ΠΑΝΕΠΙΣΤΗΜΙΟ ΘΕΣΣΑΛΙΑΣ

Σχολή Τεχνολογίας

Τμήμα Ψηφιακών Συστημάτων

Εντοπισμός Απάτης σε Διαδικτυακές Αξιολογήσεις

ΠΤΥΧΙΑΚΗ ΕΡΓΑΣΙΑ

ΚΑΡΑΣΑ ΑΛΕΞΑΝΔΡΟΣ (ΑΜ: 3120023)

Επιβλέπων: ***Φώτιος Κόκκορας, Επίκουρος Καθηγητής***

ΛΑΡΙΣΑ, Οκτώβριος 2025

UNIVERSITY OF THESSALY

School of Technology

Digital Systems Department

Detection of Fraudulent Online Reviews

BACHELOR THESIS

KARASA ALEXANDROS(RN:3120023)

Supervisor: ***Fotios Kokkoras, Assistant Professor***

LARISSA, October 2025

Υπεύθυνη Δήλωση περί Ακαδημαϊκής Δεοντολογίας και Πνευματικών Δικαιωμάτων

Με πλήρη επίγνωση των συνεπειών του νόμου περί πνευματικών δικαιωμάτων, δηλώνω ρητά ότι η παρούσα πτυχιακή εργασία, καθώς και τα ηλεκτρονικά αρχεία και πηγαίοι κώδικες που αναπτύχθηκαν ή τροποποιήθηκαν στα πλαίσια αυτής της εργασίας, αποτελεί αποκλειστικά προϊόν προσωπικής μου εργασίας, δεν προσβάλλει κάθε μορφής δικαιώματα διανοητικής ιδιοκτησίας, προσωπικότητας και προσωπικών δεδομένων τρίτων, δεν περιέχει έργα/εισφορές τρίτων για τα οποία απαιτείται άδεια των δημιουργών/δικαιούχων και δεν είναι προϊόν μερικής ή ολικής αντιγραφής, οι πηγές δε που χρησιμοποιήθηκαν περιορίζονται στις βιβλιογραφικές αναφορές και μόνον, και πληρούν τους κανόνες της επιστημονικής παράθεσης. Τα σημεία όπου έχω χρησιμοποιήσει ιδέες, κείμενο, αρχεία ή/και πηγές άλλων συγγραφέων, αναφέρονται ευδιάκριτα στο κείμενο με την κατάλληλη παραπομπή και η σχετική αναφορά περιλαμβάνεται στο τμήμα των βιβλιογραφικών αναφορών με πλήρη περιγραφή.

Αναλαμβάνω πλήρως, ατομικά και προσωπικά, όλες τις νομικές και διοικητικές συνέπειες που δύναται να προκύψουν στην περίπτωση κατά την οποία αποδειχθεί, διαχρονικά, ότι η εργασία αυτή ή τμήμα της δεν μου ανήκει διότι είναι προϊόν λογοκλοπής.

Ο/Η Δηλών/ούσα

(Υπογραφή)

ΚΑΡΑΣΑ ΑΛΕΞΑΝΔΡΟΣ

14/10/2025

Εγκρίνεται από την Επιτροπή Εξέτασης:

Επιβλέπων/πουσα	**Ονοματεπώνυμο Επιβλέποντα** Βαθμίδα/ιδιότητα επιβλέποντα, Τμήμα Ψηφιακών Συστημάτων, Πανεπιστήμιο Θεσσαλίας
Μέλος	**Ονοματεπώνυμο Μέλους 1** Βαθμίδα/ιδιότητα μέλους 1, Τμήμα/Ίδρυμα μέλους 1
Μέλος	**Ονοματεπώνυμο Μέλους 2** Βαθμίδα/ιδιότητα μέλους 2, Τμήμα/Ίδρυμα μέλους 2

Ημερομηνία έγκρισης: dd-mm-yyyy

Περίληψη

Στον σύγχρονο ψηφιακό κόσμο του διαδικτύου σημειώνεται μία ραγδαία εξάπλωση πρακτικών ψευδών αξιολογήσεων, η οποία τείνει να παραπλανάει τους χρήστες διαστρεβλώνοντας την πραγματική εικόνα διάφορων προϊόντων ή υπηρεσιών, καθιστώντας μάλιστα διάφορες επιχειρήσεις να στρέφονται σε διαδικασίες αξιολόγησης της αξιοπιστίας των κριτικών.

Η παρούσα πτυχιακή εργασία εστιάζει στη διερεύνηση και την ανάπτυξη σύνθετων διαδικασιών για τον εντοπισμό απατηλών αξιολογήσεων. Αναλύονται διαδικτυακές κριτικές και κριτές, σε συνδυασμό με χωρικά και χρονικά μεταδεδομένα των αξιολογήσεων. Στόχος της είναι ο εντοπισμός απατηλών αξιολογήσεων σε σύνολο δεδομένων αξιολόγησης πρατηρίων καυσίμων της εφαρμογής fuelGR.

Γίνεται μία συστηματική διερεύνηση των δεδομένων με χρήση SQL ερωτημάτων, για την ανίχνευση ύποπτων μοτίβων και ανωμαλιών, όπως για παράδειγμα την παρουσία υψηλής επανάληψης αρνητικών κριτικών σε μικρό χρονικό διάστημα. Για την επίτευξη αυτού του στόχου, υλοποιήθηκαν αρχικά κάποια διακριτά κριτήρια ανίχνευσης που αντιπροσωπεύουν διαφορετικές μεθόδους για την εύρεση ύποπτων συμπεριφορών. Στη συνέχεια, τα αποτελέσματα αυτά συνδυάστηκαν σε ένα ενιαίο τελικό προσδιοριστικό κριτήριο, μέσω μίας διαδικασίας τελικού φιλτραρίσματος στην οποία εφαρμόστηκαν τα κατάλληλα βάρη ώστε να κατηγοριοποιηθούν οι αξιολογήσεις ως απατηλές ή πραγματικές. Προκύπτει έτσι ένα εμπλουτισμένο dataset το οποίο πλέον περιλαμβάνει εκτίμηση για την ποιότητα κάθε αξιολόγησης.

Abstract

In the modern digital world of the internet, there is a rapid spread of fake reviews, which tend to mislead users by distorting the true image of various products or services, even making various businesses turn to procedures for evaluating the reliability of reviews.

This thesis focuses on the investigation and development of complex procedures for identifying fraudulent reviews. Online reviews and reviewers are analyzed, in combination with spatial and temporal metadata of the reviews. Its goal is to identify fraudulent reviews in a dataset of gas station reviews of the fuelGR application.

A systematic investigation of the data is carried out using SQL queries, to detect suspicious patterns and anomalies, such as the presence of a high repetition of negative reviews in a short period of time. To achieve this goal, we initially implemented some discrete detection criteria representing different methods for finding suspicious behaviors. These results were then combined into a single final identifying criterion, through a final filtering process in which the appropriate weights were applied to categorize the reviews as fraudulent or real. This results in an enriched dataset that now includes an estimate of the quality of each review.

Ευχαριστίες

Στο σημείο αυτό θα ήθελα να ευχαριστήσω θερμά τον επιβλέποντα καθηγητή, Κύριο Φώτιο Κόκκορα για την βοήθεια και την στήριξη που μου προσέφερε καθώς και για την καθοδήγηση του σε όλη τη διάρκεια της εργασίας.

Θα ήθελα συγχρόνως να ευχαριστήσω τους φίλους μου για τις όμορφες στιγμές που μοιραστήκαμε και για την αισιοδοξία που μου μετέδιδαν καθ' όλη τη διάρκεια των σπουδών μου.

Επίσης, οφείλω ένα μεγάλο ευχαριστώ στους γονείς μου και τον αδερφό μου Βαγγέλη για την συνεχή συμπαράσταση και υποστήριξή που μου προσέφεραν ώστε να ολοκληρώσω με επιτυχία τις σπουδές μου στο τμήμα των Ψηφιακών Συστημάτων.

Καράσα Αλέξανδρος

14/10/2025

Περιεχόμενα

ΠΕΡΙΛΗΨΗ ..... **VII**

ABSTRACT ..... **IX**

ΕΥΧΑΡΙΣΤΙΕΣ ..... **XI**

ΠΕΡΙΕΧΟΜΕΝΑ ..... **XIII**

1 ΕΙΣΑΓΩΓΗ **1**

2 ΘΕΩΡΗΤΙΚΟ ΥΠΟΒΑΘΡΟ **3**

2.1 Η Σημασια των Διαδικτυακων Αξιολογησεων 3

2.2 Το Προβλημα των Ψευδων Αξιολογησεων 4

2.3 Ο Ρολος της Μηχανικης Μαθησης 5

2.3.1 Επιβλεπόμενη Μάθηση (Supervised Learning) *5*

2.3.2 Μη Επιβλεπόμενη Μάθηση (Unsupervised Learning) *9*

2.3.3 Βαθιά Μάθηση (Deep Learning) *13*

2.4 Μοντελα Αξιολογησης (Evaluation Models) 15

3 ΑΝΑΣΚΟΠΗΣΗ ΒΙΒΛΙΟΓΡΑΦΙΑΣ **21**

4 ΜΕΘΟΔΟΛΟΓΙΑ **27**

4.1 Αναλυση δεδομενων 27

4.2 Αναπτυξη SQL Σεναριων 31

4.3 Διαχειριση των δεδομενων 50

5 ΣΥΜΠΕΡΑΣΜΑΤΑ **53**

ΒΙΒΛΙΟΓΡΑΦΙΑ **55**

1 Εισαγωγή

Στον σημερινό ψηφιακό κόσμο, οι διαδικτυακές αξιολογήσεις καθορίζουν σημαντικά τις αγορές συμβάλλοντας σε έναν μεγάλο βαθμό στην διαμόρφωση των αποφάσεων των καταναλωτών, οι οποίοι τείνουν να εμπιστεύονται κριτικές που παρουσιάζουν μία ταύτιση μεταξύ τους. Μία θετική εικόνα κριτικών μπορεί να λειτουργήσει σαν ένα προωθητικό μήνυμα προς τον καταναλωτή με τη δυνατότητα να τον πείσει για την επιλογή του, ενώ από την άλλη, μία παρουσία αρνητικής εικόνας μπορεί να τον αποθαρρύνει από την αγορά. Αυτή η συμπεριφορά παρατηρείται καθώς οι κριτικές διαθέτουν τον ρόλο των αυθεντικών και αμερόληπτων μαρτυριών για οποιοδήποτε προϊόν ή υπηρεσία. Ωστόσο, η εμπιστοσύνη που εκπέμπεται από το σύστημα αξιολογήσεων μπορεί να δημιουργήσει εσφαλμένες εντυπώσεις σε περίπτωση όπου το σύστημα τίθεται ευάλωτο στην κατάχρηση, αναιρώντας την πραγματική αξία των κριτικών. Με την αυξανόμενη τάση των ψευδών κριτικών στο διαδίκτυο, η διασφάλιση της αξιοπιστίας των συστημάτων αξιολόγησης καθίσταται ολοένα και πιο κρίσιμη, θέτοντας επιτακτική την ανάγκη εύρεσης αποτελεσματικών λύσεων για τον περιορισμό του φαινομένου. Για την αντιμετώπιση αυτού του φαινομένου έχει παρατηρηθεί μεγάλη αύξηση της ερευνητικής δραστηριότητας στο αντικείμενο, συνδυάζοντάς το με την επιστήμη των δεδομένων και την μηχανική μάθηση.

Ενόψει αυτών των προκλήσεων, η παρούσα πτυχιακή “Εντοπισμός Απάτης σε Διαδικτυακές Αξιολογήσεις” επικεντρώνεται στην αναζήτηση και την ανάπτυξη διερευνητικών τεχνικών για τον εντοπισμό ύποπτων μοτίβων σε αξιολογήσεις. Η εργασία στηρίζεται σε δεδομένα που παρέχονται από το dataset της εφαρμογής fuelGR, με ακριβείς πληροφορίες για τις αξιολογήσεις των χρηστών. Τα δεδομένα αναλύθηκαν αρχικά με χρήση Excel για να αποκτηθεί μία βασική κατανόηση των χαρακτηριστικών τους. Πάνω σε αυτά τα δεδομένα, δημιουργήθηκαν ερωτήματα SQL για την αναζήτηση ύποπτων μοτίβων τα οποία οδηγούν στην εύρεση των ψευδών αξιολογήσεων. Οι αξιολογήσεις, σε κάθε διαφορετικό ερώτημα χωρίστηκαν σε δύο ομάδες ανάλογα με την γνησιότητα τους, και ύστερα από την εκτέλεση όλων των ερωτημάτων εφαρμόστηκαν συντελεστές βαρύτητας για την εξαγωγή του τελικού αποτελέσματος.

Βασικό στόχο της πτυχιακής έθεσε ο εντοπισμός των ύποπτων απατηλών αξιολογήσεων ώστε τα δεδομένα να λάβουν ετικέτες που θα τα διαχωρίζουν, συμβάλλοντας στην μείωση κακόβουλων απατηλών ενεργειών.

Στα κεφάλαια της πτυχιακής που ακολουθούν, θα αναλυθεί το αντικείμενο της μελέτης, ξεκινώντας από το 2ο κεφάλαιο όπου παρουσιάζονται οι βασικές τεχνικές μηχανικής μάθησης. Συνεχίζοντας, στο 3ο κεφάλαιο πραγματοποιείται αναφορά σε σημαντικές έρευνες που έχουν διενεργηθεί σχετικά με τον περιορισμό του προβλήματος, ενώ στο 4ο κεφάλαιο τεκμηριώνεται η διαδικασία η οποία ακολουθήθηκε, καθώς και η επεξήγηση της, συνοδευόμενη από τα αποτελέσματα. Τέλος, στο 5ο κεφάλαιο, συνοψίζονται τα βασικά συμπεράσματα τα οποία αναδείχθηκαν.

2 Θεωρητικό Υπόβαθρο

Στο συγκεκριμένο κεφάλαιο, θα αναλυθούν οι κύριες τεχνικές που εφαρμόζονται για τον εντοπισμό ύποπτων αξιολογήσεων. Θα μελετηθούν προσεγγίσεις επιβλεπόμενης και μη επιβλεπόμενης μάθησης, καθώς επίσης και μοντέλα βαθιάς μάθησης. Επιπλέον θα εξεταστούν τα κύρια μοντέλα αξιολόγησης που χρησιμοποιούνται για την μέτρησή τους.

2.1 Η Σημασία των Διαδικτυακών Αξιολογήσεων

Οι διαδικτυακές αξιολογήσεις αποτελούν ένα από τα κυριότερα κριτήρια με τα οποία οι καταναλωτές διαμορφώνουν την σκέψη τους ώστε να προχωρήσουν στη λήψη της τελικής τους απόφασης σχετικά με κάποιο προϊόν. Κατέχουν επίσης κρίσιμο ρόλο στην φήμη της επιχείρησης και δημιουργούν μία αίσθηση η οποία μπορεί να ενισχύσει την βεβαιότητα του καταναλωτή για το προϊόν/την υπηρεσία, ή σε αντίθετη περίπτωση να επηρεάσει αρνητικά την αντίληψή του.

Σύμφωνα με την έρευνα (Meng Qi κ.α., 2022), οι αρνητικές αξιολογήσεις προσελκύουν μεγαλύτερη γνωστική προσοχή από τις θετικές, καθώς παρατηρείται πως οι καταναλωτές αφιερώνουν περισσότερο χρόνο πάνω σε αυτές. Επιβεβαιώνεται δηλαδή, ότι οι καταναλωτές χρησιμοποιούν τις αρνητικές αξιολογήσεις για να εντοπίσουν πιθανούς κινδύνους οι οποίοι μπορεί να αφορούν την ποιότητα του προϊόντος ή ακόμα και την αξιοπιστία του πωλητή. Με αυτόν τον τρόπο μειώνεται η πιθανότητα πραγματοποίησης μίας αποτυχημένης αγοράς. Επομένως, η μεγαλύτερη επιρροή των αρνητικών αξιολογήσεων αποδίδεται στην τάση των καταναλωτών να αποφεύγουν την αγορά προϊόντων με υψηλό κίνδυνο. Οι καταναλωτές από την άλλη, είναι πιο πιθανό να δώσουν θετικές αξιολογήσεις αν η εμπειρία τους είναι ανεκτή, δηλαδή εάν ο καταναλωτής δεν αντιμετώπισε κάποιο σοβαρό πρόβλημα κατά την χρήση του προϊόντος ή της υπηρεσίας, ακόμη και αν δεν ήταν ιδανική. Για παράδειγμα, το προϊόν μπορεί να μην ξεπερνά τις προσδοκίες, ωστόσο μπορεί να εκτελεί τη βασική του λειτουργία. Τέτοιες περιπτώσεις συχνά οδηγούν σε θετικές αξιολογήσεις διότι οι καταναλωτές έχουν τη τάση να επιβραβεύουν ανεκτές εμπειρίες διότι είναι λιγότερο επικριτικοί, ειδικά όταν περιέχεται η ομαλή διαδικασία αγοράς και παράδοσης.

Η σημασία των αξιολογήσεων υπερβαίνει την απλή παροχή πληροφοριών προς τους καταναλωτές, καθώς συνιστά βασικό μέσο για την μείωση της αβεβαιότητας της αγοράς και εργαλείο που επηρεάζει άμεσα την φήμη των επιχειρήσεων.

2.2 Το Πρόβλημα των Ψευδών Αξιολογήσεων

Ύστερα από την αναφορά της επίδρασης των διαδικτυακών αξιολογήσεων στους καταναλωτές και στις επιχειρήσεις, πρέπει να εξεταστεί ο κρίσιμος παράγοντας υπονόμευσης της αξιοπιστίας τους, τις ψευδείς αξιολογήσεις. Η αξιοπιστία των αξιολογήσεων αποτελεί ένα αμφιλεγόμενο ζήτημα καθώς οι ψευδείς διαδικτυακές κριτικές αλλοιώνουν την πραγματική εικόνα των προϊόντων. Οι ψευδείς αξιολογήσεις συχνά προέρχονται από χρήστες που επιδιώκουν να εξυπηρετήσουν εταιρικά ή προσωπικά συμφέροντα. Σύμφωνα με έρευνα (Department for Business and Trade, 2023) που πραγματοποιήθηκε , ποσοστό μεγέθους μεταξύ 11%-15% των αξιολογήσεων σε μεγάλες πλατφόρμες ηλεκτρονικού εμπορίου στο Ηνωμένο Βασίλειο οι οποίες αφορούν κατηγορίες ηλεκτρικών ειδών, είδη σπιτιού και κουζίνας και είδη αθλητισμού είναι ψευδείς, ενώ σε άλλες πλατφόρμες το ποσοστό αυτό μπορεί να φτάσει ακόμη μεγαλύτερο αριθμό της τάξης 25%-35%, επηρεασμένο όμως και από τον χαμηλότερο αριθμό χρηστών. Οι αξιολογήσεις αυτές θετικές ή αρνητικές προσπαθούν να οδηγήσουν τον καταναλωτή στην αγορά προϊόντων χαμηλότερης ποιότητας και να τον απομακρύνουν από ανταγωνιστικές επιλογές.

Μία από τις πηγές από τις οποίες προέρχονται αυτές οι κριτικές αποτελούν οι καλά οργανωμένες ομάδες που διευκολύνονται από τις διαδικτυακές πλατφόρμες και τα κοινωνικά δίκτυα. Βάσει της έρευνας (Brett Hollenbeck, Davide Proserpio & Sherry He, 2021), υπάρχουν οργανωμένες ομάδες σε κοινωνικά δίκτυα οι οποίες καταφέρνουν μέσω των διαφημίσεων σε αυτά τα δίκτυα να συγκεντρώσουν διάφορους χρήστες. Πολλοί πωλητές προϊόντων προσεγγίζουν χρήστες από αυτές τις ομάδες με σκοπό να αποκομίσουν θετικές αξιολογήσεις έναντι αμοιβής. Σε άλλες περιπτώσεις, οι καταναλωτές δίνουν ψευδείς αξιολογήσεις για ανταλλάγματα κουπονιών ή εκπτώσεων. Ωστόσο, οι ψευδείς αξιολογήσεις δεν αποτελούν πάντα αποτέλεσμα οργανωμένων ενεργειών καθώς πολλές φορές προέρχονται από μεμονωμένες απλούστερες περιπτώσεις, όπως για παράδειγμα η λήψη θετικών βαθμολογιών για κάποιον πωλητή από γνωστούς του ανθρώπους, ή στην περίπτωσή την οποία θα αναλύσουμε στην πορεία, κάποιος ιδιοκτήτης πρατηρίου καυσίμων μπορεί να βαθμολογήσει κάποιο γειτονικό ανταγωνιστικό πρατήριο με αρνητική βαθμολογία. Τέτοιες περιπτώσεις είναι δυσκολότερο να εντοπιστούν καθώς συνήθως δεν

ακολουθούν κάποιο επαναλαμβανόμενο μοτίβο. Όσο εξελίσσεται η τεχνολογία, τέτοιες μέθοδοι οι οποίες μπορούν να δημιουργήσουν μία ψεύτικη εικόνα γύρω από κάποιο προϊόν, μπορούν να γίνουν όλο και ευκολότερες κι' αυτό διότι με την χρήση της τεχνητής νοημοσύνης μπορεί να παραχθεί κείμενο το οποίο θα είναι πειστικό σε λίγα δευτερόλεπτα. Συνεπώς, είναι αναγκαία η θέσπιση και η εφαρμογή κατάλληλων μηχανισμών ασφαλείας για τις αξιολογήσεις, προκειμένου να διασφαλιστεί η αξιοπιστία, η αυθεντικότητα και η διαφάνεια των κριτικών για προϊόντα και για υπηρεσίες.

2.3 Ο Ρόλος της Μηχανικής Μάθησης

2.3.1 Επιβλεπόμενη Μάθηση (Supervised Learning)

Μία από τις συνηθέστερες πρακτικές στον εντοπισμό ψευδών αξιολογήσεων συνιστά η χρήση της επιβλεπόμενης μάθησης, για δεδομένα τα οποία έχουν ετικέτες. Η επιβλεπόμενη μάθηση αποτελεί τεχνική της μηχανικής μάθησης που στηρίζεται στην εκπαίδευση ενός μοντέλου σε labeled data, δηλαδή ισχύει πως για κάθε είσοδο υπάρχει ο αντίστοιχος συνδυασμός με την σωστή έξοδο. Στόχος του μοντέλου είναι να μάθει τη συσχέτιση μεταξύ εισόδων και εξόδων, ώστε να έχει την δυνατότητα πρόβλεψης της σωστής εξόδου για νέα άγνωστα δεδομένα. Η διαδικασία περιλαμβάνει την χρήση ενός συνόλου δεδομένων εκπαίδευσης (training data) που περιέχουν δεδομένα εισόδου γνωστά και ως χαρακτηριστικά (πληροφορίες που περιγράφουν κάθε δείγμα) και τα αντίστοιχα δεδομένα εξόδου (labels). Ύστερα, κατά την διαδικασία της μάθησης ο αλγόριθμος επεξεργάζεται αυτά τα δεδομένα εκπαίδευσης, μαθαίνοντας τις σχέσεις μεταξύ των χαρακτηριστικών εισόδου και των labels της εξόδου, μέσω της προσαρμογής των παραμέτρων του μοντέλου ώστε να ελαχιστοποιείται η διαφορά των προβλέψεων του και των ετικετών. Μετά την διαδικασία της εκπαίδευσης, αξιολογείται το μοντέλο βάσει του testing dataset,. Στο επόμενο βήμα, βελτιστοποιείται η απόδοση του μοντέλου μέσω ρύθμισης των παραμέτρων και χρήσης τεχνικών όπως την cross-validation για να επιτευχθεί ισορροπία μεταξύ μεροληψίας (bias) και διασποράς (variance) ώστε να διασφαλιστεί ότι το μοντέλο είναι αξιόπιστο στην πράξη καθώς μπορεί να διαχειριστεί και νέα άγνωστα δεδομένα με υψηλό βαθμό ακρίβειας.

