



UNIVERSITY OF
PATRAS
ΠΑΝΕΠΙΣΤΗΜΙΟ ΠΑΤΡΩΝ

Τμήμα Μηχανικών Ηλεκτρονικών Υπολογιστών και Πληροφορικής

Πολυτεχνική Σχολή

ΤΙΤΛΟΣ

Εντοπισμός Ανωμαλιών σε Δυναμικά Γραφήματα με Χρήση

Μεθόδων Βαθιάς Μηχανικής Μάθησης

Διπλωματική εργασία

Του

ΝΙΚΟΛΑΟΥ ΚΑΤΣΕΡΗ

Επιβλέπων: Τσίχλας Κωνσταντίνος

Αναπληρωτής Καθηγητής

Συν-επιβλέπων: Χριστόπουλος Κωνσταντίνος



Copyright ©- All rights reserved ΝΙΚΟΛΑΟΣ ΚΑΤΣΕΡΗΣ 2024

Με την επιφύλαξη παντός δικαιώματος.

Απαγορεύεται η αντιγραφή, αποθήκευση και διανομή της παρούσας εργασίας, εξ ολοκλήρου ή τμήματος αυτής, για εμπορικό σκοπό. Επιτρέπεται η ανατύπωση, αποθήκευση και διανομή για σκοπό μη κερδοσκοπικό, εκπαιδευτικής ή ερευνητικής φύσης, υπό την προϋπόθεση να αναφέρεται η πηγή προέλευσης και να διατηρείται το παρόν μήνυμα.

Το περιεχόμενο αυτής της εργασίας δεν απηχεί απαραίτητα τις απόψεις του Τμήματος, του Επιβλέποντα, ή της επιτροπής που την ενέκρινε.

Υπεύθυνη Δήλωση

Βεβαιώνω ότι είμαι συγγραφέας αυτής της διπλωματικής εργασίας, και ότι κάθε βοήθεια την οποία είχα για την προετοιμασία της είναι πλήρως αναγνωρισμένη και αναφέρεται στη διπλωματική εργασία. Επίσης έχω αναφέρει τις όποιες πηγές από τις οποίες έκανα χρήση δεδομένων, ιδεών ή λέξεων, είτε αυτές αναφέρονται ακριβώς είτε παραφρασμένες. Επίσης, βεβαιώνω ότι αυτή η διπλωματική εργασία προετοιμάστηκε από εμένα προσωπικά ειδικά για τις απαιτήσεις του προγράμματος σπουδών του Τμήματος Μηχανικών Ηλεκτρονικών Υπολογιστών και Πληροφορικής του Πανεπιστημίου Πατρών.

(Υπογραφή)

.....

ΝΙΚΟΛΑΟΣ ΚΑΤΣΕΡΗΣ

Περίληψη

Η τεχνητή νοημοσύνη και η μηχανική μάθηση αποτελούν πλέον τεχνολογίες αιχμής με ευρεία εφαρμογή στην καθημερινότητα. Ένας σημαντικός τομέας εφαρμογής τους είναι τα δυναμικά γραφήματα, τα οποία αποτυπώνουν σχέσεις μεταξύ οντοτήτων σε διαφορετικά χρονικά σημεία, με σκοπό την ανίχνευση ανωμαλιών, δηλαδή την αναγνώριση ασυνήθιστων προτύπων που αποκλίνουν από τους αναμενόμενους κανόνες. Η παρούσα εργασία εξετάζει τη χρήση του μοντέλου DyGED και των παραλλαγών του (DyGED-CT, DyGED-NL, DyGED-NA) για την ανίχνευση γεγονότων σε δυναμικά γραφήματα, αξιοποιώντας τόσο χρονικά όσο και δομικά χαρακτηριστικά. Αρχικά, τα μοντέλα αναλύονται και εκτελούνται σε συγκεκριμένα σύνολα δεδομένων που βασίζονται στις προτεινόμενες βάσεις των δημιουργών της πρότυπης υλοποίησης του DyGED, όπως αυτές αναφέρονται στο GitHub του έργου. Στη συνέχεια, αξιολογούνται σε δυναμικά δεδομένα που δημιουργήθηκαν μέσω της βιβλιοθήκης RDyn, για να διερευνηθεί η απόδοσή τους σε συνθήκες εξελισσόμενης δομής και ανωμαλιών, όπως οι συγχωνεύσεις και οι διαχωρισμοί κοινοτήτων. Η ανάλυση των αποτελεσμάτων με βάση τις μετρικές AUC Score και F1 Score δείχνει ότι το αρχικό μοντέλο DyGED κατατάσσεται σταθερά μεταξύ των δύο καλύτερων μοντέλων σε όλες τις περιπτώσεις, ξεχωρίζοντας για την αξιοπιστία και τη σταθερότητά του σε σύγκριση με τις παραλλαγές του. Παρά την καλή απόδοση των παραλλαγών, το βασικό μοντέλο αναδεικνύεται ως η πιο αξιόπιστη επιλογή στην ανίχνευση γεγονότων σε δυναμικά γραφήματα.

Λέξεις – Κλειδιά: Τεχνητή Νοημοσύνη, Μηχανική Μάθηση, Δυναμικά Γραφήματα, Εντοπισμός Ανωμαλιών.

Abstract

Artificial intelligence and machine learning are now cutting-edge technologies with widespread application in everyday life. One important area of application is dynamic graphs, which capture relationships between entities at different points in time in order to detect anomalies, i.e. to identify unusual patterns that deviate from the expected rules. This paper examines the use of the DyGED model and its variants (DyGED-CT, DyGED-NL, DyGED-NA) for event detection in dynamic graphs, exploiting both temporal and structural features. Initially, the models are analyzed and run on specific datasets based on the proposed databases of the authors of the DyGED model implementation, as listed on the project's GitHub. They are then evaluated on dynamic data created via the RDyn library to investigate their performance under conditions of evolving structure and anomalies such as mergers and community splits. Analysis of the results based on the AUC Score and F1 Score metrics shows that the original DyGED model consistently ranks among the two best models in all cases, standing out for its reliability and stability compared to its variants. Despite the good performance of the variants, the baseline model emerges as the most reliable choice in event detection in dynamic graphs.

Keywords: Artificial Intelligence, Machine Learning, Dynamic Graphs, Anomaly Detection.

Ευχαριστίες

Θα ήθελα να ευχαριστήσω τον επιβλέποντα καθηγητή μου, Τσίχλα Κωνσταντίνο, καθώς και τον συν-επιβλέποντα Χριστόπουλο Κωνσταντίνο, για την πολύτιμη συνεισφορά και βοήθεια τους.

Τέλος, οφείλω ένα μεγάλο ευχαριστώ στην οικογένειά μου και στους φίλους μου για τη συνεχή υποστήριξη και συμπαράστασή τους.

Πίνακας περιεχομένων

Επιβλέπων	1
Περίληψη	3
Abstract.....	3
Ευχαριστίες	4
1. Τεχνητή Νοημοσύνη	7
2. Μηχανική Μάθηση	10
2.1. Ορισμός και Αναδρομή.....	10
2.2. Κατηγορίες Μηχανικής Μάθησης	13
2.2.1. Εποπτευόμενη Μάθηση.....	13
2.2.2. Μη - Εποπτευόμενη Μάθηση	16
2.2.3. Ημι -Εποπτευόμενη Μάθηση.....	19
2.2.4. Ενισχυμένη Μάθηση	22
2.3. Προβλήματα και δυσλειτουργίες	25
3. Δυναμικά γραφήματα	28
3.1. Κατανεμημένα δυναμικά γραφήματα	30
4. Εντοπισμός Ανωμαλιών	31
4.1. Τι είναι.....	31
4.2. Είδη και Κατηγορίες Ανωμαλιών.....	32
4.3. Τρόπος εντοπισμού	36
4.3.1. Εντοπισμός ανωμαλιών στις ακμές γραφημάτων	40
4.4. Γιατί είναι σημαντική	42
5. Πειραματική διαδικασία	46
5.1. Εισαγωγή	46
5.2. Πρώτο Μέρος Πειραματικής Διαδικασίας	47
5.2.1. Χρησιμοποιούμενα Μοντέλα	48
5.2.2. Χρησιμοποιούμενα Σύνολα Δεδομένων.....	50
5.2.3. Χρησιμοποιούμενος Κώδικας	52
5.2.4. Πειραματικά Αποτελέσματα NYC Cab	55
5.2.5. Πειραματικά Αποτελέσματα Twitter Weather	56
5.2.6. Συγκριτικά Αποτελέσματα	57
5.2.7. Αξιολόγηση με F1 Score: Προκλήσεις και Αποτελέσματα	58
5.3. Δεύτερο Μέρος Πειραματικής Διαδικασίας.....	62
5.3.1. Εισαγωγή – Δημιουργία Νέα Σύνολα Δεδομένων	62
5.3.2. Χρησιμοποιούμενος Κώδικας	63
5.3.3. Πειραματικά Αποτελέσματα Συνόλων Δεδομένων Rdyn	64
6. Συμπεράσματα.....	66

6.1. Γενικά Συμπεράσματα	66
6.2. Περιορισμοί.....	67
6.3. Μελλοντικές Κατευθύνσεις	68
Βιβλιογραφία	68

Λίστα Πινάκων

Πίνακας 1: Χαρακτηριστικά και είδος γεγονότων προς εντοπισμό ανά σύνολο δεδομένων .	51
Πίνακας 2: Χρόνοι εκτέλεσης και Μετρική AUC - NYC Cab.....	55
Πίνακας 3: Χρόνοι εκτέλεσης και Μετρική AUC - Twitter Weather	56
Πίνακας 4: Μετρική AUC και F1 - Rdyn Βάσεις	64

Πίνακας Διαγραμμάτων

Διάγραμμα 1 - Ανίχνευση ανωμαλιών με μηχανική μάθηση	37
Διάγραμμα 2 - Εντοπισμός ανωμαλιών σε δεδομένα χρονοσειρών	38
Διάγραμμα 3 - Ανίχνευση ανωμαλιών με αλγόριθμο K-means	39
Διάγραμμα 4 - Ανίχνευση πιθανής απάτης	43

Πίνακας Εικόνων

Εικόνα 1: Περιβάλλον PyCharm.....	47
Εικόνα 2: Σύνολο δεδομένων NYC Cab – Μετρική AUC	55
Εικόνα 3: Σύνολο δεδομένων NYC Cab - Χρόνοι Εκτέλεσης.....	55
Εικόνα 4:Σύνολο δεδομένων Twitter Weather – Μετρική AUC.....	56
Εικόνα 5:Σύνολο δεδομένων Twitter Weather - Χρόνοι Εκτέλεσης	56
Εικόνα 6: Συγκριτικά αποτελέσματα μοντέλου DyGED_CT	57
Εικόνα 7: Συγκριτικά αποτελέσματα μοντέλου DyGED_NL	57
Εικόνα 8: Συγκριτικά αποτελέσματα μοντέλου DyGED_NA.....	57
Εικόνα 9: Συγκριτικά αποτελέσματα μοντέλου DyGED	57
Εικόνα 10:NYC Cab - Test Set (F1 Score)	59
Εικόνα 11:Twitter Weather - Test Set (F1 Score).....	59
Εικόνα 12:NYC Cab - Validation Loss (F1 Score)	59
Εικόνα 13:Twitter Weather - Validation Loss (F1 Score).....	59
Εικόνα 14: Σύγκριση Απόδοσης Μοντέλων: Scores για 500 Κόμβους	65
Εικόνα 15: Απώλεια Κατά την Εκπαίδευση (Validation Loss) για 500 Κόμβους	66

1. Τεχνητή Νοημοσύνη

Η Τεχνητή Νοημοσύνη (AI) αναφέρεται στην προσομοίωση διαδικασιών ανθρώπινης νοημοσύνης από μηχανές, ιδιαίτερα συστήματα υπολογιστών. Αυτές οι διαδικασίες περιλαμβάνουν τη μάθηση (την απόκτηση πληροφοριών και τους κανόνες για τη χρήση της πληροφορίας), τη συλλογιστική (χρήση κανόνων για την εξαγωγή κατά προσέγγιση ή οριστικών συμπερασμάτων) και την αυτοδιόρθωση. Η τεχνητή νοημοσύνη περιλαμβάνει μια ποικιλία τεχνικών και μεθόδων που επιτρέπουν στους υπολογιστές να εκτελούν εργασίες που συνήθως απαιτούν ανθρώπινη νοημοσύνη.

Ένας βασικός ορισμός της προέρχεται από τους Russell & Norvig (2020) και ορίζεται ευρέως ως η ικανότητα μιας μηχανής να μιμείται την ευφυή ανθρώπινη συμπεριφορά. Ο όρος «νοημοσύνη» σε αυτό το πλαίσιο περιλαμβάνει υπολογιστικά μοντέλα και αλγόριθμους ικανούς να αντιλαμβάνονται το περιβάλλον, να επεξεργάζονται δεδομένα, να λαμβάνουν αποφάσεις και να μαθαίνουν από τις εμπειρίες.

Υπάρχουν δύο βασικοί τύποι AI. Ο πρώτος - Narrow AI ή Weak AI αφορά συστήματα AI σχεδιασμένα να χειρίζονται μια συγκεκριμένη εργασία. Αυτά τα συστήματα είναι εξαιρετικά εξειδικευμένα και δεν διαθέτουν γενικές γνωστικές ικανότητες. Παραδείγματα περιλαμβάνουν βοηθούς φωνής όπως Siri και Alexa, συστήματα συστάσεων και αυτόνομα οχήματα. Ο δεύτερος τύπος General AI ή Strong AI σχετίζεται με συστήματα AI με γενικευμένες ανθρώπινες γνωστικές ικανότητες. Αυτά τα συστήματα μπορούν θεωρητικά να εκτελέσουν οποιοδήποτε διανοητικό έργο που μπορεί ένας άνθρωπος. Η γενική τεχνητή νοημοσύνη παραμένει μια θεωρητική κατασκευή, καθώς δεν έχει ακόμη υλοποιηθεί.

Ένας από τους βασικούς τομείς που εφαρμόζεται η Τεχνητή Νοημοσύνη είναι η Μηχανική Μάθηση που θα παρουσιαστεί ενδελεχώς στο επόμενο κεφάλαιο. Ένας δεύτερος τομέας είναι η επεξεργασία φυσικής γλώσσας (NLP) που αφορά την ικανότητα ενός προγράμματος υπολογιστή να κατανοεί, να ερμηνεύει και να δημιουργεί ανθρώπινη γλώσσα. Οι εφαρμογές περιλαμβάνουν μετάφραση γλώσσας, ανάλυση συναισθήματος και chatbots (Jurafsky & Martin, 2021). Ακόμη ένας τομέας είναι η υπολογιστική όραση - Computer Vision δηλαδή η ικανότητα των μηχανών να ερμηνεύουν και να λαμβάνουν αποφάσεις με βάση οπτικά δεδομένα από τον κόσμο. Οι εφαρμογές περιλαμβάνουν την αναγνώριση εικόνας, την αναγνώριση προσώπου και

την αυτόνομη οδήγηση (Szeliski, 2010). Ακόμη, χρησιμοποιείται στην ρομποτική και συγκεκριμένα στον σχεδιασμό και τη δημιουργία ρομπότ που μπορούν να εκτελούν εργασίες αυτόνομα ή ημιαυτόνομα. Η ρομποτική συνδυάζει την τεχνητή νοημοσύνη με τεχνολογίες μηχανικής και αισθητήρων για να κατασκευάσει μηχανές που μπορούν να αλληλεπιδράσουν με τον φυσικό κόσμο (Siciliano & Khatib, 2016).

Η έννοια της τεχνητής νοημοσύνης έχει αρχαίες ρίζες, με μύθους και ιστορίες για ευφυή αυτόματα που χρονολογούνται από την αρχαιότητα. Ωστόσο, το επίσημο πεδίο της έρευνας της τεχνητής νοημοσύνης καθιερώθηκε στα μέσα του 20ου αιώνα. Η θεμελιώδης εργασία του Alan Turing "Computing Machinery and Intelligence" (1950) έθεσε το ερώτημα εάν οι μηχανές μπορούν να σκεφτούν και εισήγαγε τη δοκιμή Turing ως μέτρο της νοημοσύνης των μηχανών. Το 1956, ο όρος «τεχνητή νοημοσύνη» επινοήθηκε στο συνέδριο Dartmouth, που διοργανώθηκε από τους John McCarthy, Marvin Minsky, Nathaniel Rochester και Claude Shannon. Αυτό το γεγονός σηματοδότησε την επίσημη γέννηση του AI ως πεδίου μελέτης (McCarthy et al., 2006). Λίγο αργότερα, στις δεκαετίες του 1960 και του 1970 παρατηρήθηκε σημαντική αισιοδοξία και χρηματοδότηση για την έρευνα της τεχνητής νοημοσύνης, που οδήγησε σε πρώιμες επιτυχίες στην επίλυση προβλημάτων και στη συμβολική λογική. Ωστόσο, η πρόοδος ήταν πιο αργή από ό,τι αναμενόταν, οδηγώντας σε περιόδους μειωμένης χρηματοδότησης και ενδιαφέροντος, γνωστές ως «χειμώνες AI». Ο 21ος αιώνας υπήρξε μάρτυρας μιας αναζωπύρωσης της τεχνητής νοημοσύνης, με γνώμονα την πρόοδο στην υπολογιστική ισχύ, τη διαθεσιμότητα μεγάλων συνόλων δεδομένων και τις ανακαλύψεις στη μηχανική μάθηση, ιδιαίτερα στη βαθιά μάθηση. Αυτή η αναζωπύρωση οδήγησε σε σημαντικές προόδους και ευρεία υιοθέτηση τεχνολογιών τεχνητής νοημοσύνης σε διάφορους κλάδους.

Η τεχνητή νοημοσύνη βρίσκει εφαρμογή σε διάφορα πεδία, στον τομέα της υγείας, στα χρηματοοικονομικά, στο μάρκετινγκ και γενικότερα τη διοίκηση επιχειρήσεων, στις πωλήσεις, σε αυτοματοποιημένα συστήματα που υπάρχουν σε διάφορους τομείς, στην επεξεργασία γλώσσας, εικόνων κ.α. που βρίσκουν εφαρμογή σε ηλεκτρονικές συσκευές π.χ. κινητά τηλέφωνα, κ.α.

Κατά τη χρήση της τεχνητής νοημοσύνης προκύπτουν διάφορα ηθικά και δεοντολογικά ζητήματα. Τα συστήματα AI μπορούν να διαιωνίσουν και να ενισχύσουν τις προκαταλήψεις που υπάρχουν στα δεδομένα εκπαίδευσης, οδηγώντας σε άδικα και

μεροληπτικά αποτελέσματα. Η διασφάλιση της δικαιοσύνης και η μείωση της προκατάληψης είναι μια σημαντική πρόκληση στην ανάπτυξη της τεχνητής νοημοσύνης (Barocas & Selbst, 2016). Επίσης, η χρήση της τεχνητής νοημοσύνης στη συλλογή και ανάλυση δεδομένων εγείρει σημαντικά ζητήματα απορρήτου. Η διασφάλιση ότι τα προσωπικά δεδομένα χρησιμοποιούνται δεοντολογικά και με ασφάλεια είναι ζωτικής σημασίας για τη διατήρηση της εμπιστοσύνης του κοινού (Shokri et al., 2017). Ακόμη, η αυτοματοποίηση των εργασιών μέσω ΑΙ μπορεί να οδηγήσει σε μετατόπιση θέσεων εργασίας και οικονομική αναστάτωση. Η αντιμετώπιση του αντίκτυπου στο εργατικό δυναμικό μέσω της επανεκπαίδευσης και των μέτρων πολιτικής είναι απαραίτητη (Bessen, 2019). Τέλος, τα συστήματα ΑΙ μπορεί να είναι πολύπλοκα και αδιαφανή, καθιστώντας δύσκολη την κατανόηση του τρόπου λήψης αποφάσεων. Η διασφάλιση της διαφάνειας και της λογοδοσίας στις διαδικασίες λήψης αποφάσεων ΑΙ είναι κρίσιμης σημασίας για την εμπιστοσύνη και τη διακυβέρνηση (Vaughan et al., 2018).

Όπως φαίνεται από τα παραπάνω, η Τεχνητή Νοημοσύνη αντιπροσωπεύει ένα μετασχηματιστικό πεδίο που έχει τη δυνατότητα να φέρει επανάσταση σε διάφορες πτυχές της κοινωνίας. Από την ενίσχυση της υγειονομικής περίθαλψης και της χρηματοδότησης μέχρι την ενεργοποίηση αυτόνομων συστημάτων και επεξεργασίας φυσικής γλώσσας, η τεχνητή νοημοσύνη προσφέρει σημαντικά οφέλη. Ωστόσο, η αντιμετώπιση των ηθικών και κοινωνικών προκλήσεων που σχετίζονται με την τεχνητή νοημοσύνη είναι απαραίτητη για τη διασφάλιση της υπεύθυνης και δίκαιης ανάπτυξής της.

2. Μηχανική Μάθηση

2.1. Ορισμός και Αναδρομή

Η μηχανική μάθηση είναι ένα κομμάτι της τεχνητής νοημοσύνης (AI) που ασχολείται με τη δημιουργία αλγορίθμων και στατιστικών μοντέλων, ώστε οι υπολογιστές να μπορούν να εκτελούν εργασίες χωρίς να χρειάζονται συγκεκριμένες εντολές για κάθε βήμα. Αντί να ακολουθούν ρητές οδηγίες, αυτά τα συστήματα αναλύουν δεδομένα, ανακαλύπτουν μοτίβα και εξάγουν συμπεράσματα από αυτά. Ο στόχος της μηχανικής μάθησης είναι να κατασκευάσει συστήματα που βελτιώνουν την απόδοσή τους σε μια συγκεκριμένη εργασία με την πάροδο του χρόνου, μαθαίνοντας από την εμπειρία.

Η μηχανική μάθηση ορίζεται ως «το πεδίο μελέτης που δίνει στους υπολογιστές τη δυνατότητα να μαθαίνουν χωρίς να είναι ρητά προγραμματισμένοι» (Samuel, 1959). Αυτό το πεδίο περιλαμβάνει διάφορες τεχνικές, συμπεριλαμβανομένης της εποπτευόμενης μάθησης, της μάθησης χωρίς επίβλεψη, της ημι-εποπτευόμενης μάθησης και της ενισχυτικής μάθησης. Αυτές οι προσεγγίσεις επιτρέπουν στις μηχανές να προσαρμοστούν σε νέα δεδομένα και να λαμβάνουν αποφάσεις ή προβλέψεις βάσει αυτών. Οι βασικές τεχνικές είναι η εποπτευόμενη, η μη-εποπτευόμενη, η ημι-εποπτευόμενη και η ενισχυτική μάθηση, οι οποίες θα αναφερθούν στη συνέχεια.

Η μηχανική εκμάθηση έχει ένα ευρύ φάσμα εφαρμογών σε διάφορους τομείς. Στην υγειονομική περίθαλψη, χρησιμοποιείται για προγνωστικές αναλύσεις, διάγνωση ασθενειών και εξατομικευμένα σχέδια θεραπείας (Esteva et al., 2019). Στα οικονομικά, οι αλγόριθμοι μηχανικής μάθησης βοηθούν στον εντοπισμό απάτης, στις αλγοριθμικές συναλλαγές και στη διαχείριση κινδύνου (Brennan & Lo, 2019). Επιπλέον, η μηχανική εκμάθηση εξουσιοδοτεί πολλές καθημερινές τεχνολογίες, όπως συστήματα συστάσεων (π.χ. Netflix, Amazon), αναγνώριση ομιλίας (π.χ. Siri, Βοηθός Google) και αναγνώριση εικόνας (π.χ. συστήματα αναγνώρισης προσώπου).

Παρά τις επιτυχίες της, η μηχανική μάθηση αντιμετωπίζει αρκετές προκλήσεις, συμπεριλαμβανομένης της ανάγκης για μεγάλους όγκους επισημασμένων δεδομένων, της ερμηνείας των μοντέλων και της πιθανότητας μεροληψίας στους αλγόριθμους (Mitchell et al., 2019). Η μελλοντική έρευνα επικεντρώνεται στη βελτίωση της αποτελεσματικότητας των δεδομένων, στην ανάπτυξη πιο διαφανών μοντέλων και στη δημιουργία δικαιότερων αλγορίθμων.

Η ιστορία της μηχανικής μάθησης μπορεί να ανιχνευθεί στα μέσα του 20ου αιώνα, με τις ρίζες της να πηγαινούν συνάμα με την ανάπτυξη της τεχνητής νοημοσύνης. Η διαδρομή της ξεκινάει τη δεκαετία του 1950 όπου τίθεται η εννοιολογική της βάση. Ο Άλαν Τούρινγκ, στη θεμελιώδη εργασία του "Υπολογιστική Μηχανή και Νοημοσύνη" (1950), πρότεινε την ιδέα μιας μηχανής που θα μπορούσε να προσομοιώσει οποιαδήποτε εργασία ανθρώπινης νοημοσύνης, μια έννοια που τώρα είναι γνωστή ως Δοκιμή Τούρινγκ. Αυτή η περίοδος είδε επίσης την ανάπτυξη του πρώτου μοντέλου νευρωνικών δικτύων, του Perceptron, από τον Frank Rosenblatt το 1958. Το Perceptron ήταν ένας δυαδικός ταξινομητής και σηματοδότησε έναν από τους πρώτους αλγόριθμους ικανούς να μαθαίνουν από δεδομένα (Rosenblatt, 1958).

Κατά τη διάρκεια των δεκαετιών του 1960 και του 1970, οι ερευνητές άρχισαν να εξερευνούν περαιτέρω τις δυνατότητες της μηχανικής μάθησης. Η εργασία του Άρθουρ Σάμουελ για την ανάπτυξη ενός προγράμματος που παίζει πούλια απέδειξε τη σκοπιμότητα των προγραμμάτων αυτομάθησης. Ο Samuel (1959) όρισε τη μηχανική μάθηση ως «το πεδίο μελέτης που δίνει στους υπολογιστές τη δυνατότητα να μαθαίνουν χωρίς να είναι ρητά προγραμματισμένοι». Ωστόσο, αυτή η εποχή αντιμετώπισε και προκλήσεις, όπως οι περιορισμοί του Perceptron, τους οποίους επισήμαναν οι Minsky και Papert στο βιβλίο τους "Perceptrons" (1969). Έδειξαν ότι το Perceptron δεν μπορούσε να λύσει μη γραμμικά προβλήματα, οδηγώντας σε προσωρινή πτώση στην έρευνα των νευρωνικών δικτύων.

Η δεκαετία του 1980 έγινε μάρτυρας μιας αναζωπύρωσης της έρευνας μηχανικής μάθησης, με γνώμονα την ανάπτυξη πιο εξελιγμένων αλγορίθμων και την εμφάνιση αυξημένης υπολογιστικής ισχύος. Αυτή η περίοδος είδε την εισαγωγή του αλγόριθμου backpropagation, ο οποίος αντιμετώπισε τους περιορισμούς των προηγούμενων νευρωνικών δικτύων επιτρέποντας σε δίκτυα πολλαπλών επιπέδων να προσαρμόσουν τα βάρη τους και να μάθουν πιο περίπλοκα μοτίβα (Rumelhart, Hinton, & Williams, 1986). Επιπλέον, το πεδίο της μηχανικής μάθησης άρχισε να διαφοροποιείται με την εισαγωγή των δέντρων αποφάσεων, των μηχανών διανυσμάτων υποστήριξης (SVM) και των δικτύων Bayes.

