ΕΘΝΙΚΟ ΜΕΤΣΟΒΙΟ ΠΟΛΥΤΕΧΝΕΙΟ

ΣΧΟΛΗ ΗΛΕΚΤΡΟΛΟΓΩΝ ΜΗΧΑΝΙΚΩΝ ΚΑΙ Μ/Υ

ΠΑΝΕΠΙΣΤΗΜΙΟ ΠΕΙΡΑΙΩΣ

ΣΧΟΛΗ ΝΑΥΤΙΛΙΑΣ ΚΑΙ ΒΙΟΜΗΧΑΝΙΑΣ

ΤΜΗΜΑΤΟΣ ΒΙΟΜΗΧΑΝΙΚΗΣ ΔΙΟΙΚΗΣΗΣ & ΤΕΧΝΟΛΟΓΙΑΣ

ΔΙΑΠΑΝΕΠΙΣΤΗΜΙΑΚΟ ΠΡΟΓΡΑΜΜΑ ΜΕΤΑΠΤΥΧΙΑΚΩΝ ΣΠΟΥΔΩΝ

«ΤΕΧΝΟ-ΟΙΚΟΝΟΜΙΚΑ ΣΥΣΤΗΜΑΤΑ»

ΔΙΠΛΩΜΑΤΙΚΗ ΕΡΓΑΣΙΑ

# Ανίχνευση bots στο Twitter με μεθόδους μηχανικής μάθησης

ΜΟΥΝΤΖΟΥΡΗΣ ΧΡΗΣΤΟΣ

ΕΠΙΒΛΕΠΩΝ

ΚΩΝΣΤΑΝΤΙΝΟΣ ΔΕΜΕΣΤΙΧΑΣ

ΙΟΥΝΙΟΣ 2023

# Πίνακας Περιεχομένων

# Περίληψη

Σήμερα, η εκρηκτική ανάπτυξη των μέσων κοινωνικής δικτύωσης έχει καλλιεργήσει ένα πρόσφορο έδαφος για την δραστηριοποίηση των bots στα κοινωνικά δίκτυα. Η ενσωμάτωση των κοινωνικών δικτύων στον πυρήνα της καθημερινότητας των ανθρώπων σε συνδυασμό με την επιρροή τους στις κοινωνικές, οικονομικές και την πολιτικές εξελίξεις, εγείρει σημαντικές ανησυχίες για τον ρόλο των bots, καθιστώντας επιτακτική την ανάγκη λήψης μέτρων για τον περιορισμό τους. Παράλληλα, η πρόοδος που συντελείται στα πλαίσια της 4ης Βιομηχανικής Επανάστασης επιτρέπει την αξιοποίηση μεγάλων δεδομένων και αναδυόμενων τεχνολογιών από τα bots για την μοντελοποίηση και την ρεαλιστική αποτύπωση ανθρώπινων μοτίβων συμπεριφοράς, δυσχεραίνοντας την διάκρισή τους από πραγματικούς χρήστες. Οι μέθοδοι μηχανικής μάθησης αποδεικνύονται σε ισχυρά εργαλεία για την ανίχνευση bots στα κοινωνικά δίκτυα, αφού η δυναμική φύση τους και η διαρκής διαδικασία εκπαίδευσής τους, επιτρέπει την συνεχή αναπροσαρμογή στα μεταβαλλόμενα συμπεριφορικά πρότυπα των πραγματικών χρηστών των κοινωνικών δικτύων. Η παρούσα εργασία εξετάζει το πρόβλημα των bots με επίκεντρο το Twitter μέσα από πραγματικά σύνολα δεδομένων, κατασκευάζοντας ταξινομητές επιβλεπόμενης μηχανικής μάθησης για την κατηγοριοποίηση των χρηστών της πλατφόρμας και αξιολογώντας την προβλεπτική ικανότητά τους.

# Abstract

Today, the explosive growth of social media has cultivated a fertile ground for bots to operate in. The integration of social networks into the core of people's daily lives, combined with their influence on social, economic and political developments, raises significant concerns about the role of bots, making it imperative to take measures to curb them. At the same time, the progress made in the context of the 4th Industrial Revolution allows the use of big data and emerging technologies to model and realistically capture human behaviour patterns, empowering the action of bots and making it more difficult to distinguish them from real users. Machine learning methods prove to be powerful tools for detecting bots in social networks, since their dynamic nature and continuous training process allows them to continuously adapt to the changing behavioral patterns of real users of social networks. This paper addresses the problem of Twitter-centric bots through real datasets, constructing supervised machine learning classifiers to categorize the platform's users and evaluating their predictive ability.

# 1. Εισαγωγή

## 1.1. Το Πρόβλημα των Bots στα Μέσα Κοινωνικής Δικτύωσης

Σήμερα, τα κοινωνικά δίκτυα έχουν καθιερωθεί ως το κυρίαρχο μέσο επικοινωνίας και αλληλεπίδρασης μεταξύ των ανθρώπων, μεταβάλλοντας σημαντικά τον τρόπο θέασης των κοινωνικών σχέσεων. Τα μέσα κοινωνικής δικτύωσης προσφέρουν την ψηφιακή υποδομή για την ανταλλαγή μηνυμάτων, περιεχομένου και απόψεων μεταξύ των χρηστών, διεισδύοντας σημαντικά στην καθημερινότητα των ανθρώπων. Υπολογίζεται πως περισσότεροι από 4.95 δισεκατομμύρια άνθρωποι διαθέτουν ενεργό λογαριασμό σε τουλάχιστον ένα μέσο κοινωνικής δικτύωσης, ενώ εκτιμάται πως την επόμενη πενταετία ο αριθμός αυτός θα αυξηθεί, ξεπερνώντας τα 6 δισεκατομμύρια (Statista, 2023). Ο χρόνος που αφιερώνει ο μέσος χρήστης στα κοινωνικά δίκτυα ανέρχεται σε 145 λεπτά ημερησίως, παραμένοντας αμετάβλητος κατά την τελευταία πενταετία, με το 36.2% των χρηστών να δηλώνει πως καλύπτει τον κενό χρόνο του μέσα από την ενασχόλησή του με αυτά (Statista, 2023).

Η πρόταση αξίας μεταξύ των κοινωνικών δικτύων διαφοροποιείται σημαντικά με αποτέλεσμα την ανισόρροπη κατανομή των χρηστών. Η πληρέστερη εμπειρία κοινωνικής δικτύωσης συναντάται στο Facebook, προσφέροντας την δυνατότητα ανάρτησης και αλληλεπίδρασης με περιεχόμενο που εκτείνεται πέραν των στενών πλαισίων μιας τυπικής σελίδας ροής, όπως σε κοινότητες αγοροπωλησιών, εκδηλώσεων, εράνων, αντιμετώπισης κρίσεων και παιχνιδιών. Το TikTok και το Instagram τοποθετούν την ψυχαγωγία στο επίκεντρο της πρότασης αξίας τους, επιτρέποντας τον διαμοιρασμό σύντομου και δημιουργικού περιεχομένου, ευνοώντας την ανάπτυξη ιογενών τάσεων. Το Twitter υιοθετεί τα χαρακτηριστικά μιας κοινότητας ιστολογίου, όπως πολυεπίπεδες αλυσίδες απαντήσεων, προωθώντας τον δημόσιο διάλογο και συμβάλλοντας στην έγκαιρη ενημέρωση γύρω ζητήματα της επικαιρότητας. Το YouTube επιτρέπει τον διαμοιρασμό, την αναζήτηση, την κοινοποίηση και την αλληλεπίδραση οπτικοακουστικού περιεχομένου μεταξύ των χρηστών του. Τέλος, το LinkedIn επικεντρώνεται στη κοινωνική δικτύωση μεταξύ επαγγελματιών, προωθώντας την προβολή της επαγγελματικής δραστηριότητας των χρηστών και απλοποιώντας την αναζήτηση θέσεων εργασίας.

Την τελευταία δεκαετία, σημειώνεται μια ραγδαία αύξηση της έντασης της δραστηριότητας των bots στις πλατφόρμες κοινωνικής δικτύωσης, εγείροντας σοβαρές κοινωνικές, οικονομικές και πολιτικές ανησυχίες. Μάλιστα, από το 2012 και μετά, η

συνολικότερη κίνηση που προέρχεται από bots στο Διαδίκτυο καταγράφει αξιοσημείωτη ανοδική τάση, ξεπερνώντας την αντίστοιχη των ανθρώπων. Το 2016, τα bots βρέθηκαν για πρώτη φορά στο επίκεντρο του ενδιαφέροντος της επιστημονικής κοινότητας, σε μία προσπάθεια συστηματικής μελέτης και ανάλυσης του ρόλου που διαδραμάτισαν στις Προεδρικές Εκλογές των ΗΠΑ. Υπολογίζεται πως η δράση των bots στο Twitter ενίσχυσε σημαντικά το εκλογικό ποσοστό του Ρεπουμπλικανικού κόμματος, συνεισφέροντας σε αύξηση κατά 3.23% (Yuriy Gorodnichenko, 2018). Υπό αυτό το πρίσμα, στα bots αποδόθηκε ένας πρώιμος ορισμός, αυτός των αυτοματοποιημένων λογαριασμών που εξυπηρετούν σκοπούς παραπληροφόρησης και προπαγάνδας. Έκτοτε, η φύση και ο ρόλος των bots μεταβάλλεται διαρκώς χάρις την αξιοποίηση της δυναμική των τεχνολογιών αιχμής, όπως η Τεχνητή Νοημοσύνη, μετασχηματιζόμενα σε οντότητες με χαρακτηριστικά που προσομοιάζουν σε εκείνα των ανθρώπων, μιμούμενα ακόμα και σύνθετες ανθρώπινες γνωστικές λειτουργίες. Έτσι, παρατηρείται μια συνεχής ενσωμάτωση συμπεριφορικών χαρακτηριστικών και εξέλιξη των υφιστάμενων, επεκτείνοντας το πεδίο δράσης τους.

Οι σκοποί που εξυπηρετούν τα bots δεν περιορίζονται στα στενά πλαίσια της παραπληροφόρησης και της προπαγάνδας, καθώς εργαλειοποιούνται για την διάδοση ψευδών ειδήσεων, την χειραγώγηση του δημόσιου διαλόγου, την διαμόρφωση της κοινής γνώμης, ακόμα και ως μέσα τέλεσης απατών και εγκλημάτων στον κυβερνοχώρο. Αυτή η πολυσχιδής δράση των bots καθιστά ιδιαίτερα δύσκολη την περιγραφή τους με περιεκτικό και πλήρες τρόπο. Αναπροσαρμόζοντας τον πρώιμο ορισμό τους και συγκλίνοντας στις κύριες συνιστώσες των ορισμών της διεθνούς βιβλιογραφίας, ως bots νοούνται οι σύνθετες προγραμματισμένες οντότητες που ενεργούν αυτόνομα και αυτοματοποιημένα στο Διαδίκτυο κατά τρόπο που μιμείται τον άνθρωπο, εξυπηρετώντας σκοπούς χειραγώγησης της πληροφορίας, διατάραξης της κοινωνικής συνοχής και διαμόρφωσης της κοινής γνώμης.

Συχνά, τα bots παρουσιάζουν υψηλή προστιθέμενη αξία σε εκφάνσεις της κοινωνικής και οικονομικής ζωής, αλλά και σε πτυχές της καθημερινότητας των ανθρώπων. Έτσι, παρατηρείται ένας άτυπος διαχωρισμός μεταξύ των bot σε «καλά» και «κακά» ή «ηθικά» και «μη-ηθικά». Ο βαθμός ψηφιακής μετάβασης των κοινωνιών, το ηθικό πλαίσιο που περιβάλλει τις κοινωνίες και οι κανόνες που θωρακίζουν την κοινωνική συνοχή οδηγούν τον διαχωρισμό αυτό. Τα bots που χρησιμοποιούνται για την ιχνηλάτιση της κίνησης και της δραστηριότητας των χρηστών σε ιστότοπους, την εξόρυξη δεδομένων και την αυτοματοποίηση λειτουργιών, ταξινομούνται ως επί τω πλείστων ως «καλά» ή «ηθικά» bots, αφού επιφέρουν μείωση δαπανών σε

υπολογιστικούς πόρους, βελτίωση της εμπειρίας του χρήστη και επιτάχυνση διαδικασιών. Χαρακτηριστικό παράδειγμα αποτελεί το chatbot της OpenAI, ChatGPT, όπου στην δημόσια σφαίρα συναντάται σύγκρουση απόψεων, αδύνατη να σταθμίσει το όφελος και τον κίνδυνο που επιφέρει αυτό το bot για τον άνθρωπο.

Η έντονη ανησυχία που εγείρεται γύρω από τον ρόλο των bots στα κοινωνικά δίκτυα, όπως το Twitter, καθώς και η συστηματική προσπάθεια για την ανίχνευση και τον περιορισμό τους, πηγάζει από την χρήση της πλατφόρμας ως ελεύθερο βήμα στον δημόσιο και πολιτικό διάλογο, αλλά και ως μέσο πρόσβασης στην ανεξάρτητη ενημέρωση γύρω από την επικαιρότητα. Τα bots διαβρώνουν την εμπιστοσύνη των ανθρώπων στις πλατφόρμες κοινωνικής δικτύωσης και στην ψηφιακή επικοινωνία, και καλλιεργούν κλίμα αμφισβήτησης για τους σκοπούς που εξυπηρετούν, με αποτέλεσμα το περιεχόμενο της πλατφόρμας να αντιμετωπίζεται με επιφυλακτικότητα από τους χρήστες. Ως συνέπεια, υπονομεύεται η αξία του δημόσιου διαλόγου και περιορίζονται οι ουσιαστικές κοινωνικές αλληλεπιδράσεις μεταξύ των χρηστών του Twitter.

Ο αντίκτυπος της δράσης των bots στα κοινωνικά δίκτυα, όπως αναφέρθηκε και παραπάνω, έχει ευρύτερες επιπτώσεις στην κοινωνία, την οικονομία και την πολιτική. Στην πολιτική, bots εργαλειοποιούνται από υποψηφίους σε προεκλογικές περιόδους για την διασπορά ψευδών ειδήσεων, την προπαγάνδα και την επιρροή της ψήφου των πολιτών, υπονομεύοντας εν γένει τις δημοκρατικές διαδικασίες. Σε περιόδους κρίσεων, όπως αυτή της υγειονομικής κρίσης του COVID-19, τα bots χρησιμοποιήθηκαν για την καλλιέργεια κλίματος φόβου και αμφισβήτησης της επιστημονικής κοινότητας, διαδίδοντας ψευδείς ειδήσεις και συνωμοσίες αναφορικά με τις επιπτώσεις του εμβολιασμού στην δημόσια υγεία και την πηγή προέλευσης του ιού, θέτοντας σε κίνδυνο ανθρώπινες ζωές. Στην οικονομία, bots χρησιμοποιούνται σε αγορές υψηλής μεταβλητότητας, όπως αυτή των κρυπτονομισμάτων, για την χειραγώγηση της αγοράς μέσω της καλλιέργειας ευνοϊκού ή δυσμενούς επενδυτικού κλίματος, αποσκοπώντας σε κέρδη μέσα από βραχυπρόθεσμες αγοροπωλησίες.

Δεδομένων αυτών των σημαντικών επιπτώσεων, κρίνεται ως ζωτικής σημασίας η κατανόηση του ρόλου και του αντίκτυπου των bots στα κοινωνικά δίκτυα, αλλά και η ανάπτυξη αποτελεσματικών στρατηγικών και εργαλείων για τον εντοπισμό και την απομόνωσή τους, μετριάζοντας τις κοινωνικές, οικονομικές και πολιτικές επιπτώσεις. Βέβαια, πρόκειται για μια πρόκληση που απαιτεί πολύπλευρη αντιμετώπιση και πλέγμα δράσεων, και όχι απλώς τεχνολογικές λύσεις, τονίζοντας την ανάγκη για ρυθμιστικά μέτρα, νομοθετική θωράκιση και εκπαίδευση των χρηστών.

Στην παρούσα εργασία, μελετάται το πρόβλημα των bots στην πλατφόρμα του Twitter, αναπτύσσοντας μοντέλα επιτηρούμενης μηχανικής μάθησης και αξιολογώντας την προβλεπτική ικανότητά τους στον εντοπισμό των bots. Επιπλέον, ερμηνεύονται τα αποτελέσματα των παραπάνω μοντέλων, εξετάζοντας τον βαθμό επίδρασης των επιμέρους γνωρισμάτων στην διάκριση του τύπου χρήστη του Twitter, που καθιστά εφικτή ακόμα και την εμπειρική αναγνώριση τέτοιων λογαριασμών.

## 1.2. Κατηγορίες των bots στα Μέσα Κοινωνικής Δικτύωσης

Κατά βάση, τα bots κατηγοριοποιούνται με βάση τον αντικειμενικό σκοπό τους, τα συμπεριφορικά χαρακτηριστικά τους και το επίπεδο πολυπλοκότητάς τους. Σε πολλές περιπτώσεις, η δυναμική φύση των bots και τα μεταβαλλόμενα μοτίβα συμπεριφοράς που υιοθετούν, καθιστούν δυσδιάκριτα τα όρια μεταξύ αυτών των κατηγοριών ή οδηγούν ακόμα και σε επικαλυπτόμενα σύνολα. Στην ενότητα αυτή, παρουσιάζονται τέσσερεις σημαντικές κατηγορίες bot που συναντώνται στην διεθνή βιβλιογραφία, και συγκεκριμένα, τα traditional spambots, τα social spambots, τα political bots και τα social bots. Τα βασικά χαρακτηριστικά κάθε επιμέρους κατηγορίας συνοψίζονται στον Πίνακας 1.

| **Χαρακτηριστικό Γνώρισμα** | **Spambots** | | **Bots** | |
|---|---|---|---|---|
| | **Traditional** | **Social** | **Political** | **Social** |
| Μίμηση ανθρώπινης συμπεριφοράς | ΟΧΙ | ΝΑΙ | ΝΑΙ | ΝΑΙ |
| Αλληλεπίδραση με άλλους χρήστες | ΟΧΙ | ΝΑΙ | ΝΑΙ | ΝΑΙ |
| Δημοσίευση πρωτότυπου περιεχομένου | ΟΧΙ | ΝΑΙ | ΝΑΙ | ΝΑΙ |
| Διάδοση spam περιεχομένου | ΝΑΙ | ΝΑΙ | ΟΧΙ | ΟΧΙ |
| Επιρροή την κοινής γνώμης | ΟΧΙ | ΟΧΙ | ΝΑΙ | ΝΑΙ |
| Δράση σε συνεργατικά δίκτυα | ΟΧΙ | ΟΧΙ | ΝΑΙ | ΝΑΙ |
| Υψηλό επίπεδο δυσκολίας εντοπισμού | ΟΧΙ | ΝΑΙ | ΝΑΙ | ΝΑΙ |

Πίνακας 1 – Χαρακτηριστικά γνωρίσματα ανά κατηγορία bot

Τα traditional spambots ανήκουν στην ευρύτερη οικογένεια των spambots, αντιπροσωπεύοντας bots που αναπαράγουν μαζικά ανεπιθύμητο περιεχόμενο στο Διαδίκτυο και στα μέσα κοινωνικής δικτύωσης. Συνήθως, εξυπηρετούν διαφημιστικούς σκοπούς, κατευθύνοντας στοχευμένα ροές χρηστών προς συγκεκριμένους ιστότοπους. Επίσης, επεκτείνουν την δράση τους στην μετάδοση κακόβουλου λογισμικού και στην υποκλοπή προσωπικών δεδομένων χρηστών. Τα traditional spambots λειτουργούν με επαναληπτικά και προβλέψιμα μοτίβα συμπεριφοράς, χωρίς να εμφανίζουν

ουσιαστικές αλληλεπιδράσεις με περιεχόμενο τρίτων χρηστών, καθιστώντας συνήθως εύκολη την ανίχνευσή τους. Έτσι, η απουσία αντιδράσεων, σχολιασμού και κοινοποιήσεων περιεχομένου από έναν λογαριασμό στα μέσα κοινωνικής δικτύωσης αποτελεί ισχυρή ένδειξη πως πρόκειται traditional spambot. Άλλα κοινά συμπεριφορικά γνωρίσματα μεταξύ τέτοιων bots είναι η μαζική και αυτοματοποιημένη αποστολή άμεσων μηνυμάτων ή μηνυμάτων ηλεκτρονικού ταχυδρομείου με το ίδιο περιεχόμενο, αλλά και η κοινοποίηση υπερσυνδέσμων που οδηγούν σε κακόβουλους ιστότοπους. Αν και ο εντοπισμός τους μπορεί να προσεγγιστεί αποδοτικά ακόμα και μέσω φίλτρων βασισμένων σε λογικούς κανόνες λόγω των ισχυρά επαναληπτικών μοτίβων, συχνά εφαρμόζονται μέθοδοι μηχανικής για την αυτοματοποιημένη αναπροσαρμογή και βελτίωση των κανόνων αυτών. Τα βασικά γνωρίσματα που εξετάζονται σε μεθόδους μηχανικής μάθησης είναι η υψηλή συχνότητα ανάρτησης περιεχομένου, ο χαμηλός ρυθμός αλληλεπίδρασης με άλλους χρήστες, ο χαμηλός λόγος ακόλουθων προς ακολούθων, η υψηλή ομοιογένεια στο περιεχόμενο και η υψηλή συχνότητα εμφάνισης υπερσυνδέσμων στα tweets.

Τα social spambots αποτελούν την προηγμένη και ευφυή εκδοχή των traditional spambots. Επίκεντρο της φιλοσοφίας τους αποτελεί η απατηλή αποτύπωση ανθρώπινων μοτίβων συμπεριφοράς, μιμούμενα την συμπεριφορά πραγματικών χρηστών στα μέσα κοινωνικής δικτύωσης ώστε να δυσχεραίνεται ο εντοπισμός τους. Έτσι, προγραμματίζονται κατά τρόπο τέτοιον που προσομοιάζει σε αληθοφανείς οντότητες, τους αποδίδονται συγκεκριμένα δημογραφικά χαρακτηριστικά, ενώ σε ορισμένες περιπτώσεις αποκτούν προσωπικότητα και συναισθήματα (S. Giorgi, 2021). Ταυτόχρονα, τα social spambots αλληλεπιδρούν ουσιαστικά με άλλους χρήστες στα κοινωνικά δίκτυα και αναρτούν περιεχόμενο παρόμοιο με εκείνο των ανθρώπων, χειριζόμενα δεξιοτεχνικά ακόμα και την φυσική γλώσσα. Σε πρόσφατη μελέτη, υπολογίστηκε πως τα social spambots ενδέχεται να αντιπροσωπεύουν εώς και το 15% της συνολικής ψηφιακής κίνησης στο Διαδίκτυο, υπογραμμίζοντας τον σημαντικό αντίκτυπο τέτοιων χρηστών στον δημόσιο διάλογο και την ανάγκη λήψης μέτρων για τον περιορισμό τους (Ferrara, 2022). Για τον εντοπισμό των social spambots χρησιμοποιούνται αποκλειστικά μοντέλα μηχανικής μάθησης, ειδικά εκείνα που προωθούν την ερμηνευσιμότητα των αποτελεσμάτων, περιγράφοντας τον βαθμό επίδρασης των επιμέρους γνωρισμάτων στο αποτέλεσμα της πρόβλεψης. Αν και η ταύτιση των συμπεριφορικών χαρακτηριστικών μεταξύ social spambots και πραγματικών χρηστών δυσχεραίνει την απομόνωση ενδεικτικών γνωρισμάτων της δράσης τους, συνήθως χαρακτηρίζονται από υψηλή αναλογία κοινοποιήσεων και

ονομαστικών αναφορών στο περιεχόμενό τους. Παράλληλα, τείνουν να ακολουθούν την χρονική κατανομή της ανθρώπινης δραστηριότητάς στα μέσα κοινωνικής δικτύωσης, παύοντας την λειτουργία τους σε χρονικά παράθυρα χαμηλής έντασης, όπως οι τυπικές ώρες γραφείου και ανάπαυσης.

Τα political bots αποτελούν μια κατηγορία bots με αντικειμενικό σκοπό την επιρροή και την διαμόρφωση της κοινής γνώμης γύρω από πολιτικά ζητήματα, εκλογικές αναμετρήσεις και πρόσωπα της πολιτικής σκηνής. Συνήθως συνιστούν κόμβους ενός ευρύτερου δικτύου από bots που συντονίζονται και συνεργάζονται για να ενισχύσουν τον αντίκτυπο και την δυναμική τους πάνω σε πολιτικά ζητήματα. Η πρώτη εμφάνιση των political bots έγινε στις Προεδρικές εκλογές των ΗΠΑ του 2016, όπως αναφέρθηκε και προηγουμένως, όπου σύμφωνα με μελέτη του Oxford Internet Institute, το 20% των πολιτικών μηνυμάτων στα μέσα κοινωνικής δικτύωσης κατά την προεκλογική περίοδο προέρχονταν από political spambots. Έτσι, υπογραμμίστηκε η ανάγκη ανίχνευσης και περιορισμού των bots από τις πλατφόρμες κοινωνικής δικτύωσης, αποσκοπώντας στον περιορισμό της πολιτικής προπαγάνδας, της επιρροής στο δημόσιο διάλογο και της διαμόρφωσης της κοινής γνώμης, και συνολικότερα, στην υπονόμευση των δημοκρατικών διαδικασιών. Τα χαρακτηριστικά των political spambots συνοψίζονται στην ανάρτηση και κοινοποίηση περιεχομένου με συγκεκριμένο πολιτικό χρωματισμό και ιδεολογικό πρόσημο, στην ενσωμάτωση hashtags σχετιζόμενων με πολιτικά συνθήματα και υπερσυνδέσμων σχετιζόμενων με ψευδείς ειδήσεις. Επίσης, τα political spambots παρουσιάζουν συντονισμένη δραστηριότητα γεωγραφικά και χρονικά, συνιστώντας μέρη ενός ευρύτερου δικτύου.

Τα social bots αποτελούν την πλέον προηγμένη μορφή bot λογαριασμών, διαφοροποιούμενη από την εκδοχή των social spambots ως προς το επίκεντρο της δραστηριότητάς τους, αφού δεν διαμοιράζουν μαζικά ανεπιθύμητο περιεχόμενο. Παρουσιάζουν έντονη αλληλεπίδραση με πραγματικούς χρήστες κατά τρόπο που προσομοιάζει με εκείνη των ανθρώπων και υιοθετούν ανθρώπινα συμπεριφορικά γνωρίσματα. Σε πρόσφατη μελέτη, εκτιμήθηκε το ποσοστό των social bots στο Twitter ενδέχεται να αντιπροσωπεύει ακόμα το 15% των συνολικών λογαριασμών στην πλατφόρμα, ποσοστό που μεταφράζεται σε σχεδόν 48 εκατομμύρια λογαριασμούς (A. Rauchfleisch, 2020). Η διάκριση των social bots από τους πραγματικούς χρήστες της πλατφόρμας αποτελεί σύνθετη διαδικασία, η οποία προσεγγίζεται κατά βάση μέσω μεθόδων μηχανικής μάθησης, ενώ σε πολλές περιπτώσεις απαιτεί συνδυαστική χειροκίνητη εξέταση και ταυτοποίηση από τον άνθρωπο. Έτσι, δεν μπορεί να προτυποποιηθεί ένα σύνολο γνωρισμάτων που χαρακτηρίζουν τα social bots.

