



Αριστοτέλειο Πανεπιστήμιο Θεσσαλονίκης  
Πολυτεχνική Σχολή  
Τμήμα Ηλεκτρολόγων Μηχανικών &  
Μηχανικών Υπολογιστών  
Τομέας Ηλεκτρονικής και Υπολογιστών

## Διπλωματική Εργασία

---

Ανάπτυξη συστήματος συνεχούς έμμεσου  
ελέγχου ταυτότητας με βάση γλωσσικά και  
συμπεριφορικά χαρακτηριστικά

---

*Εκπόνηση:*

Μυλωνάς Κωνσταντίνος  
ΑΕΜ: 10027

*Επίβλεψη:*

Καθ. Συμεωνίδης Ανδρέας  
Υπ. Δρ. Μάλαμας Νικόλας

Θεσσαλονίκη, Δεκέμβριος 2024



*Two roads diverged in a wood, and I—  
I took the one less traveled by,  
And that has made all the difference.*  
— Robert Frost, *The Road Not Taken*

*Φτάσε όπου δέν μπορείς*  
— Ν. Καζαντζάκης



---

## ΕΥΧΑΡΙΣΤΙΕΣ

---

Η ολοκλήρωση της παρούσας διπλωματικής εργασίας αποτέλεσε ένα απαιτητικό αλλά εξίσου συναρπαστικό ταξίδι, το οποίο δεν θα ήταν εφικτό χωρίς τη στήριξη και τη βοήθεια πολλών ανθρώπων, στους οποίους θα ήθελα να εκφράσω την ειλικρινή μου ευγνωμοσύνη.

Πρώτα απ' όλα, θα ήθελα να ευχαριστήσω θερμά τον καθηγητή κ. Ανδρέα Συμεωνίδη για την εμπιστοσύνη που μου έδειξε αναθέτοντάς μου την εκπόνηση αυτής της εργασίας. Επιπλέον, ο τρόπος διδασκαλίας του, η μεταδοτικότητα του και η ικανότητά του να εμπνέει το ενδιαφέρον για το αντικείμενο αποτέλεσαν σημαντική πηγή έμπνευσης για την ενασχόλησή μου με το συγκεκριμένο πεδίο.

Ιδιαίτερης μνείας αξίζει και ο υποψήφιος διδάκτορας Νικόλας Μάλαμας για την καθοδήγηση και στήριξή του καθ' όλη τη διάρκεια εκπόνησης της εργασίας. Οι πολύτιμες συμβουλές και οι εύστοχες επισημάνσεις του συνέβαλαν καταλυτικά στη ολοκλήρωση αυτής της προσπάθειας.

Το πιο εγκάρδιο ευχαριστώ οφείλω στους γονείς μου, Θεόφιλο και Στέλλα, και στην αδερφή μου Ιωάννα, για την αμέριστη αγάπη, κατανόηση και υποστήριξη καθ' όλη τη διάρκεια των σπουδών μου· με ωθήσατε να κυνηγήσω τους στόχους μου με επιμονή και αποφασιστικότητα.

Τέλος, ευγνωμονώ συγγενείς, φίλους και γνωστούς για την όποια συνεισφορά τους στην προσωπική και επαγγελματική μου πορεία, καθώς και όλες τις όμορφες στιγμές που μοιραστήκαμε και έκαναν αυτή τη διαδρομή ξεχωριστή.



---

## Περίληψη

Η αυθεντικοποίηση χρηστών αποτελεί κρίσιμη πτυχή της ασφάλειας σε κάθε ψηφιακό περιβάλλον, με τις παραδοσιακές μεθόδους, όπως οι κωδικοί πρόσβασης, τα PIN και οι ερωτήσεις ασφαλείας, να παρουσιάζουν περιορισμούς όσον αφορά την απρόσκοπτη ενσωμάτωση και τη χρηστικότητα. Αντίθετα, ο έμμεσος έλεγχος ταυτότητας επιδιώκει τη συνεχή και διακριτική ταυτοποίηση των χρηστών με βάση τις μοναδικές συμπεριφορές τους, όπως το στυλ γραφής τους, επιτυγχάνοντας υψηλότερα επίπεδα ασφάλειας χωρίς να επηρεάζεται η ροή εργασίας τους. Ωστόσο, η προσέγγιση αυτή αντιμετωπίζει σημαντικές προκλήσεις, καθώς η γραφή ενός ατόμου μπορεί να επηρεάζεται από διάφορους παράγοντες, όπως η συναισθηματική κατάσταση, η κούραση ή ο τύπος του περιεχομένου που παράγει.

Η παρούσα διπλωματική εργασία επικεντρώνεται στην ανάπτυξη ενός συστήματος συνεχούς έμμεσης αυθεντικοποίησης με βάση τα γλωσσικά χαρακτηριστικά, το οποίο βρίσκει εφαρμογές στα μέσα κοινωνικής δικτύωσης, στο ηλεκτρονικό ταχυδρομείο και στα συστήματα συνομιλίας. Χρησιμοποιεί τεχνικές Επεξεργασίας Φυσικής Γλώσσας και Μηχανικής Μάθησης για την ταυτοποίηση των χρηστών μέσω της ανάλυσης των χαρακτηριστικών του γραπτού τους κειμένου. Το σύστημα αποσκοπεί στην ενίσχυση της ασφάλειας σε περιβάλλοντα όπου οι παραδοσιακές μέθοδοι ελέγχου ταυτότητας μπορεί να είναι ακατάλληλες ή μη επαρκείς. Η προτεινόμενη λύση χρησιμοποιεί μοντέλα μηχανών διανυσμάτων υποστήριξης μίας κλάσης για την εκμάθηση εξατομικευμένων μοτίβων γραφής, αξιοποιώντας χαρακτηριστικά όπως η λεξική μορφολογία, η χρήση λεξιλογίου και τα συντακτικά και σημασιολογικά πρότυπα.

Στο πλαίσιο αυτής της εργασίας, διεξήχθη εκτεταμένη ανάλυση των χαρακτηριστικών που διαχωρίζουν μοναδικά τους χρήστες μέσω της γραφής τους, συμπεριλαμβανομένης της συχνότητας χρήσης λέξεων, της σύνθεσης προτάσεων και στατιστικών μέτρων όπως δείκτες πολυπλοκότητας και ποσοστά παύσεων. Τα δεδομένα που συλλέχθηκαν από τους χρήστες ενσωματώθηκαν σε ένα σετ δεδομένων για την εκπαίδευση και την αξιολόγηση του μοντέλου. Το σύστημα αξιολογήθηκε με βάση μετρικές όπως το ποσοστό ψευδούς απόρριψης και το ποσοστό ψευδούς αποδοχής, ενώ πραγματοποιήθηκαν πρόσθετες δοκιμές για την αξιολόγηση της ανθεκτικότητάς του σε διάφορα περιβάλλοντα και σενάρια χρήσης.

Τα ευρήματα αυτής της έρευνας καταδεικνύουν ότι το σύστημα επιτυγχάνει υψηλά επίπεδα ακρίβειας και απόκρισης, ισορροπώντας αποτελεσματικά την ασφάλεια και τη χρηστικότητα. Εξαλείφοντας την ανάγκη για συνεχή παρέμβαση του χρήστη, το σύστημα προσφέρει μια διακριτική αλλά ενσωματωμένη μέθοδο για αυθεντικοποίηση, καθιστώντας το κατάλληλο για εφαρμογές σε ένα διαρκώς εξελισσόμενο ψηφιακό περιβάλλον.





---

# Title

## Development of a Continuous Implicit Authentication System Based on Linguistic and Behavioural Features

### Abstract

User authentication is a critical aspect of security in any digital environment, with traditional authentication methods, such as passwords, PINs and security questions, having limitations in terms of seamless integration and usability. In contrast, implicit authentication seeks to identify users continuously and discreetly based on their unique behaviours, such as their writing style, thus achieving higher levels of security without affecting their workflow. However, this approach faces significant challenges, as an individual's writing can be affected by a variety of factors, such as their emotional state, level of fatigue or the type of content they produce.

This thesis focuses on the development of a continuous implicit authentication system based on linguistic features, which finds applications in social media, email and chat systems. It uses Natural Language Processing and Machine Learning techniques to identify users by analyzing the features of their written text. The system aims to enhance security in environments where traditional authentication methods may be inappropriate or insufficient. The proposed solution uses one-class support vector machine models to learn personalized writing patterns by exploiting features such as lexical morphology, vocabulary usage, and syntactic and semantic patterns.

As part of this work, an extensive analysis of the features that can be used to uniquely identify users through their writing was conducted, including word usage frequency, sentence composition, and statistical measures, such as complexity index and pause ratio. The data collected from users was incorporated into a dataset for model training and evaluation. The system was evaluated using metrics such as false rejection rate and false acceptance rate, and additional tests were conducted to evaluate its robustness in different environments and usage scenarios.

The findings of this research demonstrate that the system can achieve high levels of accuracy and adaptability, while effectively balancing safety and usability. By eliminating the need for constant user intervention, the proposed system offers an unobtrusive yet integrated method for authentication, making it suitable for applications in an ever-evolving digital environment.

Konstantinos Mylonas  
Electrical & Computer Engineering Department,  
Aristotle University of Thessaloniki, Greece  
December 2024



# Περιεχόμενα

Ευχαριστίες . . . . .	iii
Περίληψη . . . . .	v
Abstract . . . . .	vii
Ακρωνύμια . . . . .	xv
<b>1 Εισαγωγή</b>	<b>1</b>
1.1 Περιγραφή του Προβλήματος . . . . .	2
1.2 Σκοπός - Συνεισφορά της Διπλωματικής Εργασίας . . . . .	4
1.3 Διάρθρωση της Αναφοράς . . . . .	5
<b>2 Επισκόπηση της Ερευνητικής Περιοχής</b>	<b>6</b>
<b>3 Θεωρητικό Υπόβαθρο</b>	<b>11</b>
3.1 Αυθεντικοποίηση . . . . .	11
3.1.1 Ορισμός Αυθεντικοποίησης . . . . .	11
3.1.2 Κατηγορίες Τεχνικών Αυθεντικοποίησης . . . . .	12
3.1.3 Συνεχής και Έμμεση Αυθεντικοποίηση βασιζόμενη σε συμπε- ριφορικές μεθόδους . . . . .	13
3.2 Επεξεργασία Φυσικής Γλώσσας . . . . .	15
3.2.1 Εισαγωγή στην Επεξεργασία Φυσικής Γλώσσας . . . . .	15
3.2.2 Εργαλεία και Μέθοδοι Εξαγωγής Χαρακτηριστικών . . . . .	16
3.3 Σύγχρονη Μηχανική Μάθηση . . . . .	17
3.3.1 Κατηγορίες Αλγορίθμων Μηχανικής Μάθησης . . . . .	18
3.3.2 Προκλήσεις . . . . .	20
3.3.3 Το SVM και το One-Class SVM . . . . .	20
<b>4 Μεθοδολογία και Υλοποίηση</b>	<b>27</b>
4.0.1 Αρχιτεκτονική Υψηλού Επιπέδου . . . . .	27
4.0.2 Εργαλεία και Πλατφόρμες Ανάπτυξης . . . . .	28
4.1 Συλλογή και Προεπεξεργασία Δεδομένων . . . . .	28
4.1.1 Πηγή Δεδομένων . . . . .	28
4.1.2 Διαδικασία Προεπεξεργασίας Δεδομένων . . . . .	29
4.1.3 Χαρακτηριστικά του Σετ Δεδομένων . . . . .	29
4.2 Εξαγωγή Χαρακτηριστικών . . . . .	30
4.2.1 Χαρακτηριστικά που Βασίζονται στους Χαρακτήρες . . . . .	31
4.2.2 Λεξιλογικά και Συντακτικά Χαρακτηριστικά . . . . .	32
4.2.3 Δείκτες Πολυπλοκότητας και Αναγνωσιμότητας . . . . .	32
4.2.4 Δομικοί και Στυλιστικοί Δείκτες . . . . .	33

4.2.5	Σύνοψη χαρακτηριστικών . . . . .	33
4.2.6	Εργαλεία και Μέθοδοι Εξαγωγής Χαρακτηριστικών . . . . .	34
4.3	Εκπαίδευση μοντέλων . . . . .	34
4.3.1	Προεπεξεργασία Δεδομένων Εκπαίδευσης . . . . .	35
4.3.2	Εξαγωγή και Κανονικοποίηση Χαρακτηριστικών . . . . .	35
4.3.3	Εκπαίδευση Μοντέλων <i>One-Class SVM</i> . . . . .	35
4.3.4	Hyperparameter Tuning και Threshold Tuning . . . . .	36
4.4	Σύστημα απόφασης . . . . .	37
4.4.1	Εισαγωγή . . . . .	37
4.4.2	Συνάρτηση Επιπέδου Βεβαιότητας . . . . .	37
4.4.3	Υποσύστημα Ψήφοφορίας . . . . .	37
4.4.4	Παραδείγματα Εφαρμογής . . . . .	38
4.5	Σύστημα Κλειδώματος/Εμπιστοσύνης . . . . .	39
4.5.1	Συνάρτηση Εμπιστοσύνης . . . . .	39
4.5.2	Παραδείγματα Λειτουργίας . . . . .	41
4.5.3	Παρατηρήσεις . . . . .	43
4.6	Παρουσίαση Διεπαφής Χρήστη . . . . .	43
4.6.1	Παραδείγματα Χρήσης . . . . .	44
5	<b>Πειράματα - Αποτελέσματα</b> . . . . .	47
5.1	Αξιολόγηση Συστήματος . . . . .	48
5.1.1	Εισαγωγή . . . . .	48
5.1.2	Δεδομένα Δοκιμών . . . . .	48
5.1.3	Μετρικές Αξιολόγησης . . . . .	48
5.1.4	Ροή Διαδικασίας Δοκιμών . . . . .	50
5.2	Πρώτη Φάση Πειραμάτων: 1 Μοντέλο Ανά Χρήστη . . . . .	52
5.2.1	Πειραματική Διαδικασία . . . . .	52
5.2.2	Hyperparameter & Threshold Tuning . . . . .	52
5.2.3	Αποτελέσματα Ανά Χρήστη με συγκεκριμένο πλέγμα . . . . .	54
5.2.4	Παρατηρήσεις . . . . .	56
5.3	Δεύτερη Φάση Πειραμάτων: Basic Majority Voting . . . . .	56
5.3.1	Πειραματική Διαδικασία . . . . .	56
5.3.2	Αποτελέσματα Ανά Χρήστη από τη Συνάρτηση Basic Majority Voting . . . . .	57
5.3.3	Παρατηρήσεις . . . . .	58
5.4	Τρίτη Φάση Πειραμάτων: Weighted Majority Voting . . . . .	58
5.4.1	Πειραματική Διαδικασία . . . . .	58
5.4.2	Αποτελέσματα . . . . .	59
5.4.3	Παρατηρήσεις . . . . .	62
5.5	Τέταρτη Φάση Πειραμάτων: Confidence Level Function . . . . .	62
5.5.1	Πειραματική Διαδικασία . . . . .	62
5.5.2	Αποτελέσματα . . . . .	63
5.5.3	Παρατηρήσεις . . . . .	68
5.6	Πέμπτη Φάση Πειραμάτων: LOSO Cross Validation . . . . .	68
5.6.1	Αποτελέσματα . . . . .	70
5.7	Συγκριτικά Αποτελέσματα . . . . .	73

<b>6</b>	<b>Συμπεράσματα</b>	<b>77</b>
6.1	Γενικά Συμπεράσματα . . . . .	77
6.2	Προβλήματα . . . . .	78
<b>7</b>	<b>Μελλοντικές επεκτάσεις</b>	<b>81</b>
	<b>Βιβλιογραφία</b>	<b>83</b>

# Κατάλογος Σχημάτων

3.1	Πτυχές της συνεχούς και έμμεσης αυθεντικοποίησης . . . . .	14
3.2	Κλάδοι και εφαρμογές της επιστήμης της Τεχνητής Νοημοσύνης . . .	17
3.3	Διάγραμμα Venn των διαφόρων κατηγοριών μηχανικής μάθησης . . .	19
3.4	Διαφορά classification & regression . . . . .	19
3.5	Πρώτο γράφημα: underfitting, Δεύτερο γράφημα: best fit, Τρίτο γράφημα: overfitting . . . . .	20
3.6	Λειτουργία του SVM - Κατασκευή hyperplane με το maximum margin	21
3.7	Διαχωρισμός δεδομένων με τη χρήση του αλγορίθμου SVM. Με μαύρο φαίνεται η ευθεία που μεγιστοποιεί το περιθώριο ενώ με κόκκινο και πράσινο φαίνονται άλλες - μη βέλτιστες - επιλογές ευθειών διαχωρισμού	21
3.8	Μη γραμμικά διαχωρίσιμα δεδομένα μετασχηματίζονται σε διαφορετικό χώρο υψηλής διάστασης . . . . .	22
3.9	Επίδειξη λειτουργίας OC-SVM σε δισδιάστατο χώρο . . . . .	23
4.1	Αρχιτεκτονική Υψηλού Επιπέδου του Συστήματος . . . . .	28
4.2	Στιγμιότυπο οθόνης από το σετ δεδομένων που χρησιμοποιήθηκε - φαίνονται οι 4 στήλες που αναφέρονται παραπάνω καθώς και πολλαπλές καταχωρίσεις . . . . .	30
4.3	Τα χαρακτηριστικά που εξάγονται στη παρούσα εργασία . . . . .	31
4.4	Συνάρτηση Confidence Level . . . . .	39
4.5	Περιβάλλον Χρήστη του Streamlit UI . . . . .	44
4.6	Στιγμιότυπα οθόνης από το streamlit UI για επιβεβαίωση και απόρριψη αυθεντικοποίησης . . . . .	45
5.1	Ποιοτική απεικόνιση των FAR και FRR. . . . .	50
5.2	FAR, FRR, F1 για διαφορετικά πλέγματα nu, gamma & threshold . .	54
5.3	FAR & FRR ανά χρήστη για nu: 0.01, gamma: 0.05 και threshold: 0.05	55
5.4	FAR & FRR ανά χρήστη με τη χρήση της Basic Majority Voting . . .	57
5.5	FAR ανά χρήστη για διαφορετικά πλέγματα παραμέτρων . . . . .	61
5.6	Mean Accepted Prompts Before Locking for Impostors ανά χρήστη για διαφορετικά πλέγματα παραμέτρων . . . . .	61
5.7	FAR, FRR, MAPBL-G, MAPBL-I για διαφορετικές τιμές των μεταβλητών της συνάρτησης Confidence Level . . . . .	66
5.8	FAR, FRR, MAPBL-G, MAPBL-I για διαφορετικές τιμές της υπερπαραμέτρου gamma . . . . .	68
5.9	Γραφήματα επεξήγησης τεχνικής LOSO-CV . . . . .	70
5.10	Αποτελέσματα ακρίβειας (Accuracy) ανά χρήστη για την τεχνική LOSO.	71

5.11 Αποτελέσματα μέσου απόλυτου σφάλματος (MAE) ανά χρήστη για την τεχνική LOSO. . . . .	72
5.12 Μέσος όρος Accuracy & MAE. . . . .	72
5.13 Σύγκριση FAR & FRR στο υποκεφάλαιο 5.2 και στο υποκεφάλαιο 5.4	74
5.14 Σύγκριση Mean Accepted Prompts Before Locking for Impostors στο υποκεφάλαιο 5.4 με το αρχικό πλέγμα υπερπαραμέτρων, στο υποκεφάλαιο 5.4 με το δεύτερο πλέγμα υπερπαραμέτρων και στο στο υποκεφάλαιο 5.5 με την ενσωμάτωση της συνάρτησης Confidence Level	75

# Κατάλογος πινάκων

4.1	Στατιστικά Χαρακτηριστικά του Dataset . . . . .	29
4.2	Περιγραφή Σηλών του Dataset . . . . .	30
4.3	Κατηγοριοποίηση Χαρακτηριστικών Εξαγωγής . . . . .	34
4.4	Παραδείγματα Ενημέρωσης Εμπιστοσύνης . . . . .	41
4.5	Παραδείγματα Εναλλαγής Γνήσιων και Απατηλών Αποφάσεων . . . . .	42
4.6	Παραδείγματα Ενεργοποίησης Μηχανισμού Κλειδώματος . . . . .	42
4.7	Παραδείγματα Συνεχούς Ενίσχυσης Λόγω Υψηλής Βεβαιότητας . . . . .	42
4.8	Παραδείγματα Επαναλαμβανόμενου Κλειδώματος . . . . .	43
5.1	Επιλεγμένα Αποτελέσματα Tuning Παραμέτρων με Ισορροπία FAR και FRR. . . . .	53
5.2	Αποτελέσματα Ανά Χρήστη με Threshold 0.05. . . . .	55
5.3	Αποτελέσματα από τη χρήση της Basic Majority Voting για διαφορετικούς χρήστες. . . . .	57
5.4	Αποτελέσματα για $\nu \in \{0.001, 0.005\}$ και $\gamma \in \{0.05, 0.1\}$ που δείχνουν τις μετρικές FAR, FRR και τον μέσο αριθμό αποδεκτών προτροπών από απατεώνες. . . . .	59
5.5	Αποτελέσματα για $\nu \in \{0.0001, 0.0005, 0.001\}$ και $\gamma \in \{0.05, 0.1, 0.5\}$ που δείχνουν τις μετρικές FAR, FRR και τον μέσο αριθμό αποδεκτών προτροπών από απατεώνες. . . . .	60
5.6	Αποτελέσματα για $\nu \in \{0.001, 0.005, 0.01\}$ και $\gamma \in \{0.05, 0.07, 0.1, 0.2, 0.5\}$ : FAR, FRR και Μέσος Αριθμός Αποδεκτών Προτροπών από Απατεώνες. . . . .	60
5.7	Αποτελέσματα πειραμάτων για την ενσωμάτωση της συνάρτησης <i>confidence level</i> . . . . .	63
5.8	Primary Results . . . . .	64
5.9	Tighten Variables' Values . . . . .	64
5.10	Maximized boost value in confidence function . . . . .	65
5.11	Relaxation of base increase and base decrease variables' values . . . . .	65
5.12	Relaxation of all values . . . . .	65
5.13	All gamma values . . . . .	66
5.14	Include 0.3 gamma value on the grid . . . . .	66
5.15	Include 0.15 and 0.3 on the gamma grid . . . . .	67
5.16	Include 0.15 on the gamma grid . . . . .	67
5.17	Remove 0.5 from the gamma grid . . . . .	67



# Ακρωνύμια Εγγράφου

Παρακάτω παρατίθενται ορισμένα από τα πιο συχνά χρησιμοποιούμενα ακρωνύμια της παρούσας διπλωματικής εργασίας:

AI	→ Artificial Intelligence
CIA	→ Continuous Implicit Authentication
FAR	→ False Acceptance Rate
FRR	→ False Rejection Rate
ML	→ Machine Learning
DL	→ Deep Learning
NLP	→ Natural Language Processing
OCSVM	→ One Class Support Vector Machine
OS	→ Operating System
SVM	→ Support Vector Machine
UI	→ User Interface



# 1

## Εισαγωγή

Η ασφάλεια στη σύγχρονη εποχή έχει καταστεί μείζον ζήτημα, καθώς οι ψηφιακές υπηρεσίες, οι εφαρμογές και οι συσκευές που χρησιμοποιούμε σε καθημερινή βάση αποθηκεύουν και επεξεργάζονται σημαντικό όγκο προσωπικών, επαγγελματικών και ευαίσθητων δεδομένων. Από τις τραπεζικές συναλλαγές μέχρι τις επικοινωνίες και από την εργασία μέχρι την ψυχαγωγία, οι άνθρωποι στηρίζονται σε ψηφιακά μέσα που απαιτούν ασφαλή πρόσβαση και προστασία από κακόβουλες επιθέσεις. Ο αυξανόμενος κίνδυνος παραβίασης δεδομένων, απάτης και ψηφιακής κατασκοπείας, έχει ενισχύσει την ανάγκη για ισχυρά και αξιόπιστα μέτρα ασφαλείας στον κυβερνοχώρο.

Μία από τις σημαντικότερες παραμέτρους της ασφαλείας των πληροφοριών είναι η αυθεντικοποίηση, δηλαδή η διαδικασία ταυτοποίησης ενός χρήστη πριν του επιτραπεί η πρόσβαση σε έναν πόρο ή σε μια υπηρεσία. Οι παραδοσιακές μέθοδοι αυθεντικοποίησης περιλαμβάνουν συνήθως τη χρήση κωδικών πρόσβασης, προσωπικών αναγνωριστικών (PIN) ή βιομετρικών δεδομένων, όπως δακτυλικά αποτυπώματα και αναγνώριση προσώπου. Παρόλο που αυτές οι μέθοδοι αποτελούν αναπόσπαστο μέρος της ψηφιακής ασφαλείας, παρουσιάζουν ορισμένα προβλήματα και περιορισμούς. Οι κωδικοί πρόσβασης, για παράδειγμα, μπορεί να είναι εύκολο να υποκλαπούν μέσω επιθέσεων phishing, ενώ οι χρήστες συχνά δυσκολεύονται να θυμούνται πολλούς και διαφορετικούς κωδικούς για κάθε πλατφόρμα. Επιπλέον, οι βιομετρικές μέθοδοι, παρότι είναι ισχυρότερες από τους κωδικούς πρόσβασης, ενδέχεται να μην είναι πάντοτε πρακτικές και δεν αναιρούν το γεγονός πως η εκάστοτε ψηφιακή υπηρεσία παραμένει εκτεθειμένη μετά από την πρώτη και μοναδική φορά αυθεντικοποίησης του χρήστη (one-time authentication).

Στις μέρες μας, μεγάλος όγκος πληροφορίας ανταλλάσσεται μέσω κειμένου, είτε με τη μορφή συνομιλιών (chats) και μηνυμάτων ηλεκτρονικού ταχυδρομείου (emails), είτε μέσω αναρτήσεων σε μέσα κοινωνικής δικτύωσης. Αυτή η τάση φανερώνει τη σημασία της ανάλυσης του γραπτού λόγου στη κατανόηση της αλληλεπίδρασης μεταξύ χρηστών και συστημάτων. Ταυτόχρονα, οι επιθέσεις με τη χρήση bots, φεύ-

τικών προφίλ και τεχνικών όπως deepfake εντείνονται διαρκώς, αναδεικνύοντας την ανάγκη για πιο εξελιγμένα συστήματα αυθεντικοποίησης. Μια τέτοια νέα προσέγγιση θα πρέπει επομένως να ενσωματώνει γλωσσικά χαρακτηριστικά, ικανά να προσδιορίσουν τον χρήστη και να τον ταυτοποιήσουν με ασφάλεια.

Με τη ραγδαία ανάπτυξη της τεχνολογίας και την αυξημένη εξάρτηση από ψηφιακά μέσα, τα συστήματα αυθεντικοποίησης γίνονται ολοένα και πιο περίπλοκα. Οι απαιτήσεις ασφαλείας αυξάνονται, ενώ ταυτόχρονα υπάρχει η ανάγκη για μεθόδους που είναι εύχρηστες και δεν απαιτούν συνεχή παρέμβαση από τους χρήστες. Αυτή η ανάγκη καθοδηγεί την έρευνα για νέες, πιο δυναμικές και ευέλικτες μεθόδους ταυτοποίησης, οι οποίες θα μπορούν να προσαρμοστούν στις απαιτήσεις των σύγχρονων ψηφιακών εφαρμογών και στις αυξημένες απαιτήσεις ασφαλείας του σημερινού διαδικτυακού περιβάλλοντος.

Τα τελευταία χρόνια, η Μηχανική Μάθηση και η Επεξεργασία Φυσικής Γλώσσας (Natural Language Processing - NLP) έχουν αρχίσει να παίζουν καθοριστικό ρόλο στην ανάπτυξη συστημάτων ασφαλείας και ταυτοποίησης. Η Μηχανική Μάθηση επιτρέπει τη δημιουργία έξυπνων συστημάτων που μπορούν να εκπαιδεύονται και να αναγνωρίζουν πρότυπα και συμπεριφορές, ενώ οι τεχνικές NLP προσφέρουν την ικανότητα ανάλυσης και επεξεργασίας του γραπτού λόγου, δημιουργώντας νέες δυνατότητες για την ανίχνευση μοναδικών χαρακτηριστικών σε επίπεδο χρήστη. Αυτά τα εργαλεία επιτρέπουν τη δημιουργία εξατομικευμένων προφίλ, τα οποία βασίζονται σε χαρακτηριστικά όπως η γλωσσική συμπεριφορά και το στυλ γραφής του χρήστη.

Η ανάγκη για ισχυρά συστήματα ασφαλείας έχει δημιουργήσει τις προϋποθέσεις για νέες μορφές αυθεντικοποίησης, οι οποίες συνδυάζουν την ευελιξία και τη διακριτική λειτουργία με την ισχυρή ασφάλεια. Το προτεινόμενο σύστημα στην εργασία αυτή αξιοποιεί αυτές τις εξελίξεις στην τεχνολογία και ενσωματώνει μεθόδους που βασίζονται στη μηχανική μάθηση και την επεξεργασία φυσικής γλώσσας, ώστε να δημιουργήσει ένα ισχυρό πλαίσιο για την ταυτοποίηση των χρηστών.

### 1.1 ΠΕΡΙΓΡΑΦΗ ΤΟΥ ΠΡΟΒΛΗΜΑΤΟΣ

---

Η ανάγκη για αυξημένη ασφάλεια στις ψηφιακές επικοινωνίες και την προστασία δεδομένων έχει καταστεί επιτακτική, ιδίως λόγω των συνεχώς εξελισσόμενων απειλών και της ψηφιοποίησης κάθε πτυχής της ανθρώπινης δραστηριότητας. Οι σύγχρονες τεχνολογίες απαιτούν συστήματα ταυτοποίησης που να μπορούν να παρέχουν διαρκή ασφάλεια, χωρίς να παρεμβαίνουν στη ροή ενεργειών του χρήστη. Οι παραδοσιακές μέθοδοι αυθεντικοποίησης, όπως η χρήση κωδικών πρόσβασης, παρουσιάζουν σοβαρούς περιορισμούς. Από τη μία, η πολυπλοκότητα και η συχνή ανανέωση των κωδικών απαιτεί συνεχή προσοχή από τον χρήστη, ενώ από την άλλη οι μέθοδοι αυτές είναι εύαλπτες σε επιθέσεις, όπως το phishing ή το brute-force.

Καθημερινά τεράστιος όγκος πληροφοριών δημιουργείται και ανταλλάσσεται μέσω του γραπτού λόγου. Η συγκεκριμένη μορφή επικοινωνίας ωστόσο είναι ιδιαίτερα εύαλπη σε κακόβουλες ενέργειες. Καθημερινά παραδείγματα αποτελούν τα spam emails, τα οποία επιχειρούν να εξαπατήσουν χρήστες για την αποκάλυψη ευαίσθητων δεδομένων, ενώ malicious tweets και αναρτήσεις σε κοινωνικά δίκτυα συχνά

χρησιμοποιούνται για τη διασπορά παραπληροφόρησης. Παράλληλα, η διάδοση bots και deepfake τεχνικών δημιουργούν την ανάγκη για πιο εξελιγμένα μέσα ανίχνευσης και προστασίας. Σε τέτοιες περιπτώσεις, η χρήση παραδοσιακών μεθόδων αυθεντικοποίησης, όπως κωδικοί πρόσβασης ή βιομετρικά δεδομένα, δεν επαρκεί. Ακόμα και αν εξασφαλίσουν την αρχική πρόσβαση, δεν παρέχουν διαρκή προστασία καθ' όλη τη διάρκεια χρήσης της υπηρεσίας, αφήνοντας τα συστήματα ευάλωτα σε δυνητικές επιθέσεις. Είναι επιτακτική, επομένως, η ανάγκη για περισσότερο δυναμικές μεθόδους αυθεντικοποίησης, που θα ενσωματώνουν γλωσσικά χαρακτηριστικά μέσω της ανάλυσης γραφής και θα προσφέρουν μια διακριτική και αξιόπιστη λύση, επιτρέποντας τη συνεχή παρακολούθηση της ταυτότητας του χρήστη με βάση το προσωπικό του στυλ γραφής.

Η έμμεση αυθεντικοποίηση, η οποία αξιοποιεί χαρακτηριστικά της φυσικής συμπεριφοράς του χρήστη, προσφέρει μια πιο φυσική και ασφαλή λύση. Ειδικότερα, η ανάλυση γραφής, που ενσωματώνει στοιχεία του προσωπικού στυλ του χρήστη, επιτρέπει την αναγνώριση ταυτότητας με τρόπο διακριτικό και ανεξάρτητο. Καθώς κάθε άτομο έχει τον δικό του τρόπο διατύπωσης και χρήσης της γλώσσας, οι αποκλίσεις στη γραφή μπορούν να ανιχνευθούν μέσω ενός προσαρμοσμένου μοντέλου που εκπαιδεύεται και αναγνωρίζει τον αυθεντικό χρήστη.

Η χρήση τεχνικών Επεξεργασίας Φυσικής Γλώσσας επιτρέπει την εξαγωγή χαρακτηριστικών που μπορούν να χρησιμοποιηθούν ως μοναδικά "ψηφιακά αποτυπώματα" κάθε χρήστη, βασιζόμενα σε δείκτες όπως το μέσο μήκος λέξεων σε χαρακτήρες, η συχνότητα συγκεκριμένων συντακτικών δομών ή μερών του λόγου και η ποικιλία του λεξιλογίου. Οι τεχνικές NLP μετατρέπουν τον γραπτό λόγο σε ένα σύνολο ποσοτικών μετρήσεων που αναλύονται με τεχνικές μηχανικής μάθησης. Με αυτόν τον τρόπο, η ταυτοποίηση πραγματοποιείται με βάση διακριτά χαρακτηριστικά του γραπτού λόγου, επιτρέποντας την αδιάλειπτη επαλήθευση ταυτότητας, χωρίς να απαιτείται η άμεση παρέμβαση του χρήστη.

Η έμμεση αυθεντικοποίηση δεν προσφέρει μόνο ένα νέο επίπεδο ασφάλειας, αλλά και σημαντικά πλεονεκτήματα όσον αφορά τη χρησιμότητα. Αντί να διακόπτει την εμπειρία του χρήστη, ενσωματώνεται αδιάλειπτα στη διαδικασία της αλληλεπίδρασης με το σύστημα. Το γεγονός αυτό την καθιστά ιδανική για περιβάλλοντα όπου η συνεχής πρόσβαση στα δεδομένα είναι απαραίτητη και η διακοπή της ροής ενεργειών για λόγους ασφάλειας μπορεί να είναι επιζήμια ή ενοχλητική για τον χρήστη. Επιπλέον, η δυνατότητα της έμμεσης αυθεντικοποίησης να ανιχνεύει απειλές χωρίς να επιβαρύνει τον χρήστη προσφέρει μια πιο ολιστική προσέγγιση στην προστασία της ταυτότητας και των δεδομένων του.