Πριν τη μελέτη των αλγορίθμων πρέπει να τονιστεί πως η ταξινόμηση (classification), αποτελεί σημαντική διαδικασία κατά την οποία πραγματοποιείται η κατηγοριοποίηση της εξόδου σε διαφορετικές κλάσεις με βάση των μεταβλητών εισόδου. Χρησιμοποιείται

όταν η τιμή της μεταβλητής εξόδου είναι διακριτή ή κατηγορική (όπως 'Ναι' ή 'Όχι' που αποτελεί περίπτωση δυαδικής ταξινόμησης).

Η επιλογή του κατάλληλου αλγορίθμου αποτελεί καθοριστικό παράγοντα για την επιτυχία ενός μοντέλου επιβλεπόμενης μάθησης. Ο τύπος των δεδομένων καθώς και ο στόχος που επιδιώκεται, συνιστούν τους σημαντικούς παράγοντες που διαμορφώνουν την επιλογή του κατάλληλου αλγόριθμου. Ορισμένοι από τους οι οποίοι χρησιμοποιούνται συχνότερα αποτελούν:

Decision trees

Το decision tree ή αλλιώς δέντρο αποφάσεων, είναι ένας από τους πιο γνωστούς αλγόριθμους κατηγοριοποίησης και λειτουργεί ως ταξινομητής καθώς βασίζεται στον διαχωρισμό των δειγμάτων σε κατηγορίες μέσω της ιεραρχικής διάσπασης του χώρου των δεδομένων. Αποτελείται από κόμβους, με τον πρώτο κόμβο να ονομάζεται ρίζα, και κάθε κόμβος που περιέχει κάποια συνθήκη ονομάζεται εσωτερικός κόμβος. Οι τελικοί κόμβοι οι οποίοι ολοκληρώνουν το δέντρο ονομάζονται φύλλα και αντιστοιχούν σε κλάσεις που περιέχουν την τελική απόφαση. Κάθε εσωτερικός κόμβος ελέγχει ένα χαρακτηριστικό (attribute) του δείγματος εισόδου και εφαρμόζει μία συνθήκη η οποία διαιρεί τον χώρο των δειγμάτων σε 2 ή περισσότερα επιμέρους τμήματα μέσω τόξων, ανάλογα με την διακριτή συνάρτηση των τιμών εισόδου. Η πλοήγηση του δέντρου γίνεται από την ρίζα μέχρι τα φύλλα ανάλογα με τα αποτελέσματα που προκύπτουν. Έτσι κάθε κόμβος επισημαίνεται με το χαρακτηριστικό που ελέγχει, και τα κλαδιά του φέρουν τις αντίστοιχες τιμές. Κύριο χαρακτηριστικό των δέντρων αποφάσεων είναι πως μπορούν να χρησιμοποιηθούν εύκολα, εξάγοντας κανόνες που μπορούν να κατανοηθούν εύκολα από τον χρήστη.

Naïve Bayes

Ο ταξινομητής Naïve Bayes στοχεύει μέσω χρήσης του θεωρήματος Bayes στη κατηγοριοποίηση των δεδομένων βάσει των πιθανοτήτων διαφορετικών κλάσεων, τηρούμενων των χαρακτηριστικών (features) των δεδομένων, με τα χαρακτηριστικά όμως να συμβάλλουν στην πρόβλεψη χωρίς σχέση μεταξύ τους. Η ταξινόμηση, παρέχει πρακτικούς αλγόριθμους μάθησης και μπορεί να συνδυάσει παρατηρούμενα δεδομένα, παρέχοντας χρήσιμη προοπτική για κατανόηση και αξιολόγηση των αλγορίθμων. Υπολογίζει σαφείς πιθανότητες για υποθέσεις ενώ παραμένει ανθεκτικό στον θόρυβο των δεδομένων εισόδου. Αυτό σημαίνει ότι το μοντέλο μπορεί να διατηρεί την αξιόπιστη απόδοση ακόμα και όταν τα δεδομένα περιέχουν σφάλματα.

$$P(y|X) = \frac{P(X|y) \times P(y)}{P(X)}$$

Εξίσωση 1: Bayes' Theorem

Το μοντέλο Naïve Bayes χωρίζεται σε 3 κύριες κατηγορίες

1. *Gaussian Naïve Bayes*: Στην συγκεκριμένη τεχνική, συνεχείς τιμές που σχετίζονται με κάθε χαρακτηριστικό κατανέμονται σύμφωνα με Gaussian κατανομή, δηλαδή, όταν απεικονίζονται γραφικά σχηματίζεται καμπύλη σε σχήμα καμπάνας, συμμετρική ως προς τον μέσο όρο των τιμών των μεταβλητών.
2. *Multinomial Naïve Bayes*: Χρησιμοποιείται σε περιπτώσεις που features αναπαριστούν την συχνότητα όρων (όπως πλήθος λέξεων) σε κάποιο έγγραφο. Συνήθως εφαρμόζεται σε ταξινόμηση κειμένου.
3. *Bernoulli Naïve Bayes*: Βασίζεται σε δυαδικά χαρακτηριστικά όπου καθένα από αυτά υποδεικνύει αν εμφανίζεται ή όχι κάποια λέξη στο κείμενο. Συνεπώς, όπως και το Multinomial Naïve Bayes, χρησιμοποιείται κυρίως για ταξινόμηση κειμένου όμως δίνει έμφαση στην παρουσία ή απουσία όρων αντί της συχνότητάς τους.

Logistic Regression

Η λογιστική παλινδρόμηση, λειτουργεί εξάγοντας κάποιο σύνολο σταθμισμένων χαρακτηριστικών από την είσοδο, λαμβάνοντας καταγραφές (logs) και συνδυάζοντάς τα γραμμικά, που σημαίνει πως κάθε χαρακτηριστικό πολλαπλασιάζεται με ένα βάρος και τα αποτελέσματα προστίθενται μεταξύ τους. Είναι ένας τύπος παλινδρόμησης που προβλέπει την πιθανότητα εμφάνισης ενός γεγονότος προσαρμόζοντας δεδομένα σε μία λογιστική συνάρτηση, χρησιμοποιώντας αρκετές μεταβλητές πρόβλεψης είτε αριθμητικές είτε κατηγορικές.

Η κύρια διαφορά μεταξύ λογιστικής παλινδρόμησης και Naïve Bayes είναι πως η λογιστική παλινδρόμηση είναι ένας διακριτικός ταξινομητής ενώ ο Naïve Bayes είναι ένας γενετικός ταξινομητής.

Linear Regression

Στόχο της γραμμικής παλινδρόμησης αποτελεί ο εντοπισμός σχέσεων και εξαρτήσεων μεταξύ μεταβλητών. Αναπαριστά μία σχέση μοντέλου ανάμεσα σε μία συνεχόμενη εξαρτημένη μεταβλητή (γνωστή ως label ή target) και μία ή περισσότερες επεξηγηματικές

μεταβλητές που μπορούν να είναι μεταβλητές εισόδου, ανεξάρτητες μεταβλητές, χαρακτηριστικά (Vladimir Nasteski, 2017).

Support Vector Machine

Η Μηχανή Διανυσματικής Υποστήριξης (Support Vector Machine-SVM), αποτελεί αλγόριθμο της επιβλεπόμενης μηχανικής μάθησης ο οποίος χρησιμοποιείται για εργασίες ταξινόμησης και παλινδρόμησης. Σκοπός του είναι να βρει το βέλτιστο όριο (hyperplane) που χωρίζει τα δεδομένα σε διαφορετικές κλάσεις. Προσπαθεί να μεγιστοποιήσει το περιθώριο (margin) ανάμεσα σε δύο κλάσεις, όπου όσο μεγαλύτερο είναι το περιθώριο, τόσο αποδοτικότερο είναι το μοντέλο σε νέα άγνωστα δεδομένα. Η SVM θεωρείται ιδιαίτερα χρήσιμη για περιπτώσεις που αφορούν την δυαδική ταξινόμηση. Ο τρόπος με τον οποίο λειτουργεί ο αλγόριθμος, είναι τέτοιος ώστε να βρεθεί το υπέρ-επίπεδο (hyperplane) που χωρίζει αποδοτικότερα τις δύο κλάσεις, μεγιστοποιώντας το περιθώριο ανάμεσά τους. Το περιθώριο είναι η απόσταση από το υπερεπίπεδο μέχρι τα πλησιέστερα σημεία δεδομένων (διανύσματα υποστήριξης) σε κάθε πλευρά. Έτσι, το υπέρεπίπεδο που μεγιστοποιεί την απόσταση μεταξύ του υπέρεπιπέδου και των πλησιέστερων σημείων δεδομένων από κάθε κλάση, θεωρείται βέλτιστο (hard margin), διασφαλίζοντας έναν σαφή διαχωρισμό μεταξύ των κλάσεων.

Ανάλογα με την φύση του ορίου απόφασης, τα SVM μπορούν να χωριστούν σε δύο βασικές κατηγορίες.

1. *Linear SVM:* Χρησιμοποιείται γραμμικό όριο απόφασης ώστε να διαχωριστούν τα σημεία δεδομένων διαφορετικών κλάσεων. Μέσω μίας ευθείας γραμμής ή με την χρήση ενός υπερεπίπεδου (σε περισσότερες διαστάσεις), καθίσταται δυνατός ο διαχωρισμός των σημείων των δεδομένων στις αντίστοιχες κλάσεις τους.
2. *Non*-Linear SVM: Χρησιμοποιείται για την ταξινόμηση δεδομένων που δεν μπορούν να διαχωριστούν με ευθεία γραμμή. Χειρίζονται τα δεδομένα που δεν είναι γραμμικά διαχωρίσιμα μέσω χρήσης πυρηνικών συναρτήσεων (kernel functions), οι οποίες μετασχηματίζουν τα αρχικά δεδομένα εισόδου σε έναν χώρο χαρακτηριστικών μεγαλύτερων διαστάσεων όπου τα σημεία μπορούν να διαχωριστούν γραμμικά. Πάνω σε αυτόν τον νέο χώρο εφαρμόζεται Linear SVM ώστε να βρεθεί γραμμικό υπερεπίπεδο το οποίο θα έχει μη γραμμικό όριο απόφασης (όπως καμπύλη ή κύκλο).

2.3.2 Μη Επιβλεπόμενη Μάθηση (Unsupervised Learning)

Η μη επιβλεπόμενη μάθηση αποτελεί κλάδο της μηχανικής μάθησης, η οποία σε αντίθεση με την επιβλεπόμενη μάθηση που αναλύθηκε προηγουμένως, ασχολείται με μη επισημασμένα δεδομένα. Οι unsupervised learning αλγόριθμοι έχουν ως αποστολή την εύρεση μοτίβων και σχέσεων σε δεδομένα τα οποία δεν διαθέτουν προηγούμενη γνώση για την σημασιολογική τους ερμηνεία (input). Έχουν την ικανότητα να βρίσκουν 'κρυφά μοτίβα' χωρίς ανθρώπινη παρέμβαση, δηλαδή χωρίς να τους δίνεται η έξοδος (output). Ο τρόπος της εκτέλεσής του είναι τέτοιος όπου δέχεται ακατέργαστα δεδομένα και ύστερα από την αρχική ανάλυση και επεξεργασία των δεδομένων, επιλέγεται ο κατάλληλος αλγόριθμος με στόχο την εύρεση νέων άγνωστων προτύπων πληροφορίας. Τα κύρια χαρακτηριστικά των τεχνικών μη επιβλεπόμενης μάθησης περιλαμβάνουν τη δυνατότητα αναγνώρισης χαρακτηριστικών που διευκολύνουν τη δημιουργία ομάδων (clustering), την αξιολόγηση δεδομένων σε πραγματικό χρόνο κατά την εισαγωγή τους, καθώς και την ευκολότερη διαχείριση μη επισημασμένων δεδομένων από τα υπολογιστικά συστήματα, σε σύγκριση με τα επισημασμένα, τα οποία απαιτούν ανθρώπινη παρέμβαση για να λάβουν ετικέτες.

Γίνεται αντιληπτό λοιπόν, ότι οι αλγόριθμοι που χρησιμοποιούνται στη μη επιβλεπόμενη μάθηση εκτελούν συνήθως πιο σύνθετα καθήκοντα, καθώς καλούνται να ανακαλύψουν πρότυπα χωρίς την βοήθεια ετικετών. Επιπλέον, είναι συνήθως λιγότερο προβλέψιμη σε σχέση με άλλες μορφές μάθησης καθώς δεν λαμβάνει καθοδήγηση. Οι κυριότεροι τύποι αλγορίθμων που χρησιμοποιούνται στη μη επιβλεπόμενη μάθηση κατατάσσονται σε τέσσερις βασικές κατηγορίες: ομαδοποίηση (clustering), ανίχνευση ανωμαλιών (anomaly detection), νευρωνικά δίκτυα (neural networks) και μείωση διαστάσεων (dimensionality reduction).

Η ομαδοποίηση (clustering) αποτελεί μία από τις πιο διαδεδομένες τεχνικές μη επιβλεπόμενης μάθησης, καθώς χρησιμοποιείται για την εύρεση δομών ή μοτίβων από αδόμητα δεδομένα. Η διαδικασία της ομαδοποίησης, αφορά τη διαίρεση ενός συνόλου δεδομένων σε ομάδες, με τρόπο τέτοιο, ώστε τα σημεία δεδομένων που ανήκουν στην ίδια ομάδα να παρουσιάζουν παρόμοια χαρακτηριστικά, ενώ εκείνα τα σημεία που ανήκουν σε άλλες διαφορετικές ομάδες να διαφέρουν μεταξύ τους. Οι ομάδες που προκύπτουν ονομάζονται συστάδες (clusters).

Υπάρχουν διάφορες προσεγγίσεις που αφορούν την εκτέλεση της ομαδοποίησης, ωστόσο, οι κυριότερες από αυτές διακρίνονται στις ακόλουθες κατηγορίες.

Partitional Clustering (Διαχωριστική Ομαδοποίηση): Τα ακατέργαστα δεδομένα διαχωρίζονται σε ορισμένες ομάδες, τέτοιες ώστε, κάθε σημείο δεδομένων να ανήκει μόνο σε μία από αυτές.

Ένας ευρέως χρησιμοποιούμενος αλγόριθμος διαχωριστικής ομαδοποίησης είναι ο *k-means*, ο οποίος εφαρμόζεται σε μεγάλο αριθμό εφαρμογών εξαιτίας της απλότητάς και της αποτελεσματικότητάς του. Στην μέθοδο k-means τα ακατέργαστα δεδομένα κατηγοριοποιούνται σε k ακέραιες συστάδες, με κάθε συστάδα να αναπαριστάται από το κέντρο της, χαρακτηριστικό ως centroid, που αντιστοιχεί στον αριθμητικό μέσο όρο των σημείων δεδομένων της συστάδας. Η ομαδοποίηση των δεδομένων βασίζεται στον υπολογισμό και την ελαχιστοποίηση της απόστασης των σημείων δεδομένων από τα επιλεγμένα centroids, με το πλήθος των centroids να εξαρτάται από τον απαιτούμενο αριθμό ομάδων. Τα centroids, ανανεώνουν τις συστάδες προσθέτοντας τα σημεία που βρίσκονται πλησιέστερα σε αυτά, μέσω μίας επαναληπτικής διαδικασίας. Πιο συγκεκριμένα, ο αλγόριθμος k-means αρχικά, επιλέγει τον αριθμό των συστάδων (k), και στην συνέχεια επιλέγονται τα k σημεία από το σύνολο δεδομένων ως αρχικά κεντροειδή. Τα υπόλοιπα σημεία δεδομένων ταξινομούνται στη συστάδα του πλησιέστερου κεντροειδούς, συνήθως μέσω χρήσης της Ευκλείδειας απόστασης. Με κάθε προσθήκη νέου σημείου, το κέντρο κάθε συστάδας υπολογίζεται εκ νέου, ενημερώνοντας τα κεντροειδή. Η διαδικασία αυτή επαναλαμβάνεται με κάθε ανανέωση των κεντροειδών, όπου ελέγχεται ξανά σε ποια συστάδα ανήκει κάθε σημείο βάσει της απόστασης του από τα ενημερωμένα κεντροειδή. Επομένως, αν μετά την αλλαγή των κεντροειδών, κάποιο σημείο βρίσκεται κοντινότερα σε διαφορετικό κεντροειδές από πριν, μετακινείται σε νέα συστάδα. Η διαδικασία αυτή επαναλαμβάνεται μέχρι να επιτευχθεί σύγκλιση, δηλαδή μέχρι τα σημεία να μην αλλάζουν πια συστάδα σε κάθε νέα επανάληψη. Έτσι, προκύπτει ένα σύνολο k συστάδων με ετικέτες.

Hierarchical Clustering (Ιεραρχική Ομαδοποίηση): Όπως αναφέρει η ονομασία, η τεχνική αυτή, οργανώνει τα δεδομένα σε μία ιεραρχική δενδροειδή δομή από clusters. Μέσω σταδιακής συγχώνευσης ή διαίρεσης των clusters, επιδιώκεται η εύρεση μοτίβων και σχέσεων μεταξύ των σημείων δεδομένων. Το αποτέλεσμα λαμβάνει την μορφή δενδρογράμματος, αποτυπώνοντας, ανάλογα τον βαθμό ομοιότητας, τα επίπεδα συγχώνευσης ή διαχωρισμού των clusters.

Βασική προσέγγιση της ιεραρχικής ομαδοποίησης αποτελεί η *Συσσωρευτική (Agglomerative – bottom-up)*, που αποτελεί την συνηθέστερη μορφή ιεραρχικής ομαδοποίησης, κατά την οποία κάθε σημείο δεδομένων είναι αρχικά ξεχωριστό cluster, και στην πορεία τα clusters συγχωνεύονται μέχρι να σχηματιστεί μία ενιαία ομάδα.

Πιο συγκεκριμένα, στην συσσωρευτική ιεραρχική ομαδοποίηση, κάθε σημείο δεδομένων θεωρείται ανεξάρτητο cluster. Δηλαδή, για Ν σημεία δεδομένων, υπάρχουν Ν ξεχωριστές μεμονωμένες ομάδες. Στην συνέχεια, οι δύο ομάδες που βρίσκονται πιο κοντά μεταξύ τους συγχωνεύονται σε μία νέα ενιαία ομάδα μέσω κάποιου κριτηρίου απόστασης. Το βήμα αυτό επαναλαμβάνεται και κάθε φορά επιλέγονται 2 ομάδες που έχουν την μικρότερη απόσταση μεταξύ τους σε σχέση με άλλες περιπτώσεις ομάδων. Έτσι ο αριθμός των αρχικών ομάδων μειώνεται κατά ένα κάθε φορά ώσπου όλα τα δεδομένα να καταλήξουν σε μία μοναδική ομάδα. Με την οπτική απεικόνιση της διαδικασίας μέσω ενός δενδρογράμματος που αναπαριστά την δομή της ομαδοποίησης σε διάφορα επίπεδα, παρέχεται η δυνατότητα να καθοριστεί το ύψος στο οποίο θα διαχωριστούν οι ομάδες, ανάλογα με τον επιθυμητό αριθμό ομάδων.

Όπως αναφέρθηκε, η πλησιέστερη απόσταση μεταξύ των ομάδων είναι σημαντική, και οι τρόποι με τους οποίους μπορεί να μετρηθεί διαφέρουν λόγω του κριτηρίου σύνδεσης (linkage method). Μία από τις μεθόδους σύνδεσης, αποτελεί η μονή σύνδεση (single linkage), η οποία αφορά την απόσταση μεταξύ των δύο πιο κοντινών σημείων δεδομένων από διαφορετικές ομάδες. Οδηγεί σε πιο ομοιογενείς και συμπαγείς ομάδες καθώς αποτρέπεται η συγχώνευση ομάδων που περιλαμβάνουν σημεία πολύ απομακρυσμένα μεταξύ τους. Μία ακόμη μέθοδο σύνδεσης συνιστά η average, στην οποία υπολογίζεται η μέση απόσταση μεταξύ 2 ομάδων, με τον υπολογισμό της απόστασης κάθε πιθανού ζεύγους σημείων δεδομένων των 2 ομάδων και λαμβάνοντας τον μέσο όρο όλων αυτών των αποστάσεων που προκύπτουν. Επιπλέον, η μέθοδος centroid linkage, υπολογίζει την απόσταση των κεντροειδών σημείων των 2 ομάδων. Συνεπώς, ανάλογα με τη φύση των δεδομένων, κάθε μέθοδος σχηματίζει διαφορετικά την μορφή και το μέγεθος των τελικών ομάδων που προκύπτουν.

Επιπλέον προσέγγιση της ιεραρχικής ομαδοποίησης αποτελεί η *Διαιρετική (Divisive – top-down),* όπου τα δεδομένα ξεκινούν ως ένα ενιαίο cluster, το οποίο στην συνέχεια διαχωρίζεται σε μικρότερα clusters, ώσπου να επιτευχθεί η επιθυμητή ανάλυση.

Overlapping Clustering (Επικαλυπτόμενη Ομαδοποίηση): Ένα σημείο δεδομένων μπορεί να ανήκει σε περισσότερες από μία συστάδες, γεγονός που επιτρέπει μία πλουσιότερη

αναπαράσταση των δεδομένων, αναγνωρίζοντας πως κάποιο αντικείμενο μπορεί να παρουσιάζει χαρακτηριστικά που σχετίζονται με διαφορετικές ομάδες ταυτόχρονα.