Η δεκαετία του 1990 χαρακτηρίστηκε από σημαντικές θεωρητικές και πρακτικές προόδους στη μηχανική μάθηση. Η καθιέρωση πιθανοτικών μοντέλων, όπως τα Hidden Markov Models (HMMs) και Gaussian Mixture Models (GMMs), παρείχαν ισχυρά

πλαίσια για την αντιμετώπιση της αβεβαιότητας και την πραγματοποίηση προβλέψεων με βάση ελλιπή δεδομένα (Bishop, 2006). Επιπλέον, η ανάπτυξη μεθόδων συνόλου, όπως το boosting και το bagging, ενίσχυσε την ακρίβεια και την αξιοπιστία των μοντέλων μηχανικής μάθησης (Breiman, 1996; Freund & Schapire, 1997).

Η αλλαγή της χιλιετίας σηματοδότησε την αρχή μιας εποχής όπου κυριαρχούσαν τα μεγάλα δεδομένα και η βαθιά μάθηση. Με την εκθετική ανάπτυξη των ψηφιακών δεδομένων και τις προόδους στην υπολογιστική ισχύ, τα μοντέλα μηχανικής μάθησης έγιναν όλο και πιο ικανά να χειρίζονται σύνολα δεδομένων μεγάλης κλίμακας. Ο Geoffrey Hinton και οι συνεργάτες του έκαναν πρωτοποριακές συνεισφορές στη βαθιά μάθηση, ένα υποσύνολο της μηχανικής μάθησης που εστιάζει σε νευρωνικά δίκτυα με πολλά επίπεδα (Hinton, Osindero, & Teh, 2006). Αυτά τα μοντέλα βαθιάς μάθησης έχουν φέρει επανάσταση σε διάφορους τομείς, συμπεριλαμβανομένης της όρασης υπολογιστή, της επεξεργασίας φυσικής γλώσσας και της αναγνώρισης ομιλίας.

Από τα πρώτα εννοιολογικά θεμέλιά της έως την τρέχουσα κατάστασή της ως ακρογωνιαίο λίθο της σύγχρονης τεχνολογίας, η μηχανική μάθηση έχει υποστεί σημαντικούς μετασχηματισμούς. Κάθε εποχή ανάπτυξης έχει βασιστεί στις προηγούμενες εξελίξεις, οδηγώντας στους εξελιγμένους αλγόριθμους και εφαρμογές που βλέπουμε σήμερα. Καθώς η μηχανική μάθηση συνεχίζει να εξελίσσεται, υπόσχεται να οδηγήσει σε περαιτέρω καινοτομίες σε πολλούς τομείς, από την υγειονομική περίθαλψη έως τα αυτόνομα συστήματα.

2.2. Κατηγορίες Μηχανικής Μάθησης

2.2.1. Εποπτευόμενη Μάθηση

Η εποπτευόμενη μάθηση είναι μια βασική μέθοδος στη μηχανική μάθηση, όπου ο στόχος είναι η εκμάθηση μιας συνάρτησης από δεδομένα εκπαίδευσης με ετικέτα. Πρόκειται για αλγόριθμους που χρησιμοποιούνται για να κάνουν προβλέψεις για νέα, αόρατα δεδομένα. Το θεμελιώδες χαρακτηριστικό της εποπτευόμενης μάθησης είναι η παρουσία μιας μεταβλητής στόχου ή αποτελέσματος, την οποία το μοντέλο στοχεύει να προβλέψει (Montgomery, Peck, & Vining, 2012).

Βασικό συστατικό της είναι τα δεδομένα εκπαίδευσης. Το σύνολο δεδομένων που χρησιμοποιείται στην εποπτευόμενη μάθηση αποτελείται από ζεύγη εισόδου-εξόδου. Κάθε είσοδος συνδέεται με μια αντίστοιχη έξοδο, η οποία χρησιμεύει ως «εποπτεία» για τη μαθησιακή διαδικασία. Για παράδειγμα, σε ένα σύνολο δεδομένων που χρησιμοποιείται για την ταξινόμηση των μηνυμάτων ηλεκτρονικού ταχυδρομείου ως ανεπιθύμητης ή μη ανεπιθύμητης αλληλογραφίας, κάθε μήνυμα ηλεκτρονικού ταχυδρομείου (εισαγωγή) θα χαρακτηρίζεται είτε ως ανεπιθύμητο είτε ως μη ανεπιθύμητο (έξοδος). Ακόμη, χρησιμοποιεί μια ποικιλία αλγορίθμων για την εκτέλεσή της, ο καθένας από τους οποίους έχει δυνατά και τα αδύνατα σημεία. Η επιλογή του αλγορίθμου εξαρτάται συχνά από τη φύση των δεδομένων και το συγκεκριμένο πρόβλημα που αντιμετωπίζεται.

Υπάρχουν δύο κύριοι τύποι εποπτευόμενων μαθησιακών προβλημάτων, η ταξινόμηση και η παλινδρόμηση. Στις εργασίες ταξινόμησης, ο στόχος είναι η πρόβλεψη μιας διακριτής ετικέτας. Τα κοινά παραδείγματα περιλαμβάνουν τον εντοπισμό ανεπιθύμητων μηνυμάτων ηλεκτρονικού ταχυδρομείου, τη διάγνωση ασθενειών από ιατρικά δεδομένα και την αναγνώριση χειρόγραφων ψηφίων. Αλγόριθμοι όπως η λογιστική παλινδρόμηση, οι μηχανές διανυσμάτων υποστήριξης (SVM), τα δέντρα αποφάσεων και τα νευρωνικά δίκτυα χρησιμοποιούνται συχνά για εργασίες ταξινόμησης (Bishop, 2006). Στις εργασίες παλινδρόμησης, ο στόχος είναι η πρόβλεψη μιας συνεχούς παραγωγής. Στα παραδείγματα περιλαμβάνονται η πρόβλεψη των τιμών των κατοικιών με βάση χαρακτηριστικά όπως η τοποθεσία και το μέγεθος ή η πρόβλεψη των τιμών των μετοχών. Η γραμμική παλινδρόμηση, η παλινδρόμηση κορυφογραμμής και τα νευρωνικά δίκτυα χρησιμοποιούνται συνήθως για προβλήματα παλινδρόμησης (Montgomery et al., 2012).

Οι πιο βασικοί αλγόριθμοι είναι οι εξής:

- Γραμμική παλινδρόμηση: Ένας απλός αλλά ισχυρός αλγόριθμος για εργασίες παλινδρόμησης που προϋποθέτει μια γραμμική σχέση μεταξύ των χαρακτηριστικών εισόδου και της μεταβλητής εξόδου. Εκτιμά τους συντελεστές της γραμμικής εξίσωσης, που περιλαμβάνει τα χαρακτηριστικά εισόδου που προβλέπουν καλύτερα τη μεταβλητή εξόδου (Montgomery et al., 2012).
- Λογιστική παλινδρόμηση: Παρά το όνομά της, η λογιστική παλινδρόμηση χρησιμοποιείται για προβλήματα δυαδικής ταξινόμησης. Μοντελοποιεί την πιθανότητα ότι μια δεδομένη είσοδος ανήκει σε μια συγκεκριμένη κλάση και εξάγει τιμές μεταξύ 0 και 1 χρησιμοποιώντας τη λογιστική συνάρτηση (Hosmer, Lemeshow, & Sturdivant, 2013).
- Δένδρα αποφάσεων: Αυτά τα μοντέλα χρησιμοποιούν μια δομή που μοιάζει με δέντρο για τη λήψη αποφάσεων με βάση τα χαρακτηριστικά εισόδου. Κάθε εσωτερικός κόμβος αντιπροσωπεύει έναν κανόνα απόφασης και κάθε κόμβος φύλλου αντιπροσωπεύει μια έξοδο. Τα δένδρα αποφάσεων είναι διαισθητικά και μπορούν να χειριστούν εργασίες ταξινόμησης και παλινδρόμησης (Breiman et al., 1984).
- Υποστήριξη διανυσματικών μηχανών (SVM): Τα SVM είναι ισχυροί αλγόριθμοι ταξινόμησης που βρίσκουν το υπερεπίπεδο που διαχωρίζει καλύτερα τα δεδομένα σε διαφορετικές κλάσεις. Μπορούν να χειριστούν χώρους υψηλών διαστάσεων και είναι αποτελεσματικά όταν ο αριθμός των διαστάσεων υπερβαίνει τον αριθμό των δειγμάτων (Cortes & Vapnik, 1995).
- Νευρωνικά δίκτυα: Εμπνευσμένα από τον ανθρώπινο εγκέφαλο, τα νευρωνικά δίκτυα αποτελούνται από στρώματα διασυνδεδεμένων κόμβων (νευρώνες) που επεξεργάζονται τα δεδομένα εισόδου. Είναι εξαιρετικά ευέλικτοι και μπορούν να μοντελοποιήσουν σύνθετες σχέσεις, καθιστώντας τις κατάλληλες τόσο για εργασίες ταξινόμησης όσο και για εργασίες παλινδρόμησης (Goodfellow, Bengio, & Courville, 2016).

Η αξιολόγηση της απόδοσης των εποπτευόμενων μοντέλων μάθησης είναι ζωτικής σημασίας για την κατανόηση της αποτελεσματικότητάς τους. Οι κοινές μετρήσεις περιλαμβάνουν:

- Ακρίβεια: Το ποσοστό των σωστά ταξινομημένων παρουσιών στο σύνολο δεδομένων. Αν και χρήσιμο, μπορεί να μην είναι το καλύτερο μέτρο για μη ισορροπημένα σύνολα δεδομένων (Fawcett, 2006).
- Ακρίβεια και ανάκληση: Ακρίβεια είναι ο αριθμός των αληθινών θετικών αποτελεσμάτων διαιρεμένος με τον αριθμό των θετικών αποτελεσμάτων που προβλέπονται από τον ταξινομητή, ενώ η ανάκληση είναι ο αριθμός των αληθινών θετικών αποτελεσμάτων διαιρεμένος με τον αριθμό όλων των δειγμάτων που θα έπρεπε να έχουν αναγνωρισθεί ως θετικά. Η βαθμολογία F1, η οποία είναι η αρμονική μέση ακρίβεια και ανάκληση, χρησιμοποιείται επίσης συνήθως (Powers, 2011).
- Μέσο τετράγωνο σφάλμα (MSE): Για εργασίες παλινδρόμησης, το MSE μετρά τη μέση τετραγωνική διαφορά μεταξύ των παρατηρούμενων και των προβλεπόμενων τιμών. Είναι ευαίσθητο σε ακραίες τιμές και παρέχει ένα μέτρο της ακρίβειας του μοντέλου στην πρόβλεψη συνεχών αποτελεσμάτων (Hastie, Tibshirani, & Friedman, 2009).

Η εποπτευόμενη μάθηση έχει ένα ευρύ φάσμα εφαρμογών σε διάφορους τομείς όπως στην υγειονομική περίθαλψη όπου γίνεται πρόβλεψη αποτελεσμάτων ασθένειας, διάγνωση καταστάσεων με βάση ιατρικές εικόνες και εξατομίκευση σχεδίων θεραπείας (Esteva et al., 2019). Επίσης στο χρηματιστήριο όπου πραγματοποιούνται πιστωτικές βαθμολογίες, ανίχνευση απάτης και αλγοριθμική διαπραγμάτευση (Brennan & Lo, 2019). Ακόμη χρησιμοποιούνται για λόγους μάρκετινγκ όπως τμηματοποίηση πελατών, πρόβλεψη απόκλισης και στοχευμένη διαφήμιση – προσωποποιημένο μάρκετινγκ (Hastie et al., 2009).

2.2.2. Μη - Εποπτευόμενη Μάθηση

Η μάθηση χωρίς επίβλεψη είναι μια κατηγορία τεχνικών μηχανικής εκμάθησης που χρησιμοποιούνται για την εξαγωγή συμπερασμάτων από σύνολα δεδομένων που αποτελούνται από δεδομένα εισόδου χωρίς αποκρίσεις με ετικέτα. Σε αντίθεση με την εποπτευόμενη μάθηση, η οποία χρησιμοποιεί δεδομένα με ετικέτα για την εκπαίδευση μοντέλων, η μη εποπτευόμενη μάθηση λειτουργεί από μόνη της για να ανακαλύψει την υποκείμενη δομή των δεδομένων. Ο πρωταρχικός στόχος είναι να εξερευνήσει τα δεδομένα και να βρει κρυφά μοτίβα ή ομαδοποιήσεις χωρίς προηγούμενη γνώση των αποτελεσμάτων.

Όπως και στην εποπτευόμενη μάθηση, τα βασικά συστατικά της περιλαμβάνουν τα δεδομένα και τον αλγόριθμο. Σε μάθηση χωρίς επίβλεψη, το σύνολο δεδομένων περιέχει μόνο μεταβλητές εισόδου χωρίς αντίστοιχες ετικέτες εξόδου. Αυτά τα δεδομένα είναι συχνά τεράστια και πολύπλοκα, καθιστώντας τη χειροκίνητη ανάλυση ανέφικτη. Οι αλγόριθμοι που χρησιμοποιούνται στην μάθηση χωρίς επίβλεψη έχουν σχεδιαστεί για να βρίσκουν μοτίβα ή δομές απευθείας από τα δεδομένα χωρίς καθοδήγηση.

Υπάρχουν δύο κύριοι τύποι μαθησιακών προβλημάτων χωρίς επίβλεψη. Ο πρώτος είναι η ομαδοποίηση που είναι η διαίρεση του συνόλου δεδομένων σε ομάδες, ή συμπλέγματα, παρόμοιων στοιχείων. Κάθε σύμπλεγμα περιέχει στοιχεία που μοιάζουν περισσότερο μεταξύ τους παρά με στοιχεία σε άλλα συμπλέγματα. Οι κοινοί αλγόριθμοι περιλαμβάνουν το k-means, την ιεραρχική ομαδοποίηση και το DBSCAN (Density-Based Spatial Clustering of Applications with Noise) (Jain, 2010). Ο δεύτερος τύπος είναι η μείωση διαστάσεων. Η διαδικασία μείωσης του αριθμού των υπό εξέταση τυχαίων μεταβλητών, με τη λήψη ενός συνόλου κύριων μεταβλητών. Οι τεχνικές περιλαμβάνουν την Ανάλυση Κύριων Στοιχείων (PCA), την Ενσωμάτωση Στοχαστικού Γείτονα t-Distributed Stochastic Neighbor (t-SNE) και τη Γραμμική Διακριτική Ανάλυση (LDA) (Jolliffe, 2011).

Οι πιο βασικοί αλγόριθμοι της μη εποπτευόμενης μάθησης είναι οι εξής:

- Ομαδοποίηση K-means: Αυτός ο αλγόριθμος διαμερίζει το σύνολο δεδομένων σε ομάδες K αναθέτοντας επαναληπτικά κάθε σημείο δεδομένων στο σύμπλεγμα με τον πλησιέστερο μέσο όρο. Στη συνέχεια, τα μέσα ενημερώνονται με βάση τα σημεία που έχουν εκχωρηθεί στις συστάδες και η διαδικασία επαναλαμβάνεται μέχρι τη σύγκλιση (MacQueen, 1967).
- Ιεραρχική ομαδοποίηση: Αυτή η μέθοδος δημιουργεί μια ιεραρχία συστάδων χρησιμοποιώντας είτε μια προσέγγιση από κάτω προς τα πάνω (συγκεντρωτική) είτε από πάνω προς τα κάτω (διαιρετική). Στη συγκεντρωτική ομαδοποίηση, κάθε σημείο δεδομένων ξεκινά ως το δικό του σύμπλεγμα και τα ζεύγη συστάδων συγχωνεύονται καθώς κάποιος κινείται προς τα πάνω στην ιεραρχία. Στη διαιρετική ομαδοποίηση, όλα τα σημεία δεδομένων ξεκινούν σε ένα σύμπλεγμα και οι διαχωρισμοί εκτελούνται αναδρομικά καθώς κάποιος κινείται προς τα κάτω στην ιεραρχία (Murtagh & Contreras, 2012).
- DBSCAN: Αυτός ο αλγόριθμος ομαδοποίησης που βασίζεται στην πυκνότητα ομαδοποιεί σημεία που είναι στενά συσχετισμένα μεταξύ τους, ενώ επισημαίνει σημεία που βρίσκονται σε περιοχές χαμηλής πυκνότητας ως ακραία σημεία. Είναι ιδιαίτερα αποτελεσματικό στην ανακάλυψη συστάδων αυθαίρετων σχημάτων (Ester, Kriegel, Sander, & Xu, 1996).
- Ανάλυση κύριας συνιστώσας (PCA): Η PCA είναι μια τεχνική μείωσης διαστάσεων που μετατρέπει τα δεδομένα σε ένα νέο σύστημα συντεταγμένων όπου οι μεγαλύτερες αποκλίσεις από οποιαδήποτε προβολή των δεδομένων βρίσκονται στις πρώτες συντεταγμένες (τα κύρια συστατικά) και οι ελάχιστες διακυμάνσεις στις τελευταίες συντεταγμένες (Jolliffe, 2011).
- t-SNE: Η t-Distributed Stochastic Neighbor Embedding είναι μια τεχνική μείωσης διαστάσεων που είναι ιδιαίτερα κατάλληλη για την οπτικοποίηση συνόλων δεδομένων υψηλών διαστάσεων. Ελαχιστοποιεί την απόκλιση μεταξύ δύο κατανομών: μια κατανομή που μετρά ομοιότητες κατά ζεύγη των αντικειμένων εισόδου στον χώρο υψηλής διάστασης και μια παρόμοια κατανομή στο χώρο χαμηλής διάστασης (van der Maaten & Hinton, 2008).

Η αξιολόγηση μοντέλων μάθησης χωρίς επίβλεψη μπορεί να είναι δύσκολη λόγω της απουσίας ετικετών βασικής αλήθειας. Ωστόσο, διάφορες μετρήσεις μπορούν να βοηθήσουν στην αξιολόγηση της ποιότητας της ομαδοποίησης:

- Σκορ Silhouette: Μετρά πόσο παρόμοιο είναι ένα σημείο με το δικό του σύμπλεγμα σε σύγκριση με άλλα συμπλέγματα. Μια υψηλότερη βαθμολογία σιλουέτας υποδηλώνει καλύτερα καθορισμένα συμπλέγματα (Rousseeuw, 1987).
- Αδράνεια: Χρησιμοποιείται στην ομαδοποίηση k-means για τη μέτρηση του αθροίσματος των τετραγωνικών αποστάσεων των δειγμάτων στο πλησιέστερο κέντρο συστάδας τους. Η χαμηλότερη αδράνεια υποδηλώνει πιο συμπαγή συμπλέγματα.
- Δείκτης Davies-Bouldin: Αντιπροσωπεύει τη μέση αναλογία ομοιότητας κάθε συστάδας με την πιο παρόμοια συστάδα. Οι χαμηλότερες τιμές δείχνουν καλύτερη ομαδοποίηση (Davies & Bouldin, 1979).

Η μάθηση χωρίς επίβλεψη έχει ένα ευρύ φάσμα εφαρμογών σε διάφορους τομείς όπως η τμηματοποίηση πελατών όπου γίνεται προσδιορισμός διακριτών ομάδων πελατών για στοχευμένες στρατηγικές μάρκετινγκ (Ton & Raman, 2010). Ακόμη στην ανίχνευση ανωμαλιών - ανίχνευση ασυνήθιστων μοτίβων που δεν συμμορφώνονται με την αναμενόμενη συμπεριφορά, χρήσιμο για τον εντοπισμό απάτης και την ασφάλεια δικτύου (Chandola, Banerjee, & Kumar, 2009). Επίσης χρησιμοποιείται για την ανακάλυψη προτύπων γονιδιακής έκφρασης και ταξινόμηση τύπων κυττάρων (Wang, Zhu, & Pierson, 2018).

Η μάθηση χωρίς επίβλεψη είναι ένα ισχυρό εργαλείο στη μηχανική μάθηση, που επιτρέπει την ανακάλυψη προτύπων και δομών σε δεδομένα χωρίς ετικέτα. Οι τεχνικές του είναι απαραίτητες για την εξερεύνηση πολύπλοκων συνόλων δεδομένων και έχουν σημαντικές εφαρμογές σε διάφορους τομείς. Καθώς τα δεδομένα συνεχίζουν να αυξάνονται σε πολυπλοκότητα και όγκο, η μάθηση χωρίς επίβλεψη θα διαδραματίζει όλο και πιο σημαντικό ρόλο στην εξαγωγή ουσιαστικών γνώσεων.

2.2.3. Ημι-Εποπτευόμενη Μάθηση

Η ημι-εποπτευόμενη μάθηση είναι μια υβριδική προσέγγιση στη μηχανική μάθηση που χρησιμοποιεί δεδομένα τόσο με ετικέτα όσο και χωρίς ετικέτα για εκπαίδευση. Αυτή η μέθοδος βρίσκεται μεταξύ της εποπτευόμενης μάθησης, η οποία βασίζεται εξ ολοκλήρου σε δεδομένα με ετικέτα, και της μάθησης χωρίς επίβλεψη, η οποία λειτουργεί αποκλειστικά με δεδομένα χωρίς ετικέτα. Ο πρωταρχικός στόχος της ημι-εποπτευόμενης μάθησης είναι η βελτίωση της μαθησιακής ακρίβειας αξιοποιώντας τα συχνά άφθονα δεδομένα χωρίς ετικέτα μαζί με τα τυπικά σπάνια δεδομένα με ετικέτα.

Τα βασικά συστατικά της είναι τα δεδομένα με και χωρίς ετικέτα. Τα πρώτα είναι ένα υποσύνολο του συνόλου δεδομένων όπου κάθε στιγμιότυπο συσχετίζεται με μια γνωστή ετικέτα. Τα δεδομένα με ετικέτα παρέχουν άμεση επίβλεψη στον αλγόριθμο εκμάθησης. Το μεγαλύτερο μέρος του συνόλου δεδομένων είναι χωρίς ετικέτες. Ο αλγόριθμος εκμάθησης χρησιμοποιεί αυτά τα δεδομένα για να συμπεράνει τη δομή και την κατανομή των δεδομένων, τα οποία μπορούν να βοηθήσουν στη βελτίωση της απόδοσης του μοντέλου. Τα είδη της ημι-εποπτευόμενης μάθησης είναι τα εξής:

- **Αυτο-εκπαίδευση:** Στην αυτοεκπαίδευση, το μοντέλο εκπαιδεύεται αρχικά στα δεδομένα με ετικέτα. Στη συνέχεια κάνει προβλέψεις για τα δεδομένα χωρίς ετικέτα, αντιμετωπίζοντας αυτές τις προβλέψεις ως ψευδο-ετικέτες. Το μοντέλο επανεκπαίδευεται επαναληπτικά χρησιμοποιώντας τόσο τα αρχικά δεδομένα με ετικέτα όσο και τα νέα ψευδο-επισημασμένα δεδομένα (Yarowsky, 1995).
- **Συνεκπαίδευση:** Η συνεκπαίδευση περιλαμβάνει εκπαίδευση δύο ή περισσότερων μοντέλων σε διαφορετικές απόψεις των ίδιων δεδομένων. Κάθε μοντέλο εκπαιδεύεται σε ένα ξεχωριστό υποσύνολο χαρακτηριστικών και κάνει προβλέψεις για τα δεδομένα χωρίς ετικέτα. Οι προβλέψεις από το ένα μοντέλο χρησιμοποιούνται στη συνέχεια για την εκπαίδευση του άλλου μοντέλου και αυτή η διαδικασία συνεχίζεται επαναληπτικά.
- **Μέθοδοι που βασίζονται σε γράφημα:** Αυτές οι μέθοδοι αντιπροσωπεύουν τα δεδομένα ως γράφημα όπου κάθε κόμβος αντιστοιχεί σε ένα σημείο δεδομένων και οι ακμές αντιπροσωπεύουν ομοιότητες μεταξύ σημείων. Οι πληροφορίες ετικετών διαδίδονται μέσω του γραφήματος από κόμβους με ετικέτα σε μη επισημασμένους κόμβους, αξιοποιώντας την υπόθεση ότι παρόμοια σημεία είναι πιθανό να μοιράζονται την ίδια ετικέτα (Zhu & Ghahramani, 2002).

- Παραγωγικά μοντέλα: Αυτά τα μοντέλα προσπαθούν να μοντελοποιήσουν την κοινή κατανομή πιθανοτήτων των χαρακτηριστικών εισόδου και των ετικετών. Κατανοώντας τη διαδικασία παραγωγής δεδομένων, μπορούν να κάνουν πιο ακριβείς προβλέψεις ακόμη και με περιορισμένα δεδομένα με ετικέτα. Παραδείγματα περιλαμβάνουν Gaussian Mixture Models και Variational Autoencoders.

Αντίστοιχα, οι πιο κοινοί αλγόριθμοι της ημι-εποπτευόμενης μάθησης είναι :

- Αυτο-κατάρτιση: Αυτή η απλή αλλά αποτελεσματική προσέγγιση επισημαίνει επαναληπτικά τα δεδομένα χωρίς ετικέτα και βελτιώνει το μοντέλο. Συχνά χρησιμοποιείται σε συνδυασμό με άλλους ταξινομητές όπως δέντρα αποφάσεων ή μηχανές διανυσμάτων υποστήριξης (SVM).
- Συνεκπαίδευση: Απαιτεί τα δεδομένα να μπορούν να χωριστούν σε δύο διακριτές και επαρκείς προβολές. Κάθε ταξινομητής βελτιώνει τον άλλον μοιράζοντας σίγουρες προβλέψεις σε δεδομένα χωρίς ετικέτα .
- Διάδοση ετικετών: Ένας ημι-εποπτευόμενος αλγόριθμος μάθησης που βασίζεται σε γραφήματα που απλώνει ετικέτες μέσω του γραφήματος με βάση την ομοιότητα των κόμβων, χρησιμοποιώντας αποτελεσματικά τη δομή των δεδομένων για την ενημέρωση των προβλέψεων (Zhu & Ghahramani, 2002).
- Μηχανές διανυσμάτων ημι-εποπτευόμενης υποστήριξης (S3VM): Επεκτείνει τα παραδοσιακά SVM ενσωματώνοντας δεδομένα χωρίς ετικέτα στη διαδικασία εκπαίδευσης, με στόχο να βρεθεί ένα όριο απόφασης που μεγιστοποιεί το περιθώριο όχι μόνο σε δεδομένα με ετικέτα αλλά και σε ολόκληρο το σύνολο δεδομένων (Chapelle, Schölkopf και Zien , 2006).

Η αξιολόγηση των ημι-εποπτευόμενων μοντέλων μάθησης περιλαμβάνει την αξιολόγηση της απόδοσής τους σε δεδομένα με ετικέτα, καθώς και την ικανότητά τους να γενικεύουν από δεδομένα χωρίς ετικέτα. Οι κοινές μετρήσεις περιλαμβάνουν:

- Ακρίβεια: Η αναλογία των σωστά ταξινομημένων περιπτώσεων μεταξύ των δεδομένων με ετικέτα.
- Βαθμολογία F1: Ο αρμονικός μέσος ακρίβειας και ανάκλησης, που παρέχει ένα ισορροπημένο μέτρο της απόδοσης ενός μοντέλου.

- AUC-ROC: Η περιοχή κάτω από τη χαρακτηριστική καμπύλη λειτουργίας του δέκτη, που μετρά την ικανότητα του μοντέλου να διακρίνει μεταξύ των κατηγοριών.

