



ΕΘΝΙΚΟ ΜΕΤΣΟΒΙΟ ΠΟΛΥΤΕΧΝΕΙΟ
ΣΧΟΛΗ ΗΛΕΚΤΡΟΛΟΓΩΝ ΜΗΧΑΝΙΚΩΝ ΚΑΙ ΜΗΧΑΝΙΚΩΝ
ΥΠΟΛΟΓΙΣΤΩΝ
ΤΟΜΕΑΣ ΤΕΧΝΟΛΟΓΙΑΣ ΠΛΗΡΟΦΟΡΙΚΗΣ ΚΑΙ ΥΠΟΛΟΓΙΣΤΩΝ

**Δημιουργία μηχανής σύστασης για
Συνεργατική, Διασυνδεδεμένη και
Αυτοματοποιημένη Κινητικότητα**

ΔΙΠΛΩΜΑΤΙΚΗ ΕΡΓΑΣΙΑ

ΤΟΥ

ΒΙΤΖΗΛΑΙΟΥ ΓΕΩΡΓΙΟΥ

Επιβλέπων: Παναγιώτης Τσανάκας
Καθηγητής ΕΜΠ

Αθήνα, Μάιος 2025



ΕΘΝΙΚΟ ΜΕΤΣΟΒΙΟ ΠΟΛΥΤΕΧΝΕΙΟ
ΣΧΟΛΗ ΗΛΕΚΤΡΟΛΟΓΩΝ ΜΗΧΑΝΙΚΩΝ ΚΑΙ ΜΗΧΑΝΙΚΩΝ
ΥΠΟΛΟΓΙΣΤΩΝ
ΤΟΜΕΑΣ ΤΕΧΝΟΛΟΓΙΑΣ ΠΛΗΡΟΦΟΡΙΚΗΣ ΚΑΙ ΥΠΟΛΟΓΙΣΤΩΝ

Δημιουργία μηχανής σύστασης για Συνεργατική, Διασυνδεδεμένη και Αυτοματοποιημένη Κινητικότητα

ΔΙΠΛΩΜΑΤΙΚΗ ΕΡΓΑΣΙΑ

ΤΟΥ

ΒΙΤΖΗΛΑΙΟΥ ΓΕΩΡΓΙΟΥ

Επιβλέπων: Παναγιώτης Τσανάκας
Καθηγητής ΕΜΠ

Εγκρίθηκε από την τριμελή εξεταστική επιτροπή την 2η Μαΐου 2025.

(Υπογραφή)

(Υπογραφή)

(Υπογραφή)

.....
Παναγιώτης Τσανάκας
Καθηγητής ΕΜΠ

.....
Ανδρέας Σταφυλοπάτης
Ομότιμος Καθηγητής ΕΜΠ

.....
Αθανάσιος Βουλόδημος
Επικ. Καθηγητής ΕΜΠ

Αθήνα, Μάιος 2025



ΕΘΝΙΚΟ ΜΕΤΣΟΒΙΟ ΠΟΛΥΤΕΧΝΕΙΟ
ΣΧΟΛΗ ΗΛΕΚΤΡΟΛΟΓΩΝ ΜΗΧΑΝΙΚΩΝ ΚΑΙ ΜΗΧΑΝΙΚΩΝ
ΥΠΟΛΟΓΙΣΤΩΝ
ΤΟΜΕΑΣ ΤΕΧΝΟΛΟΓΙΑΣ ΠΛΗΡΟΦΟΡΙΚΗΣ ΚΑΙ ΥΠΟΛΟΓΙΣΤΩΝ

Copyright © Βιτζηλαίος Γεώργιος, 2025.

Με την επιφύλαξη παντός δικαιώματος. All rights reserved

Απαγορεύεται η αντιγραφή, αποθήκευση και διανομή της παρούσας εργασίας, εξ ολοκλήρου ή τμήματος αυτής, για εμπορικό σκοπό. Επιτρέπεται η ανατύπωση, αποθήκευση και διανομή για σκοπό μη κερδοσκοπικό, εκπαιδευτικής ή ερευνητικής φύσης, υπό την προϋπόθεση να αναφέρεται η πηγή προέλευσης και να διατηρείται το παρόν μήνυμα. Ερωτήματα που αφορούν τη χρήση της εργασίας για κερδοσκοπικό σκοπό πρέπει να απευθύνονται προς τον συγγραφέα.

Οι απόψεις και τα συμπεράσματα που περιέχονται σε αυτό το έγγραφο εκφράζουν τον συγγραφέα και δεν πρέπει να ερμηνευθεί ότι αντιπροσωπεύουν τις επίσημες θέσεις του Εθνικού Μετσόβιου Πολυτεχνείου.

Υπεύθυνη Δήλωση

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

(Υπογραφή)

.....

Βιτζηλαίος Γεώργιος

Περίληψη

Η παρούσα διπλωματική εργασία εκπονήθηκε στα πλαίσια του προπτυχιακού κύκλου σπουδών της Σχολής Ηλεκτρολόγων Μηχανικών και Μηχανικών Υπολογιστών του Εθνικού Μετσόβιου Πολυτεχνείου. Η εργασία αυτή ερευνά την ανάπτυξη και ενσωμάτωση καινοτόμων εργαλείων στο πλαίσιο του ευρωπαϊκού προγράμματος SINFONICA[1], με στόχο την προώθηση συμπεριληπτικών και βιώσιμων λύσεων Συνεργατικής Συνδεδεμένης και Αυτοματοποιημένης Κινητικότητας (Cooperative, Connected, and Automated Mobility (CCAM))[2]. Βασικός στόχος ήταν η δημιουργία ενός συστήματος συστάσεων που να απευθύνεται σε ποικίλες ομάδες χρηστών, με έμφαση σε ευάλωτους πληθυσμούς. Μέσω συστηματικής συλλογής και ανάλυσης δεδομένων, αναγνωρίστηκαν κρίσιμα συμπεράσματα, τα οποία θα αξιοποιηθούν στα πλαίσια του έργου SINFONICA και σε πιθανές επεκτάσεις του. Στη συνέχεια, υλοποιήθηκε ένα Σύστημα που Βασίζεται σε Κανόνες (Rule-Based System) με το εργαλείο Drools[3], το οποίο συνδέει τα χαρακτηριστικά των χρηστών (π.χ., ηλικία, γεωγραφική θέση) με συγκεκριμένες συστάσεις βέλτιστων πρακτικών για την ανάπτυξη CCAM solutions. Οι κανόνες σχεδιάστηκαν με βάση τα δεδομένα που συλλέξαμε από τα ερωτηματολόγια. Για να βελτιωθεί η αλληλεπίδραση με τον τελικό χρήστη, εντάχθηκαν Μεγάλα Γλωσσικά Μοντέλα (Large Language Model) όπως το OpenAI[4], σε συνδυασμό με την πλατφόρμα Pinecone[5] για διαχείριση διανυσματικών βάσεων δεδομένων. Αυτό επέτρεψε τη δημιουργία φυσικής γλώσσας για εξατομικευμένες απαντήσεις σε ερωτήσεις που βασίζονται σε δεδομένα κειμένου που έχουν εισαχθεί στο σύστημα από τον χρήστη. Τα αποτελέσματα του έργου υπογραμμίζουν τη σημασία τεχνολογικών λύσεων στον τομέα των μεταφορών με κοινωνική ευαισθησία.

Λέξεις Κλειδιά

Λύσεις Συνεργατικής Συνδεδεμένης Αυτοματοποιημένης Κινητικότητας, Οπτικοποίηση Δεδομένων, Συστήματα Βασισμένα σε Κανόνες, Μεγάλα Γλωσσικά Μοντέλα

Abstract

This thesis was conducted as part of the undergraduate program of the School of Electrical and Computer Engineering at the National Technical University of Athens. The thesis investigates the development and integration of innovative tools within the European project SINFONICA [1], aiming to advance inclusive and sustainable Cooperative, Connected, and Automated Mobility (CCAM) solutions [2]. Our primary objective was to design a recommendation system targeting diverse user groups, with emphasis on vulnerable populations. Through systematic data collection and analysis, critical insights were derived. These insights will be leveraged within the SINFONICA project and its potential extensions. Subsequently, a Rule-Based System was implemented using the Drools framework [3], linking user attributes (e.g., age, geographic location) to specific best practices for CCAM solution development. The rules were designed based on questionnaire data. To improve end-user interaction, Large Language Models (LLMs) such as OpenAI [4] were integrated, combined with the Pinecone platform [5] for vector database management. This enabled natural language generation for personalized responses to user queries based on text data input into the system. The project's results underscore the importance of socially aware technological solutions in mobility.

Keywords

Cooperative, Connected, and Automated Mobility (CCAM), Data Visualization, Rule-Based Systems, Large Language Models (LLMs)

στην Αλεξάνδρα

Ευχαριστίες

Η παρούσα διπλωματική εργασία εκπονήθηκε στο πλαίσιο της ολοκλήρωσης του προπτυχιακού κύκλου σπουδών μου, στη Σχολή Ηλεκτολόγων Μηχανικών και Μηχανικών Υπολογιστών του Εθνικού Μετσόβιου Πολυτεχνείου.

Η ολοκλήρωση της διπλωματικής αυτής εργασίας θα ήταν αδύνατη χωρίς την πολύτιμη υποστήριξη και καθοδήγηση του κ. Β. ΤΣΟΥΓΙΑΝΝΗ και του κ. Γ. ΔΡΑΪΝΑΚΗ. Ακόμα θα ήθελα να ευχαριστήσω τον καθηγητή μου, κ. Παναγιώτη Τσανάκα για την στήριξή του και την δυνατότητα να εκπονήσω την παρούσα διπλωματική εργασία.

Το γεγονός ότι κατάφερα να σπουδάσω και να ολοκληρώσω τις σπουδές μου, το οφείλω αποκλειστικά και μόνο στην οικογένειά μου και ιδιαίτερα στους γονείς μου που με στήριζαν τόσο οικονομικά με τεράστιες θυσίες όσο και με την πίστη τους στις επιλογές και στις δυνατότητές μου. Θα τους είμαι πάντα ευγνώμων για το ποιός είμαι σήμερα και για το τι έχω καταφέρει.

Ακόμα οφείλω ένα τεράστιο ευχαριστώ στους φίλους και τους συμφοιτητές μου (ιδιαίτερα στον αδελφικό μου φίλο Α.Μ.), γιατί ήταν σταθερά δίπλα μου, να μου υπενθυμίζουν ότι όποιες δυσκολίες και αν εμφανίζονται στο δρόμο μου, πρέπει πάντα να τις αντιμετωπίζω και να μην τα βάζω ποτέ κάτω. Μέσα από την προσπάθεια να ολοκληρώσω τις προπτυχιακές μου σπουδές, κατάλαβα πως ως νέος επιστήμονας έχω την ευθύνη να αξιοποιήσω τη γνώση μου προς όφελος της κοινωνίας. Η επιστήμη δεν είναι απλώς ένα εργαλείο, αλλά μια πράξη αλληλεγγύης—μια δέσμευση να συνεργαστούμε για έναν δίκαιο κόσμο, όπου η τεχνολογία εξυπηρετεί τον άνθρωπο και τις κοινωνικές ανάγκες και όχι τα κέρδη κάποιων λίγων επιχειρηματικών ομίλων, όπως γίνεται σήμερα. Δυστυχώς, ενώ η τεχνολογία και η επιστήμη έχουν φτάσει σε ένα επίπεδο που θα μπορούσαν σήμερα να καλύπτονται ολόπλευρα όλες οι κοινωνικές ανάγκες, βλέπουμε να δολοφονούνται 57 συνάνθρωποί μας ακόμα και συμφοιτητές μας επειδή μπήκαν σε ένα τρένο που δεν πληρούσε ούτε ένα μέτρο ασφάλειας. Αυτή η υπόσχεση, η επιστήμη να υπηρετεί έναν καλύτερο κόσμο στο “μπόι των ανθρώπων”, πρέπει να συνοδεύει κάθε νέο επιστήμονα σήμερα.

