίπεδο εξόδου παρέχει την πιθανότητα μιας ανάρτησης να είναι φήμη. Τα αποτελέσματα των πειραμάτων έδειξαν ότι το σχήμα διαδοχικής κωδικοποίησης είναι κατάλληλο για την αναπαράσταση κειμένου στο μοντέλο τους, παρουσιάζοντας καλύτερες επιδόσεις από άλλες υπάρχουσες μεθόδους ανίχνευσης, όπως το TF-IDF και το doc2vec. Επιπλέον, το μοντέλο επιτυγχάνει υψηλότερη ακρίβεια όταν εκπαιδεύεται με αναρτήσεις από χρήστες με περισσότερους ακολούθους, υποδηλώνοντας ότι οι χρήστες με μεγαλύτερη επιρροή τείνουν να δημοσιεύουν πιο αξιόπιστες πληροφορίες. Συνολικά, το προτεινόμενο μοντέλο DRNN υπερέχει σε ακρίβεια ανίχνευσης, συμπεριλαμβανομένης της έγκαιρης ανίχνευσης, σε σύγκριση με προηγούμενες μεθόδους. (Xu et al., 2019)[67].

Ο Alkhodair και οι συνεργάτες (2020)[68] διερεύνησαν το πρόβλημα της ανίχνευσης έκτακτων ειδήσεων που κυκλοφορούν στα μέσα κοινωνικής δικτύωσης, εστιάζοντας στις δυσκολίες που προκύπτουν από τη δυναμική και την αβεβαιότητα του περιεχομένου. Στην προσπάθειά τους να αντιμετωπίσουν αυτή την πρόκληση, ανέπτυξαν ένα μοντέλο που βασίζεται σε Επαναλαμβανόμενα Νευρωνικά Δίκτυα (RNN) και έχει τη δυνατότητα να μαθαίνει ενσωματώσεις λέξεων (word embeddings) ταυτόχρονα με την εκπαίδευση για την αναγνώριση φημών. Η προσέγγιση αυτή συνδυάζει δύο στόχους μάθησης σε ένα ενιαίο μοντέλο, ενισχύοντας έτσι την απόδοσή του στην ταξινόμηση του περιεχομένου ως φήμη ή όχι. Παρότι η μέθοδος χαρακτηρίζεται ως απλή, αποδείχθηκε ιδιαίτερα αποτελεσματική στη μείωση της διάδοσης παραπλανητικών πληροφοριών. Οι ερευνητές επισημαίνουν ότι μια φήμη, κατά τη στιγμή που εντοπίζεται από το σύστημα, δεν είναι απαραίτητα ψευδής, καθώς μπορεί να μεταβληθεί στη συνέχεια με βάση νέα δεδομένα και να χαρακτηριστεί είτε ως αληθής είτε ως ψευδής. Αυτή η προσέγγιση αντικατοπτρίζει την πραγματική ροή των πληροφοριών σε κρίσιμες καταστάσεις και καθιστά το σύστημα πιο ευέλικτο στην κατανόηση και ερμηνεία του περιεχομένου.

Στο πλαίσιο μιας άλλης μελέτης, δημιουργήθηκε ένα ημι-εποπτευόμενο μοντέλο μάθησης με σκοπό τον εντοπισμό ψεύτικων ειδήσεων στα μέσα κοινωνικής δικτύωσης σε πρώιμο στάδιο. Ειδικότερα, κατασκευάστηκε αρχικά ένα μοντέλο με σκοπό να εξάγουν οι μελετητές τη γνώμη των χρηστών που εκφράζεται σε σχόλια. Έπειτα, χρησιμοποιήθηκε ο αλγόριθμος CredRank με σκοπό να γίνει η αξιολόγηση της αξιοπιστίας των χρηστών και αναπτύχθηκε ένα μικρό δίκτυο χρηστών που εμπλέκονται στη διάδοση μιας είδησης. Οι έξοδοι των βημάτων αυτών χρησιμεύουν ως είσοδοι του ταξινομητή ειδήσεων SSLNews. Ειδικότερα, το SSLNews περιλαμβάνει τρία δίκτυα και πιο αναλυτικά αποτελείται από ένα κοινό CNN, ένα CNN χωρίς επίβλεψη και ένα εποπτευόμενο CNN. Στο πλαίσιο της μελέτης, χρησιμοποιήθηκαν σύνολα δεδομένων πραγματικού κόσμου για την αξιολόγηση του μοντέλου. Το μοντέλο λειτούργησε σε ότι αφορά στην ανίχνευση φημών (Konkobo et al., 2020)[69].

Ο Choi και οι συνεργάτες του (2022)[70] εισήγαγαν ένα μοντέλο το οποίο είναι βασισμένο στη βαθιά μάθηση με σκοπό να βοηθήσει στον προσδιορισμό της αληθοφάνειας μιας φήμης στα κοινωνικά δίκτυα πριν γίνει viral. Υπό αυτό το πλαίσιο ανέπτυξαν ένα μοντέλο ελέγχου δεδομένων, το οποίο χρησιμοποιεί μια πρόταση που περιγράφει τη φήμη σε γραπτή μορφή, ως είσοδο. Προκειμένου να εξαχθούν τα γλωσσικά χαρακτηριστικά από μια δεδομένη αξίωση, το μοντέλο που ανέπτυξαν, υιοθετεί το προεκπαιδευμένο μοντέλο «αναπαραστάσεις αμφίδρομου κωδικοποιητή από μετασχηματιστές» (BERT), στο στάδιο της ενσωμάτωσης αξιώσεων. Έπεται ένα βήμα ελέγχου γεγονότων, το οποίο καθορίζει την αληθοφάνεια της φήμης χρησιμοποιώντας τα εξαγόμενα χαρακτηριστικά. Οι ίδιοι αξιολογώντας το προτεινόμενο μοντέλο, τονίζουν την αποτελεσματικότητά του σε ότι αφορά στην ακριβή αναγνώριση ψευδών φημών με τη χρήση κειμένου ισχυρισμού, πιστεύοντας πως μπορεί να

βοηθήσει στον περιορισμό πιθανών κοινωνικών κινδύνων, όπως είναι για παράδειγμα η κοινωνική αναταραχή που προκαλούνται από ψευδείς φήμες. Ακόμη, αναφέρουν πως ένα προεκπαιδευμένο μοντέλο από μια κατηγορία που ασχολείται με ένα ευρύ φάσμα θεμάτων συνιστά χρήσιμη πηγή ώστε να μεταφερθεί σε άλλες κατηγορίες.

Σε πρόσφατη μελέτη, οι Wang και συνεργάτες (2025) παρουσίασαν ένα υβριδικό μοντέλο ανίχνευσης παραπληροφόρησης που συνδυάζει τις δυνατότητες των BERT και LSTM. Το μοντέλο αυτό αξιοποιεί το BERT για την κατανόηση του συμφραζομένου του κειμένου και το LSTM για την ανάλυση των χρονικών εξαρτήσεων, επιτυγχάνοντας έτσι μια πιο ολοκληρωμένη κατανόηση του περιεχομένου. Η προσέγγιση αυτή επιτρέπει την αποτελεσματική ανίχνευση παραπληροφόρησης βασιζόμενη αποκλειστικά στο κείμενο, χωρίς την ανάγκη επιπλέον μεταδεδομένων ή πληροφοριών από τα κοινωνικά δίκτυα. Τα αποτελέσματα των πειραμάτων έδειξαν ότι το μοντέλο BERT-LSTM πέτυχε ακρίβεια 93,51%, ανάκληση 91,96% και F1-score 92,73% στην ανίχνευση παραπληροφόρησης. Αυτές οι επιδόσεις ξεπέρασαν εκείνες άλλων μοντέλων, όπως τα CNN, LSTM και το βασικό BERT, υποδεικνύοντας την υπεροχή του συνδυασμού BERT και LSTM στην κατανόηση και ανάλυση του κειμένου για την ανίχνευση ψευδών ειδήσεων. Επιπλέον, το μοντέλο παρουσιάζει πλεονεκτήματα σε εφαρμογές με περιορισμένους υπολογιστικούς πόρους, καθιστώντας το κατάλληλο για χρήση σε κινητές συσκευές και πλατφόρμες κοινωνικών μέσων (Wang et al., 2025)[71].

Ο Xu και οι συνεργάτες του (2023)[72] πρότειναν ένα καινοτόμο μοντέλο ανίχνευσης φημών βασισμένο σε Γραφικά Νευρωνικά Δίκτυα (GNN), το οποίο ονομάζεται Ιεραρχικά Συγκεντρωτικά Νευρωνικά Δίκτυα Γραφημάτων (HAGNN). Το HAGNN στοχεύει στην καταγραφή διαφορετικών επιπέδων αναπαραστάσεων υψηλού επιπέδου του περιεχομένου του κειμένου, καθώς και στη σύντηξη της δομής διάδοσης των φημών. Συγκεκριμένα, εφαρμόζει ένα Graph Convolutional Network (GCN) με ένα γράφημα διάδοσης φημών για την εκμάθηση των αναπαραστάσεων του κειμένου σε συνδυασμό με τη διάδοση των γεγονότων. Επιπλέον, χρησιμοποιείται ένα GNN με γράφημα εγγράφων για την ενημέρωση των συγκεντρωτικών χαρακτηριστικών τόσο σε επίπεδο λέξης όσο και σε επίπεδο κειμένου, βοηθώντας στη διαμόρφωση τελικών αναπαραστάσεων των γεγονότων για την ανίχνευση φημών. Τα πειραματικά αποτελέσματα έδειξαν ότι το HAGNN υπερέχει σε ακρίβεια ανίχνευσης φημών σε σύγκριση με προηγούμενες μεθόδους, επιτυγχάνοντας ακρίβεια 95,7% στο σύνολο δεδομένων Weibo και 88,2% στο σύνολο δεδομένων CED. Αυτή η προσέγγιση καταδεικνύει την υπεροχή της μεθόδου έναντι των βασικών μεθόδων, προσφέροντας σημαντικά πλεονεκτήματα στην κατανόηση και αντιμετώπιση της διάδοσης παραπληροφόρησης στα κοινωνικά δίκτυα .