Ένας χαρακτηριστικός αλγόριθμος επικαλυπτόμενης συσταδοποίησης, αποτελεί ο αλγόριθμος Fuzzy Clustering (Ασαφής Συσταδοποίηση), ο οποίος επιτρέπει σε ένα σημείο δεδομένων να ανήκει σε παραπάνω από μία συστάδα ταυτόχρονα. Βασίζεται σε μία επαναληπτική διαδικασία βελτιστοποίησης, κατά την οποία εκχωρείται βαθμός συμμετοχής στα σημεία δεδομένων αντί για αυστηρές ετικέτες συσταδοποίησης. Πιο αναλυτικά, για κάθε σημείο δεδομένων αποδίδεται ο αντίστοιχος βαθμός συμμετοχής του για όλες τις συστάδες. Αυτές οι τιμές υποδεικνύουν τη πιθανότητα ή τον βαθμό συσχέτισης του σημείου με κάθε ομάδα, υποδηλώνοντας σε τι βαθμό ανήκει σε κάθε μία από αυτές, σε αντίθεση με αυστηρές μορφές συσταδοποίησης που ένα σημείο ανήκει μόνο σε μία ομάδα. Επομένως, επιτρέπεται η μερική συμμετοχή των δεδομένων σε διάφορες συστάδες. Στη συνέχεια, υπολογίζονται τα κεντροειδή των συστάδων με βάση το σταθμισμένο άθροισμα όλων των σημείων δεδομένων, όπου οι βαθμοί συμμετοχής λειτουργούν ως συντελεστές στάθμισης. Έτσι, βεβαιώνεται πως τα σημεία με υψηλότερη συμμετοχή σε μία συστάδα συνεισφέρουν περισσότερο στο κεντροειδές. Ύστερα από τον υπολογισμό των centroids, λαμβάνει μέρος η μέτρηση της απόστασης μεταξύ κάθε σημείου με τα κεντροειδή, ο οποίος υπολογισμός επαναπροσδιορίζει τους βαθμούς συμμετοχής. Τα σημεία που βρίσκονται πιο κοντά στο κεντροειδές, έχουν μεγαλύτερη συμμετοχή στην αντίστοιχη ομάδα. Η διαδικασία αυτή επαναλαμβάνεται έως ότου οι βαθμοί συμμετοχής σταθεροποιηθούν, σημαίνοντας ότι δεν συμβαίνουν σημαντικές μεταβολές από την μία επανάληψη στην άλλη. Σε αυτό το στάδιο, θεωρείται πως η διαδικασία της ομαδοποίησης βρίσκεται σε βέλτιστη κατάσταση. Σε μερικές περιπτώσεις, μπορεί να χρειαστεί η μετατροπή ορισμένων ασαφών συμμετοχών, κατατάσσοντας κάθε σημείο δεδομένων στην συστάδα με την οποία έχει τον μεγαλύτερο βαθμό συσχέτισης.

Probabilistic Clustering (Πιθανολογική Ομαδοποίηση): Βασίζεται στη χρήση πιθανοτήτων για την κατανομή των σημείων δεδομένων σε συστάδες. Εκτιμάται για κάθε σημείο η πιθανότητα συμμετοχής του σε κάθε διαθέσιμη συστάδα.

Το μοντέλο GMM (Gaussian Mixture Model) αποτελεί χαρακτηριστικό παράδειγμα στο οποίο η πιθανολογική ομαδοποίηση χρησιμοποιείται εκτενώς. Σε αυτό το μοντέλο, εκτιμάται πως τα δεδομένα προέρχονται από έναν συνδυασμό πολλών Gaussian κατανομών με άγνωστες παραμέτρους. Το δεδομένο δεν αναθέτεται αποκλειστικά σε μία συστάδα, αλλά μπορεί να ανήκει σε πολλαπλές συστάδες βάσει πιθανοτήτων. Κάθε

Gaussian κατανομή χαρακτηρίζεται από 3 βασικά στοιχεία. Την μέση τιμή που αποτελεί το κέντρο της κατανομής, την συνδιακύμανση η οποία περιγράφει την εξάπλωση και τον προσανατολισμό των δεδομένων και τον συντελεστή μίξης ο οποίος αφορά το ποσοστό συνεισφοράς κάθε Gaussian κατανομής στον τελικό συνδυασμό.

2.3.3 Βαθιά Μάθηση (Deep Learning)

Εκτός από τις κλασσικές μεθόδους μηχανικής μάθησης που αναλύθηκαν, η Βαθιά Μάθηση έχει αναδειχθεί ιδιαίτερα αποτελεσματική στον εντοπισμό απατηλών αξιολογήσεων. Η Βαθιά Μάθηση βασίζεται σε νευρωνικά δίκτυα πολλών επιπέδων τα οποία έχουν την ικανότητα να μαθαίνουν αυτόματα σύνθετες αναπαραστάσεις δεδομένων. Αναλυτικότερα, ένα νευρωνικό δίκτυο αποτελείται από επίπεδα διασυνδεδεμένων κόμβων ή νευρώνων που συνεργάζονται για την επεξεργασία των δεδομένων εισόδου. Σε ένα πλήρως συνδεδεμένο βαθύ νευρωνικό δίκτυο τα δεδομένα ρέουν από πολλαπλά επίπεδα όπου κάθε νευρώνας εκτελεί μη γραμμικούς μετασχηματισμούς, επιτρέποντας στο μοντέλο να μάθει περίπλοκες αναπαραστάσεις δεδομένων. Το βαθύ νευρωνικό δίκτυο, λαμβάνει τα δεδομένα στο επίπεδο εισόδου, τα οποία στη συνέχεια περνούν από κρυφά επίπεδα που τα μετασχηματίζουν ώστε να φτάσουν στο τελικό επίπεδο, το επίπεδο εξόδου όπου παράγεται η πρόβλεψη του μοντέλου. Αποτελεί την κύρια δομή στην οποία βασίζεται η βαθιά μάθηση, εξασφαλίζοντας υψηλή ακρίβεια όταν είναι διαθέσιμα μεγάλα σύνολα δεδομένων με χρήση σημαντικών υπολογιστικών πόρων και κατάλληλης αρχιτεκτονικής σχεδίασης. Σε κάθε επίπεδο του δικτύου που αναφέρθηκε μετασχηματίζονται τα δεδομένα σε πιο σύνθετες μορφές, δίνοντας στο σύστημα την δυνατότητα να αναγνωρίζει περίπλοκα μοτίβα και συσχετίσεις. Η συνήθης εφαρμογή της στον εντοπισμό ψευδών αξιολογήσεων, αφορά την αυτόματη ανάλυση σύνθετων γλωσσικών μοτίβων και την κατανόηση κειμένων καθώς οι ύποπτες αξιολογήσεις συχνά κρύβουν συχνότητες όρων όπου με την βοήθεια των βαθιών νευρωνικών δικτύων αναγνωρίζονται με ακρίβεια.

Τα μοντέλα RNNs (Recurrent Neural Networks) αποτυπώνουν μία ειδική κατηγορία βαθιάς μάθησης, καθώς διαφέρουν από τα συνηθισμένα νευρωνικά δίκτυα ως προς τον τρόπο με τον οποίο επεξεργάζονται τη πληροφορία. Τα αναδρομικά νευρωνικά δίκτυα σε κάθε βήμα τροφοδοτούν πληροφορίες πίσω στο δίκτυο (γνωστός και ως μηχανισμός ανατροφοδότησης ή αλλιώς feedback loop), ενώ στα τυπικά νευρωνικά δίκτυα η πληροφορία τροφοδοτείται ως προς μία κατεύθυνση (input to output). Αυτό συμβαίνει διότι τα RNNs διαθέτουν εσωτερική μνήμη, επομένως έχουν τη δυνατότητα να διατηρούν πληροφορίες από προηγούμενα βήματα, κατανοώντας συσχετίσεις και συμφραζόμενα. Από

την άλλη, FNN (Feedforward Neural Networks) μοντέλα όπως τα CNN (Convolutional Neural Networks) και τα MLP (Multilayer Perceptrons) δεν διαθέτουν μνήμη, επομένως διαχειρίζονται κάθε είσοδο δίχως να λαμβάνουν υπόψη τις προηγούμενες εισόδους. Αυτό δεν τα καθιστά αποδοτικά σε διαδοχικά δεδομένα (όπως τα κείμενα) καθώς απαιτούν μοντέλα με μνήμη ώστε να κατανοήσουν συσχετίσεις. Η κύρια μονάδα επεξεργασίας στα RNN μοντέλα είναι μια επαναλαμβανόμενη μονάδα (recurrent unit), όπου κρατώντας κρυφή στάση διατηρεί πληροφορίες για προηγούμενες εισόδους σε μία ακολουθία. Μπορεί να θυμάται πληροφορίες από προηγούμενα βήματα μέσω του μηχανισμού της ανατροφοδότησης, καταγράφοντας έτσι σχέσεις που υπάρχουν ανάμεσα σε δεδομένα τα οποία εμφανίζονται σε διάφορες χρονικές στιγμές. Η μονάδα αυτή αποτελεί τον αναδρομικό νευρώνα. Η διαδικασία της ανάπτυξης στα μοντέλα RNN αφορά την εξάπλωση της αναδρομικής δομής ανά χρονικά βήματα. Σε αυτή τη διαδικασία, κάθε βήμα της ακολουθίας αναπαρίσταται ως κάποιο διαφορετικό επίπεδο σε μία ακολουθία, απεικονίζοντας τον τρόπο με τον οποίο η πληροφορία ρέει σε κάθε χρονικό βήμα.

Η διαδικασία αυτή γνωστή ως unfolding, καθιστά ενεργή τη μέθοδο Backpropagation Through Time (BPTT). Στην μέθοδο αυτή της οπισθοδιάδοσης, τα σφάλματα διαδίδονται ανά χρονικά βήματα ώστε να ρυθμιστούν τα βάρη του δικτύου, βάσει των οποίων ενισχύεται η δυνατότητα του RNN να αντιλαμβάνεται εξαρτήσεις μέσα σε διαδοχικά δεδομένα. Τα Αναδρομικά Νευρωνικά Δίκτυα μπορούν να χωριστούν σε τέσσερις κατηγορίες ανάλογα τον αριθμό εισόδων και εξόδων στο δίκτυο. Η απλούστερη μορφή one-to-one RNN διαθέτει μία είσοδο και μία έξοδο και χρησιμοποιείται για βασικές εργασίες ταξινόμησης όπως την δυαδική ταξινόμηση, εφόσον δεν έχουν να κάνουν με διαδοχικά δεδομένα. Τα one-to-many RNN δέχονται μία είσοδο και παράγουν πολλαπλές εξόδους, για παράδειγμα με την είσοδο μιας εικόνας μπορεί να παραχθεί μία αλληλουχία λέξεων ως λεζάντα. Αντίθετα, στην περίπτωση many-to-one RNN, μία ακολουθία εισόδων παράγει μία μοναδική έξοδο. Επιπλέον, τα many-to-many RNN επεξεργάζονται μία ακολουθία εισόδων και παράγουν μία ακολουθία εξόδων.

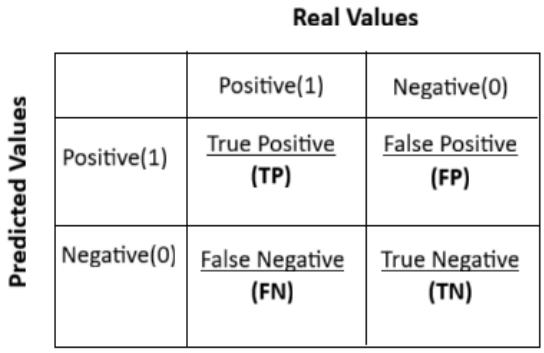
Έναν εξίσου σημαντικό τύπο νευρωνικών δικτύων αποτελούν τα CNN (Convolutional Neural Networks). Τα Συνελικτικά Νευρωνικά Δίκτυα είναι σχεδιασμένα για την επεξεργασία δεδομένων που μπορούν να αναπαρασταθούν σε μορφή πλέγματος. Η ονομασία τους προέρχεται από την μαθηματική πράξη της συνέλιξης όπου ο ρόλος της στην συγκεκριμένη περίπτωση είναι η σάρωση που πραγματοποιεί το φίλτρο στην εικόνα, και οι υπολογισμοί συνδυασμών των τιμών ώστε να αναδειχθούν συγκεκριμένα

χαρακτηριστικά. Τα δίκτυα αυτά, επιτρέπουν τον αυτόματο εντοπισμό στην ιεραρχική εξαγωγή χαρακτηριστικών μέσα από οπτικά δεδομένα όπως ακμές και πιο σύνθετα μοτίβα, τα οποία είναι απαραίτητα για την κατανόηση και την ανάλυση εικόνων. Τα κύρια χαρακτηριστικά των CNN αποτελούν τα συνελικτικά επίπεδα (convolutional layers), όπου μέσω της χρήσης φίλτρων εφαρμόζονται συνελικτικές πράξεις στις εικόνες εισόδου, και τα επίπεδα συγκέντρωσης (pooling layers) που χρησιμεύουν ιδανικά στην μείωση των χωρικών διαστάσεων της εισόδου για τον περιορισμό της υπολογιστικής πολυπλοκότητας και τον περιορισμό του αριθμού των παραμέτρων του δικτύου. Με αυτόν τον τρόπο διατηρούνται τα πιο σημαντικά χαρακτηριστικά της εικόνας, απορρίπτοντας τον θόρυβο, κάτι που καθιστά το δίκτυο πιο αποδοτικό σε μετατοπίσεις των δεδομένων. Επιπλέον, οι συναρτήσεις ενεργοποίησης (activation functions) εισάγουν ένα είδος μη γραμμικότητας στο μοντέλο, με αποτέλεσμα να μαθαίνει πολύπλοκες σχέσεις δεδομένων. Χωρίς αυτές τις συναρτήσεις, το νευρωνικό δίκτυο θα περιοριζόταν σε απλές γραμμικές συσχετίσεις, ανεξάρτητα του αριθμού των επιπέδων.

2.4 Μοντέλα Αξιολόγησης (Evaluation Models)

Εφόσον αναλύθηκαν μερικοί από τους κυριότερους αλγόριθμους μηχανικής μάθησης και προβλήθηκε η σημαντικότητα τους, καθίσταται απαραίτητη η αξιολόγηση τους, ώστε να είναι εφικτή η σύγκριση αναφορικά με την αποτελεσματικότητά τους. Η σύγκριση αυτή και η ανάλυση των διαφορών μεταξύ των μοντέλων, επιτυγχάνεται με την χρήση μιας σειράς μετρήσεων και διαδικασιών αξιολόγησης, που αποτελούν αντικειμενικό κριτήριο επιλογής του βέλτιστου μοντέλου. Τα μοντέλα αξιολόγησης εξετάζουν βασικές μετρήσεις απόδοσης με τις οποίες μπορεί να υπολογιστεί ποσοτικά η ικανότητα ενός συστήματος να διαχωρίσει τις γνήσιες από τις ύποπτες αξιολογήσεις. Μία από τις απλούστερες και θεμελιώδες μορφές μέτρησης για την αξιολόγηση ενός ταξινομητή αποτελεί το Confusion Matrix. Πρόκειται για έναν πίνακα NxN (όπου N οι κατηγορίες ταξινόμησης) διαστάσεων, αναπαριστώντας τον αριθμό των πραγματικών και των προβλεπόμενων αποτελεσμάτων. Στην περίπτωση του εντοπισμού απατηλών αξιολογήσεων που πρόκειται για ένα πρόβλημα δυαδικής ταξινόμησης (αληθές ή ψευδές) ο πίνακας αυτός είναι 2x2 διαστάσεων. Μπορεί να μας δώσει μία πλήρη εικόνα των σωστών και λανθασμένων προβλέψεων με την χρήση των 2 κατηγοριών που χωρίζονται σε πραγματικές και σε προβλεπόμενες τιμές. Ως True Positive (TP), μπορεί να χαρακτηριστεί μία αξιολόγηση που προβλέφθηκε σωστά ως ψεύτικη. Αντιθέτως, ως False Positive (FP), ονομάζεται μία αληθινή

κριτική η οποία όμως προβλέφθηκε λανθασμένα ως ψεύτικη. Τέτοιες περιπτώσεις θεωρούνται προβληματικές καθώς μπορούν να οδηγήσουν στην διαγραφή πραγματικών αξιολογήσεων, αλλοιώνοντας την αξιοπιστία του συνόλου. Με τον όρο False Negative (FN), χαρακτηρίζεται μία κριτική η οποία είναι ψευδής αλλά προβλέφθηκε λανθασμένα ως αληθινή. Μία τέτοια περίπτωση σηματοδοτεί πως δεν εντοπίστηκε η απατηλή αξιολόγηση, εντείνοντας ένα αναξιόπιστο περιεχόμενο κριτικών. Αντίθετα η περίπτωση των True Negative (TN) κριτικών, αποτελείται από τις πραγματικές αξιολογήσεις όπου προβλέφθηκαν σωστά ως αληθινές.



Εικόνα 1: Confusion Matrix

Παρόλο που ο πίνακας σύγχυσης καταγράφει μία ολοκληρωμένη εικόνα των επιδόσεων του ταξινομητή, για να ληφθούν συμπεράσματα σχετικά με την αξιοπιστία του και να είναι εφικτή η σύγκριση των αποτελεσμάτων, είναι απαραίτητος ο υπολογισμός στατιστικών μετρικών. Για τον σκοπό αυτό εφαρμόζονται συγκεκριμένες μετρικές αξιολόγησης όπως το Precision(P). Η ευστοχία (precision) μετράει την αξιοπιστία των θετικών προβλέψεων του μοντέλου, δηλαδή από όλες τις αξιολογήσεις που έκρινε ως ψεύτικες, πόσες από αυτές ήταν όντως ψευδείς (TP). Ένας υψηλός δείκτης ευστοχίας σημαίνει πως όταν το μοντέλο εντοπίζει μία ύποπτη κριτική, οι πιθανότητες να είναι ψευδής είναι μεγάλες. Έτσι ο κίνδυνος διαγραφής πραγματικών αξιολογήσεων ελαχιστοποιείται (FP).

$$Precision = \frac{TP}{TP + FP}$$

Εξίσωση 2: Ευστοχία πρόβλεψης

Η ευστοχία ως μοναδική στατιστική μετρική δεν αρκεί για μία πλήρη αξιολόγηση, καθώς δεν παρέχει πληροφορία για τον αριθμό των πραγματικών ψεύτικων αξιολογήσεων όπου αδυνατεί να εντοπίσει το μοντέλο. Για αυτό το σκοπό χρησιμοποιείται η Ευαισθησία πρόβλεψης (Sensitivity) γνωστή και ως Recall, η οποία μετρά την ικανότητα του

μοντέλου να εντοπίζει όλες τις πραγματικά ψεύτικες αξιολογήσεις. Πιο απλά, μετράει από όλες τις υπάρχουσες ψεύτικες αξιολογήσεις, πόσες από αυτές εντοπίζονται από το μοντέλο.

$$Sensitivity = \frac{TP}{TP + FN}$$

Εξίσωση 3: Ευαισθησία πρόβλεψης

Μία ακόμη θεμελιώδης μετρική αξιολόγησης αποτελεί η ακρίβεια (accuracy) η οποία ορίζεται ως το ποσοστό των σωστών προβλέψεων για τα δεδομένα δοκιμής (test data). Πρόκειται για μία εύκολα κατανοητή μετρική που παρέχει μία γενική εικόνα της απόδοσης.

$$\mathrm{A} = \frac{TP + TN}{TP + FP + TN + FN}$$

Εξίσωση 4: Ακρίβεια πρόβλεψης

Ενώ η ακρίβεια προσφέρει μία γενική εικόνα, με την ευστοχία και την ευαισθησία να εστιάζουν σε συγκεκριμένα είδη σφαλμάτων, απαιτείται συνήθως μία ενιαία μετρική όπου θα συνδυάζει αυτά τα δύο μοντέλα. Αυτό το κριτήριο εξυπηρετείται από το F1 Score. Αποτελεί τον μέσο όρο της ακρίβειας και της ανάκλησης, καθιστώντας το χρήσιμο στην επίτευξη της ισορροπίας μεταξύ τους, συνδυάζοντάς τα σε έναν αριθμό. Το εύρος του αριθμού αυτού μπορεί να είναι από το μηδέν έως το ένα και όσο μεγαλύτερο είναι το F1 score τόσο καλύτερα αποδίδει.

$$\mathrm{F1}\ Score = \frac{P * R}{P + R}$$

Εξίσωση 5: F1 Score

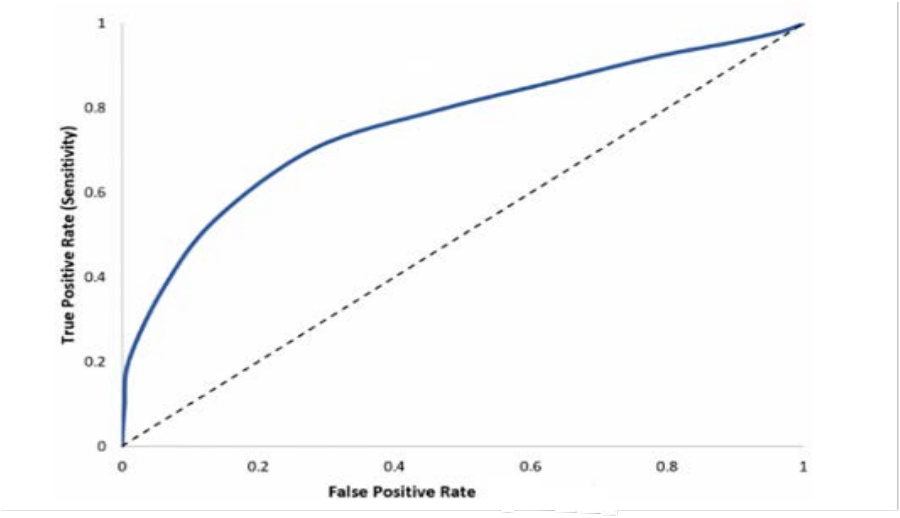
Μία ακόμη σημαντική μετρική συνιστά το ποσοστό σφάλματος (error rate), όπου μετράει την αναλογία των λανθασμένων προβλέψεων ως προς τον συνολικό αριθμό των αξιολογήσεων. Μπορεί να θεωρηθεί αντίστροφο της ακρίβειας.

$$Error\ Rate = \frac{(FP + FN)}{(TP + FP)}$$

Εξίσωση 6: Error rate

Επιπλέον, μία θεμελιώδη μετρική αποτελεί η καμπύλη λειτουργικού χαρακτηριστικού δέκτη (receiver operating characteristic – ROC) καθώς απεικονίζει την απόδοση ενός

μοντέλου σε κάθε πιθανό κατώφλι ή αλλιώς όριο απόφασης. Έτσι, μπορεί να παρέχει μία ολοκληρωμένη εικόνα της συμπεριφοράς του μοντέλου.