Η ανάπτυξη και η εξέλιξη του τομέα της συνεχούς και έμμεσης αυθεντικοποίησης έχουν ιδιαίτερη σημασία, καθώς προσφέρουν τη δυνατότητα για πιο ανθεκτικά και προσαρμοστικά συστήματα ασφάλειας. Στο πλαίσιο αυτής της εργασίας, η έμμεση και συνεχής αυθεντικοποίηση υλοποιείται μέσω τεχνικών NLP και μηχανικής μάθησης, που επιτρέπουν την εκμάθηση και ανίχνευση μοναδικών χαρακτηριστικών του γραπτού λόγου του χρήστη. Χρησιμοποιώντας ένα μοντέλο One-Class Support Vector Machine, το σύστημα αναγνωρίζει πρότυπα γραφής του χρήστη και ανιχνεύει αποκλίσεις που μπορεί να υποδηλώνουν παραβίαση στο σύστημα. Το προτεινόμενο σύστημα καταδεικνύει τη δυνατότητα των σύγχρονων τεχνολογιών να προσφέρουν λύσεις που ανταποκρίνονται στις αυξανόμενες ανάγκες για ασφάλεια, παρέχοντας

ταυτόχρονα ένα εύχρηστο και ελκυστικό προς τον χρήστη περιβάλλον.

### 1.2 ΣΚΟΠΟΣ - ΣΥΝΕΙΣΦΟΡΑ ΤΗΣ ΔΙΠΛΩΜΑΤΙΚΗΣ ΕΡΓΑΣΙΑΣ

---

Η παρούσα διπλωματική εργασία μελετά την ανάπτυξη και αξιολόγηση ενός συστήματος συνεχούς και έμμεσης αυθεντικοποίησης χρηστών, βασισμένου σε χαρακτηριστικά γραφής που εξάγονται μέσω τεχνικών επεξεργασίας φυσικής γλώσσας και μηχανικής μάθησης. Στόχος της εργασίας είναι η ανάπτυξη ενός συστήματος το οποίο μπορεί να ταυτοποιεί χρήστες με διακριτικό και συνεχόμενο τρόπο, αναγνωρίζοντας τον μοναδικό τρόπο γραφής τους. Η διαδικασία αυθεντικοποίησης πραγματοποιείται μέσω μοντέλων OC-SVM, τα οποία εκπαιδεύονται ώστε να αναγνωρίζουν αποκλίσεις από τη φυσιολογική γραφή του κάθε χρήστη, αποκλείοντας έτσι τους μη εξουσιοδοτημένους χρήστες από τη χρήση του εκάστοτε συστήματος.

Εξετάζεται η χρήση των τεχνικών NLP για την εξαγωγή γλωσσικών χαρακτηριστικών, όπως το μέσο μήκος λέξεων, η ποικιλία του λεξιλογίου και η δομή των προτάσεων, τα οποία μπορούν να αποδώσουν μια μοναδική ταυτότητα για κάθε χρήστη. Επιπλέον, παρουσιάζεται η εκπαίδευση και αξιολόγηση του μοντέλου OC-SVM για την έμμεση αυθεντικοποίηση, καθώς και η ανάλυση της αποτελεσματικότητας του προτεινόμενου συστήματος σε συνθήκες πραγματικής χρήσης. Η εργασία διερευνά την ακρίβεια και την απόδοση του συστήματος αυθεντικοποίησης, μετρώντας την αξιοπιστία του μέσω των δεικτών False Rejection Rate (FRR) και False Acceptance Rate (FAR).

Η εργασία συνεισφέρει στον τομέα της ψηφιακής ασφάλειας, προτείνοντας ένα σύστημα που προσφέρει διαρκή και αδιάλειπτη αυθεντικοποίηση χωρίς να απαιτεί συνεχείς ενέργειες από τον χρήστη, ενσωματώνοντας έτσι την επαλήθευση ταυτότητας στη φυσική ροή των καθημερινών δραστηριοτήτων. Παράλληλα, ανοίγει τον δρόμο για τη χρήση τεχνικών μηχανικής μάθησης και NLP στην αναγνώριση ταυτότητας χρηστών μέσω ανάλυσης γραφής, δημιουργώντας προοπτικές για την ανάπτυξη ασφαλών και ευέλικτων εφαρμογών σε περιβάλλοντα υψηλών απαιτήσεων.

## 1.3 ΔΙΑΡΘΡΩΣΗ ΤΗΣ ΑΝΑΦΟΡΑΣ

---

Η διάρθρωση της παρούσας διπλωματικής εργασίας είναι η εξής:

- **Κεφάλαιο 2:** Γίνεται ανασκόπηση της ερευνητικής περιοχής με έμφαση στις τεχνικές αυθεντικοποίησης και αναγνώρισης χρηστών μέσω NLP και μηχανικής μάθησης.
- **Κεφάλαιο 3:** Παρουσιάζεται το θεωρητικό υπόβαθρο της εργασίας, με ανάλυση των βασικών εννοιών της αυθεντικοποίησης, της επεξεργασίας φυσικής γλώσσας και της μηχανικής μάθησης, καθώς και του μοντέλου One-Class SVM.
- **Κεφάλαιο 4:** Παρουσιάζεται η υλοποίηση του συστήματος, συμπεριλαμβανομένων των αλγορίθμων εξαγωγής χαρακτηριστικών, της επεξεργασίας δεδομένων, της εκπαίδευσης των μοντέλων και της αξιολόγησης του συστήματος.
- **Κεφάλαιο 5:** Παρουσιάζεται η μεθοδολογία των πειραμάτων και τα αποτελέσματα αξιολόγησης του συστήματος, περιλαμβάνοντας τους δείκτες FAR και FRR.
- **Κεφάλαιο 6:** Παρουσιάζονται τα τελικά συμπεράσματα και τα προβλήματα που προέκυψαν.
- **Κεφάλαιο 7:** Προτείνονται θέματα για μελλοντική μελέτη, αλλαγές και επεκτάσεις.

## Επισκόπηση της Ερευνητικής Περιοχής

Ο συνεχής και έμμεσος έλεγχος ταυτότητας μέσω ανάλυσης γραφής αποτελεί έναν ταχύτατα αναπτυσσόμενο τομέα έρευνας, στον οποίο συνδυάζονται τεχνικές Επεξεργασίας Φυσικής Γλώσσας και Μηχανικής Μάθησης. Η συγκεκριμένη προσέγγιση παρέχει νέες προοπτικές ασφάλειας, καθώς είναι ιδιαίτερα αποτελεσματική για την ταυτοποίηση χρηστών σε πραγματικό χρόνο. Στην ενότητα αυτή, παρουσιάζονται προηγούμενες έρευνες που διαμόρφωσαν την περιοχή, υπογραμμίζοντας τη χρονολογική εξέλιξή της και τη σημερινή ερευνητική κατάσταση.

Η επεξεργασία φυσικής γλώσσας συνδυάζει γλωσσολογικές και υπολογιστικές μεθόδους για την ανάλυση και κατανόηση του γραπτού λόγου. Οι Argamon et al. [1] διερεύνησαν πώς το φύλο, το είδος, και το στυλ γραφής επηρεάζουν τα γλωσσικά χαρακτηριστικά, παρέχοντας σημαντικές γνώσεις για τη συσχέτιση κοινωνικών παραμέτρων με τη γραπτή έκφραση. Ο Van Haltern [2] ανέπτυξε μεθόδους γλωσσικού προφίλ που εστιάζουν στη συχνότητα λέξεων και φράσεων, διευκολύνοντας την ταυτοποίηση συγγραφέων. Ο Hoover [3] έδειξε ότι η συχνότητα λέξεων μπορεί να είναι ένας ισχυρός δείκτης για την αναγνώριση συγγραφικού στυλ, ενώ ο Juola [4] πρότεινε τεχνικές που βασίζονται στο μήκος αντιστοιχιών για τη βελτίωση της ακρίβειας στην ανάλυση γραπτών. Οι Kesel και Cercone [5] εισήγαγαν την προσέγγιση Common-N-Grams (CNG) με βαρύτητα ψήφων, συνδυάζοντας στατιστικές και υπολογιστικές μεθόδους για την ταυτοποίηση συγγραφέων. Οι Zhang et al. [6] χρησιμοποίησαν μοντέλα μετασχηματιστών για την αναγνώριση συγγραφικού στυλ, αποδεικνύοντας τη δύναμη της βαθιάς μάθησης στην ανάλυση κειμένου. Οι Coburn και Fitzpatrick [7] συνδύασαν τεχνικές NLP με θεωρία δικτύων για την ταυτοποίηση συγγραφέων, ενώ οι Schwartz et al. [8] ανέλυσαν τη γλώσσα κοινωνικών δικτύων, δείχνοντας πώς η προσωπικότητα και το φύλο αντικατοπτρίζονται στη χρήση της γλώσσας.

Η πρόοδος στη μηχανική μάθηση και ειδικότερα στα μοντέλα SVM έχει οδηγήσει σε σημαντικές εξελίξεις σε διάφορους τομείς ανάλυσης δεδομένων. Οι Schölkopf et al. [9] εισήγαγαν το μοντέλο One-Class SVM για την εκτίμηση κατανομών υψη-



---

λών διαστάσεων, προσφέροντας έναν καινοτόμο τρόπο ανίχνευσης αποκλίσεων σε δεδομένα. Οι Manevitz και Yousef [10] εξειδίκευσαν τη χρήση του One-Class SVM για την ταξινόμηση εγγράφων, δείχνοντας τη χρησιμότητά του σε προβλήματα με περιορισμένα παραδείγματα δεδομένων. Παράλληλα, οι Laskov et al. [11] εφάρμοσαν kernel-based learning methods για την ανίχνευση εισβολών, εισάγοντας έναν συνδυασμό πυρήνων για τη βελτίωση της ακρίβειας και της αποτελεσματικότητας των συστημάτων ασφαλείας. Οι Ferrante και Marone [12] αξιοποίησαν την τεχνολογία SVM για την ανάλυση στυλ, αναπτύσσοντας τεχνικές βαθιάς μάθησης που συνδυάζουν παραδοσιακές και σύγχρονες μεθόδους.

Η χρήση των SVM επεκτάθηκε περαιτέρω από τους Xu et al. [13], οι οποίοι ανέπτυξαν συνδυαστικές μεθόδους πυρήνων για την ανίχνευση ανωμαλιών σε πολυδιάστατα δεδομένα. Οι Baronchelli και Altmann [14] παρουσίασαν μέτρα βασισμένα στην εντροπία για τη ροή πληροφορίας σε δίκτυα, αξιοποιώντας την ικανότητα των SVM να μοντελοποιούν πολύπλοκα μοτίβα. Οι Gopalakrishnan et al. [15] χρησιμοποίησαν εξηγήσιμα μοντέλα SVM για την αυθεντικοποίηση χρηστών, ενισχύοντας τη διαφάνεια στις αποφάσεις των συστημάτων. Επιπλέον, οι Hong et al. [16] συνδύασαν SVM με Bayesian classifiers για τη βελτίωση της ακρίβειας στην ταξινόμηση βιομετρικών δεδομένων, αποδεικνύοντας τη χρησιμότητα των SVM σε εφαρμογές βιομετρίας.

Η μηχανική μάθηση έχει διαδραματίσει σημαντικό ρόλο στην ανίχνευση ανωμαλιών, με τις εφαρμογές να επεκτείνονται σε διάφορους τομείς, όπως η ασφάλεια συστημάτων, η βιομηχανική παρακολούθηση και η ανάλυση βιομετρικών δεδομένων. Οι Lu et al. [17] παρουσίασαν ένα πλαίσιο ταξινόμησης κυκλοφορίας δεδομένων με τη χρήση OC-SVM, εισάγοντας τεχνικές που επιτρέπουν τη διάκριση κανονικών και μη κανονικών μοτίβων. Παράλληλα, οι Chen et al. [18] χρησιμοποίησαν OC-SVM για την πρόβλεψη βλαβών λογισμικού, αποδεικνύοντας την αξία των μοντέλων αυτών για την αύξηση της αξιοπιστίας των συστημάτων. Επιπλέον, οι Fong και Narasimhan [19] αξιοποίησαν εργαλεία μη επιβλεπόμενης μάθησης για την ανίχνευση ανωμαλιών, επιδεικνύοντας τη δύναμη αυτών των τεχνικών στην πρόβλεψη και την παρακολούθηση υποδομών. Οι Narukawa et al. [20] ανέπτυξαν συστήματα ανίχνευσης κινδύνων σύγκρουσης σε ρομποτικά περιβάλλοντα, προσαρμόζοντας τα SVM για τη διαχείριση πραγματικού χρόνου. Οι Sun et al. [21] χρησιμοποίησαν deep OC-SVM για την ανίχνευση ανωμαλιών σε δεδομένα βίντεο, ενώ οι Seo [22] επέκτειναν τη χρήση των SVM σε εφαρμογές περιεχομένου εικόνων, εστιάζοντας στην αναγνώριση μη κανονικών προτύπων. Ειδικότερα, οι Hayashi και Ruggiero [23] παρουσίασαν hands-free αυθεντικοποίηση, αξιοποιώντας SVM και δεδομένα IoT για την ανίχνευση ανωμαλιών σε πραγματικό χρόνο. Επιπλέον, οι Rabaoui et al. [24] ανέδειξαν τη χρησιμότητα των OC-SVM για την ανάλυση ήχου, αποδεικνύοντας τη δυνατότητα εφαρμογής τους σε περιβάλλοντα επιτήρησης.

Ο συνδυασμός της μηχανικής μάθησης και του NLP έχει προσφέρει νέες προοπτικές στην ανίχνευση ανωμαλιών και τη βελτίωση των συστημάτων ασφαλείας. Οι Chatzikyriakidis και Papageorgiou [25] αξιοποίησαν μοντέλα μετασχηματιστών για την ανάλυση ελληνικών κειμένων, προσφέροντας νέες μεθόδους ανίχνευσης αποκλίσεων μέσω γλωσσικών μοτίβων. Οι Karanikiotis et al. [26] παρουσίασαν ένα σύστημα συνεχούς αυθεντικοποίησης μέσω ανάλυσης δεδομένων αφής, ενώ οι Schwartz et al. [8] εστίασαν στη γλώσσα κοινωνικών δικτύων, δείχνοντας πώς η ανάλυση της

γλώσσας μπορεί να συνδεθεί με την ταυτοποίηση χρηστών και τη βελτίωση της ασφάλειας. Οι Karpathy και Fei-Fei [27] συνδύασαν τη γλωσσική ανάλυση και την αναγνώριση εικόνων για τη δημιουργία περιγραφών, εισάγοντας συστήματα που μπορούν να αναγνωρίζουν ανωμαλίες σε πολυτροπικά δεδομένα. Επιπλέον, ο Stylios [28] ανέλυσε τη χρήση βιομετρικών χαρακτηριστικών για την ανίχνευση μη κανονικών συμπεριφορών σε συστήματα συνεχούς αυθεντικοποίησης. Οι Hong et al. [16] συνδύασαν SVM και Bayesian methods για την ταξινόμηση βιομετρικών δεδομένων, ενώ οι Ferrante και Marone [12] χρησιμοποίησαν βαθιά μάθηση και SVM για στυλιστική ανάλυση. Τέλος, οι Gopalakrishnan et al. [15] εισήγαγαν μοντέλα XAI για την ανίχνευση ανωμαλιών σε περιβάλλοντα αυθεντικοποίησης, προσφέροντας μεγαλύτερη διαφάνεια στις αποφάσεις.

Τα εργαλεία που έχουν αναπτυχθεί για την υποστήριξη εφαρμογών μηχανικής μάθησης και ανάλυσης δεδομένων αποτελούν κρίσιμο παράγοντα για την πρόοδο στον τομέα. Ο Church [29] εξέτασε το Word2Vec, τονίζοντας την αποτελεσματικότητά του στη δημιουργία σημασιολογικών αναπαραστάσεων λέξεων, που αποτελούν θεμέλιο για πολλές εφαρμογές. Το GloVe, όπως αναπτύχθηκε από τους Pennington et al. [30], εισήγαγε τη σύνδεση της στατιστικής συχνότητας λέξεων με τη σημασιολογία, παρέχοντας βελτιωμένες δυνατότητες ανάλυσης σε εφαρμογές NLP. Η πλατφόρμα TensorFlow, που αναπτύχθηκε από την ομάδα TensorFlow [31], εισήγαγε υποδομές για την εκπαίδευση μεγάλων μοντέλων μηχανικής μάθησης σε διανεμημένα περιβάλλοντα, προσφέροντας ευελιξία και αποτελεσματικότητα στην επεξεργασία μεγάλων συνόλων δεδομένων. Παράλληλα, οι Devlin et al. [32] εισήγαγαν το BERT, ένα εργαλείο βαθιάς μάθησης που άλλαξε το πεδίο της κατανόησης φυσικής γλώσσας μέσω της αμφίδρομης επεξεργασίας δεδομένων. Τέλος, οι Wolf et al. [33] παρουσίασαν τη βιβλιοθήκη Transformers, η οποία παρέχει ένα ευρύ φάσμα προ-εκπαιδευμένων μοντέλων για την επεξεργασία φυσικής γλώσσας, επιτρέποντας την εύκολη εφαρμογή σύγχρονων τεχνικών NLP.

Συνολικά, η έρευνα στον τομέα της ανάλυσης κειμένου και της έμμεσης αυθεντικοποίησης έχει εξελιχθεί από απλές μεθόδους στατιστικής ανάλυσης σε σύνθετα μοντέλα βαθιάς μάθησης και μετασχηματιστές. Το παρόν έργο αξιοποιεί αυτές τις εξελίξεις για τη δημιουργία ενός σύγχρονου συστήματος συνεχούς αυθεντικοποίησης, συνεισφέροντας στην ασφάλεια και τη βελτίωση της εμπειρίας χρήστη.





# 3

## Θεωρητικό Υπόβαθρο

Η συνεχής και έμμεση αυθεντικοποίηση μέσω ανάλυσης συμπεριφοράς αποτελεί μια σύγχρονη προσέγγιση που συνδυάζει την επεξεργασία φυσικής γλώσσας και τις τεχνικές μηχανικής μάθησης. Στόχος του κεφαλαίου αυτού είναι να παρουσιάσει τα θεωρητικά θεμέλια που καθιστούν δυνατή την υλοποίηση αυτής της προσέγγισης. Οι ενότητες που ακολουθούν αναλύουν τη διαδικασία της αυθεντικοποίησης, τις τεχνικές επεξεργασίας φυσικής γλώσσας και εξαγωγής χαρακτηριστικών, τις τεχνικές μηχανικής μάθησης, και τους αλγόριθμους που χρησιμοποιούνται για την ταξινόμηση δεδομένων και τον εντοπισμό αποκλίσεων.

### 3.1 ΑΥΘΕΝΤΙΚΟΠΟΙΗΣΗ

---

Η αυθεντικοποίηση αποτελεί θεμελιώδη διαδικασία στον τομέα της ασφάλειας πληροφοριακών συστημάτων. Στόχος της είναι η επαλήθευση της ταυτότητας ενός χρήστη ή μιας συσκευής προτού παραχωρηθεί πρόσβαση σε δεδομένα ή υπηρεσίες. Η ανάγκη για αξιόπιστες μεθόδους αυθεντικοποίησης γίνεται ολοένα και πιο επιτακτική, εξαιτίας της αυξανόμενης πολυπλοκότητας των απειλών κυβερνοασφάλειας και των επιτιθέμενων που αναζητούν διαρκώς τρόπους να παρακάμψουν τα παραδοσιακά συστήματα ελέγχου ταυτότητας.

#### 3.1.1 Ορισμός Αυθεντικοποίησης

Η αυθεντικοποίηση αναφέρεται στη διαδικασία επαλήθευσης της ταυτότητας ενός χρήστη ή συσκευής προτού επιτραπεί η πρόσβαση σε ένα σύστημα. Η διαδικασία περιλαμβάνει την ταυτοποίηση, δηλαδή τη δήλωση της ταυτότητας, και την επαλήθευση, που επιβεβαιώνει την ακρίβεια της δήλωσης. Για παράδειγμα, ένας χρήστης μπορεί να δηλώσει την ταυτότητά του μέσω του ονόματος χρήστη (ταυτοποίηση) και να την επαληθεύσει μέσω ενός κωδικού πρόσβασης ή μιας βιομετρικής

μεθόδου (επαλήθευση).

### 3.1.2 Κατηγορίες Τεχνικών Αυθεντικοποίησης

Η εξέλιξη της τεχνολογίας έχει οδηγήσει στην ανάπτυξη ποικίλων συστημάτων αυθεντικοποίησης, που απαιτούν από τον χρήστη κάτι διαφορετικό κάθε φορά για την επαλήθευση της ταυτότητάς του. Η κατηγοριοποίηση αυτή επομένως μπορεί να αναλυθεί σε:

- **Γνώση:** Αφορά κάτι που ο χρήστης γνωρίζει και συμπεριλαμβάνει κωδικούς πρόσβασης / PIN, μοτίβο ή απάντηση σε κάποια ερώτηση.
- **Κατοχή:** Αφορά κάτι που ο χρήστης κατέχει και συμπεριλαμβάνει φυσικά αντικείμενα όπως κάρτες ή tokens.
- **Βιομετρικά χαρακτηριστικά:** Αφορά έμφυτα φυσιολογικά ή συμπεριφορικά χαρακτηριστικά, γνωστά και ως βιομετρικά, που είναι μοναδικά για κάθε άτομο. Οι βιομετρικές τεχνικές χωρίζονται σε δύο κύριες κατηγορίες:

#### 1. Φυσιολογικές Μέθοδοι:

- *Αναγνώριση προσώπου:* Χρήση χαρακτηριστικών του προσώπου για την ταυτοποίηση.
- *Δακτυλικά αποτυπώματα:* Καταγραφή και αντιστοίχιση μοναδικών αποτυπωμάτων.
- *Σάρωση ίριδας:* Εξαιρετικά ασφαλής μέθοδος, αλλά απαιτεί εξειδικευμένο εξοπλισμό.

#### 2. Συμπεριφορικές Μέθοδοι:

- *Ανάλυση γραφής:* Εξαγωγή χαρακτηριστικών από τον τρόπο που γράφει ένας χρήστης.
- *Δυναμική πληκτρολόγηση:* Παρακολούθηση των μοτίβων πληκτρολόγησης.
- *Συμπεριφορά πλοήγησης:* Ανάλυση τρόπων πλοήγησης σε περιβάλλοντα χρήστη.

Οι βιομετρικές μέθοδοι, σε αντίθεση με τις παραδοσιακές, δεν μπορούν εύκολα να παραβιαστούν, καθώς βασίζονται σε εγγενή χαρακτηριστικά του χρήστη.

Ακόμη, μπορούμε να διακρίνουμε 2 κατηγορίες τεχνικών αυθεντικοποίησης ανάλογα με τη διαφάνεια του συστήματος ως προς τον τελικό χρήστη.

- **Ενεργή / Άμεση:** το σύστημα απαιτεί την εισαγωγή δεδομένων από τον χρήστη, όπως την πληκτρολόγηση ενός κωδικού ή μίας απάντησης σε μία ερώτηση ασφαλείας.
- **Παθητική / Έμμεση:** εκτελείται στο παρασκήνιο, χωρίς να χρειάζεται ενέργεια από τον χρήστη. Συστήματα αυτής της κατηγορίας ξεχωρίζουν για την δυνατότητά τους να εκτελούνται συνεχώς χωρίς να επεμβαίνουν στην λειτουργικότητα της συσκευής.

Σήμερα πολλές υπηρεσίες και εφαρμογές χρησιμοποιούν συνδυασμούς τεχνικών αυθεντικοποίησης (Multi Factor Authentication - MFA). Συχνότερο παράδειγμα αποτελεί η αυθεντικοποίηση 2 παραγόντων (two-factor authentication ή 2FA). Στη συγκεκριμένη κατηγορία εμπίπτουν η ανάληψη χρημάτων από το ATM με την χρήση κάρτας (κατοχή) και την πληκτρολόγηση του PIN (γνώση), αλλά και η σύνδεση σε λογαριασμούς ηλεκτρονικού ταχυδρομείου με τον κωδικό πρόσβασης (γνώση) και ένα συνθηματικό που αποστέλλεται σε κάποια συσκευή του χρήστη (κατοχή).

### 3.1.3 Συνεχής και Έμμεση Αυθεντικοποίηση βασιζόμενη σε συμπεριφορικές μεθόδους

Η συνεχής και έμμεση αυθεντικοποίηση αποτελεί μία καινοτόμο προσέγγιση στον τομέα της ασφάλειας πληροφοριακών συστημάτων, η οποία δίνει έμφαση στην αδιάλειπτη και μη παρεμβατική επαλήθευση της ταυτότητας του χρήστη. Σε αντίθεση με τις παραδοσιακές μεθόδους αυθεντικοποίησης, οι οποίες συχνά απαιτούν τη ρητή συμμετοχή του χρήστη, όπως η εισαγωγή κωδικών πρόσβασης ή η χρήση βιομετρικών αναγνωριστικών, η συνεχής αυθεντικοποίηση αξιοποιεί πληροφορίες που συλλέγονται από τη συμπεριφορά του χρήστη και τις αλληλεπιδράσεις του με το σύστημα.

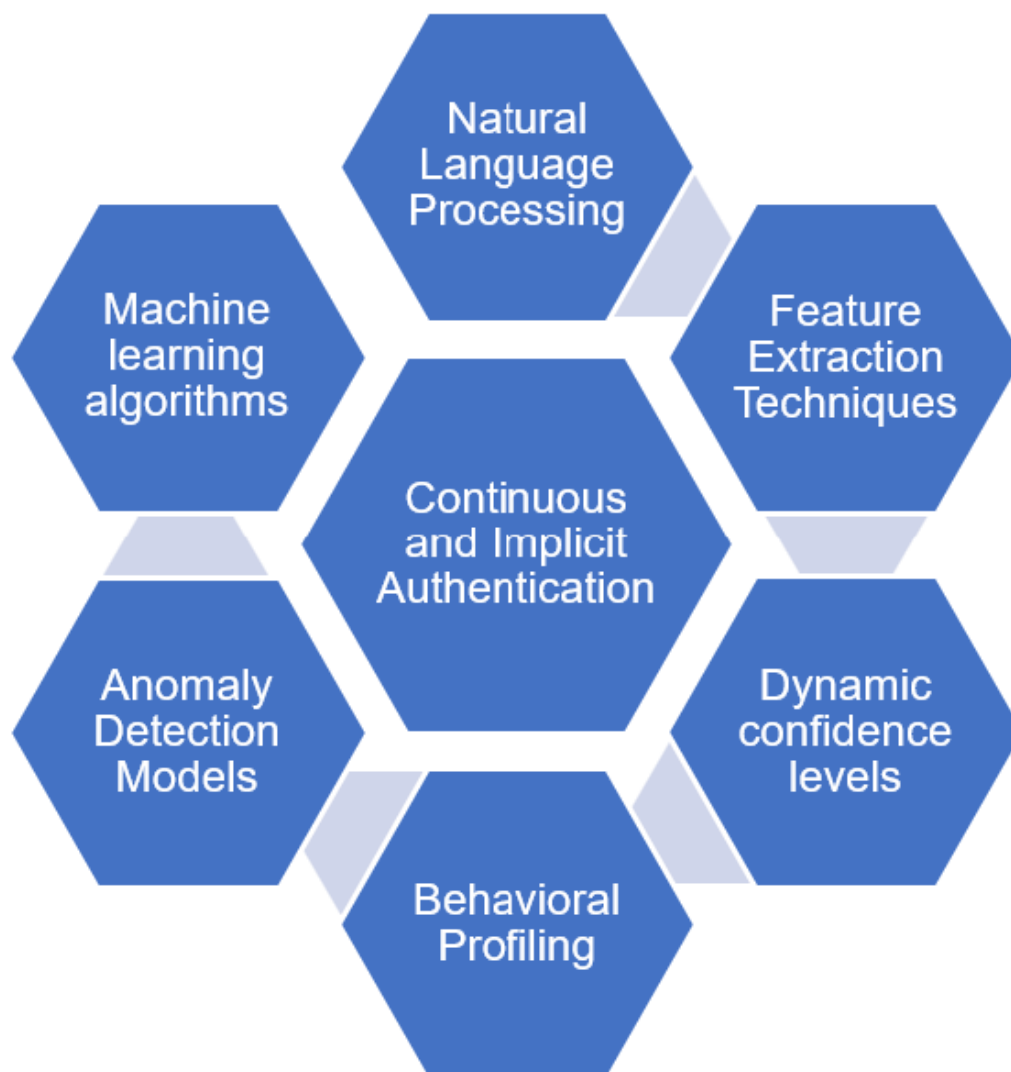
Οι τεχνικές συνεχούς και έμμεσης αυθεντικοποίησης χρησιμοποιούν δεδομένα όπως:

- Τα μοτίβα γραφής και πληκτρολόγησης, που αναλύονται μέσω τεχνικών επεξεργασίας φυσικής γλώσσας (NLP) και μηχανικής μάθησης.
- Τα μοτίβα πλοήγησης σε ψηφιακά περιβάλλοντα, τα οποία παρέχουν στοιχεία σχετικά με τη συμπεριφορά του χρήστη.
- Βιομετρικά δεδομένα χαμηλής συχνότητας, όπως η δυναμική χρήσης της συσκευής (π.χ., κλίση ή ταχύτητα κύλισης).

Η συνεχής και έμμεση αυθεντικοποίηση προσφέρει σημαντικά πλεονεκτήματα:

1. **Αυξημένη ασφάλεια:** Η συνεχής παρακολούθηση καθιστά δυσκολότερη την παραβίαση του συστήματος.
2. **Μη παρεμβατική λειτουργία:** Οι χρήστες δεν χρειάζεται να διακόπτουν τη ροή των ενεργειών τους για να επαληθεύσουν την ταυτότητά τους.
3. **Δυναμική προσαρμογή:** Τα συστήματα μπορούν να προσαρμόζονται στις αλλαγές της συμπεριφοράς του χρήστη, βελτιώνοντας τη συνολική ακρίβεια.
4. **Εύκολη ενσωμάτωση:** Μπορεί να ενσωματωθεί σε υπάρχον hardware, χωρίς να απαιτούνται πρόσθετα κόστη εξοπλισμού.
5. **Ευελιξία:** Μπορούν να χρησιμοποιηθούν πολλά διαφορετικά συμπεριφορικά χαρακτηριστικά ανάλογα με τις απαιτήσεις και το τελικό προϊόν που θα εξυπηρετεί το σύστημα.

Ένα παράδειγμα εφαρμογής αυτής της τεχνολογίας είναι η ανάλυση γραφής για αυθεντικοποίηση σε περιβάλλοντα συνομιλιών. Η τεχνική αυτή βασίζεται στην εξαγωγή χαρακτηριστικών, όπως η επιλογή λέξεων, η σύνταξη και η δομή των προτάσεων, που αποτελούν μοναδικά χαρακτηριστικά του χρήστη. Σε συνδυασμό με τεχνολογίες όπως οι αλγόριθμοι ανίχνευσης ανωμαλιών, η συνεχής αυθεντικοποίηση μπορεί να εξασφαλίσει ένα υψηλό επίπεδο ασφάλειας, χωρίς να επηρεάζει την εμπειρία του χρήστη.



Σχήμα 3.1: Πτυχές της συνεχούς και έμμεσης αυθεντικοποίησης

Η χρήση τέτοιων τεχνολογιών ανοίγει νέες προοπτικές για εφαρμογές όπως η προστασία προσωπικών δεδομένων, η ασφαλής πρόσβαση σε κρίσιμες υποδομές, και η βελτίωση της εμπειρίας των χρηστών σε ψηφιακά περιβάλλοντα.



## 3.2 ΕΠΕΞΕΡΓΑΣΙΑ ΦΥΣΙΚΗΣ ΓΛΩΣΣΑΣ

### 3.2.1 Εισαγωγή στην Επεξεργασία Φυσικής Γλώσσας

Η επεξεργασία φυσικής γλώσσας αποτελεί έναν από τους πλέον εξελισσόμενους τομείς της τεχνητής νοημοσύνης. Στόχος της είναι η κατανόηση, ανάλυση και εξαγωγή χαρακτηριστικών από κείμενα, δίνοντας τη δυνατότητα στα συστήματα να ερμηνεύουν και να επεξεργάζονται τη γλώσσα. Ο εξαιρετικά μεγάλος όγκος δεδομένων που ανταλλάσσονται διαρκώς υπό τη μορφή κειμένου καθιστά επιτακτική την ανάγκη για αυξημένη ασφάλεια. Στο πλαίσιο αυτής της εργασίας, το NLP αξιοποιείται για την εξαγωγή χαρακτηριστικών που αποτυπώνουν μοναδικές γλωσσικές και συμπεριφορικές πτυχές κάθε χρήστη.

Η κατανόηση των κειμένων περιλαμβάνει διαδικασίες όπως η αναγνώριση της δομής, η σημασιολογική ανάλυση και η δημιουργία μοναδικών γλωσσικών προφίλ. Αυτές οι διαδικασίες παίζουν κρίσιμο ρόλο στη δημιουργία συστημάτων συνεχούς αυθεντικοποίησης χρηστών, όπου ο στόχος είναι η ανίχνευση μοτίβων γραφής που επιτρέπουν την ασφαλή ταυτοποίηση.

Η εξέλιξη του κλάδου NLP ξεκίνησε με τις πρώτες γλωσσολογικές προσεγγίσεις, όπως οι τεχνικές Bag-of-Words και Term Frequency - Inverse Document Frequency (TF-IDF), οι οποίες βασίζονταν σε απλή μέτρηση συχνότητας λέξεων. Με την έλευση της μηχανικής μάθησης, η ανάλυση κειμένων πέρασε σε πιο σύνθετα επίπεδα, όπως η εξαγωγή γλωσσολογικών και σημασιολογικών χαρακτηριστικών.