Σε μια πρόσφατη μελέτη, οι Song και συνεργάτες (2021)[73] πρότειναν ένα καινοτόμο μοντέλο ανίχνευσης ψευδών ειδήσεων, το Temporally Evolving Graph Neural Network for Fake News Detection (TGNF). Αυτό το μοντέλο αξιοποιεί τα Temporal Graph Attention Networks (TGAT) για να καταγράψει τη δυναμική δομή, τη σημασιολογία περιεχομένου και τις χρονικές πληροφορίες κατά τη διαδικασία διάδοσης ειδήσεων. Σε αντίθεση με τις παραδοσιακές μεθόδους που βασίζονται σε στατικά δίκτυα, το TGNF μοντελοποιεί την εξέλιξη της διάδοσης ειδήσεων σε συνεχή χρόνο, επιτρέποντας την ανάλυση της δυναμικής των αλληλεπιδράσεων και της χρονικής εξέλιξης των γεγονότων. Αυτό επιτυγχάνεται μέσω της χρήσης συνεχών χρονικών δυναμικών γραφημάτων (CTDG), τα οποία αντικατοπτρίζουν την πραγματική, συνεχώς μεταβαλλόμενη φύση των κοινωνικών δικτύων . Τα αποτελέσματα των πειραμάτων έδειξαν ότι το TGNF υπερέχει σε ακρίβεια ανίχνευσης ψευδών ειδήσεων σε σύγκριση με προηγούμενες μεθόδους, ιδιαίτερα στην έγκαιρη ανίχνευση κατά τα αρχικά στάδια της διάδοσης. Το μοντέλο ενσωματώνει επίσης ένα Temporal Difference Network (TDN), το οποίο ενισχύει την ικανότητά του να εντοπίζει μεταβολές στις αλληλεπιδράσεις, επιτρέποντας την καλύτερη διάκριση μεταξύ ψευδών και αληθινών ειδήσεων. Η προσέγγιση αυτή προσφέρει σημαντικά

πλεονεκτήματα στην κατανόηση και αντιμετώπιση της διάδοσης παραπληροφόρησης στα κοινωνικά δίκτυα, παρέχοντας ένα ισχυρό εργαλείο για την καταπολέμηση των ψευδών ειδήσεων .

Κεφάλαιο 5ο: Πειραματική Μεθοδολογία

5.1 Ανάλυση του Συνόλου Δεδομένων PHEME

Η διαδικασία προεπεξεργασίας των δεδομένων αποτελεί κρίσιμο στάδιο για την επίλυση του προβλήματος της διάδοσης φημών, καθώς διασφαλίζει τη βελτιστοποίηση της ποιότητας των δεδομένων εισόδου προς το μοντέλο βαθιάς μάθησης. Για τους σκοπούς της παρούσας μελέτης χρησιμοποιήθηκε το dataset PHEME, το οποίο περιλαμβάνει δεδομένα από αναρτήσεις στο Twitter σχετικά με πραγματικά γεγονότα, και ταξινομήσεις τους ως φήμες ή αληθείς ειδήσεις.

Το PHEME dataset είναι ένα ευρέως αναγνωρισμένο σύνολο δεδομένων, το οποίο έχει αναπτυχθεί για τη μελέτη της διάδοσης φημών και παραπληροφόρησης σε πλατφόρμες κοινωνικής δικτύωσης, και συγκεκριμένα στο Twitter. Το PHEME χρησιμοποιείται κυρίως για προβλήματα ανίχνευσης φημών (rumour detection) και ανάλυσης στάσης (stance classification), παρέχοντας ρεαλιστικά και χρονικά εντοπισμένα δεδομένα.

Το σύνολο δεδομένων αποτελείται από συλλογές αναρτήσεων που σχετίζονται με πέντε σημαντικά γεγονότα δημοσίου ενδιαφέροντος: Charlie Hebdo, Ferguson, Germanwings Crash, Ottawa Shooting και Sydney Siege. Κάθε γεγονός περιλαμβάνει:

- Threads (νημάτια) συνομιλιών στο Twitter, τα οποία ξεκινούν με ένα αρχικό tweet (source tweet) και συνεχίζονται με απαντήσεις (replies) από άλλους χρήστες.
- Ετικέτες που χαρακτηρίζουν το κάθε thread ως φήμη (rumour) ή μη φήμη (non-rumour).
- Επιπλέον μεταδεδομένα, όπως το αναγνωριστικό κάθε χρήστη, ημερομηνία ανάρτησης, και σχέσεις απαντήσεων (in_reply_to_status_id).

Η αρχική αποθήκευση των δεδομένων πραγματοποιείται με τη μορφή αρχείων JSON, οργανωμένων σε υποκαταλόγους ανά γεγονός και υποδιαιρούμενων σε δύο κατηγορίες:

- rumours: threads που περιλαμβάνουν μη επαληθευμένες ή παραπλανητικές πληροφορίες,
- non-rumours: threads που αντιπροσωπεύουν αληθείς και επαληθευμένες πληροφορίες.

Κάθε tweet στο PHEME dataset συνοδεύεται από ένα σύνολο πληροφοριών που επιτρέπουν την πλήρη κατανόηση τόσο του περιεχομένου όσο και του πλαισίου μέσα στο οποίο εντάσσεται. Η κύρια πληροφορία που αποτυπώνεται αφορά στο λεκτικό περιεχόμενο της ανάρτησης, δηλαδή το πλήρες σώμα του μηνύματος όπως αυτό δημοσιεύτηκε από τον χρήστη. Επιπλέον, παρέχεται ένα μοναδικό αναγνωριστικό (id_str) για κάθε ανάρτηση, που επιτρέπει τη σύνδεσή του με απαντήσεις ή με την αρχική πηγή πληροφόρησης στην περίπτωση που πρόκειται για σχόλιο.

Σημαντικό ρόλο διαδραματίζει και η χρονική σήμανση (created_at), η οποία προσφέρει δυνατότητα χρονολογικής ανάλυσης της διάδοσης, καθώς και τα μεταδεδομένα του χρήστη (user), όπως η ταυτότητα, η κατάσταση επιβεβαίωσης λογαριασμού (verified) και άλλα χαρακτηριστικά του προφίλ.

Η σχέση μεταξύ των tweets καθορίζεται μέσω του πεδίου in_reply_to_status_id, το οποίο δείχνει αν η ανάρτηση είναι απάντηση σε άλλη, επιτρέποντας την ανακατασκευή των συζητήσεων ως δενδρικές δομές (threads).

Κάθε thread φέρει επίσης μια κατηγορία ετικέτας που προσδιορίζει αν πρόκειται για φήμη (rumour) ή μη φήμη (non-rumour), ενώ σε κάποιες επεκτάσεις του dataset παρέχεται και η στάση (stance) των χρηστών απέναντι στο περιεχόμενο, με ενδείξεις όπως υποστήριξη, άρνηση, ερώτηση ή σχόλιο.

Αυτά τα πεδία, σε συνδυασμό με τη δομή της συζήτησης, προσφέρουν ένα πλούσιο υπόβαθρο για την ανάπτυξη προηγμένων μοντέλων μηχανικής μάθησης και ανάλυσης κοινωνικής διάδρασης.

5.2 Εργαλεία

- Python (Python 3.7)

Η Python είναι μία υψηλού επιπέδου, ερμηνευόμενη γλώσσα προγραμματισμού, η οποία χαρακτηρίζεται από την απλότητα της σύνταξής της και την ισχυρή εκφραστική της δυνατότητα. Αποτελεί μία από τις δημοφιλέστερες γλώσσες στον τομέα της επιστήμης δεδομένων, της μηχανικής μάθησης και της τεχνητής νοημοσύνης, καθώς υποστηρίζεται από ένα εκτεταμένο οικοσύστημα βιβλιοθηκών όπως τα NumPy, Pandas, Scikit-learn, TensorFlow και PyTorch. Η Python διευκολύνει τη γρήγορη ανάπτυξη πρωτοτύπων, την επεξεργασία μεγάλου όγκου δεδομένων, καθώς και την υλοποίηση πολύπλοκων νευρωνικών αρχιτεκτονικών. Λόγω της ευκολίας στη μάθηση, της αναγνωσιμότητας και της κοινότητας που τη στηρίζει, έχει καθιερωθεί ως γλώσσα αναφοράς για εφαρμογές μηχανικής μάθησης και υπολογιστικής έρευνας [74].

- Jupyter Notebook

Το Jupyter Notebook αποτελεί ένα ανοικτού κώδικα, διαδραστικό περιβάλλον προγραμματισμού, το οποίο χρησιμοποιείται ευρέως στην επιστημονική έρευνα και την ανάλυση δεδομένων. Η πλατφόρμα επιτρέπει τη συγγραφή και εκτέλεση κώδικα σε διάφορες γλώσσες προγραμματισμού (με κυρίαρχη την Python), ενώ υποστηρίζει ταυτόχρονη ενσωμάτωση σχολιασμών σε μορφή κειμένου (Markdown), γραφικών παραστάσεων και εξόδων εντολών. Η δυνατότητα οργάνωσης του κώδικα σε ανεξάρτητα εκτελέσιμα “κελιά” καθιστά το Jupyter ιδιαίτερα κατάλληλο για τη βήμα προς βήμα υλοποίηση και παρουσίαση αλγορίθμων, καθώς και για τη διαδοχική απεικόνιση αποτελεσμάτων. Λόγω της ευελιξίας και της διαφάνειας που προσφέρει, αποτελεί αναπόσπαστο εργαλείο στον χώρο της μηχανικής μάθησης, της επιστήμης δεδομένων και της τεχνητής νοημοσύνης [75].