## 1.2. Χαρακτηριστικά των Bots στα Μέσα Κοινωνικής Δικτύωσης

Τα χαρακτηριστικά των bots στα μέσα κοινωνικής δικτύωσης ποικίλουν με βάση το επίπεδο ευφυίας τους και τον αντικειμενικό σκοπό που επιτελούν. Όπως τονίστηκε και στην προηγούμενη ενότητα, σε αρκετές περιπτώσεις είναι αδύνατη η ταξινόμηση συμπεριφορών ως ενδεικτικές των bots, αφού έχουν την τάση να αφομοιώνουν και να προσαρμόζονται στα ανθρώπινα μοτίβα συμπεριφοράς. Στην ενότητα αυτή, καταγράφονται τρείς ομάδες χαρακτηριστικών των bots στα μέσα κοινωνικής δικτύωσης, οι οποίες αξιοποιούνται για τον εντοπισμό τους μέσω μεθόδων μηχανικής μάθησης.

Τα χαρακτηριστικά περιεχομένου αναφέρονται ιδιότητες που εξάγονται άμεσα ή έμμεσα από το ίδιο το περιεχόμενο των tweets. Η συχνότητα ανάρτησης και κοινοποίησης περιεχομένου, το πλήθος των hashtags, των υπερσυνδέσμων και των ονομαστικών αναφορών στα tweets, και η ομοιογένεια της πληροφορίας αποτελούν τα κύρια γνωρίσματα αυτής της κατηγορίας. Επίσης, σημαντική συνεισφορά έχουν και παράγωγα γνωρίσματα που προκύπτουν από την συναισθηματική ανάλυση του περιεχομένου των tweets, εξορύσσοντας πληροφορία σχετικής με την κατεύθυνση και την ένταση του συναισθήματος στο περιεχόμενο. Τυπικά, μεταξύ των bots παρατηρείται υψηλότερη συχνότητα προσάρτησης hashtags και υπερσυνδέσμων στα tweets συγκριτικά με την αντίστοιχη των πραγματικών χρηστών, ενώ το περιεχόμενό τους αντανακλά μία σταθερά φορτισμένη συναισθηματικά κατάσταση, συνήθως υπεραισιόδοξη, χωρίς έντονες διακυμάνσεις και μεταβολές.

Τα χαρακτηριστικά χρήστη αναφέρονται σε ιδιότητες που συνδέονται με την οντότητα που διαχειρίζεται τον λογαριασμό. Η ηλικία του λογαριασμού, το πλήθος των ακόλουθων και ακολούθων, καθώς και το ρυθμό μεταβολής της αναλογίας αυτής στον άξονα του χρόνου αποτελούν τα κύρια γνωρίσματα αυτής της κατηγορίας. Κατά βάση, τα bots τείνουν να εμφανίζουν έντονες αυξομειώσεις στο πλήθος των λογαριασμών που ακολουθούν και ακολουθούνται, ενώ στους πραγματικούς χρήστες παρατηρείται σταθερή αναλογία. Επίσης, στην κατηγορία αυτή περιλαμβάνονται και χαρακτηριστικά που σχετίζονται με την χρήση της προκαθορισμένης φωτογραφίας προφίλ και εξωφύλλου από τον λογαριασμό.

Τα χαρακτηριστικά δικτύου αναφέρονται σε ιδιότητες που σχετίζονται με το δίκτυο συνδέσεων και αλληλεπιδράσεων μιας ομάδας χρηστών στα μέσα κοινωνικής δικτύωσης. Το πλήθος των αμοιβαίων ακολούθων μεταξύ μιας ομάδας λογαριασμών αποτελούν το κυρίαρχο γνώρισμα σε αυτή την κατηγορία. Επίσης, οι τιμές των

συντελεστών κεντρικότητας και ομαδοποίησης που προκύπτουν από την θεωρία γράφων αξιοποιούνται ως παράγωγα χαρακτηριστικά δικτύου. Τα bots τείνουν να δημιουργούν κοινότητες αμοιβαία ακολουθούμενων λογαριασμών συνδυαστικά με ένα μικρό πλήθος ακολούθων, παρουσιάζοντας χαμηλό συντελεστή ομαδοποίησης, γεγονός που υποδηλώνει πως δεν σχηματίζουν στενά συνδεδεμένες κοινότητες με τρίτους λογαριασμούς στα μέσα κοινωνικής δικτύωσης. Τέλος, ο συντελεστής κεντρικότητας του λογαριασμού στο δίκτυο αντανακλά την διασύνδεση ενός λογαριασμού με γνωστά bots, μιας και τα bots ακολουθούν μια κεντρικοποιημένη δομή στο δίκτυο.

Βέβαια, αξίζει να σημειωθεί πως οι συμπεριφορές και τα χαρακτηριστικά των bots μεταβάλλονται διαρκώς και εξελίσσονται με σκοπό την ενσωμάτωση εκείνων που προσομοιάζουν σε πραγματικούς χρήστες. Στην ουσία, από τα bots αξιοποιούνται οι ίδιες τεχνικές που εφαρμόζονται για τον εντοπισμό τους, όπως οι μέθοδοι μηχανικής μάθησης, ώστε να εντοπίσουν και να αφομοιώσουν νέα συμπεριφορικά και συναισθηματικά μοτίβα από πραγματικούς χρήστες στα μέσα κοινωνικής δικτύωσης ώστε να ενισχύσουν την συγκεκαλλυμένη ανθρώπινη υπόστασή τους.

## 1.4. Δομή Εργασίας

Η παρούσα Διπλωματική Εργασία χωρίζεται σε πέντε (5) κεφάλαια. Στο Κεφάλαιο 2, παρουσιάζεται μια εκτεταμένη βιβλιογραφική ανασκόπηση γύρω από την δράση των bots στα μέσα κοινωνικής δικτύωσης με επίκεντρο την πλατφόρμα του Twitter, εστιάζοντας σε ζητήματα δημόσιας υγείας, πολιτικής και οικονομίας με επίκεντρο το Twitter, ενώ καταγράφονται και προσεγγίσεις εντοπισμού των bots μέσω μεθόδων μηχανικής μάθησης. Στο Κεφάλαιο 3, παρουσιάζεται το μεθοδολογικό πλαίσιο που ακολουθήθηκε για τον εντοπισμό bots στη μελέτη περίπτωσης, παρουσιάζοντας τα επιμέρους βήματα της επίλυσης προβλημάτων μηχανικής μάθησης και την αρχιτεκτονική δημοφιλών ταξινομητών. Στο Κεφάλαιο 4, περιγράφονται τα στάδια συλλογής, προεπεξεργασίας και διερευνητικής ανάλυσης των δεδομένων της μελέτης περίπτωσης, το στάδιο της ρύθμισης των υπερπαραμέτρων των μοντέλων και της προσαρμογής τους στα δεδομένα εκπαίδευσης, καθώς και το στάδιο της αξιολόγησης των αποτελεσμάτων. Στο Κεφάλαιο 5, προτείνονται μελλοντικές επεκτάσεις της έρευνας.

# 2. Βιβλιογραφική Ανασκόπηση

Στο Κεφάλαιο 2, αποτυπώνεται ο αντίκτυπος των bots στην δημόσια υγεία, στην πολιτική και στην οικονομία με επίκεντρο το Twitter, μέσα από μια συστηματική μελέτη της βιβλιογραφίας. Αναδεικνύεται η εργαλειοποίηση των bots για την διάδοση ψευδών ειδήσεων και παραπληροφόρησης κατά την παγκόσμια υγειονομική κρίση του Covid-19, την προώθηση πολιτικών συμφερόντων και προπαγανδιστικού λόγου, καθώς και την χειραγώγηση του επενδυτικού κλίματος ευμετάβλητων αγορών, όπως αυτή των κρυπτονομισμάτων. Παράλληλα, καταγράφονται προσεγγίσεις εντοπισμού bots στα μέσα κοινωνικής δικτύωσης με μεθόδους μηχανικής μάθησης, αλλά και αποκρυπτογράφησης κοινών μοτίβων συμπεριφοράς και γνωρισμάτων.

## 2.1. Το Αποτύπωμα των Twitter Bots στην Δημόσια Υγεία

Οι Ahmed κ.α. (W. Ahmed, 2022) μελέτησαν την διακίνηση ψευδών ειδήσεων και θεωριών συνωμοσίας από bots στο Twitter την περίοδο της υγειονομικής κρίσης του Covid-19. Εστιάζοντας στην συνωμοσία αναφορικά με την παράλληλη εξάπλωση της πανδημίας και των τηλεπικοινωνιακών δικτύων 5G, συνέλλεξαν και ανέλυσαν δείγμα 233 tweets από λογαριασμούς που εντοπίζονται γεωγραφικά στο Ηνωμένο Βασίλειο, κατανεμημένα σε χρονικό διάστημα 7 ημερών. Η ανάλυση ανέδειξε πως το 34.8% των tweets που υιοθετούσαν την συγκεκριμένη θεωρία συνωμοσίας συνιστούσαν κόμβους ενός διασυνδεδεμένου δικτύου, με την πλειοψηφία των χρηστών να έχει μηδενική συνεισφορά σε περιεχόμενο και αλληλεπιδράσεις με τρίτους χρήστες. Έτσι, αναδείχθηκε η ενορχηστρωμένη και συνεργατική δράση των bots στα κοινωνικά δίκτυα, λειτουργώντας ως κόμβοι δικτύων προώθησης συγκεκριμένων συμφερόντων και ιδεών, αλλά και της προσπάθειας χειραγώγησης της κοινής γνώμης. Οι Z. Wang και A. Lin (Z. Weng, 2022) μελέτησαν επίσης την διακίνηση θεωριών συνωμοσίας και παραπληροφόρησης αναφορικά με την παραγωγή της μετάλλαξης του SARS-Cov-2 στα εργαστήρια της Wuhan, αποτελώντας μέρος ενός ευρύτερου σχεδίου της Κίνας για παγκόσμια αποσταθεροποίηση. Από την συλλογή και την ανάλυση δείγματος 120,118 σχετικών tweets, τα 35,945 ταξινομήθηκαν ως προερχόμενα από bots, αντιστοιχώντας σε ποσοστό 29% επί του συνολικού δείγματος, υπογραμμίζοντας τους κινδύνους που εγκυμονούνται για την δημόσια υγεία από την δράση των bots στο Twitter.

Οι Zhang κ.α. (Zhang Y, 2023) εξέτασαν την επίδραση των bots στη σφαίρα του δημόσιου διαλόγου στο Twitter αναφορικά με την υγειονομική κρίση του COVID-19. Αξιοποιώντας ένα δείγμα 56,879 μοναδικών χρηστών του Twitter που συμμετείχαν σε

σχετικές συζητήσεις στην πλατφόρμα, εργάστηκαν στην συναισθηματική ανάλυση του δείγματος. Τα αποτελέσματα έδειξαν κοινή και ισορροπημένη κατανομή της πόλωσης του περιεχομένου μεταξύ των tweets που προέρχονται από bots και πραγματικούς χρήστες, ενώ σε περιόδους έντονα αρνητικής πόλωσης, τα bots αναπροσάρμοζαν την συμπεριφορά τους ώστε να μιμείται εκείνη των πραγματικών χρηστών. Η ανάλυση πρόθεσης στο περιεχόμενο των tweets ανέδειξε τον διαφορετικό προσανατολισμό του περιεχομένου των tweets μεταξύ bots και πραγματικών χρηστών. Τα bots εμφάνιζαν ροπή προς την λέξη κλειδί «νέα», εξυπηρετώντας σκοπούς παραπληροφόρησης και διάδοσης ψευδών ειδήσεων, σε αντίθεση με τους πραγματικούς χρήστες που εμφάνιζαν ροπή προς την λέξη κλειδί «υγεία», απόρροια της ενημέρωσης στην κοινότητα για την κατάσταση της υγείας τους. Μια ακόμα πτυχή που ανέδειξε η μελέτη είναι η τάση ανακοινοποίησης περιεχομένου προερχόμενο από bots ως αποτέλεσμα των συνεργατικών κοινωνικών δικτύων που αναπτύσσουν, αλλά και της δυναμικής που έχουν στον επηρεασμό της κοινής γνώμης. Αντίθετα, τα tweets των πραγματικών χρηστών παρουσιάζονται αυξημένες αντιδράσεις και απαντήσεις, ωστόσο το μέσο πλήθος ανακοινοποιήσεων παραμένει χαμηλότερο. Αν και κατά βάση τα social bots εμφανίζουν χαμηλό αριθμό ακολούθων και ακόλουθων, καταγράφηκαν περιπτώσεις bots με τεράστιο πλήθος ακολούθων, οδηγώντας πλασματικά σε μέσο όρο υψηλότερο κατά 166 φορές συγκριτικά με εκείνον των πραγματικών χρηστών. Εν κατακλείδι, τα συμπεράσματα της μελέτης ανέδειξαν πως το 22% των tweets αναφορικά με την πανδημία του Covid-19 προέρχονταν από bots.

## 2.2. Το Αποτύπωμα των Bots στην Πολιτική

Οι Zhuravskaya κ.α. (E. Zhuravskaya, 2020) μελέτησαν τον αντίκτυπο των bots του Διαδικτύου και των μέσων κοινωνικής δικτύωσης σε διάφορες εκφάνσεις της πολιτικής ζωής, όπως στην επιρροή των ψηφοφόρων σε εκλογικές αναμετρήσεις, στην καλλιέργεια πολιτικής πόλωσης, στην επικοινωνία διαδηλώσεων και πολιτικών διαμαρτυριών, καθώς και στην διάδοση αντικυβερνητικών συνθημάτων. Ερευνήθηκαν τα αίτια πίσω από την λογοκρισία που επιβάλλεται στα μέσα κοινωνικής δικτύωσης από αυταρχικά καθεστώτα, αλλά και η εργαλειοποίηση των μέσων αυτών προς όφελος της πολιτικής προπαγάνδας και της παραπληροφόρησης. Σύμφωνα με τα ευρήματα της μελέτης, τα μέσα κοινωνικής δικτύωσης συνέβαλλαν καθοριστικά στην εξάπλωση των κινημάτων λαϊκισμού στην Ευρώπη, ενώ βοηθούν στον συντονισμό και την συσπείρωση κινημάτων διαμαρτυρίας σε αυταρχικά καθεστώτα, όπου τα παραδοσιακά μέσα ενημέρωσης βρίσκονται υπό πλήρη κυβερνητικό έλεγχο.

Οι Faveri κ.α. (F. L. De Faveri, 2023) μελέτησαν την επιρροή των bots του Twitter στη συζήτηση γύρω από τον πόλεμο της Ρωσίας στην Ουκρανία κατά τη διάρκεια των ιταλικών εκλογών του 2022. Διαπιστώθηκε πως ένα σημαντικό ποσοστό των bots λειτούργησε ως μέσο εξυπηρέτησης πολιτικών συμφερόντων συγκεκριμένων υποψηφίων με στόχο την επιρροή και την χειραγώγηση της κοινής γνώμης. Η κατανομή των bots μεταξύ των προφίλ των υποψηφίων στην εκλογική αναμέτρηση ήταν σταθερή, υποδηλώνοντας την ευρεία εργαλειοποίηση τους. Όσον αφορά τα ερευνητικά αποτελέσματα, διαπιστώθηκε πως το 9.61% των tweets για τον N. Fratoianni προερχόταν από bots, με το αντίστοιχο ποσοστό για την G. Meloni να ανέρχεται σε 15.08%. Επίσης, η μελέτη ανέδειξε πως η εισβολή της Ρωσίας στην Ουκρανία αποτέλεσε σημαντικό άξονα της προεκλογικής εκστρατείας των υποψηφίων, κυρίως από το Δημοκρατικό Κόμμα.

## 2.3. Το Αποτύπωμα των Bots στις Αγορές

Από τους O. Kraaijeveld και J. De Smedt (O. Kraaijeveld, 2020) εξετάστηκε η διαμόρφωση του επενδυτικού παλμού στην αγορά των κρυπτονομισμάτων από περιεχόμενο που δημοσιεύεται στο Twitter, καθώς και η εγγενής ικανότητά του στην πρόβλεψη της τιμής τους. Η μελέτη εστιάστηκε στην μεταβολή της απόδοσης εννέα κορυφαίων κρυπτονομισμάτων, όπως του Bitcoin, του Ethereum και του Cardano, συναρτήσει της συναισθηματικής πόλωσης που προέρχεται από tweets, δηλαδή της χειραγώγησης της αγοράς μέσα από την καλλιέργεια θετικού ή αρνητικού κλίματος. Για το σκοπό αυτό, συγκεντρώθηκαν σχετικά δεδομένα 24 εκατομμυρίων αναρτήσεων προερχόμενες από σχεδόν 2 εκατομμύρια μοναδικούς χρήστες της πλατφόρμας, με το χαμηλότερο ποσοστό tweets προερχόμενο από bots να σημειώνεται για το Bitcoin. Τα αποτελέσματα οδήγησαν στο συμπέρασμα πως το αίσθημα της κοινής γνώμης κατευθύνει σε σημαντικό βαθμό τον ημερήσιο όγκο συναλλαγών του Bitcoin, σε αντίθεση με το Ethereum που φαίνεται να παραμένει ανεπηρέαστο. Το σημαντικότερο εύρημα αφορά το Litecoin, όπου παρατηρείται σημαντική στατιστικά συσχέτιση μεταξύ του όγκου των ημερήσιων συναλλαγών και των αναρτήσεων στην πλατφόρμα του Twitter.

Από τους L. Nizzoli κ.α. (L. Nizzoli, 2020) και χαρτογραφήθηκε το τοπίο της αγοράς κρυπτονομισμάτων με έμφαση σε κλειστές διαδικτυακές επενδυτικές κοινότητες που στεγάζονται στις πλατφόρμες του Discord και του Telegram, καθώς και η προσέλκυση χρηστών σε αυτές μέσω του Twitter. Συχνά, σε τέτοιες κοινότητες καλλιεργείται πρόσφορο έδαφος για οικονομικές απάτες, όπως σχήματα Ponzi ή χειραγώγηση των

τιμών των κρυπτονομισμάτων με την μέθοδο pump and down. Το δείγμα της μελέτης αφορούσε περισσότερα από 50 εκατομμύρια tweets που συνδέονται με προσκλήσεις προς τέτοιου τύπου κοινότητες, με τα ευρήματα της ανάλυσης μέσω γράφων να αναδεικνύουν πως το 75.4% της κίνησης προς αυτές οδηγείται από 15 συντονισμένα δίκτυα με bots. Παράλληλα, οι υπερσύνδεσμοι που δημοσιεύονται στο Twitter και οδηγούν προς το Telegram αφορούν σχήματα Ponzi με κρυπτονομίσματα σε ποσοστό της τάξης του 71.4%. Έτσι, υπερτονίζεται η ανάγκη περιορισμού της δράσης των bots και της νομοθετικής θωράκισης της αγοράς των κρυπτονομισμάτων.

Από τους Kirsch και Chowdhury (David A. Kirsch, 2023) αναδείχθηκε η εργαλειοποίηση των Twitter bots για την χειραγώγηση των χρηματιστηριακών αγορών, μελετώντας την περίπτωση της μετοχής της Tesla. Η ανάλυση περισσότερων από 9.2 εκατομμυρίων tweets στην πλατφόρμα του Twitter κατά το χρονικό διάστημα μεταξύ 2010 και 2020, τα οποία είχαν αναφορά στο αναγνωριστικό της μετοχής της Tesla στο αμερικάνικο χρηματιστήριο, εντόπισε σημαντική ποσόστωση σε bots. Σχεδόν το 23% των σχετικών tweets φέρεται να προέρχεται από bots, ποσοστό αυτό κυμαίνεται στα ίδια επίπεδα με το αντίστοιχο άλλων μεγάλων εταιριών, όπως η Apple και η Amazon. Η ειδοποιός διαφορά με τα bots της Tesla έγγυται στην προώθηση ενός συγκεκριμένου επιχειρηματικού αφηγήματος για την εταιρία, ως μια συντονισμένη προσπάθεια χειραγώγησης της τιμής της μετοχής της. Εκτιμάται πως τα bots αυτά χρησιμοποιούνται για την προστασία της εταιρίας απέναντι στην κριτική, την ενίσχυση της αξιοπιστίας προς το επενδυτικό κοινό, ακόμα και ως αντισταθμιστικός παράγοντας σε περιόδους πτωτικής τάσης της τιμής της μετοχής. Μάλιστα, ιχνηλατήθηκε η πορεία της μετοχής συνδυαστικά με την εισαγωγή νέων bot στην πλατφόρμα που προωθούσαν τα συμφέροντα της εταιρίας, σημειώνοντας μέση αύξηση στην τιμή της μετοχής κατά 1.4% σε διάστημα 7 ημερών.

## 2.4. Ανίχνευση Bots στο Twitter με Μεθόδους Μηχανικής Μάθησης

Η μηχανική μάθηση αποτελεί την κυρίαρχη προσέγγιση για την ανίχνευση bots στα κοινωνικά δίκτυα, αφού επιτρέπει την συνεχή βελτίωση των μοντέλων της μέσα από μια δυναμική διαδικασία εκπαίδευσης σε νέα σύνολα δεδομένων, αναπροσαρμόζοντας τις μεθόδους εντοπισμού των bots στα συνεχώς μεταβαλλόμενα χαρακτηριστικά και στρατηγικές τους. Επίσης, προσφέρουν την δυνατότητα ανίχνευσης μοτίβων συμπεριφοράς και χαρακτηριστικών γνωρισμάτων των bots, αντικαθιστώντας παραδοσιακές προσεγγίσεις στον εντοπισμό των bots μέσα από σύνολα προκαθορισμένων κανόνων. Η υψηλή αποδοτικότητα που παρουσιάζουν στον χειρισμό

σύνθετων δεδομένων μεγάλου όγκου, ισχυροποιούν ακόμα περισσότερο την θέση τους στο σύγχρονο ψηφιακό τοπίο της υπερπληθώρας πληροφορίας.

Από τους Yang κ.α. (O. Varol) προτάθηκε ένα μεθοδολογικό πλαίσιο εξαγωγής χαρακτηριστικών δεδομένων και μεταδεδομένων λογαριασμών του Twitter για την εκπαίδευση μοντέλων μηχανικής μάθησης για την ανίχνευση bots στην πλατφόρμα. Έτσι, δημιουργήθηκε ένα εξαντλητικό σύνολο δεδομένων χρηστών του Twitter, περιλαμβάνοντας περισσότερα από χίλια γνωρίσματα, τα οποία διακρίνονται σε έξι κατηγορίες: τα γνωρίσματα μεταδεδομένων, τα γνωρίσματα περιεχομένου, τα γνωρίσματα φιλίας, τα γνωρίσματα δικτύου, τα γνωρίσματα συναισθήματος και τα γνωρίσματα χρόνου. Επάνω στα δεδομένα αυτά εκπαιδεύτηκε ένας ταξινομητής Random Forest των εκατό εκτιμητών, παρουσιάζοντας ακρίβεια της τάξης των 0.95 AUC, αξιολογούμενη ως ιδιαίτερα υψηλή. Παράλληλα, με την μέθοδο επιλογής γνωρισμάτων του Random Forest απομονώθηκαν εκατό από αυτά με την μεγαλύτερη επίδραση στην πρόβλεψη, επάνω στα οποία προσαρμόστηκε ένα μοντέλο k-Means, ταξινομώντας λογαριασμούς βάσει ομοιογενών χαρακτηριστικών. Η βέλτιστη προσαρμογή της μεθόδου σημειώθηκε για δέκα κλάσεις, ενώ σε αρκετές από αυτές παρατηρήθηκε ομαδοποίηση bot και πραγματικών χρηστών.

Από τους Efthimion κ.α. (Phillip George Efthimion, 2018) μελετήθηκε η απόδοση αλγορίθμων επιβλεπόμενης μηχανικής μάθησης στην ανίχνευση traditional και social spambots. Πιο συγκεκριμένα, εκπαιδεύτηκαν μεμονωμένα μοντέλα Λογιστικής Παλινδρόμησης και Μηχανών Διανυσμάτων Στήριξης, ενώ αξιολογήθηκε και μία συνδυαστική προσέγγιση αυτών των μεθόδων. Το συνδυαστικό μοντέλο παρουσίασε ιδιαίτερα υψηλή απόδοση, ανιχνεύοντας ορθά το 96,25% και το 95,77% των συνολικών περιπτώσεων traditional spambots και social spambots, αντίστοιχα. Η ακρίβεια του συνδυαστικού μοντέλου ως προς τα αληθώς θετικά αποτελέσματα για κάθε κατηγορία spambot κυμάνθηκε σε ακόμα υψηλότερα επίπεδα, ανιχνεύοντας ορθά το 96,81% των traditional spambots και το 97,13% των social spambots. Τα γνωρίσματα που αναδείχθηκαν ως κυρίαρχα για τον εντοπισμό των Twitter bots μέσω της απόδοσης βαρών στο μοντέλο της λογιστικής παλινδρόμησης σχετίζονταν με την γεωγραφική τοποθεσία, το πλήθος των ακολούθων, την γλώσσα και την φωτογραφία προφίλ του χρήστη. Καταλήγοντας, λογαριασμοί με λιγότερους των 30 ακόλουθων που χρησιμοποιούν την προκαθορισμένη φωτογραφία προφίλ του Twitter, αποτελούν το πιο συχνά παρατηρούμενο μοτίβο bots στο Twitter.

Από τους M. Kantepe και M.C. Ganiz (M. Kantepe, 2017) υλοποιήθηκε ένα δομημένο μεθοδολογικό πλαίσιο εντοπισμού bot λογαριασμών στο Twitter με μεθόδους

επιτηρούμενης μηχανικής μάθησης. Για τον σκοπό αυτό, μελετήθηκε δείγμα από 3,040 χρήστες της πλατφόρμας, με την ποσόστωση 78% σε πραγματικούς χρήστες. Για κάθε λογαριασμό αξιοποιήθηκαν 62 χαρακτηριστικά, μεταξύ των οποίων γνωρίσματα χρήστη (π.χ. πλήθος ακολούθων, πλήθος tweets, επιβεβαιωμένο προφίλ), γνωρίσματα tweets (π.χ. συχνότητα χρήσης hashtag, ειδικών χαρακτήρων και URL, ένταση αλληλεπίδρασης) και περιοδικά γνωρίσματα (π.χ. γλώσσα, γεωγραφική τοποθεσία, φωτογραφία προφίλ). Παράλληλα, εφαρμόστηκαν αλγόριθμοι επεξεργασίας φυσικής γλώσσας για την εξόρυξη συναισθήματος από τα tweets, δηλαδή την ταξινόμησή τους ως θετικά, αρνητικά ή ουδέτερα φορτισμένα. Με την χρήστη των μετρικών του Information Gain (IG) και του Mutual Information (MI) ποσοτικοποιήθηκε η επίδραση κάθε γνωρίσματος στην πρόβλεψη για τον τύπο χρήστη, συγκλίνοντας σε κοινά αποτελέσματα. Το πλήθος των επαναλήψεων του hashtag με την μεγαλύτερη συχνότητα εμφάνισης μεταξύ των tweets αποτελεί το κυρίαρχο γνώρισμα των bots στο Twitter, επιβεβαιώνοντας την επαναληπτική φύση της συμπεριφοράς τους. Επιπλέον, σημαντική επίδραση έχουν και τα γνωρίσματα που αντιπροσωπεύουν το χρονικό διάστημα της μεγαλύτερης συνεδρίας, το χρονικό διάστημα ανάρτησης συνεχόμενων tweets στην ίδια συνεδρία και την εντροπία της πληροφορίας των tweets. Πάνω στο σύνολο δεδομένων εκπαιδεύτηκαν τέσσερεις ταξινομητές επιτηρούμενης μηχανικής μάθησης, αυτοί της Λογιστικής Παλινδρόμησης, των Μηχανών Διανυσμάτων Στήριξης, του Naïve Bayes και των Gradient Boost Trees. Τα Gradient Boost Trees παρουσίασαν την υψηλότερη απόδοση μεταξύ των τεσσάρων αυτών μοντέλων, επιτυγχάνοντας ακρίβεια πρόβλεψης κοντά στο 86%, υπογραμμίζοντας την δυναμική των μοντέλων δέντρων απόφασης στην ανίχνευση bots στο Twitter. Τα μοντέλα της Λογιστικής Παλινδρόμησης, των Μηχανών Διανυσμάτων Στήριξης και του Naïve Bayes σημείωσαν ακρίβεια 75%, 78% και 82%, αντίστοιχα.