Εικόνα 2: Καμπύλη ROC

Στον οριζόντιο άξονα απεικονίζεται ο ρυθμός ψευδών θετικών (FPR-false positive rate) όπου ερμηνεύει το ποσοστό των αρνητικών περιπτώσεων που χαρακτηρίστηκαν λανθασμένα ως θετικές.

$$FPR = \frac{FP}{FP + TN}$$

Στον κατακόρυφο άξονα απεικονίζεται ο ρυθμός αληθινών θετικών (TPR-true positive rate) δηλαδή το sensitivity, αντιπροσωπεύοντας το ποσοστό των πραγματικά θετικών περιπτώσεων που εντοπίστηκαν σωστά.

$$TPR = \frac{TP}{TP + FN}$$

Ο άξονας X (FPR), δείχνει την αναλογία των πραγματικά αρνητικών περιπτώσεων (πραγματικές αξιολογήσεις) που ταξινομήθηκαν λανθασμένα ως θετικές (ψευδείς αξιολογήσεις).

Ο άξονας Y (TPR), δείχνει την αναλογία των πραγματικά θετικών περιπτώσεων (ψευδείς αξιολογήσεις) που εντοπίστηκαν σωστά.

Η απόδοση μπορεί να εξακριβωθεί βάσει της θέσης όπου βρίσκεται η καμπύλη μέσω της χρήσης AUC (Area Under Curve). Πιο αναλυτικά, όταν η καμπύλη βρίσκεται στην μονάδα του άξονα Y (πάνω αριστερή γωνία) τότε είναι ιδανική η αποτελεσματικότητα του ταξινομητή (AUC = 1). Είναι κατανοητό επομένως, ότι το υψηλό TPR υποδηλώνει

καλή ανίχνευση, και όσο χαμηλότερη είναι η μέτρηση του FPR υπάρχει χαμηλότερο κόστος λαθών. Σε περίπτωση που η καμπύλη βρίσκεται στην τιμή 0.5 (AUC = 0.5), η αποτελεσματικότητα του ταξινομητή ισοδυναμεί με τυχαία επιλογή, άρα δεν συνεισφέρει ουσιαστικά. Εάν η καμπύλη βρίσκεται ακόμα χαμηλότερα της τιμής 0.5 (AUC < 0.5), υποδηλώνει πως υπάρχει κάποιο σφάλμα (Petr Silhavy & Radek Silhavy, 2023; Elias Mmbongeni Sibanda, Gireen Naidu & Tranos Zuva, 2023).

Συνεπώς, τα μοντέλα αξιολόγησης συνιστούν κρίσιμες μετρικές που είναι απαραίτητες για την αντικειμενική αξιολόγηση της απόδοσης του ταξινομητή. Με την χρήση τους, υλοποιείται η σύγκριση διαφορετικών προσεγγίσεων ώστε να επιλεχθεί το βέλτιστο μοντέλο.

3 Ανασκόπηση Βιβλιογραφίας

Στο συγκεκριμένο κεφάλαιο, γίνεται αναφορά και ανάπτυξη της βιβλιογραφίας στο θέμα του εντοπισμού ψευδών αξιολογήσεων.

Πολλές μελέτες έχουν σταθεί στην ανάλυση των αξιολογήσεων με χρήση NLP(Natural Language Processing) η οποία αποτελεί διεπιστημονικό κλάδο της επιστήμης της πληροφορικής και της τεχνητής νοημοσύνης και μέσω της μηχανικής μάθησης επιτρέπει στους υπολογιστές να κατανοούν και να επικοινωνούν με ανθρώπινη γλώσσα. Η εφαρμογή της επεξεργασίας φυσικής γλώσσας, μπορεί να κατηγοριοποιηθεί σε πέντε βασικές περιοχές, πιο συγκεκριμένα στην κατανόηση φυσικής γλώσσας, στην παραγωγή φυσικής γλώσσας, στην μηχανική μετάφραση, στον γραμματικό έλεγχο και στην αναγνώριση ομιλίας ή φωνής. Μέσω NLP τεχνολογιών, είναι εφικτή η ανάλυση των δεδομένων ώστε να πραγματοποιηθεί εξόρυξη πληροφοριών από κείμενα, συμβάλλοντας στον εντοπισμό μοτίβων και συναισθημάτων που δεν είναι άμεσα εμφανή σε μεγάλα σύνολα δεδομένων. Η εφαρμογή της στο πρόβλημα των αξιολογήσεων, κατέχει την δυνατότητα αναγνώρισης λεκτικών μοτίβων και συναισθηματικής υπερβολής, που συχνά βρίσκονται σε παραπλανητικές κριτικές.

Στην μελέτη των (Ahmed Mohamed Kamel, κ.α., 2022), δόθηκε ιδιαίτερη έμφαση στον εντοπισμό ψευδών αξιολογήσεων μέσω αυτοματοποιημένων μεθόδων. Για να πραγματοποιηθεί η μελέτη, απαραίτητη ήταν η ανάγκη ανάπτυξης ενός συνόλου δεδομένων όπου υλοποιήθηκε μέσω αξιοποίησης NLP τεχνολογιών. Αρχικά, παράχθηκαν δειγματικές κριτικές με χρήση δύο γλωσσικών μοντέλων οι οποίες αξιολογήθηκαν μέσω ποσοτικών μετρήσεων και ποιοτικών εκτιμήσεων. Τα γλωσσικά μοντέλα που χρησιμοποιήθηκαν (ULMFiT, GPT-2), είναι βασισμένα στην μεταφορά μάθησης (transfer learning), η οποία στοχεύει στον περιορισμό των ειδικών ρυθμίσεων για κάθε εργασία και την αποτροπή της ανάγκης για εκπαίδευση του μοντέλου από την αρχή. Στη μεταφορά μάθησης, το γλωσσικό μοντέλο προ εκπαιδεύεται σε μεγάλο σώμα κειμένου και στη συνέχεια προσαρμόζεται λεπτομερώς (fine-tuning) σε συγκεκριμένες εργασίες για να επιτευχθεί υψηλή απόδοση. Το πρώτο μοντέλο, έχει μνήμη μακροπρόθεσμης και βραχυπρόθεσμης διάρκειας (LSTM) ως βασικό κορμό, ενώ το δεύτερο βασίζεται στην αρχιτεκτονική των μετασχηματιστών. Το τελικό dataset απαρτίστηκε από 40.000 αξιολογήσεις όπου μισές

από αυτές είναι τεχνητά παραγόμενες (ψεύτικες) και οι άλλες μισές είναι αληθινές κριτικές από ανθρώπους που χρησιμοποιήθηκαν ως δείγματα για εκπαίδευση (model training) και επικύρωση (model validation) του μοντέλου. Πάνω στο dataset που προέκυψε, πραγματοποιείται η εκπαίδευση ταξινομητών μηχανικής μάθησης για τον διαχωρισμό των πραγματικών και των παραγόμενων από υπολογιστή αξιολογήσεων. Χρησιμοποιήθηκαν ως βασικές γραμμές αναφοράς το μοντέλο NBSVM, ένας συνδυασμός Naïve-Bayes και αλγόριθμου SVM (Support Vector Machine) και το μοντέλο της OpenAI βασισμένο σε fine-tuning, εκπαιδευμένο ειδικά για την ανίχνευση ψεύτικων αξιολογήσεων, πάνω στο οποίο βασίζεται το νέο μοντέλο. Τα μοντέλα αυτά, εκπαιδεύτηκαν και αξιολογήθηκαν με αναλογία διαχωρισμού όπου το μεγαλύτερο κομμάτι του συνόλου των δεδομένων χρησιμοποιείται για εκπαίδευση (training) και το υπόλοιπο για αξιολόγηση (testing), με την απόδοσή τους να ελέγχεται μέσω μετρήσεων μηχανικής μάθησης.

Σε άλλη εξίσου σημαντική μελέτη (Aliaksandr Barushka, Michal Munk & Petr Hajek, 2020), εξετάστηκε η χρήση των βαθιών νευρωνικών δικτύων DFFNN (deep-feed-forward neural network) και CNN για την εξαγωγή πολύπλοκων χαρακτηριστικών που είναι κρυμμένα σε αναπαραστάσεις λέξεων, προτάσεων και συναισθημάτων υψηλής διάστασης. Στόχο της συγκρότησε η ανάπτυξη ενός μοντέλου αποτελεσματικής αξιοποίησης των συμφραζομένων των αξιολογήσεων, με χρήση της μεθόδου bag-of-words και με αναπαραστάσεις συναισθημάτων, πάνω σε πρότυπα σύνολα δεδομένων ενός ευρύς φάσματος προϊόντων και καταναλωτών. Τα συγκεκριμένα σύνολα δεδομένων, είναι κατασκευασμένα με τέτοιο τρόπο ώστε να είναι ισορροπημένα, δηλαδή να απαρτίζονται από τον ίδιο αριθμό γνήσιων και ψευδών αξιολογήσεων. Αυτός είναι ένας σημαντικός παράγοντας για να εξασφαλιστεί ότι τα μοντέλα που θα εκπαιδευτούν σε αυτά, δεν θα είναι μεροληπτικά σε κάποια από τις δύο κατηγορίες. Για την εξαγωγή των bag-of-words χαρακτηριστικών πραγματοποιήθηκε προεπεξεργασία με σκοπό να μειωθεί ο θόρυβος των δεδομένων και να διεξαχθεί κατάλληλα το tokenization για την κατάτμηση των κειμένων σε λέξεις. Επίσης, για την αναπαράσταση των συναισθημάτων, χρησιμοποιήθηκαν τρεις τύποι λεξικών δεικτών συναισθήματος, που αναπαραστάθηκαν από τα λεξικά που επιλέχθηκαν. Θεωρείται απαραίτητη η επιλογή λεξικών που περιέχουν ετικέτες συναισθήματος για κάθε λέξη (όπως το NRC). Με χρήση του μοντέλου Skip-Gram, πραγματοποιήθηκε η δημιουργία των ενσωματωμένων λέξεων (word embeddings) αντιστοιχίζοντας τις λέξεις σε αριθμητικά διανύσματα. Το υλοποιημένο DFFNN μοντέλο της έρευνας, εμφανίζεται ως ένα πολυεπίπεδο νευρωνικό δίκτυο με δύο κρυφά επίπεδα. Στο επίπεδο εισόδου του μοντέλου, πραγματοποιήθηκε η εξαγωγή χαρακτηριστικών από το ακατέργαστο

κείμενο των αξιολογήσεων σε κατηγορίες, βάσει της συχνότητας εμφάνισης και της σπανιότητας των όρων σε σχέση με το σύνολο όλων των κειμένων, καθώς επίσης και βάσει χαρακτηριστικών συναισθημάτων, και μέσων ενσωματώσεων λέξεων για κάθε κριτική, υπολογισμένων από το προεκπαιδευμένο μοντέλο ώστε να αναπαρίσταται ως ένα ενιαίο διάνυσμα. Τα δύο κρυφά επίπεδα, επεξεργάζονται τις σύνθετες σχέσεις μεταξύ των χαρακτηριστικών εισόδου και των κλάσεων εξόδου. Στο CNN μοντέλο, πραγματοποιήθηκε μετατροπή όλων των προτάσεων σε μία k-διάστατη αναπαράσταση λέξεων με χρήση των προεκπαιδευμένων ενσωματώσεων, όπου ως k αντιπροσωπεύεται η διάσταση (αριθμός) των ενσωματώσεων. Επίσης, για κάθε λέξη υπολογίστηκαν οι tf.idf τιμές όπου υπολογίζουν τη βαρύτητά τους, και δείκτες συναισθήματος ώστε οι λέξεις να συνοδεύονται από την συναισθηματική τους φόρτιση. Στο συνελικτικό επίπεδο διεκπεραιώθηκαν εκτεταμένες δοκιμές για την εύρεση του κατάλληλου αριθμού φίλτρων. Τελικά, χρησιμοποιήθηκε ένα τυπικό μέγεθος φίλτρου ίσο με πέντε και μία συνάρτηση ενεργοποίησης ReLU (Rectified Linear Unit), η οποία κρατά τις θετικές τιμές και μηδενίζει τις αρνητικές. Ύστερα από δοκιμές που πραγματοποιήθηκαν στην έρευνα, αποδείχθηκε πως τα CNN μοντέλα ήταν αποδοτικότερα όταν εκπαιδεύονται σε δεδομένα μίας πολικότητας συναισθήματος. Από την άλλη μεριά, το DFFNN μοντέλο, θεωρήθηκε αποτελεσματικότερο στην επεξεργασία συνδυασμένων συναισθημάτων.

Είναι άξιο αναφοράς πως η πλειοψηφία των ερευνών εστιάζει στη μελέτη των σχολίων των κριτικών καθώς η ανάλυση του γλωσσικού περιεχομένου αποτελεί το κυριότερο πεδίο μελέτης. Σαφώς λοιπόν, μεγάλο μέρος της σχετικής βιβλιογραφίας επικεντρώνεται σε τεχνικές επεξεργασίας φυσικής γλώσσας για τον εντοπισμό γλωσσικών μοτίβων, λαμβάνοντας τα κείμενα ως την κύρια πηγή πληροφορίας. Συνεπώς, άλλες πτυχές των δεδομένων όπως τα μεταδεδομένα του χρήστη δεν δέχονται ουσιαστική ανάλυση. Έτσι, ερευνητικές προσπάθειες που χρησιμοποιούν μεταδεδομένα εμφανίζουν ιδιαίτερο ενδιαφέρον καθώς δίνουν έμφαση στην αξιοποίηση πληροφοριών πέρα από το κείμενο της αξιολόγησης. Ένα χαρακτηριστικό παράδειγμα τέτοιας ερευνητικής προσέγγισης συνιστά η μελέτη (David Savage, κ.α., 2016), όπου χρησιμοποιείται δυαδική παλινδρόμηση για τον εντοπισμό αξιολογητών με ποσοστό βαθμολογιών που μπορεί να θεωρηθεί αμφισβητήσιμο, αποκλίνοντας από την γνώμη της πλειοψηφίας. Λαμβάνοντας υπόψιν ότι η πλειοψηφία των αξιολογήσεων προέρχεται από ειλικρινής χρήστες που έχουν παρόμοιες προσδοκίες για κάποιο προϊόν, εμφανίζεται σχετικά μικρή διακύμανση μεταξύ των κριτικών. Με χρήση της μέσης βαθμολογίας ως μέτρο της πλειοψηφικής γνώμης, θεωρείται

πως η χειραγώγηση της αντίληψης των καταναλωτών από spammers απαιτεί την αλλοίωση του μέσου όρου των αξιολογήσεων. Για τον εντοπισμό τέτοιων περιπτώσεων, μπορεί να μελετηθεί η πιθανότητα μίας αξιολόγησης να διαφέρει με την μέση τιμή, υπολογίζοντας τον spam βαθμό κάθε αξιολογητή με χρήση διωνυμικής κατανομής που καταλήγει σε δυαδική τιμή *επιτυχία* ή *αποτυχία*. Για την εφαρμογή της, χρησιμοποιείται ο αριθμός κριτικών κάποιου χρήστη και ο αριθμός των κριτικών του που διαφέρουν με την μέση τιμή αξιολογήσεων. Ένα σημαντικό βήμα της μελέτης αποτελεί η επαναληπτική διαδικασία καθώς διαθέτει βασικό ρόλο στην κατάλληλη μέτρηση των κριτικών που διαφωνούν με την πλειοψηφία, δίχως να χαρακτηρίζονται αφιλτράριστα ως spam. Η επαναληπτική διαδικασία προσαρμόζει επαναλαμβανόμενα τους μέσους όρους των κριτικών, χρησιμοποιώντας ως βάση την αξιοπιστία των χρηστών, και στη συνέχεια με χρήση ενός κατωφλιού εξετάζεται εάν οι αποκλίσεις των κριτικών έχουν σταθεροποιηθεί. Έτσι μπορεί να εφαρμοστεί ο διωνυμικός έλεγχος για κάθε χρήστη και να εξεταστεί αν η συμπεριφορά του αποκλίνει σημαντικά από την αναμενόμενη.

Πέρα από τη συγκεκριμένη μελέτη, η έρευνα (Arjun Mukherjee & Santosh K C, 2016) χρησιμοποιώντας ως βάση τα σύνολο δεδομένων της Yelp. Αρχικά, στόχο της μελέτης έθεσε η εύρεση του αριθμού των spam αξιολογήσεων που λειτουργούν με προωθητικό χαρακτήρα, υπολογίζοντας τα χρονικά δεδομένα της μέσης αθροιστικής βαθμολογίας των αξιολογήσεων, για να διαμορφωθεί η εικόνα της συνολικής βαθμολογίας με την πάροδο του χρόνου. Οι χρονοσειρές αυτές ευθυγραμμίστηκαν με το αρχικό χρονικό βήμα ως σημείο αναφοράς και στην συνέχεια πραγματοποιήθηκε ομαδοποίηση των πιθανών παραπλανητικών χρονικών δεδομένων των αξιολογήσεων. Έτσι κατέστη εφικτός ο εντοπισμός τριών κυρίαρχων προωθητικών τακτικών. Στην πρώτη ομάδα, αυτή του πρώιμου spamming το μοτίβο προωθητικής τακτικής εμφανίζεται νωρίς και εκτελείται σταθερά μετά τους πρώτους πέντε μήνες. Στη δεύτερη ομάδα που χαρακτηρίζεται μεσαίου spamming, τέτοια ενέργεια καθυστερεί να εμφανιστεί και ξεκινά μετά τον 14^{ο} μήνα, ενώ στην ομάδα του αργού spamming ξεκινά πολύ αργότερα με μία παύση δέκα μηνών πριν να φτάσει τη κορυφή παραπλανητικών αξιολογήσεων. Σε όλες τις παραπάνω περιπτώσεις, είναι ενδεικτικό ότι οι εταιρίες λειτουργούν με αθέμιτο τρόπο ώστε να δημιουργήσουν μία θετική εικόνα η οποία δεν είναι αντιπροσωπευτική της πραγματικότητας, καθώς καλύπτουν τις αρνητικές αξιολογήσεις που έχουν δεχτεί. Σε πολλές καταστάσεις, μέσω ανάλυσης συσχετίσεων και χρονοσειρών παρατηρήθηκε πως με την μείωση των αληθινών θετικών κριτικών, σημειώνεται αύξηση των ψευδών θετικών αξιολογήσεων λειτουργώντας ως αντιστάθμιση (buffered spamming). Μάλιστα, εφόσον η εταιρία

αποκτήσει αρκετές αληθινές θετικές αξιολογήσεις ώστε να έχει καλή εικόνα, τότε μειώνεται το ποσοστό των ψευδών αξιολογήσεων που δέχεται, γεγονός το οποίο συνδέεται άμεσα με την μείωση των γνήσιων αρνητικών αξιολογήσεων που λαμβάνει (μειωμένο spamming). Παρατηρούνται ωστόσο και περιπτώσεις όπου το spamming συμβαίνει προληπτικά ώστε να διατηρηθεί υψηλή βαθμολογία εάν λάβουν μελλοντικά αρνητικές κριτικές. Τέλος, η έρευνα με χρήση δεδομένων από το yelp και με ταξινόμηση μέσω SVM, προχώρησε στην δοκιμή των χαρακτηριστικών που αναπτύχθηκαν σε προηγούμενες μελέτες, δίνοντας έμφαση στην σημασία της ανάλυσης των χρονικών προτύπων στην πρόβλεψη ψευδών αξιολογήσεων.

Συνολικά, οι μελέτες δείχνουν πως οι τεχνικές μηχανικής μάθησης συμπεριλαμβανομένων των βαθιών νευρωνικών δικτύων και γλωσσικών μοντέλων, μπορούν να οδηγήσουν σε σημαντικές λύσεις για τον εντοπισμό ψευδών αξιολογήσεων. Το ίδιο ισχύει και για μελέτες που λειτουργούν κυρίως χρησιμοποιώντας κάποιον συνδυασμό στατιστικών μεθόδων όπως την ανάλυση χρονικών προτύπων. Η αξιοποίηση αυτών των εργαλείων είναι σημαντική για την κατανόηση των στρατηγικών που χρησιμοποιούνται για spam, ενισχύοντας την προστασία τόσο των καταναλωτών αλλά και των επιχειρήσεων.

4 Μεθοδολογία

Σε αυτό το κεφάλαιο περιγράφεται η μεθοδολογία η οποία ακολουθήθηκε για τον εντοπισμό ψευδών αξιολογήσεων στο σύνολο δεδομένων του fuelGR.

Αξιοποιήθηκαν δύο βασικά εργαλεία ανάλυσης δεδομένων, το Excel για την οπτική αναπαράσταση και τη διερεύνηση ακραίων τιμών που μπορούν να παρατηρηθούν στα δεδομένα, και το σύστημα διαχείρισης σχεσιακών βάσεων δεδομένων MySQL(8.0.42) για τον εντοπισμό και την επεξεργασία ύποπτων αξιολογήσεων. Το excel αποτελεί βασικό πρόγραμμα διαχείρισης υπολογιστικών φύλλων με δυνατότητα να οργανώνει δεδομένα σε στήλες και γραμμές και επιτρέπει την εκτέλεση ή τη χρήση διάφορων εντολών όπως την πραγματοποίηση μαθηματικών συναρτήσεων και των ταξινομήσεων. Χρησιμοποιήθηκε για την αρχική κατανόηση των βαθμολογιών καθώς και για την εξαγωγή συμπερασμάτων από τα σενάρια που εκτελέστηκαν. Η Δομημένη Γλώσσα Ερωτημάτων (SQL) συνιστά γλώσσα προγραμματισμού για τη διαχείριση σχεσιακών βάσεων δεδομένων η οποία θεωρείται απαραίτητη για εργασίες ανάλυσης και διαχείρισης δεδομένων. Συγκεκριμένα, στην εργασία χρησιμοποιήθηκαν εξειδικευμένα SQL ερωτήματα μέσω των οποίων υλοποιήθηκαν ειδικά φίλτρα που εντοπίζουν ύποπτα μοτίβα.

4.1 Ανάλυση δεδομένων

Το σύνολο δεδομένων που χρησιμοποιήθηκε από την εφαρμογή του fuelGR αποτελείται από 106816 εγγραφές, οι οποίες αφορούν κριτικές χρηστών σε πρατήρια της Ελλάδας. Κάθε εγγραφή αναπαριστά μία κριτική η οποία αποτελείται από δέκα διαφορετικά χαρακτηριστικά παρέχοντας έτσι μία ολοκληρωμένη εικόνα. Τα χαρακτηριστικά που παρέχονται από το σύνολο δεδομένων είναι τα εξής:

- *deviceID*: Το μοναδικό αναγνωριστικό της συσκευής από την οποία ο χρήστης υποβάλλει την αξιολόγηση
- *stationGUID*: Το μοναδικό αναγνωριστικό του πρατηρίου που αξιολογείται
- *clientIP*: Η διεύθυνση IP του χρήστη από την οποία υποβάλλεται η αξιολόγηση

- *rating1/ rating2/ rating3*: Κριτήρια τα οποία βαθμολογεί ο χρήστης. Το όριο στο οποίο κυμαίνονται οι βαθμολογίες είναι από 1 έως 5, με το 1 να αντιστοιχεί στην χαμηλότερη δυνατή αξιολόγηση και το 5 στην υψηλότερη. Το rating1 αφορά την ποιότητα/ποσότητα, το rating2 αφορά την εξυπηρέτηση και το rating3 αφορά την εγκατάσταση.
- *ratingWhen*: Η πλήρης ημερομηνία στην οποία υποβλήθηκε η αξιολόγηση.
- *ratingApproved*: Δυαδική τιμή που δηλώνει αν η αξιολόγηση είναι εγκεκριμένη ή μη εγκεκριμένη (1 για εγκεκριμένες και 0 για μη εγκεκριμένες).
- *month/week*: Μήνας ή εβδομάδα κατά την οποία πραγματοποιήθηκε η βαθμολογία.