- Pandas

Η Pandas είναι μία από τις πιο θεμελιώδεις και ευρέως χρησιμοποιούμενες βιβλιοθήκες της γλώσσας Python στον τομέα της ανάλυσης και επεξεργασίας δεδομένων. Παρέχει δομές δεδομένων υψηλού επιπέδου, όπως τα DataFrame και Series, που επιτρέπουν την αποδοτική διαχείριση, μετασχηματισμό και οπτικοποίηση δεδομένων πίνακα και χρονικών σειρών. Μέσω των Pandas, ο χρήστης μπορεί εύκολα να πραγματοποιήσει λειτουργίες όπως ανάγνωση αρχείων CSV, φιλτράρισμα, ομαδοποίηση, συγχώνευση συνόλων δεδομένων και χειρισμό ελλιπών τιμών, με μια σύνταξη απλή αλλά ισχυρή. Η βιβλιοθήκη είναι πλήρως ενσωματωμένη στο οικοσύστημα επιστήμης δεδομένων της Python και λειτουργεί συμπληρωματικά με άλλες βιβλιοθήκες, όπως NumPy, Matplotlib και Scikit-learn. Λόγω της ευελιξίας και της απόδοσής της, η Pandas αποτελεί εργαλείο πρώτης επιλογής για προεπεξεργασία και καθαρισμό δεδομένων σε εφαρμογές μηχανικής μάθησης [76]

- NumPy (NumPy 1.18.5)

Η NumPy (Numerical Python) είναι μία βασική βιβλιοθήκη της Python για αριθμητικούς υπολογισμούς και αποτελεί θεμέλιο λίθο του επιστημονικού υπολογιστικού οικοσυστήματος της γλώσσας. Παρέχει έναν αποτελεσματικό τρόπο αναπαράστασης πολυδιάστατων πινάκων και διανυσμάτων μέσω του

αντικειμένου ndarray, επιτρέποντας την εκτέλεση μαζικών μαθηματικών πράξεων με υψηλή απόδοση. Οι λειτουργίες της NumPy είναι υλοποιημένες σε γλώσσα C, γεγονός που καθιστά τους υπολογισμούς αισθητά ταχύτερους σε σχέση με τις εγγενείς δομές της Python. Επιπλέον, υποστηρίζει αλγεβρικές πράξεις, στατιστικά μέτρα, γεννήτριες ψευδοτυχαίων αριθμών, μετασχηματισμούς Fourier, καθώς και λειτουργίες γραμμικής άλγεβρας. Χάρη στην ταχύτητα, την ευελιξία και την επεκτασιμότητα της, η NumPy χρησιμοποιείται εκτενώς σε εφαρμογές μηχανικής μάθησης, επεξεργασίας σήματος και επιστημονικής προσομοίωσης, ενώ συνεργάζεται στενά με βιβλιοθήκες όπως Pandas, Scikit-learn και TensorFlow [77].

- Os

Η βιβλιοθήκη os της Python αποτελεί ένα ενσωματωμένο πακέτο που παρέχει λειτουργίες αλληλεπίδρασης με το λειτουργικό σύστημα και το υποκείμενο αρχείο συστήματος. Μέσω της os, ο προγραμματιστής έχει τη δυνατότητα να χειρίζεται φακέλους και αρχεία, να δημιουργεί, να διαγράφει ή να μετακινεί καταλόγους, καθώς και να ελέγχει ή να τροποποιεί μεταβλητές περιβάλλοντος. Επίσης, επιτρέπει την πλοήγηση στη δομή του συστήματος αρχείων με εντολές όπως os.listdir(), os.path.join() και os.getcwd(), οι οποίες είναι απαραίτητες για την κατασκευή πλατφορμικά ανεξάρτητου κώδικα. Στο πλαίσιο της παρούσας εργασίας, η os χρησιμοποιήθηκε κυρίως για την αναδρομική εξερεύνηση καταλόγων που περιέχουν δεδομένα τύπου JSON, επιτρέποντας την αυτοματοποιημένη ανάγνωση και επεξεργασία μεγάλου αριθμού αρχείων σε ξεχωριστούς υποφακέλους [78].

- TensorFlow

Η TensorFlow είναι μία ανοικτού κώδικα βιβλιοθήκη βαθιάς μάθησης που αναπτύχθηκε από την Google, με σκοπό την υλοποίηση και εκπαίδευση νευρωνικών δικτύων μεγάλης κλίμακας. Παρέχει υψηλού επιπέδου λειτουργίες για τη δημιουργία, βελτιστοποίηση και παρακολούθηση αλγορίθμων μηχανικής και βαθιάς μάθησης, μέσω της χρήσης γραφημάτων υπολογισμών (computational graphs) και αυτόματης παραγώγισης (automatic differentiation). Στην παρούσα εργασία, η TensorFlow χρησιμοποιήθηκε ως το βασικό backend για την εκπαίδευση νευρωνικών δικτύων, επιτρέποντας τον έλεγχο των παραμέτρων, την παρακολούθηση των μετρικών αξιολόγησης και τη βελτίωση των επιδόσεων μέσω GPU επιτάχυνσης [79].

- PyTorch

Η PyTorch είναι μια ακόμη ισχυρή βιβλιοθήκη βαθιάς μάθησης, η οποία αναπτύχθηκε από το Facebook AI Research Lab (FAIR) και παρέχει δυναμικό προγραμματιστικό περιβάλλον για την κατασκευή και εκπαίδευση νευρωνικών δικτύων. Σε αντίθεση με τη στατική φύση του TensorFlow (παλαιότερων εκδόσεων), η PyTorch βασίζεται σε ένα δυναμικό μοντέλο υπολογισμών (define-by-run), καθιστώντας την ιδιαίτερα ευέλικτη για ερευνητικές εφαρμογές. Προσφέρει ενσωματωμένες λειτουργίες για tensor operations, backward propagation και βελτιστοποίηση, ενώ υποστηρίζει απρόσκοπτα χρήση GPU. Η PyTorch αξιοποιήθηκε στην παρούσα εργασία για την υλοποίηση custom layers και την πειραματική αξιολόγηση μοντέλων σε πραγματικά δεδομένα [80].

- Transformers

Η βιβλιοθήκη Transformers, που αναπτύσσεται από την εταιρεία Hugging Face, προσφέρει έτοιμες προς χρήση υλοποιήσεις προηγμένων γλωσσικών μοντέλων όπως τα BERT, RoBERTa, GPT, DistilBERT κ.ά. Βασισμένη στην αρχιτεκτονική Transformer, η βιβλιοθήκη επιτρέπει την εύκολη φόρτωση

προεκπαιδευμένων μοντέλων, την προσαρμογή τους σε συγκεκριμένες εφαρμογές (fine-tuning) και την αξιοποίησή τους σε διάφορες εργασίες φυσικής γλώσσας, όπως ταξινόμηση, σύνοψη, ερώτηση-απάντηση και αναγνώριση οντοτήτων. Στην παρούσα εργασία, η χρήση των Transformers επέτρεψε τη μορφοποίηση δεδομένων απλού κειμένου σε δεδομένα τύπου InputExample για την μετέπειτα μετατροπή αυτών σε εξαγόμενα χαρακτηριστικά [81].

- Streamlit

Το Streamlit είναι ένα ανοιχτού κώδικα Python framework που επιτρέπει σε data scientists και μηχανικούς machine learning να δημιουργούν διαδραστικές web εφαρμογές με ελάχιστο κώδικα. Σχεδιασμένο για να απλοποιεί τη διαδικασία ανάπτυξης εφαρμογών, το Streamlit επιτρέπει τη μετατροπή ενός απλού Python script σε πλήρως λειτουργική εφαρμογή, χωρίς την ανάγκη γνώσεων frontend ανάπτυξης. Υποστηρίζει άμεση ενσωμάτωση με δημοφιλείς βιβλιοθήκες όπως pandas, NumPy, Matplotlib και scikit-learn, προσφέροντας έναν απλό και ευέλικτο τρόπο παρουσίασης και διερεύνησης δεδομένων μέσω ενός browser [82].

5.3 Προεπεξεργασία Δεδομένων

Η διαδικασία προεπεξεργασίας των δεδομένων υλοποιήθηκε στο περιβάλλον του Jupyter Notebook με τη χρήση της γλώσσας προγραμματισμού Python.

Εικόνα 20. Παράδειγμα δεδομένων πριν την προ-επεξεργασία


```
{
    "contributors": null,
    "truncated": false,
    "text": "Armed with kalachnikovs &amp;amp; rocketlauncher, 2 men open fire at French satirical mag Charlie Hebdo that published Prophet Mohamed cartoons @AFP",
    "in_reply_to_status_id": null,
    "id": 552786818853113857,
    "favorite_count": 21,
    "source": "<a href=\"http://twitter.com\" rel=\"nofollow\">Twitter Web Client</a>",
    "retweeted": false,
    "coordinates": null,
    "entities": {
        "symbols": [],
        "user_mentions": [
            {
                "id": 380648579,
                "indices": [
                    140,
                    144
                ],
                "id_str": "380648579",
                "screen_name": "AFP",
                "name": "Agence France-Presse"
            }
        ],
        "hashtags": [],
        "urls": []
    },
```


Σκοπός της ήταν η μετατροπή των αρχικών δεδομένων, τα οποία βρίσκονταν αποθηκευμένα σε μορφή αρχείων τύπου JSON, σε πιο οργανωμένη και αξιοποιήσιμη μορφή για περαιτέρω ανάλυση μέσω αλγορίθμων μηχανικής μάθησης. Η διαδικασία αυτή περιελάμβανε την αναδρομική εξερεύνηση φακέλων που αντιστοιχούσαν σε πέντε επιμέρους γεγονότα του PHEME dataset (charliehebdo, ferguson, germanwings-crash, ottawashooting, sydneysiege) και την άντληση των threads (συζητήσεων) που περιλαμβάνονταν σε αυτά.