Η ημι-εποπτευόμενη μάθηση είναι ιδιαίτερα χρήσιμη σε σενάρια όπου τα δεδομένα με ετικέτα είναι σπάνια ή ακριβά για να αποκτηθούν, αλλά τα δεδομένα χωρίς ετικέτα είναι άφθονα. Οι βασικές εφαρμογές περιλαμβάνουν την Επεξεργασία Φυσικής Γλώσσας (NLP) όπου εργασίες όπως η ταξινόμηση κειμένου και η ανάλυση συναισθήματος συχνά επωφελούνται από ημι-εποπτευόμενη μάθηση λόγω της αφθονίας των διαθέσιμων δεδομένων κειμένου χωρίς ετικέτα (Yarowsky, 1995). Επίσης χρησιμοποιείται στην ταξινόμηση εικόνων με την αξιοποίηση μεγάλων συλλογών εικόνων χωρίς ετικέτα για τη βελτίωση της απόδοσης των ταξινομητών που έχουν εκπαιδευτεί σε ένα περιορισμένο σύνολο εικόνων με ετικέτα (Rasmus et al., 2015). Ακόμη, χρησιμοποιείται για την ενίσχυση προγνωστικών μοντέλων για τη διάγνωση ασθενειών και την ανακάλυψη φαρμάκων όπου τα επισημασμένα βιοϊατρικά δεδομένα είναι περιορισμένα (Liu et al., 2008).

Η ημι-εποπτευόμενη μάθηση γεφυρώνει αποτελεσματικά το χάσμα μεταξύ εποπτευόμενης και μη εποπτευόμενης μάθησης χρησιμοποιώντας δεδομένα τόσο με ετικέτα όσο και χωρίς ετικέτα. Αυτή η προσέγγιση ενισχύει τη διαδικασία μάθησης, καθιστώντας την ιδιαίτερα πολύτιμη σε σενάρια πραγματικού κόσμου όπου τα δεδομένα με ετικέτα είναι συχνά περιορισμένα. Καθώς η διαθεσιμότητα των δεδομένων συνεχίζει να αυξάνεται, η ημι-εποπτευόμενη μάθηση θα διαδραματίσει κρίσιμο ρόλο στην ανάπτυξη πιο ακριβών και ισχυρών μοντέλων.

2.2.4. Ενισχυμένη Μάθηση

Η ενισχυτική μάθηση (RL) είναι ένας τύπος μηχανικής μάθησης όπου ένας πράκτορας μαθαίνει να λαμβάνει αποφάσεις εκτελώντας ενέργειες σε ένα περιβάλλον για τη μεγιστοποίηση της σωρευτικής ανταμοιβής. Αυτή η προσέγγιση είναι εμπνευσμένη από τη συμπεριφορική ψυχολογία, ιδιαίτερα την έννοια της μάθησης μέσω δοκιμής και λάθους, με ανταμοιβές και τιμωρίες που καθοδηγούν τη διαδικασία μάθησης. Τα βασικά συστατικά της είναι τα εξής:

- Πράκτορας: Ο μαθητής ή ο υπεύθυνος λήψης αποφάσεων που αλληλεπιδρά με το περιβάλλον.
- Περιβάλλον: Το εξωτερικό σύστημα με το οποίο αλληλεπιδρά ο πράκτορας. Παρέχει ανατροφοδότηση με τη μορφή ανταμοιβών ή κυρώσεων με βάση τις ενέργειες του πράκτορα.
- Κατάσταση: Αναπαράσταση της τρέχουσας κατάστασης ή διαμόρφωσης του περιβάλλοντος.
- Δράση: Το σύνολο όλων των πιθανών κινήσεων που μπορεί να κάνει ο πράκτορας.
- Ανταμοιβή: Ανατροφοδότηση από το περιβάλλον που χρησιμοποιείται για την αξιολόγηση των ενεργειών που έγιναν από τον πράκτορα. Μπορεί να είναι θετικό ή αρνητικό.
- Πολιτική: Μια στρατηγική που χρησιμοποιείται από τον πράκτορα για να αποφασίσει την επόμενη ενέργεια με βάση την τρέχουσα κατάσταση.
- Συνάρτηση αξίας: Εκτιμά την αναμενόμενη σωρευτική ανταμοιβή για μια δεδομένη κατάσταση ή ζεύγος ενεργειών κατάστασης, βοηθώντας τον πράκτορα να αξιολογήσει τα μακροπρόθεσμα οφέλη των ενεργειών.

Οι τύποι Ενισχυτικής Μάθησης είναι οι μέθοδοι με μοντέλα και χωρίς μοντέλα. Οι πρώτες περιλαμβάνουν την εκμάθηση ενός μοντέλου του περιβάλλοντος και τη χρήση του για τον σχεδιασμό ενεργειών. Το μοντέλο προβλέπει τις μελλοντικές καταστάσεις και τις ανταμοιβές, οι οποίες μπορούν να χρησιμοποιηθούν για τη βελτίωση της πολιτικής. Αντιθέτως, οι μέθοδοι χωρίς μοντέλα μαθαίνουν την πολιτική άμεσα ή έμμεσα χωρίς μοντέλο του περιβάλλοντος. Αυτές οι μέθοδοι χωρίζονται σε Q-Learning η οποία είναι μια μέθοδος που βασίζεται στην αξία όπου ο πράκτορας μαθαίνει την αξία των ενεργειών σε μια συγκεκριμένη κατάσταση (Watkins & Dayan, 1992) και σε μεθόδους κλίσης πολιτικής που βελτιστοποιούν την πολιτική απευθείας

προσαρμόζοντας τις παραμέτρους του δικτύου πολιτικής για να μεγιστοποιήσουν την αναμενόμενη ανταμοιβή (Sutton et al., 2000).

Οι βασικοί αλγόριθμοι είναι οι εξής:

- Q-Learning: Ένας ευρέως χρησιμοποιούμενος αλγόριθμος χωρίς μοντέλα όπου ο πράκτορας μαθαίνει μια συνάρτηση Q, η οποία εκτιμά την αξία της λήψης μιας συγκεκριμένης ενέργειας σε μια δεδομένη κατάσταση. Ο πράκτορας ενημερώνει τις τιμές Q επαναληπτικά χρησιμοποιώντας την εξίσωση Bellman (Watkins & Dayan, 1992).
- Deep Q-Networks (DQN): Μια επέκταση της Q-learning που χρησιμοποιεί βαθιά νευρωνικά δίκτυα για να προσεγγίσει τη συνάρτηση Q, επιτρέποντας στον αλγόριθμο να χειρίζεται χώρους καταστάσεων υψηλών διαστάσεων (Mnih et al., 2015).
- SARSA (State-Action-Reward-State-Action): Ένας άλλος αλγόριθμος χωρίς μοντέλα παρόμοιος με το Q-learning, αλλά ενημερώνει τις τιμές Q χρησιμοποιώντας την ενέργεια που λαμβάνει πραγματικά ο πράκτορας, αντί για τη βέλτιστη ενέργεια (Rummery & Niranjan, 1994).
- Μέθοδοι κλίσης πολιτικής: Αυτοί οι αλγόριθμοι, όπως το REINFORCE, βελτιστοποιούν άμεσα την πολιτική υπολογίζοντας τη διαβάθμιση της αναμενόμενης ανταμοιβής σε σχέση με τις παραμέτρους πολιτικής (Williams, 1992).
- Μέθοδοι κριτικής ηθοποιών: Συνδυάζουν μεθόδους που βασίζονται στην αξία και βασίζονται σε πολιτικές. Ο ηθοποιός ενημερώνει άμεσα την πολιτική ενώ ο κριτικός αξιολογεί τη δράση υπολογίζοντας τις συναρτήσεις τιμών (Konda & Tsitsiklis, 2000).

Η αξιολόγηση των αλγορίθμων ενισχυτικής μάθησης περιλαμβάνει την αξιολόγηση της απόδοσής τους με βάση διάφορα κριτήρια:

- Αθροιστική ανταμοιβή: Η συνολική ανταμοιβή που συσσωρεύεται από τον πράκτορα με την πάροδο του χρόνου. Οι υψηλότερες σωρευτικές ανταμοιβές υποδηλώνουν καλύτερη απόδοση.
- Ρυθμός εκμάθησης: Η ταχύτητα με την οποία ο πράκτορας βελτιώνει την απόδοσή του. Γενικά προτιμώνται ταχύτερα ποσοστά μάθησης.

- Σύγκλιση: Το σημείο στο οποίο σταθεροποιείται η απόδοση του πράκτορα. Ένας αλγόριθμος RL με καλή απόδοση θα πρέπει να συγκλίνει σε μια βέλτιστη ή σχεδόν βέλτιστη πολιτική.
- Sample Efficiency: Ο όγκος των δεδομένων που απαιτείται από τον πράκτορα για να μάθει μια αποτελεσματική πολιτική. Περισσότεροι αλγόριθμοι αποδοτικοί ως προς το δείγμα απαιτούν λιγότερες αλληλεπιδράσεις με το περιβάλλον.

Η ενισχυτική μάθηση έχει ένα ευρύ φάσμα εφαρμογών, όπως στα παιχνίδια όπου έχει χρησιμοποιηθεί για τη δημιουργία πρακτόρων που μπορούν να παίξουν και να διαπρέψουν σε πολύπλοκα παιχνίδια όπως το Go, το σκάκι και τα βιντεοπαιχνίδια (Silver et al., 2016). Στη ρομποτική, το RL δίνει τη δυνατότητα στα ρομπότ να μαθαίνουν πολύπλοκες εργασίες, όπως να πιάνουν αντικείμενα, να περπατούν και να πλοηγούνται σε περιβάλλοντα (Kober, et al., 2013). Επίσης, οι αλγόριθμοι RL εφαρμόζονται σε αλγοριθμικές συναλλαγές, διαχείριση χαρτοφυλακίου και διαχείριση κινδύνου για τη λήψη βέλτιστων οικονομικών αποφάσεων (Li, et al., 2020). Τέλος, στον κλάδο της υγείας, εξατομικευμένες στρατηγικές θεραπείας και βελτιστοποίηση των διαδικασιών λήψης κλινικών αποφάσεων μπορούν να αναπτυχθούν χρησιμοποιώντας RL (Yu et al., 2020).

Η ενισχυτική μάθηση είναι μια ισχυρή προσέγγιση στη μηχανική μάθηση, η οποία επιτρέπει στους πράκτορες να μάθουν βέλτιστες συμπεριφορές μέσω της αλληλεπίδρασης με το περιβάλλον τους. Εστιάζοντας στη μεγιστοποίηση των σωρευτικών ανταμοιβών, η RL έχει βρει εφαρμογές σε διάφορους τομείς, από τα παιχνίδια μέχρι τη ρομποτική και την υγειονομική περίθαλψη. Καθώς οι υπολογιστικοί πόροι και οι αλγοριθμικές τεχνικές συνεχίζουν να εξελίσσονται, η δυνατότητα RL να επιλύει όλο και πιο περίπλοκα προβλήματα συνεχίζει να αυξάνεται.

2.3. Προβλήματα και δυσλειτουργίες

Τα συστήματα μηχανικής μάθησης, αν και ισχυρά, αντιμετωπίζουν προκλήσεις, προβλήματα και πιθανές δυσλειτουργίες. Αυτά τα ζητήματα μπορεί να προκύψουν από διάφορες πηγές, όπως η ποιότητα των δεδομένων, η πολυπλοκότητα του μοντέλου που εφαρμόζουν, η ερμηνευσιμότητά τους, κ.α. Η κατανόηση αυτών των πιθανών προβλημάτων είναι ζωτικής σημασίας για την ανάπτυξη ισχυρών και αξιόπιστων εφαρμογών μηχανικής μάθησης.

Ποιότητα και ποσότητα δεδομένων:

Αυτή η κατηγορία προβλημάτων σχετίζεται με τα δεδομένα. Ένα από τα πιο συνηθισμένα ζητήματα είναι ο θόρυβος και οι ακραίες τιμές. Τα δεδομένα κακής ποιότητας που περιέχουν θόρυβο ή ακραίες τιμές μπορεί να παραπλανήσουν τη διαδικασία εκπαίδευσης, με αποτέλεσμα ανακριβή μοντέλα (Hastie, Tibshirani, & Friedman, 2009). Επίσης πρόβλημα μπορεί να προκύψει όταν υπάρχουν σύνολα δεδομένων χωρίς ισορροπία. Όταν οι κλάσεις στο σύνολο δεδομένων δεν αντιπροσωπεύονται εξίσου, το μοντέλο μπορεί να γίνει προκατειλημμένο προς την τάξη της πλειοψηφίας, οδηγώντας σε κακή απόδοση στην κατηγορία μειοψηφίας (He & Garcia, 2009). Ακόμη, το πρόβλημα των ελλιπών δεδομένων μπορεί να περιπλέξει τη διαδικασία εκπαίδευσης και να μειώσει την αποτελεσματικότητα του μοντέλου (Rubin, 1976).

Προκατάληψη δεδομένων και αντιπροσωπευτικότητα:

Ακόμη μια κατηγορία προβλημάτων που σχετίζεται με τα δεδομένα είναι η αντιπροσωπευτικότητά τους. Εάν τα δεδομένα εκπαίδευσης δεν είναι αντιπροσωπευτικά του πληθυσμού του πραγματικού κόσμου, οι προβλέψεις του μοντέλου θα είναι μεροληπτικές και όχι γενικεύσιμες (Zadrozny, Langford, & Abe, 2003). Επίσης, κάποια είδη δεδομένων όπως τα ιστορικά ενδέχεται να αντικατοπτρίζουν προκαταλήψεις και προκαταλήψεις του παρελθόντος, οι οποίες μπορούν να μάθουν ακούσια και να διαιωνιστούν από το μοντέλο σύμφωνα με τους Barocas & Selbst (2016).

Υπερπροσαρμογή και Υποπροσαρμογή:

Αυτή η κατηγορία προβλημάτων σχετίζεται με το μοντέλο. Όταν ένα μοντέλο μαθαίνει πολύ καλά τα δεδομένα εκπαίδευσης, συμπεριλαμβανομένου του θορύβου και των ακραίων στοιχείων του αν δεν έχουν αφαιρεθεί, μπορεί να έχει κακή απόδοση σε νέα, αόρατα δεδομένα τα οποία μπορεί να ήταν πιο ποιοτικά (Goodfellow, Bengio, & Courville, 2016). Αντίστοιχα, όταν ένα μοντέλο είναι πολύ απλό για να καταγράψει τα υποκείμενα μοτίβα στα δεδομένα, αποτυγχάνει να αποδώσει καλά ακόμη και στα δεδομένα εκπαίδευσης.

Πολυπλοκότητα μοντέλου:

Ακόμη μια κατηγορία που σχετίζεται με το μοντέλο αφορά την πολυπλοκότητά του. Η επιλογή των σωστών υπερπαραμέτρων είναι κρίσιμη για την απόδοση του μοντέλου. Οι κακές επιλογές μπορούν να οδηγήσουν σε υποβέλτιστα μοντέλα (Bergstra & Bengio, 2012). Επίσης, τα σύνθετα μοντέλα, ειδικά τα δίκτυα βαθιάς μάθησης, απαιτούν σημαντικούς υπολογιστικούς πόρους, οι οποίοι μπορεί να είναι περιοριστικός παράγοντας για ορισμένες εφαρμογές (Krizhevsky, Sutskever, & Hinton, 2012).

Ερμηνευσιμότητα και Διαφάνεια

Πολλά προηγμένα μοντέλα ML, ιδιαίτερα τα δίκτυα βαθιάς εκμάθησης, αναφέρονται συχνά ως "μαύρα κουτιά" λόγω της έλλειψης ερμηνείας τους. Αυτό μπορεί να είναι προβληματικό σε κρίσιμες εφαρμογές όπου η κατανόηση της διαδικασίας λήψης αποφάσεων είναι απαραίτητη (Lipton, 2018). Ακόμη υπάρχει και το ζήτημα της επεξήγησης του μοντέλου. Η αδυναμία να εξηγηθεί γιατί ένα μοντέλο έκανε μια συγκεκριμένη πρόβλεψη μπορεί να οδηγήσει σε δυσπιστία και απροθυμία στην υιοθέτηση συστημάτων ML, ειδικά σε τομείς όπως η υγειονομική περίθαλψη και τα οικονομικά (Doshi-Velez & Kim, 2017).

Ηθικές και Κοινωνικές Επιπτώσεις

Αυτά τα ζητήματα άπτονται όχι μονάχα του μοντέλου αλλά και του ίδιου του χρήστη σε περιπτώσεις που αλληλοεπιδρά με τον αλγόριθμο. Τα μοντέλα ML μπορούν άθελά τους να μάθουν και να ενισχύσουν τις κοινωνικές προκαταλήψεις που υπάρχουν στα

δεδομένα εκπαίδευσης, οδηγώντας σε άδικα και μεροληπτικά αποτελέσματα όπως αναφέρουν οι Barocas & Selbst (2016). Η χρήση προσωπικών δεδομένων στην εκπαίδευση μοντέλων ML εγείρει σημαντικά ζητήματα απορρήτου. Υπάρχει κίνδυνος παραβίασης δεδομένων και μη εξουσιοδοτημένης χρήσης ευαίσθητων πληροφοριών (Shokri et al., 2017). Επίσης, καθώς τα συστήματα ML γίνονται πιο αυτόνομα, η ανάθεση λογοδοσίας για τις αποφάσεις και τις ενέργειές τους γίνεται πολύπλοκη. Ο προσδιορισμός του ποιος είναι υπεύθυνος για τα αποτελέσματα των αποφάσεων που βασίζονται σε ML είναι μια σημαντική πρόκληση (Vaughan et al., 2018).

Θέματα Ανάπτυξης και Λειτουργίας

Αυτή η κατηγορία σχετίζεται με το μοντέλο της μηχανικής μάθησης. Η διασφάλιση ότι τα μοντέλα μπορούν να κλιμακωθούν αποτελεσματικά για να χειριστούν μεγάλους όγκους δεδομένων και απαιτήσεις υψηλής απόδοσης είναι ζωτικής σημασίας για εφαρμογές πραγματικού κόσμου (Dean et al., 2012). Ακόμη, τα μοντέλα ML μπορεί να είναι εύαλωτα σε επιθέσεις αντιπάλου όπου οι εσκεμμένα κατασκευασμένες εισροές μπορούν να εξαπατήσουν το μοντέλο να κάνει εσφαλμένες προβλέψεις (Szegedy et al., 2014). Η διασφάλιση της ευρωστίας έναντι τέτοιων επιθέσεων είναι απαραίτητη για την αξιοπιστία των συστημάτων. Τέλος, τα μοντέλα πρέπει να ενημερώνονται συνεχώς με νέα δεδομένα για να παραμένουν σχετικά και ακριβή. Η εφαρμογή αποτελεσματικών μηχανισμών για συνεχή μάθηση και προσαρμογή είναι πρόκληση (Diethe, Borchert, & Girolami, 2019).

Σε γενικές γραμμές, ενώ η μηχανική μάθηση προσφέρει δυνατότητες μετασχηματισμού σε διάφορους τομείς, συνοδεύεται από σημαντικές προκλήσεις και πιθανές δυσλειτουργίες. Η αντιμετώπιση αυτών των ζητημάτων απαιτεί μια πολύπλευρη προσέγγιση που περιλαμβάνει αξιόπιστες πρακτικές δεδομένων, προσεκτικό σχεδιασμό μοντέλων, διαφάνεια, ηθικά κριτήρια και συνεχή παρακολούθηση. Με την κατανόηση και τον μετριασμό αυτών των προκλήσεων, η αξιοπιστία και η αποτελεσματικότητα των συστημάτων ML μπορεί να βελτιωθεί σημαντικά.

3. Δυναμικά γραφήματα

Τα δυναμικά γραφήματα είναι δομές που χρησιμοποιούνται για την αναπαράσταση οντοτήτων και των σχέσεων τους με την πάροδο του χρόνου. Σε αντίθεση με τα πιο παραδοσιακά, στατικά γραφήματα που καταγράφουν σχέσεις σε ένα μόνο χρονικό σημείο, τα δυναμικά γραφήματα επιτρέπουν την αναπαράσταση και την ανάλυση των αλλαγών στη δομή και τις ιδιότητες του δικτύου καθώς εξελίσσονται. Αυτά τα γραφήματα είναι όλο και πιο σημαντικά σε τομείς όπως η ανάλυση κοινωνικών δικτύων, η βιολογία, τα δίκτυα υπολογιστών και άλλα, όπου η κατανόηση των χρονικών αλλαγών είναι ιδιαίτερα σημαντική. Τα βασικά χαρακτηριστικά των Δυναμικών Γραφημάτων είναι τα εξής (Holme and Saramäki, 2012):

1. Χρονική διάσταση: Όπως αναφέρθηκε προηγουμένως, τα δυναμικά γραφήματα ενσωματώνουν το στοιχείο του χρόνου, επιτρέποντας την ανάλυση του τρόπου με τον οποίο οι κόμβοι (οντότητες) και οι ακμές (σχέσεις) αλλάζουν με την πάροδο του χρόνου. Αυτή η μεταβολή μπορεί να σημαίνει την εμφάνιση ή την εξαφάνιση κόμβων και άκρων ή αλλαγές στη δύναμη ή τον τύπο των σχέσεων. Συνολικά η απεικόνιση αλλάζει και αυτό είναι ιδιαίτερα ορατό στον χρήστη.
2. Ακμές και κόμβοι με χρονική σήμανση: Αυτά τα δύο είναι κατηγορίες στοιχείων που παρατηρούνται εντός των διαγραμμάτων όπως αναφέρθηκαν προηγουμένως. Κάθε ακμή και κόμβος μπορεί να έχει χρονικές σημάνσεις που υποδεικνύουν τις περιόδους κατά τις οποίες υπάρχουν. Αυτή η δομή επιτρέπει την ακριβή παρακολούθηση του πότε σχηματίζονται ή διαλύονται οι σχέσεις και πότε εμφανίζονται ή φεύγουν οντότητες από το δίκτυο.
3. Στιγμιότυπα και μεταβάσεις: Τα δυναμικά γραφήματα μπορούν να αναπαρασταθούν ως μια σειρά στατικών στιγμιότυπων σε διαφορετικά χρονικά διαστήματα. Οι μεταβάσεις μεταξύ αυτών των στιγμιότυπων απεικονίζουν την εξέλιξη του δικτύου.
4. Συγκεντρωτικές και συνεχείς αναπαραστάσεις: Ανάλογα με τις απαιτήσεις ανάλυσης, τα δυναμικά γραφήματα μπορούν να συγκεντρωθούν σε συγκεκριμένα χρονικά παράθυρα για μείωση της πολυπλοκότητας ή να αναπαρασταθούν συνεχώς για λεπτομερή χρονική ανάλυση.

Τα δυναμικά γραφήματα βρίσκουν εφαρμογή σε μια σειρά από περιπτώσεις. Χρησιμοποιούνται για παράδειγμα στα κοινωνικά δίκτυα για τη μελέτη της εξέλιξης των κοινωνικών αλληλεπιδράσεων, όπως ο σχηματισμός και η διάλυση φιλιών, τα πρότυπα επικοινωνίας και η διάδοση πληροφοριών ή επιρροής με την πάροδο του χρόνου. Αυτή η χρήση μπορεί να αφορά το μάρκετινγκ των επιχειρήσεων που θέλουν να έχουν γνώση του κοινού τους. Σε ένα αρκετά διαφορετικό περιβάλλον, στη βιολογία συστημάτων, τα δυναμικά γραφήματα μοντελοποιούν τις αλληλεπιδράσεις πρωτεΐνης-πρωτεΐνης, τα ρυθμιστικά δίκτυα γονιδίων και άλλες βιολογικές διεργασίες που αλλάζουν με την πάροδο του χρόνου. Τα δυναμικά γραφήματα χρησιμοποιούνται επίσης για την αναπαράσταση δικτύων. Είτε πρόκειται για δίκτυα υπολογιστών όπου βοηθούν στην ανάλυση της συμπεριφοράς των δικτύων, όπως π.χ. η ανίχνευση επιθέσεων είτε πρόκειται για δίκτυα μεταφορών όπου τα γραφήματα μοντελοποιούν τη ροή της κυκλοφορίας και την εξέλιξη των συστημάτων μεταφορών, βοηθώντας στη βελτιστοποίηση των διαδρομών και στη διαχείριση της συμφόρησης. Μια ακόμη εφαρμογή τους είναι στον τομέα των οικονομικών όπου αναλύουν τις αλλαγές στις χρηματοπιστωτικές αγορές, τις εμπορικές σχέσεις και τους οικονομικούς δείκτες με την πάροδο του χρόνου.

Για την οπτικοποίηση των γραφημάτων αλλά και για την γενικότερη ανάλυσή τους χρησιμοποιούνται διάφορα εργαλεία, από αρκετά απλά ως ιδιαίτερα πολύπλοκα και σύγχρονα. Αρχικά για να γίνει η απεικόνιση χρησιμοποιούνται τεχνικές όπως τα κινούμενα σχέδια, οι χάρτες θερμότητας, οι δυναμικές μετρήσεις, η ανίχνευση χρονικού προτύπου, κ.α. Ως προς τα εργαλεία, χρησιμοποιούνται διάφορα λογισμικά και βιβλιοθήκες όπως τα παρακάτω (Aggarwal and Subbian, 2014):

- Gephi: Πρόκειται για λογισμικό ανοιχτού κώδικα για οπτικοποίηση και ανάλυση δικτύου που υποστηρίζει δυναμικά γραφήματα.
- GraphStream: Είναι βιβλιοθήκη Java σχεδιασμένη για τη μοντελοποίηση και ανάλυση δυναμικών γραφημάτων.
- NetworkX: Βιβλιοθήκη Python που, ενώ είναι κυρίως για στατικά γραφήματα, μπορεί να επεκταθεί για να χειριστεί δυναμικά γραφήματα.
- Pajek: Ένα εργαλείο για ανάλυση μεγάλων δικτύων που περιλαμβάνει δυνατότητες για δυναμική οπτικοποίηση δικτύου.

3.1. Κατανεμημένα δυναμικά γραφήματα

Τα κατανεμημένα δυναμικά γραφήματα είναι μια εξειδικευμένη μορφή δυναμικών γραφημάτων που έχουν σχεδιαστεί για να χειρίζονται δεδομένα μεγάλης κλίμακας, εξελισσόμενα στο χρόνο σε πολλαπλά, κατανεμημένα υπολογιστικά περιβάλλοντα. Είναι ιδιαίτερα σημαντικά σε σενάρια όπου τα δεδομένα είναι πολύ μεγάλα για να υποβληθούν σε επεξεργασία σε ένα μόνο μηχάνημα ή χρειάζεται διαχείριση και επεξεργασία σε πραγματικό χρόνο σε γεωγραφικά διασκορπισμένες τοποθεσίες. Τα κατανεμημένα δυναμικά γραφήματα έχουν σχεδιαστεί για να κλιμακώνονται σε πολλαπλές μηχανές, καθιστώντας δυνατή την αποτελεσματική επεξεργασία τεράστιων ποσοτήτων δεδομένων (big data). Μπορούν να χειριστούν ενημερώσεις σε πραγματικό χρόνο, επιτρέποντας τη συνεχή ενσωμάτωση νέων δεδομένων και την άμεση αντανάκλαση των αλλαγών στη δομή του δικτύου. Αυτά τα συστήματα μπορούν να συνεχίσουν να λειτουργούν ακόμη και αν κάποιοι κόμβοι αποτύχουν, ενισχύοντας την αξιοπιστία και την ευρωστία. Διάφοροι επεξεργαστές επιμερίζονται το φόρτο εργασίας γεγονός που μειώνει τον χρόνο επεξεργασίας αφού πραγματοποιούνται παράλληλοι υπολογισμοί.