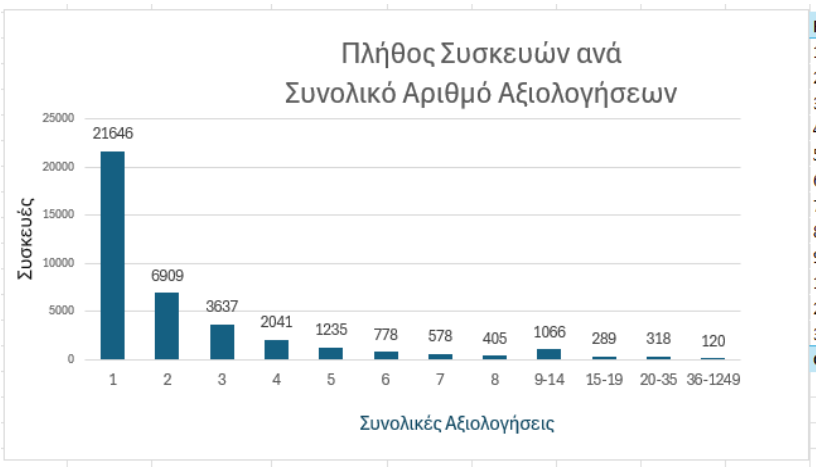
	A	B	C	D	E	F	G	H	I	J
1	deviceID	stationGUID	clientIP	rating1	rating2	rating3	ratingWhen	ratingApproved	month	week
2	android-3c166c85a3ba8626	39455D77-6132-3B66-BC0C-43B1965F	94.66.79.156	4	4	4	2017/08/07 18:15:19	1	2017-08	2017-32
3	android-3c166c85a3ba8626	60D78F46-853A-77DA-CC71-B3956E1F	94.66.79.156	4	4	4	2017/08/07 18:15:57	1	2017-08	2017-32
4	android-3c166c85a3ba8626	35B05CF4-5E99-1468-2B4D-165D485D	94.66.79.156	3	3	3	2017/08/07 18:16:28	1	2017-08	2017-32
5	android-3c166c85a3ba8626	5694D21C-E21B-3596-1A02-26284442	94.66.79.156	3	4	4	2017/08/07 18:17:29	1	2017-08	2017-32
6	android-3c166c85a3ba8626	D5198441-0DBF-2930-1BBA-516D31D3	94.66.79.156	3	4	4	2017/08/07 18:17:55	1	2017-08	2017-32
7	android-6678cff4eeb442a0	AE5EAD7B-C2E7-9C21-2F88-A71E5560	85.74.71.195	4	4	5	2017/08/07 18:18:10	1	2017-08	2017-32
8	android-3c166c85a3ba8626	B2B542F5-5C54-A28C-9672-DFA7BAD8	94.66.79.156	3	3	3	2017/08/07 18:18:42	1	2017-08	2017-32
9	android-6678cff4eeb442a0	FFFF8E9D-FAC8-07C7-DE50-1B48649A	85.74.71.195	3	3	2	2017/08/07 18:19:08	1	2017-08	2017-32
10	android-cea7f432b8cc64c6	B5DA7F5F-B824-8D40-B0BD-9802858A	62.1.92.52	5	4	5	2017/08/07 18:48:10	1	2017-08	2017-32

Εικόνα 3: Ενδεικτικό παράδειγμα του συνόλου δεδομένων.

Για την αρχική ανάλυση των δεδομένων εξετάστηκαν κάποια βασικά χαρακτηριστικά. Πιο συγκεκριμένα, μετρήθηκε ο αριθμός των διαφορετικών συσκευών που αξιολόγησαν πρατήρια και ο αριθμός των διαφορετικών πρατηρίων όπου δέχτηκαν κριτικές. Παρατηρήθηκε πως το πλήθος των συσκευών είναι ίσο με 39022 ενώ το πλήθος των πρατηρίων αντιστοιχεί σε 6701. Είναι σημαντικό να αναφερθεί πως ο χρήστης μπορεί να αξιολογήσει από την συσκευή του κάθε πρατήριο μόνο μία φορά, επομένως δεν υπάρχουν διπλότυπες αξιολογήσεις. Επομένως, ο συνδυασμός των deviceID και stationGUID είναι μοναδικός.

Στη συνέχεια, πραγματοποιήθηκε βασική ταξινόμηση των συσκευών βάσει του συνολικού αριθμού των αξιολογήσεων που έχουν υποβάλλει στα πρατήρια. Διαπιστώθηκε ότι το υψηλότερο πλήθος αξιολογήσεων που έχει πραγματοποιηθεί από μία συσκευή είναι 1249, αριθμός που δεν μπορεί να είναι πραγματικός από κανένα χρήστη. Για να μπορεί να πραγματοποιηθεί ωστόσο μία οπτική αναπαράσταση από την οποία μπορούν να παραχθούν συμπεράσματα, υπολογίστηκε το πλήθος των συσκευών για τον αντίστοιχο συνολικό αριθμό αξιολογήσεων που έχουν υποβάλλει, μέσω της χρήσης Συγκεντρωτικού Πίνακα (PivotTable). Ο συγκεντρωτικός πίνακας, αποτελεί χαρακτηριστική Excel λειτουργία με την οποία είναι δυνατή η ευκολότερη οργάνωση και ανάλυση των δεδομένων

καθώς προσφέρει την δυνατότητα πρόσθεσης ή αφαίρεσης τιμών, την εκτέλεση υπολογισμών και τη ταξινόμηση ή το φιλτράρισμα των δεδομένων. Όπως φαίνεται από την Εικόνα 4 η πλειοψηφία των συσκευών δεν έχει αξιολογήσει παραπάνω από ένα πρατήριο, ενώ επίσης, αξίζει να αναφερθεί ότι 120 συσκευές έχουν κάνει περισσότερες από 35 κριτικές.



Rating_Count	DeviceIDcount
1	21646
2	6909
3	3637
4	2041
5	1235
6	778
7	578
8	405
9-14	1066
15-19	289
20-35	318
36-1249	120
Grand Total	**39022**

Εικόνα 4: Πλήθος Συσκευών ανά Συνολικό Αριθμό Αξιολογήσεων.

Με παρόμοιο τρόπο, χρησιμοποιώντας συγκεντρωτικό πίνακα, υπολογίστηκε ο συνολικός αριθμός των πρατηρίων ανά το αντίστοιχο πλήθος αξιολογήσεων. Παρατηρήθηκε πως μεγάλο μέρος των πρατηρίων έχει δεχτεί μόνο πέντε ή λιγότερες αξιολογήσεις (Εικόνα 5).

StationRating Count	Stations Count
1	698
2	595
3	490
4	437
5	402
6	333
7-12	1295
13-18	733
19-29	720
30-50	584
51-81	255
82-293	159
Grand Total	**6701**

Εικόνα 5: Σύνολο Πρατηρίων ανά Πλήθος Αξιολογήσεων.

Έναν από τους κυριότερους υπολογισμούς αποτέλεσε η μέση βαθμολογία των αξιολογήσεων των συσκευών, καθώς αναγνωρίζεται έτσι πως οι συσκευές που έχουν πραγματοποιήσει τις περισσότερες αξιολογήσεις εμφανίζουν συνήθως και ακραίες τιμές μέσης βαθμολογίας (δηλαδή κοντά στο 1 ή στο 5). Ενδεικτικά, στην Εικόνα 6 παρατηρούνται οι 15 συσκευές με τις περισσότερες κριτικές, με ελάχιστες από αυτές να βρίσκονται κοντά στην μέση τιμή.

	A	B	C
1	DeviceID	Average Rating	Ratings Count
2	android-f2c576e48846c7e4	1.008006405	1249
3	android-89244bfd08f62e31	1.005128205	325
4	android-ea05147b9cd8ea4e	2.498971193	324
5	android-4f7b57382e7e48fe	3.146666667	300
6	android-3497b701b11a081c	4.416374269	285
7	android-18ce69176a1de339	2.298701299	154
8	android-e42d8e876f5b1af2	1.816901408	142
9	android-cf45dfbc14505eb5	2.524934383	127
10	android-39704825e341aef4	1.425474255	123
11	android-de04f21853c35271	1.511299435	118
12	android-9be68dd801cee1fe	1.868945869	117
13	android-6c3ee9ebf46a62da	4.965217391	115
14	android-ccd16fab1d5e0485	2.636636637	111
15	android-1bce1a7b76f3aabe	1.651515152	110

Εικόνα 6: Συσκευές με τις Περισσότερες Αξιολογήσεις.

Επιπρόσθετα, διεξήχθησαν και άλλες αναλύσεις για την μελέτη των συσκευών, όπως την εφαρμογή πολλαπλών ταξινομήσεων. Μία από αυτές αποτελεί η ταξινόμηση κατά φθίνουσα σειρά μέσης βαθμολογίας και δευτερεύοντος κατά φθίνουσα σειρά πλήθους αξιολογήσεων. Εξίσου χαρακτηριστική είναι και η ταξινόμηση των συσκευών κατά αύξουσα σειρά μέσης βαθμολογίας και δευτερεύοντος κατά φθίνουσα σειρά πλήθους αξιολογήσεων. Έτσι, καθίσταται δυνατή η ευκολότερη παρακολούθηση των συσκευών με υπερβολικό αριθμό ακραίων κριτικών. Επίσης, με παρόμοιο τρόπο υπολογίζοντας τον μέσο όρο των πρατηρίων, πραγματοποιήθηκε ταξινόμηση κατά φθίνουσα σειρά του πλήθους αξιολογήσεων που έχουν λάβει και ταυτόχρονα κατά φθίνουσα σειρά του μέσου όρου. Άλλες τέτοιες αναλύσεις απαρτίζουν οι συνολικές αξιολογήσεις που γίνανε ανά εβδομάδα ή ανά μήνα, και η ταξινόμηση συνολικών αξιολογήσεων που γίνανε ανά ημέρα κατά φθίνουσα σειρά πλήθους αξιολογήσεων. Η αρχική αυτή ανάλυση μέσω Excel, προσέφερε μία ολοκληρωμένη πρώτη εικόνα για τα δεδομένα αποκαλύπτοντας κάποιες βασικές τάσεις του συνόλου των δεδομένων, γεγονός που κατέστησε δυνατή την στοχευμένη εκτέλεση SQL σεναρίων.

4.2 Ανάπτυξη SQL Σεναρίων

Στο πλαίσιο της ανάλυσης αναπτύχθηκαν και εφαρμόστηκαν σύνθετα SQL(Structured Query Language) σενάρια (scripts) με στόχο τον εντοπισμό μοτίβων που συμβάλλουν ενισχύοντας στην απάτη του συστήματος αξιολογήσεων. Με την δομημένη ανάλυση και τον συνδυασμό πολλαπλών κριτηρίων όπως τις χρονικές παραμέτρους και τις στατιστικές ανωμαλίες, στόχο των ερωτημάτων έθεσε η εύρεση συντονισμένων προσπαθειών που επηρεάζουν τη πραγματική εικόνα του συνόλου των αξιολογήσεων ώστε να σχεδιαστούν βασικοί έλεγχοι οι οποίοι οδηγούν σε συστήματα ανίχνευσης μοτίβων απάτης. Δεδομένου ότι το σύνολο δεδομένων δεν διαθέτει ετικέτες για τις ψευδείς κριτικές, υιοθετήθηκε μία προσέγγιση δημιουργίας πλαστής βασικής αλήθειας (pseudo ground truth) μέσω σύνθετων και αυστηρών κανόνων. Ουσιαστικά, κάθε ερώτημα που εκτελέστηκε διαθέτει έναν ρόλο αυτόνομου ταξινομητή, όπου αποδίδει σε κάθε κριτική μία δυαδική ετικέτα, με το μηδέν να συνιστά ότι η αξιολόγηση είναι πραγματική και την μονάδα να φανερώνει ότι η αξιολόγηση είναι ψευδής. Έτσι, υλοποιήθηκε μία σύνθετη στρατηγική πολλαπλών κριτηρίων, στην οποία κάθε ερώτημα σήμανε την εισαγωγή ενός νέου κριτηρίου αξιολόγησης, δημιουργώντας με αυτόν τον τρόπο ένα σύνθετο προφίλ για κάθε αξιολόγηση ξεχωριστά. Η προσέγγιση λειτούργησε με την ανάπτυξη 15 διαδοχικών ταξινομητών (status) όπου ο καθένας έδρασε ως ένα νέο κριτήριο αξιολόγησης. Για να γίνει δυνατή η διαφοροποίηση των πραγματικών και ψευδών αξιολογήσεων, ανατέθηκε σε κάθε ταξινομητή ως προεπιλεγμένη τιμή το 0, αντικατοπτρίζοντας όλες τις κριτικές σαν πραγματικές, και στις περιπτώσεις όπου μία κριτική θεωρείται ψευδής η τιμή αυτή αλλάζει σε 1. Επιπλέον, υλοποιήθηκε μία γενική Boolean στήλη ‘status’ στην οποία εάν κάποια εγγραφή έχει επισημανθεί ως ύποπτη από οποιοδήποτε ταξινομητή (status1,…status15) τότε η status λαμβάνει για αυτή την εγγραφή την τιμή ‘fraudulent’. Αντίθετα λαμβάνει την τιμή ‘normal’.

Η έρευνα ξεκίνησε με βασικά σενάρια που αφορούν τον εντοπισμό χρηστών οι οποίοι διαθέτουν έναν σχετικά υψηλό αριθμό αξιολογήσεων όπου μεγάλο κομμάτι τους αποτελείται από αξιολογήσεις ακραίων τιμών (δηλαδή αξιολογήσεις οι οποίες είτε και στα 3 κριτήρια rating έχουν βαθμολογία ίση με μονάδα είτε ίση με πέντε). Με αυτόν τον τρόπο μπορούν να βρεθούν χρήστες που όχι μόνο είναι ιδιαίτερα δραστήριοι στην υποβολή αξιολογήσεων, αλλά τείνουν επίσης να λειτουργούν χρησιμοποιώντας συστηματικά τις ακραίες θέσεις της κλίμακας βαθμολόγησης, δίχως να χρησιμοποιούν μέσες τιμές σε ο-

ποιοδήποτε κριτήριο των κριτικών. Ο συνδυασμός αυτών των δύο παραγόντων υποδηλώνει συντονισμένη προσπάθεια του χρήστη να επηρεάσει την φήμη των πρατηρίων (Εικόνα 7).

```
1  DROP TEMPORARY TABLE temp_ratings_st1;
2  ALTER TABLE ratings ADD COLUMN status1 BOOLEAN DEFAULT FALSE;
3  CREATE TEMPORARY TABLE temp_ratings_st1 AS
4  SELECT deviceID, COUNT(*) AS ratings_count,
5  SUM(CASE WHEN rating1 = 1 AND rating2 = 1 AND rating3 = 1 THEN 1 ELSE 0 END) AS count_low,
6  SUM(CASE WHEN rating1 = 5 AND rating2 = 5 AND rating3 = 5 THEN 1 ELSE 0 END) AS count_high
7  FROM ratings
8      GROUP BY deviceID
9      HAVING ratings_count >= 20 AND ((count_low + count_high) / ratings_count >= 0.8);
10
11 UPDATE ratings
12 SET status1 = TRUE
13 WHERE deviceID IN (SELECT deviceID FROM temp_ratings_st1);
14 SELECT deviceID, ratings_count, count_low, count_high
15 FROM temp_ratings_st1;
```

Εικόνα 7: Status1

Αναλυτικά, στο συγκεκριμένο σενάριο εντοπίστηκαν χρήστες οι οποίοι έχουν πραγματοποιήσει 20 ή περισσότερες αξιολογήσεις και ένα ποσοστό μεγαλύτερο ή ίσο του 80% αυτών αποτελείται από κριτικές οι οποίες είναι απόλυτες με τα 3 πεδία των αξιολογήσεων να είναι όλα ίσα είτε μονάδα ή 5. Τροποποιήθηκαν συνολικά 11527 εγγραφές καθώς τηρούν τα κριτήρια του ερωτήματος, εκ των οποίων η προεπιλεγμένη τιμή 0 του status1 μετατράπηκε στη μονάδα. Το αποτέλεσμα φαίνεται στην Εικόνα 8.

deviceID	ratings_count	count_low	count_high
android-d5e6daa0087d6679	1249	1230	1
android-0e30210bc753a860	325	320	0
android-27c0ad95e492e48b	324	194	118
android-1cea1f3b195c6685	300	130	150
android-a620cef9b1709e8d	154	104	50
android-c4235af6a066e4d8	142	113	29
android-890f50c880355b92	127	77	47
android-a103d58669b5d717	123	109	13
android-15d5d180761a5095	118	92	9
android-a3c61d62e156386d	117	82	16
android-1c473537fd0097ee	115	1	114
android-1514c88aaa65c092	110	89	15

Εικόνα 8 Ενδεικτικό παράδειγμα των χρηστών που εντοπίστηκαν.

Ο δεύτερος ταξινομητής (status2) , όπως φαίνεται στην Εικόνα 9, αναπτύχθηκε με τέτοιο σκοπό ώστε να εντοπίσει επαναλαμβανόμενες κριτικές, οι οποίες δεν είναι απαραίτητα ακραίες όπως στην περίπτωση του προηγούμενου ταξινομητή και που έχουν ληφθεί επίσης από κάποιον χρήστη αρκετές φορές σε σχέση με τις συνολικές του αξιολογήσεις. Δεν παρατηρήθηκε μεγάλη διαφορά σε σχέση με τον προηγούμενο ταξινομητή

ωστόσο εντοπίστηκαν 9419 εγγραφές από τις οποίες 744 είναι διαφορετικές σε σχέση με το status1.

```sql
ALTER TABLE ratings ADD COLUMN status2 BOOLEAN DEFAULT FALSE;
UPDATE ratings r
JOIN (
    SELECT r.deviceID, r.rating1, r.rating2, r.rating3
    FROM (
        SELECT deviceID, rating1, rating2, rating3, COUNT(*) AS same_rating_count
        FROM ratings
        GROUP BY deviceID, rating1, rating2, rating3
    ) r
    JOIN (
        SELECT deviceID, COUNT(*) AS total_reviews
        FROM ratings
        GROUP BY deviceID
    ) tr
    ON r.deviceID = tr.deviceID
    WHERE r.same_rating_count > 15  AND (r.same_rating_count * 100.0 / tr.total_reviews) >= 60
) filtered
ON r.deviceID = filtered.deviceID
AND r.rating1 = filtered.rating1
AND r.rating2 = filtered.rating2
AND r.rating3 = filtered.rating3
SET r.status2 = TRUE;
```

Εικόνα 9: Status2

Λεπτομερώς, υπολογίστηκαν οι συσκευές που έχουν επαναλάβει την ίδια βαθμολογία περισσότερες από 15 φορές και αποτελεί τουλάχιστον το 60% των συνολικών τους αξιολογήσεων.

Για τον τρίτο ταξινομητή (status3), θεωρήθηκαν ύποπτες οι αξιολογήσεις που προέρχονται από συσκευές με περισσότερες από 10 εγγραφές και όλες έχουν ακραίες τιμές. Εντοπίστηκαν συνολικά 8006 εγγραφές, εκ των οποίων οι 2484 δεν έχουν χαρακτηριστεί ως ύποπτες προηγουμένως (Εικόνα 10).

```sql
ALTER TABLE ratings ADD COLUMN status3 BOOLEAN DEFAULT FALSE;
UPDATE ratings r
JOIN (
     SELECT deviceID
     FROM ratings
     GROUP BY deviceID
     HAVING COUNT(*) > 10
       AND COUNT(*) = SUM((rating1 = rating2) AND (rating2 = rating3) AND rating1 IN (1, 5))
)AS extreme ON r.deviceID = extreme.deviceID
SET r.status3 = TRUE;

SELECT r.deviceID, r.stationGUID, r.rating1, r.rating2, r.rating3, r.status3
FROM ratings r
JOIN (
     SELECT deviceID
     FROM ratings
     GROUP BY deviceID
     HAVING COUNT(*) > 10
       AND COUNT(*) = SUM((rating1 = rating2) AND (rating2 = rating3) AND rating1 IN (1, 5))
) AS extreme ON r.deviceID = extreme.deviceID
GROUP BY r.deviceID, r.stationGUID, r.rating1, r.rating2, r.rating3, r.statuss3;
```

Εικόνα 10: Status3

Για το status4, πραγματοποιήθηκε παρόμοια ανίχνευση με διαφορά την προσθήκη χρονικού ορίου στο οποίο πραγματοποιήθηκαν ακραίες αξιολογήσεις. Μάλιστα, σε χρονικό όριο μίας ώρας βρέθηκαν συσκευές οι οποίες εκτέλεσαν πέντε ή περισσότερες ακραίες αξιολογήσεις. Εκτελέστηκε παρόμοιο σενάριο με το status3 με την διαφορά του μοναδικού κριτηρίου, θέτοντας ύποπτες 6228 αξιολογήσεις.

```
HAVING COUNT(*) >= 5 AND TIMESTAMPDIFF(MINUTE, MIN(ratingwhen), MAX(ratingwhen)) <= 60
```

Με παρόμοιο τρόπο στο status5, εντοπίστηκαν συσκευές με πέντε ή περισσότερες αξιολογήσεις (δεν είναι απαραίτητα ακραίες) σε διάστημα μισής ώρας. Οι κριτικές οι οποίες τηρούν αυτά τα κριτήρια είναι 9342, με 3452 από αυτές να διαφέρουν σε σχέση με τις σημειωμένες ως ύποπτες κριτικές από το status4. Η αύξηση αυτή παρατηρείται καθώς υπάρχουν κριτικές οι οποίες τηρούν τους περιορισμούς που τέθηκαν, δίχως να ανήκουν στο σύνολο των κριτικών που αναφέρονται ως ακραίες.

```
HAVING COUNT(*) >= 5 AND TIMESTAMPDIFF(MINUTE, MIN(ratingwhen), MAX(ratingwhen)) <= 30
```

Το status6 λειτουργεί με παρόμοιο τρόπο και αναγνωρίζει τις κριτικές που προέρχονται από συσκευές με δέκα ή παραπάνω εγγραφές σε διάρκεια μίας ημέρας. Τέτοιες περιπτώσεις μπορούν να χαρακτηριστούν ύποπτες καθώς παρουσιάζουν έντονη δραστηριότητα η οποία είναι ασυνήθιστη. Τα κριτήρια αυτά τηρούν συνολικά 4600 εγγραφές, με 164 από αυτές να μην έχουν σημειωθεί από τις προηγούμενες περιπτώσεις.

```
HAVING COUNT(*) >= 10 AND DATEDIFF(MAX(ratingwhen), MIN(ratingwhen)) = 0
```

Στον έβδομο ταξινομητή πραγματοποιήθηκε η χρήση CTEs (Common Table Expressions) που αποτελούν απαραίτητο SQL εργαλείο για την απλοποίηση σύνθετων σεναρίων. Έχει τη δυνατότητα να ορίζει προσωρινά σύνολα αποτελεσμάτων τα οποία μπορούν να αναφερθούν πολλαπλές φορές διαχωρίζοντας έτσι κάποια περίπλοκη λογική σε διαχειρίσιμα μέρη. Συνεπώς κάποια από τα κύρια χαρακτηριστικά της συνιστά η βελτίωση της αναγνωσιμότητας καθώς και η ευχρηστία που προσφέρει, ο διαχωρισμός σύνθετων ερωτημάτων σε μικρότερα στοιχεία τα οποία μπορούν να χρησιμοποιηθούν ξανά και η ενεργοποίηση αναδρομικών λειτουργιών για ιεραρχικά δεδομένα.