Η εξέλιξη των εργαλείων NLP είναι ραγδαία. Αρχικά, το Word2Vec [34] εισήγαγε τη δυνατότητα εκμάθησης συσχετίσεων μεταξύ διανυσμάτων αναπαράστασης των λέξεων, οδηγώντας στην ανάλυση σημασιολογικών χαρακτηριστικών, πέρα από καθαρά γλωσσικών. Επίσης, το GloVe [30] βελτίωσε την προσέγγιση αυτή καθώς παρείχε στατιστικά σχετικά με τη συχνότητα των λέξεων μέσω της γεωμετρικής τους απεικόνισης σε πολυδιάστατους χώρους. Τέλος, μετασχηματιστές όπως ο BERT [32] αναπαριστούν το κείμενο ως μια σειρά από διανύσματα χρησιμοποιώντας επιβλεπόμενη μάθηση και μαθαίνουν λανθάνουσες αναπαραστάσεις των σημείων στο πλαίσιο των συμφραζομένων τους. Η τεχνική αυτή αποτέλεσε σημαντική βελτίωση έναντι των προηγούμενων μοντέλων.

Συνολικά, το NLP διαδραματίζει κεντρικό ρόλο στη δημιουργία μοναδικών προφίλ χρηστών. Μέσω της ανάλυσης γλωσσικών μοτίβων, επιτυγχάνεται η ταυτοποίηση χρηστών βασισμένη στη γραφή τους. Συστήματα αυθεντικοποίησης βασισμένα στο NLP χρησιμοποιούνται σε εφαρμογές ασφάλειας, όπου απαιτείται υψηλή ακρίβεια και αξιοπιστία.

#### Tokenization

Η διαίρεση του κειμένου σε λέξεις, φράσεις ή προτάσεις αποτελεί το πρώτο στάδιο επεξεργασίας. Το Tokenization επιτρέπει την απομόνωση σημαντικών τμημάτων του κειμένου για περαιτέρω ανάλυση.

### Stemming και Lemmatization

Η απλοποίηση λέξεων στην αρχική τους μορφή βελτιώνει την ακρίβεια της γλωσσικής ανάλυσης. Το Stemming αφαιρεί τα προσφύματα των λέξεων, ενώ το Lemmatization διατηρεί τη γραμματική ακεραιότητα.

### Part-of-Speech Tagging

Η επισήμανση της γραμματικής κατηγορίας (π.χ. ουσιαστικά, ρήματα) παρέχει σημαντικές πληροφορίες για τη σύνταξη και τη σημασιολογία.

### Named Entity Recognition (NER)

Το NER αναγνωρίζει οντότητες, όπως ονόματα, ημερομηνίες ή τοποθεσίες, βοηθώντας στη δημιουργία πλούσιων γλωσσικών προφίλ.

### Dependency Parsing

Αναλύει τις συντακτικές σχέσεις μεταξύ λέξεων, αποκαλύπτοντας τη δομή του κειμένου.

## 3.2.2 Εργαλεία και Μέθοδοι Εξαγωγής Χαρακτηριστικών

Τα χαρακτηριστικά εξάγονται με τη χρήση εργαλείων όπως:

- **NLTK**<sup>1</sup> και **spaCy**<sup>2</sup>: Για μορφολογική και συντακτική ανάλυση.
- **textstat**<sup>3</sup>: Για δείκτες αναγνωσιμότητας και πολυπλοκότητας.

---

<sup>1</sup><https://www.nltk.org/>

<sup>2</sup><https://spacy.io/>

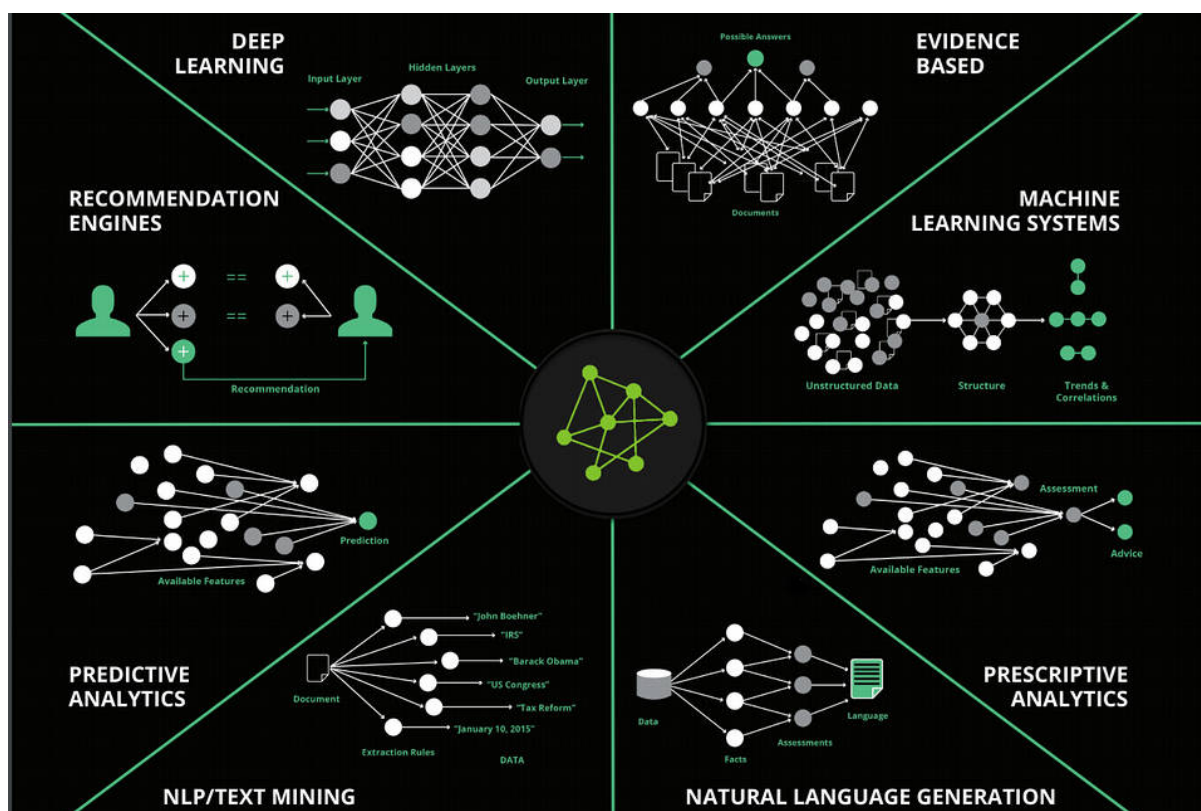
<sup>3</sup><https://texstat.org/>

### 3.3 ΣΥΓΧΡΟΝΗ ΜΗΧΑΝΙΚΗ ΜΑΘΗΣΗ

Η μηχανική μάθηση (Machine Learning, ML) αποτελεί έναν από τους πλέον δυναμικά εξελισσόμενους τομείς της επιστήμης των υπολογιστών, με εκτεταμένες εφαρμογές στην ανάλυση δεδομένων, τη λήψη αποφάσεων και την κατανόηση της φυσικής γλώσσας. Η ουσία της μηχανικής μάθησης έγκειται στη δυνατότητα των συστημάτων να «μαθαίνουν» από τα δεδομένα και να βελτιώνουν την απόδοσή τους χωρίς ρητές οδηγίες προγραμματισμού. Με τη χρήση εξελιγμένων αλγορίθμων, τα μοντέλα μηχανικής μάθησης αναπτύσσουν ικανότητες για την εξαγωγή μοτίβων και τη δημιουργία προβλέψεων σε πραγματικό χρόνο.

Η εισαγωγή του κλάδου της μηχανικής μάθησης στην επιστήμη των υπολογιστών, επέτρεψε στους υπολογιστές να μπορούν να αντιμετωπίσουν προβλήματα αντίληψης για τον πραγματικό κόσμο, όσο και να παίρνουν υποκειμενικές αποφάσεις.

Οι αλγόριθμοι ML επιτρέπουν σε συστήματα Τεχνητής Νοημοσύνης (Artificial Intelligence, AI) να προσαρμόζονται εύκολα σε καινούργια προβλήματα απαιτώντας ελάχιστη επέμβαση από τον άνθρωπο. Για παράδειγμα, ένα νευρωνικό δίκτυο που έχει εκπαιδευτεί να αναγνωρίζει γάτες σε εικόνες, δεν απαιτεί να σχεδιαστεί και να εκπαιδευτεί από το μηδέν για να έχει την ικανότητα να αναγνωρίζει και σκύλους.



Σχήμα 3.2: Κλάδοι και εφαρμογές της επιστήμης της Τεχνητής Νοημοσύνης

### 3.3.1 Κατηγορίες Αλγορίθμων Μηχανικής Μάθησης

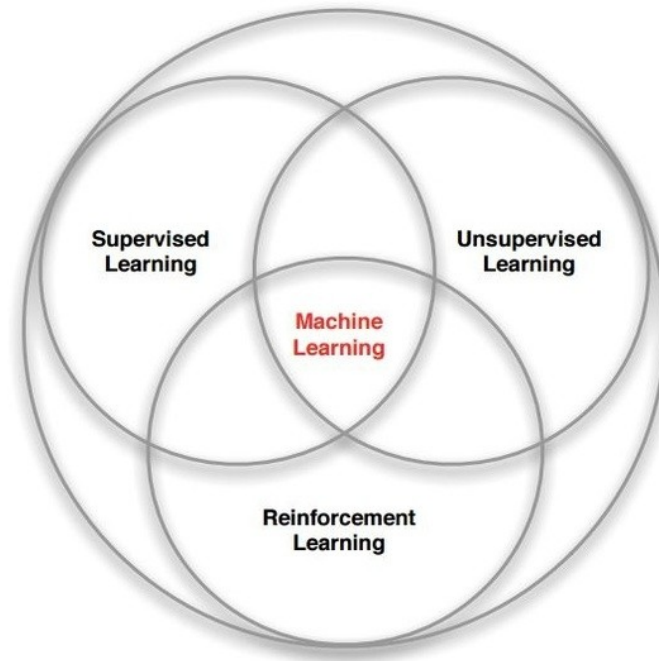
Πολλά προβλήματα που μέχρι πριν μερικά χρόνια λύνονταν με “χειρόγραφη”, προγραμματισμένη από τον άνθρωπο γνώση, σήμερα επιλύονται με χρήση αλγορίθμων ML (σχήμα 3.2). Κάποια παραδείγματα αφορούν:

- Αναγνώριση ομιλίας - Speech Recognition
- Μηχανική όραση - Computer Vision
  - Αναγνώριση αντικειμένων σε εικόνες - Object Recognition
  - Αναγνώριση και εντοπισμός της θέσης αντικειμένων σε εικόνες - Object Detection
- Αναγνώριση ηλεκτρονικών επιθέσεων στο διαδίκτυο - Cyberattack detection
- Επεξεργασία φυσικής γλώσσας - Natural Language Processing
  - Κατανόηση της φυσικής γλώσσας του ανθρώπου - Natural Language Understanding
  - Μοντελοποίηση και παραγωγή της φυσικής γλώσσας του ανθρώπου από μηχανές - Natural Language Generation
- Μηχανές αναζήτησης - Search Engines
- Αναπαράσταση γνώσης - Knowledge Representation
- Ρομποτική

Η επιστήμη της μηχανικής μάθησης μπορεί να ταξινομηθεί σε τρεις κύριες κατηγορίες:

1. **Εποπτευόμενη Μάθηση (Supervised Learning):** Το μοντέλο εκπαιδεύεται σε σύνολα δεδομένων όπου υπάρχουν ετικέτες (labels) που καθοδηγούν τη διαδικασία εκμάθησης.
2. **Μη Εποπτευόμενη Μάθηση (Unsupervised Learning):** Το μοντέλο επιχειρεί να ανακαλύψει μοτίβα και δομές από δεδομένα χωρίς προκαθορισμένες ετικέτες.
3. **Μάθηση Ενίσχυσης (Reinforcement Learning):** Το μοντέλο βελτιώνει τη συμπεριφορά του μέσω επαναλαμβανόμενης αλληλεπίδρασης με το περιβάλλον και αξιολόγησης των ενεργειών του. Ένα παράδειγμα εφαρμογής είναι η αυτόματη πλοήγηση ενός οχήματος.

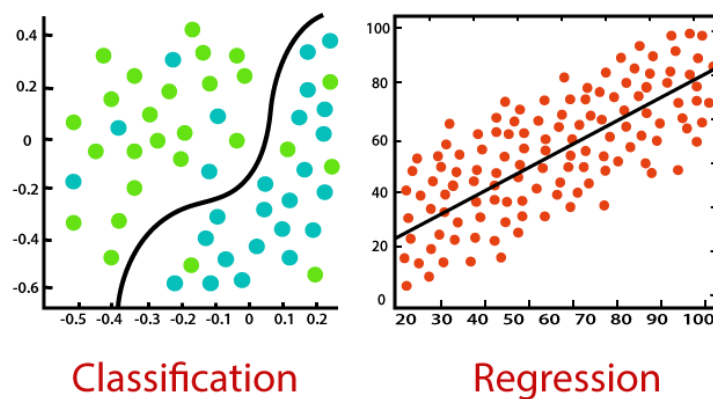
Κάποια προβλήματα είναι υβριδικά, δηλαδή συνδυασμός των πιο πάνω. Στο σχήμα 3.3 απεικονίζεται το διάγραμμα Venn των διαφόρων αλγοριθμικών κατηγοριών ML.



Σχήμα 3.3: Διάγραμμα Venn των διαφόρων κατηγοριών μηχανικής μάθησης

Επιπλέον, οι Supervised Learning αλγόριθμοι χωρίζονται σε 2 κατηγορίες, όπως φαίνεται στο [σχήμα 3.4](#), ανάλογα με την επιθυμητή μορφή της εξόδου του αλγόριθμου ML:

- Ταξινόμησης - Classification: Πρόβλεψη μίας διακριτής κατηγορίας ή κλάσης. Η έξοδος είναι μία διακριτή ετικέτα (label).
- Παλινδρόμησης - Regression: Πρόβλεψη μίας συνεχούς μεταβλητής. Η έξοδος είναι μια συνεχής τιμή. Οι αλγόριθμοι παλινδρόμησης παράγουν ένα μοντέλο που μπορεί να προβλέψει αριθμητικές τιμές.



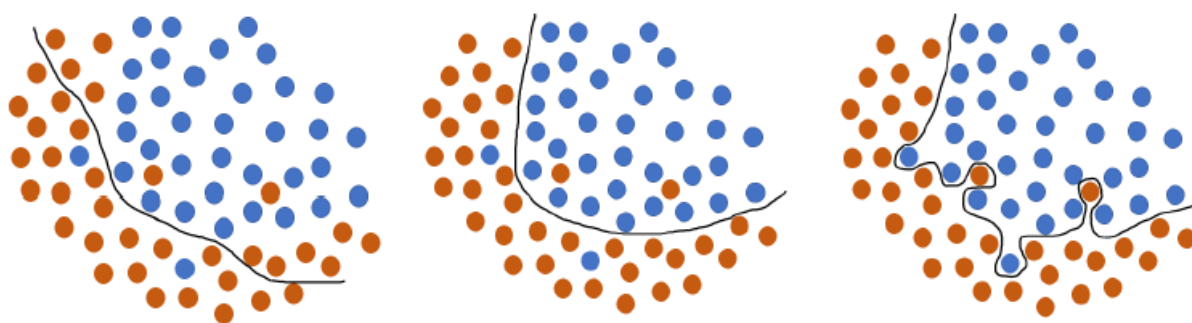
Σχήμα 3.4: Διαφορά classification & regression



### 3.3.2 Προκλήσεις

Το **overfitting** και το **underfitting** αποτελούν δύο από τις πιο σημαντικές προκλήσεις στη μηχανική μάθηση, καθώς επηρεάζουν άμεσα την ικανότητα του μοντέλου να γενικεύει σε νέα δεδομένα. Το **overfitting** προκύπτει όταν το μοντέλο μαθαίνει υπερβολικά καλά τα δεδομένα εκπαίδευσης, συμπεριλαμβανομένων των θορύβων ή σφαλμάτων, με αποτέλεσμα να αποδίδει εξαιρετικά στα δεδομένα αυτά, αλλά να αποτυγχάνει στα δεδομένα δοκιμών ή σε νέα δεδομένα. Αυτό συμβαίνει όταν το μοντέλο είναι υπερβολικά περίπλοκο, για παράδειγμα, περιέχει πολλές παραμέτρους σε σχέση με τα διαθέσιμα δεδομένα. Αντίθετα, το **underfitting** εμφανίζεται όταν το μοντέλο αποτυγχάνει να μάθει επαρκώς τα μοτίβα των δεδομένων εκπαίδευσης, με αποτέλεσμα να έχει χαμηλή απόδοση τόσο στα δεδομένα εκπαίδευσης όσο και σε νέα δεδομένα. Αυτό συμβαίνει συχνά όταν το μοντέλο είναι υπερβολικά απλό ή οι παράμετροί του δεν έχουν ρυθμιστεί σωστά.

Η αντιμετώπιση αυτών των προκλήσεων απαιτεί προσεκτική επιλογή της αρχιτεκτονικής του μοντέλου, της μεθόδου εκπαίδευσης και των υπερπαραμέτρων, καθώς και τη χρήση τεχνικών όπως η διασταυρούμενη επικύρωση, η κανονικοποίηση και η αύξηση του μεγέθους ή της ποικιλίας των δεδομένων εκπαίδευσης. Η **διασταυρούμενη επικύρωση (cross-validation, CV)** είναι μια τεχνική που χρησιμοποιείται για την αξιολόγηση της απόδοσης ενός μοντέλου και περιλαμβάνει τη διαίρεση των δεδομένων σε υποσύνολα (folds). Το μοντέλο εκπαιδεύεται επανειλημμένα σε διαφορετικά υποσύνολα και δοκιμάζεται σε αυτά που εξαιρούνται, ώστε να εκτιμηθεί η γενική απόδοσή του σε νέα δεδομένα. Η **κανονικοποίηση (regularization)** είναι η διαδικασία μετατροπής των δεδομένων σε μία ακολουθία κανονικών μορφών, οι οποίες αποτελούνται από απλές και σαφείς σχέσεις που δεν περιέχουν επαναλήψεις. Ως στόχο έχει να περιορίσει την πολυπλοκότητα του μοντέλου, αποτρέποντας έτσι το overfitting.

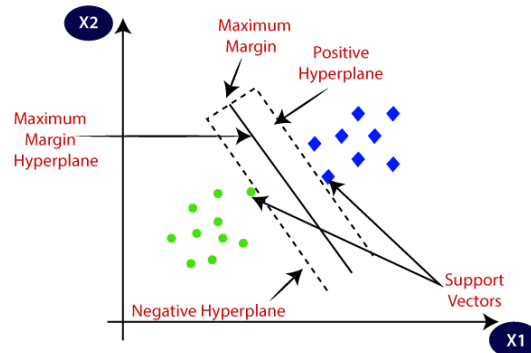


Σχήμα 3.5: Πρώτο γράφημα: underfitting, Δεύτερο γράφημα: best fit, Τρίτο γράφημα: overfitting

### 3.3.3 Το SVM και το One-Class SVM

Ένας από τους θεμελιώδεις αλγορίθμους στη μηχανική μάθηση είναι το *Support Vector Machine* (SVM), το οποίο είναι ιδιαίτερα ισχυρό για την επίλυση προβλημάτων ταξινόμησης (*classification*) και παλινδρόμησης (*regression*). Το SVM βασίζεται

στη χρήση ενός υπερεπιπέδου (*hyperplane*) που διαχωρίζει δεδομένα σε διαφορετικές κλάσεις στον χώρο χαρακτηριστικών (*feature space*). Ο αλγόριθμος επιλέγει το υπερεπίπεδο που μεγιστοποιεί το περιθώριο (*margin*) μεταξύ των δεδομένων των διαφορετικών κλάσεων, όπως φαίνεται στο [σχήμα 3.6](#).

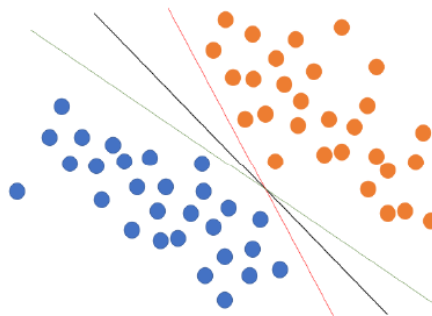


Σχήμα 3.6: Λειτουργία του SVM - Κατασκευή hyperplane με το maximum margin

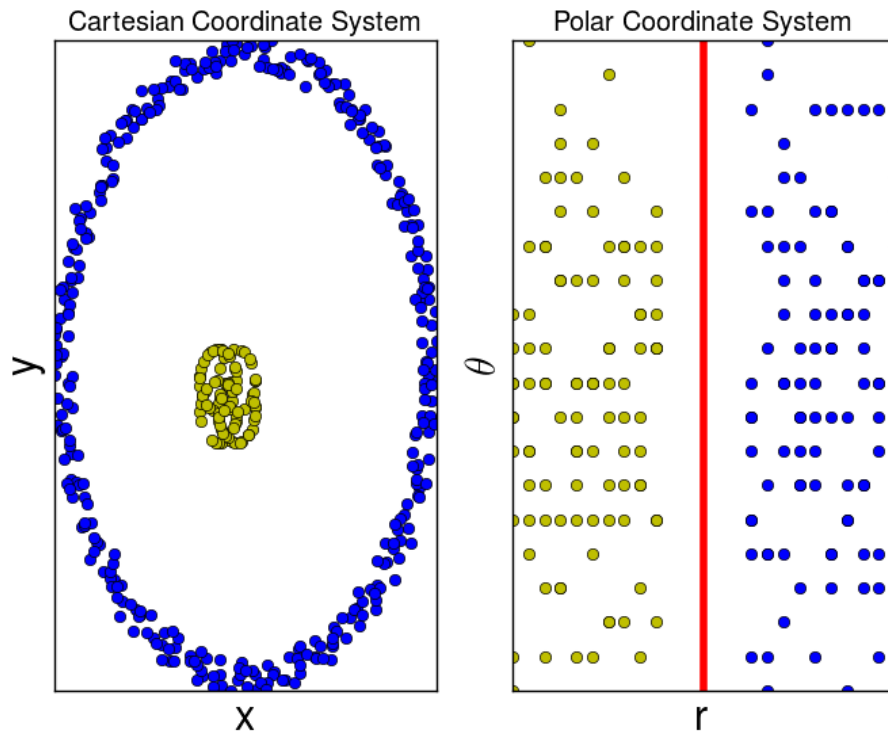
### Λειτουργία του SVM

Το SVM λειτουργεί ως εξής:

1. **Μετασχηματισμός Δεδομένων:** Τα δεδομένα μεταφέρονται σε έναν υψηλής διάστασης χώρο μέσω μιας μη γραμμικής συνάρτησης πυρήνα (*kernel function*), όπως η RBF (Radial Basis Function) ή ο πολυωνυμικός πυρήνας. Συχνά απαιτείται μετασχηματισμός των δεδομένων σε διαφορετικό σύστημα συντεταγμένων, ώστε αυτά να είναι γραμμικά διαχωρίσιμα, όπως φαίνεται στο [σχήμα 3.8](#).
2. **Κατασκευή Υπερεπιπέδου:** Το SVM επιλέγει το υπερεπίπεδο που μεγιστοποιεί το περιθώριο μεταξύ των δύο κλάσεων δεδομένων. Στο [σχήμα 3.7](#) με μαύρο χρώμα φαίνεται η ευθεία που μεγιστοποιεί αυτό το περιθώριο.
3. **Υποστήριξη Σημείων (Support Vectors):** Τα δεδομένα που βρίσκονται πλησιέστερα στο υπερεπίπεδο καλούνται *support vectors* και καθορίζουν τη θέση και τον προσανατολισμό του.



Σχήμα 3.7: Διαχωρισμός δεδομένων με τη χρήση του αλγορίθμου SVM. Με μαύρο φαίνεται η ευθεία που μεγιστοποιεί το περιθώριο ενώ με κόκκινο και πράσινο φαίνονται άλλες - μη βέλτιστες - επιλογές ευθειών διαχωρισμού



Σχήμα 3.8: Μη γραμμικά διαχωρίσιμα δεδομένα μετασχηματίζονται σε διαφορετικό χώρο υψηλής διάστασης

Το SVM είναι κατάλληλο για προβλήματα εποπτευόμενης μάθησης με δύο ή περισσότερες κλάσεις. Ωστόσο, όταν πρόκειται για ανίχνευση ανωμαλιών ή μοτίβων σε δεδομένα χωρίς ετικέτες, εισάγεται η επέκτασή του: το *One-Class SVM*.

### One-Class SVM: Εξειδίκευση για Ανίχνευση Ανωμαλιών

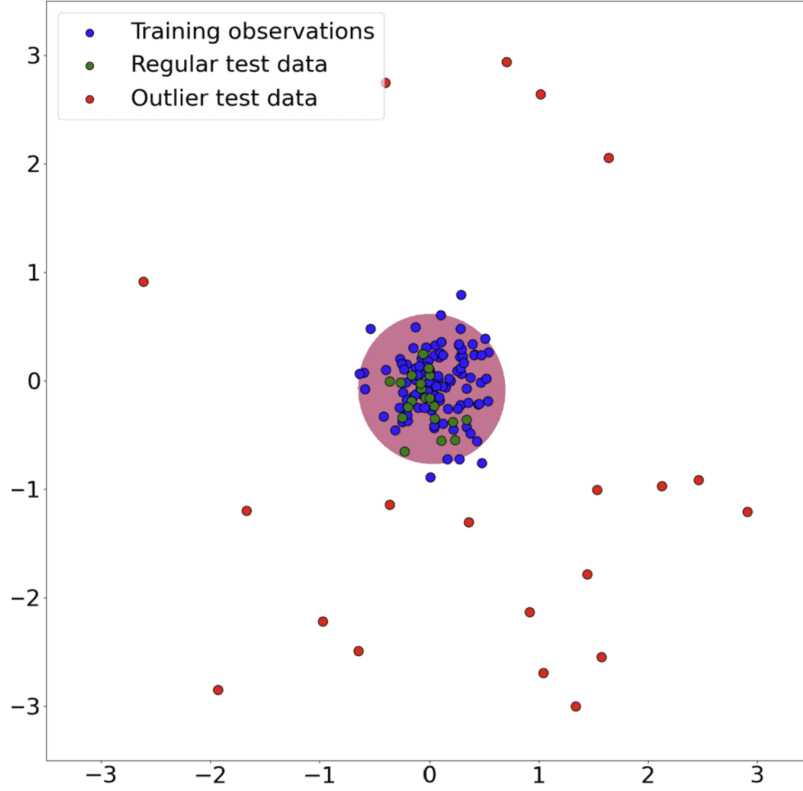
Το *One-Class Support Vector Machine* (One-Class SVM) είναι ένας αλγόριθμος μηχανικής μάθησης που ανήκει στην κατηγορία της μη εποπτευόμενης μάθησης και χρησιμοποιείται κυρίως για την ανίχνευση ανωμαλιών και την αναγνώριση μοτίβων σε δεδομένα. Αναπτύχθηκε από τους Schölkopf et al. [9], και αποτελεί μία εκτεταμένη εφαρμογή του κλασικού SVM, που έχει σχεδιαστεί για να μοντελοποιεί τη διανομή ενός μόνο κλάδου (class). Είναι μία επέκταση του SVM που χρησιμοποιείται για προβλήματα μη εποπτευόμενης μάθησης. Σκοπός του είναι να μοντελοποιήσει την κανονική κατανομή των δεδομένων και να αναγνωρίσει αποκλίσεις ή ανωμαλίες.

### Βασικές Αρχές

Το OC-SVM προσπαθεί να περικλείσει όλα τα κανονικά δεδομένα σε έναν υψηλής διάστασης χώρο μέσω μίας υπερ-επιφάνειας (*hyperplane*) ή ενός υπερσφαιρικού χώρου. Οτιδήποτε βρίσκεται εκτός αυτής της περιοχής θεωρείται ανωμαλία. Στο [σχήμα 3.9](#) με μπλε χρώμα φαίνονται τα δεδομένα εκπαίδευσης που καθορί-



ζουν τον υπερσφαιρικό χώρο. Με πράσινο φαίνονται τα δεδομένα που ανήκουν σε αυτόν τον χώρο ενώ με κόκκινο φαίνονται οι ανωμαλίες.



Σχήμα 3.9: Επίδειξη λειτουργίας OC-SVM σε δισδιάστατο χώρο

### Μαθηματική Διατύπωση

Η βασική ιδέα πίσω από το One-Class SVM είναι η αναπαράσταση του συνόλου δεδομένων σε έναν υψηλής διάστασης χώρο χαρακτηριστικών μέσω μίας μη γραμμικής συνάρτησης πυρήνα (*kernel function*). Στον χώρο αυτόν, το μοντέλο επιχειρεί να κατασκευάσει μία υπερ-επιφάνεια (*hyperplane*) που περικλείει τη μεγαλύτερη δυνατή ποσότητα των δεδομένων, ελαχιστοποιώντας παράλληλα την απόσταση των σημείων από την επιφάνεια.

Η μαθηματική διατύπωση περιλαμβάνει την επίλυση του εξής βελτιστοποιητικού προβλήματος:

$$\min_{\mathbf{w}, \xi, \rho} \frac{1}{2} \|\mathbf{w}\|^2 + \frac{1}{\nu N} \sum_{i=1}^N \xi_i - \rho,$$

υπό τους περιορισμούς:

$$(\mathbf{w} \cdot \phi(\mathbf{x}_i)) \geq \rho - \xi_i, \quad \xi_i \geq 0, \quad i = 1, \dots, N.$$

Όπου:

- $\mathbf{w}$ : Το διάνυσμα των παραμέτρων του μοντέλου.

- $\phi(\mathbf{x}_i)$ : Η συνάρτηση που μετασχηματίζει τα δεδομένα στον χώρο χαρακτηριστικών.
- $\xi_i$ : Οι μεταβλητές χαλάρωσης που επιτρέπουν ορισμένα σημεία να βρεθούν εκτός της επιφάνειας.
- $\rho$ : Η παράμετρος που καθορίζει την απόσταση του υπερεπιπέδου από την αρχή.
- $\nu$ : Ένας υπερπαράμετρος που ελέγχει το ποσοστό των σημείων που θεωρούνται εκτός της επιφάνειας.

### Διαδικασία Λειτουργίας

Η διαδικασία λειτουργίας του One-Class SVM περιλαμβάνει τα εξής στάδια:

1. **Εκπαίδευση:** Το μοντέλο εκπαιδεύεται σε ένα σύνολο δεδομένων που περιλαμβάνει μόνο τα κανονικά δεδομένα (genuine data). Ο στόχος είναι να εντοπίσει μία περιοχή στον χώρο χαρακτηριστικών που περικλείει τα δεδομένα αυτά.
2. **Αναγνώριση:** Κατά τη φάση δοκιμών, τα νέα δεδομένα αξιολογούνται με βάση την απόστασή τους από την υπερ-επιφάνεια. Σημεία που βρίσκονται εκτός της καθορισμένης περιοχής θεωρούνται ανωμαλίες ή impostor δεδομένα.
3. **Ενημέρωση:** Το μοντέλο μπορεί να προσαρμοστεί ώστε να ενσωματώσει νέα δεδομένα, βελτιώνοντας έτσι τη δυνατότητα ανίχνευσης μεταβαλλόμενων μοτίβων.

### Πλεονεκτήματα και Εφαρμογές

Το OC-SVM διακρίνεται για τα εξής:

- **Ικανότητα Ανίχνευσης Ανωμαλιών:** Το OC-SVM έχει αποδειχθεί εξαιρετικά αποτελεσματικό στην ανίχνευση ανωμαλιών, καθώς μπορεί να ανιχνεύσει μοτίβα σε δεδομένα χωρίς ετικέτες, καθιστώντας το ιδανικό για ανίχνευση απάτης ή βλάβης [9].
- **Ευελιξία:** Μπορεί να εφαρμοστεί σε διάφορους τύπους δεδομένων, από κείμενο και αριθμητικά δεδομένα μέχρι εικόνες [22].

Το One-Class SVM αποτελεί ένα θεμελιώδες εργαλείο για την ανίχνευση ανωμαλιών. Έχει εφαρμοστεί επιτυχώς σε πολλαπλά πεδία, όπως η ανίχνευση επιθέσεων σε συστήματα SCADA [11], η ανάλυση βιομετρικών δεδομένων [16], καθώς και σε προβλήματα ανίχνευσης και παρακολούθησης συστημάτων σε πραγματικό χρόνο [21]. Η χρήση του One-Class SVM στην παρούσα εργασία επιτρέπει τη δημιουργία εξατομικευμένων μοντέλων αυθεντικοποίησης και ανίχνευσης ανωμαλιών μεταξύ των χρηστών.





# 4

## Μεθοδολογία και Υλοποίηση

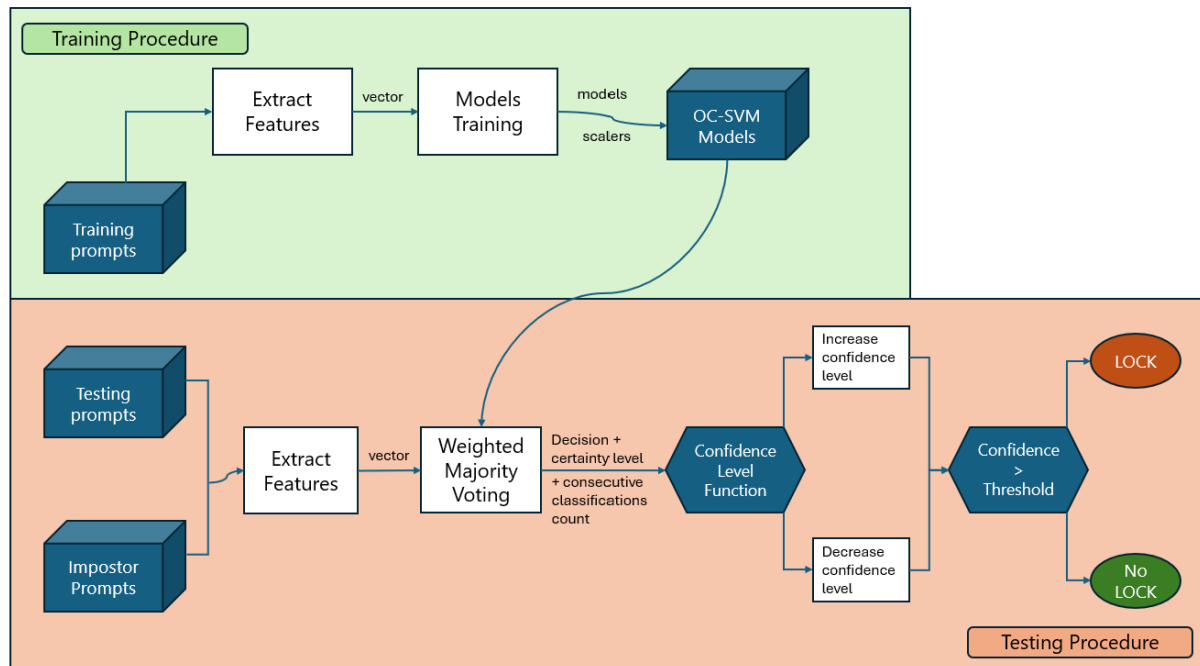
Η μεθοδολογία και η υλοποίηση αποτελούν τον πυρήνα της παρούσας εργασίας, καθώς περιγράφουν την προσέγγιση που ακολουθήθηκε για την ανάπτυξη του συστήματος αυθεντικοποίησης χρηστών. Η δομή του κεφαλαίου αναπτύσσεται με τρόπο ώστε να παρέχει μια καθολική και συστηματική παρουσίαση των διαδικασιών που ακολουθήθηκαν, από τη συλλογή των δεδομένων έως την αξιολόγηση των αποτελεσμάτων.