Περιεχόμενα

Περίληψη	7
Abstract	9
Ευχαριστίες	13
Περιεχόμενα	16
Κατάλογος Εικόνων	18
Κατάλογος Πινάκων	19
Κατάλογος με Κώδικες	21
1 Εισαγωγή	23
1.1 Σκοπός του Έργου SINFONICA	24
1.2 Σκοπός της Διπλωματικής Εργασίας	25
1.3 Δομή Διπλωματικής Εργασίας	26
2 Βιβλιογραφική Ανασκόπηση (State of the Art - SotA)	29
2.1 Σχετικές Προσεγγίσεις για CCAM solutions	30
2.2 Καινοτομία του Έργου SINFONICA	31
2.3 Η συμβολή της διπλωματικής εργασίας	32
3 Συλλογή Δεδομένων	35
3.1 Ερωτηματολόγια	38
3.2 Οπτικοποίηση Δεδομένων	39
3.3 Βιβλιοθήκες Python	41
3.4 Κώδικας Υλοποίησης Οπτικοποίησης Δεδομένων	43
3.5 Αποτελέσματα	48
3.6 Συμπεράσματα και Use Cases	57

4	Συστήματα που βασίζονται σε κανόνες (Rule Based System)	59
4.1	Το εργαλείο Drools	60
4.2	Ενσωμάτωση στο Έργο	63
4.3	Αποτελέσματα	68
4.4	Συμπεράσματα	71
5	Μεγάλα Γλωσσικά Μοντέλα (Large Language Models)	73
5.1	Ενσωμάτωση LLM στο έργο μας	74
5.2	Το μοντέλο OpenAI	75
5.3	Το εργαλείο Pinecone	76
5.4	Αποτελέσματα	78
5.5	Συμπεράσματα	80
6	Συμπεράσματα	83
6.1	Η χρησιμότητα των CCAM solutions στον τομέα των μεταφορών	84
6.2	Οι λύσεις που προσφέρει το Recommendation Engine Tool	84
6.3	Περαιτέρω Έρευνα	85

Κατάλογος Εικόνων

1.1 Συνεργατική Συνδεδεμένη και Αυτοματοποιημένη Κινητικότητα (Cooperative, Connected and Automated Mobility)	23
1.2 Αρχιτεκτονική ΚΜΕ	25
2.1 State of the Art in Mobility	29
3.1 Συλλογή - Επεξεργασία και Οπτικοποίηση Δεδομένων	35
3.2 Οι τρεις κορυφαίοι τρόποι μεταφοράς που χρησιμοποιούνται για αμοιβόμενη εργασία	49
3.3 Προτεραιότητες με τις οποίες επιλέγονται οι τρόποι μεταφοράς	50
3.4 Βαθμός αίσθησης περιορισμού των πολιτών στα ταξίδια τους σε σχέση με την διαθεσιμότητα των M.M.M.	51
3.5 Βαθμός συμφωνίας με την φράση "Αγοράζω καινούρια προϊόντα πριν τους φίλους μου και τους γείτονες μου"	52
3.6 Βαθμός συμφωνίας με τη φράση "Με ενδιαφέρει λίγο έως καθόλου η νέα τεχνολογία"	53
3.7 Χρήση ή γνώση της παρουσία του CCAM στο σύστημα κινητικότητας .	54
3.8 Μέτρα ασφάλειας σε περίπτωση που δεν υπάρχει ανθρώπινο δυναμικό	55
3.9 Επίπεδο εκπαίδευσης ανά κατηγορία πολιτών	56
3.10 Παράδειγμα Περίπτωσης Χρήσης	58
4.1 Σύστημα που βασίζεται σε κανόνες	59
4.2 Η αρχιτεκτονική του εργαλείου Drools	62
4.3 Σελίδα εισόδου στην εφαρμογή που δημιουργήσαμε	63
4.4 Φόρμα συμπλήρωσης στοιχείων χρηστών της εφαρμογής	63
4.5 1η Έξοδος της εφαρμογής με μία τυχαία είσοδο	69
4.6 2η Έξοδος της εφαρμογής με μία τυχαία είσοδο	70
4.7 3η Έξοδος της εφαρμογής με μία τυχαία είσοδο	70
5.1 Μεγάλα Γλωσσικά Μοντέλα	73
5.2 ChatBox:Ερώτηση 1	79

5.3 ChatBox:Ερώτηση 2	79
5.4 ChatBox:Ερώτηση 3	79
5.5 ChatBox:Έλεγχος Απόδοσης Πρώτη Περίπτωση	80
5.6 ChatBox:Έλεγχος Απόδοσης Δεύτερη Περίπτωση	80
6.1 Συμπεράσματα	83

Κατάλογος Πινάκων

3.1 Πίνακας Περιγραφής των Δεδομένων	48
--	----

Κατάλογος με Κώδικες

3.1	Bar Chart	44
3.2	Pie Chart	46
3.3	Histogram	48
4.1	Kie session code	65
4.2	Drool's rules	66
5.1	OpenAI code	76
5.2	Pinecone Code	77

Κεφάλαιο 1

Εισαγωγή



Εικόνα 1.1: Συνεργατική Συνδεδεμένη και Αυτοματοποιημένη Κινητικότητα (Cooperative, Connected and Automated Mobility)

Η παρούσα διπλωματική εργασία εκπονήθηκε στα πλαίσια ολοκλήρωσης του προπτυχιακού κύκλου σπουδών της σχολής Ηλεκτρολόγων Μηχανικών και Μηχανικών Υπολογιστών του Εθνικού Μετσόβιου Πολυτεχνείου. Η διπλωματική εργασία βασίζεται πάνω στο ερευνητικό έργο SINFONICA, του οποίου συνεργάτης αποτελεί και το Ερευνητικό Πανεπιστημιακό Ινστιτούτο Συστημάτων Επικοινωνιών και Υπολογιστών (ΕΠΙΣΕΥ), που αποτελεί τον ερευνητικό βραχίονα της Σχολής ΗΜΜΥ. Η παρούσα διπλωματική εργασία έχει σκοπό να ασχοληθεί με κάποιες σημαντικές εργασίες του συγκεκριμένου έργου, να βοηθήσει στην ολοκλήρωση του αλλά και να εντοπίσει ζητήματα που θα μπορούσαν να ερευνηθούν περαιτέρω στο μέλλον.

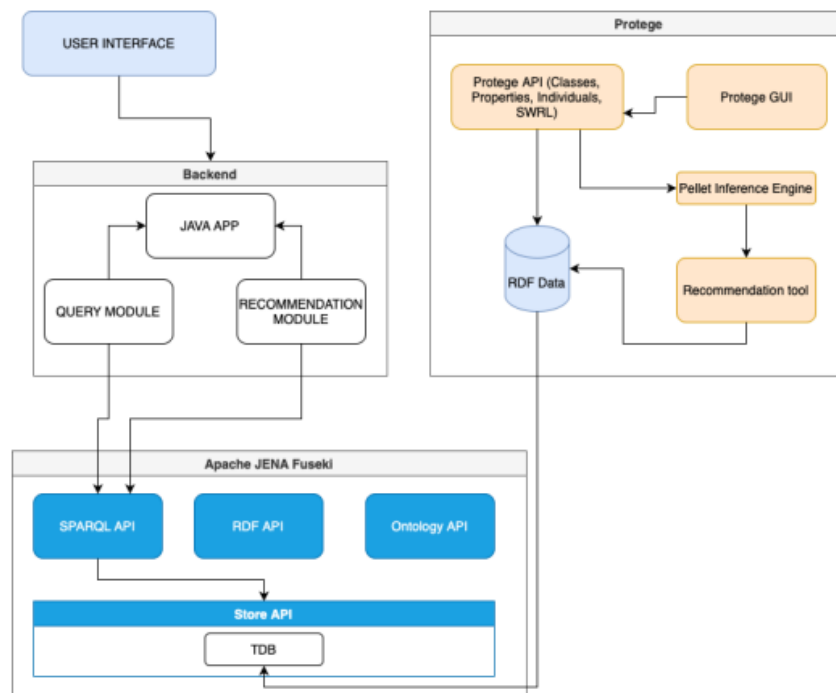
1.1 Σκοπός του Έργου SINFONICA

Το έργο SINFONICA [1], που χρηματοδοτείται από το πρόγραμμα Horizon Europe [6] της Ευρωπαϊκής Ένωσης, στοχεύει στη δημιουργία καινοτόμων μεθοδολογιών και εργαλείων για την προώθηση Συνεργατικής, Συνδεδεμένης και Αυτοματοποιημένης Κινητικότητας (Cooperative, Connected, and Automated Mobility (CCAM)) [2]. Ο κύριος στόχος είναι η ανάπτυξη συστημάτων που να εξασφαλίζουν συμπεριληπτικές και βιώσιμες λύσεις αστικής κινητικότητας, με έμφαση στις ανάγκες ευάλωτων ομάδων, όπως άτομα με αναπηρίες, ηλικιωμένοι και κάτοικοι αγροτικών περιοχών. Μέσω της συλλογής και δομημένης οργάνωσης δεδομένων από πολίτες τεσσάρων διαφορετικών πόλεων και χωρών, το έργο επιδιώκει να προσφέρει εργαλεία υποστήριξης αποφάσεων για σχεδιασμό και ανάπτυξη λύσεων CCAM που να ανταποκρίνονται στις πολυδιάστατες ανάγκες των χρηστών.

Η ανάγκη για τέτοιες λύσεις προκύπτει από την αυξανόμενη πολυπλοκότητα των αστικών μεταφορών, τις απαιτήσεις για μείωση των εκπομπών και την ανάγκη ενίσχυσης της πρόσβασης σε δίκαιες υπηρεσίες κινητικότητας [7]. Το SINFONICA δεν αναπτύσσει άμεσα CCAM συστήματα, αλλά δημιουργεί ένα πλαίσιο γνώσης και συστάσεων, βασισμένο σε οντολογίες και συστήματα συστάσεων, για να καθοδηγήσει φορείς στην υλοποίηση τέτοιων λύσεων. Βασικό εργαλείο αυτού του πλαισίου είναι ο Εξερευνητής Χαρτογράφησης Γνώσης (Knowledge Map Explorer - (KME)) [8], ο οποίος συνδυάζει semantic και rule-based συστήματα για την παροχή εξατομικευμένων συστάσεων.

Ο Knowledge Map Explorer (KME) είναι ένα εργαλείο που χρησιμοποιείται για την χαρτογράφηση, οπτικοποίηση και πλοήγηση γνώσης γύρω από ένα συγκεκριμένο αντικείμενο, όπως π.χ. η έξυπνη κινητικότητα (smart mobility) στο πλαίσιο του έργου μας. Στόχος του είναι να βοηθήσει στη δομημένη καταγραφή της υπάρχουσας γνώσης, στην αναγνώριση κενών πληροφόρησης και στη σύνδεση αναγκών των χρηστών με ενδεχόμενες λύσεις ή τεχνολογίες. Συχνά, τέτοια εργαλεία βασίζονται σε οντολογίες και χρησιμοποιούνται για υποστήριξη λήψης αποφάσεων ή για ανάλυση δεδομένων από διαφορετικές πηγές. Η αρχιτεκτονική του KME που υλοποιήθηκε στα πλαίσια του έργου Sinfonica φαίνεται στην Εικόνα 1.2. Η αρχιτεκτονική του συστήματος αποτελείται από δύο βασικά υποσυστήματα το Backend και το Protege [9]. Το Backend βασίζεται σε μια εφαρμογή Java που ενσωματώνει λειτουργικές μονάδες για διαχείριση ουρών ερωτημάτων, για παραγωγή συστάσεων, για ερωτήματα για διαχείριση δεδομένων, για προσαρμοσμένες λογικές και για αποθήκευση. Το Protege αφορά τη δημιουργία και διαχείριση οντολογιών με χρήση κλάσεων, ιδιοτήτων και ατόμων. Η αλληλεπίδραση μεταξύ των δύο υποσυστημάτων εξασφαλίζει τη συλλογή, επεξεργασία και ανάκτηση γνώσης, καθώς και την παροχή

ευφυνών συστάσεων με βάση την οντολογία και τα δεδομένα.