Αναλυτικά, στο σενάριο για το status7, δημιουργήθηκαν 3 πίνακες κοινών εκφράσεων (CTEs). Ο πίνακας CumulativeRatings λαμβάνει τα δεδομένα του συνόλου και τα μετατρέπει σε υπολογισμένες στήλες. Έτσι, δεν χρειάζεται να γραφούν οι ίδιες συναρτήσεις σε διαφορετικά υποερωτήματα και αποφεύγεται η επανάληψη των υπολογισμών τους. Συγκεκριμένα, υπολογίζει τον μέσο όρο κάποιας κριτικής καθώς επίσης και τον

συνολικό μέσο όρο των αξιολογήσεων του πρατηρίου πριν τη συγκεκριμένη κριτική. Επιπλέον, μετατρέπει την ώρα κάθε μεμονωμένης αξιολόγησης σε ένα διάστημα δύο ωρών, διαιρώντας τη με το 2 και αφού στρογγυλοποιηθεί, πολλαπλασιάζεται με το 2 ώστε να μετατραπεί το αναγνωριστικό της ομάδας πίσω στην πραγματική ώρα έναρξης του διαστήματος. Αυτή η μέθοδος είναι γνωστή ως time binning και αποτελεί μία διαδικασία κατάτμησης μίας συνεχούς τιμής σε διακριτά διαστήματα.

```
1  ALTER TABLE ratings ADD COLUMN status7 BOOLEAN DEFAULT FALSE;
2  UPDATE ratings r
3  JOIN (
4    WITH CumulativeRatings AS ( -- Υπολογισμός των συσσωρευμένων κριτικών
5      SELECT r.stationGUID, r.ratingWhen, (r.rating1 + r.rating2 + r.rating3) / 3.0 AS avg_rating,
6        AVG((r.rating1 + r.rating2 + r.rating3) / 3.0)
7        OVER (PARTITION BY r.stationGUID ORDER BY r.ratingWhen ROWS BETWEEN UNBOUNDED PRECEDING AND 1 PRECEDING)
8        AS previous_avg_rating, -- Συνολικός μέσος όρος των κριτικών εκτός της τρέχουσας κριτικής
9        CONCAT(DATE(r.ratingWhen), ' ', LPAD(FLOOR(HOUR(r.ratingWhen) / 2) * 2, 2, '0'), ':00:00')
10       AS two_hours_group -- Δημιουργία διαστήματος 2 ωρών για κάθε κριτική
11     FROM ratings r
12   ),
```

Εικόνα 11: CumulativeRatings

Συγκεκριμένα, υπολογίζει τον μέσο όρο κάποιας κριτικής καθώς επίσης και τον συνολικό μέσο όρο των αξιολογήσεων του πρατηρίου πριν τη συγκεκριμένη κριτική. Επιπλέον, μετατρέπει την ώρα κάθε μεμονωμένης αξιολόγησης σε ένα διάστημα δύο ωρών, διαιρώντας τη με το 2 και αφού στρογγυλοποιηθεί (FLOOR( )) πολλαπλασιάζεται με το 2.

```
15   avg_twoHoursRatings AS ( -- Βαθμολογίες(μ.ο) των πρατηρίων ανά διαστήματα 2 ωρών
16     SELECT
17       stationGUID, two_hours_group,
18       COUNT(*) AS current_total_reviews, -- Πλήθος κριτικών που γίνανε στο τρέχων δίωρο
19       AVG(avg_rating) AS current_avg_rating -- Μέσος όρος των κριτικών σε αυτό το διάστημα
20     FROM CumulativeRatings
21     GROUP BY stationGUID, two_hours_group
22   ),
```

Εικόνα 12: avg_twoHoursRatings

Ο πίνακας avg_twoHoursRatings στην Εικόνα 12, υπολογίζει τις βαθμολογίες (μέσο όρο) των πρατηρίων ανά διαστήματα δύο ωρών και μετράει επίσης το πλήθος κριτικών που έλαβαν σε κάθε διάστημα.

Ο τρίτος πίνακας AverageBefore, βρίσκει τον τελευταίο συνολικό μέσο όρο βαθμολογιών του πρατηρίου ακριβώς πριν από την έναρξη κάθε δίωρου διαστήματος (Εικόνα 13).

```
-- Επιλογή του τελευταίου συνολικού μέσου όρου ακριβώς πριν από την έναρξη κάθε διαστήματος δύο ωρών
AverageBefore AS (
    SELECT
    stationGUID, two_hours_group, MAX(previous_avg_rating) AS previous_avg_rating
    FROM CumulativeRatings
    GROUP BY stationGUID, two_hours_group
)
```

Εικόνα 13: AverageBefore

Μετά την εκτέλεση των CTEs, πραγματοποιείται η σύνδεση των δύο τελευταίων πινάκων και έτσι διαθέτουμε για κάθε πρατήριο το πλήθος και τον μέσο όρο των αξιολογήσεων σε κάποιο συγκεκριμένο δίωρο αλλά και τον συνολικό μέσο όρο του σταθμού πριν από αυτό. Έτσι, είναι εφικτή αργότερα η επιλογή των σταθμών που έχουν τουλάχιστον πέντε κριτικές σε κάποιο δίωρο διάστημα, όπου ο μέσος όρος αυτών, διαφέρει περισσότερο από μία μονάδα σε σχέση με τον συνολικό μέσο όρο που έχει συσσωρεύσει το πρατήριο πριν από αυτές. Συνεπώς, οι κριτικές που πραγματοποιήθηκαν σε αυτό το πρατήριο το συγκεκριμένο διάστημα σημειώνονται ως ύποπτες όπως φαίνεται στην Εικόνα 14.

```
SELECT
    r.stationGUID, r.ratingWhen
    FROM avg_twoHoursRatings avgtHR JOIN AverageBefore avgB
        ON avgtHR.stationGUID = avgB.stationGUID
        AND avgtHR.two_hours_group = avgB.two_hours_group
    JOIN ratings r
        ON avgtHR.stationGUID = r.stationGUID
        -- Έλεγχος εάν η κριτική ανήκει στο συγκεκριμένο δίωρο
        AND CONCAT(DATE(r.ratingWhen), ' ', LPAD(FLOOR(HOUR(r.ratingWhen) / 2) * 2, 2, '0'), ':00:00') = avgtHR.two_hours_group
            -- Επιλογή των σταθμών που έχουν τουλάχιστον 5 εγγραφές σε συγκεκριμένο διάστημα
            -- και ο μέσος όρος αυτών διαφέρει περισσότερο από μία μονάδα από τον προηγούμενο συνολικό μέσο όρο
            WHERE avgtHR.current_total_reviews >= 5  AND ABS(avgB.previous_avg_rating - avgtHR.current_avg_rating) > 1
) AS filteredRatings
ON r.stationGUID = filteredRatings.stationGUID
AND r.ratingWhen = filteredRatings.ratingWhen
SET r.status7 = TRUE;
```

Εικόνα 14: Status7

Εντοπίστηκαν συνολικά 73 αξιολογήσεις οι οποίες πραγματοποιήθηκαν σε 13 πρατήρια. Από αυτές, οι 50 βρέθηκαν ύποπτες πρώτη φορά. Στην Εικόνα 15 παρουσιάζονται τα πρατήρια τα οποία δέχτηκαν τις ύποπτες κριτικές.

Στο status8 πραγματοποιείται ομαδοποίηση των συσκευών και των IP διευθύνσεων που έχουν υποβάλει όλες τους τις αξιολογήσεις σε μία μέρα, και το πλήθος των αξιολογήσεών τους είναι μεγαλύτερο από 5. Επίσης, πρέπει ο μέσος όρος αυτών των κριτικών

να είναι μεγαλύτερος του 4 ή μικρότερος του 2 χαρακτηριστούν ύποπτες. Συνολικά, εντοπίστηκαν 958 συνδυασμοί συσκευών και IP διευθύνσεων που τηρούν τα κριτήρια με 11161 αξιολογήσεις.

stationGUID	two_hours_group	previous_avg_rating	current_avg_rating	rating_change	current_total_reviews
A79A238D-6229-79E3-D3B8...	2018-07-22 22:00:00	4.41667500	1.00000000	3.41667500	5
5D865FEC-3774-84EE-D8B2-...	2022-07-11 08:00:00	3.51905143	1.00000000	2.51905143	5
7755DD95-608D-B8F3-2C3B...	2019-01-26 20:00:00	3.43678276	1.00000000	2.43678276	5
D27C4761-4A7A-364A-BB41...	2022-07-11 08:00:00	3.42767358	1.00000000	2.42767358	5
28653845-4F19-A9D2-A36F-...	2021-06-23 22:00:00	3.85714286	1.48484545	2.37229741	11
BAE206C8-E2CE-9830-1D01...	2022-07-11 08:00:00	2.94444762	1.00000000	1.94444762	5
37616438-ABA0-ED78-7C7A...	2024-02-20 16:00:00	2.93869885	1.00000000	1.93869885	6
BD72964F-0C7C-A70F-F1B5...	2020-04-08 12:00:00	2.82455789	1.00000000	1.82455789	5
37616438-ABA0-ED78-7C7A...	2024-04-11 12:00:00	2.75694583	1.00000000	1.75694583	5
FB12F881-FBB7-0C85-05F7-...	2019-06-23 14:00:00	3.50980588	5.00000000	1.49019412	5
BFF43D07-12B5-1271-B4CB-...	2019-07-29 00:00:00	3.72881525	5.00000000	1.27118475	6
E694497C-84CB-C788-7C06...	2022-04-15 18:00:00	2.96491579	4.20000000	1.23508421	5
B61009BE-66CF-D540-42EA-...	2019-01-26 20:00:00	3.80000000	5.00000000	1.20000000	5

Εικόνα 15: Πρατήρια που δέχτηκαν τις ύποπτες αξιολογήσεις.

```
ALTER TABLE ratings DROP COLUMN status8;
ALTER TABLE ratings ADD COLUMN status8 BOOLEAN DEFAULT FALSE;
UPDATE ratings r
JOIN (
    SELECT deviceID, clientIP
    FROM ratings
    GROUP BY deviceID, clientIP
    HAVING COUNT(DISTINCT DATE(ratingwhen)) = 1  AND COUNT(*) > 5
    AND (AVG((rating1 + rating2 + rating3) / 3) > 4
    OR AVG((rating1 + rating2 + rating3) / 3) < 2)
) filtered
ON r.deviceID = filtered.deviceID AND r.clientIP = filtered.clientIP
SET r.status8 = TRUE;

SELECT deviceID, clientIP,
COUNT(*) AS total_count,
MIN(ratingwhen) AS first_rating,
MAX(ratingwhen) AS last_rating,
AVG((rating1 + rating2 + rating3) / 3) AS average_ratings
FROM ratings
WHERE status8 = TRUE
GROUP BY deviceID, clientIP
ORDER BY total_count DESC, average_ratings DESC;
```

Εικόνα 16: Status8

deviceID	clientIP	total_count	first_rating	last_rating	average_ratings
android-1514c88aaa65c092	85.74.13.67	100	2022/07/09 17:13:55	2022/07/09 17:30:38	1.57666700
android-d5e6daa0087d6679	79.167.43.194	85	2021/02/18 14:45:34	2021/02/18 15:13:07	1.00000000
android-8c34f750b3b2aa86	5.144.232.192	79	2018/04/29 14:12:24	2018/04/29 14:32:37	1.41350127
android-c4fd32fdd66bbe57	31.152.100.85	77	2018/09/28 19:56:38	2018/09/28 20:11:47	1.45021558
android-d5e6daa0087d6679	46.103.199.182	58	2021/03/11 22:16:32	2021/03/11 22:34:31	1.00000000
android-0d0c3c1d7da96cf4	45.139.214.66	52	2023/01/21 12:39:10	2023/01/21 12:52:01	1.30769231
android-e499ccacb128878d	87.203.207.195	45	2021/02/11 21:39:09	2021/02/11 21:52:52	1.26666667
android-519559de616e36bc	85.73.192.170	44	2019/02/23 00:36:41	2019/02/23 00:56:45	1.54545455
android-a5a6aa71c9cbf785	141.237.19.69	43	2023/05/06 23:22:01	2023/05/06 23:36:15	1.10077209
android-0e30210bc753a860	62.74.23.117	41	2023/01/11 07:17:46	2023/01/11 07:31:40	1.00000000
android-46b60f57d9fbeb92	185.51.133.57	40	2023/10/11 01:08:02	2023/10/11 01:16:14	1.10000000
android-d5e6daa0087d6679	46.12.23.219	40	2020/04/27 09:13:25	2020/04/27 14:33:03	1.00833250
android-27c0ad95e492e48b	89.44.158.181	40	2020/05/29 10:42:35	2020/05/29 11:56:18	1.00833250

Εικόνα 17: Συνδυασμοί συσκευών και IP διευθύνσεων που τηρούν τα κριτήρια.

Ο status9 έχει ως σκοπό τον εντοπισμό συσκευών με υπερβολικά υψηλό αριθμό αξιολογήσεων, οι οποίες έχουν επίσης και χαμηλό ή υψηλό μέσο όρο. Έτσι, κριτικές που προέρχονται από συσκευές με περισσότερες από 100 αξιολογήσεις και έχουν μέσο όρο χαμηλότερο ή ίσο του 2.1 ή αντίθετα υψηλότερο ή ίσο του 3.9 συνιστούν χαρακτηριστικό παράδειγμα ύποπτης συμπεριφοράς,.

```
DROP TEMPORARY TABLE IF EXISTS suspicious_devices;
ALTER TABLE ratings DROP COLUMN status9;
ALTER TABLE ratings ADD COLUMN status9 BOOLEAN DEFAULT FALSE;
CREATE TEMPORARY TABLE suspicious_devices AS
  SELECT deviceID
  FROM ratings
  GROUP BY deviceID
    HAVING COUNT(*) > 100
    AND (AVG((rating1 + rating2 + rating3) / 3) <= 2.1
    OR AVG((rating1 + rating2 + rating3) / 3) >= 3.9);

UPDATE ratings r
JOIN suspicious_devices sd ON r.deviceID = sd.deviceID
SET r.status9 = TRUE;

SELECT deviceID, COUNT(*) AS ratings_count,
AVG((rating1 + rating2 + rating3) / 3) as average_rating
FROM ratings
WHERE status9 = TRUE
GROUP BY deviceID
ORDER BY ratings_count DESC, average_rating DESC;
```

Εικόνα 18: Status9

Ο αριθμός των κριτικών που χαρακτηρίστηκαν ύποπτες είναι 3011 και προέρχονται μόνο από 13 συσκευές. Το αποτέλεσμα φαίνεται στην Εικόνα 19.

deviceID	ratings_count	average_rating
android-d5e6daa0087d6679	1249	1.00800592
android-0e30210bc753a860	325	1.00512769
android-51bdfbe8badfb54d	285	4.41637579
android-c4235af6a066e4d8	142	1.81690141
android-a103d58669b5d717	123	1.42547398
android-15d5d180761a5095	118	1.51129915
android-a3c61d62e156386d	117	1.86894615
android-1c473537fd0097ee	115	4.96521739
android-1514c88aaa65c092	110	1.65151545
android-8051f2e10b10d42d	109	1.62385321
android-b760a67a3325f06b	108	4.01543611
android-39ae6be62180d909	107	1.63551402
android-4aa8d5e2ac526e17	103	1.27184466

Εικόνα 19: Οι ύποπτες συσκευές που εντοπίστηκαν από το Status9.

Για την εύρεση ύποπτων αξιολογήσεων στο status10 εξετάστηκαν περιπτώσεις όπου παρουσιάζεται επανάληψη σε μεγάλο βαθμό στις αξιολογήσεις των χρηστών. Μία συσκευή που έχει πλήθος επανειλημμένης ίδιας αξιολόγησης μεταξύ του 10 έως 20 και ο συνολικός αριθμός των αξιολογήσεών της είναι μικρότερος του πλήθους αυτού αυξημένο κατά μισό, τότε θεωρείται ότι συμπεριφέρεται ύποπτα. Οι κριτικές αυτών των συσκευών καταγράφονται ως ύποπτες (Εικόνα 20).

```sql
ALTER TABLE ratings ADD COLUMN status10 BOOLEAN DEFAULT FALSE;
CREATE TEMPORARY TABLE suspicious_users AS
    SELECT consRatings.deviceID, consRatings.consistent_ratings_count, total.ratings_count
    FROM (
        SELECT deviceID, COUNT(*) AS ratings_count
        FROM ratings
        GROUP BY deviceID
    ) total -- Συνολικός αριθμός αξιολογήσεων για κάθε συσκευή
    JOIN (
        SELECT deviceID, COUNT(*) AS consistent_ratings_count
        FROM ratings
        GROUP BY deviceID, rating1, rating2, rating3
    ) consRatings -- Για κάθε συσκευή βρίσκει το πλήθος των εγγραφών με ακριβώς τις ίδιες τιμές rating
    ON total.deviceID = consRatings.deviceID
-- το πλήθος της επανειλημμένης ίδιας αξιολόγησης πρέπει να είναι απο 10 έως 20
-- και ο συνολικός αριθμός αξιολογήσεων της συσκευής πρέπει να είναι μικρότερος
-- από το πλήθος της επανειλημμένης ίδιας αξιολόγησης της συσκευής αυξημένο κατά μισό
WHERE consRatings.consistent_ratings_count >= 10 AND consRatings.consistent_ratings_count <= 20
AND total.ratings_count <= (1.5 * consRatings.consistent_ratings_count);
-- suspicious_users
UPDATE ratings r
JOIN suspicious_users s ON r.deviceID = s.deviceID
SET r.status10 = TRUE;
SELECT s.deviceID, s.consistent_ratings_count, s.ratings_count
FROM suspicious_users s
ORDER BY s.ratings_count DESC, s.consistent_ratings_count DESC;
```

Εικόνα 20: Status10

Εντοπίστηκαν συνολικά 329 συσκευές και καταγράφηκαν 5013 αξιολογήσεις, με ένα μέρος αυτών να παρουσιάζεται στην Εικόνα 21.

deviceID	consistent_ratings_count	ratings_count
android-b228b92720c1a8e1	20	25
android-de5908394930f0c1	17	25
android-4d279f64833bddaf	19	24
android-d69f589b484d0e48	18	24
android-8aa37d4073441222	20	23
android-952c0b1da6a6b58f	20	23
android-ac9d4adfca8a32bc	19	23
android-a1d7636812049c64	18	23

Εικόνα 21: Μέρος των ύποπτων συσκευών που εντοπίστηκαν στο status10.

Με παρόμοιο τρόπο λειτούργησε το status11 όπου πραγματοποιήθηκε η παρακάτω αλλαγή. Η διαφορά είναι ότι η συνεχής βαθμολογία της συσκευής πρέπει να επαναλαμβάνεται 20 ή παραπάνω φορές και αν διπλασιαστεί ο αριθμός αυτός χρειάζεται να είναι μεγαλύτερος από τις συνολικές αξιολογήσεις που έχει βάλει η συσκευή.

```sql
WHERE consRatings.consistent_ratings_count >= 20
AND total.ratings_count <= (2 * consRatings.consistent_ratings_count);
```

Βρέθηκαν 10646 ύποπτες κριτικές που προέρχονται από 206 συσκευές (Εικόνα 22).

deviceID	consistent_ratings_count	ratings_count
android-d5e6daa0087d6679	1230	1249
android-0e30210bc753a860	320	325
android-27c0ad95e492e48b	194	324
android-1cea1f3b195c6685	150	300
android-a620cef9b1709e8d	104	154
android-c4235af6a066e4d8	113	142
android-890f50c880355b92	77	127
android-a103d58669b5d717	109	123
android-15d5d180761a5095	92	118
android-a3c61d62e156386d	82	117
android-1c473537fd0097ee	114	115
android-1514c88aaa65c092	89	110

Εικόνα 22: Μέρος των ύποπτων συσκευών από το Status11.

Για την ανίχνευση ύποπτων αξιολογήσεων στο Status12, πραγματοποιήθηκε χρήση μεταβλητών με σκοπό να βρεθούν συσκευές που πραγματοποίησαν κάποιον υψηλό αριθμό αξιολογήσεων σε μία συνεχή ακολουθία. Η λογική βασίζεται στην εύρεση συνεχών ακολουθιών αξιολογήσεων της ίδιας συσκευής, όπου κάθε κριτική απέχει το πολύ 10 λεπτά από την προηγούμενη. Εάν η συνεχής ακολουθία αξιολογήσεων ξεπερνά τις 5 αξιολογήσεις, τότε οι εγγραφές της συσκευής που ανήκουν σε αυτή τη χρονική περίοδο θεωρούνται ύποπτες.

Οι μεταβλητές οι οποίες χρησιμοποιήθηκαν είναι οι:

- @group_id: Λειτουργεί σαν μετρητής, διαθέτοντας τον τρέχων αριθμό ομάδας.
- @prev_device: Αποθηκεύει τη συσκευή της αμέσως προηγούμενης εγγραφής, για να πραγματοποιηθεί έλεγχος της ακολουθίας.
- @prev_time: Αποθηκεύει την αμέσως προηγούμενη χρονική στιγμή, για να υπολογιστεί το χρονικό κενό μέχρι την τρέχουσα εγγραφή.

Οι μεταβλητές χρησιμοποιήθηκαν για την απομνημόνευση των στοιχείων μεταξύ των διαδοχικών εγγραφών.

```
1 ●  ALTER TABLE ratings ADD COLUMN status12 BOOLEAN DEFAULT FALSE;
2 ●  SET @group_id = 0; -- Αριθμητής ομάδων για κάθε συνεχή ακολουθία
3 ●  SET @prev_device = NULL; -- Αποθήκευση του deviceID της προηγούμενης εγγραφής για σύγκριση
4 ●  SET @prev_time = NULL; -- Αποθήκευση του προηγούμενου timestamp
5 ●  DROP TEMPORARY TABLE IF EXISTS suspicious_ratings;
6 ●  CREATE TEMPORARY TABLE suspicious_ratings AS
```

Εικόνα 23:Αρχικοποίηση μεταβλητών.