# 3. Μεθοδολογία

Στο Κεφάλαιο 3, παρουσιάζεται το μεθοδολογικό πλαίσιο που ακολουθήθηκε για την εκπόνηση της μελέτης περίπτωσης, περιγράφοντας μια σειρά από διακριτά και αυστηρά καθορισμένα στάδια επίλυσης του προβλήματος της ταξινόμησης χρηστών του Twitter. Επιπλέον, παρουσιάζεται το θεωρητικό υπόβαθρο των ταξινομητών μηχανικής μάθησης που χρησιμοποιήθηκαν για την πρόβλεψη της κατηγορίας ενός χρήστη του Twitter κατά την μελέτη περίπτωσης.

## 3.1. Στάδια στην Ανάλυση Δεδομένων και στην Μηχανική Μάθηση

### 3.1.1. Συλλογή Δεδομένων

Το πρώτο στάδιο επίλυσης προβλημάτων ανάλυσης δεδομένων και μηχανικής μάθησης είναι η συλλογή των δεδομένων. Περιλαμβάνει μια σειρά από επιμέρους συνιστώσες, όπως συνοψίζονται στα παρακάτω σημεία, που διασφαλίζουν την ακρίβεια, την αξιοπιστία και την ακεραιότητα των δεδομένων.

- Προσδιορισμός των μεταβλητών ενδιαφέροντος μέσω της αναζήτησης και της καταγραφής ποσοτικών και ποιοτικών χαρακτηριστικών που επιδρούν στη μεταβλητή στόχο του προβλήματος.
- Επιλογή αξιόπιστων πηγών δεδομένων, όπως αποθετήρια δεδομένων, βάσεις δεδομένων, δημόσια και κυβερνητικά έγγραφα, ερωτηματολόγια, συνεντεύξεις και πειράματα.
- Σχεδιασμός μέσων συλλογής δεδομένων, όπως ανάπτυξη τμημάτων κώδικα απόξεσης ιστού για την εξαγωγή πληροφορίας μέσα από το υπερκείμενο των ιστότοπων και εκτέλεσης HTTP ερωτημάτων για την πρόσβαση σε δεδομένα που προσφέρονται μέσω προγραμματιστικών διεπαφών (APIs).
- Καθορισμός στρατηγικής δειγματοληψίας για την εκλογή ενός υποσυνόλου του σετ δεδομένων, αντιπροσωπευτικό του πληθυσμού, αντισταθμίζοντας πιθανές ανισόρροπες κατανομές ως προς την μεταβλητή στόχο και μεγάλους όγκους δεδομένων.
- Καταγραφή δεδομένων σε δομημένη μορφή, όπως δομές πινάκων, δομές αντικειμένων και υπολογιστικά φύλλα, διευκολύνοντας τόσο την ανάλυση των δεδομένων μέσω λογισμικού και τμημάτων κώδικα όσο και την κατανόησή τους από τον άνθρωπο.

- Αντιμετώπιση νομικών και ηθικών ζητημάτων, διασφαλίζοντας πως τα δεδομένα που συλλέγονται δεν παραβιάζουν πολιτικές απορρήτου και άδειες χρήσης, ενώ ταυτόχρονα εναρμονίζονται με την κείμενη νομοθεσία.

### 3.1.2. Προεπεξεργασία Δεδομένων

Η δεύτερο στάδιο της επίλυσης ενός προβλήματος ανάλυσης δεδομένων και μηχανικής μάθησης είναι η προεπεξεργασία των δεδομένων. Σε αυτό το στάδιο, τα ακατέργαστα δεδομένα που συλλέχθηκαν στο προηγουμένως προετοιμάζονται και μετασχηματίζονται σε κατάλληλη μορφή για την είσοδό τους σε μοντέλα μηχανικής μάθησης. Οι βασικές συνιστώσες της προεπεξεργασίας δεδομένων συνοψίζονται στα παρακάτω σημεία.

- Καθαρισμός των δεδομένων από κενές και ακραίες τιμές, δηλαδή απόρριψή τους από το σετ δεδομένων. Υπό προϋποθέσεις, οι κενές τιμές συμπληρώνονται είτε με την χρήση κάποιου περιγραφικού στατιστικού μέτρου ή με την χρήση κάποιας προηγούμενης / επόμενης τιμής που παρουσιάζει υψηλό βαθμό αυτοσυσχέτισης με την ελλείπουσα.
- Κωδικοποίηση δεδομένων σε κατάλληλη μορφή για είσοδό τους σε μοντέλα μηχανικής μάθησης. Στα ποιοτικά χαρακτηριστικά εφαρμόζονται τεχνικές μετασχηματισμού των κατηγορηματικών μεταβλητών σε αντιπροσωπευτικές αριθμητικές αναπαραστάσεις, όπως η κωδικοποίηση ενός σημείου, η κωδικοποίηση ετικέτας και η κωδικοποίηση κατά σειρά.
- Κανονικοποίηση και τυποποίηση των ποσοτικών μεταβλητών για την αναγωγή των δεδομένων σε κοινή κλίμακα ή στο επιτρεπτό εύρος τιμών ενός μοντέλου μηχανικής μάθησης. Τυπικά, στα Νευρωνικά Δίκτυα εφαρμόζεται κανονικοποίηση στα δεδομένα ώστε να βρίσκονται στο εύρος τιμών μεταξύ του 0 και του 1, καθώς διευκολύνει την διαδικασία εκπαίδευσης του μοντέλου. Για τον σκοπό αυτό, χρησιμοποιείται η κανονικοποίηση ελαχίστου-μεγίστου, δεκαδικής ή λογαριθμικής κλιμάκωσης και z-score.
- Επιλογή χαρακτηριστικών με την πλέον ισχυρή επίδραση στην μεταβλητή στόχο του προβλήματος, απορρίπτοντας εκείνα που παρουσιάζουν χαμηλό αντίκτυπο, όπως προκύπτουν από μετρικές αξιολόγησης της σημαντικότητας των χαρακτηριστικών. Έτσι, μειώνεται η διαστατικότητα των δεδομένων και εξοικονομούνται υπολογιστικοί πόροι.
- Εφαρμογή μεθόδων μηχανικής γνωρισμάτων για τη δημιουργία παράγωγων χαρακτηριστικών από τα υφιστάμενα του σετ δεδομένων, ενισχύοντας την

προβλεπτική ικανότητα των μοντέλων. Η συνδυαστική αξιοποίηση των υφιστάμενων μεταβλητών, η δημιουργία όρων αλληλεπίδρασης μεταξύ τους και η βαθιά γνώση του τομέα συνεισφέρουν προς την κατεύθυνση αυτή.

- Μείωση διαστατικότητας δεδομένων με την ελάχιστη απώλεια πληροφορίας, συμβάλλοντας στην εξοικονόμηση υπολογιστικών πόρων και στην πρόληψη από την υπερπροσαρμογή του μοντέλου. Για τη μείωση των διαστάσεων χρησιμοποιούνται τεχνικές όπως η ανάλυση κύριων συνιστωσών (PCA) και η παραγοντική ανάλυση.
- Χειρισμός μη ισορροπημένων κατανομών ως προς τις κλάσεις της μεταβλητής στόχου μέσω της εφαρμογής υπερδειγματοληψίας ή υποδειγματοληψίας στο σετ δεδομένων.
- Διαχωρισμός δεδομένων σε σύνολα εκπαίδευσης και επικύρωσης. Το σύνολο εκπαίδευσης χρησιμοποιείται για την προσαρμογή του μοντέλου και τον συντονισμό των υπερπαραμέτρων. Το σύνολο επικύρωσης χρησιμοποιείται για την αξιολόγηση της απόδοσης του μοντέλου.

### 3.1.3. Διερευνητική Ανάλυση Δεδομένων (EDA)

Το τρίτο στάδιο σε προβλήματα ανάλυσης δεδομένων και μηχανικής μάθησης περιλαμβάνει την διερευνητική ανάλυση δεδομένων ή EDA. Στο στάδιο αυτό, εφαρμόζονται στατιστικές και αναλυτικές τεχνικές για την εξαγωγή πληροφοριών από τα δεδομένα και την κατανόηση υποκείμενων προτύπων, σχέσεων ή τάσεων σε αυτά. Οι βασικές συνιστώσες της EDA συνοψίζονται στα παρακάτω σημεία.

- Περιγραφική στατιστική, η οποία προσφέρει μια πρώτη κατατοπιστική ματιά στα δεδομένα μέσα από βασικούς στατιστικούς δείκτες, όπως ο μέσος όρος, η διάμεσος, η τυπική απόκλιση και τα ποσοστιμόρια, συμβάλλοντας στον εντοπισμό της κεντρικής τάσης, της διασποράς και της κατανομής των μεταβλητών.
- Οπτικοποίηση δεδομένων μέσω διαγραμμάτων, γραφημάτων και άλλων αναπαραστάσεων, βοηθώντας στην κατανόηση των δεδομένων και στην εξαγωγή μακροσκοπικών συμπερασμάτων για αυτά.
- Ανάλυση συσχέτισης που εντοπίζει και αξιολογεί την κατεύθυνση και την ένταση της συσχέτισης μεταξύ μεταβλητών.

### 3.1.4. Μοντελοποίηση

Το τέταρτο στάδιο σε προβλήματα ανάλυσης δεδομένων και μηχανικής μάθησης περιλαμβάνει την ρύθμιση των υπερπαραμέτρων των μοντέλων μηχανικής μάθησης

και την προσαρμογή τους στα δεδομένα εκπαίδευσης. Τα μοντέλα ταξινόμησης που χρησιμοποιήθηκαν για την μελέτη περίπτωσης παρουσιάζονται στα παρακάτω σημεία.

3.1.4.1. Δέντρα Απόφασης

Τα Δέντρα Απόφασης αναφέρονται σε δεντρικές δομές, αποτελούμενες από πιθανά μονοπάτια απόφασης, στα οποία εκτελούνται ακολουθίες ελέγχων. Οι δομές αυτές αξιοποιούνται ως ταξινομητές για την παραγωγή προβλέψεων της κλάσης της μεταβλητής στόχου σε προβλήματα επιτηρούμενης μηχανικής μάθησης μέσα από ένα σύνολο κατανοητών και ερμηνεύσιμων βημάτων από τον άνθρωπο.

Ένα Δέντρο Απόφασης αποτελείται από επιμέρους κόμβους απόφασης που αντιπροσωπεύουν αποφάσεις ή διαχωρισμούς με βάση τα χαρακτηριστικά των δεδομένων. Σημείο εκκίνησης είναι ο κόμβος ρίζα, ένας μοναδικός κόμβος που περιέχει από κάτω του το σύνολο των δεδομένων. Σε κάθε κόμβο απόφασης επιλέγεται ένα γνώρισμα που διαχωρίζει τα δεδομένα σε δύο ή περισσότερους κόμβους, γνωστούς και ως κόμβους παιδιά, με κριτήριο την μεγιστοποίηση της ομοιογένειας της μεταβλητής στόχου εντός κάθε κόμβου παιδιού. Για τον σκοπό αυτό χρησιμοποιείται η εντροπία διαμέρισης, με χαμηλές τιμές να υποδηλώνουν μικρή αβεβαιότητα.

Η διαδικασία διαμέρισης συνεχίζεται αναδρομικά, δημιουργώντας περισσότερους κόμβους μέχρι να επιτευχθεί ένα συγκεκριμένο κριτήριο διακοπής. Τα κριτήρια διακοπής περιλαμβάνουν την επίτευξη ενός μέγιστου βάθους για το δέντρο, την επίτευξη ενός ελάχιστου μεγέθους κόμβου ή την καθαρότητα των κόμβων παιδιών, δηλαδή κόμβων που περιλαμβάνουν μόνο μία κλάση. Κάθε τερματικός κόμβος στο δέντρο απόφασης, γνωστός και ως κόμβος φύλλο, παράγει και την τελική πρόβλεψη για την μεταβλητή στόχο του προβλήματος.

Τα Δέντρα Απόφασης παρουσιάζουν τρία βασικά πλεονεκτήματα που τα καθιστούν δημοφιλείς προσεγγίσεις σε προβλήματα μηχανικής μάθησης. Πρώτον, τα Δέντρα Απόφασης συνιστούν ερμηνεύσιμα μοντέλα, καθώς σε αντίθεση με την πλειοψηφία των μοντέλων μηχανικής μάθησης που λειτουργούν ως «μαύρα κουτιά», αυτά επιτρέπουν την κατανόηση της ροής λήψης μιας απόφασης από τον άνθρωπο. Δεύτερον, τα Δέντρα Απόφασης έχουν ελάχιστες απαιτήσεις για προεπεξεργασία δεδομένων, δηλαδή δεν απαιτείται κάποια κανονικοποίηση ή κωδικοποίηση των δεδομένων εισόδου, και τρίτον, τα Δέντρα Απόφασης χειρίζονται τόσο αριθμητικές όσο και κατηγορηματικές μεταβλητές.

Τα δύο βασικά μειονεκτήματα των Δέντρων απόφασης εντοπίζονται στην υπερπροσαρμογή και στην πόλωση επί των δεδομένων εκπαίδευσης. Έτσι, για την αποφυγή των ανεπιθύμητων αυτών καταστάσεων απαιτείται εκπαίδευση σε ισορροπημένα, πλήρη και συμμετρικά σύνολα εκπαίδευσης, αποφεύγοντας την εξειδίκευση και επιτρέποντας την γενίκευσή του.

Οι βασικοί υπερπαράμετροι των Δέντρων Απόφασης που απαιτούν συντονισμό κατά την επικύρωση του μοντέλου είναι το max_depth, το min_samples_split, το min_samples_leaf και το max_features, με την φυσική ερμηνεία κάθε μίας από αυτές να περιγράφεται στα παρακάτω σημεία.

- Η υπερπαράμετρος max_depth αντιπροσωπεύει το μέγιστο βάθος ενός Δέντρου Απόφασης, με την αύξηση του βάθους να συνεπάγεται υψηλότερη υπολογιστική πολυπλοκότητα και υψηλότερες απαιτήσεις σε χρόνο εκτέλεσης, αυξάνοντας παράλληλα και τον κίνδυνο υπερπροσαρμογής και πόλωσης του μοντέλου.
- Η υπερπαράμετρος max_features υποδηλώνει το κλάσμα των γνωρισμάτων του συνόλου εκπαίδευσης που εξετάζονται σε ένα δέντρο απόφασης κατά τον διαχωρισμό. Τυπικά, λαμβάνει τιμή ίση με τον λογάριθμο ή την ρίζα του πλήθους των γνωρισμάτων.
- Η υπερπαράμετρος min_samples_split αντιπροσωπεύει τον ελάχιστο αριθμό δειγμάτων που απαιτούνται για τον διαχωρισμό ενός εσωτερικού κόμβου στο μοντέλο, απαιτώντας μια ισορροπημένη τιμή, αφού υψηλές και χαμηλές τιμές οδηγούν σε υπερπροσαρμογή και υποπροσαρμογή, αντίστοιχα.
- Η υπερπαράμετρος min_samples_leaf, η οποία αντιπροσωπεύει τον ελάχιστο αριθμό δειγμάτων που απαιτούνται για την δημιουργία ενός κόμβου φύλλο.

#### 3.1.4.2. Random Forest

Το Random Forest ανήκει στην οικογένεια των μεθόδων «από κοινού εκμάθησης» και αποτελεί μια εξειδίκευση της μεθόδου των Δέντρων Απόφασης. Η οικογένεια αυτή χρησιμοποιεί ασθενή μοντέλα εκμάθησης με υψηλή πόλωση και χαμηλή διακύμανση, συναθροίζοντάς τα επιμέρους αποτελέσματα για την κατασκευή ενός ισχυρού προβλεπτικού μοντέλου. Έτσι, το Random Forest χρησιμοποιεί πολλαπλά δέντρα απόφασης, γνωστά και ως εκτιμητές, ώστε να παράγει προβλέψεις.

Η φιλοσοφία του Random Forest υιοθετεί την υπόθεση της επικρατούσας ορθότητας των συλλογικών αποφάσεων έναντι των μεμονωμένων, γνωστή και ως «σοφία του πλήθους». Το μοντέλο εκπαιδεύονται σε τυχαία υποσύνολα του συνόλου εκπαίδευσης χρησιμοποιώντας την τεχνική bootstrap, εκλέγοντας με εναπόθεση ένα καθορισμένο

πλήθος τυχαίων δειγμάτων από το σετ εκπαίδευσης για την προσαρμογή κάθε εκτιμητή. Παράλληλα, χρησιμοποιείται μειωμένη διαστατικότητα στο διάνυσμα γνωρισμάτων για την εκπαίδευση των εκτιμητών, παρόμοια με εκείνη των Δέντρων Απόφασης. Τέλος, οι επιμέρους προβλέψεις των εκτιμητών συναθροίζονται με βάση την πλειοψηφική ψήφο, διαδικασία γνωστή και ως bagging ή bootstrap aggregation.

Το βασικό πλεονέκτημα του Random Forest είναι η ανθεκτικότητα στην υπερπροσαρμογή, καθώς η εκπαίδευση πολλαπλών και ανεξάρτητων Δέντρων Απόφασης σε τυχαία υποσύνολα εκπαίδευσης, καθώς και η συνάθροιση των αποτελεσμάτων προσφέρεται για σταθερή και γενικευμένη πρόβλεψη. Παράλληλα, το Random Forest υπολογίζει την συνεισφορά κάθε γνωρίσματος στην τελική πρόβλεψη υπό την μορφή ενός βάρους σημαντικότητας, ενδυναμώνοντας την ερμηνευσιμότητα των αποτελεσμάτων του μοντέλου.

Η κύρια αδυναμία του Random Forest συνίσταται στην αδυναμία πρόβλεψης σε προβλήματα με μεγάλο πλήθος πιθανών κλάσεων της μεταβλητής στόχου. Στην ουσία, αναιρείται η ίδια η φιλοσοφία του μοντέλου, καθώς κατά την συνάθροιση των επιμέρους εξόδων των εκτιμητών δεν υφίσταται ισχυρή πλειοψηφία. Επιπλέον, το Random Forest είναι ένα αρκετά ακριβό μοντέλο σε υπολογιστικούς πόρους, αφού συνήθως αποτελείται από μεγάλο πλήθος Δέντρων Απόφασης.

Οι βασικοί υπερπαράμετροι που συντονίζονται σε ένα Random Forest είναι οι n_estimators, το min_samples_split, το max_depth και το max_features, με την φυσική ερμηνεία κάθε μίας από αυτές να περιγράφεται στα παρακάτω σημεία.

- Η υπερπαράμετρος n_estimators αναφέρεται στο πλήθος των Δέντρων Απόφασης που απαρτίζουν το Random Forest, γνωστών και ως εκτιμητών. Αν και κατά κανόνα, η αύξηση του πλήθους των εκτιμητών συνεπάγεται βελτίωση της απόδοσης του μοντέλου, απαιτείται μια ισορροπημένη προσέγγιση για την αποφυγή της υπερπροσαρμογής και της άσκοπης δαπάνης υπολογιστικών πόρων.
- Η υπερπαράμετρος max_depth αναφέρεται στο μέγιστο βάθος των επιμέρους εκτιμητών, ακολουθώντας την ίδια λογική που περιγράφηκε στα Δέντρα Απόφασης.
- Η υπερπαράμετρος min_samples_split αναφέρεται στον ελάχιστο αριθμό δειγμάτων που απαιτούνται για τον διαχωρισμό ενός εσωτερικού κόμβου στους εκτιμητές, ακολουθώντας την ίδια λογική που περιγράφηκε στα Δέντρα Απόφασης.

- Η υπερπαράμετρος max_features υποδηλώνει το κλάσμα των γνωρισμάτων του συνόλου δεδομένων εκπαίδευσης που εξετάζονται σε έναν εκτιμητή, ακολουθώντας την ίδια λογική που περιγράφηκε στα Δέντρα Απόφασης.

#### 3.1.4.3. XGBoost

Το XGBoost ανήκει στην οικογένεια των μεθόδων «από κοινού εκμάθησης» και αποτελεί μια εξειδίκευση της μεθόδου των Δέντρων Απόφασης, παρόμοια με εκείνη του Random Forest. Η μέθοδος εκπαιδεύει και προσθέτει σειριακά νέους αδύναμους learners στην αλυσίδα των υφιστάμενων, διορθώνοντας βαθμιαία το σφάλμα των προηγούμενων σε κάθε επανάληψη. Η αντικειμενική συνάρτηση του μοντέλου περιλαμβάνει ποινή σε πολύπλοκα και σύνθετα μοντέλα, μειώνοντας την πιθανότητα υπερπροσαρμογής. Η διαδικασία τερματίζεται είτε όταν το σφάλμα σταματήσει να μειώνεται ή όταν επιτευχθεί το μέγιστο όριο των δέντρων. Η τελική πρόβλεψη του XGBoost είναι ένα σταθμισμένο άθροισμα των προβλέψεων των επιμέρους learners.

Οι βασικοί υπερπαράμετροι που συντονίζονται σε ένα μοντέλο XGBoost είναι το learning_rate, το max_depth, το n_estimators, το subsample και το col_subsample, με την φυσική ερμηνεία κάθε μίας από αυτές να περιγράφεται στα παρακάτω σημεία.

- Η υπερπαράμετρος n_estimators αναφέρεται στο πλήθος των gradient boosted δέντρων που απαρτίζουν το XGBoost, γνωστών και ως εκτιμητών.
- Η υπερπαράμετρος learning_rate αναφέρεται στον ρυθμό εκμάθησης των αδύναμων learners, δηλαδή στο βήμα συρρίκνωσης που εφαρμόζεται σε κάθε ενημέρωση στη διαδικασία ενίσχυσης του μοντέλου. Οι χαμηλότερες τιμές του learning_rate είναι πιο συντηρητικές, αλλά απαιτούν περισσότερα δέντρα (n_estimators) για να μοντελοποιήσουν όλες τις σχέσεις στα δεδομένα.
- Η υπερπαράμετρος του max_depth αναφέρεται στο μέγιστο βάθος των δέντρων του μοντέλου, απαιτώντας μια ισορροπημένη τιμή για την αποφυγή της υπερπροσαρμογής του μοντέλου για μεγάλο βάθος.
- Η υπερπαράμετρος του subsample αναφέρεται στο κλάσμα των παρατηρήσεων επιλέγονται τυχαία ως δείγματα για κάθε δέντρο. Χαμηλές τιμές καθιστούν τον αλγόριθμο πιο συντηρητικό και αποτρέπουν την υπερπροσαρμογή.
- Η υπερπαράμετρος του col_subsample αναφέρεται στο κλάσμα των γνωρισμάτων που δειγματοληπτούνται τυχαία για κάθε δέντρο. Χαμηλές τιμές βοηθούν στην πρόληψη της υπερπροσαρμογής, καθώς παρέχει σε κάθε δέντρο ένα διαφορετικό υποσύνολο χαρακτηριστικών για εκπαίδευση.

#### 3.1.4.4. Λογιστική Παλινδρόμηση

Η Λογιστική Παλινδρόμηση είναι ένας στατιστικός ταξινομητής, η οποία ανήκει στην ευρύτερη οικογένεια μεθόδων επιτηρούμενης μάθησης, και βασίζεται στην προσαρμογή ενός μοντέλου παλινδρόμησης στα δεδομένα μιας δυαδικής κατηγορηματικής μεταβλητής. Η μέθοδος αυτή μοντελοποιεί την σχέση μεταξύ μιας εξαρτημένης δίτιμης κατηγορηματικής μεταβλητής και δύο ή περισσότερων ανεξάρτητων μεταβλητών, προσαρμόζοντας τα δεδομένα εκπαίδευσης στην λογιστική καμπύλη. Η καμπύλη αυτή έχει σιγμοειδή μορφή, χαρακτηριζόμενη από ένα στάδιο εκθετικής αναβάθμισης και βαθμιαίας επιβράδυνσης μέχρι την περάτωσή της στο ασυμπτωτικό σημείο κορεσμού, παράλληλο ως προς τον οριζόντια άξονα, όπως παρουσιάζεται στο Σχήμα.

Η μαθηματική αποτύπωση του μοντέλου λογιστικής παλινδρόμησης, η οποία παρουσιάζεται στην Εξίσωση 1, προσομοιάζει με εκείνη της γραμμικής παλινδρόμησης. Πιο συγκεκριμένα, η Λογιστική Παλινδρόμηση εκφράζεται ως γραμμικός συνδυασμός μεταξύ συντελεστών παλινδρόμησης $\beta_i$ και εξαρτημένων μεταβλητών $X_i$, με τα επιμέρους γινόμενα να εκφράζουν την συνεισφορά κάθε γνωρίσματος στην ολική συμπεριφορά του μοντέλου. Οι τιμές των συντελεστών παλινδρόμησης $\beta_i$ είναι ευθέως ανάλογες του μεγέθους συνεισφοράς της αντίστοιχης ανεξάρτητης μεταβλητής $X_i$ στο μοντέλο, ενώ η θετική τιμή του συντελεστή παλινδρόμησης $\beta_i$ εκφράζει αυξημένη πιθανότητα εμφάνισης του αποτελέσματος επιτυχούς έκβασης στη μεταβλητή στόχο, η οποία τυπικά υποδηλώνεται από την τιμή 1 σε δυαδικά κωδικοποιημένες κατηγορηματικές μεταβλητές.

Εξίσωση 2 - Μοντέλο Λογιστικής Παλινδρόμησης

$$logit(P) = {\log p}/{\log (1-p)} = \beta_0 + \beta_1 \times X_1 + \cdots + \beta_N \times X_N$$

Παράλληλα, η μεταβλητή p υποδηλώνει την πιθανότητα εμφάνισης του αποτελέσματος επιτυχούς έκβασης. Το κατώφλι πιθανότητας για τον χαρακτηρισμό ενός γεγονότος ως επιτυχούς πρόβλεψης είναι το 0.5.

Το βασικό πλεονέκτημα της μεθόδου της λογιστικής παλινδρόμησης έγγυται στην πιθανοτική επεξήγηση των αποτελεσμάτων πρόβλεψης, καθώς η κλάση που προβλέπει το μοντέλο είναι συνάρτηση μιας πιθανότητας, γνωρίζοντας τον βαθμό βεβαιότητας του αποτελέσματος. Παράλληλα, πρόκειται για μια απλή και αποδοτική μέθοδο, με μικρές απαιτήσεις σε υπολογιστικούς πόρους, ακόμα και στην περίπτωση μεγάλου πλήθους ανεξάρτητων μεταβλητών, επιτρέποντας την εκπαίδευση του μοντέλου σε μεγάλα σύνολα δεδομένων.