Εικόνα 1.2: Αρχιτεκτονική KME

1.2 Σκοπός της Διπλωματικής Εργασίας

Η παρούσα διπλωματική εργασία συνεισφέρει στο έργο SINFONICA μέσω της ανάπτυξης και ενσωμάτωσης βασικών στοιχείων στις εργασίες T3.6, T4.1 (Tasks T3.6, T4.1) και T4.2. Συγκεκριμένα, εστιάζει σε τρεις κύριους άξονες:

- Συλλογή και Οπτικοποίηση Δεδομένων:** Συλλογή αποτελεσμάτων από ερωτηματολόγια για τη συλλογή δεδομένων από ευάλωτες ομάδες, ανάλυση και οπτικοποίησή τους με χρήση βιβλιοθηκών της γλώσσας προγραμματισμού Python (Pandas, Matplotlib, Seaborn) [10]. Αυτό υποστηρίζει η εργασία T3.6, όπου τα δεδομένα χρησιμοποιούνται για τον εμπλουτισμό της οντολογίας και για να διαμορφωθούν ιστορίες χρηστών (user stories).
- Συστήματα που Βασίζονται σε Κανόνες (Rule-Based Systems):** Εφαρμογή του εργαλείου Drools [3] για τη δημιουργία κανόνων που συνδέουν τα χαρακτηριστικά των χρηστών με συστάσεις (CCAM). Αυτό ενισχύει την εργασία T4.2 (Task T4.2), όπου ο KME αξιοποιεί ένα εργαλείο με κανόνες (rule engine) για την παραγωγή συστάσεων βασισμένων στην οντολογία.

- **Ενσωμάτωση Μεγάλων Γλωσσικών Μοντέλων (LLMs):** Χρήση του μοντέλου OpenAI[4] και της πλατφόρμας Pinecone[5] για τη βελτίωση της αλληλεπίδρασης με τον χρήστη, προσφέροντας φυσική γλώσσα και εξατομικευμένες απαντήσεις. Αυτό επεκτείνει τις δυνατότητες του ΚΜΕ, ευθυγραμμίζοντάς το με σύγχρονες τεχνολογικές τάσεις.

Η εργασία δεν υποστηρίζει απλώς τη θεωρητική βάση του έργου, αλλά παρέχει πρακτικές υλοποιήσεις που επηρεάζουν άμεσα τη δομή και τη λειτουργία του ΚΜΕ. Για παράδειγμα, η οπτικοποίηση δεδομένων βοήθησε στην αναγνώριση κρίσιμων προτύπων για ευάλωτες ομάδες, ενώ οι κανόνες Drools κατέστησαν δυνατή την αυτοματοποίηση συστάσεων με βάση προκαθορισμένα κριτήρια.

1.3 Δομή Διπλωματικής Εργασίας

Η διπλωματική οργανώνεται σε τέσσερα κύρια κεφάλαια :

- **Κεφάλαιο 2: Βιβλιογραφική Ανασκόπηση (State of the Art - SotA).** Γίνεται μία αναφορά σε παρόμοιες προσεγγίσεις - λύσεις από άλλες ερευνητικές εργασίες σε σχέση με το έργο μας. Αντίστοιχα αναφέρονται τόσο επιστημονικά αποτελέσματα που αξιοποιήσαμε όσο και η συνεισφορά της δικής μας δουλειάς στην επιστημονική κοινότητα.
- **Κεφάλαιο 3: Συλλογή και Οπτικοποίηση Δεδομένων.** Παρουσιάζεται η μεθοδολογία συλλογής δεδομένων μέσω ερωτηματολογίων, η ανάλυσή τους με την βοήθεια εργαλείων σε περιβάλλον Python και η δημιουργία οπτικοποιήσεων που υπογραμμίζουν κρίσιμες τάσεις. Αποτελέσματα και περιπτώσεις χρήσης (use cases) καταγράφονται για να δείξουν τη σύνδεση με ιστορίες χρηστών (user stories) του έργου SINFONICA.
- **Κεφάλαιο 4: Συστήματα που Βασίζονται σε Κανόνες.** Περιγράφεται η εφαρμογή του Drools για τη δημιουργία κανόνων που συνδέουν δεδομένα χρηστών με συστάσεις, καθώς και η πιθανή ενσωμάτωσή τους στην αρχιτεκτονική του ΚΜΕ, ανοίγοντας καινούριους τομείς για περαιτέρω μελλοντική έρευνα.
- **Κεφάλαιο 5: Μεγάλα Γλωσσικά Μοντέλα.** Εξετάζεται η χρήση του εργαλείου OpenAI και του Pinecone για τη βελτίωση της αλληλεπίδρασης με τον χρήστη, με έμφαση στην επεξεργασία φυσικής γλώσσας. Δημιουργείται ένα πλαίσιο συνομιλίας (ChatBox) όπου ο χρήστης μπορεί να εισάγει διάφορες ερωτήσεις πάνω στα δεδομένα κειμένου που υπάρχουν στην βάση του εργαλείου μας και να πάρει τις κατάλληλες απαντήσεις. Σημαντικό

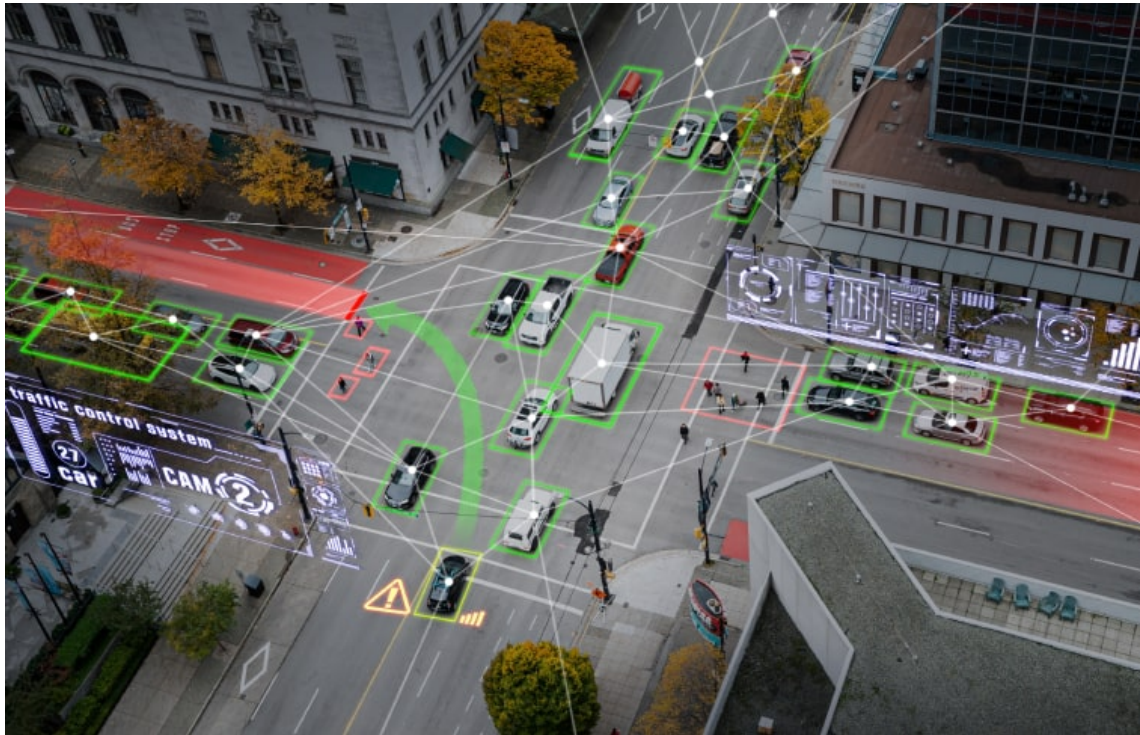
προτέρημα του εργαλείου είναι ότι με τις κατάλληλες προσαρμογές στα δεδομένα κειμένου και στην παροχή πόρων, μπορεί να αξιοποιηθεί σε αντίστοιχες εφαρμογές που χρησιμοποιούν όμως άλλα δεδομένα. Για παράδειγμα εισάγοντας επιστημονικά κείμενα που αφορούν τα τεχνολογικά επιτεύγματα στον τομέα των μεταφορών, θα μπορούσε να χρησιμοποιηθεί το εργαλείο που υλοποιήσαμε ώστε να μπορεί ο χρήστης της εφαρμογής να κάνει ερωτήσεις και να λαμβάνει άμεσα απαντήσεις για την επιστημονική έρευνα στο συγκεκριμένο κομμάτι και αναφορές πάνω σε επιστημονικά άρθρα.

- **Κεφάλαιο 6: Συμπεράσματα και Μελλοντικές Επεκτάσεις.** Συντάσσεται μια κριτική αξιολόγηση των αποτελεσμάτων, προτείνοντας τρόπους βελτίωσης και μελλοντικές εφαρμογές - επεκτάσεις.

Κάθε κεφάλαιο συνδέεται άμεσα με τις εργασίες (tasks) του έργου SINFONICA, κάνοντας φανερή τη συμβολή της εργασίας τόσο στην επίτευξη των στόχων του έργου, όσο και σε ερευνητικό επίπεδο. Η παρούσα διπλωματική εργασία εξασφαλίζει μια συνεκτική περιγραφή τόσο της θεωρητικής βάσης όσο και των πρακτικών υλοποιήσεων, προσφέροντας μια ολοκληρωμένη προοπτική για την αξία της στο πλαίσιο του έργου SINFONICA και της ακαδημαϊκής κοινότητας.

Κεφάλαιο 2

Βιβλιογραφική Ανασκόπηση (State of the Art - SotA)



Εικόνα 2.1: State of the Art in Mobility

Η ανάπτυξη συστημάτων συστάσεων με οντολογίες και κανόνες δεν είναι καινοφανής, αλλά το SINFONICA συμβάλει καθοριστικά στην εξέλιξη και την βελτίωση τους. Για παράδειγμα, παρόμοια έργα, όπως το SUaaVE[11] και το SHOW[12], ερευνούν την ανάπτυξη αυτόνομων οχημάτων. Το SINFONICA μπορεί να μην έχει ως αντικείμενο την ανάπτυξη τεχνολογικών εργαλείων που θα μπουν στην ζωή των πολιτών και θα

βελτιώσουν άμεσα τις μετακινήσεις σε αντίθεση με τα έργα που προαναφέραμε, όμως ενσωματώνει ποιοτικές μεθόδους και δεδομένα, αξιοποιεί και εξελίσει τεχνολογικά εργαλεία για να καθοδηγήσει πολιτικές και τεχνολογικές αποφάσεις και να βελτιώσει έμμεσα τον τομέα των μεταφορών και της μετακίνησης. Δείχνει δηλαδή τον δρόμο στον οποίο πρέπει να βαδίσει η επιστημονική κοινότητα, η έρευνα και η πολιτεία το επόμενο διάστημα.

2.1 Σχετικές Προσεγγίσεις για CCAM solutions

Στη συνέχεια παρατίθενται οι κυριότερες κατευθύνσεις, πρωτοβουλίες και τεχνολογίες που έχουν αναπτυχθεί έως σήμερα στον χώρο των CCAM solutions, με έμφαση σε μεθοδολογίες καινοτομίας, συμμετοχικές πρακτικές και εργαλεία υποστήριξης γνώσης και προσομοίωσης.

Μία τέτοια πρώτη τεχνολογική προσέγγιση που έχει αναπτυχθεί στον τομέα των λύσεων CCAM, είναι η προσαρμοσμένη μεθοδολογία GUEST[13]. Η GUEST είναι μια μεθοδολογία καινοτομίας που καθοδηγεί συστηματικά τον κύκλο ζωής ενός έργου, από την προετοιμασία και τον σχεδιασμό έως την υλοποίηση και την αξιολόγηση, με σαφή βήματα και μετρήσιμους δείκτες. Σε αντίθεση με δημοφιλείς προσεγγίσεις όπως Lean Startup[14], WCM[15] και AGILE[16], που απαιτούν υψηλές τεχνικές δεξιότητες και εστιάζουν κυρίως στην τεχνολογία, η GUEST προσφέρει ένα ενιαίο πλαίσιο με ξεκάθαρα βήματα και δείκτες μέτρησης. Ωστόσο, η τρέχουσα υιοθέτησή της παραμένει περιορισμένη, καθώς πολλά έργα εξακολουθούν να χρησιμοποιούν διάσπαρτα εργαλεία, δυσχεραίνοντας την αναπαραγωγή, τη σύγκριση και την αξιολόγηση των αποτελεσμάτων.[1]

Ακόμα, το Knowledge Map Explorer αποτελεί μία από τις πρώτες προσπάθειες χαρτογράφησης γνώσης ειδικά για τις συμπεριφορικές πτυχές των CCAM solutions. Αντίστοιχα έργα όπως TinnGO16[17], SUMP[18] και DIGNITY[19] έχουν παράσχει απλές χαρτογραφήσεις υποδομών, το Knowledge Map Explorer όμως στοχεύει στον εμπλουτισμό με ποιοτικά δεδομένα και οντολογικές αναπαραστάσεις. Πέρα από όλα αυτά, η ευρύτερη υιοθέτησή του είναι ακόμη περιορισμένη, γεγονός που αναδεικνύει το κενό στην εφαρμογή μοντέλων χαρτογράφησης γνώσης στη έξυπνη κινητικότητα smart mobility.