Ως προς την αρχιτεκτονική τους, τα δεδομένα αποθηκεύονται σε πολλούς κόμβους με κατανεμημένο τρόπο, συχνά χρησιμοποιώντας κατανεμημένες βάσεις δεδομένων ή συστήματα αρχείων όπως το HDFS (Hadoop Distributed File System). Το κάθε γράφημα χωρίζεται σε μικρότερα υπογραφήματα, καθένα από τα οποία χειρίζεται διαφορετικούς κόμβους. Οι αποτελεσματικές στρατηγικές κατάτμησης είναι ζωτικής σημασίας για την ελαχιστοποίηση των επιβαρύνσεων επικοινωνίας μεταξύ των κόμβων. Δομές όπως το Apache Giraph, το Apache Flink και το Apache Spark GraphX παρέχουν την υποδομή για την επεξεργασία κατανεμημένων γραφημάτων. Αυτά τα πλαίσια προσφέρουν API και εργαλεία για τη διαχείριση της διανομής δεδομένων και των υπολογισμών. Τα αποτελεσματικά πρωτόκολλα επικοινωνίας είναι απαραίτητα για τον συγχρονισμό και την ενημέρωση των κατανεμημένων κόμβων, διασφαλίζοντας τη συνέπεια και την ακρίβεια των δεδομένων του γραφήματος.

4. Εντοπισμός Ανωμαλιών

4.1. Τι είναι

Η ανίχνευση ανωμαλιών σε δυναμικά γραφήματα είναι μια διαδικασία που στοχεύει στον εντοπισμό ασυνήθιστων προτύπων, συμπεριφορών ή αλλαγών σε ένα δίκτυο που αποκλίνουν σημαντικά από τους αναμενόμενους κανόνες με την πάροδο του χρόνου. Ονομάζεται επίσης και ανίχνευση ακραίων τιμών και αφορά επίσης την αναγνώριση παρατηρήσεων, γεγονότων ή σημείων δεδομένων που αποκλίνουν από αυτό που είναι συνηθισμένο, τυπικό ή αναμενόμενο, καθιστώντας τα ασυνεπή με το υπόλοιπο σύνολο δεδομένων.

Πρόκειται για μια διαδικασία που είναι ιδιαίτερα σημαντική σε διάφορους τομείς, όπως η ασφάλεια στον κυβερνοχώρο, τα κοινωνικά δίκτυα κ.α. όπου τέτοιες ανωμαλίες μπορεί να υποδηλώνουν από δυσλειτουργίες ως πιθανές απειλές για την ασφάλεια, απάτη ή ακόμη και σημαντικά βιολογικά συμβάντα. Συγκεκριμένα, στην κυβερνοασφάλεια μέσω της διαδικασίας αυτής ανιχνεύονται ασυνήθιστα μοτίβα στην κυκλοφορία δικτύου που θα μπορούσαν να υποδεικνύουν πιθανές επιθέσεις ή παραβιάσεις (Ahmed et al., 2016). Στα μέσα κοινωνικής δικτύωσης η διαδικασία αφορά τον προσδιορισμό μη φυσιολογικών μοτίβων συμπεριφοράς, όπως το spamming ή οι ψεύτικοι λογαριασμοί (Chandola et al., 2009). Ακόμη ένας τομέας που εφαρμόζονται αυτές οι διαδικασίες είναι τα οικονομικά και συγκεκριμένα τα χρηματοοικονομικά όπου πραγματοποιείται ανίχνευση δόλιων συναλλαγών παρακολουθώντας αλλαγές στα δίκτυα συναλλαγών (Akoglu et al., 2015).

Για την πραγματοποίηση αυτών των διαδικασιών χρησιμοποιούνται διάφορες τεχνικές οι οποίες βασίζονται στα γραφήματα. Μια μέθοδος για την εξόρυξη και τη σύνοψη εξελισσόμενων ακολουθιών γραφημάτων για την ανίχνευση ανωμαλιών είναι το EagleMine (Akoglu et al., 2015). Το λογισμικό που χρησιμοποιείται για να πραγματοποιηθεί η ανεύρεση ανωμαλιών περιλαμβάνει επίσης το ELKI το οποίο είναι μια βιβλιοθήκη εργαλείων εξόρυξης δεδομένων Java ανοιχτού κώδικα που περιέχει αρκετούς αλγόριθμους ανίχνευσης ανωμαλιών, καθώς και επιτάχυνση ευρετηρίου για αυτούς. Επίσης υπάρχει το PyOD - μια βιβλιοθήκη Python ανοιχτού κώδικα που αναπτύχθηκε ειδικά για τον εντοπισμό ανωμαλιών (Zhao et al., 2019). Υπάρχει επίσης και το scikit-learn, ακόμη μια βιβλιοθήκη Python ανοιχτού κώδικα που περιέχει ορισμένους αλγόριθμους για ανίχνευση ανωμαλιών μέσω μηχανικής μάθησης χωρίς

επίβλεψη. Τέλος, το Wolfram Mathematica παρέχει λειτουργικότητα για ανίχνευση ανωμαλιών χωρίς επίβλεψη σε πολλούς τύπους δεδομένων¹.

Η διαδικασία αυτή διεξάγεται μέσω αλγορίθμων της μηχανικής μάθησης όπως η deep learning – βαθιά μάθηση και τα Νευρωνικά δίκτυα γραφημάτων – GNN. Κατά την πρώτη, γίνεται χρήση τεχνικών όπως αυτοκωδικοποιητές και επαναλαμβανόμενα νευρωνικά δίκτυα για τη μοντελοποίηση και την ανίχνευση ανωμαλιών σε δεδομένα γραφημάτων χρονοσειρών (Ding et al., 2019). Στη δεύτερη περίπτωση, στα GNN αξιοποιείται η αναπαραστατική του ικανότητα για την ανίχνευση ανωμαλιών μαθαίνοντας πολύπλοκα μοτίβα σε δεδομένα γραφημάτων όπως αναφέρουν οι Wu et al. (2020).

4.2. Είδη και Κατηγορίες Ανωμαλιών

Οι ανωμαλίες διακρίνονται κατά τύπο και κατά κατηγορία. Οι τύποι ανωμαλιών αναφέρονται στην ειδική φύση των ίδιων των ανωμαλιών, όπως οι συμπεριφορικές, οι δομικές, οι χρονικές και οι χωρικές ανωμαλίες. Οι κατηγορίες ανωμαλιών αναφέρονται στον τρόπο με τον οποίο ομαδοποιούνται οι ανωμαλίες με βάση την εμφάνισή τους ή το πλαίσιο τους, όπως οι ανωμαλίες σημείου, οι ανωμαλίες συμπραζομένων και οι συλλογικές ανωμαλίες. Ουσιαστικά, οι τύποι επικεντρώνονται στα χαρακτηριστικά των ανωμαλιών, ενώ οι κατηγορίες επικεντρώνονται στα πλαίσια ή τα μοτίβα στα οποία εμφανίζονται οι ανωμαλίες.

Η τμηματοποίηση των κατηγοριών ανωμαλιών τυπικά χωρίζεται σε τρεις κύριους τύπους: Point, Contextual & Collective anomalies. Αυτές οι κατηγορίες αναλύονται στη συνέχεια, για την κατανόηση των βασικών διαφορών μεταξύ τους.

Point anomalies - Ανωμαλίες σημείων

Οι ανωμαλίες σημείων, που αναφέρονται επίσης ως καθολικές ανωμαλίες, είναι μεμονωμένα σημεία δεδομένων που διαφέρουν σημαντικά από τα υπόλοιπα δεδομένα. Αυτές οι ανωμαλίες συχνά ανιχνεύονται μέσω στατιστικών μεθόδων, όπως ο υπολογισμός της βαθμολογίας z ή η χρήση μετρήσεων που βασίζονται στην απόσταση

¹ <https://reference.wolfram.com/language/ref/FindAnomalies.html>

για να προσδιοριστεί πόσο αποκλίνει ένα σημείο δεδομένων από τον μέσο όρο ή τη διάμεσο του συνόλου δεδομένων. Για παράδειγμα, σε ένα οικονομικό πλαίσιο, ένα ποσό συναλλαγής που είναι δραστικά υψηλότερο ή χαμηλότερο από τις τυπικές συναλλαγές θα μπορούσε να θεωρηθεί ως ανωμαλία βαθμών. Ομοίως, μια ξαφνική αύξηση στη χρήση της CPU σε ένα σύστημα παρακολούθησης δικτύου μπορεί να επισημανθεί ως ανωμαλία. Η ανίχνευση ανωμαλιών σημείων είναι σημαντική για τον εντοπισμό μεμονωμένων γεγονότων που μπορεί να υποδεικνύουν δόλια δραστηριότητα, σφάλματα συστήματος ή άλλες σημαντικές αποκλίσεις (Chandola, Banerjee, & Kumar, 2009).

Contextual anomalies

Πρόκειται για σημεία δεδομένων που θεωρούνται ανώμαλα σε ένα συγκεκριμένο πλαίσιο, αλλά μπορεί να φαίνονται φυσιολογικά σε άλλο. Το πλαίσιο ορίζεται από χαρακτηριστικά που επηρεάζουν την ερμηνεία των δεδομένων. Για παράδειγμα, στα δεδομένα χρονοσειρών, τα «συμφραζόμενα» χαρακτηριστικά θα μπορούσαν να είναι η ώρα και η ημερομηνία, όπου μια ασυνήθιστα υψηλή θερμοκρασία μπορεί να είναι φυσιολογική το καλοκαίρι αλλά ανώμαλη το χειμώνα. Η ανίχνευση τους περιλαμβάνει την κατανόηση της σχέσης μεταξύ του σημείου δεδομένων και του πλαισίου του, η οποία μπορεί να είναι πιο περίπλοκη από τον εντοπισμό ανωμαλιών σημείων. Αυτός ο τύπος ανωμαλίας είναι ιδιαίτερα σημαντικός σε εφαρμογές όπου τα δεδομένα εξαρτώνται από εξωτερικές συνθήκες, όπως η παρακολούθηση του περιβάλλοντος, η υγειονομική περίθαλψη (π.χ. παρακολούθηση ζωτικών σημείων) και τα εποχιακά δεδομένα πωλήσεων (Song, Wu, & Jermaine, 2007).

Collective anomalies - Συλλογικές ανωμαλίες

Συλλογικές ανωμαλίες συμβαίνουν όταν μια συλλογή σχετικών σημείων δεδομένων είναι ανώμαλη, παρόλο που μεμονωμένα σημεία δεδομένων εντός της συλλογής μπορεί να μην είναι. Αυτές οι ανωμαλίες είναι σημαντικές σε σενάρια όπου η σχέση μεταξύ των σημείων δεδομένων είναι κρίσιμη. Για παράδειγμα, στην ασφάλεια δικτύου, μια σειρά από συναλλαγές χαμηλής αξίας που πραγματοποιούνται μέσα σε σύντομο χρονικό διάστημα ενδέχεται να υποδηλώνουν συλλογικά δόλια συμπεριφορά, παρόλο που κάθε συναλλαγή ξεχωριστά φαίνεται φυσιολογική. Ομοίως, μια ακολουθία

ασυνήθιστων πακέτων δικτύου μπορεί να σηματοδοτεί μια παραβίαση ασφάλειας ή μια κατανεμημένη επίθεση άρνησης υπηρεσίας (DDoS). Η ανίχνευση συλλογικών ανωμαλιών περιλαμβάνει την ανάλυση των προτύπων και των σχέσεων μεταξύ των σημείων δεδομένων, που συχνά απαιτούν προηγμένες τεχνικές όπως η ανάλυση ακολουθίας, η ανάλυση χρονοσειρών και η ομαδοποίηση (Akoglu, Tong, & Koutra, 2015).

Η βιβλιογραφία περιλαμβάνει επίσης και την κατάτμηση των ανωμαλιών κατά τύπο ή είδος. Πρόκειται για μια διάκριση που αφορά διαφορετικές πτυχές των δεδομένων και απαιτεί συγκεκριμένες μεθόδους ανίχνευσης προσαρμοσμένες στα χαρακτηριστικά του.

Behavioral Anomalies - Ανωμαλίες συμπεριφοράς

Οι ανωμαλίες συμπεριφοράς αναφέρονται σε αποκλίσεις στα πρότυπα συμπεριφοράς μέσα σε ένα σύνολο δεδομένων. Αυτές οι ανωμαλίες συμβαίνουν όταν η συμπεριφορά μιας οντότητας αλλάζει απροσδόκητα σε σύγκριση με τη συνήθη ή αναμενόμενη συμπεριφορά της. Για παράδειγμα, στη δραστηριότητα των χρηστών σε έναν site, μια ανωμαλία συμπεριφοράς μπορεί να είναι ένας χρήστης που ξεκινά ξαφνικά να πραγματοποιεί μεγάλο αριθμό αγορών, κάτι που θα μπορούσε να υποδηλώνει δόλια δραστηριότητα. Ομοίως, στην παρακολούθηση δικτύου, μια ξαφνική αύξηση στη μεταφορά δεδομένων από έναν συγκεκριμένο χρήστη θα μπορούσε να επισημανθεί ως ανωμαλία συμπεριφοράς, υποδηλώνοντας μια πιθανή παραβίαση της ασφάλειας. Οι ανωμαλίες συμπεριφοράς συχνά ανιχνεύονται μέσω τεχνικών που προφίλ και παρακολουθούν τα φυσιολογικά πρότυπα συμπεριφοράς των οντοτήτων και εντοπίζουν σημαντικές αποκλίσεις από αυτά τα προφίλ (Chandola, Banerjee, & Kumar, 2009).

Structural Anomalies - Δομικές Ανωμαλίες

Οι δομικές ανωμαλίες σχετίζονται με τη δομή των δεδομένων, ιδιαίτερα εντός δικτύων ή δεδομένων που βασίζονται σε γραφήματα. Αυτές οι ανωμαλίες συμβαίνουν όταν οι σχέσεις μεταξύ των σημείων δεδομένων αποκλίνουν από την αναμενόμενη δομή. Για παράδειγμα, σε ένα κοινωνικό δίκτυο, μια απροσδόκητη σχέση φιλίας μεταξύ δύο κατά τα άλλα άσχετων ομάδων μπορεί να θεωρηθεί δομική ανωμαλία. Στην ασφάλεια του κυβερνοχώρου, οι δομικές ανωμαλίες θα μπορούσαν να εκδηλωθούν ως ασυνήθιστες συνδέσεις μεταξύ κόμβων σε ένα δίκτυο, υποδεικνύοντας πιθανώς ένα

παραβιασμένο σύστημα. Η ανίχνευση δομικών ανωμαλιών περιλαμβάνει την ανάλυση των τοπολογικών ιδιοτήτων του γραφήματος ή του δικτύου και τον εντοπισμό ασυνήθιστων μοτίβων ή υποδομών που αποκλίνουν από τον κανόνα (Akoglu, Tong, & Koutra, 2015).

Temporal Anomalies - Χρονικές ανωμαλίες

Οι χρονικές ανωμαλίες συμβαίνουν όταν τα σημεία δεδομένων αποκλίνουν από τα αναμενόμενα μοτίβα με την πάροδο του χρόνου. Αυτές οι ανωμαλίες προσδιορίζονται με βάση το χρονικό τους πλαίσιο, όπως τα δεδομένα χρονοσειρών, όπου οι ανωμαλίες μπορεί να είναι ξαφνικές αιχμές ή πτώσεις στα δεδομένα. Για παράδειγμα, μια ξαφνική αύξηση στη χρήση ηλεκτρικής ενέργειας σε μια συγκεκριμένη ώρα της ημέρας μπορεί να υποδηλώνει ανωμαλία σε ένα σύστημα ηλεκτρικού δικτύου. Ομοίως, μια απροσδόκητη πτώση των πωλήσεων κατά τη διάρκεια μιας τυπικής περιόδου υψηλών πωλήσεων θα μπορούσε να επισημανθεί ως μια χρονική ανωμαλία στα αναλυτικά στοιχεία λιανικής. Οι χρονικές ανωμαλίες συχνά ανιχνεύονται χρησιμοποιώντας μεθόδους ανάλυσης χρονοσειρών, συμπεριλαμβανομένων μοντέλων ARIMA, κινητών μέσων όρων και εποχιακής αποσύνθεσης (Song, Wu, & Jermaine, 2007).

Spatial Anomalies - Χωρικές Ανωμαλίες

Οι χωρικές ανωμαλίες αναφέρονται σε αποκλίσεις στη χωρική κατανομή των σημείων δεδομένων. Αυτές οι ανωμαλίες συμβαίνουν όταν η θέση ή η χωρική διάταξη των σημείων δεδομένων είναι ασυνήθιστη. Για παράδειγμα, στα γεωγραφικά δεδομένα, μια απροσδόκητη συστάδα περιπτώσεων ασθένειας σε μια συγκεκριμένη περιοχή μπορεί να θεωρηθεί ως χωρική ανωμαλία. Στην περιβαλλοντική παρακολούθηση, μια ασυνήθιστα υψηλή συγκέντρωση ρύπων σε μια συγκεκριμένη περιοχή θα μπορούσε να υποδηλώνει μια χωρική ανωμαλία. Η ανίχνευση χωρικών ανωμαλιών περιλαμβάνει την ανάλυση των χωρικών σχέσεων και των κατανομών των σημείων δεδομένων χρησιμοποιώντας χωρικές στατιστικές, γεωστατιστικές και τεχνικές χωρικής ομαδοποίησης (Chandola, Banerjee, & Kumar, 2009).

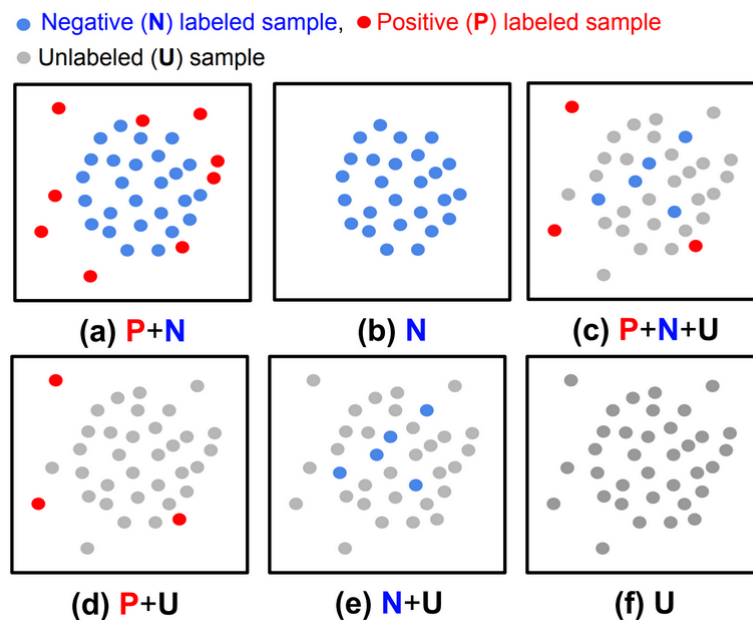
4.3. Τρόπος εντοπισμού

Η ανίχνευση των ανωμαλιών σχετίζεται με τον εντοπισμό ασυνήθιστων προτύπων ή σημείων δεδομένων που αποκλίνουν σημαντικά από την αναμενόμενη συμπεριφορά. Διάφορες μέθοδοι έχουν αναπτυχθεί για την αποτελεσματική ανίχνευση ανωμαλιών, καθεμία προσαρμοσμένη σε συγκεκριμένους τύπους δεδομένων και πλαισίων. Σε αυτό το κεφάλαιο παρουσιάζονται κάποιες από τις πιο βασικές μεθόδους ανίχνευσης, που υπάρχουν στη σχετική βιβλιογραφία. Αξιοποιώντας αυτές τις διάφορες μεθόδους ανίχνευσης, οι ερευνητές και οι επαγγελματίες μπορούν να εντοπίσουν αποτελεσματικά και να μετριάσουν τον αντίκτυπο των ανωμαλιών σε διαφορετικούς τομείς, ενισχύοντας την αξιοπιστία και την ασφάλεια των συστημάτων και των δεδομένων τους.

Οι στατιστικές μέθοδοι είναι από τις παλιότερες και πιο συχνά χρησιμοποιούμενες τεχνικές για την ανίχνευση ανωμαλιών. Αυτές οι μέθοδοι βασίζονται στην υπόθεση ότι τα κανονικά σημεία δεδομένων εμφανίζονται σε περιοχές υψηλής πιθανότητας ενός στατιστικού μοντέλου, ενώ οι ανωμαλίες εμφανίζονται σε περιοχές χαμηλής πιθανότητας. Το Z-Score είναι μια απλή αλλά αποτελεσματική μέθοδος όπου τα σημεία δεδομένων με z-score πέρα από ένα συγκεκριμένο όριο θεωρούνται ανωμαλίες. Αυτή η τεχνική είναι αποτελεσματική για μονομεταβλητά δεδομένα (Chandola, Banerjee, & Kumar, 2009). Επίσης, οι ανωμαλίες εντοπίζονται με την προσαρμογή μιας κατανομής πιθανότητας στα δεδομένα και τον εντοπισμό σημείων που εμπίπτουν στις ουρές της κατανομής (Barnett & Lewis, 1994).

Σε αυτό το πεδίο εφαρμόζονται επίσης οι μέθοδοι μηχανικής μάθησης οι οποίες έχουν κερδίσει δημοτικότητα για την ικανότητά τους να μοντελοποιούν πολύπλοκα μοτίβα και να χειρίζονται δεδομένα υψηλών διαστάσεων. Αυτές οι μέθοδοι μπορούν να κατηγοριοποιηθούν ευρέως σε εποπτευόμενες, μη εποπτευόμενες και ημι-εποπτευόμενες προσεγγίσεις. Η εποπτευόμενη μάθηση αφορά την εκπαίδευση ενός μοντέλου σε δεδομένα με ετικέτα, όπου κάθε περίπτωση επισημαίνεται ως κανονική ή ανώμαλη. Τεχνικές όπως μηχανές διανυσμάτων υποστήριξης (SVM) και νευρωνικά δίκτυα χρησιμοποιούνται συνήθως για αυτό το σκοπό όπως αναφέρουν οι Ahmed et al. (2016). Μια άλλη κατηγορία είναι η μη εποπτευόμενη μάθηση, η οποία χρησιμοποιείται συχνά διότι τα δεδομένα με ετικέτα είναι πιο σπάνια. Οι αλγόριθμοι ομαδοποίησης όπως το K-means και το DBSCAN ανιχνεύουν ανωμαλίες ως σημεία δεδομένων που δεν ταιριάζουν καλά σε κανένα σύμπλεγμα (Ester et al., 1996). Μείγμα

των δύο παραπάνω μεθόδων είναι η ημι-εποπτευόμενη μάθηση η οποία συνδυάζει πτυχές τόσο της εποπτευόμενης όσο και της μη εποπτευόμενης μάθησης χρησιμοποιώντας μια μικρή ποσότητα δεδομένων με ετικέτα μαζί με μια μεγάλη ποσότητα δεδομένων χωρίς ετικέτα. Αυτή η προσέγγιση είναι ιδιαίτερα χρήσιμη όταν η επισήμανση δεδομένων είναι δαπανηρή ή χρονοβόρα (Chapelle et al., 2006). Στη συνέχεια παρατίθεται μια περίπτωση ανίχνευσης ανωμαλιών η οποία διεξάγεται ανάλογα με τη διαθεσιμότητα του τύπου δεδομένων, είτε με αρνητική (κανονικό) είτε με θετική (ανώμαλο) ένδειξη ετικέτας. Στα παρακάτω διακρίνονται έξι διαφορετικές περιπτώσεις: (α) Πλήρως εποπτευόμενη ανίχνευση ανωμαλίας, (β) ανίχνευση ανωμαλίας μόνο με κανονικό τρόπο, (γ, δ, ε) ανίχνευση ανωμαλίας ημι-εποπτευόμενη, (στ) ανίχνευση ανωμαλίας χωρίς επίβλεψη.

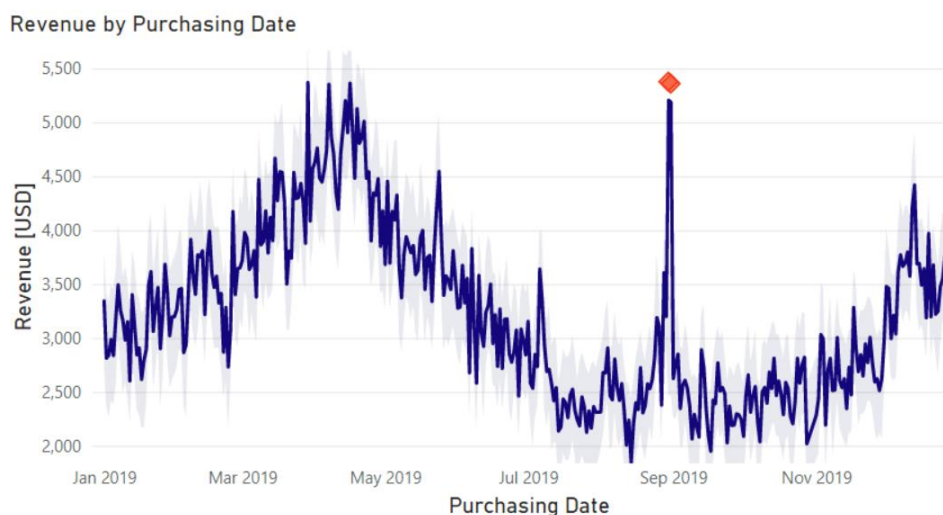


Διάγραμμα 1 - Ανίχνευση ανωμαλιών με μηχανική μάθηση

Μια τρίτη κατηγορία μεθόδων εντοπισμού είναι οι μέθοδοι που βασίζονται σε γραφήματα, οι οποίες είναι ιδιαίτερα αποτελεσματικές για την ανίχνευση ανωμαλιών σε δεδομένα που μπορούν να αναπαρασταθούν ως γράφημα, όπως κοινωνικά δίκτυα, δίκτυα επικοινωνίας και βιολογικά δίκτυα. Η ενσωμάτωση γραφήματος είναι μια τέτοια τεχνική κατά την οποία ενσωματώνεται ένα γράφημα σε χώρο χαμηλότερης διάστασης και στο οποίο ανιχνεύονται οι ανωμαλίες με βάση τις ενσωματωμένες αναπαραστάσεις (Akoglu, Tong, & Koutra, 2015). Ανώμαλοι κόμβοι ή ακμές είναι εκείνοι που αποκλίνουν σημαντικά από τον κανόνα σε αυτόν τον χώρο. Επίσης μέσω της εξόρυξης υπογραφών προσδιορίζονται ασυνήθιστες υποδομές μέσα στο γράφημα.

Για παράδειγμα, οι συχνότερες υπογραφών μπορούν να αναλυθούν για να ανιχνευθούν ανώμαλα μοτίβα (Noble & Cook, 2003).