### 4.0.1 Αρχιτεκτονική Υψηλού Επιπέδου

Η συνολική αρχιτεκτονική του συστήματος συνιστά τον συνδυασμό όλων των επιμέρους ενοτήτων που περιγράφηκαν στα προηγούμενα τμήματα, σχηματίζοντας ένα ολοκληρωμένο σύστημα αυθεντικοποίησης. Η αρχιτεκτονική έχει σχεδιαστεί με τρόπο που να διασφαλίζει την ακριβή, γρήγορη και αξιόπιστη αναγνώριση χρηστών μέσω ανάλυσης γραφής. Επιπλέον, έχει ενσωματωθεί ένα απλό αλλά λειτουργικό περιβάλλον διεπαφής χρήστη (UI) με χρήση του Streamlit, επιτρέποντας την εύκολη πρόσβαση στο σύστημα και την παρακολούθηση των αποτελεσμάτων σε πραγματικό χρόνο.

Πιο συγκεκριμένα, το κεφάλαιο ξεκινά με την ανάλυση της συλλογής και της προεπεξεργασίας των δεδομένων, αναδεικνύοντας τη σημασία της ποιότητας των δεδομένων στη συνολική απόδοση του συστήματος. Στη συνέχεια, εστιάζει στην εξαγωγή χαρακτηριστικών, μια κρίσιμη διαδικασία που συνδέει τη θεωρητική βάση της επεξεργασίας φυσικής γλώσσας με την πρακτική της εφαρμογή. Η εκπαίδευση των μοντέλων μηχανικής μάθησης και η ενσωμάτωση ενός συστήματος εμπιστοσύνης σε συνδυασμό με το σύστημα απόφασης περιγράφονται με λεπτομέρεια, παρέχοντας πληροφορίες για τις τεχνικές και τις παραμέτρους που χρησιμοποιήθηκαν. Τέλος, παρουσιάζεται το σύστημα αξιολόγησης και ένα γραφικό περιβάλλον χρήστη.

Η αρχιτεκτονική υψηλού επιπέδου του συστήματος εκπαίδευσης, απόφασης και αξιολόγησης φαίνεται στο [σχήμα 4.1](#).



Σχήμα 4.1: Αρχιτεκτονική Υψηλού Επιπέδου του Συστήματος

#### 4.0.2 Εργαλεία και Πλατφόρμες Ανάπτυξης

Για την ανάπτυξη του συστήματος χρησιμοποιήθηκαν τα εξής:

- **Γλώσσες Προγραμματισμού:** Python (βιβλιοθήκες sklearn, NLTK, textstat, Streamlit).
- **Περιβάλλον Ανάπτυξης:** PyCharm και Visual Studio Code για την ανάπτυξη και δοκιμή του κώδικα. Github για διαδικασίες Continuous Integration - Continuous Deployment.

### 4.1 ΣΥΛΛΟΓΗ ΚΑΙ ΠΡΟΕΠΕΞΕΡΓΑΣΙΑ ΔΕΔΟΜΕΝΩΝ

Η ποιότητα και η επεξεργασία των δεδομένων αποτελούν κρίσιμο στάδιο για την επιτυχία ενός συστήματος συνεχούς και έμμεσης αυθεντικοποίησης. Στο παρόν σύστημα, η διαδικασία συλλογής και προεπεξεργασίας δεδομένων έχει σχεδιαστεί ώστε να διασφαλίζει την ακεραιότητα και την αξιοπιστία των χαρακτηριστικών που χρησιμοποιούνται για την ταυτοποίηση των χρηστών.

#### 4.1.1 Πηγή Δεδομένων

Τα δεδομένα που χρησιμοποιήθηκαν αντλήθηκαν από δημόσια διαθέσιμες πηγές, και συγκεκριμένα από το *Kaggle*<sup>4</sup>, το οποίο προσφέρει ένα ευρύ φάσμα από

<sup>4</sup><https://www.kaggle.com>

σετ δεδομένων (datasets). Το συγκεκριμένο dataset περιλαμβάνει πολλαπλές εγγραφές από διάφορους χρήστες, παρέχοντας ένα πλούσιο σύνολο δεδομένων για την εξαγωγή χαρακτηριστικών και την εκπαίδευση.

Επιπλέον, τα δεδομένα που περιλαμβάνονται στο συγκεκριμένο dataset αντλούν τη θεματολογία τους από κάθε πτυχή της ανθρώπινης δραστηριότητας, διασφαλίζοντας την ποικιλομορφία στα στυλ γραφής. Η κατηγοριοποίηση βάσει χρηστών επιτρέπει τη δημιουργία εξατομικευμένων μοντέλων και εξασφαλίζει την απαραίτητη ανομοιογένεια τόσο στην εκπαίδευση των μοντέλων, όσο και στον έλεγχο και στην αξιολόγηση του συστήματος.

#### 4.1.2 Διαδικασία Προεπεξεργασίας Δεδομένων

Η διαδικασία προεπεξεργασίας περιλάμβανε τα εξής βήματα:

1. **Καθαρισμός dataset:** Αφαιρέθηκαν στήλες που παρείχαν περιττή πληροφορία σχετικά με την εργασία. Με αυτόν τον τρόπο μειώθηκε και το μέγεθος του dataset εξασφαλίζοντας ταχύτερους χρόνους ανάγνωσης και φόρτωσης των δεδομένων.
2. **Καθαρισμός κειμένων:** Αφαιρέθηκαν από τα prompts τα URLs και HTML tags. Ο καθαρισμός διασφάλισε ότι τα δεδομένα περιείχαν μόνο πληροφοριακό περιεχόμενο.

Η διαδικασία προεπεξεργασίας εφαρμόστηκε με χρήση βιβλιοθηκών Python, όπως οι NLTK, spaCy, και textstat, οι οποίες διευκόλυναν την εξαγωγή και την ανάλυση χαρακτηριστικών.

#### 4.1.3 Χαρακτηριστικά του Σετ Δεδομένων

Το dataset<sup>5</sup> περιλαμβάνει κείμενα από 14 διαφορετικούς χρήστες, με μέσο όρο 2.000 prompts ανά χρήστη. Ο πίνακας 4.1 συνοψίζει βασικά στατιστικά στοιχεία του dataset:

Χαρακτηριστικό	Τιμή
Αριθμός Χρηστών	14
Συνολικά prompts ανά χρήστη	2.000
Μέσος Όρος Λέξεων ανά Κείμενο	X
Μέγιστο Μήκος Κειμένου	X

Πίνακας 4.1: Στατιστικά Χαρακτηριστικά του Dataset

Επιπλέον, το dataset περιλαμβάνει τις παρακάτω στήλες, οι οποίες περιγράφονται στον πίνακα 4.2 και φαίνονται στο [σχήμα 4.2](#):

<sup>5</sup><https://www.kaggle.com/datasets/mmmarchetti/tweets-dataset>

Στήλη	Περιγραφή
author	Ταυτοποιητικό πεδίο του χρήστη που δημιούργησε το κείμενο.
content	Το κύριο σώμα του κειμένου, περιέχει το prompt προς ανάλυση.
date_time	Ημερομηνία και ώρα δημιουργίας του κειμένου.
id	Μοναδικό αναγνωριστικό για κάθε prompt στο dataset.

Πίνακας 4.2: Περιγραφή Στηλών του Dataset

```

1 author,content,date_time,id
2 BarackObama,"Tonight, President Obama reflects on eight years of progress. Watch the
3 BarackObama,"In the weekly address, President Obama discusses what #Obamacare has do
4 BarackObama,"Let's keep working to keep our economy on a better, stronger course.",0
5 BarackObama,"The landmark #ParisAgreement enters into force today—we must keep up the
6 BarackObama,"The economy added 161,000 jobs in October, and wages are up 2.8 percent
7 BarackObama,"There are a lot of plans out there. Check your options and lock in the o
8 BarackObama,"The positive impact of #Obamacare is undeniable, but there's one big fa
9 BarackObama,"Tens of millions of Americans have benefited from #Obamacare. Make sure
10 BarackObama,"Thanks to #Obamacare, quality health care is available to everyone. The
11 BarackObama,"Community organizing never goes out of style. Shop now:,01/11/2016 20:17
12 BarackObama,"With #Obamacare, people can focus on treatment for pre-existing conditi
13 BarackObama,"The Obamacare marketplace is now open. If you're uninsured, now is the
14 BarackObama,"Lions and Tiggers and bears! Oh my! #HappyHalloween,31/10/2016 22:09,7.9
15 BarackObama,"Usted y su familia merecen la tranquilidad de saber que están cubiertos.

```

Σχήμα 4.2: Στιγμιότυπο οθόνης από το σετ δεδομένων που χρησιμοποιήθηκε - φαίνονται οι 4 στήλες που αναφέρονται παραπάνω καθώς και πολλαπλές καταχωρίσεις

Η ποικιλομορφία στα δεδομένα επιτρέπει την εκπαίδευση μοντέλων ικανά να αναγνωρίζουν διαφορετικά στυλ γραφής. Η διασφάλιση της ποιότητας και της ομοιογένειας του dataset αποτέλεσε θεμελιώδη παράγοντα για την επιτυχία των επόμενων σταδίων.

Τα δεδομένα που προκύπτουν μετά την προεπεξεργασία αποθηκεύονται στη δομή `data/filtered_cleaned.csv` στο αποθετήριο. Η διαχείριση των δεδομένων γίνεται μέσω βιβλιοθηκών Python όπως `pandas`<sup>6</sup> για την ανάγνωση και εγγραφή αρχείων CSV και `os`<sup>7</sup> για την οργάνωση των φακέλων.

## 4.2 ΕΞΑΓΩΓΗ ΧΑΡΑΚΤΗΡΙΣΤΙΚΩΝ

Η εξαγωγή χαρακτηριστικών αποτελεί ένα από τα πιο σημαντικά βήματα στη διαδικασία ανάπτυξης ενός συστήματος αυθεντικοποίησης. Τα χαρακτηριστικά που εξάγονται από τα δεδομένα γραφής περιγράφουν μοναδικές πτυχές του στυλ γραφής κάθε χρήστη, επιτρέποντας έτσι την αναγνώρισή τους. Στην παρούσα εργασία, η εξαγωγή χαρακτηριστικών βασίζεται τόσο σε γλωσσικά όσο και σε συμπεριφορικά χαρακτηριστικά, διασφαλίζοντας ότι λαμβάνονται υπόψη η δομή, το περιεχόμενο του κειμένου αλλά και οι ιδιαιτερότητες των συγγραφικών μοτίβων.

Η διαδικασία εξαγωγής περιλαμβάνει τη χρήση βιβλιοθηκών Python, όπως οι `NLTK`<sup>8</sup>, `textstat`<sup>9</sup>, και `numpy`<sup>10</sup>, για την ανάλυση και ποσοτικοποίηση γλωσσικών

<sup>6</sup><https://pandas.pydata.org/>

<sup>7</sup><https://docs.python.org/3/library/os.html>

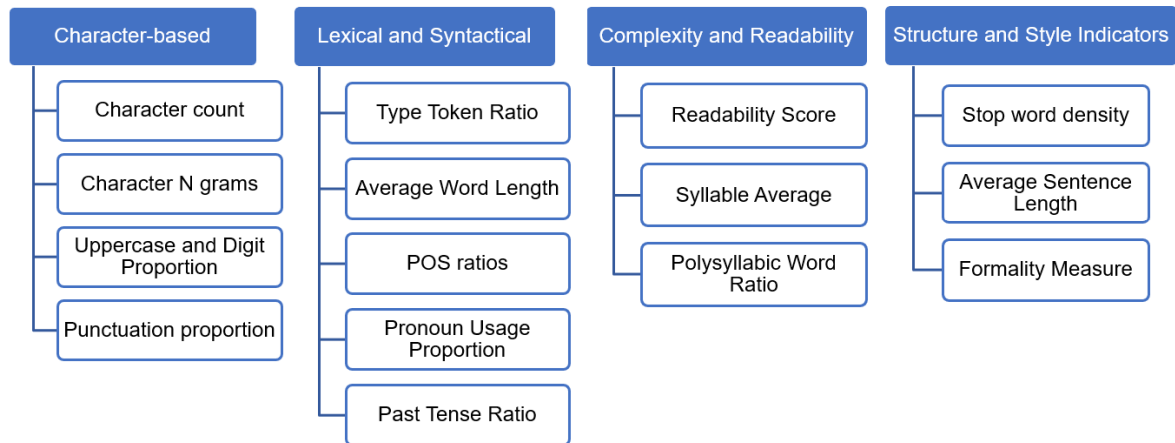
<sup>8</sup><https://www.nltk.org/>

<sup>9</sup><https://textstat.org/>

<sup>10</sup><https://numpy.org/>



και δομικών χαρακτηριστικών. Στη συνέχεια, τα χαρακτηριστικά αυτά ομαδοποιούνται σε 4 θεματικές κατηγορίες, χαρακτηριστικά που βασίζονται σε χαρακτήρες, λεξιλογικά και συντακτικά χαρακτηριστικά, δείκτες πολυπλοκότητας και αναγνωσιμότητας και δομικοί και στυλιστικοί δείκτες, όπως φαίνεται στο [σχήμα 4.3](#), διευκολύνοντας την κατανόηση της συμβολής τους στη διαδικασία αυθεντικοποίησης.



Σχήμα 4.3: Τα χαρακτηριστικά που εξάγονται στη παρούσα εργασία

### 4.2.1 Χαρακτηριστικά που Βασίζονται στους Χαρακτήρες

#### Char Count (Normalized)

Μετρά τον συνολικό αριθμό χαρακτήρων ενός κειμένου, κανονικοποιημένο με βάση το μήκος των 100 χαρακτήρων:

$$\text{Char Count Norm} = \frac{\text{Total Characters}}{100}$$

Αυτή η μέτρηση εξασφαλίζει συγκρισιμότητα μεταξύ κειμένων διαφορετικών μεγεθών.

#### Character N-grams (3-grams) Ratio

Υπολογίζει την αναλογία των τριγραμμάτων (3-grams) στους συνολικούς χαρακτήρες:

$$\text{Char 3-grams Ratio} = \frac{\text{Number of 3-grams}}{\text{Total Characters}}$$

Τα τριγράμματα είναι ακολουθίες τριών διαδοχικών χαρακτήρων (π.χ., "the", "igh"), και αποτυπώνουν μοτίβα γραφής του χρήστη.

### 4.2.2 Λεξιλογικά και Συντακτικά Χαρακτηριστικά

#### Stop Word Frequency Ratio

Υπολογίζει τον αριθμό των *stop words* που περιέχονται σε ένα κείμενο ως ποσοστό του συνολικού αριθμού λέξεων. Η εξίσωση που χρησιμοποιείται είναι η εξής:

$$\text{Stop Word Ratio} = \frac{\text{Stop Words Count}}{\text{Total Words Count}}$$

Τα *stop words* είναι λέξεις όπως: "the", "is", "in", "it", "of", "and", "to", "a", "that", "with", "as", "for", "on", "at", "by". Αυτές οι λέξεις είναι συχνές στη γλώσσα αλλά δεν παρέχουν σημαντικές πληροφορίες για το περιεχόμενο του κειμένου.

Παρά την έλλειψη νοηματικού βάρους, η κατανομή των *stop words* μπορεί να διαφέρει σημαντικά μεταξύ των χρηστών, καθώς επηρεάζεται από τον τρόπο γραφής τους. Για παράδειγμα, ορισμένοι χρήστες μπορεί να έχουν την τάση να χρησιμοποιούν *stop words* πιο συχνά για να συνδέουν φράσεις ή να δημιουργούν ροή στο κείμενο τους, γεγονός που αποτελεί χαρακτηριστικό προσωπικού ύφους.

#### Type-Token Ratio (Lexical Diversity)

Μετρά τη λεξιλογική ποικιλία του κειμένου:

$$\text{Type-Token Ratio} = \frac{\text{Unique Words Count}}{\text{Total Words Count}}$$

Μεγαλύτερη τιμή υποδηλώνει πιο ποικιλόμορφο λεξιλόγιο.

#### Part-of-Speech Ratios

Τα ποσοστά των μερών του λόγου (όπως ουσιαστικά, ρήματα, επίθετα, επιρρήματα) υπολογίζονται ως εξής:

$$\text{POS Ratio} = \frac{\text{POS Count}}{\text{Total Words Count}}$$

Για παράδειγμα, το *Adjective Ratio* αφορά τη συχνότητα των επιθέτων (JJ), ενώ το *Verb Ratio* περιλαμβάνει όλες τις μορφές ρημάτων (VB, VBD, VBG, VBN, VBP, VBZ).

### 4.2.3 Δείκτες Πολυπλοκότητας και Αναγνωσιμότητας

#### Readability Score (Flesch Reading Ease)

Ο δείκτης αναγνωσιμότητας *Flesch Reading Ease* [35] υπολογίζεται με βάση την ακόλουθη εξίσωση:

$$\text{Flesch Score} = 206.835 - 1.015 \left( \frac{\text{Total Words}}{\text{Total Sentences}} \right) - 84.6 \left( \frac{\text{Total Syllables}}{\text{Total Words}} \right)$$

Μεγαλύτερες τιμές υποδηλώνουν πιο εύκολα αναγνώσιμα κείμενα.

### Syllable Average

Ο μέσος όρος συλλαβών ανά λέξη υπολογίζεται ως:

$$\text{Syllable Avg} = \frac{\text{Total Syllables Count}}{\text{Total Words Count}}$$

### Polysyllabic Word Ratio

Μετρά την αναλογία των πολυσύλλαβων λέξεων:

$$\text{Polysyllabic Ratio} = \frac{\text{Polysyllabic Words Count}}{\text{Total Words Count}}$$

όπου πολυσύλλαβες είναι οι λέξεις με περισσότερες από τρεις συλλαβές.

## 4.2.4 Δομικοί και Στυλιστικοί Δείκτες

### Average Sentence Length

Το μέσο μήκος προτάσεων υπολογίζεται με βάση τον αριθμό λέξεων:

$$\text{Avg Sentence Length} = \frac{\text{Total Words}}{\text{Total Sentences}}$$

### Pronoun Usage Proportion

Υπολογίζει την αναλογία των αντωνυμιών στο κείμενο:

$$\text{Pronoun Proportion} = \frac{\text{Pronoun Count}}{\text{Total Words Count}}$$

Οι αντωνυμίες που περιλαμβάνονται είναι: "I", "you", "he", "she", "it", "we", "they", "me", "us", "him", "her", "them".

### Formality Measure

Ο δείκτης επιστημότητας μετράται ως εξής:

$$\text{Formality} = \frac{\text{Noun Count} + \text{Adjective Count}}{\text{Pronoun Count} + \text{Verb Count} + 0.01}$$

Μεγαλύτερη τιμή υποδηλώνει πιο επίσημο ύφος γραφής.

## 4.2.5 Σύνοψη χαρακτηριστικών

Ο πίνακας [πίνακα 4.3](#) περιλαμβάνει συνοπτικά όλα τα χαρακτηριστικά που εξάγονται στην παρούσα εργασία:

Χαρακτηριστικό	Κατηγορία
Char Count (Normalized)	Character-based Features
Character N-grams (3-grams) Ratio	Character-based Features
Uppercase Proportion	Character-based Features
Digit Proportion	Character-based Features
Punctuation Proportion	Character-based Features
Type-Token Ratio	Lexical and Syntactical Features
Average Word Length	Lexical and Syntactical Features
Part-of-Speech Ratios	Lexical and Syntactical Features
Pronoun Usage Proportion	Lexical and Syntactical Features
Past Tense Ratio	Lexical and Syntactical Features
Readability Score	Complexity and Readability Indicators
Syllable Avg	Complexity and Readability Indicators
Polysyllabic Word Ratio	Complexity and Readability Indicators
Stop Word Density	Structure and Style Indicators
Avg Sentence Length	Structure and Style Indicators
Formality Measure	Structure and Style Indicators

Πίνακας 4.3: Κατηγοριοποίηση Χαρακτηριστικών Εξαγωγής

#### 4.2.6 Εργαλεία και Μέθοδοι Εξαγωγής Χαρακτηριστικών

Η συνάρτηση `extract_features` είναι η κύρια μέθοδος εξαγωγής χαρακτηριστικών στο σύστημα αυθεντικοποίησης χρηστών. Η συνάρτηση λαμβάνει ως είσοδο ένα κείμενο (`text`), το οποίο μπορεί να περιλαμβάνει προτάσεις, παραγράφους ή μεγαλύτερα τμήματα γραπτού λόγου. Επεξεργάζεται το κείμενο και εξάγει ένα σύνολο χαρακτηριστικών, τα οποία επιστρέφονται ως ένα πολυδιάστατο διάνυσμα (`list of feature values`).

Η έξοδος της συνάρτησης περιλαμβάνει τιμές για όλα τα χαρακτηριστικά που περιγράφηκαν στις προηγούμενες υποενότητες, όπως οι αναλογίες χαρακτήρων, λέξεων, στατιστικά στυλ, αναγνωσιμότητα και άλλοι δείκτες. Το διάνυσμα αυτό αποτελεί την είσοδο για τα επόμενα στάδια ανάλυσης, εκπαίδευσης αλλά και αξιολόγησης, επιτρέποντας την αποτελεσματική αυθεντικοποίηση χρηστών.

### 4.3 ΕΚΠΑΙΔΕΥΣΗ ΜΟΝΤΕΛΩΝ

Η εκπαίδευση των μοντέλων αποτελεί έναν κρίσιμο πυλώνα του συστήματος αυθεντικοποίησης. Στη συγκεκριμένη εργασία, βασιζόμαστε στον αλγόριθμο *One-Class SVM*, όπως έχει περιγραφεί στην [ενότητα 3.3.3](#), για να εκπαιδεύσουμε τα μοντέλα που αναγνωρίζουν τα πρότυπα γραφής ενός χρήστη και απορρίπτουν τυχόν απόπειρες παραβίασης. Σε αυτή την ενότητα περιγράφεται αναλυτικά η διαδικασία εκπαίδευσης, η επιλογή των υπερπαραμέτρων και οι στρατηγικές βελτιστοποίησης.

### 4.3.1 Προεπεξεργασία Δεδομένων Εκπαίδευσης

Η προεπεξεργασία των δεδομένων περιλαμβάνει τα παρακάτω στάδια:

1. **Ανάγνωση δεδομένων:** Το αρχείο δεδομένων περιέχει τις στήλες `author` και `content`. Τα δεδομένα διαχωρίζονται με βάση τον συγγραφέα.
2. **Διαχωρισμός σε σύνολα:**
  - Σύνολο Εκπαίδευσης: 85% των δεδομένων.
  - Σύνολο Δοκιμών: 15% των δεδομένων, αποθηκεύεται για αξιολόγηση.
3. **Κανονικοποίηση δεδομένων:** Τα εξαγόμενα χαρακτηριστικά κανονικοποιούνται χρησιμοποιώντας τον `StandardScaler`, ώστε να έχουν μέσο όρο 0 και τυπική απόκλιση 1. Αυτή η διαδικασία είναι απαραίτητη, διότι διαφορετικά χαρακτηριστικά μπορεί να έχουν διαφορετικές κλίμακες (π.χ., ο μέσος όρος συλλαβών ανά λέξη κυμαίνεται από 0 έως 1, ενώ ο αριθμός χαρακτήρων μπορεί να είναι εκατοντάδες). Εάν τα δεδομένα δεν κανονικοποιηθούν, τα χαρακτηριστικά με μεγαλύτερες αριθμητικές τιμές ενδέχεται να έχουν μεγαλύτερη επίδραση στην εκπαίδευση και απόφαση του μοντέλου, καθιστώντας το μη αντικειμενικό. Η κανονικοποίηση εξασφαλίζει ότι όλα τα χαρακτηριστικά έχουν την ίδια αριθμητική βαρύτητα και συμβάλλει στη σωστή λειτουργία των αλγορίθμων μηχανικής μάθησης, όπως το `One-Class SVM`, οι οποίοι βασίζονται στη μέτρηση αποστάσεων.

### 4.3.2 Εξαγωγή και Κανονικοποίηση Χαρακτηριστικών

Τα χαρακτηριστικά που εξάγονται από τα δεδομένα εκπαίδευσης περιγράφονται αναλυτικά στο [υποκεφάλαιο 4.2](#). Μετά την εξαγωγή, εφαρμόζεται κανονικοποίηση για να διασφαλιστεί η συγκρισιμότητα μεταξύ διαφορετικών χαρακτηριστικών.

### 4.3.3 Εκπαίδευση Μοντέλων *One-Class SVM*

Η εκπαίδευση πραγματοποιείται με τη χρήση του αλγορίθμου *One-Class SVM* με πυρήνα `rbf`. Ο αλγόριθμος αυτός είναι κατάλληλος για εφαρμογές όπου υπάρχουν δεδομένα μόνο από μία κατηγορία (εν προκειμένω, του χρήστη) - *one-vs-all classification*. Το μοντέλο μαθαίνει τα πρότυπα της κατηγορίας αυτής και απορρίπτει ανωμαλίες.

#### Υπερπαράμετροι του *One-Class SVM*

Οι υπερπαράμετροι του *One-Class SVM* που ρυθμίστηκαν είναι:

- **`ν (nu)`:** Ελέγχει το ποσοστό των *outliers* που το μοντέλο θα ανεχθεί.
  - Τιμές που δοκιμάστηκαν: 0.001, 0.005, 0.01.
- **`gamma`:** Ελέγχει την ακτίνα επιρροής ενός δείγματος.

- Τιμές που δοκιμάστηκαν: 0.05, 0.07, 0.1, 0.15, 0.2, 0.5.

Για κάθε χρήστη, εκπαιδεύτηκαν 18 μοντέλα για όλους τους συνδυασμούς των παραπάνω τιμών. Τα μοντέλα αποθηκεύτηκαν στη μνήμη του συστήματος μαζί με τα μοντέλα για κανονικοποίηση νέων δεδομένων.

### 4.3.4 Hyperparameter Tuning και Threshold Tuning

#### Tuning των Υπερπαραμέτρων

Οι υπερπαραμέτροι ρυθμίστηκαν μέσω πειραματισμών:

- **$\nu$  (nu):**
  - Χαμηλές τιμές αυξάνουν την ευαισθησία του μοντέλου.
  - Υψηλές τιμές επιτρέπουν μεγαλύτερη ανοχή σε *outliers*.
- **gamma:**
  - Χαμηλές τιμές ορίζουν ευρύτερες περιοχές απόφασης.
  - Υψηλές τιμές επικεντρώνονται σε τοπικά μοτίβα.

#### Threshold Tuning

Το threshold (κατώφλι) είναι μια κρίσιμη παράμετρος στη διαδικασία απόφασης του μοντέλου, καθώς καθορίζει πώς θα ερμηνευτούν οι προβλέψεις για νέες εισόδους. Συγκεκριμένα, το threshold θέτει το σημείο στο οποίο το μοντέλο αποφασίζει αν μια νέα είσοδος ανήκει στην κατηγορία του γνήσιου χρήστη ή ερμηνεύεται ως απόρριψη.

#### Λειτουργία του threshold

- **Positive Decision:** Αν η απόσταση ή η πρόβλεψη ενός δείγματος είναι μεγαλύτερη από το threshold, το μοντέλο το αναγνωρίζει ως γνήσιο δείγμα.
- **Negative Decision:** Αν είναι μικρότερη, το μοντέλο απορρίπτει το δείγμα, θεωρώντας το ως μη γνήσιο.

**Προσαρμογή Threshold** Η ρύθμιση του threshold γίνεται με πειραματισμό σε διάφορες τιμές μέσα σε ένα εύρος, στη συγκεκριμένη περίπτωση μεταξύ  $[-0.1, 0.8]$ .

- **Χαμηλό Threshold:** Εάν το threshold είναι κοντά στο  $-0.1$ , το μοντέλο είναι πιο δεκτικό και αποδέχεται περισσότερα δείγματα ως γνήσια, αυξάνοντας την πιθανότητα λανθασμένων θετικών προβλέψεων.
- **Υψηλό Threshold:** Εάν το threshold είναι κοντά στο  $0.8$ , το μοντέλο γίνεται πιο αυστηρό, απορρίπτοντας περισσότερα δείγματα ως απατεώνες. Έτσι μειώνεται η πιθανότητα λανθασμένης αποδοχής δειγμάτων ως γνήσια, αλλά αυξάνεται η πιθανότητα λανθασμένης απόρριψης γνήσιων δειγμάτων.

Η βέλτιστη ρύθμιση του threshold επιτυγχάνει έναν αποδεκτό συμβιβασμό μεταξύ ασφάλειας και χρηστικότητας, καθιστώντας το σύστημα πιο αποτελεσματικό σε πραγματικές συνθήκες.

## 4.4 ΣΥΣΤΗΜΑ ΑΠΟΦΑΣΗΣ

---

Το σύστημα απόφασης αποτελεί κρίσιμο μέρος της συνολικής αρχιτεκτονικής του συστήματος αυθεντικοποίησης. Αξιοποιεί τα χαρακτηριστικά που εξάγονται από τα δεδομένα, καθώς και τα αποτελέσματα των μοντέλων που εκπαιδεύτηκαν, για να παρέχει τεκμηριωμένες και αξιόπιστες αποφάσεις σχετικά με τη γνησιότητα του χρήστη.

### 4.4.1 Εισαγωγή

Η διαδικασία απόφασης συνδυάζει τα αποτελέσματα των εκπαιδευμένων μοντέλων με τη χρήση της συνάρτησης επιπέδου βεβαιότητας και του αλγορίθμου σταθμισμένης πλειοψηφίας ψήφων. Το τελικό αποτέλεσμα εξαρτάται από:

- Την απόσταση από το hyperplane κάθε μοντέλου.
- Το επίπεδο βεβαιότητας που αντιστοιχεί σε κάθε απόσταση.
- Το κατώφλι απόφασης (*decision threshold*) που καθορίζει τη συμπεριφορά του συστήματος και την κατηγοριοποίηση της απόφασης.

### 4.4.2 Συνάρτηση Επιπέδου Βεβαιότητας

Η συνάρτηση επιπέδου βεβαιότητας (*certainty\_level*) υπολογίζει το πόσο σίγουρο είναι ένα μοντέλο για την απόφασή του. Ορίζεται ως εξής:

$$c(x) = \begin{cases} \frac{|d(x)|}{d_{\max}}, & \text{αν } |d(x)| < d_{\max} \\ 1, & \text{αν } d(x) > d_{\max} \text{ και } y = 1 \\ -1, & \text{αν } d(x) > d_{\max} \text{ και } y = -1 \end{cases} \quad (4.1)$$

όπου:

- $d(x)$  είναι η απόσταση του σημείου εισόδου  $x$  από το hyperplane του μοντέλου.
- $d_{\max}$  είναι το μέγιστο όριο απόστασης που καθορίζει το επίπεδο βεβαιότητας.
- $y$  είναι η προβλεπόμενη ετικέτα (1 για γνήσιος χρήστης, -1 για εισβολέας).

### 4.4.3 Υποσύστημα Ψηφοφορίας

Σε αυτή την ενότητα παρουσιάζονται δύο διαφορετικές προσεγγίσεις για τη λήψη απόφασης, η απλή πλειοψηφική συνάρτηση και η σταθμισμένη πλειοψηφική συνάρτηση. Οι δύο αυτές μέθοδοι διαφέρουν ως προς τη πολυπλοκότητα και τη φιλοσοφία της διαδικασίας λήψης αποφάσεων. Η σύγκρισή τους πραγματοποιείται σε επόμενο κεφάλαιο.

### Απλή Πλειοψηφική Συνάρτηση

Η πρώτη προσέγγιση συνδυάζει τα αποτελέσματα πολλαπλών μοντέλων με την ακόλουθη διαδικασία: κάθε μοντέλο  $i$  παράγει μια πρόβλεψη  $y_i$ : 1 για γνήσιο χρήστη και -1 για απατεώνα. Το τελικό αποτέλεσμα υπολογίζεται ως εξής:

$$\text{Decision} = \begin{cases} 1, & \text{αν } \sum_{i=1}^N y_i > 0 \\ -1, & \text{αν } \sum_{i=1}^N y_i \leq 0 \end{cases} \quad (4.2)$$

όπου:

- $N$  είναι το πλήθος των μοντέλων.
- $y_i$  είναι η πρόβλεψη του μοντέλου  $i$ .

### Σταθμισμένη Πλειοψηφική Συνάρτηση

Η δεύτερη προσέγγιση συνδυάζει τα αποτελέσματα πολλαπλών μοντέλων χρησιμοποιώντας τη σταθμισμένη πλειοψηφική συνάρτηση. Κάθε μοντέλο  $i$  παράγει μια πρόβλεψη  $y_i$  και μια βαρύτητα ψήφου  $w_i$ , που βασίζεται στο επίπεδο βεβαιότητας. Το τελικό αποτέλεσμα υπολογίζεται ως εξής:

$$\text{Decision} = \begin{cases} 1, & \text{αν } \sum_{i=1}^N w_i y_i > 0 \\ -1, & \text{αν } \sum_{i=1}^N w_i y_i \leq 0 \end{cases} \quad (4.3)$$

όπου:

- $w_i = |c_i(x)|$ , το απόλυτο επίπεδο βεβαιότητας του μοντέλου  $i$ .
- $N$  είναι το πλήθος των μοντέλων.
- $y_i$  είναι η πρόβλεψη του μοντέλου  $i$ .