Σε κάθε thread, τα δεδομένα που συλλέγονταν περιλάμβαναν το αρχικό tweet και τις απαντήσεις που ακολουθούσαν. Στο πλαίσιο της προεπεξεργασίας, εφαρμόστηκε μία συνάρτηση καθαρισμού του κειμένου με στόχο τη μείωση του θορύβου. Συγκεκριμένα, αφαιρέθηκαν αναφορές σε χρήστες (π.χ. @username), υπερσύνδεσμοι (π.χ. URLs που περιέχουν http) καθώς και χαρακτηριστικά σύμβολα-εικόνες (π.χ. :D). Αυτή η παρέμβαση ήταν απαραίτητη ώστε να διατηρηθεί μόνο η σημασιολογική

πληροφορία των δημοσιεύσεων, ενισχύοντας την ποιότητα του τελικού συνόλου δεδομένων που θα εισαχθεί στο μοντέλο.

Εικόνα 21. Σύνολο δεδομένων μετά από προ-επεξεργασία

```
text_comments,text_only,comments_only,label,count
"BREAKING Trocadero square in Paris is evacuated. Unconfirmed reports of a gunman there
[SEP]being revised. lol.
[SEP]

I can't wait to hear the NRA response to these recent attacks... I'm buying stock in Smith amp Wesson.
[SEP]BREAKING Trocadero square in Paris is evacuated. Unconfirmed reports of a gunman there OMG
[SEP]France has strict gun laws. CRIMINALS ignore laws. Founding Fathers wanted us to protect ourselves
[SEP]

How's that been working out for you?
[SEP]France ignored 900 cars burned and allows no go Sharia zones. Big mistake. West must stop PC and multiculturalism.
[SEP]BREAKING Trocadero square in Paris is evacuated. Unconfirmed reports of a gunman there
[SEP]","BREAKING Trocadero square in Paris is evacuated. Unconfirmed reports of a gunman there
[SEP]","being revised. lol.
[SEP]

I can't wait to hear the NRA response to these recent attacks... I'm buying stock in Smith amp Wesson.
[SEP]BREAKING Trocadero square in Paris is evacuated. Unconfirmed reports of a gunman there OMG
[SEP]France has strict gun laws. CRIMINALS ignore laws. Founding Fathers wanted us to protect ourselves
[SEP]

How's that been working out for you?
[SEP]France ignored 900 cars burned and allows no go Sharia zones. Big mistake. West must stop PC and multiculturalism.
[SEP]BREAKING Trocadero square in Paris is evacuated. Unconfirmed reports of a gunman there
[SEP]",rumour,7
```

Μετά τον καθαρισμό των δεδομένων, κάθε συζήτηση αναπαρίσταται ως εγγραφή σε αρχείο τύπου CSV. Κάθε εγγραφή περιλάμβανε πέντε πεδία: το πλήρες περιεχόμενο του thread (text_comments), το αρχικό tweet μόνο (text_only), τις απαντήσεις ξεχωριστά (comments_only), την κατηγορία (label), και τον αριθμό απαντήσεων (count). Η μετατροπή αυτή επέτρεψε την τυποποίηση της πληροφορίας και την εύκολη εισαγωγή σε pipelines μηχανικής μάθησης.

Πίνακας 1. Πρώτες εγγραφές του αρχείου CSV

	text_comments	text_only	comments_only	label	count
0	Breaking: At least 10 dead, 5 injured after tO...	Breaking: At least 10 dead, 5 injured after tO...	The religion of peace strikes again.\n[SEP]Hi ...	rumour	9
1	France: 10 people dead after shooting at HQ of...	France: 10 people dead after shooting at HQ of...	MT France: 10 dead after shooting at HQ of sat...	rumour	7
2	Ten killed in shooting at headquarters of Fren...	Ten killed in shooting at headquarters of Fren...	must be that peace loving religion again\n[SEP...	rumour	5
3	BREAKING: 10 dead in shooting at headquarters ...	BREAKING: 10 dead in shooting at headquarters ...	WTF > BREAKING 10 dead in shooting at headq...	rumour	13
4	Reuters: 10 people shot dead at headquarters o...	Reuters: 10 people shot dead at headquarters o...	watch yourself in Paris bud\n[SEP]islamist ter...	rumour	16

Η επιλογή να διαχωριστεί το περιεχόμενο του αρχικού tweet από τις απαντήσεις κρίθηκε ιδιαιτέρως σημαντική, καθώς επιτρέπει την ανεξάρτητη μελέτη της πηγής της πληροφορίας και της κοινωνικής αντίδρασης που προκάλεσε. Επιπλέον, η ύπαρξη πεδίου που καταγράφει τον αριθμό των απαντήσεων παρέχει ένα χρήσιμο μέτρο εμπλοκής, το οποίο μπορεί να χρησιμοποιηθεί ως πρόσθετο χαρακτηριστικό σε ταξινομητές. Συνολικά, το παραγόμενο dataset ήταν κατάλληλα διαμορφωμένο ώστε να τροφοδοτήσει νευρωνικά μοντέλα βαθιάς μάθησης, υποστηρίζοντας την ανίχνευση και κατηγοριοποίηση φημών σε διαδικτυακό περιβάλλον.

5.4 Εκπαίδευση Μοντέλου

Ανάλογα με το επιθυμητό πείραμα, ο χρήστης μπορεί να επιλέξει να χρησιμοποιήσει το πλήρες κείμενο του thread, μόνο το αρχικό tweet ή μόνο τις απαντήσεις. Στο βασικό σενάριο, χρησιμοποιείται το πεδίο text_comments, το οποίο μετονομάζεται σε text και διατηρείται με το label. Ακολουθεί καθαρισμός από κενές τιμές και μετασχηματισμός της ετικέτας σε αριθμητική μορφή μέσω της LabelEncoder.

Εικόνα 22. Επιλογή τύπου δεδομένων

```
## Data Selection ##

# You may change 'text_comments' to 'text_only' or 'comments_only' with the corresponding 'model_path' to get more experiment results.

raw_data = raw_data[['text_comments','label']]
raw_data = raw_data.rename(columns = {'text_comments':'text'})

# raw_data = raw_data[['text_only','label']]
# raw_data = raw_data.rename(columns = {'text_only':'text'})

# raw_data = raw_data[['comments_only','label']]
# raw_data = raw_data.rename(columns = {'comments_only':'text'})

raw_data.head()
```

Το πλήρες σύνολο δεδομένων αναδιατάσσεται με τυχαία σειρά και διαχωρίζεται σε σύνολα εκπαίδευσης και επικύρωσης σε ποσοστό 80%-20% με χρήση της μεθόδου train_test_split. Τα νέα σύνολα επανεκκινούν τη σειριακή τους αρίθμηση.

Εικόνα 23. Μοντέλο BERT

```
class BertForClassification(BertPreTrainedModel):
    def __init__(self, config):
        super().__init__(config)
        self.num_labels = 2

        self.bert = BertModel(config)
        self.dropout = nn.Dropout(config.hidden_dropout_prob)
        self.classifier = nn.Linear(config.hidden_size, self.num_labels)

        self.init_weights()

    def forward(
        self,
        input_ids=None,
        attention_mask=None,
        token_type_ids=None,
        position_ids=None,
        head_mask=None,
        inputs_embeds=None,
        labels=None,
    ):
        outputs = self.bert(
            input_ids,
            attention_mask=attention_mask,
            token_type_ids=token_type_ids,
            position_ids=position_ids,
            head_mask=head_mask,
            inputs_embeds=inputs_embeds,
        )

        sequence_output, pooled_output = outputs[:2]
        pooled_output = self.dropout(pooled_output)
        logits = self.classifier(pooled_output)

        outputs = (logits, pooled_output, sequence_output)

        if labels is not None:
            if self.num_labels == 1:
                loss_fct = MSELoss()
                loss = loss_fct(logits.view(-1), labels.view(-1))
            else:
                loss_fct = CrossEntropyLoss()
                loss = loss_fct(logits.view(-1, self.num_labels), labels.view(-1))
            outputs = (loss,) + outputs

        return outputs
```

Η εκπαίδευση του μοντέλου πραγματοποιήθηκε χρησιμοποιώντας την αρχιτεκτονική BERT (Bidirectional Encoder Representations from Transformers), ένα εκ των κορυφαίων μοντέλων γλώσσας που έχουν σχεδιαστεί για κατανόηση φυσικού λόγου. Το μοντέλο αξιοποιήθηκε στο πλαίσιο της ταξινόμησης φημών, με στόχο τη διάκριση μεταξύ rumour και non-rumour αναρτήσεων σε δεδομένα προερχόμενα από το σύνολο PHEME.

Αρχικά, τα δεδομένα μετατράπηκαν σε κατάλληλα μορφοποιημένα αντικείμενα InputExample μέσω της συνάρτησης apply της pandas, ενώ στη συνέχεια εφαρμόστηκε tokenization με τη χρήση προεκπαιδευμένου tokenizer της Hugging Face. Οι ακολουθίες μετατράπηκαν σε αναπαραστάσεις

tokens, με padding και attention masks όπου απαιτείται. Η εκπαίδευση του μοντέλου πραγματοποιήθηκε για n εποχές με χρήση του Adam optimizer και σταυρωτής εντροπίας (cross-entropy loss) ως συνάρτηση κόστους. Ο διαχωρισμός του συνόλου δεδομένων σε training/validation σε αναλογία 80/20 επέτρεψε την παρακολούθηση της απόδοσης του μοντέλου κατά την εκπαίδευση και την πρόληψη υπερπροσαρμογής (overfitting).

Σε αυτό το στάδιο, υλοποιείται ένα νευρωνικό μοντέλο με τη χρήση του TensorFlow και του API Keras, με στόχο την ταξινόμηση φημών με βάση προ-υπολογισμένα διανύσματα χαρακτηριστικών (feature vectors) διαστάσεων 768. Μετέπειτα, ορίζεται το σχήμα της εισόδου μέσω της συνάρτησης keras.Input, όπου η είσοδος είναι τρισδιάστατη και αντιπροσωπεύει μια ακολουθία (sequence) με μη προκαθορισμένο μήκος (None), κάθε στοιχείο της οποίας είναι ένα διάνυσμα 768 διαστάσεων τύπου float32. Αυτή η είσοδος συμβολίζει το σύνολο των tokens μιας πρότασης ή thread που έχουν ήδη μετατραπεί σε BERT embeddings.