Ωστόσο, η Λογιστική Παλινδρόμηση υποθέτει γραμμικότητα μεταξύ της μεταβλητής στόχου και των ανεξάρτητων μεταβλητών, η οποία καταρρίπτεται στον πραγματικό κόσμο. Επίσης, η πρόβλεψη του μοντέλου είναι ευαίσθητη στον θόρυβο και στην υπερπροσαρμογή, εισάγοντας απαίτηση για στατιστική προεπεξεργασία του συνόλου δεδομένων εκπαίδευσης με σκοπό την ανίχνευση της υπόθεσης περί γραμμικότητας και τον υπολογισμό των περιγραφικών στατιστικών δεικτών για τον εντοπισμό θορύβου στις παρατηρήσεις.

3.1.4.5. K-Nearest Neighbors (KNN)

Ο KNN ανήκει στην οικογένεια των μη-παραμετρικών μοντέλων επιβλεπόμενης μηχανικής μάθησης, εκείνων των μοντέλων που απομνημονεύουν το σύνολο εκπαίδευσης χωρίς να κατασκευάζουν κάποια συνάρτηση προσαρμογής στα δεδομένα. Η φιλοσοφία της μεθόδου εστιάζει στην εγγύτητα και στην ομοιότητα, ταξινομώντας νέες παρατηρήσεις με βάση παρόμοια χαρακτηριστικά και την απόστασή τους από τις υφιστάμενες παρατηρήσεις του μοντέλου. Στο KNN, ελέγχεται η απόσταση μεταξύ μιας νέας παρατήρησης με k-πλήθος υφιστάμενων παρατηρήσεων και ταξινομείται με εκείνη που παρουσιάζει την μεγαλύτερη εγγύτητα.

Ο καθορισμός της τιμής της παραμέτρου k αποτελεί την κρισιμότερη εργασία κατά την υλοποίηση ενός ταξινομητή KNN, καθώς επιδρά σημαντικά στην απόδοση του μοντέλου. Η παράμετρος k αναφέρεται στο πλήθος των κοντινότερων γειτονικών σημείων που εξετάζονται κατά την ταξινόμηση μιας νέας παρατήρησης. Μια ισορροπημένη τιμή της παραμέτρου κρίνεται απαραίτητη, μιας και για μικρές τιμές του k, το KNN επηρεάζεται από τις ακραίες τιμές, ενώ για μεγάλες τιμές του k, το KNN οδηγείται σε υπεργενίκευση. Ένας συνήθης τρόπος επιλογής της τιμής του k είναι η τετραγωνική ρίζα του πλήθους των παρατηρήσεων στο σύνολο εκπαίδευσης ή χρησιμοποιώντας τεχνικές διασταυρούμενης επικύρωσης. Ωστόσο, δεν υπάρχει καμία εγγυημένη βέλτιστη πρακτική για το "k" που να ισχύει καθολικά, ενώ πολλές φορές είναι δυνατή και η διαισθητική απόδοση τιμής.

Για τον υπολογισμό της εγγύτητας μεταξύ μιας νέας παρατήρησης και των υφιστάμενων ομαδοποιημένων παρατηρήσεων εφαρμόζονται μετρικές απόστασης. Η απόσταση αποτελεί μια μετρική αντικειμενικού υπολογισμού της σχετικής διαφοράς μεταξύ δύο παρατηρήσεων στον ίδιο χώρο. Κατά βάση, ο KNN χρησιμοποιεί την Ευκλείδεια απόσταση και την απόσταση Manhattan για τον υπολογισμό της εγγύτητας μεταξύ δύο παρατηρήσεων. Η Ευκλείδεια απόσταση, η οποία περιγράφεται από την Εξίσωση 1, αξιοποιεί το Πυθαγόρειο θεώρημα για τον υπολογισμό της ελάχιστης

απόστασης σε ευθεία γραμμή μεταξύ δύο παρατηρήσεων, αποτελώντας το πλέον διαισθητικό μέτρο απόστασης. Η απόσταση Manhattan, η οποία περιγράφεται από την Εξίσωση 2, υπολογίζει την απόσταση μεταξύ δύο παρατηρήσεων ως το άθροισμα των απόλυτων διαφορών των συντεταγμένων τους.

Εξίσωση 1 - Ευκλείδεια Απόσταση

$$E_d = \sqrt{\sum_{i=1}^{n} (q_i - p_i)^2}$$

Εξίσωση 2 - Απόσταση Manhattan

$$M_d = \sum_{i=1}^{n} |q_i - p_i|$$

### 3.1.5. Αξιολόγηση

Στην μηχανική μάθηση, για την αξιολόγηση των αποτελεσμάτων και της απόδοσης ενός ταξινομητή κατασκευάζεται ένας τετραγωνικός πίνακας σφάλματος, γνωστός και ως confusion matrix, ο οποίος παρουσιάζεται στην . Οι διαστάσεις του πίνακα αναλογούν στο πλήθος των κλάσεων ταξινόμησης, οπότε στην περίπτωση δυαδικών ταξινομητών κατασκευάζεται ένας πίνακας 2x2. Στις κλάσεις μιας δίτιμης μεταβλητής στόχου αποδίδονται οι τιμές «0» και «1», η οποίες αντιπροσωπεύουν μια συγκεκριμένη κατάσταση εξόδου. Έτσι, ο πίνακας περιλαμβάνει τέσσερεις παραμέτρους που αντιστοιχούν στο πλήθος των αληθώς θετικών (TP), ψευδώς θετικών (FP), αληθώς αρνητικών (TN) και ψευδώς αρνητικών (FN) αποτελεσμάτων, η ερμηνεία των οποίων συνοψίζεται στα παρακάτω σημεία.

- Ως TP ορίζονται τα αποτελέσματα που ανήκουν στην κλάση «1» και ταξινομήθηκαν στην κλάση «1».
- Ως FP ορίζονται τα αποτελέσματα που ανήκουν στην κλάση «0» και ταξινομήθηκαν στην κλάση «1».
- Ως TN ορίζονται τα αποτελέσματα που ανήκουν στην κλάση «0» και ταξινομήθηκαν στην κλάση «0».
- Ως FN ορίζονται τα αποτελέσματα που ανήκουν στην κλάση «1» και ταξινομήθηκαν στην κλάση «0».

Από την συνδυαστική αξιοποίηση των παραπάνω τιμών, προκύπτουν οι τιμές των τεσσάρων βασικών μετρικών σφάλματος, και συγκεκριμένα, η ορθότητα (accuracy), η

ακρίβεια (precision), η ευαισθησία (sensitivity) και η εξειδίκευση (specificity). Οι περιγραφή και η μαθηματική μοντελοποίηση των μετρικών αυτών, συνοψίζεται στα παρακάτω σημεία.

- Η **ορθότητα (accuracy)** αναφέρεται στο ποσοστό των προβλέψεων που ταξινομήθηκαν σωστά από το μοντέλο, υπολογιζόμενη ως ο λόγος των σωστών προβλέψεων, τόσο θετικών όσο και αρνητικών, προς τον συνολικό αριθμό προβλέψεων.
- Η **ευαισθησία (sensitivity)** αναφέρεται στο ποσοστό των πραγματικών θετικών περιπτώσεων που αναγνωρίστηκαν σωστά από το μοντέλο, απαντώντας στο ερώτημα «από όλες τις πραγματικές θετικές περιπτώσεις, πόσες προβλέφθηκαν σωστά;».
- Η **ακρίβεια (precision)** αναφέρεται στο ποσοστό των προβλεπόμενων θετικών περιπτώσεων που αναγνωρίστηκαν σωστά από το μοντέλο, απαντώντας στην ερώτημα «από όλες τις περιπτώσεις που προβλέψαμε ως θετικές, πόσες ήταν πραγματικά θετικές;». Στην ουσία, αντικατοπτρίζει την ικανότητα του μοντέλου να αποφεύγει την εσφαλμένη ταξινόμηση αρνητικών παραδειγμάτων ως θετικών.
- Η **εξειδίκευση (specificity)** αναφέρεται στο ποσοστό των πραγματικών αρνητικών περιπτώσεων αναγνωρίστηκαν σωστά από το μοντέλο. Στην ουσία, αντικατοπτρίζει την ικανότητα του να αναγνωρίζει μια συγκεκριμένη κατηγορία.

# 4. Μελέτη Περίπτωσης

Στο Κεφάλαιο 4, παρουσιάζεται μια μελέτη περίπτωσης που εξετάζει την απόδοση ταξινομητών επιτηρούμενης μηχανικής μάθησης στην κατηγοριοποίηση των χρηστών του Twitter και τον εντοπισμό bots μεταξύ αυτών. Για τον σκοπό αυτό, αναζητήθηκε και συλλέχθηκε σετ ανοικτών δεδομένων συναφές με το αντικείμενο της μελέτης επί του οποίου εφαρμόστηκαν τεχνικές προεπεξεργασίας δεδομένων ώστε να προετοιμαστεί σε κατάλληλη μορφή για την είσοδό του σε προβλεπτικά μοντέλα. Μέσω διασταυρούμενης επικύρωσης με αναζήτηση πλέγματος συντονίστηκαν οι υπερπαράμετροι των επιμέρους μοντέλων, τα οποία εκπαιδεύτηκαν στο 80% των παρατηρήσεων του σετ δεδομένων. Για την αξιολόγηση των επιδόσεων πρόβλεψης των ταξινομητών, αξιοποιήθηκε το 20% των παρατηρήσεων του σετ δεδομένων, υπολογίζοντας τις μετρικές σφάλματος της ακρίβειας, της ανάκλησης και του F1-score.

## 4.1. Συλλογή Δεδομένων για την Μελέτη Περίπτωσης

Η αναζήτηση και συλλογή δεδομένων από αξιόπιστες πηγές αποτελεί τον πυρήνα της διαδικασίας ανάλυσης δεδομένων και μοντελοποίησης ενός προβλήματος με μεθόδους μηχανικής μάθησης. Ποιοτικά σετ δεδομένων συνεισφέρουν στην μείωση των απαιτούμενων ενεργειών κατά την προεπεξεργασία των δεδομένων, στην αποδοτικότερη διαδικασία εκπαίδευσης των μοντέλων και στην ενίσχυση της προβλεπτικής ικανότητας των μοντέλων. Μετά από επισταμένη έρευνα σε ελεύθερες πηγές δεδομένων στο Διαδίκτυο, εντοπίστηκε το αποθετήριο Botometer (Botometer, 2023), το οποίο συνιστά μια συλλογή περισσότερων από είκοσι (20) ανοικτών σετ δεδομένων στην ερευνητική περιοχή του εντοπισμού bots στο Twitter. Τα σετ δεδομένων του Botometer έχουν αξιοποιηθεί από την επιστημονική κοινότητα στα πλαίσια συναφούς ερευνητικού έργου κατά την τελευταία δεκαετία. Η συλλογή των δεδομένων έχει πραγματοποιηθεί μέσω της επίσημης προγραμματιστικής διεπαφής που προσφέρει το Twitter, το Twitter API, διασφαλίζοντας την ακεραιότητά και την αξιοπιστίας των δεδομένων.

Τα διαθέσιμα σετ δεδομένων στο Botometer εξερευνήθηκαν ως προς το πλήθος και την ποικιλία τους σε ποιοτικά και ποσοτικά χαρακτηριστικά, αναζητώντας σετ υψηλής διαστατικότητας. Η εκπαίδευση επί πολλαπλών γνωρισμάτων βοηθάει την βέλτιστη ρύθμιση των υπερπαραμέτρων των μοντέλων και προάγει την ερμηνευσιμότητα των αποτελεσμάτων. Σε κάθε διαθέσιμο σετ δεδομένων, εξετάστηκε η ποσόστωση των κλάσεων ως προς την μεταβλητή στόχο του προβλήματος ώστε να διασφαλιστεί η

ισορροπημένη κατανομή σε πραγματικούς χρήστες και bots. Το σετ δεδομένων cresci-2017 ικανοποιεί όλες τις ανωτέρω απαιτήσεις και έτσι επιλέχθηκε ως εκείνο που θα εξεταστεί κατά την μελέτη περίπτωσης.

Το σετ δεδομένων cresci-2017 οργανώνεται σε τρία (3) επιμέρους υποσύνολα, όπως αυτά παρουσιάζονται στον Πίνακας 2. Κάθε επιμέρους υποσύνολο αντιπροσωπεύει και μια συγκεκριμένη κλάση χρηστών του Twitter. Συνολικά, περιλαμβάνονται 10,653 μοναδικοί χρήστες, ταξινομημένοι σε τρείς κατηγορίες: (α): τους πραγματικούς χρήστες, (b): τα social spambots και (γ): τα traditional spambots. Δεδομένου της αναλογίας μεταξύ μοναδικών χρηστών και συνολικών εγγραφών στο σετ δεδομένων, γίνεται αντιληπτό πως η σχέση πληθικότητας μεταξύ των ανεξάρτητων μεταβλητών και της μεταβλητής στόχου του προβλήματος είναι 1:N, οπότε σε έναν χρήστη μπορεί να αντιστοιχούν μία ή περισσότερες εγγραφές.

| | Μοναδικοί Χρήστες | Συνολικές Εγγραφές |
|---|---|---|
| Πραγματικοί Χρήστες | 3,474 | 8,377,522 |
| Social Spambots | 4,918 | 3,457,344 |
| Traditional Spambots | 2,261 | 6,028,313 |

Πίνακας 2 – Πληροφορίες για τα υποσύνολα δεδομένων του cresci-2017

Από τον Πίνακας 2, παρατηρούμε πως το πλήθος των πραγματικών χρηστών είναι δύο φορές μικρότερο από το αντίστοιχο πλήθος των spambots. Έτσι, εισάγεται η απαίτηση σχετικής δειγματοληψίας στα δεδομένα για την κατασκευή ενός ισορροπημένου δείγματος ως προς την μεταβλητή στόχο του προβλήματος. Αντίθετα, το σετ δεδομένων είναι αρκετά ισορροπημένο ως προς τις συνολικές εγγραφές μεταξύ πραγματικών χρηστών και spambots του Twitter, συνεπώς δεν απαιτείται κάποια αντίστοιχη ενέργεια.

Τα υποσύνολα του σετ δεδομένων των social spambots και των traditional spambots οργανώνονται με την σειρά τους σε επιμέρους υποσύνολα, τα οποία ομαδοποιούνται βάσει της προσέγγισης συλλογής των δεδομένων και του επίκεντρου της δράσης των spambots που περιλαμβάνονται σε αυτά. Μια συνοπτική εικόνα για τα υποσύνολα παρουσιάζεται στα παρακάτω σημεία.

- Το σετ genuine_accounts αποτελείται από 3,474 μοναδικούς χρήστες που αντιπροσωπεύουν πραγματικούς χρήστες του Twitter, περιλαμβάνοντας 8,377,522 μοναδικές εγγραφές στο σύνολο. Για την πιστοποίηση της διαχείρισης των συγκεκριμένων λογαριασμών από ανθρώπινες οντότητες, πραγματοποιήθηκε

ανταλλαγή μηνυμάτων γραμμένων σε φυσική γλώσσα, με τους χρήστες που αποκρίθηκαν επιτυχώς σε αυτά να ταξινομούνται σε αυτό το σετ δεδομένων.

- Το σετ traditional_spambots αποτελείται από 2,261 μοναδικούς χρήστες που αντιπροσωπεύουν traditional spambots του Twitter, όπως περιγράφησαν στην Ενότητα 1.3, περιλαμβάνοντας 6,028,313 μοναδικές εγγραφές στο σύνολο. Επιμερίζεται σε τέσσερα (4) περεταίρω υποσύνολα, όπως αυτά συνοψίζονται στα παρακάτω σημεία.
    - Το traditional_spambots_1 αποτελείται από 1,000 μοναδικά traditional spambots, τα οποία συλλέχθηκαν στα πλαίσια του ερευνητικού έργου των Jiang κ.α. (M. Jiang, 2016)
    - Το traditional_spambots_2 αποτελείται από 100 μοναδικά traditional spambots, τα οποία διαμοιράζονταν παραπλανητικό περιεχόμενο περί νικηφόρων διαγωνισμών.
    - Το traditional_spambots_3 αποτελείται από 433 μοναδικά traditional spambots, τα οποία διαμοιράζονταν spam περιεχόμενο αναφορικά με θέσεις εργασίας.
    - Το traditional_spambots_4 αποτελείται από 1,128 μοναδικά traditional spambots, τα οποία διαμοιράζονταν spam περιεχόμενο αναφορικά με θέσεις εργασίας, όπως προηγουμένως.
- Το social_spambots αποτελείται από 4,918 μοναδικούς χρήστες, που αντιπροσωπεύουν social spambots του Twitter, όπως περιγράφησαν στην Ενότητα 1.3, περιλαμβάνοντας 3,457,344 μοναδικές εγγραφές στο σύνολο. Επιμερίζεται σε τρία (3) περεταίρω υποσύνολα, όπως αυτά συνοψίζονται στα παρακάτω σημεία.
    - Το social_spambots_1 αποτελείται από 944 μοναδικά social spambots, εξυπηρετώντας σκοπούς προώθησης των πολιτικών συμφερόντων συγκεκριμένου υποψηφίου των δημοτικών εκλογών της Ρώμης το 2014 και καλλιέργειας θετικού κλίματος υπέρ του κοινοποιώντας μαζικά περιεχόμενο στο Twitter με προεκλογική χροιά.
    - Το social_spambots_2 αποτελείται από 3,457 μοναδικά social spambots, εξυπηρετώντας σκοπούς προώθησης VIP συνδρομητικού πακέτου της εφαρμογής TaInts.
    - Το social_spambots_3 αποτελείται από 467 μοναδικά social spambots, εξυπηρετώντας σκοπούς προώθησης εκπτωτικών προϊόντων στην πλατφόρμα του Amazon, ενσωματώνοντας σχετικά URLs στο περιεχόμενο των tweets τους.

Κάθε εγγραφή στο σετ δεδομένων αντιπροσωπεύει μια μοναδική καταχώρηση περιεχομένου από τον συγκεκριμένο χρήστη στην πλατφόρμα του Twitter, ήτοι ενός tweet. Το tweet λειτουργεί ως μοναδικό αναγνωριστικό των εγγραφών, αφού αποτελεί εκείνο το γνώρισμα που ταυτοποιεί μοναδικά μια εγγραφή του σετ δεδομένων. Αποτελείται από 17 ανεξάρτητες μεταβλητές, κάθε μία εκ των οποίων αντιστοιχεί σε ένα διακριτό γνώρισμα, οι οποίες παρουσιάζονται στον Πίνακας 3. Στην πρώτη στήλη του πίνακα καταγράφεται το όνομα της εκάστοτε ανεξάρτητης μεταβλητής και στην δεύτερη στήλη καταγράφεται ο τύπος της εκάστοτε ανεξάρτητης μεταβλητής, όπως αυτός αποδίδεται στην Python. Ο τύπος int64 υποδηλώνει ακέραια μεταβλητή, ο τύπος float64 υποδηλώνει μεταβλητή κινητής υποδιαστολής και ο τύπος object υποδηλώνει μεταβλητή αντικειμένου.

| Μεταβλητή Γνωρίσματος | Τύπος Γνωρίσματος |
|---|---|
| id | int64 |
| text | object |
| user_id | int64 |
| truncated | float64 |
| geo | float64 |
| place | object |
| contributors | float64 |
| retweet_count | int64 |
| reply_count | int64 |
| favorite_count | int64 |
| num_hashtags | int64 |
| num_urls | int64 |
| num_mentions | int64 |
| created_at | object |
| timestamp | object |
| crawled_at | object |
| updated | object |

Πίνακας 3 – Ονόματα και τύποι των ανεξάρτητων μεταβλητών του προβλήματος

Ακολούθως, αποτυπώνεται η φυσική ερμηνεία για κάθε μία εκ των 16 ανεξάρτητων μεταβλητών του προβλήματος.

1. Γνώρισμα “id”: Το “id” αντιστοιχεί στο μοναδικό αναγνωριστικό ενός tweet. Το περιεχόμενο των tweets δεν επαρκεί για να ταυτοποιήσει μοναδικά μια ανάρτηση

περιεχομένου στην πλατφόρμα, αφού δύο ή περισσότερα tweets ενός ή περισσότερων χρηστών μπορεί να παρουσιάζουν απόλυτη ταύτιση. Έτσι, κατά την ανάρτηση ενός tweet, το Twitter αποδίδει σε αυτό μια μοναδική ακολουθία τυχαίων ψηφίων που λειτουργεί ως μοναδικό αναγνωριστικό.

2. Γνώρισμα "text": Το "text" αντιπροσωπεύει την ακολουθία χαρακτήρων που συγκροτεί ένα tweet στην πλατφόρμα. Μπορεί να περιλαμβάνει οποιονδήποτε έγκυρο χαρακτήρα της κωδικοποίησης Unicode, με την τρέχουσα πολιτική χρήσης του Twitter να επιτρέπει την δημοσίευση περιεχομένου έως και 280 χαρακτήρες ανά tweet. Ωστόσο, κατά την περίοδο συλλογής των δεδομένων ο περιορισμός χαρακτήρων στα tweets ανερχόταν σε 150.
3. Γνώρισμα "user_id": Το "user_id" είναι το μοναδικό αναγνωριστικό ενός χρήστη στη πλατφόρμα του Twitter. Πρόκειται μια τυχαία αποδιδόμενη ακολουθία ψηφίων που προσφέρει την δυνατότητα ταυτοποίησης με μοναδικό τρόπο ενός χρήστη της πλατφόρμας. Το αναγνωριστικό αυτό δημιουργείται κατά την εγγραφή του χρήστη στο Twitter και παραμένει αμετάβλητο σε ολόκληρη τη διάρκεια ζωής του λογαριασμού στην πλατφόρμα.
4. Γνώρισμα "truncated": Το "truncated" αντιστοιχεί στο πλήθος των περικομμένων ενός tweet, ως αποτέλεσμα του μέγιστου επιτρεπόμενου πλήθους χαρακτήρων σε ένα tweet που επιβάλλεται από την πολιτική της πλατφόρμας. Σήμερα, το Twitter εξασφαλίζει εγγενώς με τεχνικά μέσα πως ένα tweet δεν ξεπερνά το μέγιστο επιτρεπτό όριο, αδυνατώντας να ολοκληρώσει επιτυχώς την ανάρτηση περιεχομένου περισσότερων των 280 χαρακτήρων.
5. Γνώρισμα geo: Το "geo" αντιπροσωπεύει την τρέχουσα τοποθεσία του χρήστη που ανιχνεύεται κατά την ανάρτηση περιεχομένου στην πλατφόρμα. Πρόκειται για πληροφορία που εξάγεται από την συνδυαστική ανάλυση του δικτύου, των πρωτοκόλλων επικοινωνίας και τον φυλλομετρητή. Ωστόσο, αποτελεί άκρως ευαίσθητο προσωπικό δεδομένο, η προστασία του οποίου εμπίπτει στην αυστηρή πολιτική απορρήτου της πλατφόρμας, καθιστώντας το μη-ορατό σε τρίτους χρήστες της πλατφόρμας.
6. Γνώρισμα place: Το "place" αντιπροσωπεύει την γεωγραφική τοποθεσία που δηλώνεται συμπληρωματικά και προαιρετικά από τον χρήστη κατά την ανάρτηση ενός tweet.
7. Γνώρισμα contributors: Το "contributors" δηλώνει το πλήθος τρίτων συντελεστών με συνεισφορά στο περιεχόμενο του tweet. Οι συνεργατικές αναρτήσεις είναι αποτέλεσμα από κοινού δημοσίευσης δύο ή περισσότερων λογαριασμών και αποτελεί μια λειτουργία που έχει ενσωματωθεί από ορισμένα κοινωνικά δίκτυα

προσφάτως. Έτσι, μένει να διαπιστωθεί αν αυτή η λειτουργία προσφερόταν από το Twitter κατά το χρονικό διάστημα συλλογής των δεδομένων.

8. Γνώρισμα favorite_count: Το “favorite_count” αντιπροσωπεύει το πλήθος των θετικών αντιδράσεων από τρίτους χρήστες του Twitter στο συγκεκριμένο περιεχόμενο. Οι θετικές αντιδράσεις καταγράφονται ως «μου αρέσει» και αποτελούν το πιο χαρακτηριστικό παράδειγμα αλληλεπίδρασης μεταξύ χρηστών στην πλατφόρμα. Σύμφωνα με την βιβλιογραφική ανασκόπηση, ανάγεται σε κρίσιμο γνώρισμα για την ταξινόμηση ενός χρηστών του Twitter.
9. Γνώρισμα retweet_count: Το “retweet_count” αντιπροσωπεύει το πλήθος των κοινοποιήσεων από τρίτους χρήστες του Twitter στο συγκεκριμένο περιεχόμενο. Σύμφωνα με την βιβλιογραφική ανασκόπηση, ανάγεται σε κρίσιμο γνώρισμα για την ταξινόμηση των χρηστών στο Twitter.
10. Γνώρισμα num_hashtags: Το “num_hashtags” αντιπροσωπεύει το πλήθος των hashtags στο συγκεκριμένο tweet, τα οποία αντανακλούν μια συγκεκριμένη τάση στην πλατφόρμα. Τα hashtags αποτελούν έναν δημοφιλή τρόπο εγγενούς προώθησης του περιεχομένου των χρηστών στο Twitter, ευνοώντας την συμμετοχή στον δημόσιο διάλογο.
11. Γνώρισμα num_urls: Το “num_urls” αντιπροσωπεύει το πλήθος των URLs στο συγκεκριμένο tweet. Τα URLs αποτελούν σημαίνον γνώρισμα για την ταξινόμηση των χρηστών του Twitter, με μεγάλο ποσοστό των bots να τα χρησιμοποιεί URLs με σκοπό να κατευθύνει κίνηση πραγματικών χρηστών προς κακόβουλους ιστότοπους ή να εξυπηρετήσεις σκοπούς διαφήμισης.
12. Γνώρισμα num_mentions: Το “num_mentions” αντιπροσωπεύει το πλήθος των ονομαστικών αναφορών που περιλαμβάνονται στο συγκεκριμένο tweet. Μέσω των ονομαστικών αναφορών, αποστέλλονται προσωποποιημένες ειδοποιήσεις στους χρήστες που επισημαίνονται, λειτουργώντας ως μια επιθετικότερη προσπάθεια προσέγγισης πραγματικών χρηστών από bots του Twitter.
13. Γνώρισμα created_at: Το “created_at” αντιπροσωπεύει την χρονοσφραγίδα δημιουργίας του λογαριασμού του συγκεκριμένου χρήστη στο Twitter.
14. Γνώρισμα timestamp: Το “timestamp” αντιπροσωπεύει την χρονοσφραγίδα ανάρτησης του συγκεκριμένου tweet από τον χρήστη στο Twitter.
15. Γνώρισμα crawled_at: Το “crawled_at” αντιπροσωπεύει την χρονοσφραγίδα διάχυσης του συγκεκριμένου tweet στην πλατφόρμα. Η ανάρτηση του περιεχομένου μπορεί να πραγματοποιηθεί ασύγχρονα, προγραμματίζοντάς την ανάρτηση στην επιθυμητή χρονική στιμγή.