#### 4.4.4 Παραδείγματα Εφαρμογής

Για να κατανοηθεί καλύτερα η λειτουργία του συστήματος λήψης αποφάσεων, παρατίθενται παραδείγματα εφαρμογής. Στο πλαίσιο αυτό, οι  $c_1, c_2, c_3, \dots, c_n$  αντιπροσωπεύουν τις αποφάσεις που λαμβάνονται από διαφορετικά μοντέλα για ένα δείγμα  $x$ . Κάθε  $c_i(x)$  εκφράζει τη βαθμολογία εμπιστοσύνης (*confidence score*) που προκύπτει από το αντίστοιχο μοντέλο, με θετικές τιμές να υποδηλώνουν γνήσιο χρήστη (*genuine user*) και αρνητικές τιμές να υποδηλώνουν εισβολέα (*impostor*).

- **Περίπτωση Γνήσιου Χρήστη:**

$$c_1(x) = 0.8, \quad c_2(x) = 0.7, \quad c_3(x) = 0.6$$

- **Απλή Πλειοψηφία:** Όλα τα μοντέλα αποφασίζουν ότι πρόκειται για γνήσιο χρήστη και αναθέτουν τη τιμή 1 στα  $c_i(x)$ :

$$\text{Απλή Απόφαση} = 1 + 1 + 1 = 3 > 0 \quad (\text{Γνήσιος Χρήστης})$$



- **Σταθμισμένη Πλειοψηφία:** Όλα τα μοντέλα αποφασίζουν ότι πρόκειται για γνήσιο χρήστη και αναθέτουν τη τιμή 1 πολλαπλασιασμένη με το βάρος της βεβαιότητας κάθε μοντέλου:

Σταθμισμένη Απόφαση =  $0.8 * 1 + 0.7 * 1 + 0.6 * 1 = 2.1 > 0$  (Γνήσιος Χρήστης)

• **Περίπτωση Εισβολέα:**

$$c_1(x) = -0.8, \quad c_2(x) = -0.7, \quad c_3(x) = -0.6$$

- **Απλή Πλειοψηφία:** Όλα τα μοντέλα αποφασίζουν ότι πρόκειται για εισβολέα και αναθέτουν τη τιμή -1 στα  $c_i(x)$ :

$$\text{Απλή Απόφαση} = (-1) + (-1) + (-1) = -3 < 0 \quad (\text{Εισβολέας})$$

- **Σταθμισμένη Πλειοψηφία:** Όλα τα μοντέλα αποφασίζουν ότι πρόκειται για εισβολέα και αναθέτουν τη τιμή -1 πολλαπλασιασμένη με το βάρος της βεβαιότητας κάθε μοντέλου:

$$\text{Σταθμισμένη Απόφαση} = 0.8 * (-1) + 0.7 * (-1) + 0.6 * (-1) = -2.1 < 0 \quad (\text{Εισβολέας})$$

Με αυτόν τον τρόπο, η σταθμισμένη πλειοψηφική ψήφος λαμβάνει υπόψη τη βαρύτητα κάθε  $c_i(x)$ , ενώ η βασική ψήφος βασίζεται αποκλειστικά στο πρόσημο της απόφασης του κάθε μοντέλου.

## 4.5 ΣΥΣΤΗΜΑ ΚΛΕΙΔΩΜΑΤΟΣ/ΕΜΠΙΣΤΟΣΥΝΗΣ

---

### 4.5.1 Συνάρτηση Εμπιστοσύνης

Η ασφάλεια ενός συστήματος αυθεντικοποίησης εξαρτάται όχι μόνο από την ακρίβεια των μοντέλων μηχανικής μάθησης, αλλά και από τη δυνατότητά του να διαχειρίζεται καταστάσεις όπου οι αποφάσεις μπορεί να είναι αβέβαιες ή να υπόκεινται σε διαδοχικά λάθη. Το σύστημα κλειδώματος/εμπιστοσύνης ενσωματώνει έναν μηχανισμό παρακολούθησης του επιπέδου εμπιστοσύνης ( $C$ ) του συστήματος προς τον χρήστη. Αυτό το επίπεδο αυξάνεται ή μειώνεται ανάλογα με την απόδοση του χρήστη, και σε περίπτωση που το  $C$  πέσει κάτω από ένα όριο (`confidence_threshold`), το σύστημα ενεργοποιεί μηχανισμούς κλειδώματος για την προστασία από κακόβουλη χρήση.

#### Συνάρτηση Εμπιστοσύνης

Η συνάρτηση εμπιστοσύνης περιγράφεται μαθηματικά ως:

$$C = \begin{cases} C + \text{base\_increase} + \begin{cases} \text{highCertaintyBoost\_increase}, & \text{if } \text{certaintyScore} > \text{highCertaintyThreshold} \\ 0, & \text{otherwise} \end{cases} + \begin{cases} \text{consecutiveGenuineBoost}, & \text{if } \text{consec\_genuine} = 2 \\ 0, & \text{otherwise} \end{cases}, & \text{for genuine} \\ C - \text{base\_decrease} - \begin{cases} \text{highCertaintyBoost\_decrease}, & \text{if } \text{certaintyScore} > \text{highCertaintyThreshold} \\ 0, & \text{otherwise} \end{cases} - \begin{cases} \text{consecutiveImpostorPenalty}, & \text{if } \text{consec\_impostor} = 3 \\ 0, & \text{otherwise} \end{cases}, & \text{for impostor} \end{cases}$$

Σχήμα 4.4: Συνάρτηση Confidence Level

### Βασικές Σταθερές

Υστερα από μεγάλο αριθμό δοκιμών, οι τιμές των βασικών παραμέτρων της συνάρτησης του σχήματος 4.4 καθορίστηκαν στις παρακάτω. Σε επόμενο κεφάλαιο συγκρίνεται η απόδοση της συνάρτησης εμπιστοσύνης με διαφορετικές τιμές των βασικών παραμέτρων.

- $\text{base}_{\text{increase}} = 0.06$
- $\text{base}_{\text{decrease}} = 0.12$
- $\text{high\_certainty\_threshold} = 0.7$
- $\text{high\_certainty\_boost\_factor} = 0.4$
- $\text{consecutive\_genuine\_boost} = 0.04$
- $\text{consecutive\_impostor\_penalty} = 0.05$
- $\text{confidence\_threshold} = 0.3$
- αρχικό επίπεδο εμπιστοσύνης:  $C_0 = 0.6$

### Υπολογισμός Ενισχύσεων Υψηλής Βεβαιότητας

Οι ενισχύσεις λόγω υψηλής βεβαιότητας υπολογίζονται ως:

$$\text{high\_certainty\_boost\_increase} = \text{base}_{\text{increase}} \times \text{high\_certainty\_boost\_factor} = 0.06 \times 0.4 = 0.024$$

$$\text{high\_certainty\_boost\_decrease} = \text{base}_{\text{decrease}} \times \text{high\_certainty\_boost\_factor} = 0.12 \times 0.4 = 0.048$$

**Επίπεδο Βεβαιότητας (certaintyScore)** Το επίπεδο βεβαιότητας υπολογίζεται με βάση την `certainty_level_function` ως:

$$\text{certaintyScore} = \frac{|\text{απόσταση απόφασης}|}{\text{μέγιστη απόσταση απόφασης}}$$

### Ενίσχυση Διαδοχικών Αποφάσεων

Ο όρος της ενίσχυσης διαδοχικών αποφάσεων προστίθεται ως ένα μέσο σταθερότητας στη συνάρτηση, ώστε να αποτρέπεται η υπερβολική ευαισθησία του συστήματος σε μεμονωμένες ανωμαλίες ή θορύβους. Με αυτόν τον τρόπο ενισχύεται η εμπιστοσύνη στον χρήστη όσο περισσότερο ο ίδιος χρησιμοποιεί το σύστημα, ενώ μειώνεται η εμπιστοσύνη όσο συχνότερα το χρησιμοποιεί κάποιος εισβολέας. Βελτιώνεται, συνεπώς, και η συνολικότερη ακρίβεια του συστήματος.

- Για τις γνήσιες αποφάσεις, η ενίσχυση λόγω διαδοχικών γνήσιων αποφάσεων εφαρμόζεται όταν ο αριθμός των διαδοχικών γνήσιων αποφάσεων ( $\text{consec}_{\text{genuine}}$ ) είναι πολλαπλάσιο του 3:

$$\text{consecutive\_genuine\_boost} = \begin{cases} 0.04, & \text{if } \text{consec}_{\text{genuine}} \bmod 3 = 0 \\ 0, & \text{otherwise} \end{cases}$$

- Για τις αποφάσεις απατεώνα, η ποινή λόγω διαδοχικών αποφάσεων απατεώνα εφαρμόζεται όταν ο αριθμός των διαδοχικών αποφάσεων απατεώνα ( $\text{consec}_{\text{impostor}}$ ) είναι πολλαπλάσιο του 2:

$$\text{consecutive\_impostor\_penalty} = \begin{cases} 0.05, & \text{if } \text{consec}_{\text{impostor}} \bmod 2 = 0 \\ 0, & \text{otherwise} \end{cases}$$

### Μηχανισμός Κλειδώματος

Ο μηχανισμός κλειδώματος ενεργοποιείται όταν  $C < \text{confidence\_threshold}$ . Σε αυτή την περίπτωση:

- Το σύστημα απαιτεί επανεξουσιοδότηση μέσω πρόσθετων στοιχείων ταυτοποίησης.
- Ο δείκτης  $C$  επανέρχεται στο  $C_0$  μετά από επιτυχημένη επανεξουσιοδότηση.

### 4.5.2 Παραδείγματα Λειτουργίας

Παρακάτω παρουσιάζονται εκτενώς παραδείγματα λειτουργίας του συστήματος κλειδώματος/εμπιστοσύνης για διαφορετικά σενάρια. Κάθε σενάριο περιλαμβάνει ακολουθία αποφάσεων, το επίπεδο βεβαιότητας ( $\text{certaintyScore}$ ), την αλλαγή στο επίπεδο εμπιστοσύνης ( $\Delta C$ ), και το τελικό επίπεδο εμπιστοσύνης ( $C$ ).

**Παράδειγμα 1: Διαδοχικές Γνήσιες Αποφάσεις** Σε αυτό το παράδειγμα, ο χρήστης λαμβάνει διαδοχικές γνήσιες αποφάσεις με διαφορετικά επίπεδα βεβαιότητας ( $\text{certaintyScore}$ ). Παρατηρούμε ότι κάθε τρίτη διαδοχική γνήσια απόφαση ενισχύεται με τον παράγοντα  $\text{consecutive\_genuine\_boost}$ :

$$\Delta C = \text{base}_{\text{increase}} + \text{high\_certainty\_boost\_increase} + \text{consecutive\_genuine\_boost}.$$

Η επίδραση των τιμών φαίνεται στον πίνακα 4.4, όπου το επίπεδο εμπιστοσύνης αυξάνεται σημαντικά μετά από κάθε απόφαση.

Απόφαση	Certainty	Διαδοχικές	True Label	$\Delta C$	$C$
Γνήσιος	0.8	1	Γνήσιος	+0.06 + 0.024	0.684
Γνήσιος	0.9	2	Γνήσιος	+0.06 + 0.024	0.768
Γνήσιος	0.85	3	Γνήσιος	+0.06 + 0.024 + 0.04	0.892
Απατεώνας	0.82	1	Απατεώνας	-0.12 - 0.048	0.724
Απατεώνας	0.6	2	Γνήσιος	-0.12 - 0.05	0.554
Γνήσιος	0.55	1	Απατεώνας	+0.06	0.614
Απατεώνας	0.8	1	Απατεώνας	-0.12 - 0.048	0.446

Πίνακας 4.4: Παραδείγματα Ενημέρωσης Εμπιστοσύνης

**Παράδειγμα 2: Εναλλαγή Γνήσιων και Απατεώνων** Σε αυτή την περίπτωση, εναλλάσσονται γνήσιες και απατηλές αποφάσεις. Ο πίνακας 4.5 δείχνει τη σταδιακή μείωση του επιπέδου εμπιστοσύνης λόγω αποφάσεων απατεώνων:

Απόφαση	Certainty	Διαδοχικές	True Label	$\Delta C$	$C$
Γνήσιος	0.75	1	Γνήσιος	+0.06 + 0.024	0.684
Απατεώνας	0.65	1	Γνήσιος	-0.12	0.564
Γνήσιος	0.8	1	Γνήσιος	+0.06 + 0.024	0.648
Απατεώνας	0.55	1	Γνήσιος	-0.12	0.528
Γνήσιος	0.9	1	Γνήσιος	+0.06 + 0.024	0.612
Απατεώνας	0.78	1	Απατεώνας	-0.12 - 0.048	0.444

Πίνακας 4.5: Παραδείγματα Εναλλαγής Γνήσιων και Απατηλών Αποφάσεων

**Παράδειγμα 3: Απώλεια Εμπιστοσύνης και Επανεξουσιοδότηση** Εάν το επίπεδο εμπιστοσύνης πέσει κάτω από το κατώφλι  $\text{confidence\_threshold} = 0.3$ , ενεργοποιείται ο μηχανισμός κλειδώματος, όπως φαίνεται στον πίνακα 4.6. Επίσης, φαίνεται πώς ο μηχανισμός επαναφέρει το επίπεδο εμπιστοσύνης μετά από επιτυχημένη επανεξουσιοδότηση:

Απόφαση	Certainty	Διαδοχικές	True Label	$\Delta C$	$C$
Απατεώνας	0.6	2	Απατεώνας	-0.12 - 0.048 - 0.05	0.102
Κλείδωμα	-	-	-	-	0.6
Γνήσιος	0.85	1	Γνήσιος	+0.06 + 0.024	0.684

Πίνακας 4.6: Παραδείγματα Ενεργοποίησης Μηχανισμού Κλειδώματος

**Παράδειγμα 4: Συνεχής Ενίσχυση λόγω Υψηλής Βεβαιότητας** Σε αυτό το σενάριο, όλες οι αποφάσεις είναι γνήσιες και συνοδεύονται από υψηλή βεβαιότητα ( $\text{certaintyScore} > \text{high\_certainty\_threshold}$ ). Στον πίνακα 4.7 παρατηρούμε τη σημαντική ενίσχυση του επιπέδου εμπιστοσύνης:

Απόφαση	Certainty	Διαδοχικές	True Label	$\Delta C$	$C$
Γνήσιος	0.9	1	Γνήσιος	+0.06 + 0.024	0.684
Γνήσιος	0.95	2	Γνήσιος	+0.06 + 0.024	0.768
Γνήσιος	0.92	3	Γνήσιος	+0.06 + 0.024 + 0.04	0.892
Γνήσιος	0.94	1	Γνήσιος	+0.06 + 0.024	0.976
Γνήσιος	0.97	2	Γνήσιος	+0.06 + 0.024	1.060

Πίνακας 4.7: Παραδείγματα Συνεχούς Ενίσχυσης Λόγω Υψηλής Βεβαιότητας

**Παράδειγμα 5: Επανάληψη Κλειδώματος Λόγω Απατηλών Αποφάσεων** Σε αυτό το σενάριο, ο χρήστης λαμβάνει συστηματικά αποφάσεις απατεώνων, προκαλώντας επαναλαμβανόμενη ενεργοποίηση του μηχανισμού κλειδώματος, όπως φαίνεται στον πίνακα 4.8:

Απόφαση	Certainty	Διαδοχικές	True Label	$\Delta C$	$C$
Απατεώνας	0.86	1	Απατεώνας	$-0.12 - 0.048$	0.432
Απατεώνας	0.92	2	Απατεώνας	$-0.12 - 0.048 - 0.05$	0.214
Κλείδωμα	-	-	-	-	0.6
Απατεώνας	0.91	1	Απατεώνας	$-0.12 - 0.048$	0.432
Απατεώνας	0.83	2	Απατεώνας	$-0.12 - 0.048 - 0.05$	0.214
Κλείδωμα	-	-	-	-	0.6

Πίνακας 4.8: Παραδείγματα Επαναλαμβανόμενου Κλειδώματος

Το παράδειγμα δείχνει ότι η συνεχής απώλεια εμπιστοσύνης λόγω αποφάσεων απατεώνα οδηγεί σε συχνή ενεργοποίηση του μηχανισμού κλειδώματος. Αυτό εξασφαλίζει την προστασία του συστήματος από κακόβουλη χρήση.

### 4.5.3 Παρατηρήσεις

Τα παραπάνω παραδείγματα αναδεικνύουν τη λειτουργία του συστήματος εμπιστοσύνης σε διαφορετικά σενάρια χρήσης. Οι μαθηματικοί υπολογισμοί και οι διαδοχικές αποφάσεις παρουσιάζουν τη δυναμική φύση του μηχανισμού εμπιστοσύνης, ο οποίος μπορεί να προσαρμοστεί σε διαφορετικά μοτίβα συμπεριφοράς χρηστών. Ο συνδυασμός υψηλής βεβαιότητας, διαδοχικών αποφάσεων και μηχανισμού κλειδώματος διασφαλίζει την ασφάλεια και την αξιοπιστία του συστήματος.

Το σύστημα κλειδώματος/εμπιστοσύνης παρέχει ένα δυναμικό μέσο διαχείρισης αποφάσεων, ενισχύοντας την ασφάλεια και την αξιοπιστία. Η μαθηματική του θεμελίωση το καθιστά ικανό να προσαρμόζεται σε διαφορετικά σενάρια χρήσης.

## 4.6 ΠΑΡΟΥΣΙΑΣΗ ΔΙΕΠΑΦΗΣ ΧΡΗΣΤΗ

Το κεφάλαιο αυτό παρουσιάζει το γραφικό περιβάλλον χρήστη που αναπτύχθηκε μέσω της βιβλιοθήκης Streamlit<sup>11</sup>, το οποίο σχεδιάστηκε για να υλοποιεί τη διαδικασία συνεχούς και έμμεσης αυθεντικοποίησης. Το περιβάλλον αυτό επιτρέπει την αλληλεπίδραση του χρήστη με το σύστημα μέσω απλών και κατανοητών λειτουργιών.

Η ενσωμάτωση του Streamlit UI επιτρέπει την άμεση αλληλεπίδραση των χρηστών με το σύστημα. Οι βασικές λειτουργίες περιλαμβάνουν:

- **Εισαγωγή Username:** Ο χρήστης εισάγει ένα username μέσω του UI.
- **Εισαγωγή Prompt:** Ο χρήστης εισάγει ένα prompt μέσω του UI.
- **Κουμπί Αυθεντικοποίησης:** Το σύστημα εμφανίζει τα αποτελέσματα της απόφασης (γνήσιος χρήστης ή απατεώνας) και το επίπεδο εμπιστοσύνης.

<sup>11</sup><https://streamlit.io/>



Σχήμα 4.5: Περιβάλλον Χρήστη του Streamlit UI

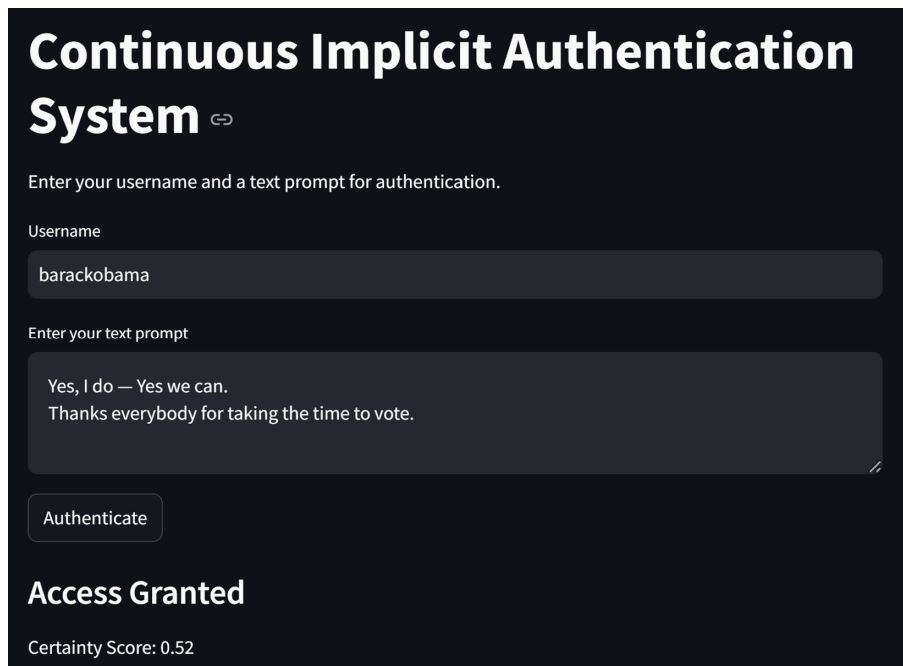
Η διαδικασία αυθεντικοποίησης βασίζεται στην πρόβλεψη που πραγματοποιεί το μοντέλο, ενώ η εφαρμογή επιστρέφει είτε "Access Granted" είτε "Access Denied" συνοδευόμενη από τη βαθμολογία βεβαιότητας (*certainty score*) της απόφασης. Ο πλήρης κώδικας βρίσκεται στο *Github*<sup>12</sup>.

### 4.6.1 Παραδείγματα Χρήσης

Παρακάτω παρουσιάζονται παραδείγματα από τη λειτουργία του Streamlit UI.

---

<sup>12</sup><https://github.com/conmylo/master-thesis/tree/main/final>



**Continuous Implicit Authentication System** ⇄

Enter your username and a text prompt for authentication.

Username

barackobama

Enter your text prompt

Yes, I do — Yes we can.  
Thanks everybody for taking the time to vote.

Authenticate

**Access Granted**

Certainty Score: 0.52

(α') Παράδειγμα έγκρισης αυθεντικοποίησης με υψηλό certainty score.



**Continuous Implicit Authentication System**

Enter your username and a text prompt for authentication.

Username

barackobama

Enter your text prompt

This is a test prompt for access denial.

Authenticate

**Access Denied**

Certainty Score: 0.35

(β') Παράδειγμα απόρριψης αυθεντικοποίησης με χαμηλό certainty score.

Σχήμα 4.6: Στιγμιότυπα οθόνης από το streamlit UI για επιβεβαίωση και απόρριψη αυθεντικοποίησης

Στην εικόνα 4.6α', ο χρήστης "barackobama" εισάγει ένα prompt που το σύστημα αναγνωρίζει ως έγκυρο, παραχωρώντας την πρόσβαση με certainty score 0.52.

Στην εικόνα 4.6β', ο ίδιος χρήστης εισάγει ένα prompt που το σύστημα αξιολογεί ως μη έγκυρο, απορρίπτοντας την πρόσβαση με certainty score 0.35.





# 5

## Πειράματα - Αποτελέσματα

Η παρούσα ενότητα επικεντρώνεται στην παρουσίαση των αποτελεσμάτων που προέκυψαν από τα πειράματα και στην αξιολόγηση του συστήματος συνολικά. Σκοπός της ενότητας είναι να προσφέρει μια σαφή εικόνα της απόδοσης του συστήματος υπό διαφορετικές συνθήκες και παραμέτρους, αλλά και να συγκρίνει τις διαφορετικές προσεγγίσεις που χρησιμοποιήθηκαν.

Αρχικά, περιγράφεται η διαδικασία αξιολόγησης και αναλύονται οι μετρικές που χρησιμοποιούνται για τη μέτρηση της απόδοσης του συστήματος, όπως το *False Rejection Rate (FRR)*, το *False Acceptance Rate (FAR)*, και οι μέσοι αριθμοί προτροπών για γνήσιους χρήστες και απατεώνες. Η επιλογή των συγκεκριμένων μετρικών βασίζεται στη δυνατότητά τους να αποτυπώνουν με ακρίβεια την απόδοση του συστήματος σε διαφορετικά σενάρια.

Η ανάλυση των αποτελεσμάτων δομείται σε διαφορετικές φάσεις, κάθε μία από τις οποίες εστιάζει σε συγκεκριμένες πτυχές της απόδοσης του συστήματος.

Στην πρώτη φάση, αξιολογείται η απόδοση του συστήματος όταν χρησιμοποιείται ένα μοντέλο OC-SVM ανά χρήστη. Παρουσιάζονται τα αποτελέσματα για διάφορους συνδυασμούς των υπερπαραμέτρων  $nu$ ,  $\gamma$ , και *threshold*, αποτυπώνοντας την ευαισθησία του μοντέλου στις παραμέτρους αυτές.

Η δεύτερη φάση αφορά την τεχνική *basic majority voting*, όπου συνδυάζονται πολλαπλά μοντέλα OC-SVM ανά χρήστη. Παρουσιάζονται οι βελτιώσεις στην απόδοση και αναλύεται η συνεισφορά της πλειοψηφικού μοντέλου.

Η τρίτη φάση ενσωματώνει τη μέθοδο *weighted majority voting* και τη συνάρτηση επιπέδου βεβαιότητας (*certainty level*), αναδεικνύοντας πώς ο συνδυασμός των τεχνικών αυτών βελτιώνει τη συνολική απόδοση του συστήματος. Γίνεται χρήση της σταθμισμένης πλειοψηφίας, ώστε κάθε μοντέλο να συνεισφέρει στο τελικό αποτέλεσμα με βάση τη βεβαιότητά του για την απόφαση.

Στην τέταρτη φάση, παρουσιάζεται η εισαγωγή της συνάρτησης επιπέδου εμπιστοσύνης (*confidence level function*). Αναλύεται πώς συμβάλλουν στην σταθεροποίηση του συστήματος και την μεγαλύτερη ακρίβειά του τόσο η ενίσχυση των διαδοχικών

αποφάσεων, όσο και ο διαχωρισμός μεταξύ των περισσότερων βέβαιων αποφάσεων από τις υπόλοιπες.

Στην πέμπτη φάση, αξιολογείται η γενίκευση του συστήματος μέσω της τεχνικής Leave-One-Subject-Out Cross Validation. Τα αποτελέσματα αποτυπώνουν την απόδοση του συστήματος όταν εκπαιδεύεται σε δεδομένα άλλων χρηστών και δοκιμάζεται στα δεδομένα του αποκλεισμένου χρήστη.

Τέλος, πραγματοποιείται σύγκριση των αποτελεσμάτων από όλες τις φάσεις, αναδεικνύοντας τις βελτιώσεις που προκύπτουν από κάθε μέθοδο. Η συνολική εικόνα που παρέχεται επιτρέπει την αποτίμηση της απόδοσης του συστήματος και τη διερεύνηση δυνατοτήτων περαιτέρω βελτίωσης.

Για λόγους πληρότητας, ακολουθεί η περιγραφή του συστήματος που χρησιμοποιήθηκε στην εκτέλεση των πειραμάτων:

Μονάδα	# πυρήνων	RAM
i7-1260p	16	32

## 5.1 ΑΞΙΟΛΟΓΗΣΗ ΣΥΣΤΗΜΑΤΟΣ

### 5.1.1 Εισαγωγή

Το σύστημα δοκιμών έχει σχεδιαστεί για να αξιολογήσει την απόδοση και την αξιοπιστία του συστήματος αυθεντικοποίησης σε διαφορετικά σενάρια. Εστιάζουμε στη συμπεριφορά του συστήματος όταν διαχειρίζεται γνήσιους χρήστες και απατεώνες, αναλύοντας την ακρίβεια και την αποτελεσματικότητά του.

### 5.1.2 Δεδομένα Δοκιμών

Τα δεδομένα που χρησιμοποιήθηκαν στις δοκιμές περιλαμβάνουν:

- **Γνήσια δεδομένα χρηστών:** Κείμενα από γνήσιους χρήστες που εκπαιδεύτηκαν στο σύστημα.
- **Δεδομένα απατεώνων:** Κείμενα από άλλους χρήστες του dataset εκτός του προφίλ του εκάστοτε γνήσιου χρήστη.

### 5.1.3 Μετρικές Αξιολόγησης

Για την αξιολόγηση του συστήματος χρησιμοποιούνται οι παρακάτω μετρικές:

- **F1 Score):**

Η μετρική F1 Score είναι μια σταθμισμένη μέση τιμή της ανάκλησης (**Recall**) και της ακρίβειας (**Precision**), η οποία χρησιμοποιείται για την αξιολόγηση της απόδοσης ενός συστήματος. Στο πλαίσιο του συστήματος αυθεντικοποίησης, η F1 υπολογίζεται ως εξής:

$$F1 = 2 \cdot \frac{\text{Precision} \cdot \text{Recall}}{\text{Precision} + \text{Recall}}$$

Όπου:

- **Precision (Ακρίβεια):**

$$\text{Precision} = \frac{\text{True Positives (TP)}}{\text{True Positives (TP)} + \text{False Positives (FP)}}$$

Η ακρίβεια αντιπροσωπεύει το ποσοστό των προτροπών που ταξινομήθηκαν ως γνήσιες (*genuine*) και ήταν πράγματι γνήσιες. Υψηλή ακρίβεια σημαίνει ότι το σύστημα αποφεύγει λανθασμένες αποδοχές απατεώνων (*impostors*).

- **Recall (Ανάκληση):**

$$\text{Recall} = \frac{\text{True Positives (TP)}}{\text{True Positives (TP)} + \text{False Negatives (FN)}}$$

Η ανάκληση μετρά το ποσοστό των πραγματικά γνήσιων προτροπών που αναγνωρίστηκαν σωστά από το σύστημα. Υψηλή ανάκληση δείχνει ότι το σύστημα αποφεύγει να απορρίψει γνήσιες προτροπές ως απατεώνες.

- **True Positives (TP):** Προτροπές που το σύστημα ταξινόμησε σωστά ως γνήσιες.
- **False Positives (FP):** Προτροπές που το σύστημα ταξινόμησε λανθασμένα ως γνήσιες ενώ ήταν απατεώνες.
- **False Negatives (FN):** Προτροπές που το σύστημα ταξινόμησε λανθασμένα ως απατεώνες ενώ ήταν γνήσιες.
- **False Rejection Rate (FRR):** Υπολογίζεται ως:

$$\text{FRR} = \frac{\text{False Rejections}}{\text{Total Genuine Prompts}}$$

- **False Acceptance Rate (FAR):** Υπολογίζεται ως:

$$\text{FAR} = \frac{\text{False Acceptances}}{\text{Total Impostor Prompts}}$$

- **Accuracy (Ακρίβεια Συνολική):**

$$\text{Accuracy} = \frac{\text{True Positives (TP)} + \text{True Negatives (TN)}}{\text{Total Predictions}}$$

Η συνολική ακρίβεια μετρά το ποσοστό των προτροπών που ταξινομήθηκαν σωστά, είτε ως γνήσιες είτε ως απατεώνες. Υψηλή τιμή ακρίβειας δείχνει τη συνολική αποτελεσματικότητα του συστήματος.

- **MAE (Μέσο Απόλυτο Σφάλμα):**

$$\text{MAE} = \frac{\sum_{i=1}^n |\hat{y}_i - y_i|}{n}$$

Το MAE μετρά το μέσο απόλυτο σφάλμα μεταξύ των προβλεπόμενων τιμών ( $\hat{y}_i$ ) και των πραγματικών τιμών ( $y_i$ ). Μια χαμηλή τιμή MAE υποδηλώνει ότι οι προβλέψεις του συστήματος είναι κοντά στις πραγματικές τιμές, καταδεικνύοντας την ακρίβεια και την αξιοπιστία του μοντέλου.

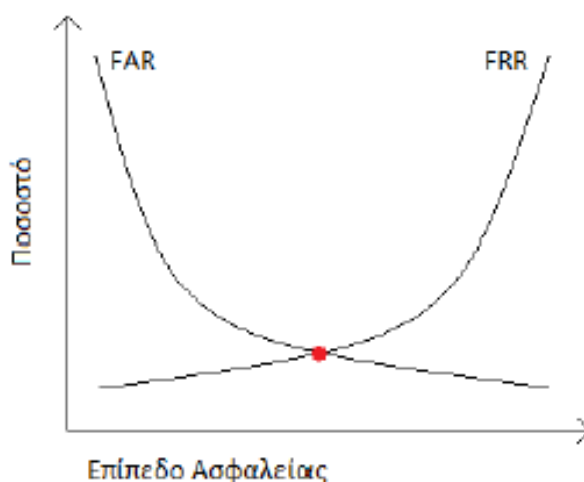
- **Μέσος Αριθμός Προτροπών για Γνήσιους Χρήστες:**

$$\text{Mean Genuine Prompts Before Lock} = \frac{\text{Σύνολο Προτροπών Γνήσιων Χρηστών}}{\text{Συνολικός Αριθμός Locks για Γνήσιους}}$$

- **Μέσος Αριθμός Προτροπών για Απατεώνες:**

$$\text{Mean Impostor Prompts Before Lock} = \frac{\text{Σύνολο Προτροπών Απατεώνων}}{\text{Συνολικός Αριθμός Locks για Απατεώνες}}$$

Οι μετρικές FAR και FRR είναι αντιστρόφως ανάλογες που σημαίνει πως όταν η μία αυξάνεται, η άλλη μειώνεται όπως φαίνεται στο παρακάτω σχήμα:



Σχήμα 5.1: Ποιοτική απεικόνιση των FAR και FRR.

Το κόκκινο σημείο ονομάζεται Equal Error Rate - EER και αναπαριστά το επίπεδο ασφαλείας για το οποίο οι τιμές των FAR και FRR είναι ίσες, όπως φαίνεται στο [σχήμα 5.1](#). Ερμηνεύοντας το σχήμα από μια πιο πρακτική προσέγγιση, γίνεται αντιληπτό πως όσο ασφαλέστερο είναι ένα σύστημα, τόσο λιγότερο βολικό θα είναι για τον χρήστη καθώς θα κλειδώνει συχνότερα και, αντίστοιχα, όσο λιγότερη ασφάλεια το χαρακτηρίζει, τόσο ευκολότερο θα είναι για τον πραγματικό χρήστη να το χειριστεί, αλλά και για τον υποκλοπέα να αυθεντικοποιηθεί από το σύστημα.