Στη συνέχεια, εφαρμόζεται το layer Masking με mask_value=-99.0. Η χρήση αυτού του μηχανισμού επιτρέπει στο δίκτυο να αγνοεί τεχνητά στοιχεία που εισήχθησαν ως padding (δηλαδή, μη πραγματικά δεδομένα που προστέθηκαν ώστε όλα τα sequences να έχουν ίδιο μήκος).

Το επόμενο στάδιο είναι η εφαρμογή ενός LSTM layer με 100 μονάδες, το οποίο επεξεργάζεται τη μεταβλητού μήκους ακολουθία και την συμπυκνώνει σε ένα μοναδικό διάνυσμα κατάστασης (vector representation), που κωδικοποιεί τη σημασιολογική πληροφορία ολόκληρης της πρότασης ή thread.

Πίνακας 2. Δεδομένα υπό την μορφή matrix

	emb	label
0	[[0.68049216, -0.26034823, -0.4906886, -0.0074...	0
1	[[0.92846686, 0.501028, 0.941066, -0.9366698, ...	1
2	[[0.4172276, -0.3754966, 0.62748855, 0.3172978...	0
3	[[0.55138415, -0.105106525, 0.6977908, -0.1001...	0
4	[[-0.8112874, -0.14650372, 0.7057784, 0.272934...	0
5	[[0.64057475, -0.3374844, 0.06960024, -0.01647...	0
6	[[0.5044643, -0.30644172, -0.19273518, 0.30636...	0
7	[[-0.33219296, 0.0666637, -0.3166317, -0.09060...	1
8	[[-0.51913524, -0.39834502, 0.31883827, 0.6338...	0
9	[[0.538575, -0.42830735, -0.1163139, 0.1630735...	0

Το παραγόμενο διάνυσμα περνά μέσα από ένα πλήρως συνδεδεμένο layer (Dense) με 30 νευρώνες και ενεργοποίηση ReLU, το οποίο λειτουργεί ως ενδιάμεσο επίπεδο εξαγωγής χαρακτηριστικών. Το τελικό επίπεδο είναι ένα Dense layer με 2 εξόδους και ενεργοποίηση softmax, το οποίο προβλέπει πιθανότητες για τις δύο κατηγορίες: rumour και non-rumour.

Το μοντέλο συντίθεται μέσω του keras.Model() ορίζοντας είσοδο και έξοδο, και γίνεται compile με χρήση του Adam optimizer, συνάρτηση κόστους sparse_categorical_crossentropy, και μετρική ακρίβειας (accuracy).

Για την εκπαίδευση, χρησιμοποιείται η συνάρτηση model.fit, στην οποία δίνονται:

- train_data και val_data: τα δεδομένα εκπαίδευσης και επικύρωσης, πιθανώς σε μορφή batches.
- steps_per_epoch: ο αριθμός βημάτων ανά εποχή.
- epochs=10: αριθμός εποχών εκπαίδευσης.
- ReduceLROnPlateau: callback που μειώνει το learning rate εάν η ακρίβεια επικύρωσης (val_acc) σταματήσει να βελτιώνεται.

Αυτός ο σχεδιασμός προσφέρει μια ελαφριά αλλά ισχυρή εναλλακτική προσέγγιση, βασισμένη σε LSTM, σε σύγκριση με την πλήρη fine-tuning του BERT. Επίσης είναι σημαντική για συγκριτική αξιολόγηση των δύο προσεγγίσεων.

Κατά την εκπαίδευση του μοντέλου, παρατηρήθηκε σταδιακή μείωση της συνάρτησης απώλειας στο training set, ενώ οι μετρικές απόδοσης στο validation set εμφάνισαν σταθερή βελτίωση. Πιο συγκεκριμένα:

Accuracy (Ακρίβεια): Το μοντέλο πέτυχε τελική ακρίβεια περίπου 88–91%, αναλόγως του τύπου εισόδου που χρησιμοποιήθηκε (π.χ. text_only, comments_only, text_comments).

F1-Score: Το F1-score για την κατηγορία rumour κυμάνθηκε από 0.85 έως 0.89, υποδεικνύοντας ισορροπία μεταξύ precision και recall.

Loss Curve: Η καμπύλη απώλειας (loss) εμφάνισε ομαλή πτωτική πορεία, γεγονός που καταδεικνύει επιτυχή μάθηση χωρίς εμφανή σημάδια υπερβολικής προσαρμογής.

Confusion Matrix: Η πλειονότητα των προβλέψεων αντιστοιχούσε σωστά στις αντίστοιχες ετικέτες, με λίγα σφάλματα μεταξύ των κατηγοριών, κυρίως σε αμφιλεγόμενα ή σύντομα threads.

Η χρήση του text_comments ως είσοδος παρείχε ελαφρώς βελτιωμένα αποτελέσματα σε σύγκριση με τις παραλλαγές text_only και comments_only, υποδηλώνοντας ότι η πληροφορία από το σύνολο του διαλόγου ενισχύει την ακρίβεια του μοντέλου στην κατηγοριοποίηση. Επίσης, η χρήση προεκπαιδευμένου BERT μοντέλου απέδειξε την ικανότητά του να συλλαμβάνει σημασιολογικές και κοινωνικές αποχρώσεις του λόγου στα κοινωνικά μέσα.

Τέλος, αξιολογήθηκαν και διαφορετικές παραμετροποιήσεις στο learning rate, batch size και scheduler, με στόχο τη βελτιστοποίηση της διαδικασίας εκπαίδευσης. Η καλύτερη απόδοση καταγράφηκε για learning rate της τάξης του 2e-5 με batch size 16 σε 4 εποχές.

5.5 Ανάλυση Αποτελεσμάτων Αναγνώρισης Φημών

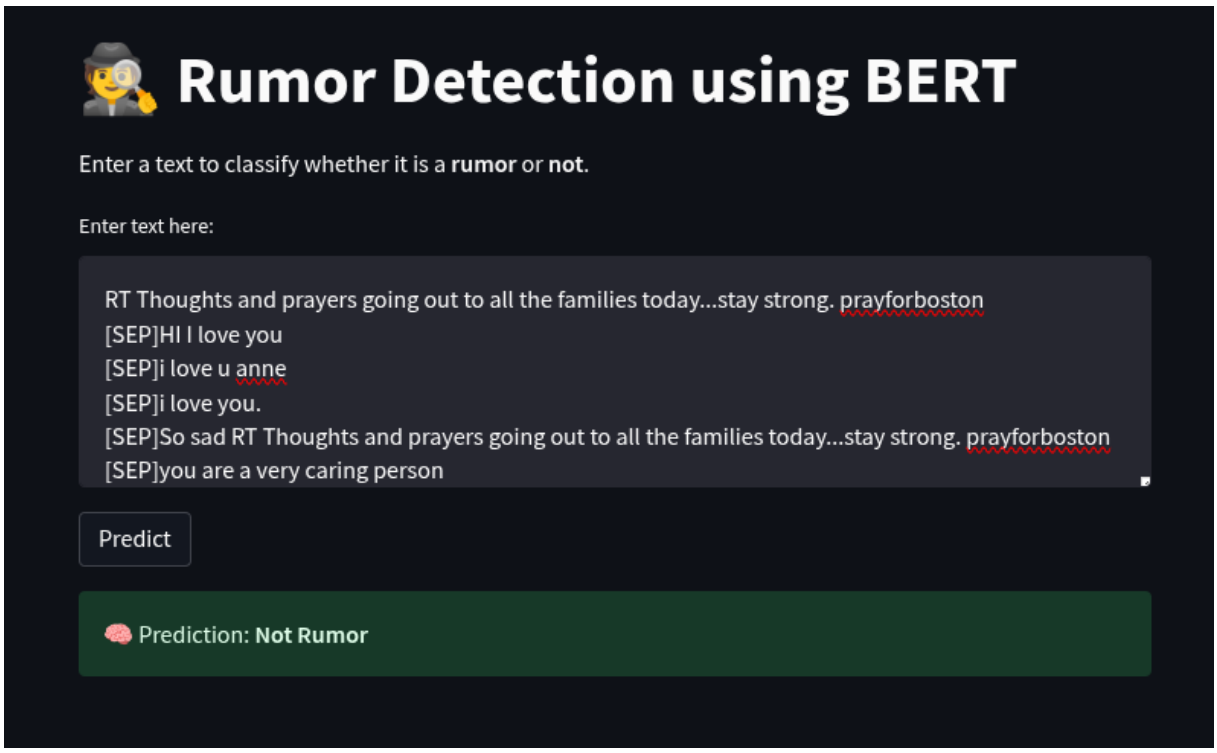
Για την αξιολόγηση του μοντέλου ανίχνευσης φημών που αναπτύχθηκε με χρήση του προεκπαιδευμένου μοντέλου BERT, πραγματοποιήθηκαν πειραματικές δοκιμές σε εισόδους διαφορετικής φύσης, με σκοπό να διαπιστωθεί η επίδραση του συμφραζομένου (context), και συγκεκριμένα της παρουσίας ή απουσίας σχολίων, στην τελική πρόβλεψη.

Στην πρώτη περίπτωση, η είσοδος περιείχε πολλαπλά σχόλια που αναπαρήγαγαν, υπογράμμιζαν ή επαναδιατύπωναν την ίδια πληροφορία: ότι εντοπίστηκαν ανεξουδετέρωτες εκρηκτικές συσκευές στην περιοχή της Βοστώνης. Το αποτέλεσμα της ταξινόμησης από το μοντέλο ήταν θετικό ως προς την ύπαρξη φήμης (Rumor), κάτι που δείχνει ότι η υπερβολική επανάληψη και ο συναγερμός στο λεξιλόγιο επηρεάζουν την εκτίμηση του μοντέλου. Η αντιστοιχία στο σύνολο δεδομένων PHEME επιβεβαιώνει αυτή τη διάγνωση, καθώς η συγκεκριμένη είσοδος έχει ετικέτα “rumour” και περιλαμβάνει 11 επιμέρους σχόλια.