Αντίστοιχα, η ένταξη ευάλωτων χρηστών σε διαδικασίες συν-δημιουργίας(δηλαδή σε συνεργατικές φάσεις όπου οι ίδιοι οι χρήστες συμμετέχουν ενεργά στο σχεδιασμό, την ανάπτυξη και την αξιολόγηση υπηρεσιών ή προϊόντων), αποτελεί κρίσιμο ζήτημα για την κοινωνικά δίκαιη ανάπτυξη CCAM. Παρά τη θεωρητική αναγνώριση της ανάγκης, λίγες μελέτες έχουν έως σήμερα πραγματικά ενσωματώσει άτομα με

αναπηρίες, ηλικιωμένους ή κοινωνικά ευάλωτες ομάδες σε καίριες φάσεις του σχεδιασμού. Οι περισσότερες πρωτοβουλίες επικεντρώνονται στην τεχνολογική καινοτομία, παραμελώντας τη συμβολή των κοινωνικών επιστημών και των ανθρωπιστικών σπουδών (Social Sciences and Humanities - SSH) στην κατανόηση των εμπειριών και των περιορισμών αυτών των χρηστών.

Οι συνεργατικές διαδικασίες σχεδιασμού με πολίτες και ευάλωτες ομάδες αποτελούν τη θεμελιώδη προϋπόθεση για συμπερίληψη και αποδοχή των CCAM solutions. Ωστόσο, η πράξη έχει δείξει ότι η δημιουργία ουσιαστικής συνεργασίας αντιμετωπίζει εμπόδια εμπιστοσύνης, γλωσσικά και πολιτισμικά εμπόδια, καθώς και δυσκολίες πρόσβασης σε ψηφιακά μέσα. Οι προσπάθειες που έχουν γίνει μέχρι στιγμής απαιτούν πιο δομημένες προσεγγίσεις και προσαρμοσμένα εργαλεία διευκόλυνσης της επικοινωνίας.

Τέλος, τα εργαλεία προσομοίωσης για την κλιμάκωση μεταφορικών μέτρων που έχουν χρησιμοποιηθεί σε αντίστοιχα έργα, συνήθως εστιάζουν σε κλασικούς δείκτες κυκλοφορίας και εκπομπών. Σπάνια περιλαμβάνουν εκ των προτέρων τις ειδικές ανάγκες ευάλωτων χρηστών ή στόχους προσβασιμότητας. Η έλλειψη προσομοιώσεων που λαμβάνουν υπόψη ποικίλες μορφές ασφάλειας, άνεσης και συμμετοχής καθιστά αναγκαία την ανάπτυξη νέων, πιο ευέλικτων πλατφορμών προσομοίωσης, ικανών να ενσωματώσουν κοινωνικά και ανθρωποκεντρικά κριτήρια.

2.2 Καινοτομία του Έργου SINFONICA

Παρακάτω αναλύονται οι βασικές καινοτομίες που εισάγει το SINFONICA με στόχο την υποστήριξη συμπεριληπτικών, ανθρωποκεντρικών και μετρήσιμων CCAM λύσεων.

Πρώτη καινοτομία: Το SINFONICA υιοθετεί τη μεθοδολογία GUEST, μια απλή επιχειρησιακή προσέγγιση που καθοδηγεί ολόκληρη τη διαδικασία ανάπτυξης καινοτομιών. Η GUEST σχεδιάστηκε ώστε να είναι άμεσα κατανοητή ακόμη και από χρήστες με περιορισμένη εμπειρία στην καινοτομία, μειώνοντας τη μαθησιακή καμπύλη και διευκολύνοντας τη συνεργασία μεταξύ τεχνικών και μη ειδικών. Πέραν των τρεχόντων μεθοδολογιών (Lean Startup, WCM, AGILE), η GUEST εξειδικεύεται ώστε να αναλύει και να μετρά τις επιπτώσεις των CCAM λύσεων σε επιμέρους ομάδες χρηστών και ενδιαφερόμενους φορείς, με ιδιαίτερη έμφαση στην ένταξη ατόμων με ειδικές ανάγκες στο σχεδιασμό πολιτικών και υπηρεσιών.

Δεύτερη καινοτομία: Μέσω συστηματικής βιβλιογραφικής ανάλυσης, το SINFONICA διερεύνησε την υπάρχουσα γνώση για τη συμμετοχή ευάλωτων

ομάδων σε διαδικασίες συν-δημιουργίας. Ενώ λίγες πρωτοβουλίες συνδυάζουν τεχνολογία και κοινωνικές επιστήμες (SSH), εδώ εξασφαλίζεται η ενεργή εμπλοκή ψυχολόγων, κοινωνιολόγων και άλλων ειδικών, οι οποίοι επεξεργάζονται τα δεδομένα από ανθρωπιστική σκοπιά. Αυτή η διεπιστημονική προσέγγιση ξεπερνά το τρέχον επίπεδο γνώσης, προσφέροντας νέα εργαλεία και μεθόδους ανάλυσης των κοινωνικών επιπτώσεων των CCAM λύσεων.

Τρίτη καινοτομία: Κεντρικό στοιχείο του έργου είναι η δημιουργία μιας γνωσιακής βάσης με οντολογίες, η οποία χαρτογραφεί τις ανάγκες, τις προτεραιότητες και τα σενάρια χρήσης των CCAM λύσεων. Σε αυτό το πλαίσιο, αναπτύχθηκε το εργαλείο Knowledge Map Explorer, το οποίο καταγράφει ποσοτικά και ποιοτικά δεδομένα, παρέχοντας συμφραζόμενες συστάσεις για συμπεριφορικές πτυχές κινητικότητας. Η δυνατότητα εμπλουτισμού με κανόνες και LLM-based ερωταποκρίσεις καθιστά το εργαλείο μοναδικό στο state of the art.

Τέταρτη καινοτομία: Τέλος, το SINFONICA επέκτεινε υπάρχουσα πλατφόρμα προσομοίωσης για να περιλαμβάνει στόχους προσβασιμότητας και περιορισμούς ευάλωτων χρηστών. Πέραν των κλασικών δεικτών κυκλοφορίας, ενσωματώθηκαν αλγόριθμοι διαχείρισης στόλου οχημάτων που λαμβάνουν υπόψη τεχνικούς, φυσικούς και κοινωνικούς περιορισμούς, προσφέροντας βελτιστοποιημένες λύσεις κλιμάκωσης με έμφαση στην ισότητα και την ανθρώπινη εμπειρία.

2.3 Η συμβολή της διπλωματικής εργασίας

Η παρούσα διπλωματική εργασία ενισχύει καθοριστικά το SINFONICA με μια ολοκληρωμένη υλοποίηση που ξεκινά από τη συστηματική συλλογή και ανάλυση δεδομένων και φτάνει έως την παροχή έξυπνων συστάσεων σε φυσική γλώσσα. Αρχικά, τα ευρήματα από τα ερωτηματολόγια καθώς και οι οπτικοποιήσεις που παράχθηκαν με Python, Matplotlib και Seaborn αποτύπωσαν με σαφήνεια τις κρίσιμες τάσεις στις ανάγκες των χρηστών, επιβεβαιώνοντας και επεκτείνοντας τα user stories του έργου. Με τη δημιουργία συγκεκριμένων use cases, η ανάλυση δεδομένων μετατράπηκε σε πρακτικά σενάρια χρήσης, βελτιώνοντας την ικανότητα των φορέων να σχεδιάζουν στοχευμένες CCAM λύσεις.

Στη συνέχεια, η ενσωμάτωση ενός rule-based μηχανισμού με το Drools έφερε μια άμεση εφαρμογή των κανόνων που ορίζει το Knowledge Map Explorer. Με την κατανοητή διατύπωση κανόνων if-else και την απρόσκοπτη ενοποίηση στο Java Maven περιβάλλον, το εργαλείο πλέον αντιστοιχεί άμεσα κάθε συνδυασμό χώρας, τύπου χρήστη, περιοχής και εισοδήματος σε εξατομικευμένες συστάσεις,

ανοίγοντας παράλληλα το πεδίο για αυτόματη ενημέρωση και διεύρυνση του συνόλου κανόνων.

Τέλος, η ανάπτυξη ενός διαδραστικού ChatBox με χρήση του OpenAI API και της υποδομής Pinecone ενίσχυσε την εμπειρία του τελικού χρήστη. Μέσω Retrieval-Augmented Generation, το σύστημα απαντά σε ερωτήσεις φυσικής γλώσσας αξιοποιώντας τα αποθηκευμένα έγγραφα, παρέχοντας συνεπείς, συμφραζόμενες απαντήσεις. Αυτή η διπλή προσέγγιση — data-driven visualization, rule-based logic και LLM-enhanced dialogue — δημιουργεί ένα ολοκληρωμένο εργαλείο που καλύπτει ολόκληρο τον κύκλο από τη συγκέντρωση δεδομένων έως τη δυναμική υποστήριξη αποφάσεων σε πραγματικό χρόνο.

Κεφάλαιο 3

Συλλογή Δεδομένων



Εικόνα 3.1: Συλλογή - Επεξεργασία και Οπτικοποίηση Δεδομένων

Στον σημερινό κόσμο της επιστήμης τα δεδομένα παίζουν τον πιο σημαντικό ρόλο. Ποτέ στο παρελθόν δεν έχουν υπάρξει τόσα πολλά δεδομένα όσο σήμερα. Ζούμε σε μια εποχή όπου τα δεδομένα έχουν γίνει το απαραίτητο καύσιμο που οδηγεί την ανάπτυξη των εταιρειών και ολόκληρων βιομηχανιών. Καθημερινά, παράγονται περίπου 463 exabyte (EB) δεδομένων σε παγκόσμια κλίμακα, αριθμός που αναμένεται να αυξηθεί περαιτέρω μέχρι το 2025, καθιστώντας τον όγκο των δεδομένων σαν το «καύσιμο» της ψηφιακής οικονομίας [20]. Παράλληλα, στις αρχές του 2025, 5,56 δισεκατομμύρια άνθρωποι – δηλαδή το 67,9% του παγκόσμιου πληθυσμού – χρησιμοποιούν το διαδίκτυο σε καθημερινή βάση, με 136 εκατομμύρια νέους χρήστες να προστίθενται μόνο μέσα στο 2024 [21]. Επιπλέον, οι ταυτότητες ενεργών χρηστών κοινωνικών μέσων έχουν ξεπεράσει τα 5 δισεκατομμύρια (62,3% του πληθυσμού), σημειώνοντας ετήσια αύξηση 5,6% (266

εκατ. νέοι λογαριασμοί)[21].

Σήμερα, οι οργανισμοί που μπορούν να συλλέγουν, να αναλύουν και να αξιοποιούν αποτελεσματικά τα δεδομένα αποκτούν ανταγωνιστικό πλεονέκτημα, επιτρέποντάς τους να λαμβάνουν ενημερωμένες επιχειρηματικές αποφάσεις και να βελτιώνουν τις λειτουργίες τους. Αντίστοιχα, το 60% των εργαζομένων αναμένεται να χρησιμοποιεί τεχνολογίες τεχνητής νοημοσύνης (Artificial Intelligence - AI) για την εκτέλεση καθημερινών εργασιών μέσα στο 2024, ενισχύοντας την παραγωγικότητα και την καινοτομία στην εργασία [22]. Τέλος, η κοινότητα της εκπαίδευσης δεδομένων έχει ξεπεράσει τα 12 εκατομμύρια μαθήματα αλληλεπίδρασης, με το 80% των εταιρειών του Fortune 1000 να εμπιστεύονται πλατφόρμες εκπαίδευσης δεδομένων για την ανάπτυξη δεξιοτήτων του προσωπικού τους[22]. Μπορεί λοιπόν να ειπωθεί ότι τα δεδομένα είναι ο χρυσός του 21ου αιώνα και παρά την αυξανόμενη ποσότητα που είναι διαθέσιμη κάθε χρόνο, γίνονται όλο και πιο πολύτιμα.