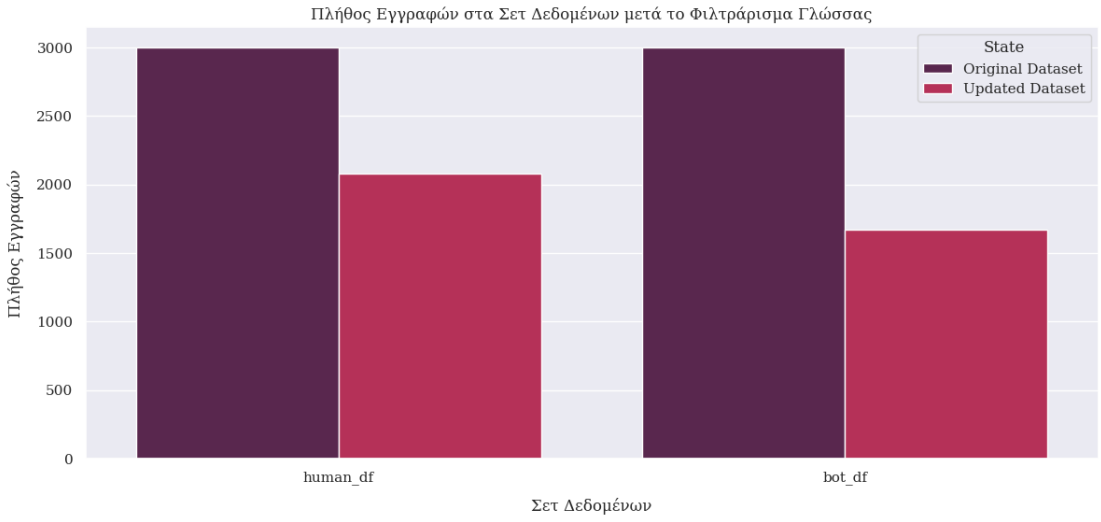
16. Γνώρισμα updated: Το “updated” αντιπροσωπεύει την χρονοσφραγίδα ανανέωσης του συγκεκριμένου tweet στην πλατφόρμα. Το Twitter επιτρέπει την επεξεργασία του περιεχομένου ενός tweet ακόμα και μετά την ανάρτησή του, επισημαίνοντας σε τρίτους χρήστες πως στο τρέχον περιεχόμενο έχουν εφαρμοστεί αλλαγές.

Ο υψηλός όγκος δεδομένων του cresci-2017 εισάγει σημαντικό εμπόδιο στην αποδοτική μοντελοποίηση του προβλήματος με μεθόδους μηχανικής μάθησης, με την εκπαίδευση των μοντέλων να απαιτεί υψηλή διαθεσιμότητα σε υπολογιστικούς πόρους. Επιπλέον, η προσαρμογή των μοντέλων μηχανικής μάθησης σε σετ δεδομένων μεγάλου όγκου ενδέχεται να οδηγήσει σε πόλωση έναντι κάποιας κλάσης και υπερπροσαρμογή, δυσχεραίνοντας την γενίκευση των προβλεπτικών μοντέλων και υποβαθμίζοντας την ακρίβεια της πρόβλεψης, αντίστοιχα. Έτσι, αποφασίστηκε η εφαρμογή τυχαίας δειγματοληψίας στα υποσύνολα δεδομένων για κάθε κατηγορία spambot με σκοπό την κατασκευή ενός ισορροπημένου σετ δεδομένων ως προς την μεταβλητή στόχο, σημαντικά μικρότερου πλήθος εγγραφών. Στην μελέτη περίπτωσης, τα traditional spambots και τα social spambots αντιμετωπίζονται ως μία κοινή κατηγορία χρηστών με την ευρύτερη έννοια των bots, αποσκοπώντας επί της αρχής στην κατασκευή ενός προβλεπτικού μοντέλου που δεν εξειδικεύεται σε κάποια συγκεκριμένη έκφρασή τους. Από κάθε κλάση της εξαρτημένης μεταβλητής, ήτοι πραγματικών χρηστών και spambots, επιλέχθηκαν με τυχαίο τρόπο 3,000 εγγραφές, κατασκευάζοντας ένα σύνολο δεδομένων εκπαίδευσης και επικύρωσης με 6,000 δείγματα στο σύνολο.

Απαραίτητη προϋπόθεση για την αποκρυπτογράφηση και την ερμηνεία του συναισθήματος με μεθόδους μηχανικής μάθησης είναι ο εντοπισμός της γλωσσικής ρίζας ενός κειμένου σε φυσική γλώσσας, όπως τα tweets. Για τον σκοπό αυτό, χρησιμοποιούνται προεκπαιδευμένα μοντέλα μηχανικής μάθησης, προσαρμοσμένα σε σύνολα εκπαίδευσης μεγάλου όγκου, ικανά να προβλέψουν με υψηλή ακρίβεια την γλώσσα μιας ακολουθίας συμβολοσειρών. Τυπικά, η απόδοση τέτοιων μοντέλων βελτιστοποιείται για κείμενο γραμμένο στην αγγλική γλώσσα, μιας και η πλειοψηφία των πακέτων εκπαίδευσης έχουν ως σημείο αναφοράς τα αγγλικά. Συνεπώς, κρίνεται σκόπιμη η απόρριψη των εγγραφών σε γλώσσα διαφορετική της αγγλικής για το σετ δεδομένων του cresci-2017.

Η ανίχνευση της γλώσσας ενός κειμένου πραγματοποιείται με την βιβλιοθήκη langdetect της Python, η οποία είναι βασισμένη στο Cloud Translation API της Google, προσεγγίζοντας τον εντοπισμό της γλώσσας ενός κειμένου σε φυσική γλώσσα όπως ακριβώς και το Google Translate. Η μέθοδος detect_lang της βιβλιοθήκης langdetect

δέχεται ως όρισμα μια ακολουθία συμβολοσειρών, επιστρέφοντας μια πρόβλεψη σχετικά με εκτιμώμενη γλώσσα του κειμένου εισόδου. Δεδομένης της στοχαστικής φύσης τέτοιων προβλέψεων, απαιτείται ο προσδιορισμός ενός κατωφλίου πιθανότητας που υποδηλώνει την εμπιστοσύνη του αποτελέσματος της πρόβλεψης. Εν προκειμένου, η τιμή του κατωφλίου ορίστηκε στο 0.90, ενισχύοντας την πιθανότητα θετικής έκβασης της πρόβλεψης του μοντέλου, με τις εγγραφές που αντιστοιχούσαν σε χαμηλότερες πιθανότητες να απορρίπτονται.




Σχήμα 1 - Πλήθος εγγραφών ανά σετ δεδομένων, πριν και μετά την εφαρμογή φιλτραρίσματος των δεδομένων βάσει της αγγλικής γλώσσας.

Το ανανεωμένο σετ δεδομένων εκπαίδευσης και επικύρωσης που προκύπτει μετά την συλλογή των δεδομένων και το φιλτράρισμά τους βάσει της γλωσσικής ρίζας του περιεχομένου, περιλαμβάνοντας συνολικά 3,756 tweets στα αγγλικά. Το σετ δεδομένων είναι αρκετά ισορροπημένο μεταξύ των δύο κλάσεων της μεταβλητής στόχου, με τις 1,673 εγγραφές να αναφέρονται σε bots και τις 2,083 εγγραφές να αναφέρονται σε πραγματικούς χρήστες. Έτσι, οι πραγματικοί χρήστες του Twitter καταλαμβάνουν το 44,54% του ανανεωμένου δείγματος, ενώ τα bots το υπόλοιπο 55,46%. Από το αρχικό σετ δεδομένων που συλλέχθηκε, απορρίφθηκαν 2,244 εγγραφές κατά το φιλτράρισμα γλώσσας των tweet υπό την απαίτηση να είναι γραμμένο στα αγγλικά με εμπιστοσύνη 90%, ποσοστό που αντιστοιχεί στο 37,4% επί του συνόλου των εγγραφών του αρχικού σετ δεδομένων. Τέλος, το ανανεωμένο σετ δεδομένων διασπάται σε δύο επιμέρους υποσύνολα, το σύνολο εκπαίδευσης και το σύνολο επικύρωσης με σκοπό την προσαρμογή και την αξιολόγηση των αποτελεσμάτων των μεθόδων, αντίστοιχα. Στο σετ εκπαίδευσης αποδίδεται το 80% των συνολικών εγγραφών, περιλαμβάνοντας 1,666 εγγραφές πραγματικών χρηστών και 1,338 εγγραφές bot του Twitter, ενώ το υπόλοιπο 20% συνιστά το σετ επικύρωσης.

## 4.2. Προεπεξεργασία Δεδομένων για την Μελέτη Περίπτωσης

Η προεπεξεργασία των δεδομένων συνιστά ένα κρίσιμο βήμα στη μοντελοποίηση προβλημάτων με τεχνικές μηχανικής μάθησης. Σε αυτό, το σετ δεδομένων προετοιμάζεται και μετασχηματίζεται σε συμβατή μορφή με εκείνη των εισόδων των μοντέλων. Οι κενές τιμές που περιλαμβάνονται στο σετ δεδομένων συμπληρώνονται ή απορρίπτονται κατά τρόπο που υποδεικνύεται από την ίδια την φύση των δεδομένων. Οι ποιοτικές μεταβλητές που αντιπροσωπεύουν κατηγορικά γνωρίσματα κωδικοποιούνται σε αριθμητικές αναπαραστάσεις ώστε να είναι διαχειρίσισμες από τα μοντέλα. Τέλος, παράγονται και νέα γνωρίσματα, εξαγόμενα από τα υφιστάμενα γνωρίσματα του σετ δεδομένων, συνεισφέροντας στην βελτίωση της προβλεπτικής ικανότητας των μοντέλων.

### 4.2.1. Διαχείριση Κενών Τιμών

Στα προβλήματα μηχανικής μάθησης, η διαχείριση των κενών τιμών δεν αντιμετωπίζεται με ενιαίο και οριζόντιο τρόπο, αλλά είναι προϊόν κριτικής σκέψης και αντίληψης. Η στάθμιση του ρίσκου μεταξύ της απώλειας πληροφορίας από την απόρριψη και της εισαγωγής θορύβου από την συμπλήρωση των κενών τιμών καθοδηγεί την προσέγγιση που ακολουθείται. Έτσι, σε σετ δεδομένων όπου οι τιμές είναι διατεταγμένες στον χρόνο ή παρουσιάζουν μία χρονική αλληλουχία, οι κενές τιμές συμπληρώνονται με την προηγούμενη ή την επόμενη τιμή, εισάγοντας κατάλληλη χρονική υστέρηση βάσει της έντασης της αυτοσυσχέτισης της μεταβλητής. Εναλλακτικά, οι κενές τιμές συμπληρώνονται με κάποιο μέτρο περιγραφικής στατιστικής, όπως ο μέσος όρος, η διάμεσος και η τιμή με την υψηλότερη συχνότητα. Οι κενές τιμές του σετ δεδομένων cresci-2017 παρουσιάζονται στον Πίνακας 4.

| Μεταβλητή Γνωρίσματος | Πλήθος Μη-Κενών Τιμών |
|---|---|
| id | 3,756 |
| text | 3,756 |
| user_id | 3,756 |
| truncated | 0 |
| geo | 0 |
| place | 159 |
| contributors | 0 |
| retweet_count | 3,756 |
| reply_count | 3,756 |

| | |
|---|---|
| favorite_count | 3,756 |
| num_hashtags | 3,756 |
| num_urls | 3,756 |
| num_mentions | 3,756 |
| created_at | 3,756 |
| timestamp | 3,756 |
| crawled_at | 3,756 |
| updated | 3,756 |

Πίνακας 4 - Πλήθος μη-κενών τιμών ανά γνώρισμα του σετ δεδομένων

Από τα αποτελέσματα του πίνακα, παρατηρούμε η πλειοψηφία των γνωρισμάτων του σετ δεδομένων είναι πλήρης, αναδεικνύοντας τα οφέλη από την αξιοποίηση αξιόπιστων πηγών δεδομένων, όπως το Twitter API. Αναφορικά με τις κενές τιμές, αυτές εντοπίζονται σε τέσσερα γνωρίσματα, και συγκεκριμένα σε τρία από αυτά, το "truncated", το "geo" και το "contributors" περιλαμβάνονται αποκλειστικά κενές τιμές, ενώ στο γνώρισμα "place" περιλαμβάνεται μικρό πλήθος μη-κενών τιμών.

- Το γνώρισμα "truncated" αντιπροσωπεύει το πλήθος των περικομμένων χαρακτήρων από το tweet ως αποτέλεσμα του περιορισμού χαρακτήρων που επιβάλλει η πλατφόρμα στην ανάρτηση περιεχομένου. Αποτελείται αποκλειστικά από κενές τιμές, εικάζοντας πως οφείλεται στον εγγενή τεχνικό περιορισμό της πλατφόρμας να μην επιτρέπει την ανάρτηση tweets με πλεονάζον πλήθος χαρακτήρων. Το γνώρισμα αυτό αποφασίζεται να απορριφθεί πλήρως από το σετ δεδομένων, καθώς δεν μπορεί να εισφέρει ουσιαστικά στην εκπαίδευση των μοντέλων.
- Το γνώρισμα "geo" αντιπροσωπεύει την γεωγραφική τοποθεσία που εντοπίζεται ο χρήστης. Αποτελείται αποκλειστικά από κενές τιμές, εικάζοντας πως οφείλεται στους αυστηρούς περιορισμούς της πολιτικής απορρήτου της πλατφόρμας για τα προσωπικά δεδομένα των χρηστών, απαγορεύοντας την δημόσια πρόσβαση στην πληροφορία αυτή μέσω του Twitter API. Και σε αυτή την περίπτωση, το γνώρισμα αποφασίζεται να απορριφθεί πλήρως από το σετ δεδομένων, καθώς δεν μπορεί να εισφέρει ουσιαστικά στην εκπαίδευση των μοντέλων.
- Το γνώρισμα "contributors" αντιπροσωπεύει την κοινή συνεισφορά δύο ή περισσότερων χρηστών σε ένα tweet. Αποτελείται αποκλειστικά από κενές τιμές, εικάζοντας πως είναι απόρροια της εγγενούς λειτουργίας της πλατφόρμας μην υποστηρίζει την από κοινού ανάρτηση περιεχομένου κατά την περίοδο

συλλογής των δεδομένων. Και σε αυτή την περίπτωση, το γνώρισμα αποφασίζεται να απορριφθεί πλήρως από το σετ δεδομένων, καθώς δεν μπορεί να εισφέρει ουσιαστικά στην εκπαίδευση των μοντέλων.

- Το γνώρισμα "place" αντιπροσωπεύει την τοποθεσία που δηλώνει προαιρετικά ο χρήστης στο tweet, διαφοροποιούμενη από το γνώρισμα "geo" καθώς δεν συλλέγεται με τεχνικά μέσα από την πλατφόρμα. Αποτελείται από 159 μη-κενές τιμές, ήτοι το 96% των τιμών του "place" είναι κενές. Η φύση του γνωρίσματος δεν επιτρέπει την συμπλήρωση των μη-κενών τιμών μέσω κάποιας κοινής πρακτικής και προσέγγισης που χρησιμοποιείται σε αριθμητικά δεδομένα. Για την διαχείριση των κενών τιμών με τρόπο που συνεισφέρει ουσιαστικά στην εκπαίδευση δεδομένων, αποφασίστηκε η δημιουργία του παράγωγου δίτιμου γνωρίσματος "tweet.has_place", λαμβάνοντας τιμή "1" εφόσον το tweet περιλαμβάνει τοποθεσία.

Η εξερεύνηση των κενών τιμών του σετ δεδομένων οδήγησε στην αφαίρεση τεσσάρων γνωρισμάτων, αυτών του "truncated", "geo", "contributors" και "place", ενώ κατασκευάστηκε και νέο παράγωγο γνώρισμα, αυτό του "tweet.has_place" που αντιπροσωπεύει αν ένα tweet ενσωματώνει πληροφορία σχετική με την τοποθεσία. Τελικά, το ανανεωμένο σετ δεδομένων αποτελείται από 14 γνωρίσματα.

### 4.2.2. Παράγωγα Γνωρίσματα των Tweets

Κατά την βιβλιογραφική ανασκόπηση, υπογραμμίστηκε η προοπτική σημαντικής αναβάθμισης της απόδοσης των μοντέλων μηχανικής μάθησης από την εμβάθυνση στο περιεχόμενο των tweets μέσα από την εξαγωγή παράγωγων γνωρισμάτων. Το πλήθος των hashtags, των ονομαστικών αναφορών και των ειδικών χαρακτήρων αποτελούν παραδείγματα παράγωγων γνωρισμάτων που συνιστούν ισχυρές ενδείξεις ταξινόμησης ενός χρήστη του Twitter σε κάποια κλάση. Μέρος της πληροφορίας αυτής ενσωματώνεται ήδη στο σετ δεδομένων, όπως το πλήθος των hashtags, των URLs και των ονομαστικών αναφορών. Για την εξαγωγή περισσότερων παράγωγων γνωρισμάτων μέσα από τα tweets, υλοποιήθηκαν συναρτήσεις λογικών εκφράσεων στην Python, οι οποίες υπολογίζουν και επιστρέφουν τα αντιπροσωπευτικά πλήθη αναφορικά με το πλήθος των ψηφίων, των χαρακτήρων emoji και του μήκους του tweet.

- Το παράγωγο γνώρισμα "tweet.num_digits" αντιπροσωπεύει το πλήθος των ψηφίων που περιλαμβάνονται σε ένα tweet. Για την εξαγωγή του,

κατασκευάστηκε κατάλληλη λογική έκφραση που αναζητά ψηφία μεταξύ των χαρακτήρων ενός tweet.

- Το παράγωγο γνώρισμα “tweet.num_emoji” αντιπροσωπεύει το πλήθος των emoji που περιλαμβάνονται σε ένα tweet. Για την εξαγωγή του, χρησιμοποιήθηκε η εξωτερική βιβλιοθήκη emoji της Python, η οποία προσφέρει την μέθοδο emoji_count που επιστρέφει το πλήθος των emoji μεταξύ των χαρακτήρων ενός tweet.
- Το παράγωγο γνώρισμα “tweet.length” αντιπροσωπεύει το μήκος ενός tweet σε χαρακτήρες. Για την εξαγωγή του, χρησιμοποιήθηκε η μέθοδος len της Python.

Παράλληλα, υπολογίστηκε και η εντροπία της πληροφορίας των tweets, η οποία αποδόθηκε στο παράγωγο γνώρισμα “tweet.entropy”. Η εντροπία αποτελεί συνδυαστικό μέτρο της τυχαιότητας, της αβεβαιότητας και της αταξίας σε ένα σύνολο χαρακτήρων, εν προκειμένου στο tweet, εξετάζοντας την διάταξη των χαρακτήρων, και όχι την σημασιολογική έννοια των λέξεων που συνιστούν.

### 4.2.3. Παράγωγα Γνωρίσματα των Χρηστών

Τα παράγωγα γνωρίσματα χρηστών του Twitter αναφέρονται σε ιδιότητες του λογαριασμό που αντιστοιχεί στον χρήστη που ανάρτησε το tweet, και όχι στο περιεχόμενό του. Στο σετ δεδομένων, μετά την αφαίρεση του “geo”, προσφέρονται μόλις δύο τέτοια γνωρίσματα, το “user_id” που αντιπροσωπεύει το μοναδικό αναγνωριστικό του χρήστη και το “created_at” που αντιπροσωπεύει την ημερομηνία δημιουργίας του λογαριασμού. Το πληροφοριακό περιεχόμενο του “user_id” είναι μια τυχαία διάταξη ψηφίων που αποδίδεται στον χρήστη από το Twitter κατά την δημιουργία του λογαριασμού, αδυνατώντας σε ουσιαστική συνεισφορά κατά την εκπαίδευση των προβλεπτικών μοντέλων. Ομοίως, η ανεξάρτητη μεταβλητή “created_at” βρίσκεται σε μορφή αντικειμένου τύπου ημερομηνίας, μορφή που δεν μπορεί να κατανοηθεί από τα μοντέλα μηχανικής μάθησης. Το “user_id” απορρίφθηκε σαν γνώρισμα, ενώ από το “created_at” δημιουργήθηκαν δύο νέα παράγωγα γνωρίσματα, τα οποία ενσωματώνουν ουσιαστική πληροφορία για τον χρήστη σε αξιοποιήσιμη μορφή από τα μοντέλα ταξινόμησης.

- Το παράγωγο γνώρισμα “tweet.account_age” αντιπροσωπεύει την ηλικία του συγκεκριμένου λογαριασμού στο Twitter, μετρημένη σε ημέρες. Προκύπτει ως η διαφορά μεταξύ της χρονοσφραγίδας της δημιουργίας του λογαριασμού, ήτοι “created_at”, και της ανάρτησης του τελευταίου tweet, ήτοι “crawled_at”. Αν

και η προσέγγιση αυτή δεν είναι απόλυτα ακριβής, προσφέρει μια αρκετά κατατοπιστική εικόνα για το χρονικό διάστημα που δραστηριοποιείται ο συγκεκριμένος λογαριασμός στην πλατφόρμα.

- Το παράγωγο γνώρισμα "tweet.time_update" αντιπροσωπεύει το χρονικό διάστημα μεταξύ της ανάρτησης ενός tweet και της ανανέωσής του από τον συγκεκριμένο λογαριασμό, προκύπτοντας ως η αντίστοιχη διαφορά μεταξύ των δύο γνωρισμάτων. Στην πλειοψηφία των εγγραφών λαμβάνει μηδενική τιμή, ωστόσο πιθανολογείται πως αρκετοί λογαριασμοί bots μεταβάλλουν εν μέρει ή καθολικά το περιεχόμενό τους, εξυπηρετώντας σκοπούς παραπλάνησης.

### 4.2.4. Συναισθηματική Ανάλυση των Tweets

Η συναισθηματική ανάλυση αποτελεί μέθοδο επεξεργασίας φυσικής γλώσσας (NLP) με στατιστικές μεθόδους και μεθόδους μηχανικής μάθησης για την αναγνώριση συναισθηματικών προτύπων σε κείμενο φυσικής γλώσσας, καθώς και τον εντοπισμό και την εξαγωγή υποκειμενικών πληροφοριών μέσα από αυτό. Αναγνωρίζοντας την κατεύθυνση της πόλωσης σε μια ολοκληρωμένη πρόταση σε φυσική γλώσσα, αποδίδει σε αυτή μια διακριτή συναισθηματική κατάσταση. Τυπικά, οι μέθοδοι συναισθηματικής ανάλυσης ταξινομούν την πόλωση του κειμένου σε θετική, αρνητική και ουδέτερη.

Η ανάλυση συναισθήματος των tweets αποτελεί ένα ακόμα κρίσιμο κομμάτι της μελέτης περίπτωσης, αφού τα tweets που προέρχονται από bots τείνουν να παρουσιάζουν κοινό μοτίβο συναισθήματος, συνήθως υπέρμετρα θετικό ή αρνητικό. Έτσι, η ανάλυση του κειμένου των tweets με τεχνικές επεξεργασίας φυσικής γλώσσας για την ταξινόμησή τους ως θετικά, αρνητικά και ουδέτερα φορτισμένα κρίνεται απαραίτητη. Για τον σκοπό αυτό, χρησιμοποιήθηκε η βιβλιοθήκη Textblob της Python, η οποία προσφέρει μεθόδους για τον συμβολισμό (tokenization), την συντακτική ανάλυση και την συναισθηματική ανάλυση κειμένου σε φυσική γλώσσα. Στα πλαίσια της μελέτης, η Textblob χρησιμοποιήθηκε αποκλειστικά για την αποκωδικοποίηση του συναισθήματος των tweets. Ωστόσο, για την βέλτιστη απόδοση της μεθόδου απαιτούνται δύο βήματα προεπεξεργασίας του περιεχομένου τους, όπως περιγράφονται στα παρακάτω σημεία.

- Αφαίρεση περιεχομένου χωρίς ουσιαστική συνεισφορά στην πόλωση

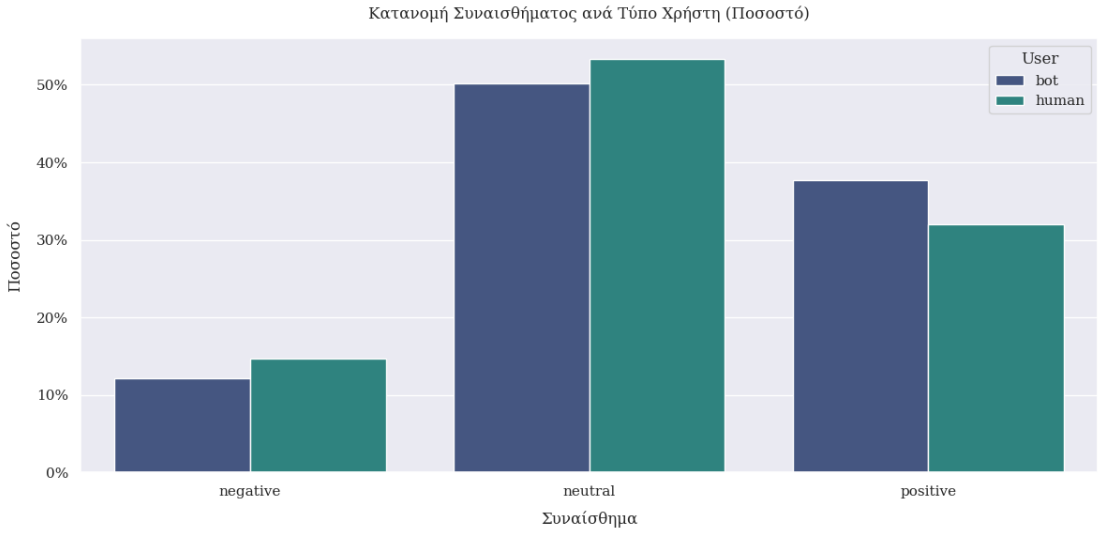
Η ανίχνευση του συναισθήματος απαιτεί εμβάθυνση στην σημασιολογία των λέξεων, αφού αντιμετωπίζει τις συμβολοσειρές χαρακτήρων ως ενιαία νοηματικά αντικείμενα.

Έτσι, τμήματα κειμένου με νοηματικό κενό πρέπει να αφαιρεθούν για να διευκολύνουν την ανάλυση και να βελτιώσουν την απόδοση των μοντέλων. Οι ειδικοί χαρακτήρες, τα σημεία στίξης, τα τονικά σημάδια και τα URLs αποτελούν τυπικά παραδείγματα περιεχομένου στα tweets χωρίς ουσιαστική συνεισφορά στον εντοπισμό της κατεύθυνσης της πόλωσης, συνεπώς κρίνεται σκόπιμη η αφαίρεση τους. Συγκεκριμένα για τον αυτοματοποιημένο εντοπισμό των URLs, κατασκευάστηκαν κατάλληλες συναρτήσεις που σαρώνουν το περιεχόμενο των tweets, αναζητώντας μοτίβα κανονικών εκφράσεων που ταυτίζονται με εκείνα των URLs, όπως το πρόθεμα του πρωτοκόλλου HTTP και το πρόθεμα διευθύνσεων ιστού WWW. Η ίδια διαδικασία ακολουθήθηκε και για τα emojis, τα οποία αν και έχουν συναισθηματικό πρόσημο, επί της ουσίας δεν μπορούν να ερμηνευθούν από αυτές τις μεθόδους. Επίσης, εντοπίστηκαν και αφαιρέθηκαν τα άκλιτα μέρη του λόγου που χρησιμοποιούνται για τον προσδιορισμό ρημάτων και επιρρημάτων, όπως άρθρα, προθέσεις, σύνδεσμοι, επιφωνήματα και μόρια. Για τον σκοπό αυτό, χρησιμοποιήθηκαν τα λεξικά stopwords και punkt της βιβλιοθήκης NLTK, και μέσω παράλληλης αναζήτησης, αφαιρέθηκαν μέρη αυτά από το περιεχόμενο των tweets.