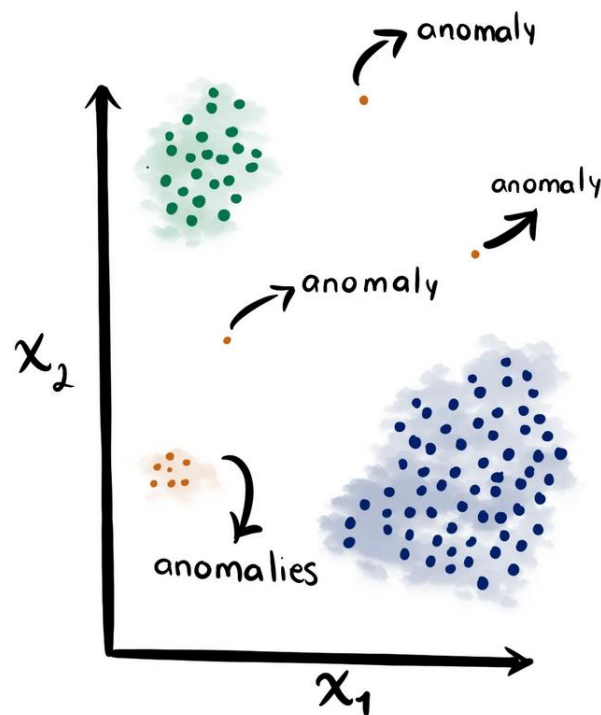
Η ανάλυση χρονοσειρών είναι ακόμη μια μέθοδος ιδιαίτερης σημασίας για τον εντοπισμό ανωμαλιών σε δεδομένα που συλλέγονται με την πάροδο του χρόνου, όπως οικονομικά δεδομένα, αναγνώσεις αισθητήρων και κίνηση δικτύου. Τα μοντέλα AutoRegressive Integrated Moving Average (ARIMA) χρησιμοποιούνται για την πρόβλεψη μελλοντικών σημείων δεδομένων και τον εντοπισμό ανωμαλιών ως σημεία που αποκλίνουν σημαντικά από αυτές τις προβλέψεις (Box & Jenkins, 1970). Επίσης η εποχιακή αποσύνθεση είναι μια μέθοδος αποσυνθέτει τα δεδομένα χρονοσειρών σε εποχιακά, τάσεις και υπολειπόμενα συστατικά και ανιχνεύονται ανωμαλίες στην υπολειπόμενη συνιστώσα (Cleveland et al., 1990). Ακολουθεί παράδειγμα ανάλυσης χρονοσειρών για την ανίχνευση πιθανής απάτης. Το σημείο που επισημαίνεται φαίνεται ότι είναι εντελώς εκτός της ροής των δεδομένων εκείνη την περίοδο.



Διάγραμμα 2 - Εντοπισμός ανωμαλιών σε δεδομένα χρονοσειρών

Οι μέθοδοι που βασίζονται σε ομαδοποίηση διεξάγονται μέσω αλγορίθμων της μηχανικής μάθησης και ομαδοποιούν τα σημεία δεδομένων σε συστάδες και προσδιορίζουν τις ανωμαλίες ως σημεία που δεν ανήκουν σε κανένα σύμπλεγμα ή ανήκουν σε μικρά, αραιά συμπλέγματα. Κατά τους αλγορίθμους K-means, τα σημεία δεδομένων μακριά από το πλησιέστερο κέντρο συστάδας θεωρούνται ανωμαλίες (MacQueen, 1967) ενώ κατά τους DBSCAN, οι ανωμαλίες ανιχνεύονται ως σημεία σε περιοχές χαμηλής πυκνότητας (Ester et al., 1996). Αντίστοιχα, οι μέθοδοι που

βασίζονται σε απόσταση είναι άλλη μια εφαρμογή της μηχανικής μάθησης με τις οποίες ανιχνεύονται ανωμαλίες με βάση την απόστασή τους από άλλα σημεία δεδομένων. Τα σημεία που βρίσκονται μακριά από τους γείτονές τους θεωρούνται ανωμαλίες. Οι K-Nearest Neighbors (KNN) αλγόριθμοι μετρούν την απόσταση κάθε σημείου δεδομένων από τους k-πλησιέστερους γείτονές του και επισημαίνει σημεία με μεγάλες μέσες αποστάσεις ως ανωμαλίες (Ramaswamy, Rastogi, & Shim, 2000). Στη συνέχεια παρατίθεται μια διαγραμματική απεικόνιση της ανίχνευσης ανωμαλιών με αλγόριθμο K-means όπου οι παρατηρήσεις που δεν συμπεριλαμβάνονται στην ομαδοποίηση (clustering) επισημαίνονται ως ανωμαλίες.



Διάγραμμα 3 - Ανίχνευση ανωμαλιών με αλγόριθμο K-means

Η πρόσφατη μελέτη των Kosal et al. (2023) προτείνει διάφορες μεθόδους ανίχνευσης για ανωμαλίες σε δυναμικά γραφήματα, ιδιαίτερα μέσω της χρήσης ενός νέου μοντέλου βαθιάς μάθησης που ονομάζεται DyGED (Dynamic Graph Event Detection). Οι μέθοδοι που προτείνουν αφορούν πέντε διαφορετικές προσεγγίσεις. Τα GCN (Graph Convolutional Network) χρησιμοποιούνται για την εκμάθηση ενσωματώσεων κόμβων αξιοποιώντας τη δομή του γραφήματος και τα χαρακτηριστικά του κόμβου. Αυτή η προσέγγιση βοηθά στη συλλογή πληροφοριών τοπικής γειτονιάς (k-n). Ο τελεστής συγκέντρωσης (pooling operator), συγκεντρώνει τις ενσωματώσεις

κόμβων σε μια ενιαία ενσωμάτωση σε επίπεδο γραφήματος. Αυτό το βήμα είναι κρίσιμο για τη μετατροπή των αναπαραστάσεων σε επίπεδο κόμβου σε μια σύνοψη σε επίπεδο γραφήματος. Το επαναλαμβανόμενο νευρωνικό δίκτυο (RNN) και ιδιαίτερα τα δίκτυα μακράς βραχυπρόθεσμης μνήμης (LSTM), χρησιμοποιείται για την καταγραφή της χρονικής δυναμικής του γραφήματος με την επεξεργασία ακολουθιών ενσωματώσεων γραφημάτων. Οι χρονικοί μηχανισμοί αυτοπροσοχής (Temporal Self-Attention) χρησιμοποιούνται για να σταθμίσουν τη σημασία των διαφορετικών χρονικών βημάτων, επιτρέποντας στο μοντέλο να εστιάζει στις πιο σχετικές πληροφορίες του παρελθόντος κατά την πρόβλεψη γεγονότων. Τέλος, το Multi-Layer Perceptron (MLP) είναι ένα απλό νευρωνικό δίκτυο προς τα εμπρός που λαμβάνει την τελική ενσωμάτωση σε επίπεδο γραφήματος και παράγει την έξοδο ανίχνευσης συμβάντων. Συνολικά οι Kosal et al. (2023) παρουσίασαν μια εκτενή πειραματική αξιολόγηση που καταδεικνύει ότι το DyGED ξεπερνά τις προϋπάρχουσες λύσεις όσον αφορά την ακρίβεια ανίχνευσης συμβάντων ενώ είναι πιο κλιμακωτό. Βασικά χαρακτηριστικά, όπως η επεκτασιμότητα και η ικανότητα αποτελεσματικής εκμάθησης τόσο δομικών όσο και χρονικών προτύπων επισημαίνονται μέσω διαφόρων συνόλων δεδομένων που καλύπτουν την κινητικότητα, την επικοινωνία και τα δεδομένα περιεχομένου που δημιουργούνται από τους χρήστες.

4.3.1. Εντοπισμός ανωμαλιών στις ακμές γραφημάτων

Οι Ma et al. (2021) αναφέρουν ότι υπάρχουν συγκεκριμένες τεχνικές ανίχνευσης ανώμαλων ακμών, εστιάζοντας ιδιαίτερα στις μεθόδους που ισχύουν τόσο για στατικά όσο και για δυναμικά γραφήματα. Ακολουθούν οι κύριες κατηγορίες/τύποι ανίχνευσης ανωμαλιών στις ακμές:

1. Τεχνικές βασισμένες σε βαθύ νευρωνικό δίκτυο (DNN). Σε αυτή την κατηγορία αναφέρονται δύο τεχνικές. Η ενοποιημένη ανίχνευση βάσει ενσωμάτωσης γραφήματος (UGED), μια μέθοδος μοντελοποιεί την κατανομή των ακμών μέσω μοντέλων σε βάθος. Προβλέπει την πιθανότητα κάθε ακμής και προσδιορίζει τις ακμές με τη μικρότερη πιθανότητα ως ανωμαλίες. Η βαθμολογία ανωμαλίας υπολογίζεται με βάση τον μέσο όρο των αποκλίσεων από τις αναμενόμενες πιθανότητες και των δύο κόμβων στην άκρη. Η δεύτερη είναι η ενσωμάτωση δικτύου με επίγνωση ανωμαλιών (AANE), τεχνική που ενημερώνει επαναληπτικά

τις ενσωματώσεις κόμβων και τα αποτελέσματα ανίχνευσης κατά τη διάρκεια της εκπαίδευσης. Προσδιορίζει ανώμαλες ακμές συγκρίνοντας την προβλεπόμενη πιθανότητα σύνδεσης με τη μέση πιθανότητα σύνδεσης που σχετίζεται με τον κόμβο, προσαρμόζοντας τις τυπικές αποκλίσεις

2. Τεχνικές βασισμένες σε συνελικτικό δίκτυο (GCN) όπως η AANE κατηγοριοποιείται εδώ, χρησιμοποιεί επίπεδα GCN για τη δημιουργία ενσωματώσεων κόμβων και έναν πίνακα δείκτη για τον εντοπισμό πιθανών ανώμαλων ακμών, τιμωρώντας τις πιθανότητες πρόβλεψης για ανιχνευμένες ανωμαλίες.
3. Τεχνικές με βάση την αναπαράσταση δικτύου. Η εκμάθηση αναπαράστασης ακμών δημιουργεί αναπαραστάσεις ακμών απευθείας από το γράφημα. Οι αποτελεσματικές αναπαραστάσεις άκρων διατηρούν τη δομή του γραφήματος και το περιεχόμενο αλληλεπίδρασης, οδηγώντας ενδεχομένως σε βελτιωμένη απόδοση ανίχνευσης. Παραδείγματα περιλαμβάνουν έργα των Xu et al., τα οποία δείχνουν πολλά υποσχόμενα αποτελέσματα, αλλά δεν έχουν σχεδιαστεί ειδικά για ανίχνευση ανωμαλιών γραφήματος.
4. Τεχνικές δυναμικού γραφήματος: Η μέθοδος NetWalk κωδικοποιεί ακμές σε έναν λανθάνοντα χώρο χρησιμοποιώντας ενσωματώσεις κόμβων και προσδιορίζει ανωμαλίες με βάση τις αποστάσεις τους από τα πλησιέστερα κέντρα συμπλέγματος ακμών. Ενημερώνει τις αναπαραστάσεις ακμών δυναμικά καθώς εξελίσσεται το γράφημα. Η μέθοδος AddGraph συνδυάζει χρονικές, δομικές και πληροφορίες χαρακτηριστικών, χρησιμοποιεί GCN και Gated Recurrent Units (GRU) με προσοχή για την εκχώρηση βαθμολογιών ανωμαλιών σε ακμές σε δυναμικά γραφήματα. Υποθέτει ότι όλες οι υπάρχουσες ακμές στο δυναμικό γράφημα είναι κανονικές κατά τη διάρκεια της προπόνησης και χρησιμοποιεί δειγματοληπτικά μη υπάρχοντα άκρα ως ανωμαλίες.
5. Τεχνικές μη βαθιάς μάθησης: Τεχνικές όπως αυτές που προτείνουν οι Eswaran και Faloutsos, οι οποίες μοντελοποιούν δυναμικά γραφήματα ως ροές ακμών, εκχωρούν βαθμολογίες ανωμαλιών με βάση τις αλλαγές στη δομή του γραφήματος και τα μοτίβα εμφάνισης ακμών. Αυτές οι μέθοδοι χρησιμοποιούν συχνά στατιστικές μετρήσεις και πιθανοτικά μοντέλα για να χειρίζονται αποτελεσματικά τα edge streams.

Αυτές οι προσεγγίσεις υπογραμμίζουν τις διαφορετικές μεθοδολογίες που χρησιμοποιούνται για την ανίχνευση ανώμαλων ακμών τόσο σε στατικά όσο και σε δυναμικά περιβάλλοντα γραφημάτων, αξιοποιώντας μοντέλα βαθιάς μάθησης, ενσωματώσεις δικτύου και στατιστική ανάλυση για τον εντοπισμό ασυνήθιστων μοτίβων και σχέσεων.

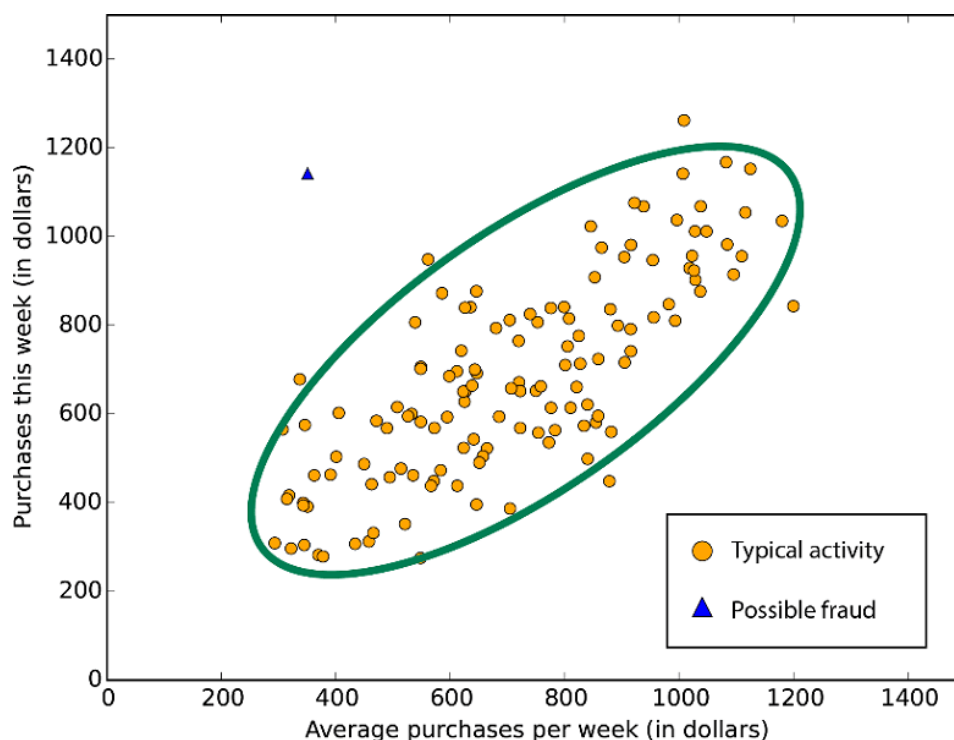
4.4. Γιατί είναι σημαντική

Η σημασία της ανίχνευσης ανωμαλιών εκτείνεται σε διάφορους τομείς όπως έχει αναφερθεί και προηγουμένως. Η πολλαπλή σημαντικότητα της ανίχνευσης δείχνει την κρισιμότητά της. Σε αυτό το τμήμα του κεφαλαίου θα διερευνήσουμε την πολύπλευρη σημασία της ανίχνευσης ανωμαλιών, όπως αυτή προκύπτει από τη βιβλιογραφία.

Στον τομέα της κυβερνοασφάλειας, ο εντοπισμός ανωμαλιών παίζει καθοριστικό ρόλο στον εντοπισμό κακόβουλων δραστηριοτήτων, μη εξουσιοδοτημένης πρόσβασης και πιθανών παραβιάσεων της ασφάλειας. Οι επιθέσεις στον κυβερνοχώρο εκδηλώνονται συχνά ως ανωμαλίες στα μοτίβα κυκλοφορίας δικτύου, στις συμπεριφορές του συστήματος ή στις δραστηριότητες των χρηστών. Ο έγκαιρος εντοπισμός αυτών των ανωμαλιών μπορεί να αποτρέψει παραβιάσεις δεδομένων, οικονομικές απώλειες και ζημιά στη φήμη του οργανισμού. Ο Ahmed et al. (2016) τονίζουν ότι οι τεχνικές ανίχνευσης ανωμαλιών είναι απαραίτητες για τον εντοπισμό επιθέσεων μηδενικής ημέρας και προηγμένων επίμονων απειλών (APTs) που μπορεί να παραλείψουν τα παραδοσιακά συστήματα ανίχνευσης που βασίζονται σε υπογραφές. Με τη συνεχή παρακολούθηση και ανάλυση της κυκλοφορίας του δικτύου, τα συστήματα ανίχνευσης ανωμαλιών μπορούν να εντοπίσουν ασυνήθιστα μοτίβα ενδεικτικά απειλών στον κυβερνοχώρο, επιτρέποντας έτσι έγκαιρες απαντήσεις για τον μετριασμό πιθανών ζημιών (Ahmed, Mahmood, & Hu, 2016).

Στον χρηματοπιστωτικό τομέα, ο εντοπισμός ανωμαλιών είναι ζωτικής σημασίας για τον εντοπισμό δόλιων δραστηριοτήτων, όπως μη εξουσιοδοτημένες συναλλαγές, ξέπλυμα χρήματος και συναλλαγές εμπιστευτικών πληροφοριών. Τα χρηματοπιστωτικά ιδρύματα βασίζονται σε εξελιγμένους αλγόριθμους ανίχνευσης ανωμαλιών για την παρακολούθηση συναλλαγών σε πραγματικό χρόνο και την επισήμανση ύποπτων δραστηριοτήτων για περαιτέρω διερεύνηση. Οι Akoglu et al. (2015) υπογραμμίζουν την αποτελεσματικότητα των μεθόδων ανίχνευσης ανωμαλιών που βασίζονται σε γραφήματα στην αποκάλυψη δόλιων προτύπων σε δίκτυα

συναλλαγών. Αυτές οι μέθοδοι αναλύουν τις σχέσεις μεταξύ οντοτήτων και εντοπίζουν ανωμαλίες που μπορεί να υποδηλώνουν δόλια συστήματα, προστατεύοντας έτσι τα χρηματοοικονομικά συστήματα από σημαντικές απώλειες και διατηρώντας την εμπιστοσύνη μεταξύ των πελατών (Akoglu, Tong, & Koutra, 2015). Ακολουθεί ένα παράδειγμα που δείχνει την συνηθισμένη δραστηριότητα και την περίπτωση πιθανής απάτης σε ένα σύνολο δεδομένων που προκύπτει από γράφημα στο οποίο καταγράφονται οι μέσοι όροι για τις συναλλαγές ανά εβδομάδα και για τη συγκεκριμένη εβδομάδα.



Διάγραμμα 4 - Ανίχνευση πιθανής απάτης

Η ανίχνευση ανωμαλιών είναι επίσης ζωτικής σημασίας στον κλάδο της υγειονομικής περίθαλψης, όπου βοηθά στην έγκαιρη διάγνωση, την παρακολούθηση των ασθενών και τον εντοπισμό ιατρικών λαθών. Ο εντοπισμός ανωμαλιών στα ζωτικά σημεία του ασθενούς, τα εργαστηριακά αποτελέσματα και τα ιατρικά αρχεία μπορεί να οδηγήσει σε πρώιμες παρεμβάσεις, σώζοντας δυνητικά ζωές και βελτιώνοντας τα αποτελέσματα των ασθενών. Οι Chandola et al. (2009) συζητούν πώς χρησιμοποιούνται τεχνικές ανίχνευσης ανωμαλιών για την παρακολούθηση ασθενών

σε μονάδες εντατικής θεραπείας (ΜΕΘ). Αναλύοντας συνεχείς ροές φυσιολογικών δεδομένων, όπως ο καρδιακός ρυθμός και η αρτηριακή πίεση, αυτά τα συστήματα μπορούν να ανιχνεύσουν πρώιμα σημάδια επιδείνωσης και να ειδοποιήσουν το ιατρικό προσωπικό να λάβει προληπτικές ενέργειες (Chandola, Banerjee, & Kumar, 2009).

Στη βιομηχανία και σε συναφείς κλάδους, η ανίχνευση ανωμαλιών είναι απαραίτητη για τη διατήρηση της ποιότητας του προϊόντος και της λειτουργικής αποτελεσματικότητας. Η ανίχνευση ανωμαλιών σε δεδομένα αισθητήρων από μηχανήματα και γραμμές παραγωγής μπορεί να βοηθήσει στον εντοπισμό δυσλειτουργιών του εξοπλισμού, αποκλίσεων στη διαδικασία και πιθανών ελαττωμάτων στα προϊόντα. Οι Hodge και Austin (2004) περιγράφουν πώς εφαρμόζονται μέθοδοι ανίχνευσης ανωμαλιών στην προγνωστική συντήρηση στην κατασκευή. Με τη συνεχή παρακολούθηση της απόδοσης του εξοπλισμού και τον εντοπισμό αποκλίσεων από τις κανονικές συνθήκες λειτουργίας, οι κατασκευαστές μπορούν να προγραμματίσουν τη συντήρηση πριν συμβούν κρίσιμες βλάβες, μειώνοντας έτσι το χρόνο διακοπής λειτουργίας και το σχετικό κόστος (Hodge & Austin, 2004).

Στα κοινωνικά δίκτυα, η ανίχνευση ανωμαλιών βοηθά στον εντοπισμό ανώμαλων συμπεριφορών, όπως ανεπιθύμητη αλληλογραφία, διάδοση παραπληροφόρησης και συντονισμένες μη αυθεντικές δραστηριότητες. Οι πλατφόρμες μέσω κοινωνικής δικτύωσης χρησιμοποιούν αλγόριθμους ανίχνευσης ανωμαλιών για την παρακολούθηση των δραστηριοτήτων και του περιεχομένου των χρηστών για να διατηρήσουν την ακεραιότητα και την αξιοπιστία των πλατφορμών τους. Χρησιμοποιείται όμως και για λόγους ανίχνευσης της συμπεριφοράς καταναλωτών και για λόγους μάρκετινγκ από τις επιχειρήσεις. Οι Aggarwal και Subbian (2012) συζητούν τη σημασία της ανίχνευσης ανωμαλιών στα κοινωνικά ρεύματα για τον εντοπισμό αναδυόμενων γεγονότων, τάσεων και πιθανών απειλών. Αναλύοντας τις αλληλεπιδράσεις και το περιεχόμενο των χρηστών, αυτά τα συστήματα μπορούν να εντοπίσουν ασυνήθιστα μοτίβα που μπορεί να υποδεικνύουν συντονισμένες καμπάνιες ή εξάπλωση επιβλαβούς περιεχομένου (Aggarwal & Subbian, 2012).

Η ανίχνευση ανωμαλιών χρησιμοποιείται επίσης σε ένα ιδιαίτερο ζήτημα της εποχής, στην περιβαλλοντική παρακολούθηση για τον εντοπισμό ασυνήθιστων μοτίβων στα κλιματικά δεδομένα, τις μετρήσεις ποιότητας αέρα και νερού και δείκτες

φυσικών καταστροφών. Η ανίχνευση ανωμαλιών στα περιβαλλοντικά δεδομένα μπορεί να βοηθήσει στα συστήματα έγκαιρης προειδοποίησης και να παροτρύνει τις απαραίτητες ενέργειες για τον μετριασμό των επιπτώσεων των φυσικών καταστροφών. Οι Chen και Neill (2014) υπογραμμίζουν την εφαρμογή της ανίχνευσης ανωμαλιών στην παρακολούθηση δεδομένων δημόσιας υγείας και περιβάλλοντος. Η εργασία τους δείχνει πώς οι στατιστικές μη παραμετρικής σάρωσης μπορούν να χρησιμοποιηθούν για τον εντοπισμό αναδυόμενων συστάδων εστιών ασθενειών ή συμβάντων ρύπανσης, επιτρέποντας έγκαιρες απαντήσεις στη δημόσια υγεία (Chen & Neill, 2014).

Συμπερασματικά, ο εντοπισμός ανωμαλιών είναι μια κρίσιμη διαδικασία σε διάφορους τομείς, παρέχοντας σημαντικά οφέλη για την ενίσχυση της κυβερνοασφάλειας, τον εντοπισμό χρηματοοικονομικής απάτης, τη βελτίωση των αποτελεσμάτων της υγειονομικής περίθαλψης, τη διασφάλιση της ποιότητας παραγωγής, την παρακολούθηση των κοινωνικών δικτύων και τη διευκόλυνση της περιβαλλοντικής παρακολούθησης. Εντοπίζοντας αποκλίσεις από τα κανονικά πρότυπα, τα συστήματα ανίχνευσης ανωμαλιών επιτρέπουν έγκαιρες παρεμβάσεις, αποτρέπουν πιθανές απειλές και ενισχύουν τις διαδικασίες λήψης αποφάσεων. Καθώς τα δεδομένα συνεχίζουν να αυξάνονται σε όγκο και πολυπλοκότητα, η σημασία των αποτελεσματικών μεθόδων ανίχνευσης ανωμαλιών θα αυξηθεί μόνο, οδηγώντας σε προόδους σε αλγόριθμους και εφαρμογές σε πολλαπλά πεδία.

5. Πειραματική διαδικασία

5.1. Εισαγωγή

Ο βασικός στόχος του πρακτικού μέρους της διπλωματικής εργασίας είναι η εξοικείωση και ανάλυση μοντέλων εντοπισμού γεγονότων (event detection) σε δυναμικά γραφήματα, με βάση την πρότυπη υλοποίηση που διατίθεται στο (<https://github.com/mertkosan/DyGED?tab=readme-ov-file>),. Σημειώνεται ότι ο αρχικός κώδικας, συμπεριλαμβανομένων των αρχείων `models.py` και `layers.py` καθώς και των δύο βάσεων δεδομένων του πρώτου μέρους, τροποποιήθηκε και εμπλουτίστηκε με διαδικασίες εκπαίδευσης (train) και αξιολόγησης (test) για την εκπλήρωση των απαιτήσεων της μελέτης μας.

Το πειραματικό μέρος της εργασίας διαρθρώνεται σε δύο κύριες φάσεις:

- Στην πρώτη φάση, πραγματοποιείται ανάλυση των διαθέσιμων μοντέλων εντοπισμού γεγονότων και των δύο βάσεων δεδομένων που προτείνονται στην πρότυπη υλοποίηση του DyGED. Στη συνέχεια, ο κώδικας επεκτάθηκε ώστε να περιλαμβάνει μετρήσεις απόδοσης (AUC) και χρονομετρήσεις, προσφέροντας μια πλήρη διαδικασία εκπαίδευσης και αξιολόγησης των μοντέλων.
- Στη δεύτερη φάση, διερευνάται η απόδοση των μοντέλων σε δυναμικά σύνολα δεδομένων που δημιουργήθηκαν με τη βιβλιοθήκη RDyn, εστιάζοντας στην ανάλυση της συμπεριφοράς τους υπό συνθήκες εξελισσόμενων δομών και ανωμαλιών, όπως οι συγχωνεύσεις (merge) και οι διαχωρισμοί (split) κοινοτήτων.

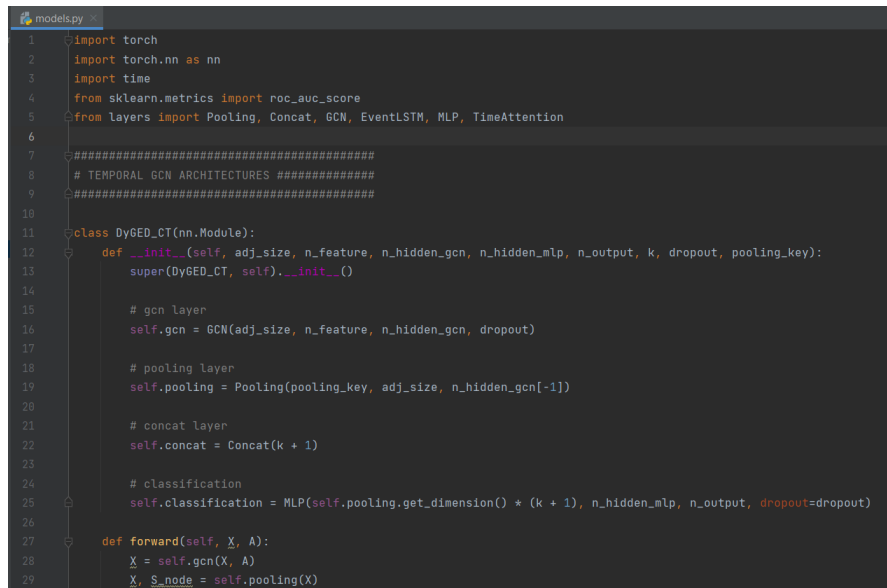
Σημειώνεται ότι στο αποθετήριο GitHub της δικής μας υλοποίησης, παρατίθεται ο πλήρης και τροποποιημένος κώδικας, χωρίς όμως να περιλαμβάνονται τα αρχεία που παρέχονται στην πρότυπη υλοποίηση του DyGED.