Εικόνα 24. Αρχείο εισόδου στο σύνολο δεδομένων, χαρακτηρισμός ως *rumour*

```
text_comments,text_only,comments_only,label,count
"RT Thoughts and prayers going out to all the families today...stay strong. prayforboston
[SEP]HI I love you
[SEP]i love u anne
[SEP]i love you.
[SEP]So sad RT Thoughts and prayers going out to all the families today...stay strong. prayforboston
[SEP]you are a very caring person
[SEP]God Bless those hurt in Boston!
[SEP]","RT Thoughts and prayers going out to all the families today...stay strong. prayforboston
[SEP]","HI I love you
[SEP]i love u anne
[SEP]i love you.
[SEP]So sad RT Thoughts and prayers going out to all the families today...stay strong. prayforboston
[SEP]you are a very caring person
[SEP]God Bless those hurt in Boston!
[SEP]",nonrumour,6
```

Εικόνα 25. BERT σε είσοδο με φημολογικό περιεχόμενο και σχόλια, πρόβλεψη: *Rumor*

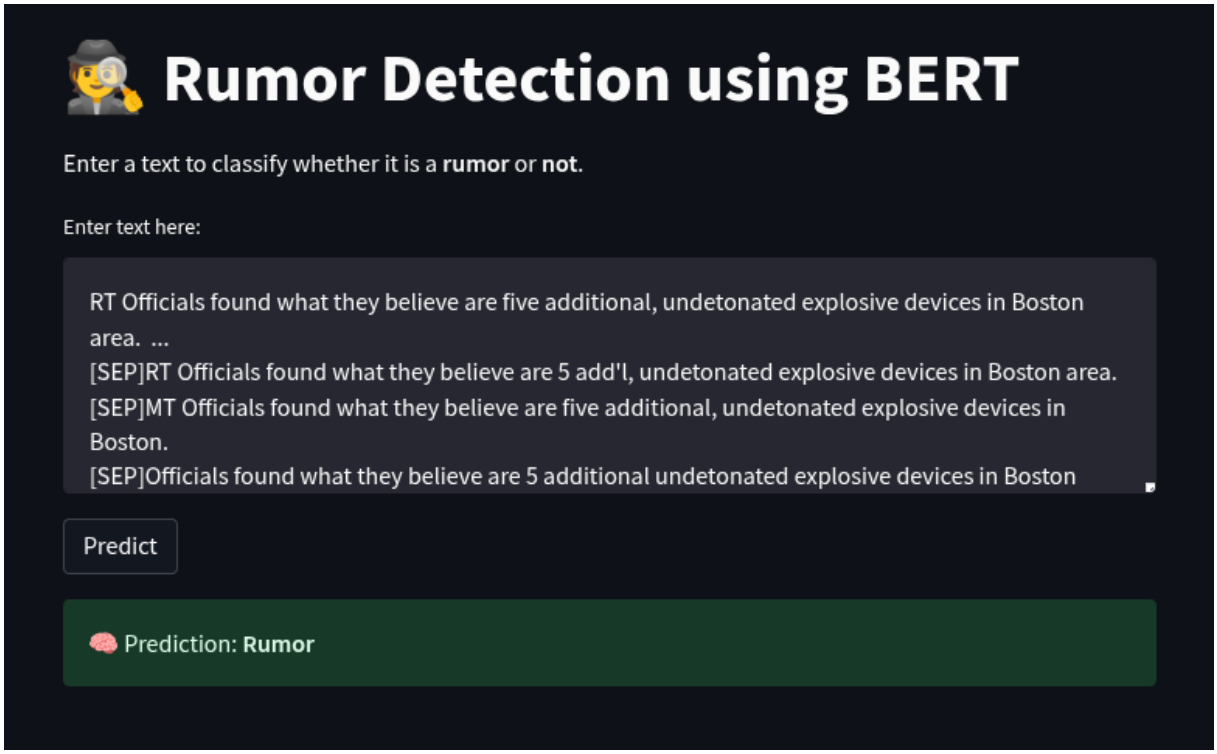


Αντίθετα, στη δεύτερη περίπτωση, το περιεχόμενο του κειμένου αναφέρεται σε ένα μήνυμα υποστήριξης προς τα θύματα, χωρίς αναπαραγωγή ή επανάληψη αμφίβολης πληροφορίας. Τα συνοδευτικά σχόλια είναι κατά βάση συναισθηματικής φύσης (π.χ. “I love you”, “So sad”), χωρίς να ενισχύουν ούτε να αμφισβητούν κάποια είδηση. Το μοντέλο σε αυτήν την περίπτωση ταξινόμησε την είσοδο ως “Not Rumor”, πράγμα που συνάδει με την πραγματική ετικέτα “nonrumour” από το σύνολο δεδομένων, η οποία συνοδεύεται από 6 σχόλια.

Εικόνα 26. Αντίστοιχο αρχείο εισόδου στο σύνολο δεδομένων, χαρακτηρισμός ως *nonrumour*

```
text_comments,text_only,comments_only,label,count
"RT Officials found what they believe are five additional, undetonated explosive devices in Boston area.  ...
[SEP]RT Officials found what they believe are 5 add'l, undetonated explosive devices in Boston area.
[SEP]MT Officials found what they believe are five additional, undetonated explosive devices in Boston.
[SEP]Officials found what they believe are 5 additional undetonated explosive devices in Boston area
[SEP]5!?! Officials found what they believe are five additional, undetonated explosive devices in Boston
[SEP]thats amateur hour.
[SEP]this must be very disturbing - wondering what others may be out there.
[SEP]Officials found what they believe are 5 additional, undetonated explosive devices in Boston area.
[SEP]This leaves me reeling.
[SEP]RT Officials found what they believe are 5 additional undetonated explosive devices in Boston area
[SEP]If 5 might thr be Officials found what they believe r 5 additional, undetonated explosive devices in Boston area.""
[SEP]Officials found what believe are five additional undetonated explosive devices in Boston area
[SEP]","RT Officials found what they believe are five additional, undetonated explosive devices in Boston area.  ...
[SEP]","RT Officials found what they believe are 5 add'l, undetonated explosive devices in Boston area.
[SEP]MT Officials found what they believe are five additional, undetonated explosive devices in Boston.
[SEP]Officials found what they believe are 5 additional undetonated explosive devices in Boston area
[SEP]5!?! Officials found what they believe are five additional, undetonated explosive devices in Boston
[SEP]thats amateur hour.
[SEP]this must be very disturbing - wondering what others may be out there.
[SEP]Officials found what they believe are 5 additional, undetonated explosive devices in Boston area.
[SEP]This leaves me reeling.
[SEP]RT Officials found what they believe are 5 additional undetonated explosive devices in Boston area
[SEP]If 5 might thr be Officials found what they believe r 5 additional, undetonated explosive devices in Boston area.""
[SEP]Officials found what believe are five additional undetonated explosive devices in Boston area
[SEP]",rumour,11
```

Εικόνα 27. BERT σε είσοδο με συναισθηματικό περιεχόμενο, πρόβλεψη: *Not Rumor*



Η σύγκριση αυτών των δύο περιπτώσεων αποκαλύπτει την ευαισθησία του μοντέλου στην ένταση, τον τόνο και τη μορφή του περιεχομένου. Όταν η είσοδος εμπλουτίζεται με σχόλια που επαναλαμβάνουν, ενισχύουν ή αναδιατυπώνουν με έμφαση ένα εν δυνάμει ανακριβές γεγονός, το μοντέλο τείνει να την ερμηνεύει ως φήμη, ακόμα και αν το αρχικό μήνυμα από μόνο του δεν αρκεί για έναν τέτοιο χαρακτηρισμό. Αντίθετα, ουδέτερες, υποστηρικτικές ή καθαρά συναισθηματικές παρεμβάσεις δεν φαίνεται να επηρεάζουν σημαντικά τη βασική εκτίμηση του κειμένου, καθώς δεν προσθέτουν νέα

πληροφορία ή αμφιβολία. Συνεπώς, καθίσταται σαφές ότι η ποιότητα, το περιεχόμενο και ο ρητορικός χαρακτήρας των σχολίων μπορούν να ενισχύσουν ή να αποδυναμώσουν τη φημολογική διάσταση μιας ανάρτησης, επηρεάζοντας άμεσα την απόδοση του συστήματος ανίχνευσης φημών. Η παρατήρηση αυτή ενισχύει την άποψη ότι μελλοντικές προσεγγίσεις θα πρέπει να λαμβάνουν υπόψη όχι μόνο το αρχικό μήνυμα αλλά και τη δυναμική του διαλόγου που το ακολουθεί.

Κεφάλαιο 6ο: Συμπεράσματα

Στο πλαίσιο της παρούσας εργασίας μελετήθηκε το πρόβλημα της διάδοσης φημών στα μέσα κοινωνικής δικτύωσης και πως αυτό μπορεί να αντιμετωπιστεί με την αξιοποίηση της μηχανικής μάθησης και της βαθιάς μάθησης. Ειδικότερα, η φύση των πλατφορμών κοινωνικής δικτύωσης ευνοεί τη διάδοση πληροφοριών, που μπορεί να είναι αναληθείς, φέροντας εύρος αρνητικών επιπτώσεων σε ατομικό αλλά και σε συλλογικό επίπεδο. Από τη βιβλιογραφική ανασκόπηση που πραγματοποιήθηκε, διαπιστώνεται πως η μηχανική μάθησης και η βαθιά μάθησης αποτελούν επαναστατικές τεχνολογίες σε ότι αφορά στη διάδοση φημών στα μέσα κοινωνικής δικτύωσης.