Για να αξιοποιηθεί πλήρως το δυναμικό των δεδομένων, είναι απαραίτητο να εφαρμόζονται κατάλληλες μέθοδοι και εργαλεία ανάλυσης. Ποικίλες αναλυτικές τεχνικές μπορούν να βοηθήσουν τις επιχειρήσεις, τους διάφορους οργανισμούς και όποιον ασχολείται με την ανάλυση δεδομένων, να εξάγουν πολύτιμες πληροφορίες από τα δεδομένα τους.

Μία βασική τεχνική είναι η περιγραφική ανάλυση, η οποία συνοψίζει και απεικονίζει τα δεδομένα, καθιστώντας τα χαρακτηριστικά και τις τάσεις τους πιο κατανοητά. Η οπτικοποίηση δεδομένων, όπως η δημιουργία γραφημάτων και θερμικών χαρτών (heatmaps), μπορεί να βοηθήσει στην αναγνώριση προτύπων και σχέσεων.

Μια άλλη δημοφιλής τεχνική είναι η ανάλυση τάσεων, η οποία παρακολουθεί αλλαγές στο χρόνο και προβλέπει μελλοντικές συμπεριφορές με βάση ιστορικά δεδομένα. Οι τεχνικές αυτές είναι ιδιαίτερα χρήσιμες για προβλέψεις ζήτησης, σχεδιασμό αποθεμάτων και εντοπισμό εποχικών προτύπων πωλήσεων.

Η ανάλυση δεδομένων δεν πρέπει να θεωρείται ως μια εφάπαξ εργασία αλλά ως συνεχής διαδικασία. Καθώς προκύπτουν νέα δεδομένα, πρέπει να ενημερώνονται οι αναλύσεις και να προσαρμόζονται οι στρατηγικές. Εξίσου σημαντικό είναι να διατηρούνται ενημερωμένα τα αναλυτικά μοντέλα, ώστε να εξελίσσονται και να βελτιώνονται σε ακρίβεια. Οι εταιρείες που αναπτύσσουν κουλτούρα βασισμένη στα δεδομένα και μαθαίνουν συνεχώς από αυτά, αποκτούν ανταγωνιστικό πλεονέκτημα.

Με την πρόοδο της Τεχνητής Νοημοσύνης (Artificial Intelligence - AI) και της Μηχανικής Μάθησης (Machine Learning - ML), τα δεδομένα γίνονται κρίσιμος παράγοντας για την υλοποίηση προηγμένων λύσεων. Χωρίς κατάλληλα δεδομένα, η

εφαρμογή AI λύσεων είναι περιορισμένη και δεν μπορεί να αξιοποιηθεί πλήρως το δυναμικό τους.

Οι αλγόριθμοι AI και ML βασίζονται στα δεδομένα για να εκπαιδεύσουν μοντέλα, τα οποία στη συνέχεια μπορούν να λαμβάνουν αποφάσεις ή να προβλέπουν αποτελέσματα. Όσο περισσότερα ποιοτικά δεδομένα υπάρχουν, τόσο πιο ακριβή και αποτελεσματικά θα είναι τα αποτελέσματα των λύσεων αυτών.

Πριν από την υλοποίηση λύσεων AI, είναι απαραίτητο να συγκεντρωθούν και να προετοιμαστούν σωστά τα δεδομένα. Αυτό περιλαμβάνει τη συλλογή από σχετικούς τομείς, τον καθαρισμό τους από άχρηστες ή ελλιπείς πληροφορίες, αφαίρεση διπλοτύπων, συμπλήρωση κενών ώστε να μπορούν να αξιοποιηθούν από τα διάφορα συστήματα.

Η σωστή επεξεργασία δεδομένων είναι θεμελιώδης πρακτική για την ακρίβεια και αξιοπιστία των αποτελεσμάτων. Για αυτό είναι απαραίτητα τα εργαλεία και οι διαδικασίες που υποστηρίζουν την αποτελεσματική διαχείριση δεδομένων.

Συμπερασματικά, τα δεδομένα αποτελούν πλέον έναν απαραίτητο πόρο για τον σύγχρονο κόσμο. Επιτρέπουν τη λήψη τεκμηριωμένων αποφάσεων, τη βελτιστοποίηση διαδικασιών και τη βαθύτερη κατανόηση των αναγκών των χρηστών.

Για να αξιοποιήσουμε αυτό το δυναμικό, πρέπει να εφαρμόσουμε κατάλληλες μεθόδους συλλογής, επεξεργασίας και ανάλυσης δεδομένων, επενδύοντας σε συστήματα AI και ML. Παράλληλα, πρέπει να δώσουμε προσοχή σε προκλήσεις όπως η ασφάλεια, η ποιότητα και η προστασία της ιδιωτικότητας.

Ειδικότερα στην δική μας εργασία που αφορά CCAM solutions, τα δεδομένα παίζουν πρωτεύοντα ρόλο. Με αξιόπιστα δεδομένα μπορούμε να εξασφαλίσουμε ότι οι συστάσεις που θα δίνει το εργαλείο μας θα είναι σωστές και έγκυρες και θα αποσκοπούν στο να λύνουν τα ζητήματα που αντιμετωπίζουν οι χρήστες οχημάτων με πρακτικές τεχνητής νοημοσύνης και αυτοματοποίησης.

Το πρώτο βήμα που χρειάζεται να εξασφαλίσουμε λοιπόν, για την δημιουργία του εργαλείου μηχανής σύστασης (Recommendation Engine) είναι η συλλογή και σωστή οπτικοποίηση δεδομένων. Για να γίνει πράξη το παραπάνω, δημιουργήθηκε ένα ερωτηματολόγιο το οποίο συμπληρώθηκε από διάφορες ηλικιακές ομάδες σε τέσσερις διαφορετικές χώρες. Πιο συγκεκριμένα τα δεδομένα προήλθαν από 4 περιοχές ενδιαφέροντος (Τρίκαλα, Αμβούργο, Ολλανδία, Αγγλία) σε τρεις φάσεις (round1, round2, round3).

Για την καλύτερη διαχείριση των δεδομένων που μας δόθηκαν, οι απαντήσεις

συγκεντρώθηκαν, ενοποιήθηκαν-ομαδοποιήθηκαν και τις διαχειριστήκαμε-επεξεργαστήκαμε μέσω υπολογιστικών εργαλείων όπως το Excel.

3.1 Ερωτηματολόγια

Τα ερωτηματολόγια που χρησιμοποιήσαμε χωρίζονται σε επτά ενότητες. Παρακάτω, παρατίθενται συνοπτικά ανά κατηγορία ερωτήσεων, το περιεχόμενο τους, ώστε να γίνουν καλύτερα κατανοητά τα δεδομένα που θέλουμε να συλλέξουμε.

- Η πρώτη κατηγορία ερωτήσεων έχει στόχο να εντοπίσει ποιά είναι το κύριο μεταφορικό μέσο των πολιτών για διάφορες καθημερινές δραστηριότητες και να εντοπίσει προσεγγιστικά την απόσταση αυτών των δραστηριοτήτων από το σπίτι του κάθε πολίτη.
- Η δεύτερη κατηγορία ερωτήσεων αποσκοπούν στο να εντοπίσουν τα κριτήρια με τα οποία οι πολίτες επιλέγουν τον τρόπο με τον οποίο θα μετακινηθούν.
- Η τρίτη κατηγορία ερωτήσεων προσπαθεί να εντοπίσει τις ελλείψεις που υπάρχουν σε διάφορους κρίσιμους τομείς στην μετακίνηση όπως η ασφάλεια, το κόστος μετακίνησης, η διαθεσιμότητα στα ΜΜΜ, ο σχεδιασμός κάποιου μακρινού ταξιδιού.
- Στην τέταρτη κατηγορία ερωτήσεων, γίνονται ερωτήσεις που αφορούν την επαφή και την σχέση των πολιτών με την τεχνολογία, τόσο με την πρόσβαση τους σε αυτή, όσο και με την χρήση της για θέματα μετακίνησης(π.χ. αγορά εισιτηρίου online κ.α.).
- Η πέμπτη κατηγορία ερωτήσεων, αφορά την γνώση και την επαφή των πολιτών με τα CCAM solutions.
- Η έκτη κατηγορία ερωτήσεων, προσπαθεί να εντοπίσει τα συναισθήματα και τις σκέψεις των πολιτών γύρω από την χρήση της τεχνολογίας στις μετακινήσεις και την χρήση αυτόματων οχημάτων. Ακόμα, να εντοπίσει ποιά οι πολίτες πιστεύουν ότι είναι τα πλεονεκτήματα και τα μειονεκτήματα από την χρήση της τεχνολογίας στην μετακίνηση.
- Η έβδομη και τελευταία κατηγορία ερωτήσεων αφορά τα στοιχεία των πολιτών που συμπλήρωσαν τα ερωτηματολόγια (ηλικία, φύλο, οικονομική κατάσταση, εκπαίδευση, εθνικότητα κ.α.)

Οι παραπάνω ερωτήσεις σε συνδυασμό με το γεγονός ότι γίνονται σε τέσσερις διαφορετικές χώρες και πόλεις (Τρίκαλα-Ελλάδα, Noord-Brabant-Ολλανδία, West Midlands-Ηνωμένο Βασίλειο, Hamburg-Γερμανία), βοηθούν καθοριστικά στο να

εντοπίσουμε με όσο το δυνατόν μεγαλύτερη ακρίβεια τις επιλογές των πολιτών στην μετακίνηση, τα προβλήματα που αντιμετωπίζουν, την σχέση τους με την τεχνολογία και την άποψη τους στο ενδεχόμενο να δωθούν λύσεις σε αυτά τα προβλήματα μέσω της τεχνολογίας, της τεχνητής νοημοσύνης, του Διαδικτύου των Πραγμάτων Internet of Things, της αυτοματοποίησης των οχημάτων κ.α. Στην παρούσα διπλωματική εργασία, θα αξιοποιήσουμε τα παραπάνω δεδομένα που μας δόθηκαν, ώστε να δημιουργήσουμε συστάσεις που θα ανταποκρίνονται στην απαιτήσεις των πολιτών και θα δίνουν αξιόπιστες λύσεις στους πολίτες αλλά και σε εταιρίες ή κρατικές αρχές που μπορούν με τις ενέργειες τους να ανοίξουν τον δρόμο για ένα καλύτερο μέλλον στον τομέα της μετακίνησης των πολιτών.

3.2 Οπτικοποίηση Δεδομένων

Η οπτικοποίηση δεδομένων περιλαμβάνει την παρουσίαση δεδομένων σε γραφική ή εικονογραφική μορφή, καθιστώντας την πληροφορία ευκολότερη στην κατανόηση. Βοηθά στην επεξήγηση γεγονότων και στον καθορισμό δράσεων. Μπορεί να ωφελήσει κάθε τομέα που απαιτεί καινοτόμους τρόπους παρουσίασης μεγάλων και πολύπλοκων πληροφοριών[23]. Αυτό μπορεί να γίνει με την παρουσίαση των δεδομένων σε διάφορα γραφήματα.

Υπάρχει λοιπόν, η ανάγκη εμφάνισης τεράστιων ποσοτήτων δεδομένων με τρόπο εύκολα προσβάσιμο και κατανοητό. Οι οργανισμοί παράγουν καθημερινά δεδομένα, με αποτέλεσμα τη δραματική αύξηση των διαθέσιμων δεδομένων στο Διαδίκτυο. Η κατανόηση αυτών των δεδομένων είναι δύσκολη για τους χρήστες.

Η ικανότητα οπτικοποίησης δεδομένων είναι κρίσιμη για την επιστημονική έρευνα. Οι υπολογιστές σήμερα μπορούν να επεξεργάζονται μεγάλες ποσότητες δεδομένων. Η οπτικοποίηση δεδομένων αφορά τον σχεδιασμό, την ανάπτυξη και την εφαρμογή γραφικών αναπαραστάσεων που δημιουργούνται από υπολογιστή. Παρέχει αποτελεσματική αναπαράσταση δεδομένων από διάφορες πηγές, επιτρέποντας σε αυτούς που παίρνουν αποφάσεις να βλέπουν την ανάλυση σε οπτική μορφή και να εξάγουν συμπεράσματα[23]. Τους βοηθά να ανακαλύπτουν πρότυπα, να κατανοούν την πληροφορία και να σχηματίζουν άποψη.