- Λημματοποίηση του περιεχομένου των tweets

Η λημματοποίηση είναι η διαδικασία επεξεργασίας κειμένου σε φυσική γλώσσα για την αναγωγή λέξεων στην βασική ή στην ριζική μορφή τους. Έτσι, τα ρήματα, τα ουσιαστικά και τα επίθετα ανάγονται σε μία κοινή πτωτική κλίση ή χρόνο, επιστρέφοντας στη βασική λεξικογραφική μορφή τους, γνωστή και ως λήμμα. Πρόκειται για μία αρκετά σημαντική διαδικασία στο πλαίσιο της συναισθηματικής ανάλυσης, καθώς μειώνει σημαντικά τον όγκο των δεδομένων, επιταχύνοντας την ανάλυση, ενώ χρησιμεύει και στην ποσοτικοποίησης της συνεισφοράς συγκεκριμένων λέξεων στα αποτελέσματα της ανάλυσης. Η βιβλιοθήκη NLTK προσφέρει την μέθοδο WordNetLemmatizer για την λημματοποίηση περιεχομένου, όπου υλοποιώντας μια επαναληπτική ρουτίνα, συγχωνεύτηκαν σε κοινή μορφή λέξεις με το ίδιο μορφολογικό περιεχόμενο.

Στο Σχήμα 2, παρουσιάζεται η ποσοστιαία κατανομή των διακριτών καταστάσεων συναισθηματικής πόλωσης μεταξύ για τις κλάσεις των πραγματικών χρηστών και των bots του δείγματος.




Σχήμα 2 - Κατανομή συναισθήματος μεταξύ των κατηγοριών χρηστών.

Παρατηρούμε πως η συχνότητα δημοσίευσης ουδέτερα φορτισμένων tweets είναι κατά 7% υψηλότερη σε λογαριασμούς που διαχειρίζονται από πραγματικούς χρήστες, και συγκεκριμένα ανέρχεται σε 55.29% έναντι 48.50% από bots. Η σημαντική ποσόστωση σε ουδέτερα πολωμένα tweets από bots, ερμηνεύεται ως προσπάθεια μίμησης των ανθρώπινων μοτίβων συμπεριφοράς. Αντίθετα, αναφορικά με τα θετικά συναισθηματικά φορτισμένα tweets, τα bots εμφανίζουν υψηλότερη συχνότητα που κυμαίνεται στο 39.40% έναντι 30.95% της αντίστοιχης συχνότητας των πραγματικών χρηστών. Τα αποτελέσματα αυτά επιβεβαιώνουν τα σχετικά βιβλιογραφικά ευρήματα που παρουσιάστηκαν στο Κεφάλαιο 2, τονίζοντας την τάση των bots να αναρτούν tweets με αισιόδοξο περιεχόμενο. Τέλος, οι πραγματικοί χρήστες εμφανίζουν ελαφρώς υψηλότερη τάση ανάρτησης tweets με αρνητικό συναισθηματικό πρόσημο συγκριτικά με τα bots.

### 4.2.5. Συσχέτιση Γνωρισμάτων

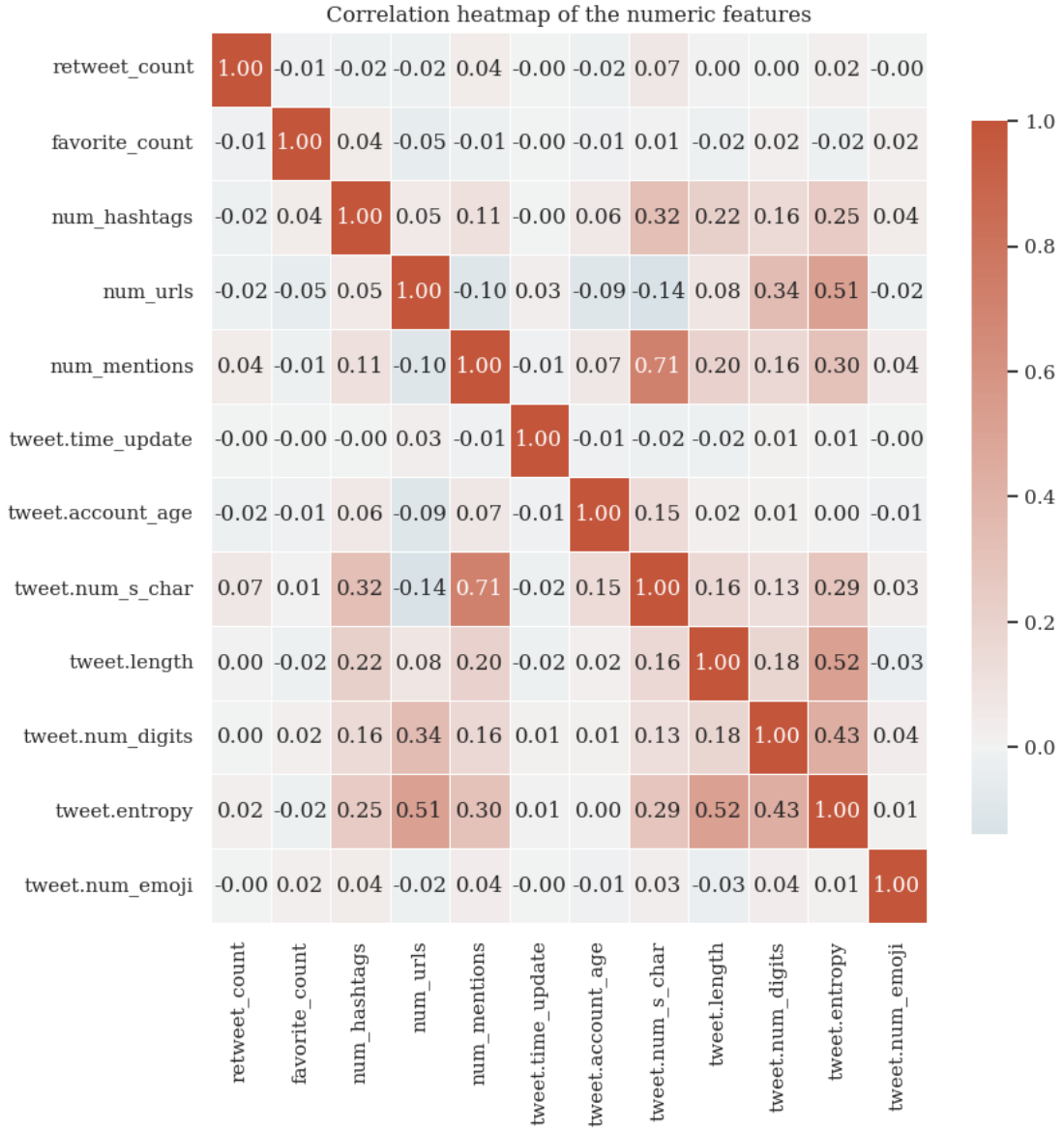
Η πολυσυγγραμμικότητα αντιπροσωπεύει την υψηλή γραμμική συσχέτιση που παρουσιάζουν δύο ή περισσότερες μεταβλητές σε ένα υπόδειγμα παλινδρόμησης, η οποία χρήζει εντοπισμού και διαχείρισης κατά το στάδιο της προεπεξεργασίας των δεδομένων σε προβλήματα μηχανικής μάθησης. Η πλειοψηφία των στατιστικών μεθόδων και των μεθόδων μηχανικής μάθησης αδυνατεί να προσαρμοστεί αποδοτικά σε έντονα γραμμικά συσχετισμένα δεδομένα. Παράλληλα, δυσχεραίνεται η ερμηνευσιμότητα των αποτελεσμάτων, αφού είναι αδύνατον να αξιολογηθεί η επίδραση μιας ανεξάρτητης μεταβλητής στην μεταβλητή στόχο, όταν δύο ή περισσότερες ανεξάρτητες μεταβλητές παρουσιάζουν γραμμικά συσχετισμένη

μεταβολή. Στην ουσία, δεν μπορεί να ποσοτικοποιηθεί το αποτύπωμα κάθε ανεξάρτητης μεταβλητής στην εξαρτημένη μεταβλητή, αφού η μεταβολή της τιμής μιας ανεξάρτητης μεταβλητής επηρεάζει εγγενώς την έτερη συσχετισμένη μεταβλητή, αναιρώντας την υπόθεση περί σταθερότητας. Επίσης, οι έντονα συσχετισμένες μεταβλητές συνιστούν πλεονασμό στο σετ δεδομένων, εισάγοντας υψηλότερη διαστατικότητα και πολυπλοκότητα στο σετ δεδομένων, χωρίς ουσιαστική συνεισφορά στην βελτίωση της απόδοσης του μοντέλου.

Σε στατιστικά μοντέλα και μοντέλα μηχανικής μάθησης που στηρίζονται σε γραμμικές υποθέσεις και προωθούν την ερμηνευσιμότητα των αποτελεσμάτων, αναζητείται η ένταση και η κατεύθυνση της γραμμικής συσχέτισης μεταξύ των ανεξάρτητων μεταβλητών, και βάσει των σχετικών αποτελεσμάτων αφαιρούνται ή διατηρούνται οι μεταβλητές αυτές στο σετ δεδομένων. Παραδείγματα τέτοιων μοντέλων είναι η Γραμμική Παλινδρόμηση, η Λογιστική Παλινδρόμηση, οι Μηχανές Διανυσμάτων Στήριξης και τα Νευρωνικά Δίκτυα, ορισμένα από τα οποία εξετάζονται και στην παρούσα μελέτη περίπτωσης. Αντίθετα, η αφαίρεση των πολυσυγγραμμικών μεταβλητών από σετ δεδομένων που εκπαιδεύουν μοντέλα της οικογένειας των Δέντρων Απόφασης, όπως το Random Forest και τα Gradient Boost Trees, δεν αποτελεί απαραίτητη προϋπόθεση.

Στην μελέτη πολυσυγγραμμικότητας των ανεξάρτητων μεταβλητών ενός σετ δεδομένων, υπολογίζεται ο συντελεστής γραμμικής συσχέτισης, γνωστός και ως συντελεστής Pearson. Ο συντελεστής Pearson λαμβάνει τιμές στο εύρος μεταξύ του -1 και του +1, με την θετική τιμή για ένα ζεύγος μεταβλητών να υποδηλώνει πως η αύξηση της τιμής της μιας μεταβλητής οδηγεί σε αύξηση της συσχετισμένης μεταβλητής σε βαθμό ευθέως ανάλογο της τιμής του, και την αρνητική τιμή να υποδηλώνει πως η αύξηση της τιμής της μιας μεταβλητής οδηγεί σε μείωσης της συσχετισμένης μεταβλητής σε βαθμό ευθέως ανάλογο της τιμής του. Αν η τιμή του συντελεστή Pearson προσεγγίζει το μηδέν, έπεται πως οι μεταβλητές είναι γραμμικά ασυσχέτιστες, ενώ αν η τιμή προσεγγίζει τα άκρα του διαστήματος δυνατών τιμών, έπεται απόλυτη γραμμική συσχέτιση μεταξύ των μεταβλητών. Καταχρηστικά, ένα ζεύγος μεταβλητών θεωρείται απόλυτα θετικά και αρνητικά συσχετισμένο για τιμές του συντελεστή Pearson μεγαλύτερες του 0.9 και -0.9, αντίστοιχα. Ομοίως, ισχυρή θετική και αρνητική γραμμική συσχέτιση νοείται για τιμές του συντελεστή Pearson που κυμαίνονται στα διαστήματα 0.7 εώς 0.9 και -0.7 έως -0.9, αντίστοιχα. Φυσικά, η απουσία γραμμικής συσχέτισης σε ένα ζεύγος μεταβλητών δεν αποκλείει την ύπαρξη άλλου τύπου συσχέτισης, όπως μονοτονική ή διατεταγμένη συσχέτιση.

Η ποσοτικοποιημένη ένταση της συσχέτισης παρουσιάζεται μέσω του πίνακα συσχετίσεων, ο οποίος αποτελεί έναν τετραγωνικό πίνακα με το κάθε κελί να αντιπροσωπεύει την τιμή της έντασης της συσχέτισης μεταξύ των αντίστοιχων μεταβλητών. Η γραμμική συσχέτιση εξετάζεται αποκλειστικά για ποσοτικές μεταβλητές, δηλαδή αριθμητικές μεταβλητές που λαμβάνουν τιμές σε ένα συνεχές ή διακριτό διάστημα. Για την οπτικοποίηση χρησιμοποιείται ένας χάρτης θερμότητας, με την διαφάνεια χρωματισμού σε κάθε κελί του πίνακα να είναι ευθέως ανάλογη της έντασης της γραμμικής συσχέτισης μεταξύ των αντίστοιχων μεταβλητών. Στην περίπτωση της μελέτης περίπτωσης, τα αποτελέσματα των οποίων συνοψίζονται στο Σχήμα 3, με τις αποχρώσεις του κόκκινου και του μπλε να αναπαριστούν θετικές και αρνητικές συσχετίσεις, αντίστοιχα. Στα παρακάτω σημεία, παρουσιάζονται κάποια βασικά συμπεράσματα που προκύπτουν από την μελέτη του χάρτη θερμότητας.



Σχήμα 3 - Πίνακας Συσχέτισης Αριθμητικών Μεταβλητών

- Ο συντελεστής γραμμικής συσχέτισης για τις ανεξάρτητες μεταβλητές του πλήθους των ονομαστικών αναφορών και του πλήθος των ειδικών χαρακτήρων σε ένα tweet είναι 0.69. Η τιμή αυτή υποδηλώνει μέτρια προς ισχυρή ένταση θετικής συσχέτισης μεταξύ των δύο μεταβλητών, δηλαδή η αύξηση του πλήθους των ονομαστικών αναφορών σε ένα tweet τείνει να οδηγεί σε αύξηση του πλήθους των ειδικών χαρακτήρων. Δεδομένου ότι η εισαγωγή ονομαστικής αναφοράς σε ένα tweet απαιτεί την εισαγωγή κατάλληλου ειδικού συμβόλου που κατατάσσεται στην κατηγορία των ειδικών χαρακτήρων, ερμηνεύεται ως αναμενόμενη. Συνδυάζοντας την φύση και την ένταση της συσχέτισης, κρίνεται πως δεν αποτελεί πηγή πλεονασμού για τα μοντέλα, καθώς μόνο ένα μέρος της ανεξάρτητης μεταβλητής του πλήθους των ειδικών χαρακτήρων σε ένα tweet μπορεί να ερμηνευθεί μέσω του πλήθους των ονομαστικών αναφορών, κατά συνέπεια η διατήρηση αμφότερων των γνωρισμάτων μπορεί να συνεισφέρει στην βελτίωση της προσαρμογής των μοντέλων της μελέτης περίπτωσης.
- Ο συντελεστής γραμμικής συσχέτισης μεταξύ της εντροπίας της πληροφορίας και του μήκους χαρακτήρων ενός tweet λαμβάνει τιμή ίση με 0.53, με την αντίστοιχη τιμή για την εντροπία της πληροφορίας και του πλήθους των URLs ενός tweet να είναι ίση με 0.51. Αμφότερες οι τιμές του συντελεστή Pearson υποδηλώνουν μέτριας έντασης θετική γραμμική συσχέτιση μεταξύ των γνωρισμάτων. Τα αποτελέσματα αυτά οδηγούν στο συμπέρασμα πως η συνεισφορά της εντροπίας της πληροφορίας ερμηνεύεται σε μεγάλο βαθμό από τα παραπάνω γνωρίσματα, συνεπώς απαιτείται περεταίρω διερεύνηση του γνωρίσματος αυτού με μέτρα περιγραφικής στατιστικής για την επίδρασή του στην τιμή της μεταβλητής στόχου του προβλήματος.
- Ο συντελεστής γραμμικής συσχέτισης μεταξύ της εξαρτημένης μεταβλητής του τύπου χρήστη του Twitter, δηλαδή του user_type, και των ανεξάρτητων μεταβλητών του πλήθους των ονομαστικών αναφορών και των ειδικών χαρακτήρων σε ένα tweet παρουσιάζουν χαμηλής προς μέτριας έντασης αρνητική συσχέτιση, λαμβάνοντας τιμές -0.48 και -0.40, αντίστοιχα. Οι κατηγορηματικές μεταβλητές, όπως η εξαρτημένη μεταβλητή του μελέτης περίπτωσης, δεν μπορούν να ερμηνεύσουν την κατεύθυνση και την ένταση της συσχέτισης, καθώς λαμβάνουν τιμές σε ένα διακριτό σύνολο. Αξίζει να σημειωθεί πως τα γνωρίσματα που αντιπροσωπεύουν αυτές οι ανεξάρτητες μεταβλητές, αποτελούν εκείνα τα γνωρίσματα με την πλέον αξιοσημείωτη επίδραση στην μεταβλητή στόχο του

προβλήματος, ενώ οι τιμές αυτές υποδεικνύουν αδυναμία αποδοτικής εφαρμογής γραμμικών προβλεπτικών μοντέλων.

## 4.3. EDA Ανάλυση για την Μελέτη Περίπτωσης

Η EDA ανάλυση χρησιμοποιεί μέτρα περιγραφικής στατιστικής και βασικούς στατιστικούς δείκτες που βοηθούν στην βαθύτερη κατανόηση των δεδομένων και στην εξαγωγή μακροσκοπικών συμπερασμάτων από αυτά.

### 4.3.1. Δείκτες Περιγραφικής Στατιστικής

Οι δείκτες περιγραφικής στατιστικής αποτελούν βασικούς περιγραφικούς στατιστικούς δείκτες, ενδεικτικούς της κεντρικής τάσης και της διασποράς των παρατηρήσεων ενός δείγματος, της συχνότητας και της κατανομής τους, καθώς και των ακραίων τιμών. Συνήθως, για τη απεικόνιση της κατανομής των τιμών των παρατηρήσεων μιας συνεχούς μεταβλητής χρησιμοποιούνται θηκογράμματα, ενώ για διακριτές μεταβλητές χρησιμοποιούνται διαγράμματα συχνότητας.

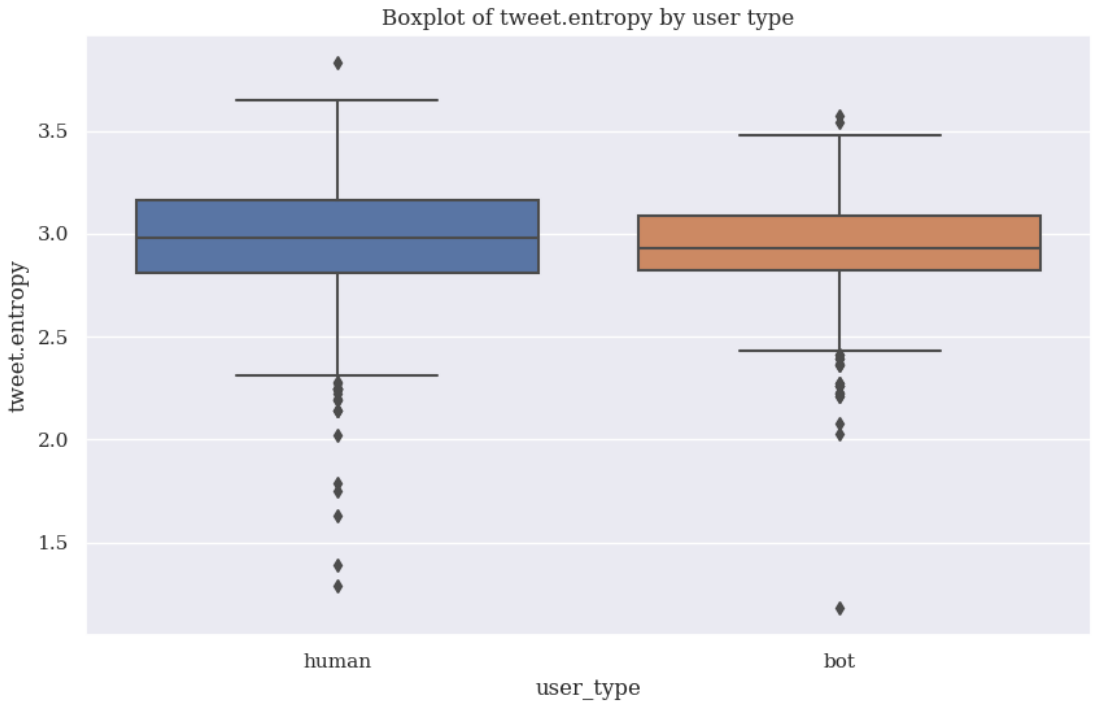
**1. Εντροπία Πληροφορίας (tweet.entropy)**

Οι δείκτες περιγραφικής στατιστικής για την ανεξάρτητη μεταβλητή της εντροπίας της πληροφορίας ενός tweet με βάση τον τύπο χρήστη παρουσιάζονται στον Πίνακας 5.

| user_type | mean | std | min | 25% | 50% | 75% | max |
|---|---|---|---|---|---|---|---|
| bot | 2.94 | 0.21 | 1.18 | 2.82 | 2.93 | 3.09 | 3.57 |
| human | 2.97 | 0.26 | 1.29 | 2.81 | 2.98 | 3.16 | 3.83 |

Πίνακας 5 - Δείκτες περιγραφικής στατιστικής για την εντροπία πληροφορίας

Παρατηρούμε πως η μέση τιμή της εντροπίας της πληροφορίας μεταξύ πραγματικών χρηστών και bots κυμαίνεται στα ίδια επίπεδα, και συγκεκριμένα, λαμβάνει τιμή 2.97 και 2.94, αντίστοιχα. Οι τιμές της διαμέσου της εντροπίας πληροφορίας σχεδόν ταυτίζονται, με τιμή 2.98 για τους πραγματικούς χρήστες και 2.93 για τα bots. Από την συνδυαστική αξιοποίηση των ποσοστιμορίων 25% και 75% με την μέγιστη και ελάχιστη παρατήρηση στο δείγμα ανά χρήστη, προκύπτει πως δεν υπάρχουν σημαντικά ακραίες τιμές στο δείγμα.




Σχήμα 4 - Θηκόγραμμα εντροπίας πληροφορίας

Οπτικοποιώντας την πληροφορία αυτή μέσω του boxplot του Σχήμα 4, επιβεβαιώνεται η ομοιόμορφη κατανομή της εντροπίας της πληροφορίας μεταξύ των δύο τύπων χρηστών. Ωστόσο, στην περίπτωση των bots, η κατανομή των ακραίων τιμών της εντροπίας της πληροφορίας συμπιέζεται κοντά στο πρώτο και το τρίτο ποσοστιμόριο, σε αντίθεση με την αντίστοιχη των πραγματικών χρηστών που παρουσιάζουν μεγαλύτερη διασπορά.

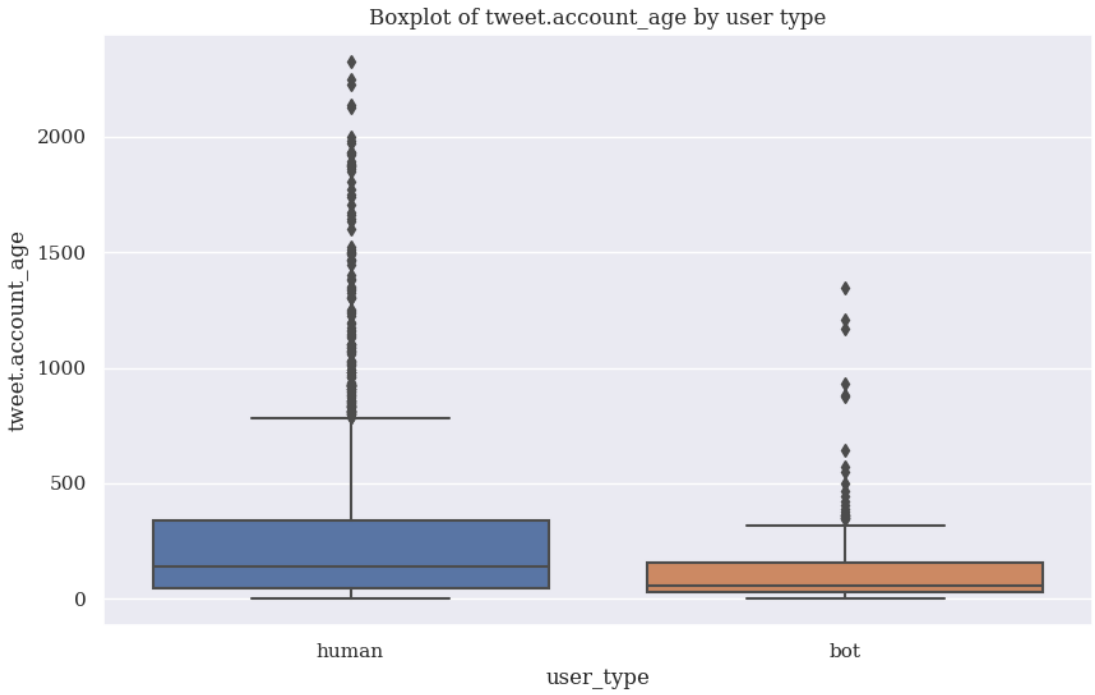
### 2. Ηλικία Λογαριασμού (tweet.account_age)

Οι δείκτες περιγραφικής στατιστικής για την ανεξάρτητη μεταβλητή της ηλικίας του λογαριασμού με βάση τον τύπο χρήστη παρουσιάζονται στον Πίνακας 6.

| user_type | mean | std | min | 25% | 50% | 75% | max |
|---|---|---|---|---|---|---|---|
| bot | 99.91 | 100.64 | 0.14 | 30.16 | 58.40 | 154.86 | 1345.06 |
| human | 260.32 | 338.62 | 0.31 | 44.32 | 138.67 | 339.94 | 2323.64 |

Πίνακας 6 - Δείκτες περιγραφικής στατιστικής για την ηλικία του λογαριασμού

Παρατηρούμε πως η μέση ηλικία ενός λογαριασμού στο Twitter που διαχειρίζεται από άνθρωπο έναντι λογαριασμού που διαχειρίζεται από bot είναι κατά 2.6 φορές υψηλότερη. Συγκεκριμένα, η μέση ηλικία ενός λογαριασμού πραγματικού χρήστη είναι 260.32 ημέρες, ενώ η αντίστοιχη των bots είναι μόλις 99.91 ημέρες.




Σχήμα 5 - Θηκόγραμμα ηλικίας λογαριασμού

Από την κατανομή των τιμών στα ποσοστιμόρια 25%, 50% και 75%, η οποία οπτικοποιείται μέσω του θηκογράμματος του Σχήμα 5, παρατηρούμε πως οι πραγματικοί χρήστες παρουσιάζουν έκτοπες τιμές προς τα επάνω, δηλαδή ηλικίες λογαριασμού αρκετά μεγαλύτερες από το 75% των παρατηρήσεων στο δείγμα. Συνεπώς, μπορούμε να εικάσουμε πως η υψηλή ηλικία λογαριασμού στην πλατφόρμα αποτελεί ισχυρή ένδειξη πως ένας χρήστης του Twitter ανήκει στην κλάση των πραγματικών χρηστών.