Επιπλέον, αξίζει να αναφερθεί ότι η πειραματική διαδικασία πραγματοποιήθηκε σε περιβάλλον PyCharm. Όπως είναι προφανές με βάση τα παραπάνω, η γλώσσα υλοποίησης είναι η Python, και απαραίτητες βιβλιοθήκες για την επιτυχή εκτέλεση του κώδικα είναι οι παρακάτω:

- **torch**: Η βιβλιοθήκη torch αποτελεί ένα ισχυρό εργαλείο για τη δημιουργία και εκπαίδευση μοντέλων μηχανικής μάθησης, ιδιαίτερα νευρωνικών δικτύων. Επιπλέον, παρέχει ενσωμάτωση με GPU για ταχύτερους υπολογισμούς, γεγονός ιδανικό για εφαρμογές με μεγάλα δεδομένα και περίπλοκες αρχιτεκτονικές.
- **time**: Η συγκεκριμένη βιβλιοθήκη παρέχει εργαλεία για την καταμέτρηση χρονικών διαστημάτων. Επιτρέπει τον εύκολο χειρισμό timestamps, την αναστολή εκτέλεσης ενός προγράμματος για καθορισμένο χρόνο με τη λειτουργία `sleep()`, καθώς και τη μέτρηση του χρόνου εκτέλεσης κώδικα μέσω της συνάρτησης `time()`. Αυτές οι δυνατότητες καθιστούν τη βιβλιοθήκη

πολύτιμη για προγραμματιστές που χρειάζονται ακριβείς χρονικούς ελέγχους στα προγράμματά τους.

- **sklearn:** Η βιβλιοθήκη sklearn (scikit-learn) είναι ένα δημοφιλές εργαλείο για την υλοποίηση αλγορίθμων μηχανικής μάθησης στην Python. Παρέχει εύκολες στη χρήση συναρτήσεις για προεπεξεργασία δεδομένων, εκπαίδευση και αξιολόγηση μοντέλων, όπως ταξινομητές, αλγόριθμους συστάσεων και μοντέλα παλινδρόμησης.



```
1 import torch
2 import torch.nn as nn
3 import time
4 from sklearn.metrics import roc_auc_score
5 from layers import Pooling, Concat, GCN, EventLSTM, MLP, TimeAttention
6
7 #####
8 # TEMPORAL GCN ARCHITECTURES #####
9 #####
10
11 class DyGED_CT(nn.Module):
12     def __init__(self, adj_size, n_feature, n_hidden_gcn, n_hidden_mlp, n_output, k, dropout, pooling_key):
13         super(DyGED_CT, self).__init__()
14
15         # gcn layer
16         self.gcn = GCN(adj_size, n_feature, n_hidden_gcn, dropout)
17
18         # pooling layer
19         self.pooling = Pooling(pooling_key, adj_size, n_hidden_gcn[-1])
20
21         # concat layer
22         self.concat = Concat(k + 1)
23
24         # classification
25         self.classification = MLP(self.pooling.get_dimension() * (k + 1), n_hidden_mlp, n_output, dropout=dropout)
26
27     def forward(self, X, A):
28         X = self.gcn(X, A)
29         X, S_node = self.pooling(X)
```

Εικόνα 1: Περιβάλλον PyCharm

5.2. Πρώτο Μέρος Πειραματικής Διαδικασίας

Ο στόχος αυτού του πρώτου πειραματικού μέρους είναι η αρχική αξιολόγηση και ανάλυση των τεσσάρων μοντέλων DyGED, DyGED_CT, DyGED_NL και DyGED_NA για την ανίχνευση γεγονότων σε δυναμικά γραφήματα. Κάθε μοντέλο αξιοποιεί χρονικά και δομικά χαρακτηριστικά για να εντοπίζει ανωμαλίες, όπως συγχωνεύσεις και διαχωρισμούς κοινοτήτων, σε εξελισσόμενα περιβάλλοντα. Για την αξιολόγηση της απόδοσής τους, χρησιμοποιούνται τα σύνολα δεδομένων [NYC Cab](#) και [Twitter Weather](#), τα οποία παρέχουν δυναμικές πληροφορίες για μεταβαλλόμενες σχέσεις μεταξύ οντοτήτων. Η εκπαίδευση και εκτέλεση των μοντέλων γίνεται με διάκριση των δεδομένων σε train και test sets, ενώ ο πλήρης κώδικας αυτής της διαδικασίας διατίθεται στο GitHub αποθετήριο. Τα αποτελέσματα αξιολογούνται με τις μετρικές AUC score και χρόνο εκτέλεσης προσφέροντας σημαντικά δεδομένα για την αποτελεσματικότητα και απόδοση κάθε μοντέλου.

5.2.1. Χρησιμοποιούμενα Μοντέλα

5.2.1.1. Μοντέλο DyGED_CT

Το μοντέλο DyGED_CT αντιπροσωπεύει ένα νευρωνικό δίκτυο για την ενσωμάτωση δυναμικών γραφημάτων με χρονικό πλαίσιο. Ακολουθεί η ανάλυση των στρωμάτων από τα οποία αποτελείται:

- **Στρώμα GCN:** Το στρώμα GCN (Graph Convolutional Network) χρησιμοποιείται για την επεξεργασία των εισόδων X και του πίνακα γειτνίασης A ώστε να παραχθούν ενσωματώσεις κόμβων που λαμβάνουν υπόψη τη δομή του γραφήματος. Το GCN μαθαίνει αναπαραστάσεις κόμβων συλλέγοντας πληροφορίες από τους γειτονικούς κόμβους.
- **Στρώμα Pooling :** Αυτό το στρώμα εφαρμόζει μια λειτουργία pooling, η οποία μειώνει τη διάσταση της αναπαράστασης του γραφήματος. Ο τύπος του pooling (καθορίζεται από το `pooling_key`) ορίζει πώς οι ιδιότητες των κόμβων συγκεντρώνονται. Η έξοδος περιλαμβάνει τα pooled χαρακτηριστικά των κόμβων X και ένα σκορ δομής S_{node} .
- **Στρώμα Συγχώνευσης:** Το στρώμα συγχώνευσης συνδυάζει πολλαπλά χρονικά βήματα ενσωματώσεων κόμβων σε μία ενιαία αναπαράσταση.
- **Στρώμα Ταξινόμησης:** Το τελικό στρώμα είναι ένα MLP (Multi-Layer Perceptron) που χρησιμοποιείται για την ταξινόμηση. Λαμβάνει την ενσωματωμένη αναπαράσταση και την επεξεργάζεται για να παράγει την τελική έξοδο `outs`.

5.2.1.2. Μοντέλο DyGED_NL

Το μοντέλο DyGED_NL προσθέτει ένα στρώμα προσοχής χρόνου, το οποίο επιτρέπει στο μοντέλο να εστιάζει σε σημαντικές χρονικές στιγμές. Περιέχει τα παρακάτω στρώματα:

- **Στρώμα GCN:** Όπως και στο DyGED_CT, το GCN μαθαίνει ενσωματώσεις κόμβων βάσει της δομής του γραφήματος.
- **Στρώμα Pooling:** Όπως και προηγουμένως, εφαρμόζεται pooling για τη μείωση της διάστασης των δεδομένων.

- **Στρώμα Προσοχής Χρόνου:** Αυτό το στρώμα χρησιμοποιείται για να εντοπίσει σημαντικές χρονικές στιγμές στα δεδομένα. Ουσιαστικά, το μοντέλο μαθαίνει να δίνει μεγαλύτερη σημασία σε συγκεκριμένες χρονικές στιγμές.
- **Στρώμα Ταξινόμησης:** Τέλος, η έξοδος του στρώματος προσοχής περνάει από το MLP για να παραχθεί η τελική πρόβλεψη.

5.2.1.3. Μοντέλο DyGED_NA

Το μοντέλο DyGED_NA προσθέτει ένα LSTM στρώμα για να αποτυπώσει τη διαχρονική δυναμική των γραφημάτων.

- **Στρώμα GCN:** Όπως και στα άλλα μοντέλα, μαθαίνει ενσωματώσεις βάσει των γειτονικών κόμβων.
- **Στρώμα Pooling:** Εφαρμόζεται pooling για τη μείωση της διάστασης των δεδομένων.
- **Στρώμα LSTM:** Το LSTM στρώμα επεξεργάζεται τις ενσωματώσεις για να αποτυπώσει διαχρονικές σχέσεις και δυναμική στα δεδομένα του γραφήματος.
- **Στρώμα Ταξινόμησης:** Η τελική αναπαράσταση περνάει από το MLP για την τελική πρόβλεψη.

5.2.1.4. Μοντέλο DyGED

Το μοντέλο DyGED είναι η πιο ολοκληρωμένη εκδοχή, συνδυάζοντας LSTM και προσοχή χρόνου και περιέχει:

- **Στρώμα GCN:** Μαθαίνει τις ενσωματώσεις των κόμβων από το γράφημα.
- **Στρώμα Pooling :** Εφαρμόζει pooling για τη μείωση της διάστασης.
- **Στρώμα LSTM:** Επεξεργάζεται τις ενσωματώσεις για να αποτυπώσει διαχρονικές σχέσεις.
- **Στρώμα Προσοχής Χρόνου:** Εστιάζει σε σημαντικές χρονικές στιγμές στα δεδομένα.
- **Στρώμα Ταξινόμησης:** Τελικά, η αναπαράσταση περνάει από το MLP για την πρόβλεψη.

5.2.2. Χρησιμοποιούμενα Σύνολα Δεδομένων

Για την αρχική αξιολόγηση των 4 μοντέλων, χρησιμοποιήθηκαν δύο σύνολα δεδομένων, τα οποία δίνονται στην πρότυπη υλοποίηση:

- NYC Cab
- Twitter Weather

Όσο αφορά το πρώτο σύνολο δεδομένων (NYC Cab), πρόκειται για ένα σύνολο δεδομένων που βασίζεται σε αθλητικές εκδηλώσεις και τον ωριαίο αριθμό επιβατών που μεταφέρονται μεταξύ ζωνών ταξί (taxi zones) στη Νέα Υόρκη. Στο σύνολο αυτό, οι κόμβοι αναπαριστούν τις ζώνες ταξί, οι ακμές αντιπροσωπεύουν τα ταξίδια μεταξύ αυτών των ζωνών και τα βάρη τους εκφράζουν τον αριθμό των επιβατών που μεταφέρονται. Τα στατικά χαρακτηριστικά των κόμβων περιλαμβάνουν τις γεωγραφικές συντεταγμένες, τους δήμους, τα μήκη, τις επιφάνειες και τις ζώνες εξυπηρέτησης. Από την άλλη, τα δυναμικά χαρακτηριστικά περιλαμβάνουν τον βαθμό του κόμβου, την κεντρικότητα ενδιαμεσότητας (betweenness centrality) και τον συντελεστή συσταδοποίησης, τα οποία υπολογίζονται ανά χρονική περίοδο μίας ώρας. Το πιο σημαντικό που αφορά το πρακτικό μέρος της παρούσας εργασίας είναι ότι οι αθλητικοί αγώνες μπέιζμπολ που περιλαμβάνουν τους Yankees ή τους Mets στη Νέα Υόρκη αποτελούν τα κύρια γεγονότα ενδιαφέροντος.

Όσο αφορά το δεύτερο σύνολο δεδομένων (Twitter Weather), πρόκειται για ένα σύνολο δεδομένων που αποτελείται από tweets σχετικά με τον καιρό και σημαντικά καιρικά φαινόμενα στις ΗΠΑ από το 2012 έως το 2018. Τα tweets έχουν εξαχθεί από ένα μεγάλο σύνολο δεδομένων που παρέχεται από το Internet Archive. Οι κόμβοι αντιπροσωπεύουν αγγλικές λέξεις, οι ακμές αντιστοιχούν σε συμπτώσεις λέξεων και τα βάρη των ακμών υποδηλώνουν τον αριθμό αυτών των συμπτώσεων. Ξεκινώντας από ένα μικρό σύνολο λέξεων σχετικών με τον καιρό, χρησιμοποιήθηκε ένας αλγόριθμος για να επεκταθεί το σύνολο σε 300 λέξεις. Επιπλέον, έχουν εφαρμοστεί προεκπαιδευμένα vectors από το word2vec ως στατικά χαρακτηριστικά των κόμβων, ενώ τα δυναμικά χαρακτηριστικά περιλαμβάνουν τον βαθμό του κόμβου, την κεντρικότητα ενδιαμεσότητας (betweenness centrality) και τον συντελεστή συσταδοποίησης, που υπολογίζονται σε χρονικά διαστήματα μίας ημέρας. Αυτό που έχει σημασία για το πρακτικό μέρος της παρούσας εργασίας είναι ότι τα γεγονότα προς εντοπισμό αφορούν καιρικά φαινόμενα.

Επομένως, τα μοντέλα τα οποία θα αναλυθούν και θα εξεταστούν πλέον ως προς τη συμπεριφορά τους σε δύο πραγματικά σύνολα δεδομένων, ακολουθούν λεπτομέρειες για τα σύνολα στον παρακάτω πίνακα:

Χαρακτηριστικό	NYC Cab	Twitter Weather
Nodes (μέσος όρος)	263	300
Edges (μέσος όρος)	3717	1142
Static Features	6	300
Dynamic Features	3	3
Snapshots	4464	2557
Events	162	287
Είδος Γεγονότων	Αγώνες μπίιζμπολ	Ακραία καιρικά φαινόμενα

Πίνακας 1: Χαρακτηριστικά και είδος γεγονότων προς εντοπισμό ανά σύνολο δεδομένων

Όσο αφορά τις μετρικές με βάση τις οποίες αξιολογήθηκαν τα 4 μοντέλα πάνω στα δύο σύνολα δεδομένων είναι:

- Χρόνος εκτέλεσης
- Μετρική AUC

Επίσης γίνεται σαφές από την αρχή ότι δεν είναι σκοπός να μιμηθεί η παρούσα πειραματική διαδικασία την πειραματική διαδικασία πάνω στην οποία βασίστηκε το paper της πρότυπης υλοποίησης. Σκοπός είναι να συμπληρωθεί και επεκταθεί ο διαθέσιμος κώδικας, ώστε να μπορούν να εκτελεστούν ολοκληρωμένα σενάρια εκτέλεσης ώστε να μπορούν να προκύπτουν συγκριτικά αποτελέσματα.

Τέλος, αναφέρεται ότι στα παρεχόμενα αποτελέσματα, παρατίθενται οι καλύτερες τιμές των μετρικών μετά από πολλαπλές εκτελέσεις.

5.2.3. Χρησιμοποιούμενος Κώδικας

Ο ολοκληρωμένος κώδικας για τα αρχεία *models_NYC.py* και *models_TW.py*, που περιλαμβάνει την υλοποίηση των τεσσάρων μοντέλων DyGED και τις παραλλαγές τους, βρίσκεται στο [GitHub αποθετήριο](#). Ακολουθεί ενδεικτικό τμήμα κώδικα για τη δημιουργία των μοντέλων αυτών.

```
models = {
    "DyGED_CT": DyGED_CT(adj_size=num_nodes, n_feature=1, n_hidden_gcn=(num_hidden,),
n_hidden_mlp=(num_hidden,),
                        n_output=1, k=0, dropout=dropout_rate, pooling_key='expert'),
    "DyGED_NL": DyGED_NL(adj_size=num_nodes, n_feature=1, n_hidden_gcn=(num_hidden,),
attention_expert=1,
                        n_hidden_mlp=(num_hidden,), n_output=1, k=0, dropout=dropout_rate, pooling_key='expert'),
    "DyGED_NA": DyGED_NA(adj_size=num_nodes, n_feature=1, n_hidden_gcn=(num_hidden,),
n_hidden_lstm=(num_hidden,),
                        n_hidden_mlp=(num_hidden,), n_output=1, dropout=dropout_rate, pooling_key='expert'),
    "DyGED": DyGED(adj_size=num_nodes, n_feature=1, n_hidden_gcn=(num_hidden,),
n_hidden_mlp=(num_hidden,),
                  n_hidden_lstm=(num_hidden,),
                  attention_expert=1, n_output=1, k=0, dropout=dropout_rate, pooling_key='expert')
```

Ακολουθεί το τμήμα του κώδικα για την φόρτωση των δεδομένων. Στην συγκεκριμένη περίπτωση φορτώνονται 3 αρχεία δεδομένων:

- **graphs.pt** : Περιέχει την δομή των γραφημάτων
- **dynamic_attrs.pt**: Περιέχει πληροφορίες για τις δυναμικές ιδιότητες των γραφημάτων
- **events.pt**: Περιέχει τις ετικέτες (θα χρησιμοποιηθούν για την σύγκριση με τις προβλεφθείσες τιμές και τον υπολογισμό της μετρικής AUC)

Αξιοσημείωτο είναι επίσης ότι τα 3 παραπάνω αρχεία βρίσκονται και στα 2 σύνολα δεδομένων και ότι όλα τα δεδομένα φορτώνονται σε tensors.

```
#Φόρτωση των αρχείων που έχουν δημιουργηθεί από το custom_data_format.py
G = torch.load('data/nyc_cab/raw/graphs.pt')
G = G.type(torch.FloatTensor)

X = torch.load('data/nyc_cab/raw/dynamic_attrs.pt')
X = X.type(torch.FloatTensor)

y_true = torch.load('data/nyc_cab/raw/events.pt')
y_true = y_true.type(torch.FloatTensor)
```

Ακολουθεί μια αναλυτική περιγραφή της διαδικασίας εκπαίδευσης και αξιολόγησης των μοντέλων DyGED_CT, DyGED_NL, DyGED_NA, και DyGED, συμπεριλαμβανομένων των ρυθμίσεων και των τεχνικών επιλογών που αποτελούν το υπόλοιπο κομμάτι του κώδικα :

1. **Διαχωρισμός Δεδομένων:** Αρχικά, τα δεδομένα διαχωρίζονται σε training και testing sets (π.χ. `train_ratio = 0.8334`), χωρίς ανακάτεμα, για να διατηρηθεί η αλληλουχία των στιγμιοτύπων. Έτσι, το μοντέλο προσαρμόζεται σε δεδομένα από το παρελθόν και αξιολογείται σε δεδομένα από το μέλλον, προσομοιώνοντας ένα πραγματικό σενάριο πρόβλεψης.

2. **Ρυθμίσεις Εκπαίδευσης:**

- **Learning Rate (0.005):** Το learning rate καθορίζει το βήμα προσαρμογής των βαρών στη διάρκεια της εκπαίδευσης. Η τιμή 0.005 επιλέγεται ώστε το μοντέλο να προσαρμόζεται σταδιακά, αποφεύγοντας μεγάλες αλλαγές που μπορεί να οδηγήσουν σε αστάθεια.
- **Dropout Rate (0.02):** Το dropout rate είναι μια τεχνική κανονικοποίησης (regularization), όπου με πιθανότητα 0.02 "απενεργοποιούνται" τυχαία ορισμένοι neurons (νευρώνες) σε κάθε βήμα εκπαίδευσης, συμβάλλοντας στην αποφυγή υπερπροσαρμογής (overfitting) και βοηθώντας το μοντέλο να γενικεύει καλύτερα σε μη ορατά δεδομένα.
- **num_hidden (64):** Ο αριθμός των νευρώνων στα κρυφά επίπεδα του μοντέλου. Εδώ, έχουμε 64 νευρώνες σε κάθε κρυφό επίπεδο του νευρωνικού δικτύου.
- **Αριθμός Εποχών (Epochs) :** Η εκπαίδευση γίνεται για όσες εποχές έχουμε ορίσει, ώστε το μοντέλο να έχει αρκετές ευκαιρίες να μάθει τα μοτίβα στα δεδομένα, ενώ αποφεύγεται η υπερβολική προσαρμογή που μπορεί να προκαλέσει overfitting.

3. Βελτιστοποίηση με Adam Optimizer:

Ο Adam (Adaptive Moment Estimation) είναι ένας ευρέως χρησιμοποιούμενος αλγόριθμος βελτιστοποίησης που συνδυάζει τα χαρακτηριστικά του Stochastic Gradient Descent (SGD) και της προσαρμοστικής προσαρμογής βήματος (adaptive learning rates). Βασίζεται στην ταυτόχρονη παρακολούθηση τόσο της μέσης τιμής όσο και της διακύμανσης των gradients, κάνοντας την εκπαίδευση πιο αποδοτική και σταθερή.

4. Συνάρτηση Απώλειας (Binary Cross-Entropy with Sigmoid):

- **Binary Cross-Entropy Loss:** Η συνάρτηση απώλειας που χρησιμοποιείται είναι η binary cross-entropy, η οποία είναι κατάλληλη για δυαδικά προβλήματα, όπως η ανίχνευση γεγονότων (events) ή μη. Η binary cross-entropy μετρά την απόκλιση μεταξύ των προβλέψεων του μοντέλου και των πραγματικών ετικετών, με στόχο τη μεγιστοποίηση της ακρίβειας στην ταξινόμηση.
- **Sigmoid Activation:** Η συνάρτηση ενεργοποίησης sigmoid χρησιμοποιείται για να μετατρέψει τις εξόδους του μοντέλου σε πιθανότητες (από 0 έως 1), διευκολύνοντας έτσι την ερμηνεία της πρόβλεψης κάθε κόμβου.

5. Αξιολόγηση Αποτελεσμάτων:

- **AUC Score:** Για την αξιολόγηση της απόδοσης των μοντέλων χρησιμοποιείται το AUC (Area Under the Curve), που μετρά την ακρίβεια των προβλέψεων ανεξαρτήτως ορίου. Ένα υψηλότερο AUC υποδεικνύει καλύτερη διάκριση μεταξύ των θετικών και αρνητικών δειγμάτων.
- **Χρονική Μέτρηση:** Καταγράφεται ο συνολικός χρόνος από την έναρξη της εκπαίδευσης μέχρι την ολοκλήρωση της αξιολόγησης, δίνοντας μια εικόνα της αποδοτικότητας και ταχύτητας του μοντέλου για διαφορετικά μεγέθη δεδομένων και ρυθμίσεις.

Με αυτές τις ρυθμίσεις και τις τεχνικές επιλογές, τα μοντέλα DyGED δοκιμάζονται και αξιολογούνται, παρέχοντας μια πρώτη ανάλυση της αποτελεσματικότητας και των επιδόσεών τους στη συγκεκριμένη εφαρμογή. Βρίσκουμε ότι η εκπαίδευση χρησιμοποιώντας την βελτιστοποίηση Adam με ρυθμό μάθησης, ποσοστό dropout και αριθμό νευρώνων στα κρυφά επίπεδα να είναι 0.005, 0.02 και 64 αντίστοιχα, λειτουργεί καλά για τις μεθόδους μας. Για αυτό χρησιμοποιούνται τόσο στο πρώτο όσο και στο δεύτερο πειραματικό κομμάτι.

5.2.4. Πειραματικά Αποτελέσματα NYC Cab

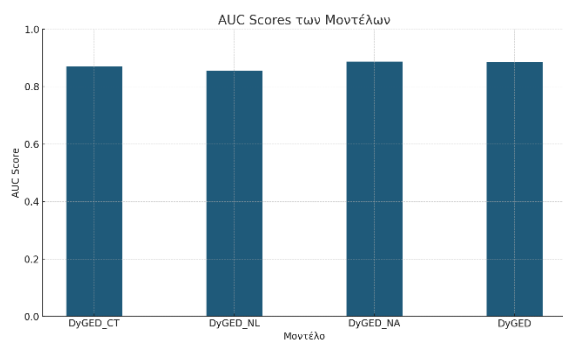
Αρχικά, παρατίθεται ο πίνακας με τους χρόνους εκτέλεσης των τεσσάρων μοντέλων (με αριθμό επαναλήψεων του train = 100) καθώς και τα αντίστοιχα συγκριτικά αποτελέσματα ως προς τη μετρική AUC, τα οποία προκύπτουν από τον μέσο όρο πολλαπλών εκτελέσεων.

Μοντέλο	AUC	Χρόνος Εκτέλεσης (seconds)
DyGED_CT	0.8696	135.28
DyGED_NL	0.8560	142.84
DyGED_NA	0.8873	147.14
DyGED	0.8857	149.08

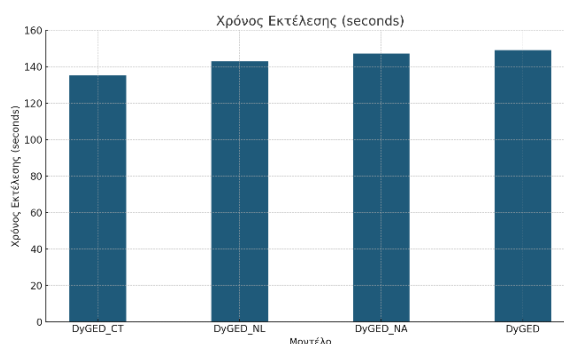
Πίνακας 2: Χρόνοι εκτέλεσης και Μετρική AUC - NYC Cab

Από τον παραπάνω πίνακα, διαπιστώνεται ότι οι τιμές της μετρικής AUC των μοντέλων είναι αρκετά υψηλές με καλύτερη αυτή του DyGED_NA. Επίσης διαπιστώνεται ότι το DyGED που έχει τα περισσότερα layers έχει το μεγαλύτερο χρόνο εκτέλεσης.

Ακολουθούν και τα αντίστοιχα γραφήματα για καλύτερη οπτικοποίηση των αποτελεσμάτων:



Εικόνα 2: Σύνολο δεδομένων NYC Cab – Μετρική AUC



Εικόνα 3: Σύνολο δεδομένων NYC Cab - Χρόνοι Εκτέλεσης

5.2.5. Πειραματικά Αποτελέσματα Twitter Weather

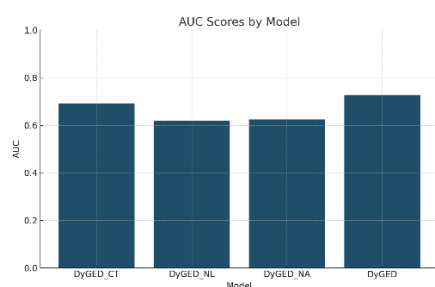
Όπως και στο προηγούμενο σύνολο δεδομένων, παρουσιάζονται οι συγκριτικοί χρόνοι (με αριθμό επαναλήψεων του train = 50), καθώς και τα αντίστοιχα συγκριτικά αποτελέσματα ως προς τη μετρική AUC. Τα αποτελέσματα προκύπτουν από τον μέσο όρο πολλαπλών εκτελέσεων.

Μοντέλο	AUC	Χρόνος Εκτέλεσης (seconds)
DyGED_CT	0.6926	46.99
DyGED_NL	0.6203	46.83
DyGED_NA	0.6245	47.10
DyGED	0.7257	49.63

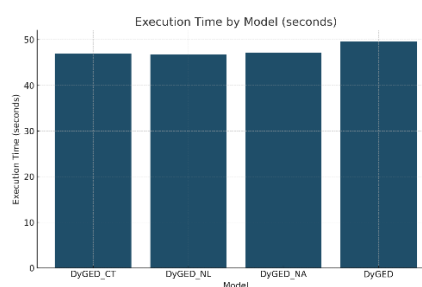
Πίνακας 3: Χρόνοι εκτέλεσης και Μετρική AUC - Twitter Weather

Από το παραπάνω πίνακα, παρατηρείται παρόμοια συμπεριφορά με το προηγούμενο σύνολο δεδομένων όσο αφορά τον χρόνο εκτέλεσης, με το μοντέλο DyGED να απαιτεί τον περισσότερο χρόνο εκτέλεσης. Στο συγκεκριμένο σύνολο δεδομένων το μοντέλο DyGED έχει την καλύτερη τιμή στην μετρική AUC .