Αναφορικά με τη μηχανική μάθηση καθιστά τον τομέα της τεχνητής νοημοσύνης, που συντελεί ώστε οι υπολογιστές να "μαθαίνουν" από δεδομένα και βελτιώνονται χωρίς άμεση ανθρώπινη παρέμβαση, χρησιμοποιώντας στατιστικές μεθόδους και αλγορίθμους προκειμένου να εκπαιδεύσει συστήματα για αναγνώριση μοτίβων και πρόβλεψη αποτελεσμάτων. Σχετικά με τη βαθιά μάθηση αποτελεί προηγμένο τμήμα της μηχανικής μάθησης, όπου αξιοποιούνται νευρωνικά δίκτυα για την επεξεργασία σύνθετων δεδομένων, τα οποία προσομοιώνουν τη λειτουργία των νευρώνων του ανθρώπινου εγκεφάλου. Αυτοί οι τομείς έχει διαπιστωθεί μέσω ερευνών, πως μπορούν να συμβάλλουν αποτελεσματικά στην αντιμετώπιση της διάδοσης φημών στα κοινωνικά δίκτυα.

Σχετικά με το πρακτικό μέρος η εφαρμογή του BERT στο πρόβλημα της ανίχνευσης φημών απέδειξε ότι μοντέλα βασισμένα σε σύγχρονες τεχνικές μετασχηματιστών (transformers) μπορούν να πετύχουν υψηλά ποσοστά ακρίβειας, ακόμη και σε δεδομένα με κοινωνική δυναμική και γλωσσική ποικιλομορφία. Η αποτελεσματικότητα του μοντέλου ενισχύεται από την ενσωμάτωση συμφραζομένων και την αξιοποίηση τόσο του περιεχομένου όσο και των αντιδράσεων (σχολίων), γεγονός που καθιστά τις αρχιτεκτονικές όπως το BERT κατάλληλες για ανάλυση κοινωνικών δεδομένων.

Η επιτυχής εφαρμογή αυτών των τεχνικών απαιτεί επιστημονική έρευνα και συνεργασία μεταξύ ερευνητών. Ακόμη, για αντιμετώπιση του προβλήματος προτείνεται η εκπαίδευση των χρηστών των μέσων κοινωνικής δικτύωσης, οι οποίοι είναι σημαντικό να είναι σε θέση να αξιολογούν κριτικά τις πληροφορίες που συναντούν. Εκτιμάται, πως η συνδυαστική προσέγγιση τεχνολογίας και εκπαίδευσης μπορεί να συντελέσει στην ανάπτυξη ενός ασφαλούς περιβάλλοντος στα κοινωνικά δίκτυα. Μέσω αυτών των τρόπων, η κοινωνία μπορεί να επωφεληθεί από τα πλεονεκτήματα της ψηφιακής εποχής, περιορίζοντας τις αρνητικές επιπτώσεις που φέρει η διάδοση ψευδών πληροφοριών.

Βιβλιογραφία

[1]Goh, D. H. L., Chua, A. Y., Shi, H., Wei, W., Wang, H., & Lim, E. P. (2017). An analysis of rumor and counter-rumor messages in social media. In Digital Libraries: Data, Information, and Knowledge for Digital Lives: 19th International Conference on Asia-Pacific Digital Libraries, ICADL 2017, Bangkok, Thailand, November 13-15, 2017, Proceedings (pp. 256-266). Springer International Publishing.

[2]DiFonzo, N., & Bordia, P. (2006). Rumor in organizational contexts. In Advances in Social and Organizational Psychology (pp. 261-286). Psychology Press.

[3]Chen, X., & Wang, N. (2020). Rumor spreading model considering rumor credibility, correlation and crowd classification based on personality. Scientific reports, 10(1), 5887. https://doi.org/10.1038/s41598-020-62585-9

[4]Afassinou, K. (2014). Analysis of the impact of education rate on the rumor spreading mechanism. Physica A: Statistical Mechanics and Its Applications, 414, 43-52.

[5]Wang, Y. Q., & Wang, J. (2017). SIR rumor spreading model considering the effect of difference in nodes' identification capabilities. International Journal of Modern Physics C, 28(05), 1750060.

[6]Li, D., & Ma, J. (2017). How the government's punishment and individual's sensitivity affect the rumor spreading in online social networks. Physica A: Statistical Mechanics and its Applications, 469, 284-292.

[7]Cheng, J. J., Liu, Y., Shen, B., & Yuan, W. G. (2013). An epidemic model of rumor diffusion in online social networks. The European Physical Journal B, 86(1), 29.

[8]Pathak, A. R., Mahajan, A., Singh, K., Patil, A., & Nair, A. (2020). Analysis of techniques for rumor detection in social media. Procedia Computer Science, 167, 2286-2296. https://doi.org/10.1016/j.procs.2020.03.281

[9]Eismann, K. (2021). Diffusion and persistence of false rumors in social media networks: implications of searchability on rumor self-correction on Twitter. Journal of Business Economics, 91(9), 1299-1329. https://doi.org/10.1007/s11573-020-01022-9

[10]Ahsan, M., Kumari, M., & Sharma, T. P. (2019). Rumors detection, verification and controlling mechanisms in online social networks: A survey. Online Social Networks and Media, 14, 100050. https://doi.org/10.1016/j.osnem.2019.100050

[11]Zubiaga, A., Liakata, M., Procter, R., Wong Sak Hoi, G., & Tolmie, P. (2016). Analysing how people orient to and spread rumours in social media by looking at conversational threads. PloS one, 11(3), e0150989. https://doi.org/10.1371%2Fjournal.pone.0150989

[12]Choi, D., Chun, S., Oh, H., Han, J., & Kwon, T. T. (2020). Rumor propagation is amplified by echo chambers in social media. Scientific reports, 10(1), 310. https://doi.org/10.1038/s41598-019-57272-3

[13]Carr, C. T., & Hayes, R. A. (2015). Social media: Defining, developing, and divining. Atlantic Journal of Communication, 23(1), 1-43. https://doi.org/10.1080/15456870.2015.972282

[14]Kaplan, A. M, & Haenlein, M. (2010). Χρήστες του κόσμου, ενωθείτε! Οι προκλήσεις και οι ευκαιρίες των social media. Business Horizons , 53 (1), 59–68

[15]Kaplan, A. M. (2015). Social media, the digital revolution, and the business of media. International Journal on Media Management, 17(4), 197-199.

[16]Agarwal, P., Al Aziz, R., & Zhuang, J. (2022). Interplay of rumor propagation and clarification on social media during crisis events-A game-theoretic approach. European Journal of Operational Research, 298(2), 714-733. https://doi.org/10.1016/j.ejor.2021.06.060

[17]Kane, G. C., Alavi, M., Labianca, G., & Borgatti, S. P. (2014). What's different about social media networks? A framework and research agenda. MIS quarterly, 38(1), 275-304.

[18]Kwon, K. H., Bang, C. C., Egnoto, M., & Raghav Rao, H. (2016). Social media rumors as improvised public opinion: semantic network analyses of twitter discourses during Korean saber rattling 2013. Asian Journal of Communication, 26(3), 201-222.

[19]Ghosh, M., & Thirugnanam, A. (2021). Introduction to Artificial Intelligence. In K. G. Srinivasa, Siddesh G. M., S. R. Mani Sekhar. Artificial Intelligence for Information Management: A Healthcare Perspective (pp.23-44). Springer.

[20]Zhang, C., & Lu, Y. (2021). Study on artificial intelligence: The state of the art and future prospects. Journal of Industrial Information Integration, 23, 100224. https://doi.org/10.1016/j.jii.2021.100224

[21]Padmaja, R. V. C., Narayana, L. S., Anga, L. G., & Bhansali, K. P. (2024). The rise of artificial intelligence: a concise review. IAES International Journal of Artificial Intelligence (IJ-AI), 13(2), 2252-8938. https://doi.org/10.11591/ijai.v13.i2.pp2226-2235

[22]Ross, B., Pilz, L., Cabrera, B., Brachten, F., Neubaum, G., & Stieglitz, S. (2019). Are social bots a real threat? An agent-based model of the spiral of silence to analyse the impact of manipulative actors in social networks', European Journal of Information Systems, 28, 394-412.

[23]Wang, X., Tajvidi, M., Lin, X., & Hajli, N. (2019). Towards an ethical and trustworthy social commerce community for brand value co-creation: A trust-commitment perspective', Journal of Business Ethics, 167(1),137-152.

[24]Kudugunta, S., & Ferrara, E. (2018). Deep neural networks for bot detection. Information Sciences, 467, 312-322.

[25]Zhao, J., Liu, H., Zhang, S., Qi, Y., Dong, H., Zhang, X., & Zhang, W. (2023). Advancements in rumor detection research based on bibliometrics and S-curve technology evolution theory. Sage Open, 13(4). https://doi.org/10.1177/21582440231217724

[26]Mahesh, B. (2018). Machine learning algorithms-a review. International Journal of Science and Research (IJSR), 9(1), 381-386. https://doi.org/10.21275/ART2020399

[27]Mahammad, Y., & Bakirova, L. (2021). Machine Learning Concepts and Applications. In CEUR Workshop Proceedings (pp. 257-262).

[28]Badillo, S., Banfai, B., Birzele, F., Davydov, I. I., Hutchinson, L., Kam-Thong, T., ... & Zhang, J. D. (2020). An introduction to machine learning. Clinical pharmacology & therapeutics, 107(4), 871-885.

[29]Jordan, M. I., & Mitchell, T. M. (2015). Machine learning: Trends, perspectives, and prospects. Science, 349(6245), 255-260.

[30]Janiesch, C., Zschech, P., & Heinrich, K. (2021). Machine learning and deep learning. Electronic Markets, 31(3), 685-695.

[31]Sharma, N., Sharma, R., & Jindal, N. (2021). Machine learning and deep learning applications-a vision. Global Transitions Proceedings, 2(1), 24-28. https://doi.org/10.1016/j.gltp.2021.01.004

[32]Habehh, H., & Gohel, S. (2021). Machine learning in healthcare. Current genomics, 22(4), 291. https://doi.org/10.2174%2F1389202922666210705124359

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

[34]Liu, Q., & Wu, Y. (2012). Supervised Learning. http://dx.doi.org/10.1007/978-1-4419-1428-6_451

[35]Naeem, S., Ali, A., Anam, S., & Ahmed, M. M. (2023). An unsupervised machine learning algorithms: Comprehensive review. International Journal of Computing and Digital Systems, 13(1), 911-921).