### 3. Μήκος Tweet (tweet.length)

Οι δείκτες περιγραφικής στατιστικής για την ανεξάρτητη μεταβλητή του μήκους του tweet σε χαρακτήρες με βάση με τον τύπο χρήστη παρουσιάζονται στον Πίνακας 7.

| user_type | mean | std | min | 25% | 50% | 75% | max |
|---|---|---|---|---|---|---|---|
| bot | 80.36 | 30.00 | 8.0 | 60.0 | 77.0 | 103.0 | 150.0 |
| human | 81.47 | 38.51 | 6.0 | 48.0 | 77.0 | 117.0 | 148.0 |

Πίνακας 7 - Δείκτες περιγραφικής στατιστικής για το μήκος ενός tweet σε χαρακτήρες

Στην περίπτωση του γνωρίσματος του μήκους του tweet σε χαρακτήρες, παρατηρούμε πως η μέγιστη τιμή είναι 150 και 148 χαρακτήρες για πραγματικούς χρήστες και bots, αντίστοιχα. Οι 150 χαρακτήρες αποτελούσαν εγγενή περιορισμό της πλατφόρμας κατά τη περίοδο συλλογής των δεδομένων, ήτοι το 2017. Επιπλέον, οι πραγματικοί χρήστες

αναρτούν περιεχόμενο με μόλις 1.11 περισσότερους χαρακτήρες ανά tweet συγκριτικά με τα bots, δηλωτικό πως και οι δύο κατηγορίες χρηστών ακολουθούν κοινό μοτίβο σύνταξης περιεχομένου. Η διάμεσος της ανεξάρτητης μεταβλητής και στις δύο περιπτώσεις ταυτίζεται, λαμβάνοντας τιμή 77, ενισχύοντας την εικασία περί κοινού μοτίβου στο περιεχόμενο των tweets. Το ποσοστιμόριο 25% τοποθετείται στους 60 χαρακτήρες για τα bots και στους 48 χαρακτήρες για τους πραγματικούς χρήστες, υποδηλώνοντας μια ελαφρά τάση των πραγματικών χρηστών να αναρτούν συντομότερα tweets.

### 4. Πλήθος Ψηφίων στο Tweet (tweet.num_digits)

Οι δείκτες περιγραφικής στατιστικής για την ανεξάρτητη μεταβλητή του πλήθους των ψηφίων στο tweet με βάση τον τύπο χρήστη παρουσιάζονται στον Πίνακας 8.

| user_type | mean | std | min | 25% | 50% | 75% | max |
|---|---|---|---|---|---|---|---|
| bot | 0.70 | 1.32 | 0.0 | 0.0 | 0.0 | 1.0 | 13.0 |
| human | 1.11 | 2.11 | 0.0 | 0.0 | 0.0 | 2.0 | 50.0 |

Πίνακας 8 - Δείκτες περιγραφικής στατιστικής για το πλήθος των ψηφίων σε ένα tweet

Παρατηρούμε πως τα bots εισάγουν κατά μέσο όρο 0.70 ψηφία σε ένα tweet με τυπική απόκλιση 1.32, ενώ οι πραγματικοί χρήστες 1.11 ψηφία με τυπική απόκλιση 2.11. Έτσι, συμπεραίνεται πως τα bots έχουν μια ελαφρώς ασθενέστερη τάση να περιλαμβάνουν ψηφία στα tweets τους. Επίσης, αξίζει να σημειωθεί πως η διάμεσος και για τις δύο κατηγορίες χρηστών τοποθετείται στο 0, δηλαδή το 50% των tweets και για τις δύο κατηγορίες χρηστών δεν περιλαμβάνει κανένα ψηφίο.

### 5. Πλήθος Ειδικών Χαρακτήρων στο Tweet (tweet.num_s_char)

Οι δείκτες περιγραφικής στατιστικής για την ανεξάρτητη μεταβλητή του πλήθους των ειδικών χαρακτήρων στο tweet με βάση τον τύπο χρήστη παρουσιάζονται στον Πίνακας 9.

| user_type | mean | std | min | 25% | 50% | 75% | max |
|---|---|---|---|---|---|---|---|
| bot | 0.19 | 0.40 | 0.0 | 0.0 | 0.0 | 0.0 | 1.0 |
| human | 0.69 | 0.46 | 0.0 | 0.0 | 1.0 | 1.0 | 1.0 |

Πίνακας 9 - Δείκτες περιγραφικής στατιστικής για το πλήθος των ειδικών χαρακτήρων σε ένα tweet

Παρατηρούμε πως οι πραγματικοί χρήστες εμφανίζουν 3.6 φορές υψηλότερο μέσο πλήθος ειδικών χαρακτήρων ανά tweet, και συγκεκριμένα 0.69 ειδικούς χαρακτήρες ανά tweet έναντι 0.40 από τα bots. Επίσης, στο δείγμα δεν παρατηρήθηκε πάνω από

ένας ειδικός χαρακτήρας ανά tweet για κάθε κατηγορία χρηστών, όπως προκύπτει από την μέγιστη τιμή για κάθε κατηγορία χρήστη. Ωστόσο, σημειώνεται πως η αναζήτηση ειδικών χαρακτήρων δεν έλαβε υπόψιν εκείνους που αντιπροσωπεύουν ονομαστικές αναφορές, emoji και σημεία στίξης. Από την θέση των ποσοστιμορίων, παρατηρείται πως το 75% των παρατηρήσεων του δείγματος των bots δεν περιλαμβάνουν ειδικό χαρακτήρα, σε αντίθεση με το ποσοστιμόριο 50% που τοποθετείται στην θέση μηδέν για τους πραγματικούς χρήστες.

### 6. Πλήθος Emoji στο Tweet (tweet.num_emoji)

Οι δείκτες περιγραφικής στατιστικής για την ανεξάρτητη μεταβλητή του πλήθους των emoji στο tweet με βάση τον τύπο χρήστη παρουσιάζονται στον Πίνακας 10.

| user_type | mean | std | min | 25% | 50% | 75% | max |
|---|---|---|---|---|---|---|---|
| bot | 0.00 | 0.02 | 0.0 | 0.0 | 0.0 | 0.0 | 1.0 |
| human | 0.02 | 0.19 | 0.0 | 0.0 | 0.0 | 0.0 | 4.0 |

Πίνακας 10 - Δείκτες περιγραφικής στατιστικής για το πλήθος emoji στα tweet

Παρατηρούμε πως το πλήθος των emoji ανά tweet για τα bots είναι μηδέν, ενώ η μέγιστη τιμή των emoji που παρατηρήθηκε στο δείγμα των bots είναι μόλις 1. Ωστόσο, η στρογγυλοποίηση έχει γίνει στο δεύτερο δεκαδικό ψηφίο, που σημαίνει πως ενδέχεται να είναι ελαφρώς υψηλότερη του μηδενός. Αντίθετα, σημαντικά υψηλότερη μέγιστη τιμή παρατηρήθηκε για τους πραγματικούς χρήστες, και συγκεκριμένα 4 emoji σε ένα tweet, με μία ανεπαίσθητα υψηλότερη μέση τιμή, της τάξης του 0.02. Έτσι, συμπεραίνουμε πως μικρός αριθμός πραγματικών χρηστών του δείγματος χρησιμοποιεί δυσανάλογα υψηλό πλήθος emoji στα tweets του. Και για τις δύο κλάσεις χρηστών, το 75% των tweets δεν περιλαμβάνουν emoji, όπως προκύπτει από τις σχετικές θέσεις των ποσοστιμορίων.

### 7. Πλήθος Ονομαστικών Αναφορών στο Tweet (num_mentions)

Οι δείκτες περιγραφικής στατιστικής για την ανεξάρτητη μεταβλητή του πλήθους των ονομαστικών αναφορών στο tweet με βάση τον τύπο χρήστη παρουσιάζονται στον Πίνακας 11.

| user_type | mean | std | min | 25% | 50% | 75% | max |
|---|---|---|---|---|---|---|---|
| bot | 0.17 | 0.44 | 0.0 | 0.0 | 0.0 | 0.0 | 6.0 |
| human | 0.80 | 0.86 | 0.0 | 0.0 | 1.0 | 1.0 | 7.0 |

Πίνακας 11 - Δείκτες περιγραφικής στατιστικής για το πλήθος ονομαστικών αναφορών σε ένα tweet

Παρατηρούμε πως η μέση τιμή των ονομαστικών αναφορών ανά tweet για τους πραγματικούς χρήστες είναι 4.7 φορές υψηλότερη συγκριτικά με εκείνη των bots. Παράλληλά, το 75% των παρατηρήσεων αναφορικά με το πλήθος ονομαστικών αναφορών ανά tweet για τα bots είναι μηδέν, με την τιμή αυτή να τοποθετείται στο ποσοστιμόριο 25% για τα πραγματικούς χρήστες. Έτσι, εικάζεται πως η ονομαστική αναφορά σε ένα tweet είναι κατά βάση γνώρισμα των πραγματικών χρηστών.

**8. Πλήθος URLs στο Tweet (num_urls)**

Οι δείκτες περιγραφικής στατιστικής για την ανεξάρτητη μεταβλητή του πλήθους των URLs στο tweet με βάση τον τύπο χρήστη παρουσιάζονται στον Πίνακας 12.

| user_type | mean | std | min | 25% | 50% | 75% | max |
|---|---|---|---|---|---|---|---|
| bot | 0.26 | 0.44 | 0.0 | 0.0 | 0.0 | 1.0 | 2.0 |
| human | 0.16 | 0.38 | 0.0 | 0.0 | 0.0 | 0.0 | 2.0 |

Πίνακας 12 - Πίνακας περιγραφικής στατιστικής για το πλήθος των URLs σε ένα Tweet

Παρατηρούμε πως το μέθοδος πλήθος των URLs σε ένα tweet για τα bots είναι υψηλότερο από το αντίστοιχο των πραγματικών χρηστών. Ο μέσος πραγματικός χρήστης του δείγματος χρησιμοποιεί 0.16 URLs ανά tweet, ενώ το μέσο bot χρησιμοποιεί 0.26 URLs ανά tweet. Και για τις δύο κατηγορίες χρηστών, η διάμεσος τοποθετείται στο μηδέν. Ωστόσο, στο 25% των παρατηρήσεων στην κλάση των bots καταγράφεται τιμή μεγαλύτερη ή ίση του 1, υποδηλώνοντας πως τα bots τείνουν να εμφανίζουν συχνότερα και μεγαλύτερη ένταση URLs στα tweets τους.

## 4.4. Ανάπτυξη Ταξινομητών Μηχανικής Μάθησης

### 4.4.1. Ταξινομητής Δέντρου Απόφασης

#### 4.4.1.1. Ρύθμιση Υπερπαραμέτρων του Ταξινομητή Δέντρου Απόφασης

Για την βέλτιστη ρύθμιση των υπερπαραμέτρων του ταξινομητή Δέντρου Απόφασης πραγματοποιήθηκε αναζήτηση πλέγματος με χρήση διασταυρούμενης επικύρωσης στα δεδομένα εκπαίδευσης. Μέσω της εφαρμογής μιας πενταπλής διασταυρούμενης επικύρωσης, προέκυψαν από τα δεδομένα εκπαίδευσης πέντε διαφορετικοί και τυχαίοι συνδυασμοί σετ δεδομένων εκπαίδευσης και επικύρωσης. Τα τέσσερα από τα πέντε αυτά σετ αξιοποιήθηκαν για την προσαρμογή των υπερπαραμέτρων του μοντέλου και το ένα από τα πέντε για την αξιολόγηση της προβλεπτικής ικανότητάς του. Κατά την διαδικασία αυτή, διερευνήθηκε ένα πλέγμα τιμών για τρείς υπερπαραμέτρους, και συγκεκριμένα του μέγιστου βάθους (max_depth), του ελάχιστου πλήθους δειγμάτων

για διαχωρισμό εσωτερικού κόμβου (min_samples_split) και του ελάχιστου πλήθους δειγμάτων για κόμβο φύλλο (min_samples_leaf), όπως παρουσιάζονται στον Πίνακας 13.

| **Υπερπαράμετρος** | | **Τιμές Πλέγματος** | | | |
|---|---|---|---|---|---|
| Μέγιστο βάθος | max_depth | 5 | 10 | 15 | 20 |
| Ελάχιστο πλήθος δειγμάτων για διαχωρισμό εσωτερικού κόμβου | min_samples_split | 2 | 5 | 10 | 20 |
| Ελάχιστο πλήθος δειγμάτων για κόμβο φύλλο | min_samples_leaf | 1 | 2 | 5 | 10 |

Πίνακας 13 - Πλέγμα υπερπαραμέτρων για το μοντέλο Δέντρου Απόφασης

Κατά την διασταυρούμενη επικύρωση των παραπάνω υπερπαραμέτρων, το μοντέλο παρουσίασε την βέλτιστη προσαρμογή σε Δέντρο Απόφασης για μέγιστο βάθος ίσο με 10, ελάχιστο πλήθος δειγμάτων για διαχωρισμό εσωτερικού κόμβου ίσο με 20 και ελάχιστου πλήθος δειγμάτων για κόμβο φύλλο ίσο με 5. Η ακρίβεια του ταξινομητή Δέντρων Απόφασης για τις παραπάνω τιμές των υπερπαραμέτρων προέκυψε ίση με 0.86 ή 86%, η οποία συνιστά ένα πολύ υψηλό επίπεδο απόδοσης.

4.4.1.2. Προβλεπτική Ικανότητα του μοντέλου Δέντρου Απόφασης

Η προβλεπτική ικανότητα του μοντέλου Δέντρου Απόφασης στην ταξινόμηση των χρηστών του Twitter ποσοτικοποιείται μέσω των μετρικών σφάλματος της ακρίβειας, της ανάκλησης και του F1-score. Στον Πίνακας 14, παρουσιάζεται ο πίνακας σύγχυσης του ταξινομητή Δέντρου Απόφασης, ενώ στον Πίνακας 15 παρουσιάζεται η αναφορά ταξινόμησης για το μοντέλο αυτό.

| | **Predicted Human** | **Predicted Bot** |
|---|---|---|
| **Actual Human** | 388 | 49 |
| **Actual Bot** | 45 | 270 |

Πίνακας 14 - Πίνακας σύγχυσης για την ταξινομητή του Δέντρου Απόφασης

| | **Ακρίβεια** | **Ανάκληση** | **F1-score** |
|---|---|---|---|
| **Human** | 0.90 | 0.89 | 0.89 |
| **Bot** | 0.85 | 0.86 | 0.85 |

Πίνακας 15 - Αναφορά ταξινόμησης για τον ταξινομητή Δέντρου Απόφασης

Για την κλάση Human, προέκυψε ακρίβεια ίση με 0.90, δηλαδή το 90% των περιπτώσεων που προβλέφθηκαν ως Human ανήκαν πράγματι στην κλάση αυτή. Από την μετρική της ανάκλησης προέκυψε πως το μοντέλο αναγνωρίζει με ακρίβεια 89%

τις πραγματικών περιπτώσεις της κλάσης Human. Το F1-score προέκυψε ίσο με 0.89, υποδηλώνοντας μια καλή ισορροπία μεταξύ ακρίβειας και ανάκλησης.

Για την κλάση Bots, προέκυψε ακρίβεια ίση με 0.85, δηλαδή το 85% των περιπτώσεων που προβλέφθηκαν ως Bots ανήκαν πράγματι στην κλάση αυτή. Από την μετρική της ανάκλησης προέκυψε πως το μοντέλο αναγνωρίζει με ακρίβεια 89% τις πραγματικών περιπτώσεις της κλάσης Bot. Το F1-score προέκυψε ίσο με 0.89, υποδηλώνοντας μια ελαφρά μικρότερη ισορροπία μεταξύ ακρίβειας και ανάκλησης συγκριτικά με την κλάση Human.

Συνολικά, το μοντέλο επιδεικνύει σταθερή και υψηλή απόδοση τόσο για τις ταξινομήσεις πραγματικών χρηστών όσο και για τις ταξινομήσεις bots.

4.4.1.3. Σημαντικότητα Γνωρισμάτων του μοντέλου Δέντρων Απόφασης

Η επίδραση κάθε επιμέρους γνωρίσματος του σετ δεδομένων στην επίδοση του μοντέλου των Δέντρων Απόφασης, ποσοτικοποιείται και υπολογίζεται μέσω του δείκτη σημαντικότητας, όπως παρουσιάζονται στον Πίνακας 16. Από τα αποτελέσματα προκύπτει πως μόνο δύο γνωρίσματα έχουν αξιοσημείωτο αντίκτυπο στην προβλεπτική ικανότητα του μοντέλου. Πρόκειται για την ηλικία του λογαριασμού και το πλήθος των ειδικών χαρακτήρων στο tweet με δείκτες σημαντικότητας ίσους με 0.376 και 0.335, αντίστοιχα. Από εκεί και πέρα, τα υπόλοιπα γνωρίσματα του σετ δεδομένων παρουσιάζουν αρκετές φορές μικρότερο δείκτη σημαντικότητας, χωρίς να προκύπτει κάποια αξιόλογη παρατήρηση. Τέλος, η σημαντικότητα των γνωρισμάτων του πλήθους των hashtags και των emoji στο tweet είναι μηδενική, δηλαδή δεν έχουν συνεισφορά στην πρόβλεψη.

| **Μεταβλητή Γνωρίσματος** | **Δείκτης Σημαντικότητας** |
|---|---|
| tweet.account_age | 0.376 |
| tweet.num_s_char | 0.335 |
| favorite_count | 0.075 |
| tweet.length | 0.073 |
| retweet_count | 0.054 |
| tweet.entropy | 0.042 |
| num_urls | 0.023 |
| num_mentions | 0.020 |
| tweet.num_digits | 0.002 |
| num_hashtags | 0.000 |

| | |
|---|---|
| tweet.num_emoji | 0.000 |

Πίνακας 16 – Δείκτες σημαντικότητας γνωρισμάτων για τον ταξινομητή του Δέντρου Απόφασης

### 4.4.2. Ταξινομητής Random Forest

#### 4.4.2.1. Ρύθμιση Υπερπαραμέτρων του Ταξινομητή Random Forest

Για την βέλτιστη ρύθμιση των υπερπαραμέτρων του ταξινομητή Random Forest πραγματοποιήθηκε αναζήτηση πλέγματος με χρήση διασταυρούμενης επικύρωσης στα δεδομένα εκπαίδευσης. Μέσω της εφαρμογής μιας πενταπλής διασταυρούμενης επικύρωσης, προέκυψαν από τα δεδομένα εκπαίδευσης πέντε διαφορετικοί και τυχαίοι συνδυασμοί σετ δεδομένων εκπαίδευσης και επικύρωσης. Τα τέσσερα από τα πέντε αυτά σετ αξιοποιήθηκαν για την προσαρμογή των υπερπαραμέτρων του μοντέλου και το ένα από τα πέντε για την αξιολόγηση της προβλεπτικής ικανότητάς του. Κατά την διαδικασία αυτή, διερευνήθηκε ένα πλέγμα τιμών για τέσσερεις υπερπαραμέτρους, και συγκεκριμένα του πλήθος εκτιμητών (n_estimators), του μέγιστου βάθους (max_depth), του μέγιστου πλήθους γνωρισμάτων (max_features) και του ελάχιστου πλήθους δειγμάτων για διαχωρισμού εσωτερικού κόμβου (min_samples_split) όπως παρουσιάζονται στον Πίνακας 17.

| **Υπερπαράμετρος** | | **Τιμές Πλέγματος** | | | |
|---|---|---|---|---|---|
| Πλήθος εκτιμητών | n_estimators | 50 | 75 | 100 | 125 |
| Μέγιστο βάθος | max_depth | 2 | 5 | 10 | 12 |
| Μέγιστο πλήθος γνωρισμάτων | max_features | auto | sqrt | log2 | None |
| Ελάχιστο πλήθος δειγμάτων για διαχωρισμό | min_samples_split | 2 | 5 | 7 | 9 |

Πίνακας 17 - Πλέγμα υπερπαραμέτρων για το μοντέλο Random Forest

Κατά την διασταυρούμενη επικύρωση των παραπάνω υπερπαραμέτρων, το μοντέλο παρουσίασε βέλτιστη επίδοση για πλήθος 100 εκτιμητών με μέγιστο βάθος ίσο με 12, μέγιστο πλήθος γνωρισμάτων ίσο με το πλήθος των γνωρισμάτων του σετ εκπαίδευσης και ελάχιστο πλήθος δειγμάτων για διαχωρισμό ίσο με 9. Η αξιολόγηση κατά την επικύρωση του μοντέλου για τις παραπάνω παραμέτρους ήταν 0.88 ή 88%, υποδεικνύοντας υψηλό επίπεδο απόδοσης.

#### 4.4.2.2. Προβλεπτική Ικανότητα του Random Forest

Η προβλεπτική ικανότητα του μοντέλου Random Forest στην ταξινόμηση των χρηστών του Twitter ποσοτικοποιείται μέσω των μετρικών σφάλματος της ακρίβειας, της

ανάκλησης και του F1-score. Στον Πίνακας 18 παρουσιάζεται ο πίνακας σύγχυσης του ταξινομητή Random Forest, ενώ στον Πίνακας 19 παρουσιάζεται η αναφορά ταξινόμησης για το μοντέλο αυτό.

| | **Predicted Human** | **Predicted Bot** |
|---|---|---|
| **Actual Human** | 386 | 51 |
| **Actual Bot** | 39 | 276 |

Πίνακας 18 - Πίνακας σύγχυσης για την ταξινομητή Random Forest.

| | **Ακρίβεια** | **Ανάκληση** | **F1-score** |
|---|---|---|---|
| **Human** | 0.91 | 0.88 | 0.90 |
| **Bot** | 0.84 | 0.88 | 0.86 |

Πίνακας 19 - Αναφορά ταξινόμησης για τον ταξινομητή Random Forest

Για την κλάση Human, προέκυψε ακρίβεια ίση με 0.91, δηλαδή το 91% των περιπτώσεων που προβλέφθηκαν ως Human ανήκαν πράγματι στην κλάση αυτή. Από την μετρική της ανάκλησης προέκυψε πως το μοντέλο αναγνωρίζει με ακρίβεια 88% τις πραγματικών περιπτώσεις της κλάσης Human. Το F1-score προέκυψε ίσο με 0.90, υποδηλώνοντας μια καλή ισορροπία μεταξύ ακρίβειας και ανάκλησης.

Για την κλάση Bots, προέκυψε ακρίβεια ίση με 0.84, δηλαδή το 84% των περιπτώσεων που προβλέφθηκαν ως Bots ανήκαν πράγματι στην κλάση αυτή. Από την μετρική της ανάκλησης προέκυψε πως το μοντέλο αναγνωρίζει με ακρίβεια 88% τις πραγματικών περιπτώσεις της κλάσης Bot. Το F1-score προέκυψε ίσο με 0.86, υποδηλώνοντας μια ελαφρά μικρότερη ισορροπία μεταξύ ακρίβειας και ανάκλησης συγκριτικά με την κλάση Human.

Συνολικά, το μοντέλο επιδεικνύει σταθερή και πολύ υψηλή απόδοση τόσο για τις ταξινομήσεις πραγματικών χρηστών όσο και για τις ταξινομήσεις bots.

4.4.2.3. Σημαντικότητα Γνωρισμάτων του Random Forest

Σχετικά με την επίδραση των μεμονωμένων γνωρισμάτων στην επίδοση του μοντέλου, τα αποτελέσματα παρουσιάζονται στον Πίνακας 20. Η ηλικία του λογαριασμού του χρήστη παρουσιάζει την σημαντικότερη επίδραση στο μοντέλο με τιμή ίση με 0.351, σχεδόν τρείς φορές μεγαλύτερη από το πλήθος των ειδικών χαρακτήρων στο tweet και το μήκος του tweet, με τιμές 0.121 και 0.120, αντίστοιχα.

| **Γνώρισμα** | **Σημαντικότητα** |
|---|---|
| tweet.account_age | 0.351 |

| | |
|---|---|
| tweet.num_s_char | 0.121 |
| tweet.length | 0.120 |
| num_mentions | 0.098 |
| retweet_count | 0.089 |
| tweet.entropy | 0.086 |
| favorite_count | 0.066 |
| tweet.num_digits | 0.025 |
| num_hashtags | 0.022 |
| num_urls | 0.016 |
| tweet.num_emoji | 0.001 |

Πίνακας 20 - Σημαντικότητα γνωρισμάτων για το μοντέλο Random Forest

### 4.4.3. Ταξινομητής XGBoost

#### 4.4.3.1. Ρύθμιση Υπερπαραμέτρων XGBoost

Για την βέλτιστη ρύθμιση των υπερπαραμέτρων του ταξινομητή Random Forest πραγματοποιήθηκε αναζήτηση πλέγματος με χρήση διασταυρούμενης επικύρωσης στα δεδομένα εκπαίδευσης. Μέσω της εφαρμογής μιας πενταπλής διασταυρούμενης επικύρωσης, προέκυψαν από τα δεδομένα εκπαίδευσης πέντε διαφορετικοί και τυχαίοι συνδυασμοί σετ δεδομένων εκπαίδευσης και επικύρωσης. Τα τέσσερα από τα πέντε αυτά σετ αξιοποιήθηκαν για την προσαρμογή των υπερπαραμέτρων του μοντέλου και το ένα από τα πέντε για την αξιολόγηση της προβλεπτικής ικανότητάς του. Κατά την διαδικασία αυτή, διερευνήθηκε ένα πλέγμα τιμών για πέντε υπερπαραμέτρους, και συγκεκριμένα, του ρυθμού εκμάθησης (learning_rate), του μέγιστου βάθους (max_depth), του πλήθους εκτιμητών (n_estimators), του μεγέθους υποδείγματος (subsample) και του μεγέθους των γνωρισμάτων (col_subsample), όπως παρουσιάζεται στον Πίνακα 1.

| **Υπερπαράμετρος** | | **Τιμές Πλέγματος** | | | |
|---|---|---|---|---|---|
| Ρυθμός Εκμάθησης | learning_rate | 0.01 | 0.02 | 0.05 | 0.10 |
| Μέγιστο Βάθος | max_depth | 3 | 5 | 7 | 9 |
| Πλήθος Εκτιμητών | n_estimators | 25 | 50 | 75 | - |
| Μέγεθος Υποδείγματος | subsample | 0.25 | 0.50 | 0.75 | - |
| Μέγεθος Γνωρισμάτων | col_subsample | 0.25 | 0.50 | 0.75 | - |

Πίνακας 21 - Πλέγμα υπερπαραμέτρων για το μοντέλο XGBoost

Κατά την διασταυρούμενη επικύρωση των παραπάνω υπερπαραμέτρων, το μοντέλο XGBoost παρουσίασε βέλτιστη επίδοση για πλήθος 50 εκτιμητών με ρυθμό εκμάθησης 0.02, μέγιστο βάθος ίσο με 9, μέγεθος υποδείγματος 0.5 και μέγεθος γνωρισμάτων 0.75. Η αξιολόγηση κατά την επικύρωση του μοντέλου για τις παραπάνω παραμέτρους ήταν 0.80 ή 80%, υποδεικνύοντας υψηλό επίπεδο απόδοσης.

4.4.3.2. Προβλεπτική Ικανότητα του XGBoost

Η προβλεπτική ικανότητα του μοντέλου XGBoost στην ταξινόμηση των χρηστών του Twitter ποσοτικοποιείται μέσω των μετρικών σφάλματος της ακρίβειας, της ανάκλησης και του F1-score. Στον Πίνακας 22 παρουσιάζεται ο πίνακας σύγχυσης του ταξινομητή Δέντρου Απόφασης, ενώ στον Πίνακας 23 παρουσιάζεται η αναφορά ταξινόμησης για το μοντέλο αυτό.

| | **Predicted Human** | **Predicted Bot** |
|---|---|---|
| **Actual Human** | 391 | 46 |
| **Actual Bot** | 36 | 279 |

Πίνακας 22 - Πίνακας σύγχυσης για τον ταξινομητή XGBoost.

| | **Ακρίβεια** | **Ανάκληση** | **F1-score** |
|---|---|---|---|
| **Human** | 0.92 | 0.89 | 0.91 |
| **Bot** | 0.86 | 0.89 | 0.87 |

Πίνακας 23 - Αναφορά ταξινόμησης για τον ταξινομητή XGBoost

Για την κλάση Human, προέκυψε ακρίβεια ίση με 0.92, δηλαδή το 92% των περιπτώσεων που προβλέφθηκαν ως Human ανήκαν πράγματι στην κλάση αυτή. Από την μετρική της ανάκλησης προέκυψε πως το μοντέλο αναγνωρίζει με ακρίβεια 89% τις πραγματικών περιπτώσεις της κλάσης Human. Το F1-score προέκυψε ίσο με 0.91, υποδηλώνοντας μια καλή ισορροπία μεταξύ ακρίβειας και ανάκλησης.