#### 5.1.4 Ροή Διαδικασίας Δοκιμών

Η διαδικασία δοκιμών ακολουθεί συγκεκριμένα βήματα, διασφαλίζοντας την αξιολόγηση κάθε μοντέλου και χρήστη ξεχωριστά. Παρακάτω περιγράφονται τα στάδια της ροής:

### Προετοιμασία δεδομένων

Η προετοιμασία περιλαμβάνει τη δημιουργία δύο συνόλων δεδομένων για κάθε χρήστη:

- **Διαχωρισμός Δεδομένων Χρήστη:** Για κάθε χρήστη, το 15% των δεδομένων του αφαιρείται από το σύνολο εκπαίδευσης και διατίθεται ως σύνολο δοκιμών (*test set*). Το υπόλοιπο 85% έχει ήδη χρησιμοποιηθεί για την εκπαίδευση των μοντέλων, όπως περιγράφεται στο Κεφάλαιο 4.3.
- **Δημιουργία Impostor Dataset:** Για κάθε χρήστη, επιλέγεται ένα σύνολο από κείμενα που ανήκουν σε άλλους χρήστες. Αυτά τα δεδομένα σχηματίζουν το *impostor dataset*, το οποίο χρησιμοποιείται για την αξιολόγηση της ικανότητας του συστήματος να εντοπίζει μη γνήσιους χρήστες. Το μέγεθος του impostor test set αυτού είναι ίσο με το μέγεθος του genuine test set του κάθε χρήστη (50-50 split).

### Χρήση Εκπαιδευμένων Μοντέλων

Τα εκπαιδευμένα μοντέλα χρησιμοποιούνται για να αξιολογήσουν τα κείμενα του *genuine test set* κάθε χρήστη και του *impostor test set*. Για κάθε κείμενο:

- Εξάγονται τα χαρακτηριστικά της κάθε εγγραφής των 2 test set (*genuine & impostor*) μέσω της συνάρτησης `extract_features`, όπως περιγράφεται στο Κεφάλαιο 4.2.
- Τα χαρακτηριστικά εισάγονται στα εκπαιδευμένα μοντέλα (*One-Class SVM*) για την παραγωγή απόφασης και του επιπέδου βεβαιότητας (*certainty level*).

### Συλλογή Αποτελεσμάτων

Η συλλογή των αποτελεσμάτων γίνεται σε δύο επίπεδα:

- **Αποτελέσματα Ανά Χρήστη:** Για κάθε χρήστη καταγράφονται:
  - Οι τιμές των μετρικών αξιολόγησης (*FRR*, *FAR*, μέσος αριθμός προτροπών πριν το κλείδωμα για γνήσιους χρήστες, μέσος αριθμός προτροπών πριν το κλείδωμα για απατεώνες).
  - Οι τιμές βεβαιότητας (*certainty scores*) για κάθε κείμενο.
- **Συνολικά Αποτελέσματα:** Τα αποτελέσματα όλων των χρηστών συνδυάζονται για την εξαγωγή συνολικών στατιστικών, παρέχοντας μια ολοκληρωμένη εικόνα της απόδοσης του συστήματος.

Η διαδικασία δοκιμών διασφαλίζει την αντικειμενική αξιολόγηση της απόδοσης του συστήματος και την καταγραφή των μετρικών σε επίπεδο χρήστη αλλά και συνολικά για το σύστημα.

## 5.2 ΠΡΩΤΗ ΦΑΣΗ ΠΕΙΡΑΜΑΤΩΝ: 1 ΜΟΝΤΕΛΟ ΑΝΑ ΧΡΗΣΤΗ

---

### 5.2.1 Πειραματική Διαδικασία

Στη πρώτη φάση πειραμάτων, χρησιμοποιούμε ένα μοντέλο OC-SVM ανά χρήστη με σαφείς καθορισμένες παραμέτρους. Η πειραματική διαδικασία μπορεί να αναλυθεί σε:

- Προετοιμασία train & test set για κάθε χρήστη.
- Εκπαίδευση μοντέλων με διαφορετικούς συνδυασμούς  $\nu$  &  $\gamma$  για όλους τους χρήστες.
- Έλεγχος test prompts για όλους τους χρήστες με διαφορετικές τιμές για το threshold.
- Μετρικές αξιολόγησης - F1 & FAR & FRR - για συνδυασμούς παραμέτρων.

### 5.2.2 Hyperparameter & Threshold Tuning

**Διαδικασία Tuning** Η διαδικασία tuning είχε στόχο τη βελτιστοποίηση των παραμέτρων  $\nu$ ,  $\gamma$  και του κατωφλίου (*threshold*) για την εξασφάλιση της μέγιστης απόδοσης του συστήματος. Αξιοποιήθηκε η τεχνική grid search, κατά την οποία δίνουμε στο σύστημα ένα πλέγμα τιμών (grid) και μας επιστρέφει αποτελέσματα για κάθε δυνατό συνδυασμό τιμών του πλέγματος. Οι παράμετροι δοκιμάστηκαν στις εξής τιμές:

- $\nu$ : {0.0001, 0.001, 0.005, 0.01, 0.05, 0.1}
- $\gamma$ : {0.05, 0.1, 0.2, 0.5, 1.0}
- Κατώφλι: {-0.01, 0.0, 0.01, 0.02, 0.03, 0.04, 0.05, 0.06, 0.07, 0.08}

Οι μετρικές που αξιολογήθηκαν, όπως αναφέρεται και στο [υποκεφάλαιο 5.1](#), περιλαμβάνουν:

- **F1 Score:** Ο σταθμισμένος μέσος όρος της ακρίβειας και της ανάκλησης.
- **False Acceptance Rate (FAR):** Ποσοστό προτροπών απατεώνων που έγιναν δεκτές.
- **False Rejection Rate (FRR):** Ποσοστό γνήσιων προτροπών που απορρίφθηκαν.

**Αποτελέσματα Tuning** Τα αποτελέσματα της αναζήτησης πλέγματος παρουσιάζονται συνοπτικά στον Πίνακα [5.1](#). Οι καλύτερες τιμές των μετρικών για κάθε συνδυασμό παραμέτρων εμφανίζονται με έμφαση.

$\nu$	$\gamma$	Threshold	F1 Score	FAR (%)	FRR (%)
0.01	0.5	0.05	<b>0.49707</b>	<b>44.18</b>	<b>49.58</b>
0.005	0.2	0.07	<b>0.51769</b>	<b>43.85</b>	<b>46.72</b>
0.01	0.1	0.05	0.67351	86.02	5.73
0.005	0.1	0.07	0.67279	85.75	5.97
0.001	0.2	0.0	0.68526	86.38	3.18
0.01	0.2	0.03	0.67170	81.88	8.37
0.01	0.5	0.07	<b>0.41699</b>	<b>35.20</b>	<b>39.90</b>
0.005	0.1	0.03	0.47475	41.65	46.38
0.01	0.2	0.05	<b>0.66543</b>	<b>39.85</b>	<b>38.92</b>
0.01	0.5	0.06	0.49714	41.58	45.62

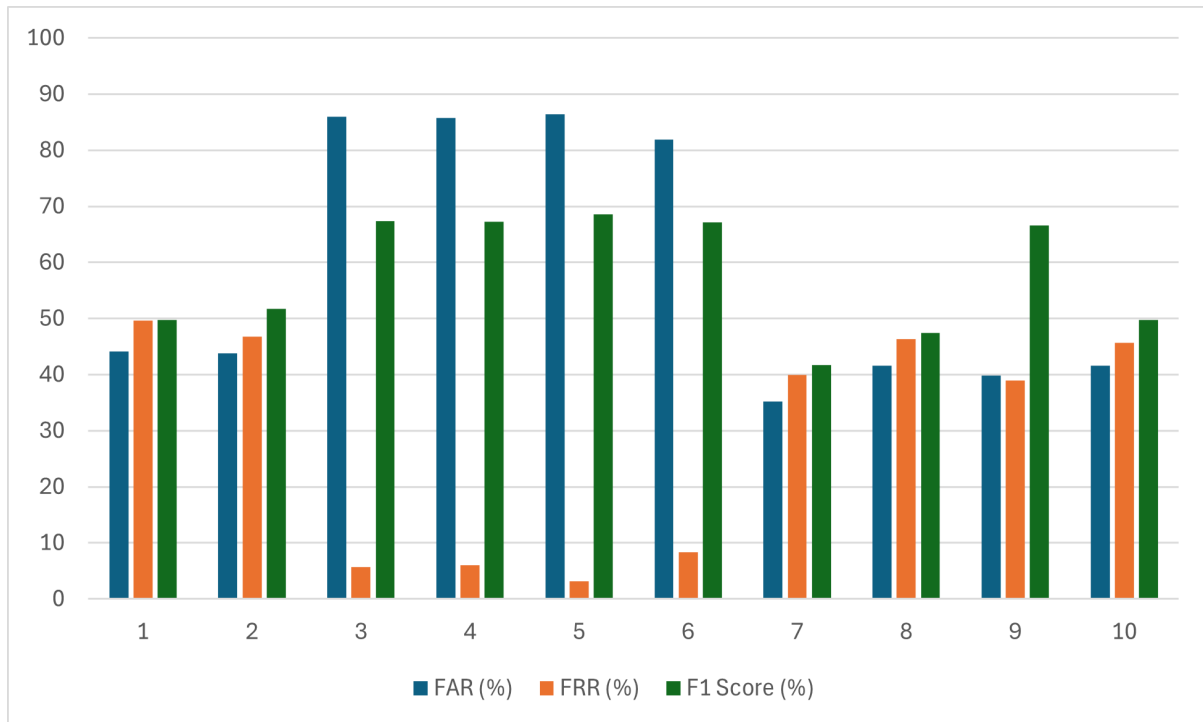
Πίνακας 5.1: Επιλεγμένα Αποτελέσματα Tuning Παραμέτρων με Ισορροπία FAR και FRR.

#### Παραδείγματα Ισορροπίας FAR και FRR

- ( **$\nu = 0.01$ ,  $\gamma = 0.5$ , threshold = 0.05**): Παρουσιάζεται ισχυρή ισορροπία μεταξύ FAR (44.18%) και FRR (49.58%), αν και το F1 είναι σχετικά χαμηλό. Υποδεικνύεται ότι το σύστημα διαχειρίζεται ομοιόμορφα τους genuine και impostor χρήστες, ωστόσο τα ποσοστά των μετρικών δεν είναι ικανοποιητικά χαμηλά.
- ( **$\nu = 0.01$ ,  $\gamma = 0.5$ , threshold = 0.07**): Τα FAR και FRR είναι πιο χαμηλά (35.20%, 39.90% αντίστοιχα), υποδεικνύοντας καλύτερη διαχείριση τόσο των γνήσιων όσο και των απατεώνων χρηστών.
- ( **$\nu = 0.01$ ,  $\gamma = 0.2$ , threshold = 0.05**): Το παράδειγμα αυτό παρουσιάζει εξαιρετικά ισορροπημένες τιμές, με FAR (39.85%) και FRR (38.92%).

#### Παραδείγματα Ακραίων Τιμών

- ( **$\nu = 0.001$ ,  $\gamma = 0.2$ , threshold = 0.0**): Παρόλο που το F1 είναι σχετικά υψηλό (0.68526), το FAR (86.38%) είναι πολύ μεγαλύτερο από το FRR (3.18%). Αυτό δείχνει ότι το σύστημα είναι πολύ επιεικές με genuine prompts, αλλά δυσκολεύεται να απορρίψει impostors.
- ( **$\nu = 0.001$ ,  $\gamma = 0.05$ , threshold = 0.01**): Σε αυτή τη περίπτωση το FAR είναι εξαιρετικά χαμηλό, υπονοώντας πως το σύστημα καταφέρνει με επιτυχία να αναγνωρίσει impostors, αλλά η τιμή του FRR (97.62%) υποδηλώνει πως το σύστημα σπάνια αναγνωρίζει τα genuine prompts. Το F1 παραμένει υψηλό.



Σχήμα 5.2: FAR, FRR, F1 για διαφορετικά πλέγματα  $\nu$ ,  $\gamma$  & threshold

Στο [σχήμα 5.2](#) βλέπουμε τις συνολικές τιμές των μετρικών για διαφορετικά πλέγματα παραμέτρων.

### 5.2.3 Αποτελέσματα Ανά Χρήστη με συγκεκριμένο πλέγμα

Ο παρακάτω πίνακας παρουσιάζει τις επιδόσεις του συστήματος για κάθε χρήστη με ρύθμιση κατωφλίου 0.05, τη τιμή του  $\nu$  0.001 και του  $\gamma$  0.05. Οι μετρικές περιλαμβάνουν την ακρίβεια (*Precision*), την ανάκληση (*Recall*), το F1 Score, το *False Acceptance Rate* (FAR), και το *False Rejection Rate* (FRR). Επιπλέον, εμφανίζονται τα συνολικά FAR και FRR.

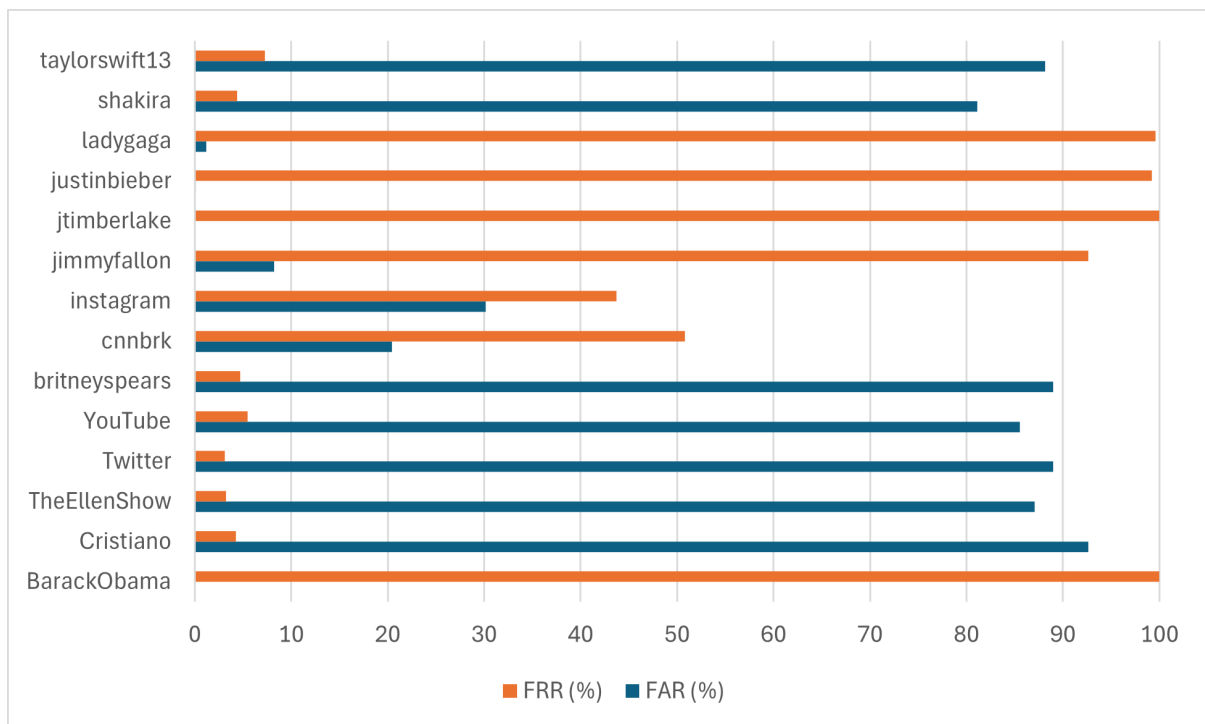


## 5.2. ΠΡΩΤΗ ΦΑΣΗ ΠΕΙΡΑΜΑΤΩΝ: 1 ΜΟΝΤΕΛΟ ΑΝΑ ΧΡΗΣΤΗ

User	Precision	Recall	F1 Score	FAR (%)	FRR (%)
BarackObama	1.00000	0.00000	0.00000	0.00	100.00
Cristiano	0.50828	0.95722	0.66399	92.60	4.28
TheEllenShow	0.52645	0.96759	0.68189	87.04	3.24
Twitter	0.52127	0.96898	0.67787	88.99	3.10
YouTube	0.52500	0.94508	0.67502	85.50	5.49
britneyspears	0.51703	0.95274	0.67030	88.99	4.73
cnnbrk	0.70671	0.49186	0.58003	20.41	50.81
instagram	0.65110	0.56267	0.60366	30.15	43.73
jimmyfallon	0.47433	0.07401	0.12805	8.20	92.60
jtimberlake	1.00000	0.00000	0.00000	0.00	100.00
justinbieber	0.88235	0.00763	0.01514	0.10	99.24
ladygaga	0.27027	0.00435	0.00856	1.17	99.57
shakira	0.54099	0.95644	0.69109	81.15	4.36
taylorswift13	0.51259	0.92723	0.66020	88.17	7.28
<b>Overall</b>	-	-	-	<b>61.68</b>	<b>55.82</b>

Πίνακας 5.2: Αποτελέσματα Ανά Χρήστη με Threshold 0.05.

Όπως φαίνεται στον 5.2, το σύστημα παρουσιάζει σημαντικές διαφοροποιήσεις στις επιδόσεις του μεταξύ των χρηστών. Οι διακυμάνσεις στις τιμές FAR και FRR είναι σημαντικές, υποδηλώνοντας την επίδραση της κατανομής δεδομένων και της ποικιλομορφίας στα χαρακτηριστικά γραφής του κάθε χρήστη. Στο σχήμα 5.3 παρουσιάζονται τα παραπάνω αποτελέσματα.



Σχήμα 5.3: FAR & FRR ανά χρήστη για nu: 0.01, gamma: 0.05 και threshold: 0.05

### 5.2.4 Παρατηρήσεις

- Τα καλύτερα αποτελέσματα για ισορροπία παρατηρούνται όταν  $\nu$  είναι σχετικά μικρό (0.005–0.01),  $\gamma$  είναι χαμηλό (0.2–0.5), και το threshold είναι προσεκτικά ρυθμισμένο κοντά στο 0.05–0.07.
- Τα παραδείγματα του [πίνακα 5.1](#) με **bold** υποδεικνύουν ιδανικές ισορροπίες για εφαρμογές όπου το FAR και το FRR πρέπει να είναι ισοδύναμα.

## 5.3 ΔΕΥΤΕΡΗ ΦΑΣΗ ΠΕΙΡΑΜΑΤΩΝ: BASIC MAJORITY VOTING

---

### 5.3.1 Πειραματική Διαδικασία

Η πειραματική διαδικασία της δεύτερης φάσης επικεντρώνεται στη χρήση πολλαπλών μοντέλων για κάθε χρήστη και την εφαρμογή του απλού πλειοψηφικού μοντέλου (basic majority voting) όπως έχει παρουσιαστεί στο [υποκεφάλαιο 4.4](#). Παρακάτω περιγράφονται αναλυτικά τα βήματα της διαδικασίας.

#### Διαδικασία Ελέγχου

Για κάθε χρήστη του συστήματος, δημιουργούνται και εκπαιδεύονται πολλαπλά μοντέλα *One-Class SVM* με διαφορετικούς συνδυασμούς υπερπαραμέτρων. Οι παράμετροι που χρησιμοποιούνται είναι:

- **Nu** ( $\nu$ ): Η υπερπάρμετρος  $\nu$  ελέγχει το ποσοστό των υποδειγμάτων που θεωρούνται outliers. Οι τιμές που εξετάζονται είναι:

$$\nu \in \{0.001, 0.005, 0.01\}$$

- **Γάμμα** ( $\gamma$ ): Η παράμετρος  $\gamma$  επηρεάζει τη μορφή της συνάρτησης πυρήνα (*kernel function*). Οι τιμές που δοκιμάζονται είναι:

$$\gamma \in \{0.01, 0.05, 0.1\}$$

- **Κατώφλι Αποδοχής** (*Acceptance Threshold*): Σύμφωνα με το [υποκεφάλαιο 5.2](#) καθορίζεται σε:

$$\text{Threshold} = 0.05$$

#### Καταγραφή και Ανάλυση Αποτελεσμάτων

Για κάθε χρήστη, καταγράφονται:

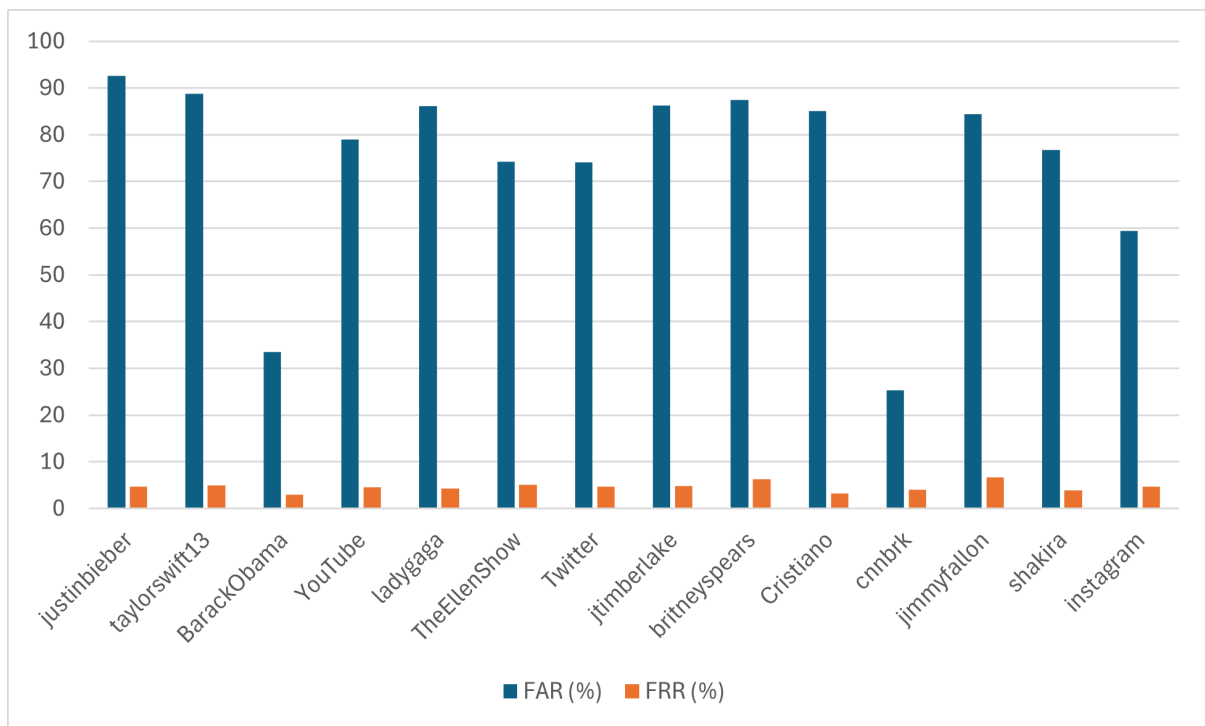
- Οι τιμές των μετρικών *False Rejection Rate (FRR)* και *False Acceptance Rate (FAR)*.
- Ο αριθμός από γνήσια και μη γνήσια prompts στον οποίο εφαρμόζεται ο έλεγχος.

### 5.3.2 Αποτελέσματα Ανά Χρήστη από τη Συνάρτηση Basic Majority Voting

Στον παρακάτω πίνακα παρουσιάζονται τα αποτελέσματα για διαφορετικούς χρήστες, όπως καταγράφηκαν κατά τη διάρκεια των πειραμάτων:

User	FAR (%)	FRR (%)	Total Genuine Tests	Total Impostor Tests
justinbieber	92.54	4.75	295	295
taylorswift13	88.78	4.95	303	303
BarackObama	33.49	3.02	430	430
YouTube	79.00	4.55	462	462
ladygaga	86.09	4.35	345	345
TheEllenShow	74.21	5.07	473	473
Twitter	74.13	4.65	344	344
jtimberlake	86.29	4.84	372	372
britneyspears	87.50	6.25	416	416
Cristiano	85.11	3.19	376	376
cnnbrk	25.27	3.97	277	277
jimmyfallon	84.43	6.61	469	469
shakira	76.78	3.96	379	379
instagram	59.43	4.65	387	387

Πίνακας 5.3: Αποτελέσματα από τη χρήση της Basic Majority Voting για διαφορετικούς χρήστες.



Σχήμα 5.4: FAR & FRR ανά χρήστη με τη χρήση της Basic Majority Voting

### 5.3.3 Παρατηρήσεις

Η παραπάνω ανάλυση δείχνει σημαντική ποικιλομορφία στις επιδόσεις του συστήματος, κυρίως στη μετρική FAR, ανάλογα με τον χρήστη. Συγκεκριμένα:

- Παρατηρείται μεγάλη ανισορροπία μεταξύ των τιμών FAR (πολύ υψηλές) και των τιμών FRR (ικανοποιητικά χαμηλές). Η μεταβλητότητα στις μετρικές δείχνει την ανάγκη για πιο εξελιγμένες τεχνικές λήψης απόφασης, όπως η *Weighted Majority Voting*, που αναλύεται στα επόμενα υποκεφάλαια.
- Χρήστες όπως οι *justinbieber* και *taylorswift13* παρουσιάζουν πολύ υψηλά ποσοστά FAR, κάτι που δείχνει αυξημένη πιθανότητα αποδοχής απατεώνων.
- Οι χρήστες *cnnbrk* και *BarackObama* έχουν εξαιρετικά χαμηλά ποσοστά FAR, κάτι που υποδεικνύει ότι τα μοντέλα τους είναι πιο αυστηρά απέναντι σε απατεώνες.

Τα παραπάνω αποτελέσματα υποδηλώνουν την ανάγκη βελτιστοποίησης του συστήματος μέσω διαφορετικών τεχνικών.

## 5.4 ΤΡΙΤΗ ΦΑΣΗ ΠΕΙΡΑΜΑΤΩΝ: WEIGHTED MAJORITY VOTING

---

Η χρήση της μεθόδου *Weighted Majority Voting* σε συνδυασμό με τη συνάρτηση *Certainty Level* παρέχει μια πιο λεπτομερή και προσαρμοστική προσέγγιση στη διαδικασία λήψης αποφάσεων. Σε αυτό το υποκεφάλαιο παρουσιάζονται τα αποτελέσματα που προέκυψαν από τη χρήση των παραπάνω μεθόδων και η αξιολόγηση της απόδοσής τους σε σύγκριση με την *Basic Majority Voting*.

### 5.4.1 Πειραματική Διαδικασία

#### Διαδικασία Ελέγχου

Με παρόμοιο τρόπο με νωρίτερα, για κάθε χρήστη του συστήματος, δημιουργούνται και εκπαιδεύονται πολλαπλά μοντέλα *One-Class SVM* με διαφορετικούς συνδυασμούς υπερπαραμέτρων. Οι παράμετροι που χρησιμοποιούνται καθ' όλη τη διάρκεια της διαδικασίας είναι:

- **Nu** ( $\nu$ ): Οι τιμές που εξετάζονται είναι:

$$\nu \in \{0.001, 0.005, 0.01, 0.05, 0.13, 0.25\}$$

- **Γάμμα** ( $\gamma$ ): Οι τιμές που εξετάζονται είναι:

$$\gamma \in \{0.00005, 0.0009, 0.0005, 0.001, 0.005, 0.01, 0.05, 0.07, 0.1, 0.15, 0.2, 0.3, 0.5\}$$

### Καταγραφή και Ανάλυση Αποτελεσμάτων

Τα αποτελέσματα που προέκυψαν από τη μέθοδο *Weighted Majority Voting* με τη χρήση της συνάρτησης *Certainty Level* παρουσιάζονται στους Πίνακες 5.4, 5.5, 5.6. Οι μετρικές που αξιολογούνται περιλαμβάνουν:

- *False Acceptance Rate (FAR)*: Ποσοστό προτροπών απατεώνων που λανθασμένα χαρακτηρίστηκαν ως γνήσιες.
- *False Rejection Rate (FRR)*: Ποσοστό γνήσιων προτροπών που λανθασμένα απορρίφθηκαν.
- Μέσος αριθμός μη γνήσιων προτροπών που γίνονται αποδεκτές πριν από το κλείδωμα.

#### 5.4.2 Αποτελέσματα

User	FAR (%)	FRR (%)	Mean Accepted Prompts by Impostor
justinbieber	88.81	7.80	$\infty$
taylorswift13	86.14	6.60	$\infty$
<b>BarackObama</b>	<b>28.60</b>	<b>3.49</b>	<b>3.77</b>
YouTube	73.59	5.41	$\infty$
ladygaga	80.29	6.38	$\infty$
TheEllenShow	67.44	7.82	$\infty$
Twitter	68.60	7.56	$\infty$
jtimberlake	80.11	9.41	$\infty$
britneyspears	83.89	8.17	$\infty$
Cristiano	78.72	9.31	$\infty$
<b>cnnbrk</b>	<b>20.58</b>	<b>5.42</b>	<b>2.00</b>
jimmyfallon	78.25	9.17	$\infty$
shakira	71.24	6.86	$\infty$
<b>instagram</b>	<b>52.20</b>	<b>7.24</b>	<b>29.00</b>

Πίνακας 5.4: Αποτελέσματα για  $\nu \in \{0.001, 0.005\}$  και  $\gamma \in \{0.05, 0.1\}$  που δείχνουν τις μετρικές FAR, FRR και τον μέσο αριθμό αποδεκτών προτροπών από απατεώνες.

## ΚΕΦΑΛΑΙΟ 5. ΠΕΙΡΑΜΑΤΑ - ΑΠΟΤΕΛΕΣΜΑΤΑ

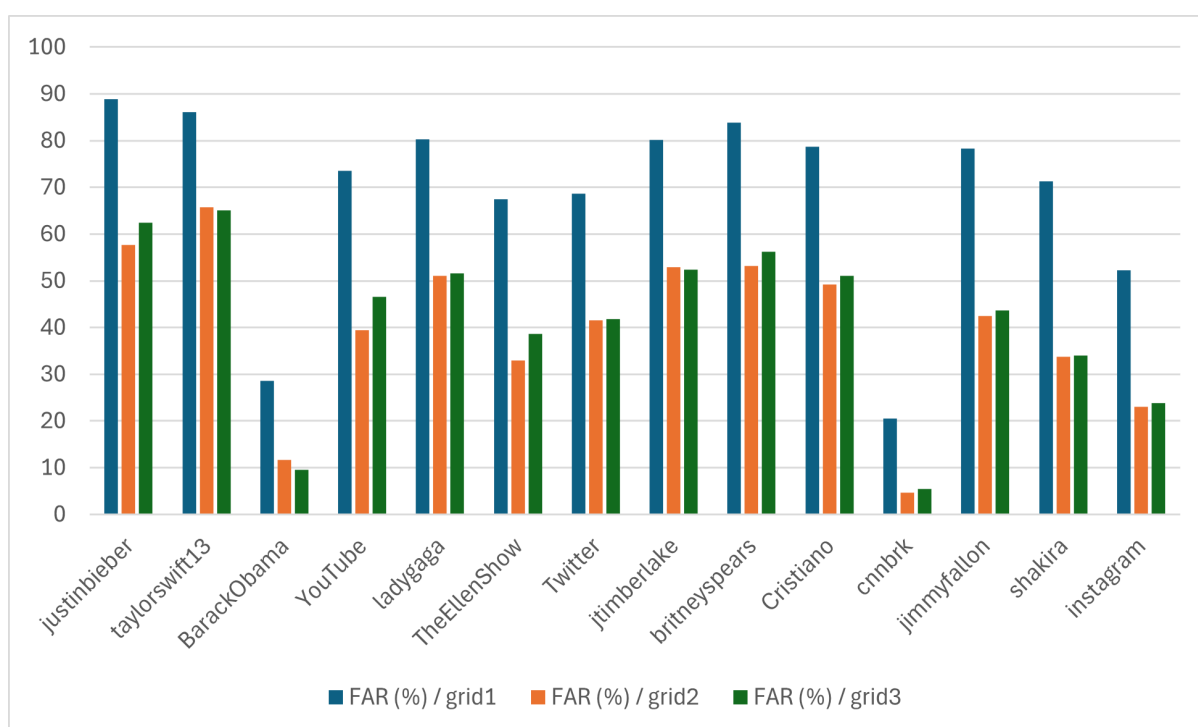
User	FAR (%)	FRR (%)	Mean Accepted Prompts by Impostor
justinbieber	57.63	19.66	$\infty$
taylorswift13	65.68	29.04	$\infty$
<b>BarackObama</b>	<b>11.63</b>	<b>27.21</b>	<b>0.87</b>
YouTube	39.39	30.95	10.31
ladygaga	51.01	30.14	$\infty$
<b>TheEllenShow</b>	<b>32.98</b>	<b>29.39</b>	<b>5.74</b>
Twitter	41.57	29.94	11.00
jtimberlake	52.96	34.68	16.50
britneyspears	53.12	28.61	$\infty$
Cristiano	49.20	30.85	6.00
<b>cnnbrk</b>	<b>4.69</b>	<b>40.43</b>	<b>0.29</b>
jimmyfallon	42.43	33.90	16.00
shakira	33.77	32.19	5.62
<b>instagram</b>	<b>23.00</b>	<b>29.46</b>	<b>2.50</b>

Πίνακας 5.5: Αποτελέσματα για  $\nu \in \{0.0001, 0.0005, 0.001\}$  και  $\gamma \in \{0.05, 0.1, 0.5\}$  που δείχνουν τις μετρικές FAR, FRR και τον μέσο αριθμό αποδεκτών προτροπών από απατεώνες.