Σε αντίθεση με την στατική απεικόνιση, η διαδραστική επιτρέπει στους χρήστες να καθορίζουν τον τρόπο εμφάνισης των δεδομένων.

Συνηθισμένες τεχνικές είναι[24]:

- Γραφική παράσταση γραμμής: δείχνει τη σχέση μεταξύ στοιχείων και χρησιμοποιείται για συγκρίσεις στο χρόνο.

- Ραβδόγραμμα: συγκρίνει ποσότητες μεταξύ διαφορετικών κατηγοριών.
- Διάγραμμα διασποράς: δισδιάστατο γράφημα που δείχνει τη μεταβλητότητα δύο στοιχείων. Με τον όρο μεταβλητότητα εννοούμε τον βαθμό στον οποίο μια μεταβλητή διαφέρει ή αποκλίνει από τη μέση τιμή της, υποδεικνύοντας πόσο απλωμένες είναι οι τιμές της σε ένα σύνολο δεδομένων.
- Κυκλικό γράφημα: συγκρίνει τα μέρη ενός συνόλου.

Οι μορφές μπορεί να είναι ραβδογράμματα, κυκλικά γραφήματα, γραμμικά διαγράμματα κ.ά. Είναι σημαντικό να γνωρίζουμε ποια μορφή ταιριάζει καλύτερα στα δεδομένα μας.

Η οπτικοποίηση χρησιμοποιεί γραφικά για την ανάδειξη των δεδομένων, των τάσεων και σχέσεων μεταξύ στοιχείων και για αυτό το λόγο είναι κρίσιμο να γίνεται με τρόπο που βοηθάει τους χρήστες. Ακόμα και τα χρώματα πρέπει να επιλέγονται προσεκτικά για να διεκρίνονται τα δεδομένα μεταξύ τους ή να δίνετε ακόμα και μεγαλύτερη έμφαση σε κάποια.

Στο recommendation engine εργαλείο που υλοποιούμε οπτικοποιήσαμε τα δεδομένα μας πρώτα ανά ερώτηση και έπειτα ανά κατηγορία πολιτών (target groups) που θέλουμε να εστιάσουν οι συστάσεις μας. **Οι κατηγορίες πολιτών που επιλέξαμε είναι οι εξής:**

- Ηλικιωμένοι (Elderly)
- Άτομα με νοητικές αναπηρίες (Cognitive Disabilities)
- Ψηφιακά ευάλωτα άτομα (Digital Vulnerable People)
- Γυναίκες και άτομα ευάλωτα με βάση το φύλο τους (Women and gender related vulnerabilities)
- Νέοι (18-25) (Young (18-25))
- Μετανάστες (Migrant)
- Μονογονεϊκή οικογένεια (Single Parent Family)
- Κάτοικοι αγροτικής περιοχής (Rural inhabitant)
- Ποδηλάτες (Cyclist)
- Άτομα με σωματικές αναπηρίες (Physical Disabilities)
- Χαμηλόμισθοι (Low Income)
- Φοιτητές Πανεπιστημίου (University Student)

Εν κατακλείδι, η οπτικοποίηση δεδομένων είναι η γραφική ή εικονογραφική αναπαράσταση δεδομένων με σαφή και αποτελεσματικό τρόπο. Είναι ισχυρό εργαλείο για ανάλυση και ερμηνεία πολύπλοκων δεδομένων και προσφέρει έναν γρήγορο, παγκόσμια κατανοητό τρόπο παρουσίασης. Πρέπει να μεταδίδει πολύπλοκες ιδέες με σαφήνεια, ακρίβεια και αποτελεσματικότητα. Αυτά τα πλεονεκτήματα την καθιστούν χρήσιμη σε πολλούς επιστημονικούς τομείς.

3.3 Βιβλιοθήκες Python

Για την υλοποίηση της οπτικοποίησης δεδομένων, χρησιμοποιήσαμε δύο βιβλιοθήκες της γλώσσας προγραμματισμού python, που μας δίνουν την δυνατότητα να μπορέσουμε να παρουσιάσουμε τα δεδομένα που συλλέξαμε. Έχει αξία να αναφέρουμε πως μέσω της δημιουργίας κώδικα στην γλώσσα προγραμματισμού python και αξιοποιώντας τις δύο βιβλιοθήκες που θα αναφέρουμε στην συνέχεια, καταφέραμε να αυτοματοποιήσουμε τις διαδικασίες της οπτικοποίησης των δεδομένων, με αποτέλεσμα να μπορεί ο κώδικας να επαναχρησιμοποιηθεί σε όλο το φάσμα των δεδομένων. Παρακάτω παραθέτουμε τις δύο αυτές βιβλιοθήκες και τον τρόπο χρήσης τους.

Η βιβλιοθήκη Matplotlib

Το πακέτο matplotlib[25] αποτελεί μία από τις πιο δημοφιλείς βιβλιοθήκες για τη δημιουργία γραφικών και διαγραμμάτων στη γλώσσα προγραμματισμού Python. Το γεγονός ότι είναι ένα έργο ανοιχτού κώδικα με μεγάλη κοινότητα συμβάλλει στη συνεχή βελτίωσή του και στην εκτενή τεκμηρίωση, κάνοντάς το κατάλληλο τόσο για αρχάριους όσο και για έμπειρους χρήστες της Python.

Δημιουργήθηκε με σκοπό να παρέχει λειτουργίες παρόμοιες με εκείνες του MATLAB[26], επιτρέποντας στους χρήστες να παράγουν υψηλής ποιότητας διαγράμματα με σχετικά απλό και ευέλικτο κώδικα. Μέσω της βασικής διεπαφής του pyplot, προσφέρει εργαλεία για τη δημιουργία γραμμικών διαγραμμάτων, ραβδογραμμάτων, διαγραμμάτων διασποράς, ιστογραμμάτων και πολλών άλλων τύπων οπτικοποιήσεων.

Ένα από τα σημαντικότερα πλεονεκτήματα του matplotlib είναι η δυνατότητα πλήρους ελέγχου κάθε στοιχείου του διαγράμματος, όπως ο τίτλος, οι άξονες, οι ετικέτες, τα χρώματα και οι γραμματοσειρές. Επιπλέον, το matplotlib υποστηρίζει ενσωμάτωση με άλλες βιβλιοθήκες, όπως το NumPy[27] και το pandas[28], διευκολύνοντας την απεικόνιση δεδομένων από πίνακες ή πίνακες δεδομένων. Μέσω των υποπακέτων του, όπως το mpl_toolkits, είναι εφικτή ακόμη και η

δημιουργία τρισδιάστατων γραφικών ή γεωγραφικών απεικονίσεων.

Τέλος, το `matplotlib` προσφέρει πολλαπλές επιλογές εξαγωγής των γραφικών σε διάφορες μορφές αρχείων, όπως PNG, PDF, SVG, και EPS, καθιστώντας το ιδιαίτερα χρήσιμο για την ενσωμάτωσή τους σε επιστημονικές δημοσιεύσεις ή παρουσιάσεις.

Η βιβλιοθήκη Seaborn

Το Seaborn[24] είναι μια βιβλιοθήκη για τη δημιουργία στατιστικών γραφημάτων στην Python. Βασίζεται στο `matplotlib` και ενσωματώνεται στενά με τις δομές δεδομένων της `pandas`.

Το Seaborn μας βοηθά να εξερευνήσουμε και να κατανοήσουμε τα δεδομένα μας. Οι συναρτήσεις σχεδίασης λειτουργούν πάνω σε `dataframes` και `arrays` που περιέχουν ολόκληρα σύνολα δεδομένων, πραγματοποιώντας εσωτερικά την απαραίτητη σημασιολογική αντιστοίχιση και τη στατιστική συγκέντρωση ώστε να παραχθούν κατατοπιστικά γραφήματα. Η δηλωτική (declarative) διεπαφή προγραμματισμού εφαρμογών (Application Programming Interface - API) του σε αφήνει να εστιάζεις στο τι σημαίνουν τα στοιχεία του γραφήματος και όχι στο πώς σχεδιάζονται.

Στο παρασκήνιο, το `seaborn` χρησιμοποιεί το `matplotlib` για τη σχεδίαση των γραφημάτων. Για διαδραστική χρήση, προτείνεται να χρησιμοποιείται Jupyter/IPython σε λειτουργία `matplotlib` ή διαφορετικά να καλείται η `matplotlib.pyplot.show()` για να εμφανιστεί το γράφημα.

Δεν υπάρχει μία "σωστή" απεικόνιση για όλα τα δεδομένα. Ανάλογα με την ερώτηση, απαιτείται διαφορετικό γράφημα. Το `seaborn` κάνει εύκολη την εναλλαγή μεταξύ διαφορετικών απεικονίσεων, με μια συνεπή, βασισμένη σε `dataset`, API.

Υπάρχουν διάφορες χρήσεις:

- **Στατιστική Εκτίμηση:**

Συχνά θέλουμε να δούμε τον μέσο όρο μιας μεταβλητής σε σχέση με άλλες. Το `seaborn` μπορεί να υπολογίσει αυτόματα τις στατιστικές τιμές. Το `seaborn` εφαρμόζει bootstrapping και σχεδιάζει `error bars` που δείχνουν την αβεβαιότητα της εκτίμησης. Για ακόμα πιο προηγμένη στατιστική οπτικοποίηση, μπορούμε να προσθέσουμε γραμμικά μοντέλα με `lmplot()`.

- **Αναπαράσταση Κατανομών:**

Η στατιστική ανάλυση απαιτεί γνώση των κατανομών των μεταβλητών. Η `displot()` υποστηρίζει διάφορες μεθόδους απεικόνισης, όπως ιστογράμματα ή πυκνότητα `kernel`. Η πυκνότητα κερνελ (Kernel Density Estimation - KDE) είναι μια μη παραμετρική μέθοδος για την εκτίμηση της πιθανότητας κατανομής μιας συνεχούς μεταβλητής. Με απλά λόγια, αντί να ομαδοποιεί τα δεδομένα σε διαστήματα όπως κάνει το ιστογράμμα, η KDE δημιουργεί μια ομαλή καμπύλη που δείχνει πού συγκεντρώνονται οι περισσότερες τιμές.

- **Γραφήματα Κατηγορικών Δεδομένων:**

Η `catplot()` παρέχει απεικονίσεις ειδικά για κατηγορικά δεδομένα. Ένα παράδειγμα είναι το `swarm plot`, που μετακινεί ελαφρώς τα σημεία ώστε να μην επικαλύπτονται

- **Πολυμεταβλητές Απεικονίσεις:**

Κάποιες συναρτήσεις, όπως η `jointplot()`, συνδυάζουν πολλαπλά γραφήματα για συνοπτικές παρουσιάσεις. Η `jointplot()` απεικονίζει τη σχέση μεταξύ δύο μεταβλητών καθώς και τις οριακές τους κατανομές

- **Εργαλεία χαμηλότερου επιπέδου για τη δημιουργία διαγραμμάτων:**

Αυτά τα εργαλεία λειτουργούν συνδυάζοντας συναρτήσεις σχεδίασης σε επίπεδο άξονα (`axes-level`) με αντικείμενα που διαχειρίζονται τη διάταξη του διαγράμματος, συνδέοντας τη δομή ενός `dataset` με ένα πλέγμα από άξονες. Και τα δύο αυτά στοιχεία αποτελούν μέρος του δημόσιου API και μπορούν να χρησιμοποιηθούν απευθείας για να δημιουργήσεις περίπλοκα διαγράμματα με λίγες επιπλέον γραμμές κώδικα.

Τέλος, η ενσωμάτωση του `seaborn` με το `matplotlib` μας επιτρέπει να το χρησιμοποιήσουμε σε όλα τα περιβάλλοντα που υποστηρίζει το `matplotlib`. Για παράδειγμα η δική μας ανάλυση και οπτικοποίηση δεδομένων έγινε σε περιβάλλον `jupyter notebook` για ευκολία χρήσης και επαναχρησιμοποίηση.

3.4 Κώδικας Υλοποίησης Οπτικοποίησης Δεδομένων

Με βάση όλα τα παραπάνω παραθέτουμε κάποια παραδείγματα κώδικα με τα οποία υλοποιήσαμε την ανάλυση και την οπτικοποίηση των δεδομένων μας.