Για την κλάση Bots, προέκυψε ακρίβεια ίση με 0.86, δηλαδή το 86% των περιπτώσεων που προβλέφθηκαν ως Bots ανήκαν πράγματι στην κλάση αυτή. Από την μετρική της ανάκλησης προέκυψε πως το μοντέλο αναγνωρίζει με ακρίβεια 89% τις πραγματικών περιπτώσεις της κλάσης Bot. Το F1-score προέκυψε ίσο με 0.87, υποδηλώνοντας μια ελαφρά μικρότερη ισορροπία μεταξύ ακρίβειας και ανάκλησης συγκριτικά με την κλάση Human.

Συνολικά, το μοντέλο επιδεικνύει σταθερή και υψηλή απόδοση τόσο για τις ταξινομήσεις πραγματικών χρηστών όσο και για τις ταξινομήσεις bots.

#### 4.4.3.3. Σημαντικότητα Γνωρισμάτων του XGBoost

Σχετικά με την επίδραση των μεμονωμένων γνωρισμάτων στην επίδοση του μοντέλου, τα αποτελέσματα παρουσιάζονται στον Πίνακας 24. Το πλήθος των ειδικών χαρακτήρων στο tweet εμφανίζει σημαντικότητα 0.445, τρείς φορές υψηλότερη συγκριτικά με το επόμενο κατά σειρά γνώρισμα, εκείνο του πλήθους των ονομαστικών αναφορών στο tweet.

| **Γνώρισμα** | **Σημαντικότητα** |
|---|---|
| tweet.num_s_char | 0.445 |
| num_mentions | 0.143 |
| favorite_count | 0.119 |
| tweet.account_age | 0.066 |
| retweet_count | 0.059 |
| tweet.num_emoji | 0.036 |
| num_hashtags | 0.033 |
| num_urls | 0.030 |
| tweet.length | 0.027 |
| tweet.entropy | 0.024 |
| tweet.num_digits | 0.019 |

Πίνακας 24 - Σημαντικότητα γνωρισμάτων για το μοντέλο XGBoost

### 4.4.4. Ταξινομητής Λογιστικής Παλινδρόμησης

#### 4.4.4.1. Ρύθμιση Υπερπαραμέτρων Μοντέλου Λογιστικής Παλινδρόμησης

Για την αναζήτηση και την ρύθμιση των βέλτιστων υπερπαραμέτρων του μοντέλου Λογιστικής Παλινδρόμησης πραγματοποιήθηκε αναζήτηση πλέγματος με χρήση διασταυρούμενης επικύρωσης. Πιο συγκεκριμένα, εφαρμόστηκε πενταπλή διασταυρούμενη επικύρωση, δηλαδή το σετ δεδομένων εκπαίδευσης χωρίστηκε σε πέντε επιμέρους υποσύνολα εκπαίδευσης και επικύρωσης, αξιολογούμενο στο ένα από αυτά. Κατά την διαδικασία αυτή, διερευνήθηκε ένα πλέγμα τιμών για τις υπερπαραμέτρους του βάρους κανονικοποίησης και του τύπου ποινής, όπως παρουσιάζεται στον Πίνακας 25.

| Βάρος Κανονικοποίησης | C | 0.01 | 0.1 | 1 | 10 |
|---|---|---|---|---|---|
| Τύπος Ποινής | penalty | L1 | L2 | - | - |

Πίνακας 25 - Πλέγμα υπερπαραμέτρων για το μοντέλο Λογιστικής Παλινδρόμησης

Κατά την διασταυρούμενη επικύρωση των παραπάνω υπερπαραμέτρων, το μοντέλο Λογιστικής Παλινδρόμησης παρουσίασε την βέλτιστη επίδοση για βάρος κανονικοποίησης ίσο με 1 και για τύπο ποινής L1. Η αξιολόγηση κατά την επικύρωση του μοντέλου για τις παραπάνω παραμέτρους ήταν 0.80 ή 80%, υποδεικνύοντας υψηλό επίπεδο απόδοσης.

4.4.4.2. Προβλεπτική Ικανότητα του Μοντέλου Λογιστικής Παλινδρόμησης

Η προβλεπτική ικανότητα του μοντέλου Λογιστικής Παλινδρόμησης στην ταξινόμηση ενός χρήστη του Twitter μεταξύ πραγματικού χρήστη και bot, παρουσιάζεται μέσω του πίνακα σύγχυσης και των βασικών μετρικών σφάλματος στους Πίνακες 1, 2.

| | **Predicted Human** | **Predicted Bot** |
|---|---|---|
| **Actual Human** | 353 | 84 |
| **Actual Bot** | 58 | 257 |

Πίνακας 26 - Πίνακας σύγχυσης για την ταξινομητή Λογιστικής Παλινδρόμησης

| | **Ακρίβεια** | **Ανάκληση** | **F1-score** |
|---|---|---|---|
| **Human** | 0.86 | 0.81 | 0.83 |
| **Bot** | 0.75 | 0.82 | 0.78 |

Πίνακας 27 - Αναφορά Ταξινόμησης για τον Ταξινομητή Λογιστικής Παλινδρόμησης

Για την κλάση Human, προέκυψε ακρίβεια ίση με 0.86, δηλαδή το 86% των περιπτώσεων που προβλέφθηκαν ως Human ανήκαν πράγματι στην κλάση αυτή. Από την μετρική της ανάκλησης προέκυψε πως το μοντέλο αναγνωρίζει με ακρίβεια 81% τις πραγματικών περιπτώσεις της κλάσης Human. Το F1-score προέκυψε ίσο με 0.83, υποδηλώνοντας μια καλή ισορροπία μεταξύ ακρίβειας και ανάκλησης.

Για την κλάση Bots, προέκυψε ακρίβεια ίση με 0.75, δηλαδή το 75% των περιπτώσεων που προβλέφθηκαν ως Bots ανήκαν πράγματι στην κλάση αυτή. Από την μετρική της ανάκλησης προέκυψε πως το μοντέλο αναγνωρίζει με ακρίβεια 82% τις πραγματικών περιπτώσεις της κλάσης Bot. Το F1-score προέκυψε ίσο με 0.78, υποδηλώνοντας μια ελαφρά μικρότερη ισορροπία μεταξύ ακρίβειας και ανάκλησης συγκριτικά με την κλάση Human.

Συνολικά, το μοντέλο επιδεικνύει υψηλή απόδοση στον εντοπισμό πραγματικών χρηστών, αλλά σημαντικά χαμηλότερη απόδοση αναφορικά με την ταξινόμηση των bots.

4.4.4.3. Σημαντικότητα Γνωρισμάτων του Μοντέλου Λογιστικής Παλινδρόμησης

Σχετικά με την επίδραση των μεμονωμένων γνωρισμάτων στην επίδοση του μοντέλου, υπολογίστηκαν οι συντελεστές παλινδρόμησης, οι οποίοι παρουσιάζονται στον Πίνακας 28. Οι θετικοί συντελεστές συνεπάγονται αύξηση της πιθανότητας για ταξινόμηση στην θετική κατηγορία, εν προκειμένου σε αυτή των bots, ενώ αρνητικοί συντελεστές την μειώνουν. Έτσι, οι θετικοί και υψηλοί συντελεστές παλινδρόμησης αναφορικά με το πλήθος των URLs σε ένα tweet και την εντροπία της πληροφορίας σε αυτό, υποδηλώνουν σημαντική συνεισφορά στην ταξινόμηση ενός χρήστη ως bots. Αντίθετα, οι αρνητικοί και υψηλοί συντελεστές παλινδρόμησης αναφορικά με το πλήθος των αντιδράσεων, των ονομαστικών αναφορών και του πλήθους των emoji αυξάνουν την πιθανότητα ταξινόμησης ενός χρήστη ως πραγματικού.

| **Γνώρισμα** | **Συντελεστής Παλινδρόμησης** |
|---|---|
| tweet.num_s_char | -0.703 |
| num_mentions | -1.223 |
| favorite_count | -1.298 |
| tweet.account_age | -0.005 |
| retweet_count | -0.001 |
| tweet.num_emoji | -1.780 |
| num_hashtags | -0.568 |
| num_urls | 0.246 |
| tweet.length | 0.008 |
| tweet.entropy | 0.216 |
| tweet.num_digits | -0.153 |

Πίνακας 28 – Συντελεστές παλινδρόμησης για το μοντέλο Λογιστικής Παλινδρόμησης

### 4.4.5. Ταξινομητής K-Nearest Neighbors

Για την βέλτιστη ρύθμιση των υπερπαραμέτρων του ταξινομητή των K-Nearest Neighbors πραγματοποιήθηκε αναζήτηση πλέγματος με χρήση διασταυρούμενης επικύρωσης στα δεδομένα εκπαίδευσης. Μέσω της εφαρμογής μιας πενταπλής διασταυρούμενης επικύρωσης, από τα δεδομένα εκπαίδευσης προέκυψαν πέντε διαφορετικοί και τυχαίοι συνδυασμοί σετ δεδομένων εκπαίδευσης και επικύρωσης. Τα τέσσερα από τα πέντε αυτά σετ αξιοποιήθηκαν για την προσαρμογή των υπερπαραμέτρων του μοντέλου και το ένα από τα πέντε για την αξιολόγηση της προβλεπτικής ικανότητάς του. Κατά την διαδικασία αυτή, διερευνήθηκε ένα πλέγμα τιμών αποκλειστικά για την υπερπαράμετρο του πλήθους των γειτόνων, όπως παρουσιάζεται στον Πίνακας 29.

| Υπερπαράμετρος | | Τιμές Πλέγματος | | | | | |
|---|---|---|---|---|---|---|---|
| Πλήθος γειτόνων | n_neighbors | 3 | 5 | 7 | 9 | 11 | 13 |

Πίνακας 29 - Πλέγμα υπερπαραμέτρων για τον ταξινομητή K-nearest neighbors

Κατά την διασταυρούμενη επικύρωση των παραπάνω υπερπαραμέτρων, το μοντέλο παρουσίασε την βέλτιστη προσαρμογή για πλήθος γειτόνων ίσο με 5. Η ακρίβεια του ταξινομητή K-nearest neighbors για τις παραπάνω τιμές των υπερπαραμέτρων ήταν ίση με 0.79 ή 79%, η οποία συνιστά ένα οριακά αποδεκτό επίπεδο απόδοσης.

4.4.1.2. Προβλεπτική Ικανότητα του μοντέλου Δέντρου Απόφασης

Η προβλεπτική ικανότητα του μοντέλου των K-nearest neighbors στην ταξινόμηση των χρηστών του Twitter ποσοτικοποιείται μέσω των μετρικών σφάλματος της ακρίβειας, της ανάκλησης και του F1-score. Στον Πίνακας 30 παρουσιάζεται ο πίνακας σύγχυσης του μοντέλου των K-nearest neighbors που κατασκευάστηκε, ενώ στον Πίνακας 31 παρουσιάζεται η αναφορά ταξινόμησης για το μοντέλο αυτό.

| | **Predicted Human** | **Predicted Bot** |
|---|---|---|
| **Actual Human** | 320 | 117 |
| **Actual Bot** | 58 | 257 |

Πίνακας 30 - Πίνακας σύγχυσης για τον ταξινομητή Λογιστικής Παλινδρόμησης

| | **Ακρίβεια** | **Ανάκληση** | **F1-score** |
|---|---|---|---|
| **Human** | 0.84 | 0.73 | 0.78 |
| **Bot** | 0.69 | 0.81 | 0.74 |

Πίνακας 31 - Αναφορά ταξινόμησης για τον ταξινομητή Λογιστικής Παλινδρόμησης

Για την κλάση Human, προέκυψε ακρίβεια ίση με 0.84, δηλαδή το 84% των περιπτώσεων που προβλέφθηκαν ως Human ανήκαν πράγματι στην κλάση αυτή. Από την μετρική της ανάκλησης προέκυψε πως το μοντέλο αναγνωρίζει με ακρίβεια 73% τις πραγματικών περιπτώσεις της κλάσης Human. Το F1-score προέκυψε ίσο με 0.78, υποδηλώνοντας μια ελαφρά ανισορροπία μεταξύ ακρίβειας και ανάκλησης.

Για την κλάση Bots, προέκυψε ακρίβεια ίση με 0.69, δηλαδή το 69% των περιπτώσεων που προβλέφθηκαν ως Bots ανήκαν πράγματι στην κλάση αυτή. Από την μετρική της ανάκλησης προέκυψε πως το μοντέλο αναγνωρίζει με ακρίβεια 81% τις πραγματικών περιπτώσεις της κλάσης Bot. Το F1-score προέκυψε ίσο με 0.74, υποδηλώνοντας μια μεγαλύτερη ανισορροπία μεταξύ ακρίβειας και ανάκλησης συγκριτικά με την κλάση Human.

Συνολικά, το μοντέλο επιδεικνύει οριακά αποδεκτή απόδοση για τις ταξινομήσεις των πραγματικών χρηστών και μη-ικανοποιητική απόδοση για τις ταξινομήσεις των bots.

## 4.5. Αξιολόγηση Μοντέλων για την Μελέτη Περίπτωσης

Το μοντέλο XGBoost παρουσίασε την υψηλότερη συνολική ακρίβεια μεταξύ των ταξινομητών που εξετάστηκαν για την μελέτη περίπτωσης, προβλέποντας σωστά τον τύπο του λογαριασμού στο 89% των περιπτώσεων. Η ακρίβεια, η ανάκληση και το F1-score του ήταν επίσης τα υψηλότερα και για τις δύο κατηγορίες χρηστών μεταξύ των μοντέλων, ιδιαίτερα για την ταξινόμηση των πραγματικών χρηστών.

Τα μοντέλα Δέντρου Απόφασης και Random Forest παρουσίασαν αξιόλογες επιδόσεις, ανταγωνιζόμενα την επίδοση του μοντέλου XGBoost, με συνολική ακρίβεια 88%. Παρόλο που δεν έφτασαν την ίδια μέγιστη απόδοση με το XGBoost, επέδειξαν αποδοτικότητα και ισορροπημένη απόδοση στην ταξινόμηση των χρηστών. Το Random Forest έδειξε ελαφρώς καλύτερη ανάκληση για τα bots σε σύγκριση με το Δέντρο Απόφασης, αλλά χαμηλότερη ακρίβεια για την κατηγορία αυτή, γεγονός που υποδηλώνει ότι αναγνώρισε σωστά ένα υψηλό ποσοστό πραγματικών θετικών αποτελεσμάτων, αλλά παρήγαγε επίσης περισσότερα ψευδώς θετικά αποτελέσματα.

Το μοντέλο Λογιστικής Παλινδρόμησης είχε ελαφρώς χαμηλότερη συνολική ακρίβεια από τα παραπάνω μοντέλα, και συγκεκριμένα 81%. Ωστόσο, ήταν πιο ισορροπημένο μεταξύ ακρίβειας και ανάκλησης και για τις δύο κατηγορίες χρηστών σε σύγκριση με τα άλλα παραπάνω μοντέλα, αλλά παρατηρήθηκε σημαντική μείωση της ακρίβειας για την κατηγορία bot. Παρ' όλα αυτά, η απλότητα, η ερμηνευσιμότητα και οι γρήγοροι χρόνοι εκπαίδευσης θα μπορούσαν να το καταστήσουν ως μία αξιόλογη εναλλακτική επιλογή.

Το μοντέλο K-Nearest Neighbors είχε τη χαμηλότερη συνολική ακρίβεια μεταξύ των μοντέλων που εξετάστηκαν με 77%. Ενώ η ανάκληση του μοντέλου κυμάνθηκε σε ικανοποιητικά επίπεδα, ιδίως για την κατηγορία bot, η ακρίβειά του ήταν σημαντικά χαμηλότερη από τους υπόλοιπους ταξινομητές.

Για την ανίχνευση ανθρώπων, το μοντέλο XGBoost παρουσίασε την καλύτερη απόδοση με ακρίβεια 0,92, που σημαίνει ότι όταν προέβλεψε σωστά ότι ένας λογαριασμός διαχειρίζεται από άνθρωπο στο 92% των περιπτώσεων. Παράλληλα, η ανάκληση κυμάνθηκε σε 0,89, που μεταφράζεται σε ορθή αναγνώριση του 89% όλων των λογαριασμών που διαχειρίζονται από άνθρωπο. Το F1-score, το οποίο είναι ο

αρμονικός μέσος όρος της ακρίβειας και της ανάκλησης, ήταν 0,91, υποδεικνύοντας μια αξιοσημείωτη ισορροπία μεταξύ ακρίβειας και ανάκλησης.

Για τον εντοπισμό των bots, η απόδοση του μοντέλου παρουσίασε ελαφρώς χαμηλότερη απόδοση. Το XGBoost εξακολουθεί να ανιχνεύει bot καλύτερα από κάθε άλλο μοντέλο που εξετάστηκε όσων αφορά την ακρίβεια, προβλέποντας σωστά το 86% των περιπτώσεων bots. Ωστόσο, όταν εξετάζεται η ανάκληση, δηλαδή το πόσα από τα συνολικά bots αναγνωρίζει σωστά, το μοντέλο Random Forest παρουσιάζει ελαφρώς καλύτερη απόδοση από το XGBoost, με ανάκληση 0,88 έναντι 0,89.

Αναλύοντας την επίδραση των γνωρισμάτων στα τρία μοντέλα με τις καλύτερες επιδόσεις, ήτοι του XGBoost, του Δέντρου Απόφασης και του Random Forest, εκτιμάται η μηχανική των μοντέλων πίσω από τις προβλέψεις τους.

Το tweet.account_age είναι το πιο κρίσιμο χαρακτηριστικό τόσο για το μοντέλο Δέντρου Απόφασης όσο και για το μοντέλο Random Forest, τονίζοντας ότι η ηλικία ενός λογαριασμού Twitter είναι ένας παράγοντας με μεγάλη επιρροή στη διάκριση των bots από τους ανθρώπους. Αντίθετα, το μοντέλο XGBoost θεωρεί το tweet.num_s_char, δηλαδή το πλήθος των ειδικών χαρακτήρων σε ένα tweet, ως το πλέον επιδραστικό χαρακτηριστικό για την πρόβλεψη, το οποίο κατατάσσεται δεύτερο από τα μοντέλα πρόβλεψης του Δέντρου Απόφασης και του Random Forest.

Ένα βασικό εύρημα σε όλα τα μοντέλα σχετίζεται με την υψηλή σημασία των γνωρισμάτων που εξάγονται μέσα από το περιεχόμενο των tweets, και συγκεκριμένα του tweet.num_s_char και του num_mentions. Ένα άλλο αξιοσημείωτο μοτίβο είναι η χαμηλή σημασία του tweet.time_update σε όλα τα μοντέλα, το οποίο εικάζεται πως οφείλεται στο μικρό πλήθος μη-μηδενικών τιμών του συνόλου εκπαίδευσης αναφορικά με το γνώρισμα αυτί. Μια ενδιαφέρουσα παρατήρηση αφορά και στην διαφοροποίηση της επίδρασης του favorite_count, αφού το μοντέλο XGBoost και το μοντέλο Δέντρου Απόφασης το κατατάσσουν ως το τρίτο πιο σημαντικό χαρακτηριστικό σε αντιδιαστολή με το μοντέλο Random Forest που το κατατάσσει προς το τέλος. Έτσι, υπογραμμίζεται ο τρόπος με τον οποίο οι διαφορετικοί αλγόριθμοι μπορούν να ερμηνεύσουν τη σημασία του ίδιου χαρακτηριστικού με διαφορετικούς τρόπους.

# 5. Συζήτηση

Στην παρούσα Διπλωματική Εργασία, αξιολογήθηκε η απόδοση μιας σειράς μοντέλων ταξινόμησης με μηχανική μάθηση στο ερευνητικό αντικείμενο του εντοπισμού bots στο Twitter, τα οποία εκπαιδεύτηκαν πάνω σε πραγματικά δεδομένα χρηστών της πλατφόρμας. Στο σύνολό τους, οι μέθοδοι ταξινόμησης παρουσίασαν αξιοσημείωτη προβλεπτική ικανότητα, ειδικά όσων αφορά τον επιτυχή εντοπισμό πραγματικών χρηστών στην πλατφόρμα, ενώ ελαφρώς μειωμένη ήταν η απόδοσή τους αναφορικά με τον εντοπισμό των bots. Ως τα πλέον ισχυρά προβλεπτικά μοντέλα αναδείχθηκαν εκείνα που αξιοποιούν αδύναμους learners για την παραγωγή προβλέψεων που συναθροίζονται σε μια συνολική πρόβλεψη, όπως το μοντέλο XGBoost και το μοντέλο Random Forest.

Αν και τα αποτελέσματα της έρευνας θεμελιώνουν ένα ισχυρό υπόβαθρο στην ταξινόμηση χρηστών του Twitter, κρίνεται σκόπιμη η αναζήτηση τρόπων ενίσχυσης της προβλεπτικής ικανότητάς τους αναφορικά με την κλάση των bots που αξιοποιούν και προεκτείνουν τα αποτελέσματα της παρούσας έρευνας. Έτσι, μια πιθανή μελλοντική προέκταση της έρευνας κατευθύνεται προς την μηχανική γνωρισμάτων. Η εμβάθυνση στη μελέτη των χαρακτηριστικών των bots και η διαρκής αναζήτηση νέων συμπεριφορικών προτύπων που ακολουθούν μπορεί να εισφέρει σημαντικά στην βελτίωση της ακρίβειας των μοντέλων ταξινόμησης. Ακόμη, η ενσωμάτωση πληροφορίας αναφορικά με γνωρίσματα της δικτυακής δομής των bots και της γλωσσικού περιεχομένου των tweets του, κρίνονται απαραίτητα για την εκπαίδευση ισχυρότερων προβλεπτικών μοντέλων με έμφαση στην ορθή ταξινόμηση των bots.

Παράλληλα, η συνδυαστική αξιοποίηση των υφιστάμενων μοντέλων, και συγκεκριμένα εκείνων που εμφάνισαν την υψηλότερη απόδοση στον εντοπισμό των bots κρίνεται σκόπιμο να μελετηθεί ως μελλοντική προέκταση της παρούσας έρευνας. Έτσι, μια συνδυαστική προσέγγιση μεταξύ των μοντέλων του XGBoost, του Random Forest και του Δέντρου Απόφασης για την παραγωγή μιας συνολικής πρόβλεψης μέσω της απόδοσης κατάλληλων βαρών στις επιμέρους προβλέψεις ή της συνάθροισης με την πλειοψηφική πρόβλεψη, ενδέχεται να εμφάνιζε ακόμη υψηλότερη προβλεπτική ικανότητα. Πέρα από τους ταξινομητές επιτηρούμενης μάθησης που διερευνήθηκαν στην παρούσα εργασία, η μελέτη της απόδοσης μεθόδων βαθιάς μάθησης, όπως τα Νευρωνικά Δίκτυα, εικάζεται ότι μπορεί προσφέρει βαθύτερη γνώση σχετικά με τη δομή και τη σημασιολογία της φυσικής γλώσσας των tweets, και κατά συνέπεια, στην αποτελεσματικότερη ανίχνευση bots στο Twitter.

Τέλος, κρίνεται σκόπιμη η εφαρμογή του μοντέλου σε δεδομένα πραγματικού χρόνου στο Twitter ώστε να αξιολογηθεί η προβλεπτική ικανότητα των μοντέλων σε ένα εκτεταμένο σύνολο δεδομένων, βοηθώντας παράλληλα και την εξερεύνηση της μεταβολής των συμπεριφορικών μοτίβων και μοντέλων, αναπροσαρμόζοντας τα υφιστάμενα μοντέλα στις τρέχουσες τάσεις των bots.

# Παράρτημα – Σχήματα



# Παράρτημα – Πίνακες

# Βιβλιογραφία

A. Rauchfleisch, J. K. (2020). The False positive problem of automatic bot detection in social science research . *Plos One*.

Botometer. (2023, Ιούνιος 7). *Bot Repository*. Ανάκτηση από Botometer: https://botometer.osome.iu.edu/bot-repository/

David A. Kirsch, M. A. (2023). Fanbois and Fanbots: Tesla's Entrepreneurial Narratives and Corporate Computational Propaganda on Social Media. *World Electr. Veh*.

E. Zhuravskaya, M. P. (2020). Political Effects of the Internet and Social Media. *Annual Review of Economics*.

F. L. De Faveri, L. C. (2023). Twitter Bots Influence on the Russo-Ukrainian War During the 2022 Italian General Elections.

Ferrara, E. (2022). Twitter spam and false accounts prevalence, detection, and characterization: A survey. *First Monday*.

L. Nizzoli, S. T. (2020). Charting the Landscape of Online Cryptocurrency Manipulation. *Proceedings of the ACM Internet Measurement Conference*. IEEE.

M. Jiang, P. C. (2016). Catching synchronized behaviors in large networks: A graph mining approach. *ACM Trans*.

M. Kantepe, M. G. (2017). Preprocessing framework for Twitter bot detection. *International Conference on Computer Science and Engineering.*

O. Kraaijeveld, J. D. (2020). The predictive power of public Twitter sentiment for forecasting cryptocurrency prices. *Journal of International Financial Markets, Institutions and Money*.

O. Varol, E. F. (χ.χ.). Online Human-Bot Interactions:Detection, Estimation, and Characterization. *Proceedings of the Eleventh International AAAI Conference on Web and Social Media* . 2017.

Phillip George Efthimion, S. P. (2018). Supervised Machine Learning Bot Detection Techniques to Identify Social Twitter Bots. *SMU Data Science Review*.

S. Giorgi, L. U. (2021). Characterizing Social Spambots by their Human Traits. *Findings 2021*.

Statista. (2023, Ιούνιος 8). *Most popular reasons for internet users worldwide to use social media as of 3rd quarter 2022* . Ανάκτηση από Statista: https://www.statista.com/statistics/715449/social-media-usage-reasons-worldwide/

Statista. (2023, Ιούνιος 5). *Social Media & User-Generated Content*. Ανάκτηση από Statista: https://www.statista.com/markets/424/topic/540/social-media-user-generated-content/#statistic2

W. Ahmed, J. V.-A. (2022). COVID-19 and the 5G Conspiracy Theory: Social Network Analysis of Twitter Data. *J Med Internet Res*.

Yuriy Gorodnichenko, T. P. (2018). Social Media, Sentiment, and Public Opinions: Evidence from #Brexit and #USElection. *National Bureau of Economic Research*.

Z. Weng, A. L. (2022). Public Opinion Manipulation on Social Media: Social Network Analysis of Twitter Bots during the COVID-19 Pandemic . *Int. J. Environ. Res. Public Health*.

Zhang Y, S. W. (2023). Social Bots' Role in the COVID-19 Pandemic Discussion on Twitter . *Int J Environ Res Public Health* .