Ύστερα από την αρχικοποίηση των μεταβλητών, δημιουργήθηκε ο προσωρινός πίνακας suspicious_ratings (Εικόνα 24) για την αποθήκευση των προσωρινών αποτελεσμάτων από τις ύποπτες ομάδες όπου θα χρησιμοποιηθούν για το update. Για κάθε ύποπτη

ακολουθία επιλέγει την συσκευή, την χρονική στιγμή της πρώτης αλλά και της τελευταίας αξιολόγησης και τον αριθμό των αξιολογήσεων σε αυτή.

```
6 ●   CREATE TEMPORARY TABLE suspicious_ratings AS
7       SELECT
8       deviceID,
9       MIN(ratingWhen) AS first_rating, -- Πρώτη χρονική στιγμή κάθε ομάδας(rating_group)
10      MAX(ratingWhen) AS last_rating,  -- Τελευταία χρονική στιγμή κάθε ομάδας(rating_group)
11      COUNT(*) AS rating_count -- Αριθμός αξιολογήσεων κάθε ομάδας(rating_group)
```

Εικόνα 24: suspicious_ratings.

Τα στοιχεία αυτά τα αντλεί από την εσωτερική SELECT στην Εικόνα 25 στην οποία για κάθε συσκευή και χρονική στιγμή πραγματοποιείται έλεγχος για τον διαχωρισμό των ομάδων. Στο κομμάτι αυτό πραγματοποιείται ο έλεγχος της ακολουθίας για την τρέχουσα συσκευή και αν παραμένει η ίδια τότε ελέγχεται εάν το χρονικό κενό με την αμέσως προηγούμενή της εγγραφή είναι μικρότερο των 10 λεπτών. Αν ισχύουν αυτά τα 2 κριτήρια τότε ο μετρητής μένει στον ίδιο αριθμό (η εγγραφή εντάσσεται στην ίδια ομάδα), αλλιώς η εγγραφή μπαίνει σε νέα ομάδα.

```
12      FROM (
13        SELECT
14        deviceID,
15        ratingWhen,
16          -- Έλεγχος για τον διαχωρισμό των ομάδων.
17          -- Εάν πρόκειται για το ίδιο deviceID και τηρεί το κριτήριο, τότε παραμένει στην ίδια ομάδα.
18          -- Αλλιώς αυξάνεται ο αριθμός των ομάδων.
19          @group_id := IF(
20          deviceID = @prev_device
21          AND TIMESTAMPDIFF(MINUTE, @prev_time, ratingWhen) <= 10,
22            @group_id,
23            @group_id + 1
24          ) AS rating_group,
```

Εικόνα 25: Κριτήρια Status12.

Για την συνέχεια, η τρέχουσα συσκευή αποθηκεύεται στην @prev_device και η χρονική της στιγμή στην @prev_time. Ύστερα, με την ομαδοποίηση των εγγραφών ανά συσκευή και ομάδα, επιλέγονται οι ομάδες που έχουν περισσότερες από 5 αξιολογήσεις.

```
24          ) AS rating_group,
25        @prev_device := deviceID, -- Αποθήκευση της τρέχουσας συσκευής για την επόμενη εγγραφή
26        @prev_time := ratingWhen -- Αποθήκευση της τρέχουσας χρονικής στιγμής για την επόμενη εγγραφή
27        FROM ratings
28        ORDER BY deviceID, ratingWhen
29      ) AS grouped_ratings
30    GROUP BY deviceID, rating_group
31    -- Επιλογή όσων συσκευών έχουν 6+ αξιολογήσεις σε ομάδα όπου η κάθε μία έχει διαφορά λιγότερη των 10 λεπτών από την προηγούμενη
32    HAVING rating_count > 5; -- suspicious_ratings
```

Εικόνα 26: Κριτήρια Status12

Τέλος, με την εκτέλεση του UPDATE, οι εγγραφές που βρίσκονται στις ύποπτες ομάδες καταγράφονται.

```
33  UPDATE ratings r
34  JOIN suspicious_ratings sr
35  ON r.deviceID = sr.deviceID
36  AND r.ratingWhen BETWEEN sr.first_rating AND sr.last_rating
37  SET r.status12 = TRUE;
38  SELECT * FROM suspicious_ratings;
```

Εικόνα 27: Status12 Update

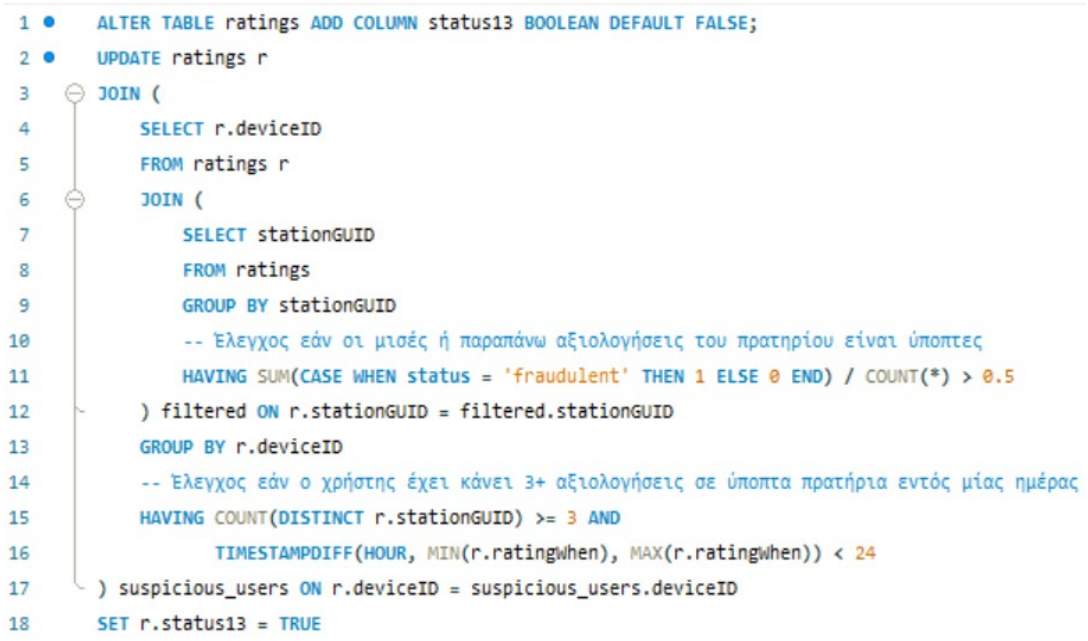
Ως αποτέλεσμα, βρέθηκαν συνολικά 1938 ομάδες οι οποίες αφορούν 21355 εγγραφές (Εικόνα 28).

deviceID	first_rating	last_rating	rating_count
android-d5e6daa0087d6679	2020/03/31 09:51:01	2020/03/31 10:15:10	144
android-d5e6daa0087d6679	2020/03/28 07:20:02	2020/03/28 08:02:00	130
android-1514c88aaa65c092	2022/07/09 17:13:55	2022/07/09 17:30:38	100
android-d5e6daa0087d6679	2020/03/28 17:18:15	2020/03/28 18:06:13	95
android-d5e6daa0087d6679	2021/02/18 14:45:34	2021/02/18 15:13:07	85
android-8c34f750b3b2aa86	2018/04/29 14:12:24	2018/04/29 14:32:37	79
android-c4fd32fdd66bbe57	2018/09/28 19:56:38	2018/09/28 20:11:47	77
android-d5e6daa0087d6679	2020/03/30 22:52:20	2020/03/30 23:14:11	66
android-d4ae9506a824668a	2022/04/04 04:08:28	2022/04/04 04:34:03	63
android-d5e6daa0087d6679	2021/03/11 22:16:32	2021/03/11 22:34:31	58
android-d5e6daa0087d6679	2020/03/27 20:58:11	2020/03/27 21:16:48	57
android-f60309ed7ed87690	2022/07/07 23:56:23	2022/07/08 00:07:42	56
android-1eea7b67c5124675	2019/02/22 23:53:39	2019/02/23 00:22:43	55

Εικόνα 28: Εγγραφές του Status12.

Στο status13, εντοπίστηκαν συσκευές που έχουν αξιολογήσει σε μία ημέρα 3 ή παραπάνω πρατήρια, τα οποία έχουν περισσότερο από 50% των αξιολογήσεων τους ως ύποπτες.

Εντοπίστηκαν συνολικά 6780 ύποπτες κριτικές.

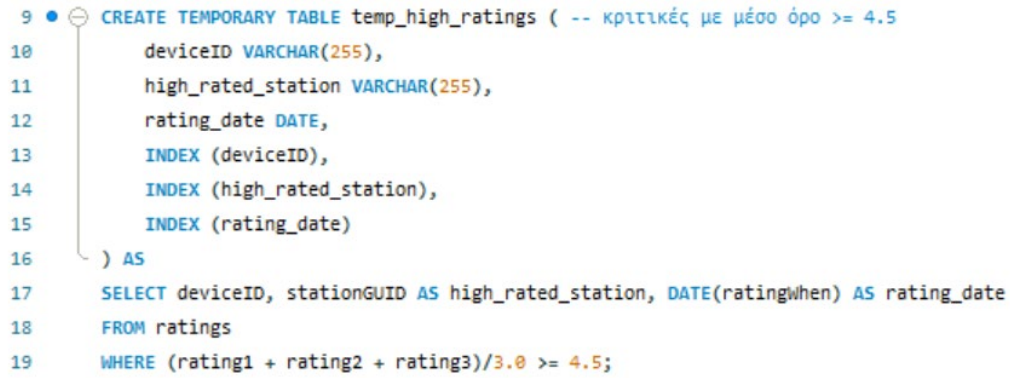


```sql
ALTER TABLE ratings ADD COLUMN status13 BOOLEAN DEFAULT FALSE;
UPDATE ratings r
JOIN (
    SELECT r.deviceID
    FROM ratings r
    JOIN (
        SELECT stationGUID
        FROM ratings
        GROUP BY stationGUID
        -- Έλεγχος εάν οι μισές ή παραπάνω αξιολογήσεις του πρατηρίου είναι ύποπτες
        HAVING SUM(CASE WHEN status = 'fraudulent' THEN 1 ELSE 0 END) / COUNT(*) > 0.5
    ) filtered ON r.stationGUID = filtered.stationGUID
    GROUP BY r.deviceID
    -- Έλεγχος εάν ο χρήστης έχει κάνει 3+ αξιολογήσεις σε ύποπτα πρατήρια εντός μίας ημέρας
    HAVING COUNT(DISTINCT r.stationGUID) >= 3 AND
           TIMESTAMPDIFF(HOUR, MIN(r.ratingWhen), MAX(r.ratingWhen)) < 24
) suspicious_users ON r.deviceID = suspicious_users.deviceID
SET r.status13 = TRUE
```

Εικόνα 29: Status13

Για το status14 αναπτύχθηκε ένα πολύπλοκο σενάριο πολλαπλών προσωρινών πινάκων για τον εντοπισμό ύπουλων αξιολογήσεων που έχουν ως σκοπό την δυσφήμηση συγκεκριμένων πρατηρίων και της προώθησης άλλων. Χρησιμοποιήθηκαν δείκτες για να επιτευχθεί η ταχύτερη εκτέλεση των εντολών και γρηγορότερες αναζητήσεις.

Αρχικά, δημιουργήθηκε ο προσωρινός πίνακας temp_high_ratings που παρουσιάζεται στην Εικόνα 30, στον οποίο αποθηκεύονται τα στοιχεία των υψηλών αξιολογήσεων (δηλαδή των εγγραφών με μέσο όρο μεγαλύτερο ή ίσο του 4.5)



```sql
CREATE TEMPORARY TABLE temp_high_ratings ( -- κριτικές με μέσο όρο >= 4.5
    deviceID VARCHAR(255),
    high_rated_station VARCHAR(255),
    rating_date DATE,
    INDEX (deviceID),
    INDEX (high_rated_station),
    INDEX (rating_date)
) AS
SELECT deviceID, stationGUID AS high_rated_station, DATE(ratingWhen) AS rating_date
FROM ratings
WHERE (rating1 + rating2 + rating3)/3.0 >= 4.5;
```

Εικόνα 30: temp_high_ratings

Αμέσως μετά, υλοποιήθηκε ο πίνακας temp_low_ratings (Εικόνα 31) για την εύρεση των κριτικών με μέσο όρο χαμηλότερο ή ίσο του 1.5 από χρήστες που έχουν δώσει περισσότερες από 2 τέτοιες αξιολογήσεις σε μία μέρα. Αποθηκεύει την συσκευή από την

οποία προέρχονται αυτές οι αξιολογήσεις, τα πρατήρια που δέχτηκαν τις χαμηλές βαθμολογίες και την μέρα.

```sql
-- Κριτικές με μ.ο. <= 1.5 από χρήστες που έδωσαν χαμηλή αξιολόγηση σε περισσότερα από 2 πρατήρια την ίδια μέρα
CREATE TEMPORARY TABLE temp_low_ratings (
    deviceID VARCHAR(255),
    low_rated_station VARCHAR(255),
    rating_date DATE,
    INDEX (deviceID),
    INDEX (low_rated_station),
    INDEX (rating_date)
) AS
  SELECT r.deviceID, r.stationGUID AS low_rated_station, DATE(r.ratingWhen) AS rating_date
  FROM ratings r
  JOIN (
    -- Αποθήκευση χρηστών που έχουν >2 χαμηλές αξιολογήσεις σε μία μέρα
    SELECT deviceID, DATE(ratingWhen) AS rating_date
    FROM ratings
    WHERE (rating1 + rating2 + rating3)/3.0 <= 1.5
    GROUP BY deviceID, DATE(ratingWhen)
    HAVING COUNT(*) > 2
) AS low_ratings_activity
  ON r.deviceID = low_ratings_activity.deviceID
 AND DATE(r.ratingWhen) = low_ratings_activity.rating_date
 -- Επιλογή μόνο των χαμηλών ακραίων βαθμολογιών από τον χρήστη για αυτή τη μέρα
WHERE (r.rating1 + r.rating2 + r.rating3)/3.0 <= 1.5;
```

Εικόνα 31: temp_low_ratings

Τέτοιες περιπτώσεις, μπορεί να είναι αποτέλεσμα συντονισμένης επίθεσης σε συγκεκριμένα πρατήρια για την μείωση της βαθμολογίας τους με πρακτικές αθέμιτου ανταγωνισμού.

```sql
-- Χρήστες με >= 1 ακραία υψηλή βαθμολογία και > 2 ακραίες χαμηλές βαθμολογίες σε μία μέρα
CREATE TEMPORARY TABLE temp_both_ratings_users (
    deviceID VARCHAR(255),
    high_rated_station VARCHAR(255),
    low_rated_station VARCHAR(255),
    rating_date DATE,
    INDEX (deviceID),
    INDEX (high_rated_station),
    INDEX (low_rated_station))
  AS
  SELECT
    highR.deviceID,
    highR.high_rated_station,
    lowR.low_rated_station,
    highR.rating_date
  FROM temp_high_ratings highR
  -- Σύνδεση των χρηστών που τηρούν τα κριτήρια των προηγούμενων temporary tables
  JOIN temp_low_ratings lowR ON highR.deviceID = lowR.deviceID
    AND highR.rating_date = lowR.rating_date;
```

Εικόνα 32: temp_both_ratings_users

Ο επόμενος πίνακας που δημιουργήθηκε είναι ο ‘temp_both_ratings_users’ (Εικόνα 32) όπου αφορά χρήστες που έχουν δώσει και πολύ υψηλές και πολύ χαμηλές βαθμολογίες την ίδια μέρα. Μάλιστα, πραγματοποιείται σύνδεση των 2 προηγούμενων πινάκων ώστε να βρεθούν οι χρήστες με τουλάχιστον 1 υψηλή βαθμολογία και 3 ή περισσότερες χαμηλές βαθμολογίες την ίδια μέρα. Οι χρήστες αυτοί λειτουργούν επαινώντας συγκεκριμένα πρατήρια ενώ παράλληλα βάζουν χαμηλές βαθμολογίες σε ανταγωνιστές τους. Είναι πιθανό τέτοιες περιπτώσεις να αφορούν επιχειρήσεις που προσπαθούν να μειώσουν τις βαθμολογίες των γειτονικών σταθμών.

Στη συνέχεια, δημιουργείται ο πίνακας ‘temp_suspicious_stations’ (Εικόνα 33) που αποθηκεύει τους ύποπτους σταθμούς που δέχονται υψηλές βαθμολογίες από 10 ή παραπάνω χρήστες του προηγούμενου πίνακα ‘temp_both_ratings_users’ (δηλαδή χρήστες με τουλάχιστον 1 υψηλή βαθμολογία και 3 ή περισσότερες χαμηλές βαθμολογίες την ίδια μέρα).

```
65      -- Σταθμοί που δέχονται υψηλές βαθμολογίες από >= 10 διαφορετικούς χρήστες του temp_both_ratings_users
66      CREATE TEMPORARY TABLE temp_suspicious_stations (
67          stationGUID VARCHAR(255),
68          INDEX (stationGUID))
69      AS
70      SELECT
71          high_rated_station AS stationGUID FROM temp_both_ratings_users
72      GROUP BY high_rated_station
73      HAVING COUNT(DISTINCT deviceID) >= 10;
```

Εικόνα 33: temp_suspicious_stations

```
75      -- Σταθμοί που δέχονται ύποπτες χαμηλές βαθμολογίες
76      CREATE TEMPORARY TABLE temp_targeted_stations ( -- Σταθμοί που λαμβάνουν χαμηλές βαθμολογίες από >= 5 διαφορετικούς χρήστες του temp_both_ratings_users
77          stationGUID VARCHAR(255),
78          INDEX (stationGUID))
79      AS
80      SELECT
81          low_rated_station AS stationGUID
82      FROM temp_both_ratings_users
83      -- Επιλογή των εγγραφών όπου ο χρήστης έχει δώσει τουλάχιστον μία υψηλή βαθμολογία(>=4.5) σε ύποπτο σταθμό
84      WHERE high_rated_station IN (SELECT stationGUID FROM temp_suspicious_stations)
85      GROUP BY low_rated_station
86      -- Ο σταθμός θεωρείται στοχευμένος αν έχει λάβει χαμηλές βαθμολογίες(<=1.5) από >= 5 διαφορετικούς χρήστες
87      -- όπου παράλληλα έχουν δώσει υψηλή βαθμολογία(>=4.5) σε κάποιον ύποπτο σταθμό την ίδια μέρα
88      HAVING COUNT(DISTINCT deviceID) >= 5;
```

Εικόνα 34: temp_targeted_stations

Στον επόμενο προσωρινό πίνακα ‘temp_targeted_stations’ αποθηκεύονται οι σταθμοί που δέχονται χαμηλές αξιολογήσεις από τουλάχιστον 5 διαφορετικούς χρήστες, οι οποίοι την ίδια μέρα έχουν δώσει και τουλάχιστον μία υψηλή βαθμολογία σε ύποπτο

σταθμό. Οι σταθμοί αυτοί μπορούν να θεωρηθούν στοχευμένοι καθώς δέχονται αρνητικές αξιολογήσεις από ύποπτους χρήστες.

Ο τελευταίος πίνακας 'temp_suspicious_users', υλοποιήθηκε ώστε να εντοπίζει χρήστες οι οποίοι μέσα σε μία μέρα πραγματοποιούν συνδυασμό ύποπτων ακραίων αξιολογήσεων. Ο συνδυασμός αυτών των αξιολογήσεων είναι, είτε οι χρήστες να δίνουν υψηλή κριτική σε ύποπτο σταθμό και τουλάχιστον μία χαμηλή σε στοχευμένο, είτε, να δίνουν χαμηλή κριτική σε στοχευμένο σταθμό και τουλάχιστον μία υψηλή σε κάποιο ύποπτο σταθμό. Αν οι χρήστες έχουν τουλάχιστον τέσσερις ακραίες αξιολογήσεις, και πράγματι υπάρχει κάποιος συνδυασμός υψηλών και χαμηλών αξιολογήσεων τότε, οι χρήστες συμπεριλαμβάνονται στον πίνακα και ο αριθμός των ακραίων αξιολογήσεων προστίθεται στον μετρητή(suspicious_rating_count). Παρουσιάζεται παρακάτω στην Εικόνα 35.

```
90    -- Χρήστες που έχουν τουλάχιστον 4 ύποπτες αξιολογήσεις (συνδυασμός θετικών και αρνητικών) συνολικά
91    CREATE TEMPORARY TABLE temp_suspicious_users (
92        deviceID VARCHAR(255),
93        suspicious_rating_count INT,
94        INDEX (deviceID)
95    ) AS
96    SELECT
97        r.deviceID,
98        COUNT(*) AS suspicious_rating_count -- Σύνολο των ύποπτων βαθμολογιών του χρήστη
99    FROM ratings r
100   -- Χρήστες με τουλάχιστον μία υψηλή αξιολόγηση σε ύποπτο σταθμό (temp_suspicious_stations)
101   -- και την ίδια μέρα τουλάχιστον μία αρνητική αξιολόγηση σε στοχευμένο σταθμό (temp_targeted_stations)
102   WHERE ((r.stationGUID IN (SELECT stationGUID FROM temp_suspicious_stations) AND (r.rating1 + r.rating2 + r.rating3)/3.0 >= 4.5
103           AND EXISTS (
104               SELECT 1 -- Έλεγχος εάν ο χρήστης την ίδια μέρα έβαλε τουλάχιστον μία χαμηλή κριτική σε στοχευμένο σταθμό
105               FROM temp_low_ratings tlr
106               WHERE tlr.deviceID = r.deviceID
107               AND tlr.rating_date = DATE(r.ratingWhen)
108               AND tlr.low_rated_station IN (SELECT stationGUID FROM temp_targeted_stations)
109           )
110       )
111   OR -- Χρήστες με τουλάχιστον μία χαμηλή αξιολόγηση σε στοχευμένο σταθμό(temp_targeted_stations)
112      -- και την ίδια μέρα τουλάχιστον μία υψηλή αξιολόγηση σε ύποπτο σταθμό (temp_suspicious_stations)
113       (r.stationGUID IN (SELECT stationGUID FROM temp_targeted_stations) AND (r.rating1 + r.rating2 + r.rating3)/3.0 <= 1.5
114        AND EXISTS (
115             SELECT 1 -- Έλεγχος εάν ο χρήστης την ίδια μέρα έβαλε τουλάχιστον μία υψηλή κριτική σε ύποπτο σταθμό
116             FROM temp_high_testratings highR
117             WHERE highR.deviceID = r.deviceID
118             AND highR.rating_date = DATE(r.ratingWhen) AND highR.high_rated_station IN (SELECT stationGUID FROM temp_suspicious_stations)
119        )
120       )
121   )
122   GROUP BY r.deviceID
123   HAVING COUNT(*) >= 4; -- Χρήστες που έχουν συνολικά 4 ή περισσότερες ακραίες αξιολογήσεις με συνδυασμό ακραίων τιμών
```

Εικόνα 35: temp_suspicious_users

Για την εκτέλεση του update για το status14, έγινε επανάληψη του ελέγχου WHERE ώστε να μην σημειωθούν όλες οι εγγραφές των χρηστών ως ύποπτες αλλά μόνο αυτές οι οποίες περιέχουν τον συνδυασμό ακραίων βαθμολογιών όπως αναφέρθηκε στο τελευταίο πίνακα. Έτσι, εξασφαλίζεται ότι μόνο οι αξιολογήσεις που πραγματικά συγκροτούν ύποπτη ενέργεια ενημερώνονται ως τέτοιες, αποφεύγοντας την ενημέρωση για κριτικές που έγιναν από χρήστες με ύποπτη δραστηριότητα αλλά οι ίδιες δεν εμπίπτουν στα κριτήρια (Εικόνα 36).