Τα τρία παραδείγματα κώδικα αφορούν τρία διαφορετικά διαγράμματα που υλοποιήσαμε.

```
1 import pandas as pd
2 import matplotlib.pyplot as plt
3 # Define priority levels and initialize an empty DataFrame
4 # to store aggregated results
5 priority_levels = {
6     'First Priority': 1,
7     'Second Priority': 2,
8     'Third Priority': 3
9 }
10 aggregated_data = pd.DataFrame()
11
12 # Loop through each priority level and calculate distributions
13 for priority_name, priority_value in priority_levels.items():
14     # Select and clean the data for the current priority
15     data = greece_raw[[
16         'PRIO_TIME_SAVINGS',
17         'PRIO_AVAILABILITY_TIME',
18         'PRIO_SAFETY_SECURITY',
19         'PRIO_TRAVEL_COSTS',
20         'PRIO_RELIABILITY',
21         'PRIO_ENVIRONMENTAL_R',
22         'PRIO_REACHABILITY',
23         'PRIO_COMFORT',
24         'PRIO_HEALTH_R',
25         'PRIO_CLEANNES',
26         'PRIO_OTHER'
27     ]]
28     cleaned_data = data.applymap(lambda x: 1 if x == \
29     priority_value else 0)
30     # Calculate the distribution and store it
31     priority_distribution = cleaned_data.sum()
32     aggregated_data[priority_name] = priority_distribution
33
34 # Define labels for categories
35 # (removing unused ones based on cleaned data)
36 labels = [
37     'TIME_SAVINGS',
38     'AVAILABILITY_TIME',
```

```
39     'SAFETY_SECURITY',
40     'TRAVEL_COSTS',
41     'RELIABILITY',
42     'ENVIRONMENTAL',
43     'REACHABILITY',
44     'COMFORT',
45     'HEALTH_R',
46     'CLEANNESS',
47     'OTHER'
48 ]
49 aggregated_data.index = labels
50 # Plot a grouped bar chart
51 plt.figure(figsize=(14, 8))
52 bar_width = 0.25
53 x_positions = range(len(aggregated_data.index))
54 # Create grouped bars for each priority
55 for i, priority_name in enumerate(aggregated_data.columns):
56     plt.bar(
57         [x + i * bar_width for x in x_positions],
58         aggregated_data[priority_name],
59         bar_width,
60         label=priority_name
61     )
62 # Customize the chart
63 plt.xticks(
64     [x + (bar_width * len(priority_levels)) / 2 for x in \
65      x_positions],
66     aggregated_data.index,
67     rotation=45,
68     ha='right'
69 )
70 plt.xlabel('Priority Categories')
71 plt.ylabel('Number of Respondents')
72 plt.title('Priorities Distribution Across Categories')
73 plt.legend(title='Priority')
74 plt.grid(axis='y')
75 plt.tight_layout()
76 # Show the chart
77 plt.show()
```

Κώδικας 3.1: Bar Chart

```
1 import pandas as pd
2 import matplotlib.pyplot as plt
3
4 # 1. Load the dataset and select rows 0 to 68
5 # (i.e., the first 69 entries)
6 # Focus only on the 'MOB_APP_USE' column,
7 # which contains responses related to mobile app usage
8 raw_data = pd.read_excel('FINAL_DATASET.xlsx')
9 data = raw_data.loc[0:68, ['MOB_APP_USE']]
10
11 # 2. Clean the data: Replace numeric codes
12 # with corresponding descriptive labels
13 # This improves readability and interpretability
14 # in the visualization
15 data['MOB_APP_USE'] = data['MOB_APP_USE'].replace({
16     1:'I have never done this',
17     2:'I last did this more than 3 months ago',
18     3:'Less than once a week but at least once in the last 3 months',
19     4:'At least once a week',
20     5:'Every day or almost every day',
21     6:'Many times in a day',
22     0:'Dont know/prefer not to answer'
23 })
24 # Define all possible answer categories,
25 # including those that might not appear in the actual responses
26 # This ensures the pie chart accounts for every potential option
27 all_categories = [
28     'I have never done this',
29     'I last did this more than 3 months ago',
30     'Less than once a week but at least once in the last 3 months',
31     'At least once a week',
32     'Every day or almost every day',
33     'Many times in a day',
34     'Dont know/prefer not to answer'
35 ]
36 # 3. Calculate the distribution (frequency count)
37 # of each response category
38 # Ensure missing categories are represented with a count of zero
39 ans_distribution = data['MOB_APP_USE'].value_counts(). \
40     reindex(all_categories, fill_value=0)
```

```
41
42 # Filter the distribution to keep only
43 # categories with a non-zero count
44 # This avoids displaying empty segments in the pie chart
45
46 filtered_distribution = ans_distribution[ans_distribution > 0]
47 # Assign a specific color to each category
48 # in the filtered list for better visual clarity
49 filtered_colors = [
50     color for label, color in zip(
51         all_categories,
52         ['red',
53          'lightcoral',
54          'lightblue',
55          'lightgreen',
56          'green',
57          'yellow',
58          'grey']
59     ) if label in filtered_distribution.index
60 ]
61
62 # 4. Create a pie chart using the cleaned and filtered data
63 # Display percentage labels on each slice
64 # and use consistent coloring
65
66 plt.figure(figsize=(10, 8))
67 plt.pie(
68     filtered_distribution,
69     labels=filtered_distribution.index,
70     autopct='%1.1f%%',
71     colors=filtered_colors
72 )
73 # Remove the y-axis label for a cleaner look
74
75 plt.ylabel('')
76 plt.show()
```

Κώδικας 3.2: Pie Chart

```

1 import pandas as pd
2 import matplotlib.pyplot as plt
3 import seaborn as sns
4 # Create the DataFrame ()
5 data = section_g
6 # 2. Visualization
7 # Age distribution
8 plt.figure(figsize=(8, 6))
9 sns.histplot(data['AGE'], bins=10, kde=True)
10 plt.title('Age Distribution')
11 plt.xlabel('Age')
12 plt.ylabel('Frequency')
13 plt.show()

```

Κώδικας 3.3: Histogram

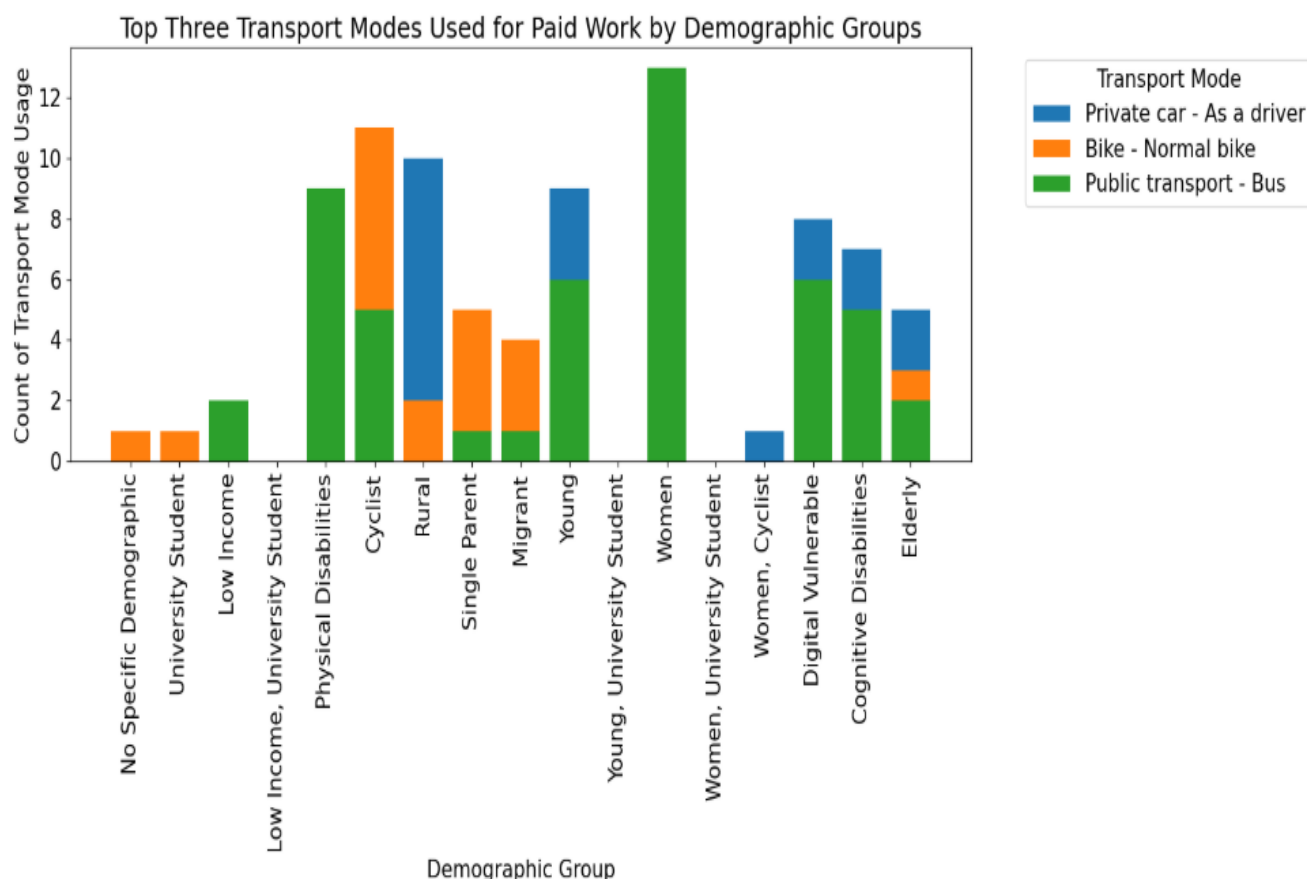
3.5 Αποτελέσματα

Σε αυτή την ενότητα θα παρουσιάσουμε τα αποτελέσματα της οπτικοποίησης των δεδομένων που συλλέξαμε από τα ερωτηματολόγια που αναφέραμε παραπάνω. Στον Πίνακα 3.1 παρουσιάζουμε μία πιο εκτενή ποσοτικά περιγραφή των δεδομένων μας.

Πίνακας 3.1: Πίνακας Περιγραφής των Δεδομένων

Πόλεις	Πολίτες που Ρωτήθηκαν	Μέγεθος Δεδομένων(kByte)
Τρίκαλα	69	579,8
Noord-Brabant	81	597,5
West Midlands	69	572,5
Hamburg	69	731,3