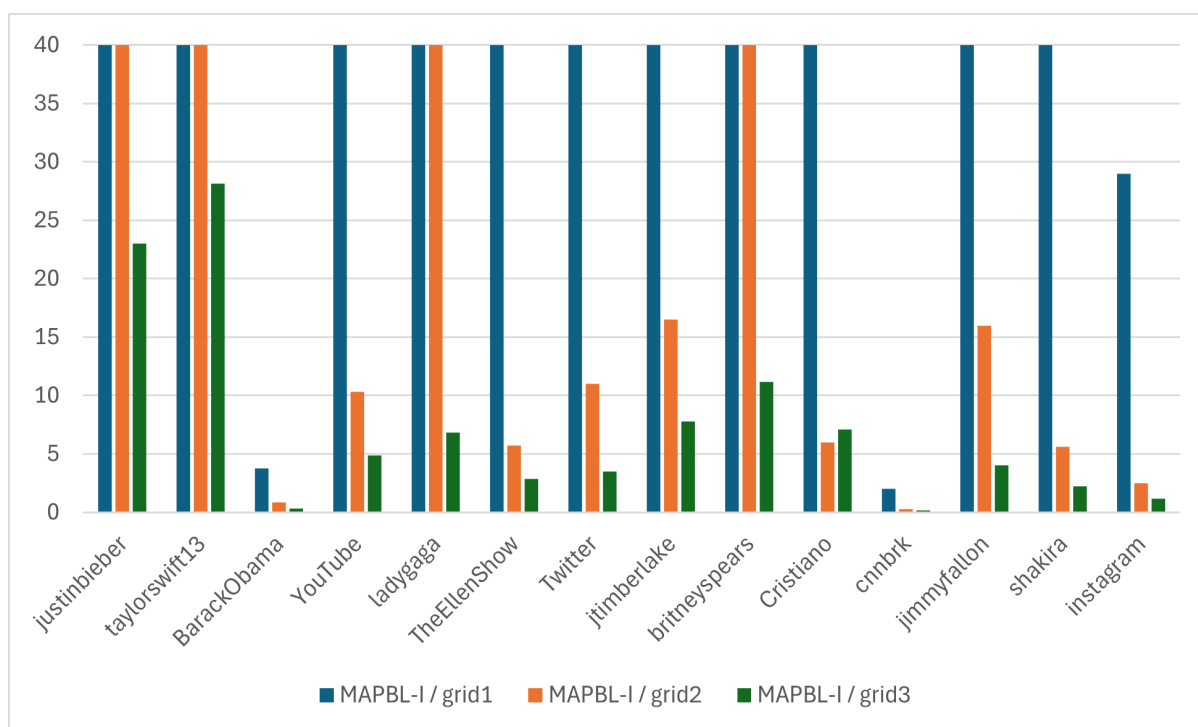
User	FAR (%)	FRR (%)	Mean Accepted Prompts by Impostor
justinbieber	62.37	21.69	23.00
taylorswift13	65.02	25.74	28.14
<b>BarackObama</b>	<b>9.53</b>	<b>29.07</b>	<b>0.34</b>
YouTube	46.54	27.92	4.89
ladygaga	51.59	27.54	6.85
TheEllenShow	38.69	26.22	2.86
Twitter	41.86	32.56	3.51
jtimberlake	52.42	33.60	7.80
britneyspears	56.25	26.92	11.14
Cristiano	51.06	29.79	7.11
<b>cnnbrk</b>	<b>5.42</b>	<b>37.91</b>	<b>0.18</b>
jimmyfallon	43.71	30.92	4.02
shakira	34.04	32.19	2.22
instagram	23.77	31.27	1.15

Πίνακας 5.6: Αποτελέσματα για  $\nu \in \{0.001, 0.005, 0.01\}$  και  $\gamma \in \{0.05, 0.07, 0.1, 0.2, 0.5\}$ : FAR, FRR και Μέσος Αριθμός Αποδεκτών Προτροπών από Απατεώνες.

#### 5.4. ΤΡΙΤΗ ΦΑΣΗ ΠΕΙΡΑΜΑΤΩΝ: WEIGHTED MAJORITY VOTING



Σχήμα 5.5: FAR ανά χρήστη για διαφορετικά πλέγματα παραμέτρων



Σχήμα 5.6: Mean Accepted Prompts Before Locking for Impostors ανά χρήστη για διαφορετικά πλέγματα παραμέτρων

### 5.4.3 Παρατηρήσεις

Η ανάλυση των αποτελεσμάτων για τη λειτουργία της μεθόδου weighted majority voting σε συνδυασμό με τη certainty level function αποδεικνύει τη σταδιακή βελτίωση της απόδοσης σε σχέση με τις προηγούμενες φάσεις. Οι τιμές των μετρικών FAR (False Acceptance Rate) και FRR (False Rejection Rate) αναδεικνύουν τη δυνατότητα του συστήματος να επιτυγχάνει καλύτερη ισορροπία μεταξύ αποδοχής γνήσιων προτροπών και απόρριψης μη γνήσιων. Μάλιστα το FAR μειώθηκε από 80.31% σε 49.11%, παρότι το FRR αυξήθηκε από 4.54% σε 24.00%.

Συγκεκριμένα, στο τρίτο πλέγμα παραμέτρων ( $\nu \in 0.001, 0.005, 0.01$  και  $\gamma \in 0.05, 0.07, 0.1, 0.2, 0.5$ ), παρατηρούνται εντυπωσιακά αποτελέσματα, όπως για τον χρήστη *cnhbrk*, όπου η τιμή FAR μειώθηκε στο εξαιρετικά χαμηλό επίπεδο του 5.42%, με μέση αποδοχή 0.18 prompts από impostors, ενώ το FRR παραμένει σχετικά υψηλό στο 37.91%. Αυτό υποδεικνύει τη δυνατότητα του συστήματος να εντοπίζει με ακρίβεια μη έγκυρες εισόδους.

Η προσαρμογή των παραμέτρων  $\nu$  και  $\gamma$  φαίνεται να επηρεάζει άμεσα την απόδοση του συστήματος. Χαμηλότερες τιμές  $\nu$  οδηγούν σε χαμηλότερο FAR, ενώ υψηλότερες τιμές  $\gamma$  συμβάλλουν στη μείωση του FRR, με προφανή αντίκτυπο στην ακρίβεια και την ευαισθησία του συστήματος. Συνολικά, τα αποτελέσματα αναδεικνύουν την πρόοδο που έχει επιτευχθεί, με την ενσωμάτωση της weighted majority voting, και την επίτευξη βέλτιστων συνδυασμών παραμέτρων που εξισορροπούν τις ανάγκες ακρίβειας και ευαισθησίας.

## 5.5 ΤΕΤΑΡΤΗ ΦΑΣΗ ΠΕΙΡΑΜΑΤΩΝ: CONFIDENCE LEVEL FUNCTION

Η τέταρτη φάση των πειραμάτων επικεντρώνεται στην ενσωμάτωση της συναρτήσης *confidence level*, η οποία αποσκοπεί στη βελτίωση της ακρίβειας και της απόδοσης του συστήματος λήψης αποφάσεων. Η προσέγγιση αυτή επεκτείνει τη μέθοδο της σταθμισμένης πλειοψηφίας, προσθέτοντας έναν επιπλέον δείκτη αξιοπιστίας στις αποφάσεις, βασιζόμενο σε ένα δυναμικά μεταβαλλόμενο επίπεδο εμπιστοσύνης του συστήματος.

### 5.5.1 Πειραματική Διαδικασία

Για την αξιολόγηση της ενσωμάτωσης της συναρτήσης *confidence level*, ακολουθήθηκε η εξής διαδικασία:

- Χρήση του πλέγματος υπερπαραμέτρων:

$$\nu \in \{0.001, 0.005, 0.01\}, \quad \gamma \in \{0.05, 0.07, 0.1, 0.2, 0.5\}$$

- Εκτέλεση της διαδικασίας ελέγχου για κάθε συνδυασμό υπερπαραμέτρων και καταγραφή των παρακάτω μετρικών:
  - **MAPBL-G**: Μέσος αριθμός αποδεκτών προτροπών από γνήσιους χρήστες πριν την ενεργοποίηση του μηχανισμού κλειδώματος.



## 5.5. ΤΕΤΑΡΤΗ ΦΑΣΗ ΠΕΙΡΑΜΑΤΩΝ: CONFIDENCE LEVEL FUNCTION

- **MAPBL-I**: Μέσος αριθμός αποδεκτών προτροπών από μη γνήσιους χρήστες (impostors) πριν την ενεργοποίηση του μηχανισμού κλειδώματος.
- Αξιολόγηση της ακρίβειας και της αξιοπιστίας του συστήματος με βάση τα παραπάνω δεδομένα.

### 5.5.2 Αποτελέσματα

Τα αποτελέσματα της πειραματικής διαδικασίας συνοψίζονται στους παρακάτω πίνακες.

Αρχικά βλέπουμε τα αποτελέσματα των περιαιμάτων για κάθε χρήστη μεμονωμένα για μια συγκεκριμένη confidence function.

User	FAR (%)	FRR (%)	MAPBL-Genuine	MAPBL-Impostor
justinbieber	62.37	21.69	64.00	23.00
taylorswift13	65.02	25.74	78.00	28.14
<b>BarackObama</b>	<b>9.53</b>	<b>29.07</b>	<b>125.00</b>	<b>0.34</b>
YouTube	46.54	27.92	64.50	4.89
ladygaga	51.59	27.54	95.00	6.85
<b>TheEllenShow</b>	<b>38.69</b>	<b>26.22</b>	<b>124.00</b>	<b>2.86</b>
Twitter	41.86	32.56	56.00	3.51
jtimberlake	52.42	33.60	20.83	7.80
britneyspears	56.25	26.92	112.00	11.14
Cristiano	51.06	29.79	56.00	7.11
<b>cnnbrk</b>	<b>5.42</b>	<b>37.91</b>	<b>13.12</b>	<b>0.18</b>
jimmyfallon	43.71	30.92	145.00	4.02
shakira	34.04	32.19	61.00	2.22
instagram	23.77	31.27	60.50	1.15

Πίνακας 5.7: Αποτελέσματα πειραμάτων για την ενσωμάτωση της συνάρτησης confidence level.

Αλλάζοντας τις τιμές των μεταβλητών της συνάρτησης confidence function, μπορούμε να προσαρμόσουμε την αυστηρότητα της συνάρτησης και τις μετρικές MAPBL-G & MAPBL-I. Ακολουθούν πίνακες των μετρικών με διαφορετικές τιμές παραμέτρων της confidence level:

- Στον [πίνακα 5.8](#), οι τιμές των παραμέτρων είναι:
  - Base Increase: 0.09
  - Base Decrease: 0.08
  - High-Certainty Boost factor: 0.4
  - Confidence Threshold: 0.3
- Στον [πίνακα 5.9](#), οι τιμές των παραμέτρων είναι:
  - Base Increase: 0.07

- Base Decrease: 0.1
- High-Certainty Boost factor: 0.5
- Confidence Threshold: 0.4
- Στον [πίνακα 5.10](#), οι τιμές των παραμέτρων είναι:
  - Base Increase: 0.07
  - Base Decrease: 0.1
  - High-Certainty Boost factor: 1.00
  - Confidence Threshold: 0.4
- Στον [πίνακα 5.11](#), οι τιμές των παραμέτρων είναι:
  - Base Increase: 0.1
  - Base Decrease: 0.06
  - High-Certainty Boost factor: 0.5
  - Confidence Threshold: 0.4
- Στον [πίνακα 5.12](#), οι τιμές των παραμέτρων είναι:
  - Base Increase: 0.12
  - Base Decrease: 0.09
  - High-Certainty Boost factor: 0.4
  - Confidence Threshold: 0.3

Μετρική	Τιμή
Total Users Tested	14
Mean FRR (%)	31.94
Mean FAR (%)	42.55
Mean Genuine Rejected Prompts Before Lock	75.41
Mean Impostor Accepted Prompts Before Lock	6.90

Πίνακας 5.8: Primary Results

Μετρική	Τιμή
Total Users Tested	14
Mean FRR (%)	39.48
Mean FAR (%)	34.40
Mean Genuine Rejected Prompts Before Lock	18.55
Mean Impostor Accepted Prompts Before Lock	3.03

Πίνακας 5.9: Tighten Variables' Values

## 5.5. ΤΕΤΑΡΤΗ ΦΑΣΗ ΠΕΙΡΑΜΑΤΩΝ: CONFIDENCE LEVEL FUNCTION

Μετρική	Τιμή
Total Users Tested	14
Mean FRR (%)	48.57
Mean FAR (%)	26.87
Mean Genuine Rejected Prompts Before Lock	7.35
Mean Impostor Accepted Prompts Before Lock	1.70

Πίνακας 5.10: Maximized boost value in confidence function

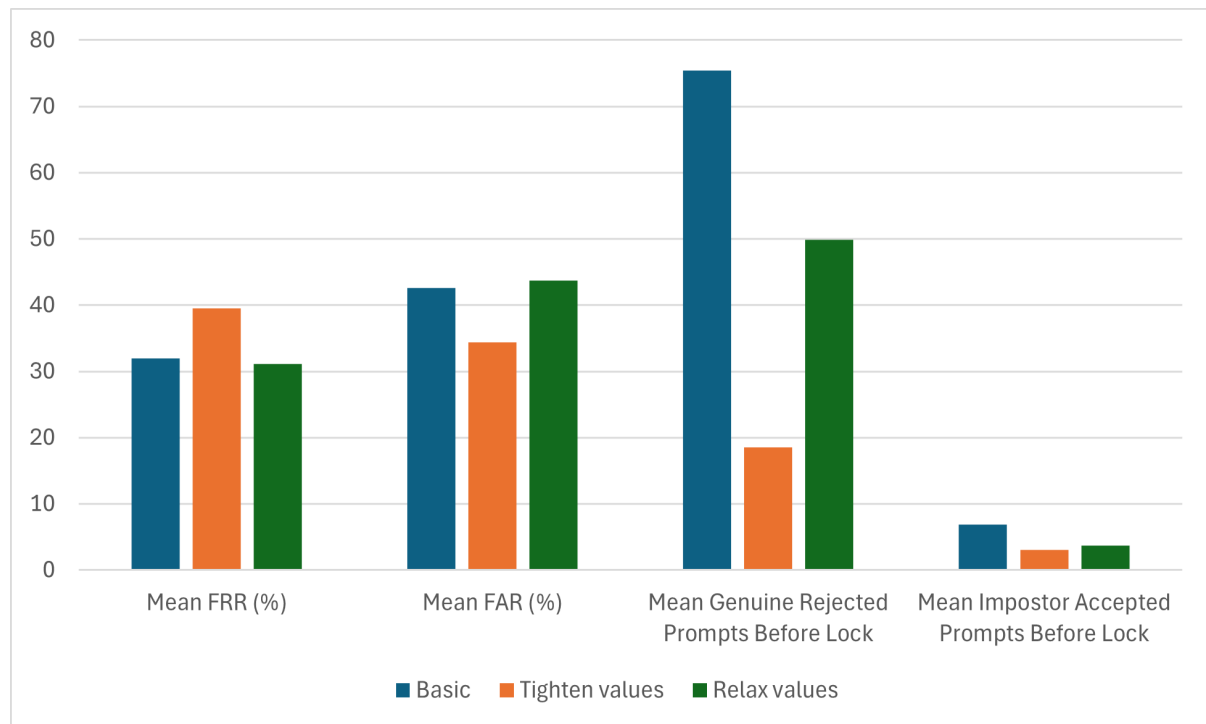
Μετρική	Τιμή
Total Users Tested	14
Mean FRR (%)	31.12
Mean FAR (%)	43.74
Mean Genuine Rejected Prompts Before Lock	49.87
Mean Impostor Accepted Prompts Before Lock	3.66

Πίνακας 5.11: Relaxation of base increase and base decrease variables' values

Μετρική	Τιμή
Total Users Tested	14
Mean FRR (%)	31.12
Mean FAR (%)	43.74
Mean Genuine Rejected Prompts Before Lock	123.25
Mean Impostor Accepted Prompts Before Lock	6.85

Πίνακας 5.12: Relaxation of all values

## ΚΕΦΑΛΑΙΟ 5. ΠΕΙΡΑΜΑΤΑ - ΑΠΟΤΕΛΕΣΜΑΤΑ



Σχήμα 5.7: FAR, FRR, MAPBL-G, MAPBL-I για διαφορετικές τιμές των μεταβλητών της συνάρτησης Confidence Level

Ενδιαφέρουσα είναι και η περίπτωση αλλαγής της υπερπαραμέτρου  $\gamma$  σε συνάρτηση με τις μετρικές που μελετάμε. Παρακάτω ακολουθούν κάποια αποτελέσματα από τέτοιες περιπτώσεις.

Μετρική	Τιμή
Total Users Tested	14
Mean FRR (%)	31.12
Mean FAR (%)	43.74
Mean Genuine Rejected Prompts Before Lock	72.38
Mean Impostor Accepted Prompts Before Lock	6.76

Πίνακας 5.13: All gamma values

Μετρική	Τιμή
Total Users Tested	14
Mean FRR (%)	37.42
Mean FAR (%)	36.83
Mean Genuine Rejected Prompts Before Lock	40.25
Mean Impostor Accepted Prompts Before Lock	2.61

Πίνακας 5.14: Include 0.3 gamma value on the grid

## 5.5. ΤΕΤΑΡΤΗ ΦΑΣΗ ΠΕΙΡΑΜΑΤΩΝ: CONFIDENCE LEVEL FUNCTION

---

Μετρική	Τιμή
Total Users Tested	14
Mean FRR (%)	35.87
Mean FAR (%)	38.40
Mean Genuine Rejected Prompts Before Lock	42.22
Mean Impostor Accepted Prompts Before Lock	3.03

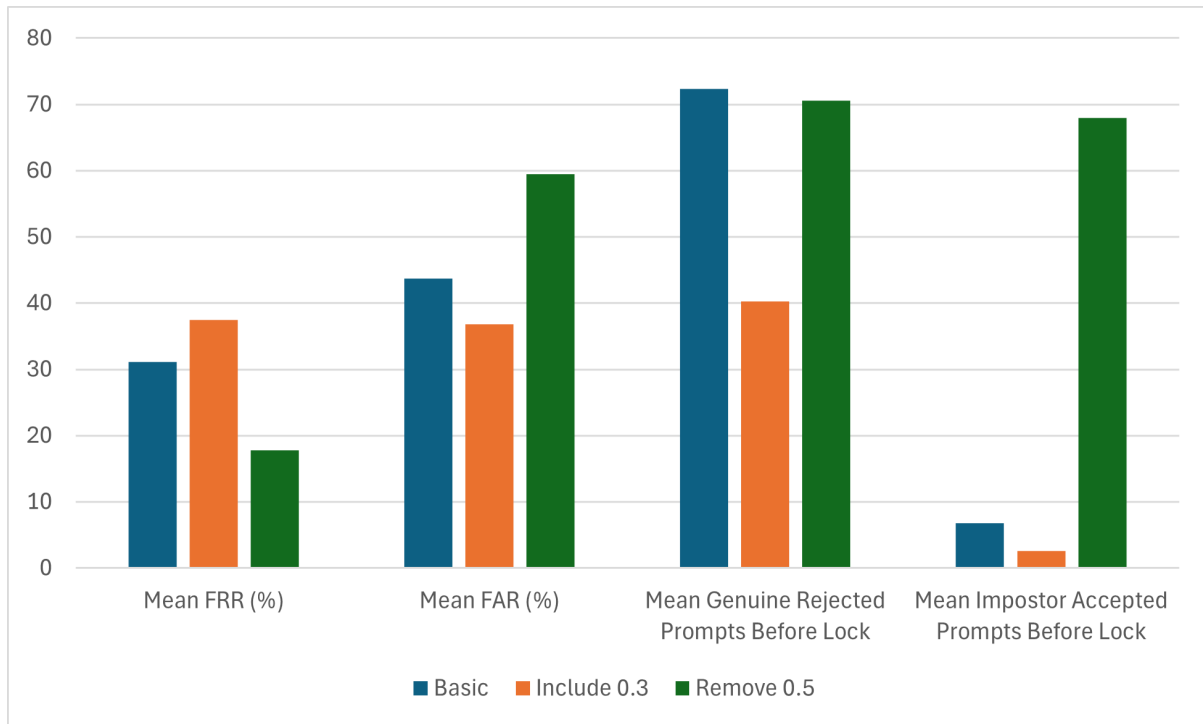
Πίνακας 5.15: Include 0.15 and 0.3 on the gamma grid

Μετρική	Τιμή
Total Users Tested	14
Mean FRR (%)	30.20
Mean FAR (%)	44.29
Mean Genuine Rejected Prompts Before Lock	86.85
Mean Impostor Accepted Prompts Before Lock	6.08

Πίνακας 5.16: Include 0.15 on the gamma grid

Μετρική	Τιμή
Total Users Tested	14
Mean FRR (%)	17.80
Mean FAR (%)	59.52
Mean Genuine Rejected Prompts Before Lock	70.62
Mean Impostor Accepted Prompts Before Lock	68.00

Πίνακας 5.17: Remove 0.5 from the gamma grid



Σχήμα 5.8: FAR, FRR, MAPBL-G, MAPBL-I για διαφορετικές τιμές της υπερπαραμέτρου  $\gamma$

### 5.5.3 Παρατηρήσεις

Η ενσωμάτωση της συναρτήσης *confidence level* προσέφερε ποσοτικά σημαντικά οφέλη στις επιδόσεις του συστήματος:

- **Μείωση FAR:** Η μέση τιμή του FAR μειώθηκε από 42.55% σε 26.87% κατά τις πειραματικές δοκιμές, αποδεικνύοντας ότι το σύστημα μπορεί να αποφεύγει με μεγαλύτερη ακρίβεια τις λανθασμένες αποδοχές μη γνήσιων χρηστών.
- **Ισορροπία μεταξύ FRR και MAPBL:** Το FRR παρέμεινε σε διαχειρίσιμα επίπεδα με μέση τιμή 31.12%, ενώ οι μέσοι αριθμοί αποδεκτών προτροπών από impostors (MAPBL-I) μειώθηκαν σημαντικά, επιτυγχάνοντας καλύτερη ισορροπία.

Η παραπάνω προσέγγιση επιβεβαιώνει τη σημασία της χρήσης μιας δυναμικής μεθόδου όπως το *confidence level*.

## 5.6 ΠΕΜΠΤΗ ΦΑΣΗ ΠΕΙΡΑΜΑΤΩΝ: LOSO CROSS VALIDATION

Η πέμπτη φάση αφορά την τεχνική Leave-One-Subject-Out Cross Validation, η οποία στοχεύει στην ανάδειξη της δυνατότητας γενίκευσης του μοντέλου.

Η τεχνική Leave-One-Subject-Out Cross-Validation (LOSO-CV) αποτελεί μια εξειδικευμένη προσέγγιση του ευρύτερου πλαισίου της διαδικασίας **Cross-Validation**.

Το Cross-Validation χρησιμοποιείται συχνά στη μηχανική μάθηση για την αξιολόγηση της δυνατότητας γενίκευσης ενός μοντέλου. Σκοπός του είναι να επιβεβαιώσει πως τα αποτελέσματα που παράγονται από ένα μοντέλο δεν είναι προσαρμοσμένα αποκλειστικά στο σύνολο των δεδομένων εκπαίδευσης αλλά μπορούν να γενικευτούν και σε άγνωστα δεδομένα.

Κατά τη διαδικασία Cross-Validation, το σύνολο δεδομένων διαχωρίζεται σε υποκατηγορίες (*folds*). Ένα από τα *folds* χρησιμοποιείται για την αξιολόγηση (*validation set*), ενώ τα υπόλοιπα χρησιμοποιούνται για την εκπαίδευση (*training set*). Αυτή η διαδικασία επαναλαμβάνεται ώσπου κάθε *fold* να έχει χρησιμοποιηθεί ως σύνολο αξιολόγησης.

Συγκεκριμένα στη τεχνική LOSO, αντί για τυχαία *folds*, κάθε επανάληψη του Cross-Validation επικεντρώνεται στον διαχωρισμό όλων των δεδομένων ενός συγκεκριμένου χρήστη από όλων των υπολοίπων, δηλαδή ολόκληρο το προφίλ ενός χρήστη αποκλείεται κατά τη διάρκεια της εκπαίδευσης και χρησιμοποιείται αποκλειστικά για την αξιολόγηση. Η προσέγγιση αυτή βρίσκει εφαρμογή και στην ανάλυση βιομετρικών δεδομένων, όπου η γενίκευση σε άγνωστα προφίλ χρηστών αποτελεί κριτήριο για την αποδοτικότητα του συστήματος.

### Περιγραφή της Διαδικασίας LOSO

Η διαδικασία LOSO λειτουργεί ως εξής:

1. Το σύνολο δεδομένων χωρίζεται με βάση τον χρήστη. Κάθε χρήστης θεωρείται μία μοναδική κατηγορία δεδομένων.
2. Σε κάθε επανάληψη, τα δεδομένα ενός χρήστη (*testing subject*) αφαιρούνται πλήρως από το σύνολο εκπαίδευσης και χρησιμοποιούνται αποκλειστικά για την αξιολόγηση.
3. Τα υπόλοιπα δεδομένα των υπόλοιπων χρηστών (*training subjects*) χρησιμοποιούνται για την εκπαίδευση των μοντέλων.
4. Μετά την εκπαίδευση, τα μοντέλα αξιολογούνται στα δεδομένα του αποκλεισμένου χρήστη.
5. Η διαδικασία επαναλαμβάνεται για κάθε χρήστη, διασφαλίζοντας ότι όλοι οι χρήστες έχουν χρησιμοποιηθεί μία φορά ως *testing subjects*.

Στο [σχήμα 5.9α'](#) φαίνεται η διαδικασία της τεχνικής LOSO για πολυδιάστατα δεδομένα (π.χ. δεδομένα από πολλαπλές πηγές ή συσκευές), ενώ στο [σχήμα 5.9β'](#) παρουσιάζεται η βασική αρχή αυτής της τεχνικής, όπου κάθε χρήστης αποκλείεται διαδοχικά για να χρησιμοποιηθούν τα δεδομένα του ως (*testing subject*).



(α') Επεξήγηση της διαδικασίας LOSO για δεδομένα από διαφορετικές πηγές. Η κάθε πηγή δεδομένων αποκλείεται διαδοχικά και τα υπόλοιπα δεδομένα χρησιμοποιούνται για εκπαίδευση.

(β') Παραδείγματα του τρόπου λειτουργίας της τεχνικής LOSO. Κάθε χρήστης αποκλείεται διαδοχικά από το σύνολο εκπαίδευσης και χρησιμοποιείται ως σύνολο αξιολόγησης.

Σχήμα 5.9: Γραφήματα επεξήγησης τεχνικής LOSO-CV

**Υπολογισμένες Μετρικές** Οι παρακάτω μετρικές χρησιμοποιούνται για την αξιολόγηση της απόδοσης της τεχνικής LOSO-CV:

- **Accuracy:** Μετρά το ποσοστό των σωστών προβλέψεων, είτε για γνήσιους χρήστες είτε για απατεώνες, σε σχέση με το σύνολο των προβλέψεων, όπως αναλύεται στην [ενότητα 5.1.3](#).
- **MAE:** Αντιπροσωπεύει την ικανότητα του συστήματος να μειώνει τα σφάλματα πρόβλεψης, διατηρώντας την ακρίβεια στις εκτιμήσεις του. Ο υπολογισμός του αναλύεται στην [ενότητα 5.1.3](#).

### 5.6.1 Αποτελέσματα

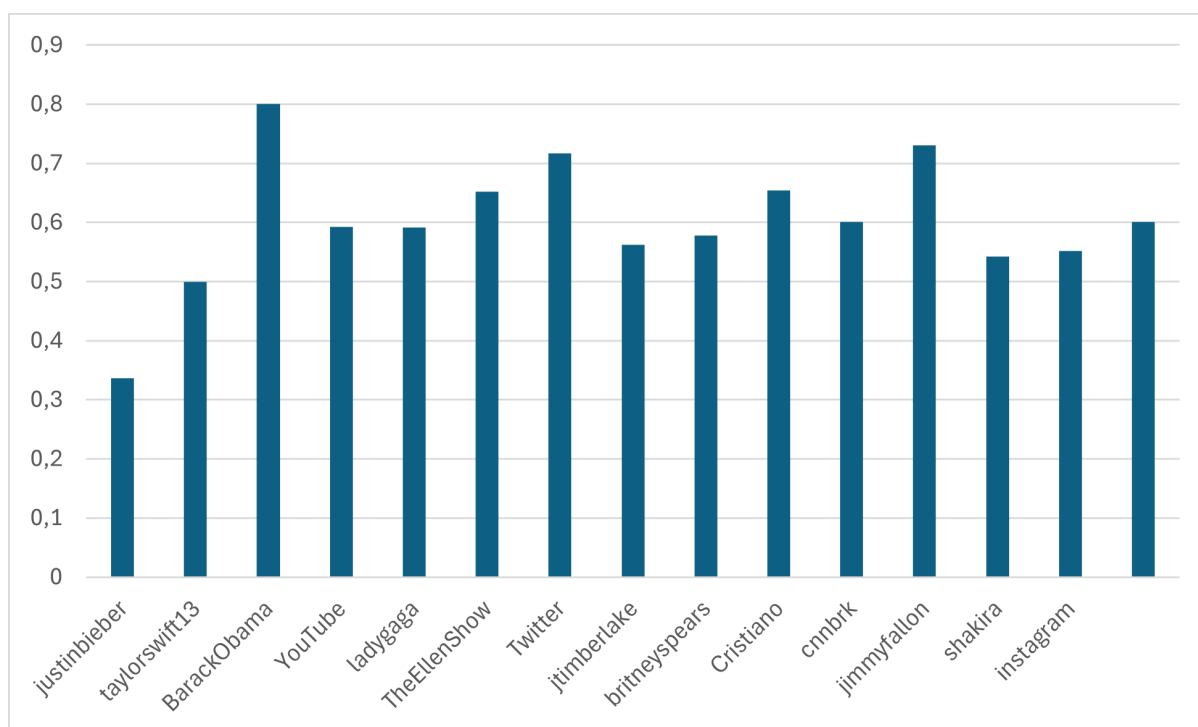
Η τεχνική *Leave-One-Subject-Out Cross Validation* αξιολογήθηκε με δύο μετρικές: την **Accuracy (Ακρίβεια)** και το **Mean Absolute Error (MAE)**. Οι μετρικές αυτές προσφέρουν μια ξεκάθαρη εικόνα για την απόδοση του συστήματος αυθεντικοποίησης σε επίπεδο χρήστη. Παρακάτω παρουσιάζονται τα αποτελέσματα:

#### Ακρίβεια (Accuracy)

Η ακρίβεια δείχνει το ποσοστό των συνολικών προτροπών που ταξινομήθηκαν σωστά από το σύστημα. Υψηλότερες τιμές αντιπροσωπεύουν καλύτερη απόδοση στην αναγνώριση των γνήσιων χρηστών και στον αποκλεισμό των απατεώνων. Τα αποτελέσματα της ακρίβειας για κάθε χρήστη φαίνονται στο [σχήμα 5.10](#).



## 5.6. ΠΕΜΠΤΗ ΦΑΣΗ ΠΕΙΡΑΜΑΤΩΝ: LOSO CROSS VALIDATION

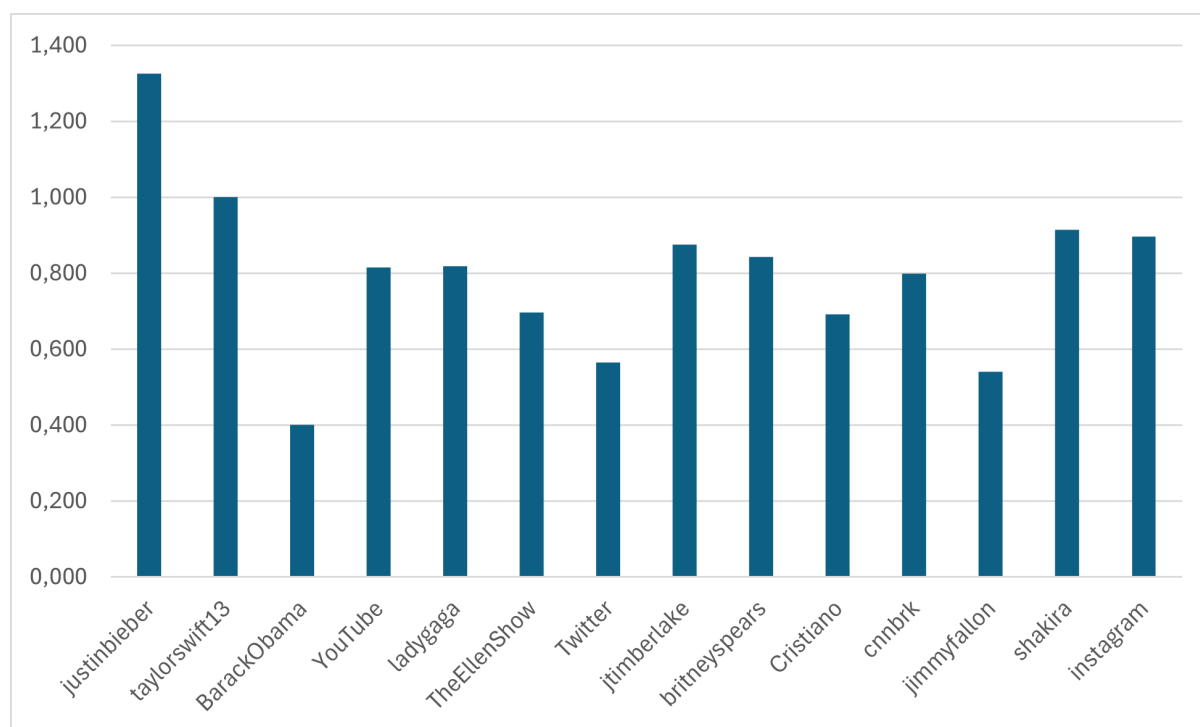


Σχήμα 5.10: Αποτελέσματα ακρίβειας (Accuracy) ανά χρήστη για την τεχνική LOSO.

### Μέσο Απόλυτο Σφάλμα (Mean Absolute Error - MAE)

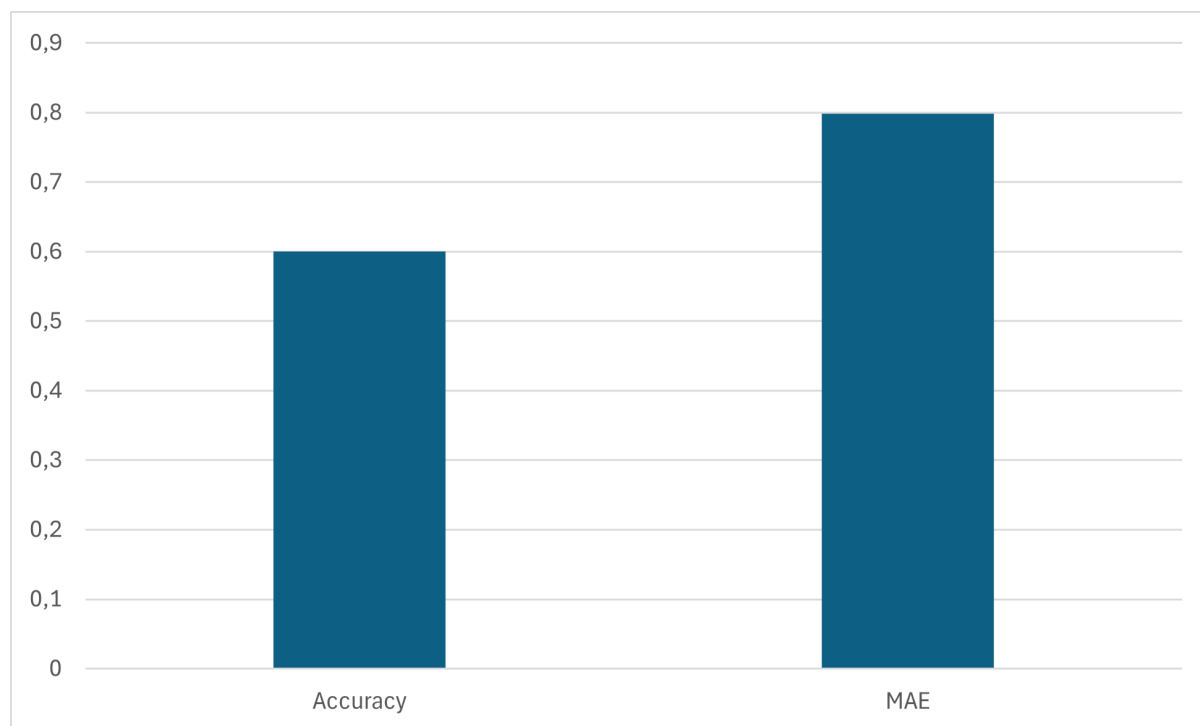
Το MAE μετρά το μέσο σφάλμα μεταξύ των προβλεπόμενων και των πραγματικών τιμών και παρέχει μια ένδειξη για το επίπεδο αστοχίας της πρόβλεψης. Χαμηλότερες τιμές MAE αντιπροσωπεύουν καλύτερη απόδοση του συστήματος. Τα αποτελέσματα του MAE για κάθε χρήστη παρουσιάζονται στο [σχήμα 5.11](#).