Έτσι, προέκυψαν 2553 απατηλές αξιολογήσεις.

```
125     -- Update status14
126     ALTER TABLE ratings ADD COLUMN status14 TINYINT(1) DEFAULT 0;
127     UPDATE ratings r
128     SET r.status14 = 1
129     WHERE r.deviceID IN (SELECT deviceID FROM temp_suspicious_users)
130     AND (
131         (r.stationGUID IN (SELECT stationGUID FROM temp_suspicious_stations)
132          AND (r.rating1 + r.rating2 + r.rating3)/3.0 >= 4.5
133          AND EXISTS (
134              SELECT 1
135              FROM temp_low_ratings tlr
136              WHERE tlr.deviceID = r.deviceID
137              AND tlr.rating_date = DATE(r.ratingWhen)
138              AND tlr.low_rated_station IN (SELECT stationGUID FROM temp_targeted_stations)
139          ))
140         OR
141         (r.stationGUID IN (SELECT stationGUID FROM temp_targeted_stations)
142          AND (r.rating1 + r.rating2 + r.rating3)/3.0 <= 1.5
143          AND EXISTS (
144              SELECT 1
145              FROM temp_high_ratings highR
146              WHERE highR.deviceID = r.deviceID
147              AND highR.rating_date = DATE(r.ratingWhen)
148              AND highR.high_rated_station IN (SELECT stationGUID FROM temp_suspicious_stations)
149          ))
150     );
```

Εικόνα 36: Status14 Update

Από τα αποτελέσματα που προέκυψαν είναι προφανής η συμπεριφορά αθέμιτου ανταγωνισμού που παρουσιάζεται από διάφορες συσκευές μεταξύ των σταθμών.

Αρχικά, όπως μπορεί να παρατηρηθεί από ένα μέρος των αποτελεσμάτων στην Εικόνα 37, βλέπουμε ότι η συσκευή με αναγνωριστικό ‘android-01f98e3778aa3bed’ εμφανίζει στις αξιολογήσεις της μία επανειλημμένη σειρά μονάδων σε κάθε πεδίο, με μόνη διαφορά στην αξιολόγηση της σε συγκεκριμένο πρατήριο το οποίο βαθμολογεί με την υψηλότερη βαθμολογία. Επίσης μπορεί να παρατηρηθεί ότι οι αξιολογήσεις της συμβαίνουν σε πολύ μικρό χρονικό διάστημα (spam). Κοιτώντας την δεύτερη περίπτωση που αφορά άλλη συσκευή, παρατηρούμε πάλι ότι όλες οι βαθμολογίες λαμβάνουν την μονάδα σαν μέσο όρο εκτός από την μοναδική κριτική που διαφέρει ξανά για το ίδιο πρατήριο. Επιπλέον, οι αξιολογήσεις που έδωσε ως μονάδα, αφορούν στο μεγαλύτερο κομμάτι τα

ίδια πρατήρια και φαίνεται να υποβλήθηκαν την ίδια ημέρα και ώρα. Επομένως, παρατηρούμε πως ο σταθμός με χρήση διάφορων deviceID επιδιώκει την αύξηση της βαθμολογίας του, ενώ παράλληλα με τα ίδια deviceID στοχοποιεί άλλα πρατήρια με σκοπό να μειώσει την συνολική τους βαθμολογία. Πιθανότατα πρόκειται για σταθμούς που βρίσκονται στην ίδια περιοχή με το συγκεκριμένο πρατήριο και με αυτόν τον τρόπο προωθείται σε αντίθεση με τα υπόλοιπα. Πραγματοποιώντας έλεγχο των συσκευών μέσω της οπτικής αναπαράστασής σε χάρτη που διατίθεται από το σύστημα διαχείρισης του fuelGR (Εικόνα 38), διαπιστώνεται πως πράγματι αυτές οι ύποπτες αξιολογήσεις αφορούν πρατήρια τα οποία βρίσκονται σε κοντινή απόσταση.

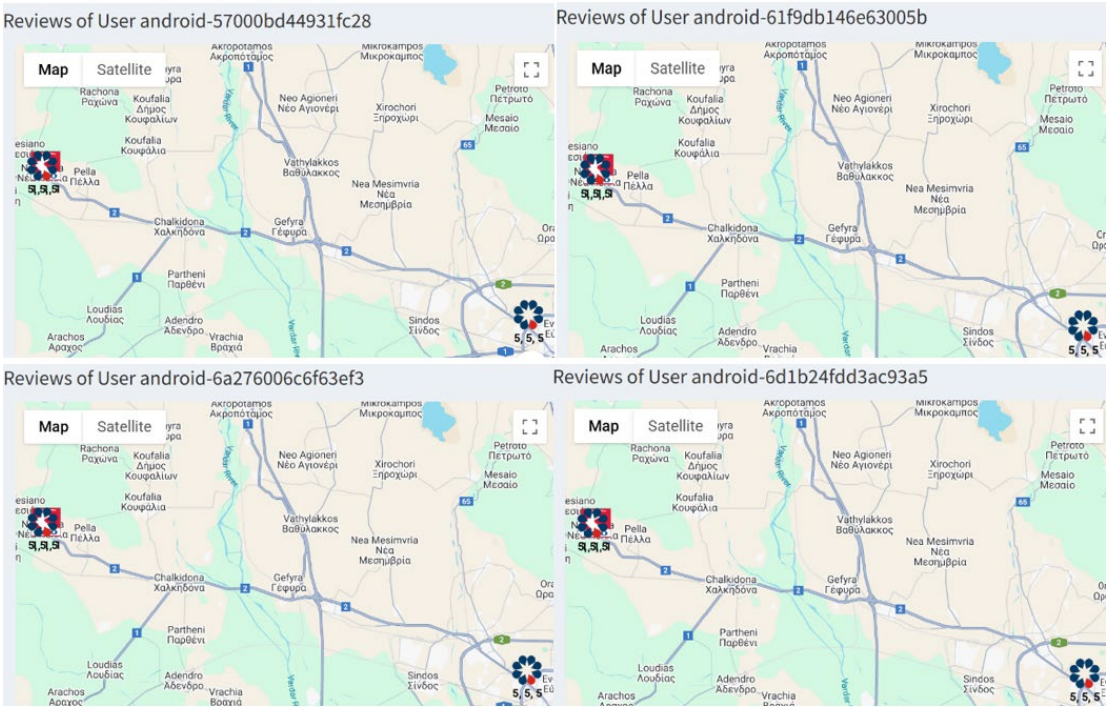
deviceID	stationGUID	clientIP	rating1	rating2	rating3	ratingWhen
android-01f98e3778aa3bed	E57237C9-DC71-E91A-3886-51409E3C	185.51.133.57	5	5	5	2023/09/27 22:09:22
android-01f98e3778aa3bed	A83824E4-6A76-FB49-6C84-88F24B00	185.51.133.57	1	1	1	2023/09/27 22:09:32
android-01f98e3778aa3bed	7C29B156-D5AE-928E-23AF-7B805B9C	185.51.133.57	1	1	1	2023/09/27 22:09:41
android-01f98e3778aa3bed	56657B2E-9752-8936-86BA-9077AC5B	185.51.133.57	1	1	1	2023/09/27 22:09:56
android-01f98e3778aa3bed	D631DFFC-8957-AC4A-7047-7EE9E289	185.51.133.57	1	1	1	2023/09/27 22:10:06
android-01f98e3778aa3bed	076A6564-FE05-0491-7C9D-4BAE6A58	185.51.133.57	1	1	1	2023/09/27 22:10:21
android-01f98e3778aa3bed	6DCABADD-9A26-843F-0AD8-B69E15E0	185.51.133.57	1	1	1	2023/09/27 22:10:49
android-01f98e3778aa3bed	490C9077-EE7D-C07F-0D87-16C391DD	185.51.133.57	1	1	1	2023/09/27 22:10:59
android-01f98e3778aa3bed	8A32BA0D-B0D9-6884-6744-C7864E8D	185.51.133.57	1	1	1	2023/09/27 22:11:24
android-01f98e3778aa3bed	2E09555F-3470-0072-7F00-41040340	185.51.133.57	1	1	1	2023/09/27 22:11:31

deviceID	stationGUID	clientIP	rating1	rating2	rating3	ratingWhen
android-0b1faf0fd87feda6	E57237C9-DC71-E91A-3886-51409E3C	185.51.133.57	5	5	5	2023/09/27 21:57:20
android-0b1faf0fd87feda6	A83824E4-6A76-FB49-6C84-88F24B00	185.51.133.57	1	1	1	2023/09/27 21:57:29
android-0b1faf0fd87feda6	7C29B156-D5AE-928E-23AF-7B805B9C	185.51.133.57	1	1	1	2023/09/27 21:57:59
android-0b1faf0fd87feda6	56657B2E-9752-8936-86BA-9077AC5B	185.51.133.57	1	1	1	2023/09/27 22:00:00
android-0b1faf0fd87feda6	D631DFFC-8957-AC4A-7047-7EE9E289	185.51.133.57	1	1	1	2023/09/27 22:00:59
android-0b1faf0fd87feda6	076A6564-FE05-0491-7C9D-4BAE6A58	185.51.133.57	1	1	1	2023/09/27 22:01:12
android-0b1faf0fd87feda6	6DCABADD-9A26-843F-0AD8-B69E15E0	185.51.133.57	1	1	1	2023/09/27 22:01:22
android-0b1faf0fd87feda6	490C9077-EE7D-C07F-0D87-16C391DD	185.51.133.57	1	1	1	2023/09/27 22:01:39
android-0b1faf0fd87feda6	5AE32E0B-653A-DDDD-2C79-8EB38167	185.51.133.57	1	1	1	2023/09/27 22:02:32

Εικόνα 37: Παράδειγμα αθέμιτης συμπεριφοράς.

Εικόνα 38: Ύποπτες κριτικές με αναπαράσταση στον χάρτη.

Στο παράδειγμα στην Εικόνα 39, εντοπίστηκε περίπτωση όπου διαφορετικές συσκευές βαθμολογούν με την μέγιστη βαθμολογία τα ίδια δύο πρατήρια, και στοχεύουν συγκεκριμένο πρατήριο βαθμολογώντας το με την ελάχιστη βαθμολογία. Αυτές οι αξιολογήσεις είναι και οι μοναδικές που έχουν εκτελέσει οι ορισμένες συσκευές, παρουσιάζοντας έτσι μία ύποπτη συμπεριφορά.



Εικόνα 39: Ύποπτη συμπεριφορά χρηστών με αναπαράσταση στον χάρτη.

Το τελικό script status15 υπολογίζει τον αριθμό των ύποπτων κριτικών για κάθε χρήστη και αν είναι μεγαλύτερος του 10 με ποσοστό status = ‘fraudulent’ 50% ή περισσότερο ,τότε ο χρήστης θεωρείται ύποπτος και μετατρέπει όλες τις αξιολογήσεις του στο status15 ως ψεύτικες.

```
ALTER TABLE ratings DROP COLUMN status15;
ALTER TABLE ratings ADD COLUMN status15 BOOLEAN DEFAULT FALSE;
UPDATE ratings r
JOIN (
    SELECT deviceID
    FROM ratings
    GROUP BY deviceID
    HAVING COUNT(*) > 10
       AND SUM(CASE WHEN status = 'fraudulent' THEN 1 ELSE 0 END) >
           SUM(CASE WHEN status != 'fraudulent' THEN 1 ELSE 0 END)
) s ON r.deviceID = s.deviceID
SET r.status15 = TRUE;
```

Εικόνα 40: Status15

Με την ολοκλήρωση αυτών των ερωτημάτων, διαθέτουμε πλέον ένα ισχυρό σύνολο εργαλείων για την ανίχνευση και τον έλεγχο ανωμαλιών των δεδομένων, ενισχύοντας την αξιοπιστία και την προστασία του συστήματος από παραπλανητικές συμπεριφορές.

4.3 Διαχείριση των δεδομένων

Με την ολοκλήρωση της διαδικασίας κατά την οποία πραγματοποιήθηκε η επεξεργασία των δεδομένων, ο πίνακας ‘ratings’ έχει ενημερωθεί με νέα πεδία που συνιστούν τις πρόσθετες πληροφορίες που προέκυψαν από τις εκτελέσεις των SQL ερωτημάτων. Κάθε εγγραφή αξιολόγησης, κατηγοριοποιήθηκε βάσει της συμπεριφοράς της σε νέες στήλες (status1,…,status15), με κάθε μία να ενεργεί χρησιμοποιώντας διαφορετικό σύνολο κριτηρίων, με το οποίο ορίζει εάν τελικά τηρεί τα κριτήρια. Επίσης, η νέα στήλη ‘status’ δρα ως την προσωρινή τελική κατάσταση κάθε αξιολόγησης, καταγράφοντάς την ως ‘fraudulent’ σε περίπτωση όπου τηρεί οποιοδήποτε κριτήριο από τα status ελέγχου, ενώ όσες δεν εμφάνισαν ύποπτη συμπεριφορά καταγράφονται ως ‘normal’. Με την εξαγωγή του επεξεργασμένου με ανανεωμένα πεδία πίνακα, ακολουθεί η διερευνητική του ανάλυση ώστε να υλοποιηθεί μία διαδικασία καθαρισμού και περαιτέρω φιλτραρίσματος στις εγγραφές του με σκοπό να μειώσει τον αριθμό των αξιολογήσεων που έως τώρα αποκαλούνται αβάσιμες. Η μείωση αυτή αποσκοπεί στην δημιουργία ενός τελικού και αντιπροσωπευτικού συνόλου δεδομένων, με σκοπό να χωριστεί σε υποσύνολα εκπαίδευσης (training) και ελέγχου (testing).

	Status	Fraudulent_count	Normal_count
2	status1	11527	95288
3	status2	9419	97396
4	status3	8006	98809
5	status4	6228	100587
6	status5	9342	97473
7	status6	4600	102215
8	status7	73	106742
9	status8	11161	95654
10	status9	3011	103804
11	status10	5013	101802
12	status11	10646	96169
13	status12	21355	85460
14	status13	6780	100035
15	status14	2553	104262
16	status15	22053	84762

Εικόνα 41: Αριθμός απατηλών/πραγματικών αξιολογήσεων ανά status.

Με την εισαγωγή του νέου συνόλου δεδομένων στο περιβάλλον του Excel, αρχικά πραγματοποιήθηκε μία αρχική μέτρηση των αξιολογήσεων ανά κατάσταση status. Παρατηρήθηκε, λοιπόν, ένας διαχωρισμός ποσοστού 71.26% των κριτικών που θεωρούνται πραγματικές και 28.38% των αξιολογήσεων που θεωρήθηκε πως λειτουργούν παραπλανητικά. Τα ποσοστά αυτά ανέρχονται σε 76.507 και 30.308 αντίστοιχα από τις 106.815 συνολικές αξιολογήσεις. Στη συνέχεια υπολογίστηκε για κάθε κατάσταση status ο αριθμός των κριτικών ανά κατηγορία, όπου διαπιστώθηκε μία γενική εικόνα των αποτελεσμάτων.

Στη συνέχεια προστέθηκε μία νέα κενή γραμμή πάνω από τις υπάρχουσες γραμμές του πίνακα ώστε να προστεθούν οι συντελεστές βαρύτητας των status πάνω από αυτά. Επιπλέον δημιουργήθηκε μία νέα στήλη filter_score η οποία λειτουργεί σαν φίλτρο των status πολλαπλασιάζοντάς τα με το βάρος που τους αντιστοιχεί και αθροίζοντάς τα. Για την εύρεση των τιμών του φίλτρου εκτελέστηκε ο παρακάτω υπολογισμός (AA3=L$1*L3 + M$1*M3 + N$1*N3 + O$1*O3 + P$1*P3 + Q$1*Q3 + R$1*R3 + S$1*S3 + T$1*T3 + U$1*U3 + V$1*V3 + W$1*W3 + X$1*X3 + Y$1*Y3 + Z$1*Z3)για κάθε κελί της στήλης AA. Στη συνέχεια προστέθηκε ένα όριο στη θέση AC2 που ήταν απαραίτητο ώστε να λειτουργεί σαν ελεγκτής για την νέα στήλη status_final. Για να υπολογιστεί εν τέλη αν μία εγγραφή τελικά είναι ψευδής ή όχι, πραγματοποιείται σύγκριση της τιμής του φίλτρου κάθε εγγραφής με το threshold που έχει οριστεί (=IF(AA3>=AC2,1,0)). Αν η τιμή του φίλτρου είναι μεγαλύτερη ή ίση από το threshold τότε η εγγραφή οριστικοποιείται ως ψευδής, λαμβάνοντας τη τιμή 1.

Στόχο της εφαρμογής του τελικού φιλτραρίσματος στις 15 καταστάσεις που δημιουργήθηκαν στο SQL, έθεσε η δημιουργία ενός τελικού ταξινομητή status_final όπου θα κρίνει τελικά ποιες αξιολογήσεις είναι ψευδείς ή αληθινές, με σκοπό να καταλήξει σε ένα ποσοστό που θα χωρίζει τις εγγραφές σε 20% έως 25% και 75% ως 80% αντίστοιχα. Αυτό διότι, θα απορριφθούν τυχών αυστηρές ύποπτες ταυτοποιήσεις κριτικών που έχουν συμβεί καθώς και διότι αποτελεί έναν ικανοποιητικό διαχωρισμό των δεδομένων που μπορεί να τα καταστήσει έτοιμα για την εκτέλεση διαδικασίας testing και training κάποιου μελλοντικού μοντέλου. Για να βρεθούν τα κατάλληλα βάρη που οδήγησαν σε τέτοια ποσοστά πραγματοποιήθηκαν πολλές προσπάθειες με αλλαγές των βαρών.

Τελικά χρησιμοποιήθηκαν οι συντελεστές βαρύτητας και το threshold που φαίνονται στην Εικόνα 42, οδηγώντας σε έναν αριθμό 21382 ψευδών και 85433 πραγματικών αξιο-λογήσεων. Πρόκειται δηλαδή για ποσοστά 20,02% και 79,98% αντίστοιχα.

L	M	N	O	P	Q	R	S	T
3	2	1.5	5	4.25	3.5	3	3.5	6
status1	status2	status3	status4	status5	status6	status7	status8	status9

U	V	W	X	Y	Z	AA	AB	AC
1	2	4	2	6	0.5			Threshold
status10	status11	status12	status13	status14	status15	filter_score	status_fina	6

Εικόνα 42: Βάρη και threshold.

5 Συμπεράσματα

Συνοψίζοντας, η πτυχιακή εργασία επικεντρώνεται στην δημιουργία και εφαρμογή δομημένων ερωτημάτων για την ανάλυση των δεδομένων, μελετώντας σε βάθος ύποπτα μοτίβα με σκοπό να ανιχνευτούν ύποπτες ή δόλιες αξιολογήσεις. Για κάθε έναν συνδυασμό ερωτημάτων που εκτελέστηκε, δημιουργήθηκε ένας διαφορετικός δείκτης κατηγοριοποίησης στον οποίον αποθηκεύτηκε η τιμή αξιολόγησης για κάθε κριτική. Στη συνέχεια, αυτοί οι δείκτες κατηγοριοποίησης συνδυάστηκαν με τα αντίστοιχα βάρη που ανατέθηκαν στο καθένα ώστε να προκύψει η τελική τους ετικέτα κατηγοριοποίησης σε απατηλές ή πραγματικές αξιολογήσεις.

Η πτυχιακή εργασία κατέληξε σε ένα ικανοποιητικό ποσοστό ψευδών και πραγματικών αξιολογήσεων όπου μπορεί να χρησιμοποιηθεί σε μελλοντική εργασία για την υλοποίηση ενός μοντέλου επιβλεπόμενης μάθησης. Μελλοντικά θα μπορούσε επίσης να αποτελέσει μία εργασία η ενσωμάτωση των αναπτυγμένων ερωτημάτων στην εφαρμογή ώστε να ενημερώνονται οι χρήστες για την αξιοπιστία των αξιολογήσεων σε πραγματικό χρόνο.

Βιβλιογραφία

[1] Chen T, Samaranayake P, Cen X, Qi M and Lan Y-C, "The Impact of Online Reviews on Consumers' Purchasing Decisions: Evidence From an Eye-Tracking Study", 2022.

[2] UK Department for Business & Trade, 'Estimating the prevalence and impact of fake online reviews', 2023.

[3] Brett Hollenbeck, Davide Proserpio, Sherry He, "The Market for Fake Reviews", 2021

[4] https://www.ibm.com

[5] Joni Salminen, Chandrashekhar Kandpal, Ahmed Mohamed Kamel, Soon-qyo Jung, Bernard J.Jansen "Creating and detecting fake reviews of online products" , 2022.

[6] https://www.geeksforgeeks.org – Πλατφόρμα διαδικτυακής εκπαίδευσης.

[7] Vladimir Nasteski, "An overview of the supervised machine learning methods", 2017.

[8] Dr. Kamlesh Lakhwani, Dr Ruchi, Kamal Kant Hiran, Ritesh Kumar Jain, ,"Machine Learning: Master Supervised and Unsupervised Learning Algorithms with Real Examples", 2021.

[9] Aliaksandr Barushka, Michal Munk & Petr Hajek, "Fake consumer review detection using deep neural networks integrating word embeddings and emotion mining", 2020.

[10] Elias Mmbongeni Sibanda, Gireen Naidu & Tranos Zuva, "A Review of Evaluation Metrics in Machine Learning Algorithms", 2023.

[11] Petr Silhavy, Radek Silhavy, "Artificial Intelligence Application in Networks and Systems", 2023.

[12] Arjun Mukherjee, Santosh K C, "On the temporal Dynamics of Opinion Spamming: Case Studies on Yelp", 2016.

[13] David Savage, Pauline Chou, Qingmai Wang, Xinghuo Yu, Xiuzhen Zhang, "Detection of opinion spam based on anomalous rating deviation", 2016.

[14] www.w3schools.com - Πλατφόρμα διαδικτυακής εκπαίδευσης.