Ακολουθούν και τα αντίστοιχα γραφήματα :



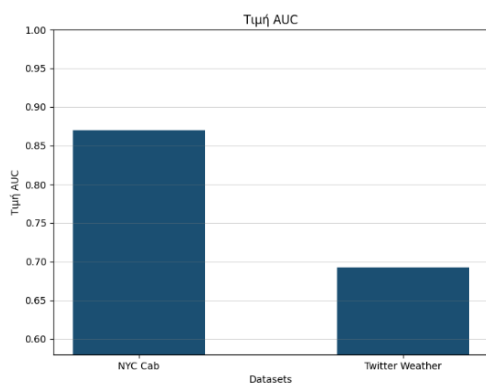
Εικόνα 4:Σύνολο δεδομένων Twitter Weather – Μετρική AUC



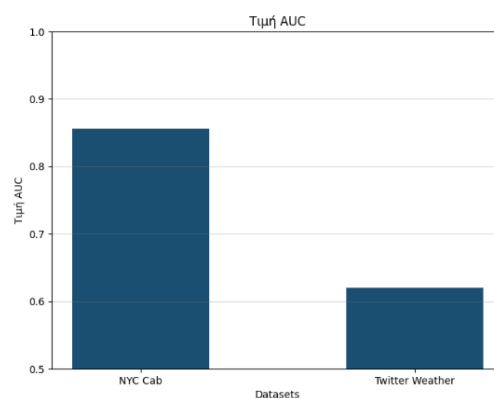
Εικόνα 5:Σύνολο δεδομένων Twitter Weather - Χρόνοι Εκτέλεσης

5.2.6. Συγκριτικά Αποτελέσματα

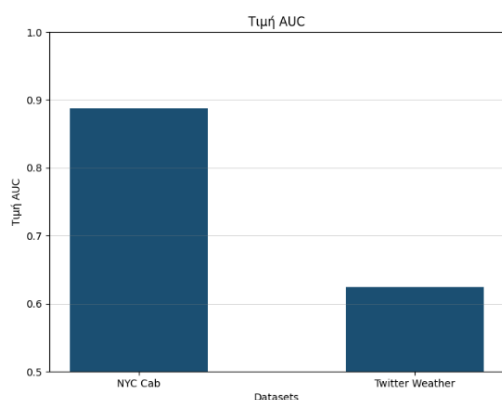
Στα πλαίσια της συγκεκριμένης ενότητας, θα παρατεθούν συγκριτικά αποτελέσματα όσο αφορά τη μετρική AUC για τα 4 μοντέλα (για τα δύο σύνολα δεδομένων):



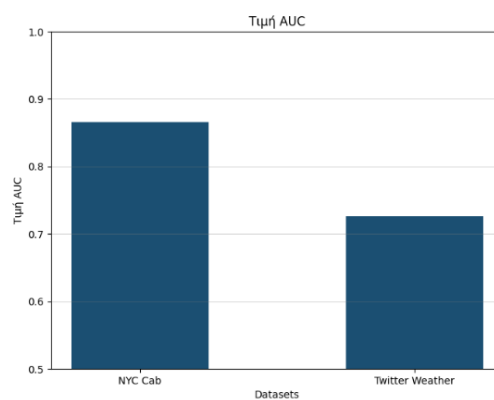
Εικόνα 6: Συγκριτικά αποτελέσματα μοντέλου DyGED_CT



Εικόνα 7: Συγκριτικά αποτελέσματα μοντέλου DyGED_NL



Εικόνα 8: Συγκριτικά αποτελέσματα μοντέλου DyGED_NA



Εικόνα 9: Συγκριτικά αποτελέσματα μοντέλου DyGED

Από τα παραπάνω συγκριτικά γραφήματα συμπεραίνεται ότι στην περίπτωση του συνόλου δεδομένων NYC Cab καταγράφηκε καλύτερη τιμή της μετρικής AUC για όλα τα μοντέλα συγκριτικά με τον σύνολο δεδομένων Twitter Weather.

5.2.7. Αξιολόγηση με F1 Score: Προκλήσεις και Αποτελέσματα

Στην ενότητα αυτή, επιχειρούμε την ανάλυση της απόδοσης των μοντέλων με τη χρήση της **μετρικής F1**, που αποτελεί βασικό δείκτη για την ισορροπία μεταξύ **Precision** και **Recall**. Η μετρική αυτή αναδεικνύει σημαντικές προκλήσεις στις βάσεις δεδομένων **NYC Cab** και **Twitter Weather**, ενώ δοκιμάστηκαν διάφοροι συνδυασμοί παραμέτρων για τη βελτίωση της απόδοσης. Ο κώδικας για την εφαρμογή της f1 στο **NYC Cab** και **Twitter Weather** βρίσκεται στο [models_f1.py](#).

1. ΕΙΣΑΓΩΓΗ

Η προσθήκη της F1 Score στην πειραματική διαδικασία έχει στόχο να αναδείξει την ποιότητα των τελικών προβλέψεων των μοντέλων. Ενώ το **AUC (Area Under the Curve)** μετρά την ικανότητα των μοντέλων να διαχωρίζουν θετικά και αρνητικά γεγονότα ανεξαρτήτως threshold, το **F1 Score** εξαρτάται από την επιλογή του threshold και την ισορροπία μεταξύ Precision και Recall.

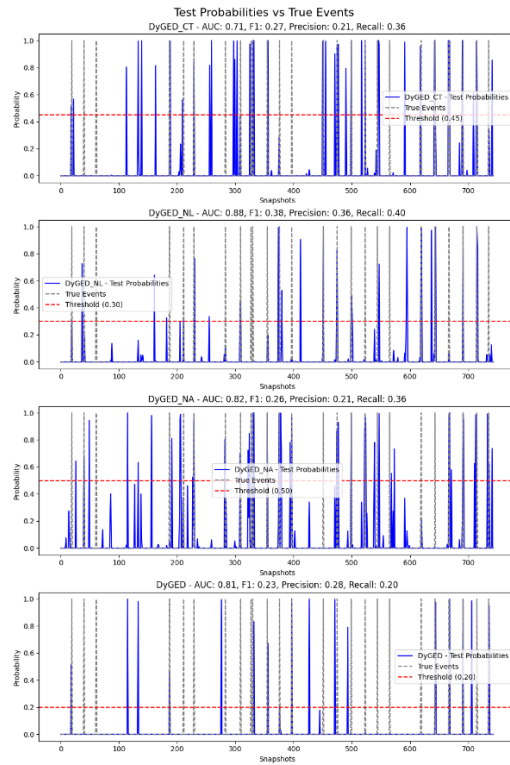
Στις συγκεκριμένες βάσεις δεδομένων, παρατηρούνται χαμηλές τιμές F1, ακόμα και σε περιπτώσεις όπου το AUC παραμένει υψηλό. Αυτό υποδεικνύει ότι οι τελικές προβλέψεις επηρεάζονται από:

- Την ανισορροπία των δεδομένων.
- Την ευαισθησία των πιθανοτήτων στο threshold.
- Τη δυναμική φύση των δεδομένων.

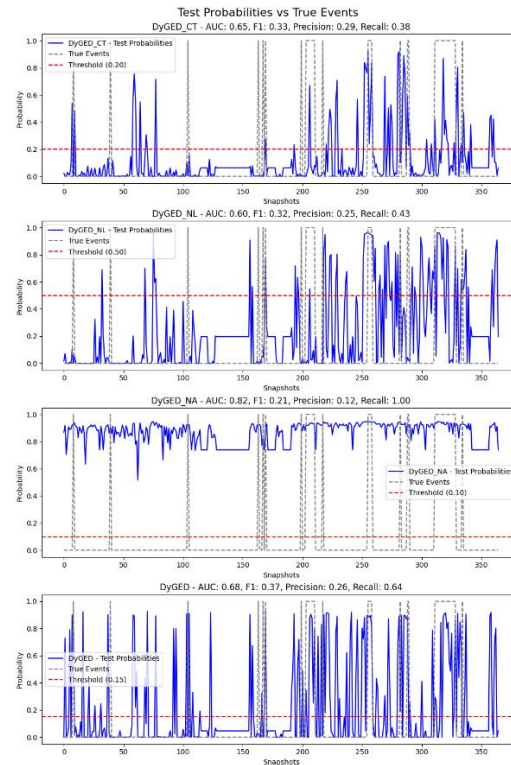
Παρακάτω παρουσιάζονται τα αποτελέσματα, συνοδευόμενα από γραφήματα που απεικονίζουν την απόδοση των μοντέλων. Οι καλύτερες επιδόσεις κάθε μοντέλου για τα δύο σύνολα δεδομένων συνοψίζονται στον συγκεντρωτικό πίνακα που ακολουθεί:

Μοντέλο	Βάση Δεδομένων	F1	Threshold	Precision	Recall	AUC	Learning Rate	Epochs
DyGED_CT	NYC Cab	0.27	0.45	0.21	0.36	0.71	0.001	150
DyGED_CT	Twitter Weather	0.33	0.20	0.29	0.38	0.65	0.005	50
DyGED_NL	NYC Cab	0.38	0.30	0.36	0.40	0.88	0.005	100
DyGED_NL	Twitter Weather	0.32	0.50	0.25	0.43	0.60	0.005	150
DyGED_NA	NYC Cab	0.26	0.50	0.21	0.36	0.82	0.005	150
DyGED_NA	Twitter Weather	0.21	0.10	0.12	1.00	0.82	0.001	50
DyGED	NYC Cab	0.23	0.20	0.28	0.20	0.81	0.001	100
DyGED	Twitter Weather	0.37	0.15	0.26	0.64	0.68	0.005	150

Ακολουθούν τα διαγράμματα που δείχνουν τις Προβλέψεις Πιθανοτήτων σε σύγκριση με τα Πραγματικά Γεγονότα (Test Set):



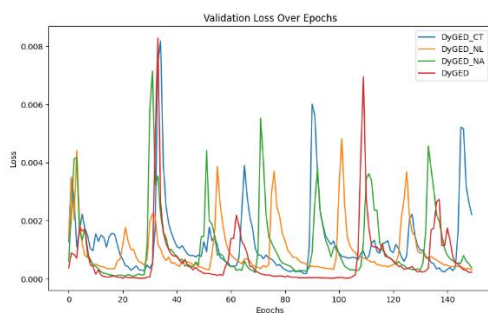
Εικόνα 10: NYC Cab - Test Set (F1 Score)



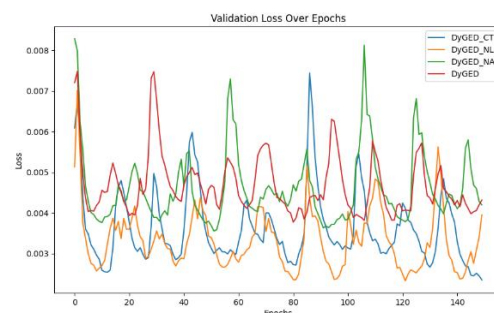
Εικόνα 11: Twitter Weather - Test Set (F1 Score)

Τα διαγράμματα παρουσιάζουν τις πιθανότητες που προβλέφθηκαν από τα μοντέλα, συγκριτικά με τα πραγματικά γεγονότα (true events) στο test set. Αποκαλύπτουν την ικανότητα των μοντέλων να ανιχνεύουν θετικά γεγονότα και τη συμπεριφορά τους απέναντι στα διαφορετικά thresholds.

Ακολουθούν τα διαγράμματα που αποτυπώνουν τη μεταβολή του Validation Loss ανά Εποχή:



Εικόνα 12: NYC Cab - Validation Loss (F1 Score)



Εικόνα 13: Twitter Weather - Validation Loss (F1 Score)

Οι παραπάνω πληροφορίες ενισχύουν τη συνολική ανάλυση των αποτελεσμάτων, παρέχοντας τόσο αριθμητική όσο και οπτική κατανόηση της απόδοσης των μοντέλων.

2. Ανάλυση Precision και Recall

Precision:

- Το **Precision** μετρά την ακρίβεια στις θετικές προβλέψεις.
 - **Υψηλό Precision** (π.χ., DyGED_NL στο NYC Cab): Το μοντέλο αποφεύγει τα **False Positives**, προβλέποντας θετικά γεγονότα μόνο όταν είναι σίγουρο.
 - **Χαμηλό Precision** (π.χ., DyGED_NA στο Twitter Weather): Υποδεικνύει ότι το μοντέλο προβλέπει πολλά **False Positives**, μειώνοντας την ακρίβεια των προβλέψεων.

Recall:

- Το **Recall** μετρά την κάλυψη των θετικών γεγονότων.
 - **Υψηλό Recall** (π.χ., DyGED_NA στο Twitter Weather): Το μοντέλο ανιχνεύει σχεδόν όλα τα θετικά γεγονότα, ακόμα κι αν παράγει περισσότερα False Positives.
 - **Χαμηλό Recall** (π.χ., DyGED στο NYC Cab): Το μοντέλο χάνει πολλά θετικά γεγονότα, οδηγώντας σε αυξημένα **False Negatives**.

F1 Score:

- Η **F1 Score** συνδυάζει το Precision και το Recall.
 - Υψηλή F1 (π.χ., DyGED_NL στο NYC Cab, $F1 = 0.38$): Δείχνει ισορροπία.
 - Χαμηλή F1 (π.χ., DyGED_NA στο Twitter Weather, $F1 = 0.21$): Δείχνει προβλήματα στο μοντέλο.

AUC:

- Το **AUC** αποκαλύπτει την ικανότητα διαχωρισμού:
 - Στο NYC Cab, το DyGED_NL πέτυχε **0.88**, δείχνοντας ισχυρή διαχωριστική ικανότητα.
 - Στο Twitter Weather, το AUC ήταν έως **0.82**, αλλά οι προβλέψεις ήταν πιο ασταθείς λόγω της φύσης των δεδομένων.

3. Δοκιμές Παραμέτρων και Thresholds

Για την επίτευξη καλύτερης απόδοσης, δοκιμάστηκαν διάφοροι συνδυασμοί παραμέτρων:

1. Learning Rate (0.001 έως 0.005):

Μικρότερες τιμές έδωσαν σταθερότερη εκπαίδευση, ενώ μεγαλύτερες επιτάχυναν την εκπαίδευση αλλά αύξησαν την αστάθεια.

2. Αριθμός Εποχών (50 έως 150):

Περισσότερες εποχές βελτίωσαν το Recall, αλλά η πρόοδος σταμάτησε μετά από συγκεκριμένο σημείο.

3. Thresholds (0.1 έως 0.5):

Τα thresholds επηρεάζουν σημαντικά το F1 Score:

- **Χαμηλά thresholds:** Αύξησαν το Recall αλλά μείωσαν το Precision.
- **Υψηλά thresholds:** Αύξησαν το Precision αλλά μείωσαν το Recall.

Για όλα τα πειράματα, χρησιμοποιήσαμε σταθερές τιμές για το **batch size (100)** και το **dropout_rate (0.02)**. Επιπλέον, ορίστηκε ο αριθμός των κρυφών μονάδων σε **num_hidden = 64**. Για την αντιμετώπιση της ανισορροπίας των δεδομένων, χρησιμοποιήθηκε σταθμισμένη συνάρτηση απώλειας (weighted loss) μέσω της παραμέτρου θετικού βάρους (pos_weight), η οποία υπολογίστηκε με βάση την αναλογία των θετικών και αρνητικών παραδειγμάτων στο σύνολο εκπαίδευσης.

4. Επεξήγηση Χαμηλών Τιμών F1

Οι χαμηλές τιμές F1 οφείλονται:

1. **Ανισορροπία στη Βάση:**
 - Τα θετικά γεγονότα είναι πολύ λιγότερα από τα αρνητικά.
 - Το μοντέλο δυσκολεύεται να ανιχνεύσει τα θετικά γεγονότα, μειώνοντας το Recall.
2. **Χαμηλές Προβλέψεις Πιθανοτήτων:**
 - Τα περισσότερα θετικά γεγονότα προβλέπονται με πιθανότητες κοντά στο threshold, οδηγώντας σε αυξημένα False Negatives.
3. **Ευαισθησία στο Threshold:**
 - Το threshold έχει καθοριστική επίδραση στο F1, ιδιαίτερα σε βάσεις με ανισόρροπη κατανομή

5. Συμπέρασμα

1. **Υψηλό AUC αλλά Χαμηλό F1:**
 - Το AUC δείχνει ότι τα μοντέλα έχουν καλές δυνατότητες διαχωρισμού.
 - Οι χαμηλές τιμές F1 οφείλονται στη δυσκολία διατήρησης ισορροπίας Precision-Recall.
2. **Καλύτερα Αποτελέσματα:**
 - Στο **NYC Cab**, το **DyGED_NL** είχε την καλύτερη F1 Score (**0.38**) και υψηλό AUC (**0.88**).
 - Στο **Twitter Weather**, το **DyGED** πέτυχε την υψηλότερη F1 (**0.37**) με καλό Recall (**0.64**).
3. **Βελτίωση Μοντέλων:**
 - Τεχνικές όπως η **υπερδειγματοληψία (oversampling)** ή η χρήση συνθετικών δειγμάτων θα μπορούσαν να δοκιμαστούν για την αντιμετώπιση της ανισορροπίας στα δεδομένα.
 - Η καλύτερη ρύθμιση των **thresholds** μπορεί να συμβάλει στη βελτίωση της ισορροπίας μεταξύ Precision και Recall. Η περαιτέρω ανάπτυξη των μοντέλων μπορεί να εστιάσει στη δυναμική προσαρμογή των thresholds και στη βελτίωση των χαρακτηριστικών των δεδομένων, για πιο αξιόπιστες προβλέψεις.

5.3. Δεύτερο Μέρος Πειραματικής Διαδικασίας

5.3.1. Εισαγωγή – Δημιουργία Νέα Σύνολα Δεδομένων

Σε αυτό το μέρος της πειραματικής διαδικασίας που χρησιμοποιούμε τη βιβλιοθήκη RDyn (<https://github.com/GiulioRossetti/RDyn>) για τη δημιουργία δυναμικών συνόλων δεδομένων γραφημάτων, τα οποία εξελίσσονται διαδοχικά μέσω πολλαπλών στιγμιότυπων (snapshots). Το RDyn παράγει έναν φάκελο results, ο οποίος περιέχει τα αποτελέσματα για κάθε διαμόρφωση του μοντέλου σε ξεχωριστό υποφάκελο. Για παράδειγμα, ο υποφάκελος '150_100_15_0.6_0.8_0.2_1' αναφέρεται σε ένα μοντέλο που έχει 150 κόμβους και 100 στιγμιότυπα. Κάθε υποφάκελος περιέχει τα εξής αρχεία:

- **graph-*.txt:** Παρέχει τη λίστα ακμών του γράφου για κάθε χρονική στιγμή, προσφέροντας πληροφορίες συνδεσιμότητας μεταξύ των κόμβων.
- **communities-*.txt:** Περιγράφει τις κοινότητες, αναφέροντας τους κόμβους που ανήκουν σε κάθε κοινότητα.
- **events.txt:** Συμπεριλαμβάνει μια σύνοψη των γεγονότων συγχωνεύσεων και διαχωρισμών (merge/split) που συμβαίνουν ανά χρονική στιγμή, επισημαίνοντας τις αλλαγές στη δομή των κοινοτήτων. Αυτές οι αλλαγές ουσιαστικά είναι οι ανωμαλίες μας.

Τα παραγόμενα αρχεία .txt αποτελούν τη βάση για τη δημιουργία των τανυστών εισόδου στα μοντέλα μας, παρέχοντας πληροφορίες που διευκολύνουν την ανίχνευση των αλλαγών στις κοινότητες και τις συνδέσεις του γράφου. Για να διατηρήσουμε τον αριθμό των events χαμηλό σε κάθε βάση δεδομένων, επιλέγουμε να κρατάμε περιορισμένο το πλήθος των snapshots. Με τον τρόπο αυτό, μειώνεται η συχνότητα των γεγονότων (merge και split events), τα οποία αποτελούν στιγμές σημαντικών αλλαγών στις κοινότητες και τις συνδέσεις του γράφου. Ο περιορισμός αυτός επιτρέπει στα μοντέλα μας να επικεντρωθούν σε πιο ουσιαστικές ανιχνεύσεις και διαφοροποιήσεις, αποφεύγοντας την υπερβολική πυκνότητα γεγονότων που μπορεί να οδηγήσει σε ασαφή ή λιγότερο ακριβή αποτελέσματα κατά την εκπαίδευση και αξιολόγηση.

Για τη συγκεκριμένη διαδικασία, θα δοκιμάσουμε τα δεδομένα από τους εξής υποφάκελους: results/30_60_15_0.6_0.8_0.2_1, results/150_100_15_0.6_0.8_0.2_1 και results/500_60_15_0.6_0.8_0.2_1. Για την αναπαραγωγικότητα των πειραμάτων, τα αποτελέσματα αποθηκεύτηκαν σε ένα αρχείο [results.zip](#) στο αποθετήριο GitHub, το οποίο περιέχει ακριβώς τις βάσεις δεδομένων που χρησιμοποιήθηκαν.

5.3.2. Χρησιμοποιούμενος Κώδικας

Αρχικά, παραθέτουμε ενδεικτικό κώδικα για τη δημιουργία των δεδομένων με τη χρήση της βιβλιοθήκης RDyn.

```
from rdyn import RDyn

# Παράδειγμα διαμόρφωσης RDyn για τη δημιουργία δυναμικών δεδομένων
rdb = RDyn(size=150, iterations=100)
rdb.execute()
```

Έπειτα, πραγματοποιείται η μετατροπή των δεδομένων από τα παραγόμενα αρχεία σε τανυστές για την εκπαίδευση των μοντέλων. Η δημιουργία των τανυστών G (πίνακας γειτνίασης), X (χαρακτηριστικά κόμβων) και y_true (labels γεγονότων) αναλύεται παρακάτω, ενώ ο πλήρης κώδικας υλοποίησης βρίσκεται στο αρχείο [custom_data_format.py](#).

1. Δημιουργία του τανυστή G :

Ο τανυστής G αναπαριστά τον πίνακα γειτνίασης για κάθε στιγμιότυπο (snapshot) του γραφήματος. Για κάθε αρχείο `graph-*.txt`, το οποίο περιέχει τις συνδέσεις των κόμβων, ενημερώνουμε τις αντίστοιχες θέσεις στον τανυστή G για να καταγράψουμε τις συνδέσεις κάθε στιγμιότυπου.

2. Δημιουργία του τανυστή X :

Ο τανυστής X αναπαριστά τα χαρακτηριστικά των κόμβων, βασισμένα στις κοινότητες στις οποίες ανήκουν κάθε χρονική στιγμή. Τα αρχεία `communities-*.txt` περιέχουν τα δεδομένα των κοινοτήτων, τα οποία μετατρέπονται στον τανυστή X αποδίδοντας σε κάθε κόμβο ένα αναγνωριστικό κοινότητας.

3. Δημιουργία του τανυστή y_true :

Ο τανυστής y_true δημιουργείται με βάση τα συμβάντα που καταγράφονται στο αρχείο `events.txt`, όπου στιγμιότυπα με MERGE και SPLIT γεγονότα λαμβάνουν την τιμή 1, ενώ τα υπόλοιπα καταχωρούνται ως 0. Για κάθε τέτοιο γεγονός, η τιμή στο y_true αντιστοιχεί στην χρονική στιγμή (`current_time_step` μείον 1), διευκολύνοντας την ακριβή αντιστοίχιση των πραγματικών συμβάντων με τις θέσεις των snapshots. Αυτός ο τανυστής αποτελεί θεμελιώδη βάση για την εκπαίδευση του μοντέλου και χρησιμοποιείται επίσης στην αξιολόγηση του test set, όπου υπολογίζονται κατάλληλες μετρικές για μια ολοκληρωμένη εκτίμηση της απόδοσης του μοντέλου.

Ακόμη παρουσιάζουμε μια σύντομη περιγραφή των βασικών βημάτων, τα οποία ακολουθούν τη διαδικασία που έχουμε ήδη αναλύσει στο προηγούμενο υλοποιητικό μέρος. Τα βήματα αυτά υλοποιούνται στον κώδικα του αρχείου [models_in_custom.py](#) και περιλαμβάνουν τα εξής:

- **Φόρτωση δεδομένων:** Τα αρχεία G , X και y_true , που περιέχουν τον πίνακα γειτνίασης, τα χαρακτηριστικά των κόμβων και τις ετικέτες των γεγονότων, φορτώνονται από προηγούμενα βήματα επεξεργασίας.

- **Διαχωρισμός σε training και test sets:** Τα δεδομένα διαχωρίζονται χωρίς ανακάτεμα, διατηρώντας την αλληλουχία τους.
- **Ρύθμιση παραμέτρων:** Καθορίζονται οι βασικές παράμετροι, όπως ο ρυθμός εκμάθησης και το dropout, για την εκπαίδευση των μοντέλων.
- **Ορισμός μοντέλων:** Ορίζονται διαφορετικές παραλλαγές του DyGED για την ανίχνευση γεγονότων.
- **Εκπαίδευση και αξιολόγηση:** Κάθε μοντέλο εκπαιδεύεται σε πολλαπλές εποχές και αξιολογείται με τις μετρικές AUC και F1 για τον έλεγχο της αποτελεσματικότητάς του.

5.3.3. Πειραματικά Αποτελέσματα Συνόλων Δεδομένων Rdyn

Σε αυτή την ενότητα παρουσιάζονται τα πειραματικά αποτελέσματα που προέκυψαν από τις βάσεις δεδομένων που δημιουργήθηκαν με τη βιβλιοθήκη RDyn, όπως αναλύσαμε παραπάνω.

Ο πίνακας παρουσιάζει την καλύτερη απόδοση των μοντέλων για τα γεγονότα merge και split μετά από αρκετές εκτελέσεις, ενώ γίνεται σύγκριση των επιδόσεων ανάμεσα στις διάφορες παραλλαγές του μοντέλου DyGED.