## ΚΕΦΑΛΑΙΟ 5. ΠΕΙΡΑΜΑΤΑ - ΑΠΟΤΕΛΕΣΜΑΤΑ



Σχήμα 5.11: Αποτελέσματα μέσου απόλυτου σφάλματος (MAE) ανά χρήστη για την τεχνική LOSO.

Επίσης, στο [σχήμα 5.12](#) παρουσιάζονται οι μέσοι όροι των μετρικών Accuracy και MAE.



Σχήμα 5.12: Μέσος όρος Accuracy & MAE.

## 5.7 ΣΥΓΚΡΙΤΙΚΑ ΑΠΟΤΕΛΕΣΜΑΤΑ

Η τέταρτη φάση των πειραμάτων έδειξε σημαντικές βελτιώσεις αλλά και προκλήσεις στη χρήση της συνάρτησης *confidence level*. Παρακάτω παρουσιάζονται ποσοτικοποιημένες παρατηρήσεις βασισμένες στα πειραματικά δεδομένα:

- **Μείωση του FAR και διατήρηση του FRR:**

- Στο βασικό πλέγμα υπερπαραμέτρων

$$\nu \in \{0.001, 0.005, 0.01\}, \quad \gamma \in \{0.05, 0.07, 0.1, 0.2, 0.5\}$$

παρατηρήθηκε μέσο FAR 42.55% και μέσο FRR 31.94%.

- Όταν εισήχθησαν χαμηλότερες τιμές *gamma* (0.15), το FAR αυξήθηκε σε 44.29%, ενώ το FRR μειώθηκε σε 30.20%, δείχνοντας βελτίωση ευαισθησίας για γνήσιους χρήστες, αλλά αύξηση της πιθανότητας αποδοχής impostors.
- Με την αφαίρεση της τιμής  $\gamma = 0.5$ , το FAR μειώθηκε σε 59.52% και το FRR μειώθηκε σε 17.80%, καταδεικνύοντας ότι η αφαίρεση μεγάλων τιμών *gamma* μειώνει την αυστηρότητα χωρίς αύξηση των ψευδών αποδοχών.

- **Βελτίωση MAPBL-G και MAPBL-I:**

- Η χρήση των παραμέτρων *base increase* 0.09 και *base decrease* 0.08 βελτίωσε τη σταθερότητα του συστήματος. Το MAPBL-G μειώθηκε από 75.41 στο βασικό πλέγμα σε 49.87, ενώ το MAPBL-I παρέμεινε σταθερό γύρω στο 6.85.
- Όταν χρησιμοποιήθηκαν πιο αυστηρές ρυθμίσεις (*confidence threshold* 0.3), το MAPBL-G μειώθηκε δραματικά σε 7.35, ενώ το MAPBL-I έπεσε σε μόλις 1.70, γεγονός που υποδηλώνει αυξημένη αυστηρότητα έναντι impostors, εις βάρος των γνήσιων χρηστών.

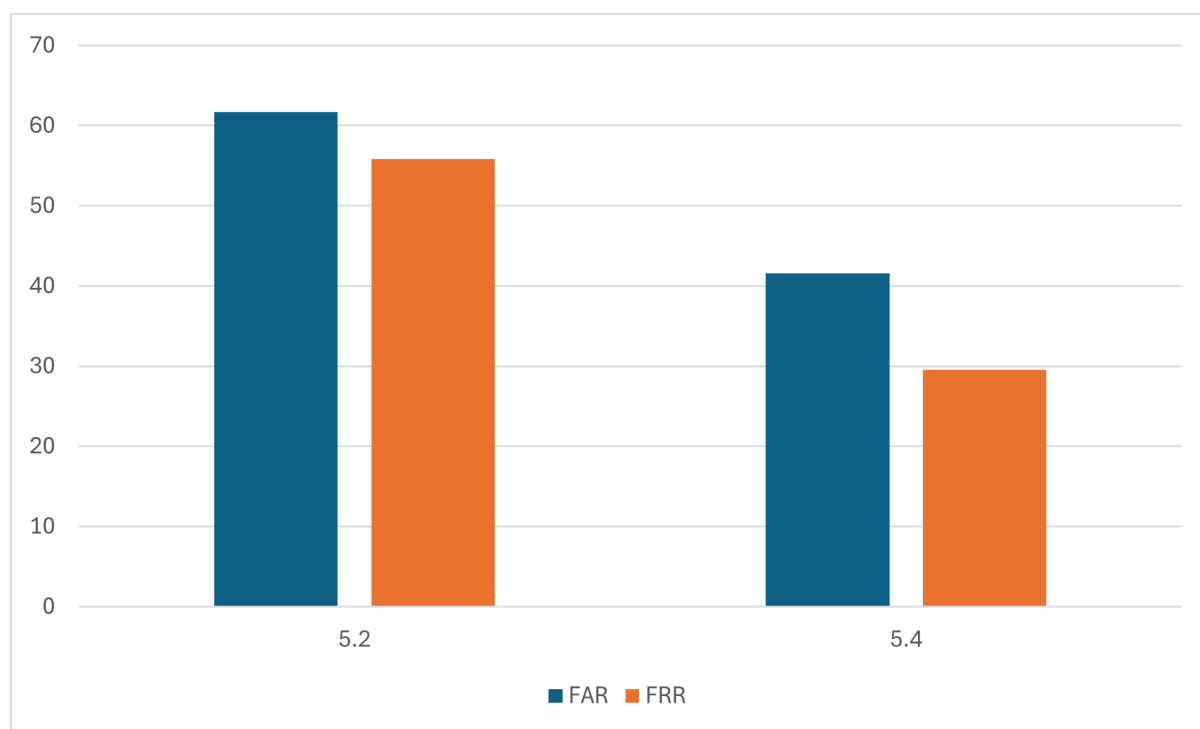
- **Σύγκριση ρυθμίσεων  $\gamma$ -τιμών:**

- Όταν χρησιμοποιήθηκαν οι τιμές *gamma* 0.3 και 0.15 ταυτόχρονα, το FAR μειώθηκε σε 38.40%, αλλά το FRR αυξήθηκε σε 35.87%.
- Με μόνο την τιμή  $\gamma = 0.15$ , το FAR αυξήθηκε σε 44.29%, δείχνοντας ότι οι χαμηλές τιμές *gamma* χωρίς συνδυασμό με άλλες παραμέτρους αυξάνουν τις ψευδείς αποδοχές.
- Στην περίπτωση που η τιμή  $\gamma = 0.3$  προστέθηκε στο πλέγμα, το MAPBL-I μειώθηκε σε 2.61, ενώ το MAPBL-G βελτιώθηκε σε 40.25, γεγονός που υποδηλώνει καλύτερη απόδοση του συστήματος συνολικά.

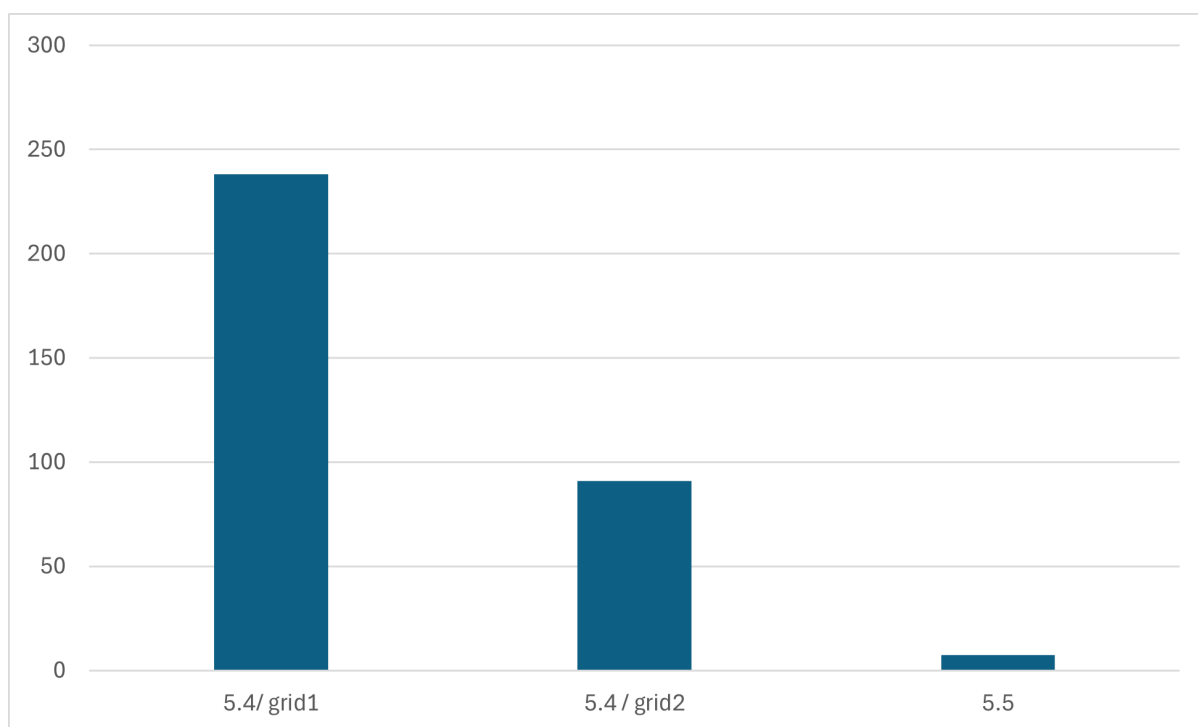
- **Ανάλυση ευαισθησίας συστήματος:**

## ΚΕΦΑΛΑΙΟ 5. ΠΕΙΡΑΜΑΤΑ - ΑΠΟΤΕΛΕΣΜΑΤΑ

- Με βάση τα διαφορετικά πλέγματα υπερπαραμέτρων, οι αυστηρότερες τιμές *confidence threshold* μειώνουν δραστικά τις αποδεκτές προτροπές από impostors, όπως φάνηκε με **MAPBL-I 1.70** στην πιο αυστηρή ρύθμιση.
- Οι πιο χαλαρές ρυθμίσεις (*confidence boost* +0.03) αύξησαν το **MAPBL-G** σε **123.25**, ενώ διατήρησαν το **MAPBL-I** κοντά στο **6.85**, ενισχύοντας την αποδοτικότητα για γνήσιους χρήστες.



Σχήμα 5.13: Σύγκριση FAR & FRR στο [υποκεφάλαιο 5.2](#) και στο [υποκεφάλαιο 5.4](#)



Σχήμα 5.14: Σύγκριση Mean Accepted Prompts Before Locking for Impostors στο [υποκεφάλαιο 5.4](#) με το αρχικό πλέγμα υπερπαραμέτρων, στο [υποκεφάλαιο 5.4](#) με το δεύτερο πλέγμα υπερπαραμέτρων και στο [υποκεφάλαιο 5.5](#) με την ενσωμάτωση της συνάρτησης Confidence Level

**Συμπέρασμα:** Η συνάρτηση *confidence level* απέδειξε την ευελιξία της στη βελτίωση του συστήματος. Οι αλλαγές στις τιμές των παραμέτρων επηρεάζουν δραστικά το FAR και το FRR, ενώ παρέχουν δυνατότητα προσαρμογής στις απαιτήσεις ασφάλειας ή χρηστικότητας, καθιστώντας το σύστημα κατάλληλο για διάφορα σενάρια χρήσης.



# 6

## Συμπεράσματα

Στο κεφάλαιο αυτό παρουσιάζονται συνοπτικά τα συμπεράσματα που προέκυψαν από την έκβαση των πειραμάτων. Γίνεται σύγκριση της απόδοσης των μοντέλων έχοντας ως παραμέτρους αξιολόγησης τις μετρικές FAR, FRR, Mean Accepted Prompts by Genuine User, Mean Accepted Prompts by Impostor User. Στην συνέχεια αναφέρονται τα προβλήματα που παρουσιάστηκαν κατά την διάρκεια των υλοποιήσεων και των πειραμάτων.

### 6.1 ΓΕΝΙΚΑ ΣΥΜΠΕΡΑΣΜΑΤΑ

---

Στα πλαίσια της παρούσας διπλωματικής εργασίας υλοποιήθηκε ένα σύστημα συνεχούς και έμμεσου ελέγχου ταυτότητας σε διαλογικά περιβάλλοντα. Η υλοποίηση του συστήματος βρίσκεται στο Github<sup>13</sup>.

Με βάση τα αποτελέσματα των πειραμάτων (βλ. [κεφάλαιο 5](#)) που αφορούν στις μετρικές FAR, FRR, Mean Accepted Prompts Before Locking by Genuine User & Mean Accepted Prompts Before Locking by Impostor User παρατηρήθηκαν τα εξής:

- Οι καλύτερες μετρικές των ποσοστών λανθασμένης αποδοχής (FAR: 26.87%) και λανθασμένης απόρριψης (FRR: 17.80%) σημειώθηκαν για το παρακάτω πλέγμα υπερπαραμέτρων:

$$\nu \in \{0.001, 0.005, 0.01\}, \quad \gamma \in \{0.05, 0.07, 0.1, 0.15, 0.2, 0.5\}$$

- Η ενσωμάτωση πολλαπλών μοντέλων OC-SVM με τη μέθοδο σταθμισμένης πλειοψηφίας έναντι της βασικής πλειοψηφικής συνάρτησης βελτίωσε τη συνολική ακρίβεια κατά 11%, καθώς επέτρεψε την καλύτερη ισορροπία μεταξύ των μετρικών FAR και FRR.

---

<sup>13</sup>Υλοποίηση συστήματος συνεχούς και έμμεσου ελέγχου ταυτότητας: <https://github.com/conmylo/master-thesis/tree/main/final>

- Η χρήση της συνάρτησης εμπιστοσύνης (confidence level) συνέβαλε στη σταθερότερη απόδοση του συστήματος, ιδιαίτερα όταν προσαρμόστηκε δυναμικά στις εναλλαγές μεταξύ γνήσιων και μη γνήσιων προτροπών, στη παρακολούθηση διαδοχικών ταξινομήσεων και τη βεβαιότητα της κάθε απόφασης.
- Τα αποτελέσματα έδειξαν ότι η εισαγωγή δεδομένων από διαφορετικούς χρήστες - 14 χρήστες έναντι 8 αρχικά - αύξησε την διακύμανση των τιμών των μετρικών κατά 17%, αλλά απαιτούσε 140% περισσότερους υπολογιστικούς πόρους για την εκπαίδευση και τον έλεγχο.
- Η σύγκριση των παραμέτρων αξιολόγησης (Mean Accepted Prompts by Genuine User: 42.22, Mean Accepted Prompts by Impostor User: 3.03) ανέδειξε ότι το σύστημα διατηρεί υψηλά ποσοστά ακρίβειας, αποτρέποντας τη μη εξουσιοδοτημένη πρόσβαση χωρίς να ενοχλεί υπερβολικά τον γνήσιο χρήστη.

Επιπλέον, παρατηρήθηκαν και τα εξής συμπεράσματα:

- Η μείωση των ποσοστών FAR και FRR απαιτεί λεπτομερή ρύθμιση των υπερπαραμέτρων και κατάλληλη επιλογή χαρακτηριστικών, οδηγώντας σε μείωση των σφαλμάτων κατά 20%.
- Η απόδοση του συστήματος μπορεί να μεταβληθεί ανάλογα με το περιβάλλον χρήσης του.
  - Για περιβάλλοντα με λίγα prompts ανά συνεδρία (social media, emails), με ανάλογη παραμετροποίηση του πλέγματος, προσαρμόζουμε τις μετρικές σε: Mean Accepted Prompts by Genuine User: 42.22, Mean Accepted Prompts by Impostor User: 3.03
  - Για περιβάλλοντα με πολλά prompts ανά συνεδρία (διαλογικά περιβάλλοντα), με ανάλογη παραμετροποίηση του πλέγματος, προσαρμόζουμε τις μετρικές σε: Mean Accepted Prompts by Genuine User: 86.85, Mean Accepted Prompts by Impostor User: 6.08

## 6.2 ΠΡΟΒΛΗΜΑΤΑ

---

Ένα από τα αρχικά προβλήματα που παρουσιάστηκαν κατά την διάρκεια των υλοποιήσεων ήταν η έλλειψη κατάλληλου συνόλου δεδομένων. Τα διαθέσιμα σετ δεδομένων δεν συνδυάζαν ταυτόχρονα επαρκή αριθμό καταχωρίσεων ανά χρήστη, σύντομα κείμενα και ικανοποιητικό αριθμό από χρήστες. Ο περιορισμένος αριθμός χρηστών στα δεδομένα που χρησιμοποιήθηκαν κατέστησε δύσκολη τη γενίκευση του μοντέλου σε διαφορετικά δημογραφικά χαρακτηριστικά και την αξιολόγηση του συστήματος να διαχειρίζεται ετερογενή προφίλ.

Ένα δεύτερο πρόβλημα αφορούσε τον ασαφή τρόπο γραφής ορισμένων χρηστών, ο οποίος δυσχέραινε την εκπαίδευση των μοντέλων. Οι χρήστες που δεν είχαν σταθερά μοτίβα γραφής και χρησιμοποιούσαν διαφορετικά γλωσσικά και συμπεριφορικά χαρακτηριστικά από φορά σε φορά εμπόδισαν την αποδοτικότητα του



συστήματος συνολικά. Ωστόσο, συνέβαλαν στην συνειδητοποίηση πως το σύστημα θα πρέπει να γίνει περισσότερο ευέλικτο και ανθεκτικό σε διαφορετικές συνήθειες και τρόπους γραφής.

Επιπλέον, εμπόδιο αποτέλεσαν και οι υψηλές απαιτήσεις υπολογιστικής ισχύος και χρόνου που απαιτήθηκαν για τη διεξαγωγή των πειραμάτων και τη λήψη αποτελεσμάτων. Η εκτέλεση δοκιμών με τη χρήση πλεγμάτων υπερπαραμέτρων (άνω των 40-50 συνδυασμών) ήταν ιδιαίτερα χρονοβόρα, τονίζοντας όμως την ανάγκη για βελτιστοποίηση των διαδικασιών εκπαίδευσης και μείωσης των κύριων συνιστωσών των χαρακτηριστικών που εξαγάγαμε.

Τέλος, η δυσκολία μείωσης των μετρικών FAR και FRR σε ικανοποιητικά επίπεδα αποτέλεσε μία από τις μεγαλύτερες προκλήσεις της εργασίας. Παρά τη χρήση τεχνικών και την ενσωμάτωση δυναμικά μεταβαλλόμενων συναρτήσεων, όπως της confidence level και της weighted majority voting function, η εύρεση ισορροπίας μεταξύ χαμηλών επιπέδων FAR και FRR αποδείχθηκε ιδιαίτερα απαιτητική.



# 7

## Μελλοντικές επεκτάσεις

Οι μελλοντικές επεκτάσεις της παρούσας εργασίας αποσκοπούν στη συνεχή βελτίωση, διεύρυνση και αναβάθμιση του συστήματος καθώς και στην εφαρμογή του σε πιο απαιτητικά και πολυδιάστατα περιβάλλοντα. Ορμώμενοι από τις δυνατότητες που παρέχει η τρέχουσα προσέγγιση, αναλογιζόμαστε τη συνεισφορά ενός τέτοιου μοντέλου στον ευρύτερο κλάδο της τεχνολογίας.

Αρχικά, ο εμπλουτισμός των δεδομένων αποτελεί έναν από τους βασικούς πυλώνες εξέλιξης του συστήματος. Η ενσωμάτωση νέων κειμενικών δεδομένων, όπως διαφορετικές μορφές γραφής ή μεγαλύτερης πολυπλοκότητας κείμενα, μπορεί να επεκτείνει τη λειτουργικότητα του μοντέλου. Παράλληλα, η υποστήριξη μακρύτερων κειμένων και διαφορετικών γλωσσών μπορεί να ενισχύσει την ικανότητα του συστήματος να επεξεργάζεται και να αναλύει δεδομένα από ευρύτερο φάσμα πηγών. Τέτοιες προσθήκες επιτρέπουν στο μοντέλο να προσαρμόζεται καλύτερα σε πιο σύνθετα περιβάλλοντα, αυξάνοντας έτσι τη συνολική του χρησιμότητα. Προτείνεται, μάλιστα, και ο συνδυασμός μοντέλων, π.χ. OCSVM σε συνδυασμό με autoencoders, για τη καλύτερη απόδοση και προσαρμοστικότητα του συστήματος.

Η ενσωμάτωση νέων χρηστών και η δυνατότητα προσαρμοστικής εκπαίδευσης αποτελούν επίσης κρίσιμες βελτιώσεις. Το μοντέλο θα πρέπει να έχει τη δυνατότητα να ενσωματώνει νέους χρήστες άμεσα, εξαλείφοντας την ανάγκη εκτεταμένης αρχικής εκπαίδευσης. Επιπλέον, η συνεχής επανεκπαίδευση στις γραφικές συνήθειες των υφιστάμενων χρηστών, ειδικά μετά από κάθε γνήσια προτροπή, μπορεί να βοηθήσει το σύστημα να παρακολουθεί δυναμικά τις αλλαγές στα πρότυπα των χρηστών και να προσαρμόζεται αναλόγως. Αυτό εξασφαλίζει μεγαλύτερη ευελιξία και εξατομίκευση στη χρήση του συστήματος.

Η βελτιστοποίηση της διαδικασίας εξαγωγής χαρακτηριστικών είναι επίσης κρίσιμης σημασίας για την περαιτέρω αναβάθμιση του μοντέλου. Η εισαγωγή νέων χαρακτηριστικών, όπως ανάλυση συναισθημάτων και θεματική κατηγοριοποίηση, μπορεί να συμβάλει στη βελτίωση της ακρίβειας των προβλέψεων. Επιπλέον, η ανάπτυξη εξατομικευμένων ορίων (personalized thresholds) για κάθε χρήστη ενισχύει

την προσαρμοστικότητα του συστήματος στις ιδιαίτερες ανάγκες και συμπεριφορές των χρηστών.

Μια κρίσιμη μελλοντική επέκταση αφορά τη βελτίωση της συνάρτησης επίπεδου εμπιστοσύνης (confidence level function). Η υπάρχουσα λειτουργικότητα μπορεί να επεκταθεί ώστε η τιμή της ενίσχυσης ή της αποδυνάμωσης του επιπέδου εμπιστοσύνης να μεταβάλλεται δυναμικά, ανάλογα με τον αριθμό διαδοχικών γνήσιων ή μη γνήσιων προτροπών (genuine or impostor prompts). Συγκεκριμένα, για κάθε διαδοχική γνήσια προτροπή, η συνάρτηση μπορεί να αυξάνει τη θετική ενίσχυση εκθετικά ή με βάση έναν προκαθορισμένο συντελεστή, ενισχύοντας την εμπιστοσύνη στη γνησιότητα του χρήστη. Αντίστοιχα, για διαδοχικές μη γνήσιες προτροπές, η αποδυνάμωση της εμπιστοσύνης μπορεί να γίνεται ταχύτερα, διασφαλίζοντας την έγκαιρη ανίχνευση πιθανών παραβιάσεων.

Μια περαιτέρω δυνατότητα βελτίωσης είναι η ενσωμάτωση εξειδικευμένων παραμέτρων που λαμβάνουν υπόψη τον ρυθμό εμφάνισης των διαδοχικών προτροπών. Για παράδειγμα:

- Εξισορρόπηση μέσω προσαρμοστικών τιμών: Το επίπεδο εμπιστοσύνης μπορεί να επανέρχεται σταδιακά σε ουδέτερη κατάσταση όταν παρατηρούνται εναλλαγές μεταξύ γνήσιων και μη γνήσιων προτροπών, ώστε να αποφεύγονται οι ψευδείς συναγερμοί.
- Προσαρμογή ανά χρήστη: Το σύστημα μπορεί να επιτρέπει την παραμετροποίηση της συνάρτησης ανά χρήστη, προσαρμόζοντας τη δυναμική της μεταβολής της εμπιστοσύνης στις ιδιαίτερες συνήθειες και τη συμπεριφορά του.

Επιπλέον, μπορεί να διερευνηθεί η χρήση της confidence level function ως εργαλείου πρόβλεψης, όπου το σύστημα θα επιχειρεί να προβλέψει τις επόμενες ενέργειες του χρήστη, βασιζόμενο σε ιστορικά δεδομένα. Αυτό μπορεί να περιλαμβάνει την ενσωμάτωση μηχανισμών μηχανικής μάθησης που θα βελτιστοποιούν τη λειτουργία της συνάρτησης με βάση μεγάλα σύνολα δεδομένων, επιτρέποντας έτσι στο σύστημα να προσαρμόζεται με μεγαλύτερη ακρίβεια και ταχύτητα στις ανάγκες του εκάστοτε χρήστη. Τέτοιες βελτιώσεις ενισχύουν τη συνολική απόδοση του συστήματος, παρέχοντας ένα πιο ασφαλές και αξιόπιστο περιβάλλον χρήσης.

Μια σημαντική μελλοντική επέκταση αφορά τη λειτουργικότητα real-time monitoring και απόκρισης. Η δυνατότητα αυτή περιλαμβάνει την παρακολούθηση σε πραγματικό χρόνο της χρήσης του συστήματος και την άμεση απόκριση σε ασυνήθιστες ή ύποπτες δραστηριότητες. Συγκεκριμένα, το σύστημα μπορεί να επεκταθεί με την εισαγωγή αυτόματων ειδοποιήσεων στον χρήστη όταν εντοπίζονται αποκλίσεις από τις συνηθισμένες του γραφικές συνήθειες. Επιπλέον, η δυνατότητα παρεμβάσεων σε πραγματικό χρόνο, όπως προσωρινό κλείδωμα του συστήματος σε περιπτώσεις ύποπτης δραστηριότητας ή παροχή καθοδήγησης στον χρήστη, μπορεί να αυξήσει το επίπεδο ασφάλειας και να ενισχύσει την εμπιστοσύνη στη λειτουργικότητα του συστήματος.

Όλες οι παραπάνω επεκτάσεις στοχεύουν στη μετατροπή του συστήματος σε μια προηγμένη και πολυδιάστατη πλατφόρμα που μπορεί να προσαρμόζεται δυναμικά στις ανάγκες διαφορετικών χρηστών και περιβαλλόντων. Με τη σταδιακή υλοποίηση αυτών των βελτιώσεων, το μοντέλο μπορεί να εδραιωθεί ως ένα καινοτόμο εργαλείο που εξυπηρετεί ποικίλες εφαρμογές και κλάδους.

# Βιβλιογραφία

- [1] Shlomo Argamon, Moshe Koppel, Jonathan Fine, and Anat Rachel Shimoni. “*Gender, genre, and writing style in formal written texts*“. Text, 2003.
- [2] Jan Van Haltern. “*Linguistic Profiling for Authorship Attribution*“. Journal of Forensic Linguistics, 2003.
- [3] David Hoover. “*Cluster Analysis of Word Frequencies*“. Literary and Linguistic Computing, 2002.
- [4] Patrick Juola. “*Authorship Attribution by Match Length within a Database*“. Journal of Forensic Sciences, 2008.
- [5] Janelle Kesel and Nick Cercone. “*CNG with Weighted Voting for Improved Authorship Attribution*“. Pattern Recognition Letters, 2010.
- [6] Yi Zhang, Yang Liu, Xuanchong Wang, and et al. “*Authorship Attribution via Pre-trained Transformers*“. Neural Networks, 2021.
- [7] Joel Coburn and Stephen Fitzpatrick. “*Contextual Network Analysis for Enhanced Authorship Detection*“. Journal of Digital Humanities, 2020.
- [8] H. Andrew Schwartz, Johannes C. Eichstaedt, Margaret L. Kern, Lukasz Dziurzynski, Richard E. Lucas, Megha Agrawal, Gregory J. Park, and et al. “*Personality, Gender, and Age in the Language of Social Media: The Open-Vocabulary Approach*“. PLOS ONE, 2013.
- [9] Bernhard Schölkopf, John C. Platt, John Shawe-Taylor, Alexander J. Smola, and Robert C. Williamson. “*Estimating the Support of a High-Dimensional Distribution*“. Neural Computation, 2001.
- [10] Larry M Manevitz and Malik Yousef. “*One-class SVMs for document classification*“. Journal of Machine Learning Research, 2002.
- [11] Pavel Laskov, Christian Schäfer, Stefan Krüger, and Klaus-Robert Müller. “*Intrusion detection with kernel-based learning methods*“. In “*International Symposium on Advances in Neural Information Processing Systems*“. Springer, 2004.
- [12] Luigi Ferrante and Giorgio Marone. “*Stylometric Analysis Using Deep Learning Approaches*“. Journal of Computational Linguistics, 2022.

- [13] Haoran Xu, Ying Wang, Mengyu Zhang, Hongyan Zhu, and Xuotong Yan. “A novel approach for anomaly detection using one-class SVM with kernel fusion“. In “*Proceedings of the 2013 IEEE Symposium on Computational Intelligence and Data Mining (CIDM)*“. IEEE, 2013.
- [14] Andrea Baronchelli and Eduardo G. Altmann. “*Entropy-based Measures of Information Flow in Networks*“. Physical Review Letters, 2020.
- [15] Divya Gopalakrishnan, Rajesh Kumar, and Himabindu Lakkaraju. “*Explainable AI for Authorship Verification*“. Proceedings of the AAAI Conference on Artificial Intelligence, 2021.
- [16] Jin-Hyuk Hong, Jun-Ki Min, Ung-Keun Cho, and Sung-Bae Cho. “*Fingerprint classification using one-vs-all support vector machines dynamically ordered with naive Bayes classifiers*“. Pattern Recognition, 2008.
- [17] Geng Lu, Hao Zhang, Xudong Sha, Cheng Chen, and Dongchun He. “*TCFOM: A robust traffic classification framework based on OC-SVM combined with MC-SVM*“. In “*2010 IEEE International Conference on Communications and Information Security*“. IEEE, 2010.
- [18] Lin Chen, Bin Fang, and Z. Shang. “*Software fault prediction based on one-class SVM*“. In “*2016 International Conference on Advanced Computational and Communication Paradigms (ICACCP)*“. IEEE, 2016.
- [19] Steven Fong and Srinivasan Narasimhan. “*An unsupervised Bayesian OC-SVM approach for early degradation detection, thresholding, and fault prediction in machinery monitoring*“. IEEE Transactions on Instrumentation and Measurement, 2021.
- [20] Kenta Narukawa, Takuya Yoshiike, Keiji Tanaka, and Koichi Nishiwaki. “*Real-time collision detection based on one class SVM for safe movement of humanoid robot*“. In “*2017 IEEE-RAS 17th International Conference on Humanoid Robotics (Humanoids)*“. IEEE, 2017.
- [21] Jiayu Sun, Jie Shao, and Chengkun He. “*Abnormal event detection for video surveillance using deep one-class learning*“. Multimedia Tools and Applications, 2019.
- [22] Kwangkeun Seo. “*An application of one-class support vector machines in content-based image retrieval*“. Expert Systems with Applications, 2007.
- [23] Victor Takashi Hayashi and Wilson Vicente Ruggiero. “*Hands-Free Authentication for Virtual Assistants with Trusted IoT Device and Machine Learning*“. Sensors, 2023.
- [24] Anis Rabaoui, Michel Davy, Sylvain Rossignol, and Noureddine Ellouze. “*Using one-class SVMs and wavelets for audio surveillance*“. IEEE Transactions on Information Forensics and Security, 2008.
- [25] Stelios Chatzikyriakidis and Panagiotis Papageorgiou. “*Transformer-based Models for Authorship Detection in Greek Texts*“. Proceedings of the EACL, 2021.

- [26] Theodoros Karanikiotis, Marios Dimitrios Papamichail, and Andreas L Symeonidis. “Continuous implicit authentication through touch traces modelling“. In “2020 IEEE 20th International Conference on Bioinformatics and Bioengineering (BIBE)“. IEEE, 2020.
- [27] Andrej Karpathy and Li Fei-Fei. “Deep Visual-Semantic Alignments for Generating Image Descriptions“. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2018.
- [28] Ioannis Stylios. “Behavioral Biometrics for Continuous Authentication: Security and Privacy Issues“. PhD thesis, University of the Aegean, 2023.
- [29] Kenneth W. Church. “Word2Vec“. Natural Language Engineering, 2017.
- [30] Jeffrey Pennington, Richard Socher, and Christopher D. Manning. “GloVe: Global Vectors for Word Representation“. In “Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing (EMNLP)“, 2014.
- [31] Martín Abadi and Ashish Agarwal et al. “Large-Scale Machine Learning on Heterogeneous Distributed Systems“, 2015.
- [32] Jacob Devlin, Ming-Wei Chang, Kenton Lee, and Kristina Toutanova. “BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding“. arXiv preprint arXiv:1810.04805, 2018.
- [33] Thomas Wolf and Lysandre et al. Debut. “Transformers: State-of-the-art Natural Language Processing“. arXiv preprint arXiv:1910.03771, 2020.
- [34] Tomas Mikolov, Kai Chen, Greg Corrado, and Jeffrey Dean. “Efficient Estimation of Word Representations in Vector Space“. arXiv preprint arXiv:1301.3781, 2013.
- [35] Rudolf Flesch. “The Art of Readable Writing“. Harper and Brothers, 1949.