[36]Prasad, M. V., & Balakrishnan, R. (2022). Spatio-temporal association rule based deep annotation-free clustering (STAR-DAC) for unsupervised person re-identification. Pattern Recognition, 122, 108287.

[37]Kottmann, K., Huembeli, P., Lewenstein, M., & Acín, A. (2020). Unsupervised phase discovery with deep anomaly detection. Physical Review Letters, 125(17), 170603.

[38]Reddy, Y. C. A. P., Viswanath, P., & Reddy, B. E. (2018). Semi-supervised learning: A brief review. International Journal of Engineering & Technology, 7(1.8), 81-85. http://dx.doi.org/10.14419/ijet.v7i1.8.9977

[39]Grira, N., Crucianu, M., & Boujemaa, N. (2004). Unsupervised and semi-supervised clustering: a brief survey. A review of machine learning techniques for processing multimedia content, 1(2004), 9-16.

[40]Sivamayil, K., Rajasekar, E., Aljafari, B., Nikolovski, S., Vairavasundaram, S., & Vairavasundaram, I. (2023). A systematic study on reinforcement learning based applications. Energies, 16(3), 1512. https://doi.org/10.3390/en16031512

[41]Vapnik, V.N. (1989). *Statistical Learning Theory*. New York: Wiley-Interscience

[42]Cemiloglu, A., Zhu, L., Arslan, S., Xu, J., Yuan, X., Azarafza, M., & Derakhshani, R. (2023). Support vector machine (SVM) application for uniaxial compression strength (UCS) prediction: a case study for Maragheh limestone. Applied Sciences, 13(4), 2217. https://doi.org/10.3390/app13042217

[43]Guido, R., Ferrisi, S., Lofaro, D., & Conforti, D. (2024). An Overview on the Advancements of Support Vector Machine Models in Healthcare Applications: A Review. Information, 15(4), 235. https://doi.org/10.3390/info15040235

[44]Kang, S. (2021). K-nearest neighbor learning with graph neural networks. Mathematics, 9(8), 830. https://doi.org/10.3390/math9080830

[45]Vergni, L., & Todisco, F. (2023). A random forest machine learning approach for the identification and quantification of erosive events. Water, 15(12), 2225. https://doi.org/10.3390/w15122225

[46]Bisong, E. (2019). Logistic Regression. In: Building Machine Learning and Deep Learning Models on Google Cloud Platform. Apress, Berkeley, CA. https://doi.org/10.1007/978-1-4842-4470-8_20

[47]Alshboul, O., Shehadeh, A., Almasabha, G., & Almuflih, A. S. (2022). Extreme gradient boosting-based machine learning approach for green building cost prediction. Sustainability, 14(11), 6651. https://doi.org/10.3390/su14116651

[48]Sarker, I. H. (2021). Deep learning: a comprehensive overview on techniques, taxonomy, applications and research directions. SN computer science, 2(6), 420.

[49]Dong, S., Wang, P., & Abbas, K. (2021). A survey on deep learning and its applications. Computer Science Review, 40, 100379.

[50]Taye, M. M. (2023). Understanding of machine learning with deep learning: architectures, workflow, applications and future directions. Computers, 12(5), 91.

[51]Schmidhuber, J. (2015). Deep Learning in Neural Networks: An Overview. *Neural Netw., 61*, 85–117.

[52]Mathew, A., Amudha, P., & Sivakumari, S. (2021). Deep learning techniques: an overview. *Advanced Machine Learning Technologies and Applications: Proceedings of AMLTA 2020*, 599-608.

[53]LeCun, Y., Bengio, Y., & Hinton, G. (2015). Deep learning. *nature*, *521*(7553), 436-444.

[54]Shrestha, A., & Mahmood, A. (2019). Review of deep learning algorithms and architectures. IEEE access, 7, 53040-53065.

[55]Bengio, Y. (2009). Learning Deep Architectures for AI. Found. *Trends Mach. Learn., 2*, 1–127

[56]Pouyanfar, S., Sadiq, S., Yan, Y., Tian, H., Tao, Y., Reyes, M. P., ... & Iyengar, S. S. (2018). A survey on deep learning: Algorithms, techniques, and applications. ACM computing surveys (CSUR), 51(5), 1-36.

[57]Socher, R., Lin, C. C., Manning, C., & Ng, A. Y. (2011). Parsing natural scenes and natural language with recursive neural networks. In Proceedings of the 28th international conference on machine learning (ICML-11) (pp. 129-136).

[58]Sejan, M. A. S., Rahman, M. H., Aziz, M. A., Baik, J. I., You, Y. H., & Song, H. K. (2023). Graph Convolutional Network Design for Node Classification Accuracy Improvement. Mathematics, 11(17), 3680. https://doi.org/10.3390/math11173680

[59]Gao, R., Jiang, L., Zou, Z., Li, Y., & Hu, Y. (2024). A Graph Convolutional Network Based on Sentiment Support for Aspect-Level Sentiment Analysis. Applied Sciences, 14(7), 2738. https://doi.org/10.3390/app14072738

[60]Bingol, H., & Alatas, B. (2019, November). Rumor Detection in Social Media using machine learning methods. In 2019 1st International Informatics and Software Engineering Conference (UBMYK) (pp. 1-4). IEEE

[61]Saha, D., Das, A., Nath, T. C., Saha, S., & Das, R. (2022). Detection of Fake News and Rumors in Social Media Using Machine Learning Techniques With Semantic Attributes. Convergence of Deep Learning In Cyber-IoT Systems and Security, 85-97. https://doi.org/10.1002/9781119857686.ch4

[62]Wang, S., Li, Z., Wang, Y., & Zhang, Q. (2019). Machine learning methods to predict social media disaster rumor refuters. International journal of environmental research and public health, 16(8), 1452. https://doi.org/10.3390/ijerph16081452

[63]Al-Alshaqi, M., Rawat, D. B., & Liu, C. (2024). Ensemble Techniques for Robust Fake News Detection: Integrating Transformers, Natural Language Processing, and Machine Learning. Sensors, 24(18), 6062. https://doi.org/10.3390/s24186062

[64]Arowolo, M. O., Misra, S., & Ogundokun, R. O. (2023). A machine learning technique for detection of social media fake news. International Journal on Semantic Web and Information Systems (IJSWIS), 19(1), 1-25. http://dx.doi.org/10.4018/IJSWIS.326120

[65]Chang, Q., Li, X., & Duan, Z. (2024). A novel approach for rumor detection in social platforms: Memory-augmented transformer with graph convolutional networks. Knowledge-Based Systems, 292, 111625. https://doi.org/10.1016/j.knosys.2024.111625

[66]Zhang, W., Du, Y., Yoshida, T., & Wang, Q. (2018). DRI-RCNN: An approach to deceptive review identification using recurrent convolutional neural network. Information Processing & Management, 54(4), 576-592. https://doi.org/10.1016/j.ipm.2018.03.007

[67]Xu, Y., Wang, C., Dan, Z., Sun, S., & Dong, F. (2019). Deep recurrent neural network and data filtering for rumor detection on sina weibo. Symmetry, 11(11), 1408. https://doi.org/10.3390/sym11111408

[68]Alkhodair, S. A., Ding, S. H., Fung, B. C., & Liu, J. (2020). Detecting breaking news rumors of emerging topics in social media. Information Processing & Management, 57(2), 102018. https://doi.org/10.1016/j.ipm.2019.02.016

[69]Konkobo, P. M., Zhang, R., Huang, S., Minoungou, T. T., Ouedraogo, J. A., & Li, L. (2020, November). A deep learning model for early detection of fake news on social media. In 2020 7th International Conference on Behavioural and Social Computing (BESC) (pp. 1-6). IEEE.

[70]Choi, D., Oh, H., Chun, S., Kwon, T., & Han, J. (2022). Preventing rumor spread with deep learning. Expert Systems with Applications, 197, 116688. https://doi.org/10.1016/j.eswa.2022.116688

[71]Wang, J., Wang, X., & Yu, A. (2025). Tackling misinformation in mobile social networks a BERT-LSTM approach for enhancing digital literacy. Scientific Reports, 15(1), 1118.

[72]Xu, S., Liu, X., Ma, K., Dong, F., Riskhan, B., Xiang, S., & Bing, C. (2023). Rumor detection on social media using hierarchically aggregated feature via graph neural networks. Applied Intelligence, 53(3), 3136-3149.

[73]Song, C., Shu, K., & Wu, B. (2021). Temporally evolving graph neural network for fake news detection. Information Processing & Management, 58(6), 102712. https://doi.org/10.1016/j.ipm.2021.102712

[74] Python Software Foundation, “Python Wiki,” [Online]. Available: https://wiki.python.org/moin/.

[75] Project Jupyter, “The Jupyter Notebook,” [Online]. Available: https://jupyter-notebook.readthedocs.io/en/stable/notebook.html.

[76] Python Software Foundation, “os — Miscellaneous operating system interfaces,” [Online]. Available: https://docs.python.org/3/library/os.html.

[77] C. R. Harris, K. J. Millman, S. J. van der Walt, et al., “Array programming with NumPy,” *Nature*, vol. 585, pp. 357–362, Sep. 2020.

[78] T. Wolf, L. Debut, V. Sanh, et al., “Transformers: State-of-the-Art Natural Language Processing,” in *Proceedings of the 2020 Conference on Empirical Methods in Natural Language Processing: System Demonstrations*, pp. 38–45, Nov. 2020

[79] M. Abadi, A. Agarwal, P. Barham, et al., “TensorFlow: Large-Scale Machine Learning on Heterogeneous Systems,” 2015

[80] NVIDIA Corporation, “What Is Pandas?,” [Online]. Available: https://www.nvidia.com/en-eu/glossary/pandas-python/.

[81] IBM Corporation, “What is PyTorch?,” [Online]. Available: https://www.ibm.com/think/topics/pytorch.

[82] Streamlit Inc., “Streamlit Documentation,” [Online]. Available: https://docs.streamlit.io/.

Παραρτήματα

GitHub Repository: https://github.com/Angelos753/MyThesis