	DyGED_CT	DyGED_NL	DyGED_NA	DyGED
RDyn (30 Κόμβοι, 60 Στιγμιότυπα) num_epochs = 50	AUC: 0.973 F1: 0.800	AUC: 0.971 F1: 0.800	AUC: 0.973 F1: 0.800	AUC: 0.975 F1: 0.800
RDyn (150 Κόμβοι, 100 Στιγμιότυπα) num_epochs = 100	AUC: 0.936 F1: 0.909	AUC: 0.895 F1: 0.909	AUC: 0.902 F1: 0.910	AUC: 0.945 F1: 0.909
RDyn (500 Κόμβοι, 60 Στιγμιότυπα) num_epochs = 150	AUC: 0.981 F1: 0.863	AUC: 0.962 F1: 0.754	AUC: 0.963 F1: 0.861	AUC: 0.967 F1: 0.868
Ρυθμίσεις εκπαίδευσης	train_ratio = 0.65 learning_rate = 0.005 dropout_rate = 0.02 num_hidden = 64	train_ratio = 0.65 learning_rate = 0.005 dropout_rate = 0.02 num_hidden = 64	train_ratio = 0.65 learning_rate = 0.005 dropout_rate = 0.02 num_hidden = 64	train_ratio = 0.65 learning_rate = 0.005 dropout_rate = 0.02 num_hidden = 64

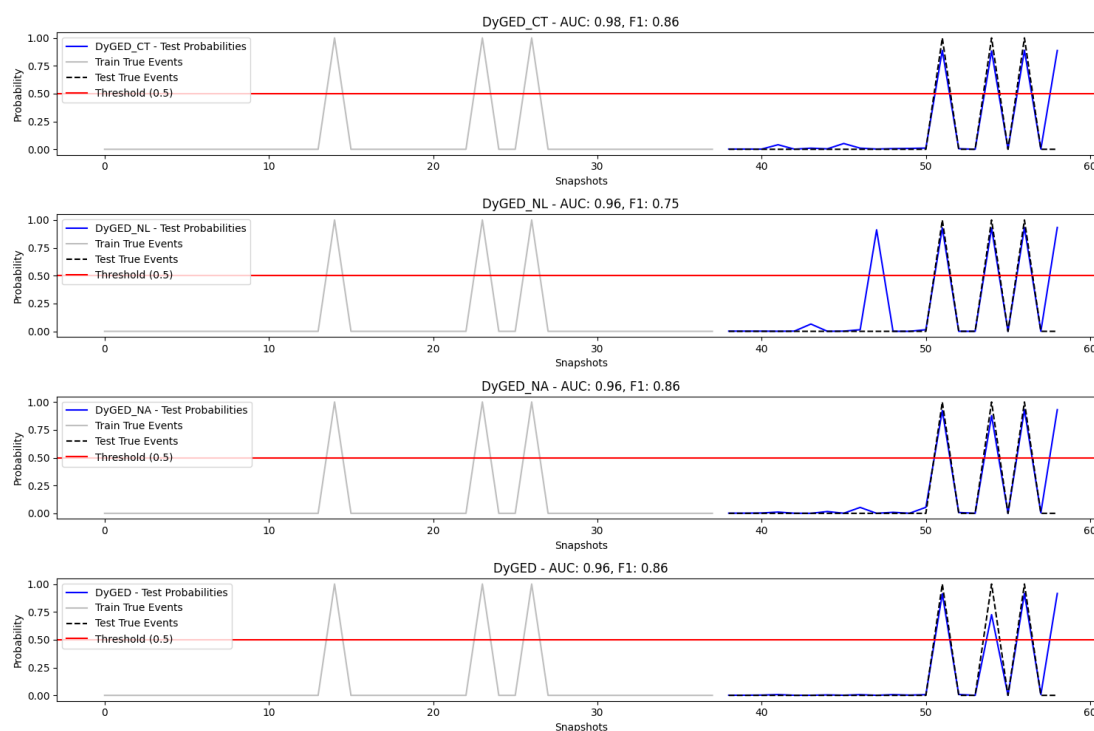
Πίνακας 4: Μετρική AUC και F1 - Rdyn Βάσεις

Τα αποτελέσματα που παρουσιάζονται δείχνουν ότι τα μοντέλα επιτυγχάνουν υψηλές επιδόσεις στα δεδομένα RDyn με 30, 150 και 500 κόμβους, παρέχοντας σταθερά υψηλές τιμές AUC και F1. Όταν αυξάνεται ο αριθμός των κόμβων και των στιγμιότυπων, παρατηρείται μια μικρή πτώση στις τιμές F1, κυρίως λόγω της αυξημένης πολυπλοκότητας των δεδομένων, αλλά τα μοντέλα εξακολουθούν να

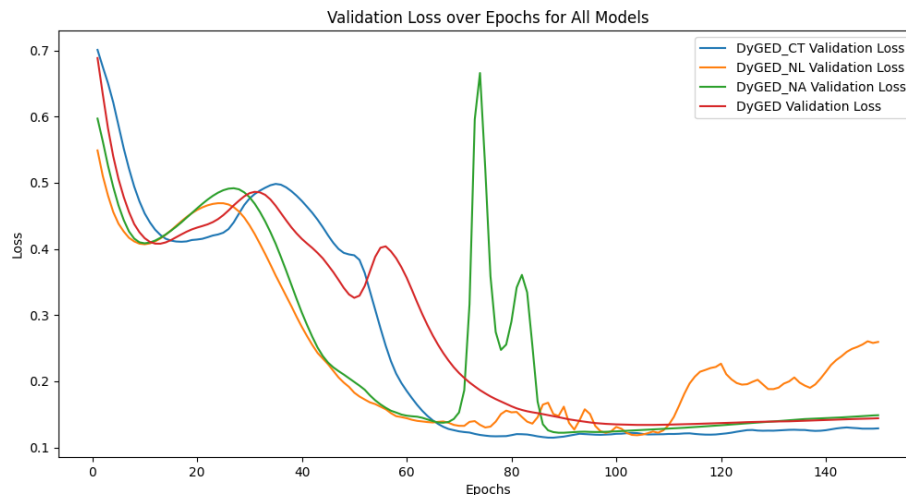
δείχνουν υψηλές επιδόσεις, με AUC και F1 να παραμένουν σε πολύ ικανοποιητικά επίπεδα.

Οι παράμετροι εκπαίδευσης που χρησιμοποιήθηκαν φαίνεται να είναι αποτελεσματικές για τη σταθεροποίηση και την απόδοση των μοντέλων. Η τιμή του learning rate (0.005) επιτρέπει στα μοντέλα να μαθαίνουν σταδιακά και με σταθερότητα, ενώ το dropout rate (0.02) παρέχει επαρκή τακτικότητα χωρίς να διακόπτει τη μάθηση. Επιπλέον, ο αριθμός των κρυφών μονάδων ($\text{num_hidden} = 64$) φαίνεται να είναι κατάλληλος, προσφέροντας επαρκή υπολογιστική ικανότητα για την επεξεργασία των χαρακτηριστικών. Συνολικά, τα αποτελέσματα δείχνουν ότι οι ρυθμίσεις αυτές αποτελούν καλές επιλογές για τα μοντέλα, προσφέροντας σταθερά αποδοτικά αποτελέσματα σε διαφορετικά σενάρια δεδομένων.

Για καλύτερη οπτικοποίηση, ακολουθούν διαγράμματα που παρουσιάζουν τα scores όλων των μοντέλων, καθώς και τις απώλειες (validation loss) κατά τη διάρκεια της εκπαίδευσης, για τη βάση δεδομένων με 500 κόμβους. Αυτά τα διαγράμματα παρέχουν μια συγκριτική εικόνα της απόδοσης των μοντέλων κατά την ανίχνευση συμβάντων και την εκπαίδευσή τους στην εν λόγω βάση δεδομένων.



Εικόνα 14: Σύγκριση Απόδοσης Μοντέλων: Scores για 500 Κόμβους



Εικόνα 15: Απώλεια Κατά την Εκπαίδευση (Validation Loss) για 500 Κόμβους

6. Συμπεράσματα

6.1. Γενικά Συμπεράσματα

Στα πλαίσια του πρακτικού μέρους της διπλωματικής εργασίας, εμπλουτίστηκε ο πειραματικός κώδικας με βάση πρότυπη υλοποίηση η οποία δόθηκε, και κατέστη εφικτή η αξιολόγηση τεσσάρων διαφορετικών μοντέλων ανίχνευσης γεγονότων σε δυναμικά γραφήματα. Αξιοσημείωτο είναι το γεγονός ότι πραγματοποιήθηκε πειραματική αξιολόγηση σε δύο διαφορετικά σύνολα δεδομένων, τα οποία ήταν μεγάλα σε μέγεθος και περίπλοκα στη διαχείριση.

Από την πειραματική αξιολόγηση, προέκυψαν χρήσιμα συμπεράσματα σχετικά με τη συμπεριφορά των μοντέλων, τα οποία οπτικοποιήθηκαν μέσω συγκριτικών γραφημάτων. Σημαντικό ρόλο στην επιτυχή εξέλιξη της πειραματικής διαδικασίας διαδραμάτισε η γλώσσα Python και, ειδικότερα, η βιβλιοθήκη PyTorch, μέσω της δυνατότητας χρήσης τανυστών (tensors) που διαθέτει.

Με την προσθήκη του δεύτερου πειραματικού μέρους, επεκτάθηκε η ανάλυση των μοντέλων ανίχνευσης γεγονότων σε δυναμικά γραφήματα με τη χρήση δεδομένων που δημιουργήθηκαν από τη βιβλιοθήκη RDyn. Αυτή η προσέγγιση επέτρεψε τη διερεύνηση της αποτελεσματικότητας των μοντέλων σε ένα ακόμα πιο σύνθετο και δυναμικά εξελισσόμενο σύνολο δεδομένων, εμπλουτίζοντας τα συμπεράσματα με χρήσιμες παρατηρήσεις σχετικά με τη συμπεριφορά των μοντέλων σε περιβάλλοντα με μεγάλο αριθμό στιγμιότυπων και αυξημένη πολυπλοκότητα συμβάντων. Οι επιδόσεις των μοντέλων αξιολογήθηκαν χρησιμοποιώντας μετρικές AUC και F1, ενισχύοντας τη συγκριτική ανάλυση των αποτελεσμάτων.

Ωστόσο, χρειάζεται περαιτέρω ενίσχυση της πειραματικής διαδικασίας, ώστε να προσεγγιστούν τα βέλτιστα αποτελέσματα της δημοσίευσης πάνω στην οποία στηρίχθηκε η πρότυπη υλοποίηση που δόθηκε.

6.2. Περιορισμοί

Παρότι η παρούσα εργασία ανέδειξε σημαντικά ευρήματα, υπάρχουν ορισμένοι περιορισμοί που πρέπει να επισημανθούν:

- 1. Υποστήριξη PyTorch Geometric και Προετοιμασία Δεδομένων:**
Το framework που χρησιμοποιήθηκε για την εκπαίδευση και αξιολόγηση των μοντέλων δεν υποστηρίζει απευθείας PyTorch Geometric, απαιτώντας τη μετατροπή των δεδομένων σε ακατέργαστη μορφή. Αυτό αυξάνει την πολυπλοκότητα της διαδικασίας προετοιμασίας δεδομένων και ενδέχεται να επηρεάσει την αποδοτικότητα της πειραματικής διαδικασίας.
- 2. Προκαθορισμένες Παράμετροι:**
Η επιλογή υπερπαραμέτρων (π.χ. learning rate, batch size) πραγματοποιήθηκε εμπειρικά για τα συγκεκριμένα σύνολα δεδομένων που χρησιμοποιήθηκαν. Ωστόσο, σε άλλα μελλοντικά σύνολα δεδομένων, μπορεί να απαιτηθεί διαφορετική παραμετροποίηση για την επίτευξη βέλτιστων αποτελεσμάτων. Πιθανές βελτιώσεις μπορούν να προκύψουν μέσω εκτενέστερης διερεύνησης των τιμών αυτών με τεχνικές όπως το grid search ή το Bayesian optimization.
- 3. Σταθερότητα Κατά την Εκπαίδευση:**
Κατά τη διάρκεια της εκπαίδευσης, είναι πιθανό κάποια τρεξίματα να μην αποδώσουν καλά αποτελέσματα λόγω των αρχικών συνθηκών ή τυχαιότητας. Για να διασφαλιστεί ότι τα αποτελέσματα δεν είναι τυχαία, συνιστώνται πολλαπλά τρεξίματα για την εξαγωγή αξιόπιστων και σταθερών συμπερασμάτων.
- 4. Συχνότητα Γεγονότων:**
Η συχνότητα των γεγονότων (events) επηρεάζει σημαντικά την απόδοση των μοντέλων. Στο δεύτερο μέρος της πειραματικής διαδικασίας, παρατηρήθηκε ότι τα μοντέλα παρουσιάζουν πολύ καλές επιδόσεις όταν το ποσοστό των γεγονότων (π.χ. ανωμαλίες) είναι κάτω από 20%. Ως εκ τούτου, θα πρέπει να ληφθεί υπόψη η επιλογή συνόλων δεδομένων με κατάλληλα ποσοστά γεγονότων ή η δημιουργία νέων βάσεων δεδομένων που προσαρμόζονται στις συγκεκριμένες συνθήκες.
- 5. Υπολογιστική Ισχύς:**
Ορισμένα από τα μοντέλα που δοκιμάστηκαν απαιτούν υψηλή υπολογιστική ισχύ, γεγονός που ενδέχεται να περιορίσει την εφαρμογή τους σε περιβάλλοντα με περιορισμένους πόρους. Οι πειραματισμοί που πραγματοποιήθηκαν βασίστηκαν σε υπολογιστή με επεξεργαστή 12th Gen Intel(R) Core(TM) i7-1255U στα 1.70 GHz και 16 GB RAM. Ενδέχεται, επομένως, σε συστήματα με χαμηλότερη ισχύ, τα αποτελέσματα ή η ταχύτητα εκτέλεσης να επηρεαστούν σημαντικά.

6.3. Μελλοντικές Κατευθύνσεις

Σαν μελλοντικές κατευθύνσεις προτείνονται:

- Περαιτέρω πειραματική αξιολόγηση με χρήση διαφορετικών τιμών των παραμέτρων των μοντέλων.
- Συνέχιση της προσπάθειας για εύρεση εναλλακτικών συνόλων δεδομένων, ώστε να εξεταστεί η γενίκευση των μοντέλων.
- Εφαρμογή επιπλέον μοντέλων ανίχνευσης γεγονότων, τα οποία μπορεί να προσφέρουν καλύτερες επιδόσεις.
- Χρήση περισσότερων μετρικών αξιολόγησης που θα προσφέρουν μια ευρύτερη και βαθύτερη κατανόηση των αποτελεσμάτων και της συμπεριφοράς των μοντέλων σε πολυδιάστατα σύνολα δεδομένων.
- Η επέκταση της μελέτης σε κατανεμημένα δυναμικά γραφήματα, επιτρέποντας την εφαρμογή σε κλιμακούμενα και πραγματικού χρόνου περιβάλλοντα, κάτι που θα ενίσχυε τη χρησιμότητα των μεθόδων σε συστήματα μεγάλης κλίμακας.

Βιβλιογραφία

Aggarwal, C., & Subbian, K. (2014). Evolutionary network analysis: A survey. *ACM Computing Surveys (CSUR)*, 47(1), 1-36.

Ahmed, M., Mahmood, A. N., & Hu, J. (2016). A survey of network anomaly detection techniques. *Journal of Network and Computer Applications*, 60, 19-31.
<https://doi.org/10.1016/j.jnca.2015.11.016>

Akoglu, L., Tong, H., & Koutra, D. (2015). Graph-based anomaly detection and description: A survey. *Data Mining and Knowledge Discovery*, 29(3), 626-688.
<https://doi.org/10.1007/s10618-014-0365-y>

Arroyo, I. (2022). Unsupervised Anomaly Detection on Spotify Data: K-means vs. Local Outlier Factor.

Barnett, V., & Lewis, T. (1994). *Outliers in Statistical Data*. John Wiley & Sons.

Barocas, S., & Selbst, A. D. (2016). Big data's disparate impact. *California Law Review*, 104(3), 671-732.

- Bergstra, J., & Bengio, Y. (2012). Random search for hyper-parameter optimization. *Journal of Machine Learning Research*, 13(1), 281-305.
- Bessen, J. E. (2019). *AI and Jobs: The role of demand*. National Bureau of Economic Research.
- Bishop, C. M. (2006). *Pattern recognition and machine learning*. Springer.
- Blum, A., & Mitchell, T. (1998). Combining labeled and unlabeled data with co-training. In *Proceedings of the Eleventh Annual Conference on Computational Learning Theory* (pp. 92-100).
- Box, G. E., & Jenkins, G. M. (1970). *Time Series Analysis: Forecasting and Control*. Holden-Day.
- Breiman, L. (1996). Bagging predictors. *Machine Learning*, 24(2), 123-140.
- Breiman, L., Friedman, J. H., Olshen, R. A., & Stone, C. J. (1984). *Classification and regression trees*. Wadsworth.
- Brennan, T., & Lo, A. W. (2019). *The origin of behavior*. MIT Press.
- Chapelle, O., Schölkopf, B., & Zien, A. (Eds.). (2006). *Semi-supervised learning*. MIT Press.
- Chandola, V., Banerjee, A., & Kumar, V. (2009). Anomaly detection: A survey. *ACM Computing Surveys*, 41(3), 1-58.
- Chen, F., & Neill, D. B. (2014). Non-parametric scan statistics for event detection and forecasting in heterogeneous social media graphs. In *Proceedings of the 20th ACM SIGKDD international conference on Knowledge discovery and data mining* (pp. 1166-1175). <https://doi.org/10.1145/2623330.2623660>
- Cleveland, R. B., Cleveland, W. S., McRae, J. E., & Terpenning, I. (1990). STL: A seasonal-trend decomposition procedure based on loess. *Journal of Official Statistics*, 6(1), 3-73.

Cortes, C., & Vapnik, V. (1995). Support-vector networks. *Machine Learning*, 20(3), 273-297.

Davies, D. L., & Bouldin, D. W. (1979). A cluster separation measure. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, PAMI-1(2), 224-227.

Dean, J., Corrado, G., Monga, R., Chen, K., Devin, M., Le, Q. V., ... & Ng, A. Y. (2012). Large scale distributed deep networks. In *Advances in neural information processing systems* (pp. 1223-1231).

Diethe, T., Borchert, T., & Girolami, M. (2019). Continual learning in practice. *NeurIPS Continual Learning Workshop*.

Ding, K., Li, J., Bhanushali, R., & Liu, H. (2019). Deep anomaly detection on attributed networks. Proceedings of the 2019 SIAM International Conference on Data Mining, 594-602. <https://doi.org/10.1137/1.9781611975673.67>

Dorfer, T. A. (2023). Advanced Time-Series Anomaly Detection with Deep Learning in Power BI. Towards Data Science.

Doshi-Velez, F., & Kim, B. (2017). Towards a rigorous science of interpretable machine learning. *arXiv preprint arXiv:1702.08608*.

Ester, M., Kriegel, H. P., Sander, J., & Xu, X. (1996). A density-based algorithm for discovering clusters in large spatial databases with noise. In *Proceedings of the 2nd International Conference on Knowledge Discovery and Data Mining* (pp. 226-231).

Esteva, A., Robicquet, A., Ramsundar, B., Kuleshov, V., DePristo, M., Chou, K., Cui, C., Corrado, G. S., Thrun, S., & Dean, J. (2019). A guide to deep learning in healthcare. *Nature Medicine*, 25(1), 24-29.

Fawcett, T. (2006). An introduction to ROC analysis. *Pattern Recognition Letters*, 27(8), 861-874.

Freund, Y., & Schapire, R. E. (1997). A decision-theoretic generalization of on-line learning and an application to boosting. *Journal of Computer and System Sciences*, 55(1), 119-139.

- Goodfellow, I., Bengio, Y., & Courville, A. (2016). *Deep learning*. MIT Press.
- Hastie, T., Tibshirani, R., & Friedman, J. (2009). *The elements of statistical learning: Data mining, inference, and prediction*. Springer.
- He, H., & Garcia, E. A. (2009). Learning from imbalanced data. *IEEE Transactions on Knowledge and Data Engineering*, 21(9), 1263-1284.
- Hinton, G. E., Osindero, S., & Teh, Y. W. (2006). A fast learning algorithm for deep belief nets. *Neural Computation*, 18(7), 1527-1554.
- Hodge, V. J., & Austin, J. (2004). A survey of outlier detection methodologies. *Artificial Intelligence Review*, 22(2), 85-126.
<https://doi.org/10.1023/B:AIRE.0000045502.10941.a9>
- Holme, P., & Saramäki, J. (2012). Temporal networks. *Physics Reports*, 519(3), 97-125.
- Hosmer, D. W., Lemeshow, S., & Sturdivant, R. X. (2013). *Applied logistic regression*. John Wiley & Sons.
- Jain, A. K. (2010). Data clustering: 50 years beyond K-means. *Pattern Recognition Letters*, 31(8), 651-666.
- Jolliffe, I. T. (2011). *Principal component analysis*. Springer.
- Jurafsky, D., & Martin, J. H. (2021). *Speech and language processing*. Pearson.
- Kober, J., Bagnell, J. A., & Peters, J. (2013). Reinforcement learning in robotics: A survey. *The International Journal of Robotics Research*, 32(11), 1238-1274.
- Konda, V. R., & Tsitsiklis, J. N. (2000). Actor-critic algorithms. In *Advances in Neural Information Processing Systems* (pp. 1008-1014).
- Krizhevsky, A., Sutskever, I., & Hinton, G. E. (2012). Imagenet classification with deep convolutional neural networks. In *Advances in neural information processing systems* (pp. 1097-1105).

- Li, Y., Deng, Y., & Cass, B. (2020). Applications of reinforcement learning in finance. *IEEE Computational Intelligence Magazine*, 15(1), 17-25.
- Lipton, Z. C. (2018). The mythos of model interpretability. *Communications of the ACM*, 61(10), 36-43.
- Litman, T. (2020). Autonomous vehicle implementation predictions: Implications for transport planning. *Victoria Transport Policy Institute*.
- Liu, B., Dai, Y., Li, X. L., Lee, W. S., & Yu, P. S. (2008). Building text classifiers using positive and unlabeled examples. In *Proceedings of the International Conference on Data Mining (ICDM)* (pp. 179-188).
- Ma, X., Xue, S., Yang, J., Zhou, C., Sheng, Q. and Xiong, H. (2021) A Comprehensive Survey on Graph Anomaly Detection with Deep Learning. JOURNAL OF LATEX CLASS FILE.
- MacQueen, J. (1967). Some methods for classification and analysis of multivariate observations. In *Proceedings of the Fifth Berkeley Symposium on Mathematical Statistics and Probability* (Vol. 1, pp. 281-297).
- Minsky, M., & Papert, S. (1969). *Perceptrons: An introduction to computational geometry*. MIT Press.
- Mitchell, T. M., Brynjolfsson, E., Brynjolfsson, E., Rock, D., & Syverson, C. (2019). *Artificial intelligence and the modern productivity paradox: A clash of expectations and statistics*. University of Chicago Press.
- Mnih, V., Kavukcuoglu, K., Silver, D., Rusu, A. A., Veness, J., Bellemare, M. G., ... & Hassabis, D. (2015). Human-level control through deep reinforcement learning. *Nature*, 518(7540), 529-533.
- Montgomery, D. C., Peck, E. A., & Vining, G. G. (2012). *Introduction to linear regression analysis*. John Wiley & Sons.

Murtagh, F., & Contreras, P. (2012). Algorithms for hierarchical clustering: An overview. *Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery*, 2(1), 86-97.

Noble, C. C., & Cook, D. J. (2003). Graph-based anomaly detection. *Proceedings of the 9th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, 631-636.

Perez Veiga, A. (2018). Applications of Artificial Intelligence (AI) to Network Security

Powers, D. M. (2011). Evaluation: From precision, recall and F-measure to ROC, informedness, markedness and correlation. *Journal of Machine Learning Technologies*, 2(1), 37-63.

Ramaswamy, S., Rastogi, R., & Shim, K. (2000). Efficient algorithms for mining outliers from large data sets. *Proceedings of the ACM SIGMOD International Conference on Management of Data*, 427-438.

Rasmus, A., Berglund, M., Honkala, M., Valpola, H., & Raiko, T. (2015). Semi-supervised learning with ladder networks. In *Advances in Neural Information Processing Systems* (pp. 3546-3554).

Rosenblatt, F. (1958). The perceptron: A probabilistic model for information storage and organization in the brain. *Psychological Review*, 65(6), 386-408.

Rousseeuw, P. J. (1987). Silhouettes: A graphical aid to the interpretation and validation of cluster analysis. *Journal of Computational and Applied Mathematics*, 20, 53-65.

Rubin, D. B. (1976). Inference and missing data. *Biometrika*, 63(3), 581-592.

Rummery, G. A., & Niranjan, M. (1994). On-line Q-learning using connectionist systems. University of Cambridge, Department of Engineering.

Rumelhart, D. E., Hinton, G. E., & Williams, R. J. (1986). Learning representations by back-propagating errors. *Nature*, 323(6088), 533-536.

Samuel, A. L. (1959). Some studies in machine learning using the game of checkers. *IBM Journal of Research and Development*, 3(3), 210-229.

Shokri, R., Stronati, M., Song, C., & Shmatikov, V. (2017). Membership inference attacks against machine learning models. In *2017 IEEE Symposium on Security and Privacy (SP)* (pp. 3-18). IEEE.

Siciliano, B., & Khatib, O. (Eds.). (2016). *Springer handbook of robotics*. Springer.

Silver, D., Huang, A., Maddison, C. J., Guez, A., Sifre, L., Van Den Driessche, G., ... & Hassabis, D. (2016). Mastering the game of Go with deep neural networks and tree search. *Nature*, 529(7587), 484-489.

Song, X., Wu, M., & Jermaine, C. (2007). Conditional anomaly detection. *IEEE Transactions on Knowledge and Data Engineering*, 19(5), 631-645. <https://doi.org/10.1109/TKDE.2007.1015>

Sutton, R. S., McAllester, D. A., Singh, S. P., & Mansour, Y. (2000). Policy gradient methods for reinforcement learning with function approximation. In *Advances in Neural Information Processing Systems* (pp. 1057-1063).

Szegedy, C., Zaremba, W., Sutskever, I., Bruna, J., Erhan, D., Goodfellow, I., & Fergus, R. (2014). Intriguing properties of neural networks. *arXiv preprint arXiv:1312.6199*.

Szeliski, R. (2010). *Computer vision: Algorithms and applications*. Springer.

Ton, Z., & Raman, A. (2010). The effect of product variety and inventory levels on retail store sales: A longitudinal study. *Production and Operations Management*, 19(5), 546-560.

Turing, A. M. (1950). Computing machinery and intelligence. *Mind*, 59(236), 433-460.

van der Maaten, L., & Hinton, G. (2008). Visualizing data using t-SNE. *Journal of Machine Learning Research*, 9, 2579-2605.

Vaughan, J. W., Wallach, H., Wortman Vaughan, J., & Wallach, H. (2018). A human-centered agenda for intelligible machine learning. *arXiv preprint arXiv:1801.00004*.

- Wang, D., Zhu, J., & Pierson, E. (2018). Unsupervised learning in genomics and systems biology. *PLoS Computational Biology*, 14(11), e1006578.
- Watkins, C. J., & Dayan, P. (1992). Q-learning. *Machine Learning*, 8(3-4), 279-292.
- Williams, R. J. (1992). Simple statistical gradient-following algorithms for connectionist reinforcement learning. *Machine Learning*, 8(3-4), 229-256.
- Wu, Z., Pan, S., Chen, F., Long, G., Zhang, C., & Philip, S. Y. (2020). A comprehensive survey on graph neural networks. *IEEE Transactions on Neural Networks and Learning Systems*, 32(1), 4-24.
- Yarowsky, D. (1995). Unsupervised word sense disambiguation rivaling supervised methods. In *Proceedings of the 33rd Annual Meeting on Association for Computational Linguistics* (pp. 189-196).
- Yu, C., Liu, J., & Nemati, S. (2020). Reinforcement learning in healthcare: A survey. *ACM Computing Surveys*, 53(3), 1-36.
- Yoon, J., & Arik, S. O. (2023). Unsupervised and Semi-Supervised Anomaly Detection with Data-Centric Machine Learning.
- Zadrozny, B., Langford, J., & Abe, N. (2003). Cost-sensitive learning by cost-proportionate example weighting. In *Third IEEE International Conference on Data Mining* (pp. 435-442). IEEE.
- Zhao, Y., Nasrullah, Z., Li, Z. (2019). "Pyod: A python toolbox for scalable outlier detection" (PDF). *Journal of Machine Learning Research*. 20. <https://www.jmlr.org/papers/volume20/19-011/19-011.pdf>
- Zhu, X., & Ghahramani, Z. (2002). Learning from labeled and unlabeled data with label propagation. In *Proceedings of the 15th Annual Conference on Neural Information Processing Systems* (pp. 912-919).