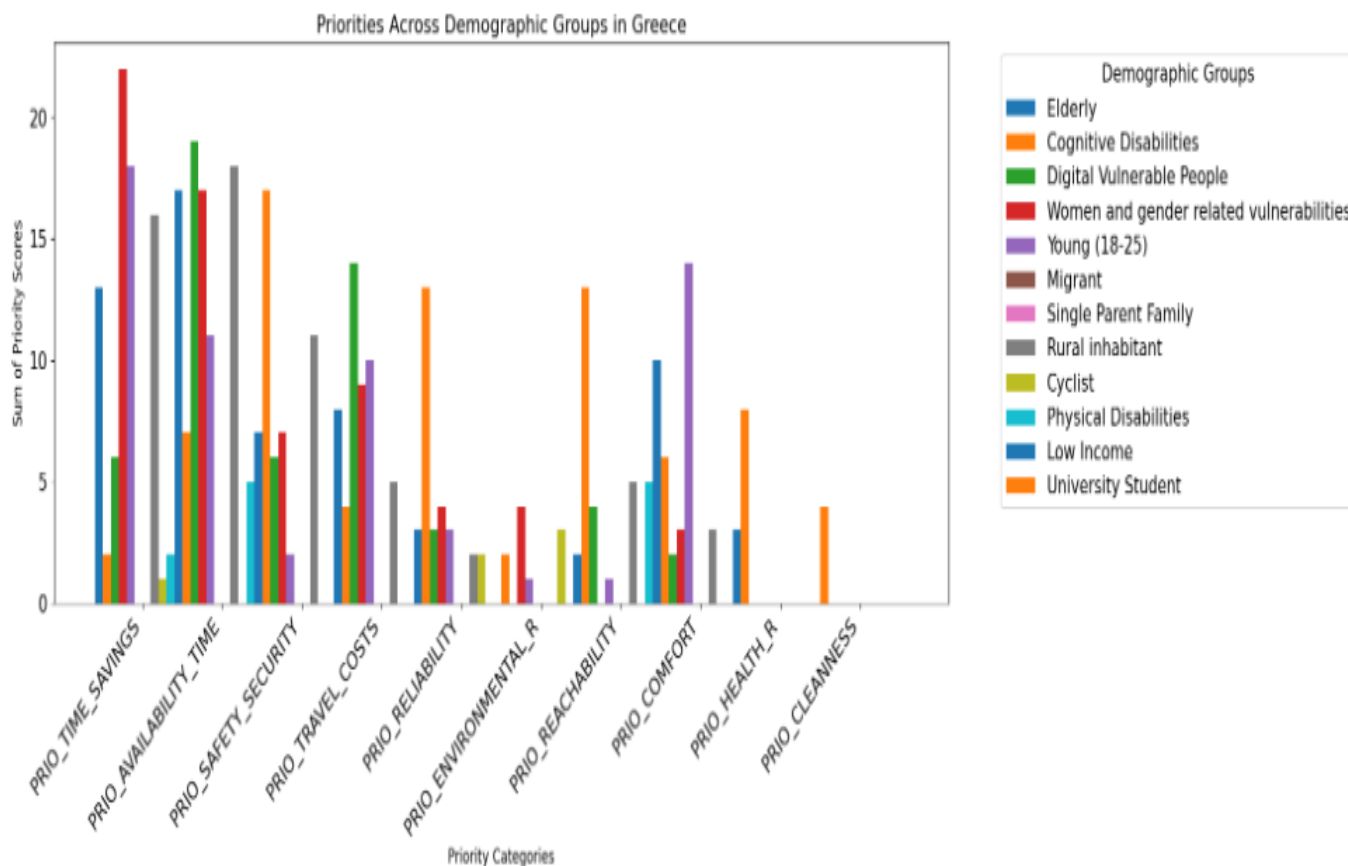
Δεδομένου των πολλαπλών ερωτημάτων και του μεγέθους των δεδομένων, είναι δύσκολο να παρουσιάσουμε αναλυτικά τα διάφορα διαγράμματα που υλοποιήσαμε. Για αυτό το λόγο παρουσιάζουμε μόνο ένα διάγραμμα ανά κατηγορία ερωτήσεων που εμφανίζει οπτικοποιημένα τα δεδομένα ανά target group. Επίσης, για λόγους καλύτερης κατανόησης των δεδομένων, όλα τα διαγράμματα που παραθέτουμε είναι από τα δεδομένα που συλλέξαμε από τα Τρίκαλα (Ελλάδα).



Εικόνα 3.2: Οι τρεις κορυφαίοι τρόποι μεταφοράς που χρησιμοποιούνται για αμειβόμενη εργασία

Στην Εικόνα 3.2 τα αποτελέσματα δείχνουν σημαντικές διαφορές στη χρήση μέσων μεταφοράς ανά δημογραφική ομάδα. Η πιο διαδεδομένη επιλογή είναι το ιδιωτικό αυτοκίνητο (ως οδηγός), ιδιαίτερα σε ομάδες όπως Low Income, Single Parent, και Rural. Το ποδήλατο χρησιμοποιείται κυρίως από Cyclists, Young, και Φοιτητές, ενώ το λεωφορείο είναι δημοφιλές σε ομάδες με φυσικούς περιορισμούς (Physical Disabilities) και χαμηλό εισόδημα.

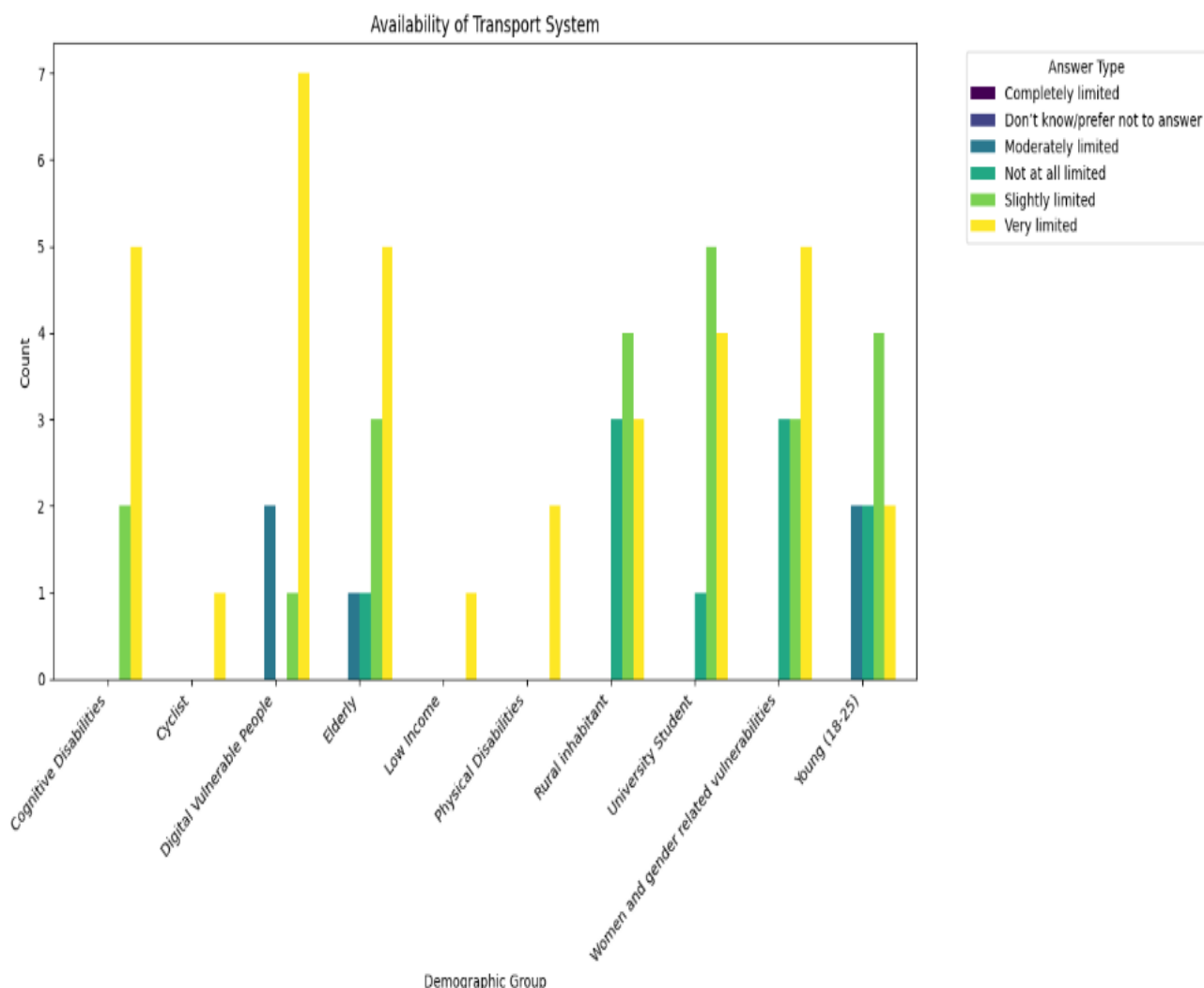
Για να υποστηριχθούν οι ομάδες με έλλειψη πρόσβασης σε ιδιωτικό αυτοκίνητο, προτείνεται η επέκταση δρομολογίων λεωφορείων σε αγροτικές περιοχές και η δημιουργία ασφαλών ποδηλατοδρόμων σε αστικές ζώνες. Επίσης, η χορήγηση επιδοτήσεων για ηλεκτρικά ποδήλατα σε φοιτητές και οικογένειες με χαμηλό εισόδημα θα μείωνε την εξάρτηση από το αυτοκίνητο.



Εικόνα 3.3: Προτεραιότητες με τις οποίες επιλέγονται οι τρόποι μεταφοράς

Η Εικόνα 3.3 απεικονίζει τη συνολική βαθμολογία προτεραιοτήτων όπως εξοικονόμηση χρόνου, ασφάλεια, μείωση κόστους ταξιδιών κ.ά. Παρατηρείται ότι ομάδες όπως οι νέοι (18-25), οι γυναίκες και άτομα με φυσικές αναπηρίες δίνουν υψηλή έμφαση στην εξοικονόμηση χρόνου και στην ασφάλεια. Αντίθετα, περιβαλλοντικά ζητήματα ή η καθαριότητα καταγράφουν χαμηλές βαθμολογίες προτεραιότητας σε όλες τις ομάδες. Οι φοιτητές πανεπιστημίου φαίνεται να διαφοροποιούνται ελαφρώς δίνοντας περισσότερη έμφαση σε θέματα άνεσης.

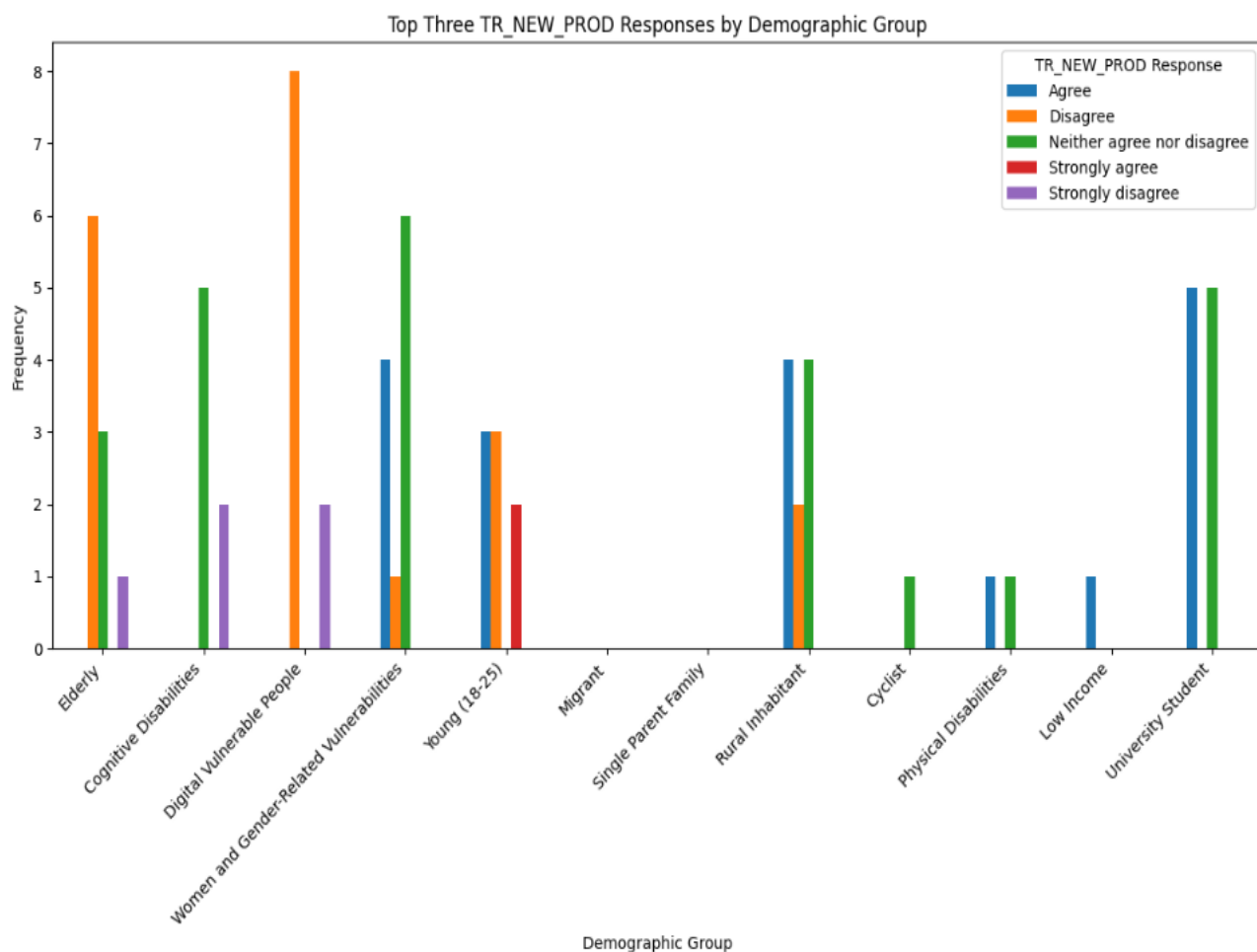
Δεδομένων των ευρημάτων, προτείνεται να δοθεί έμφαση στον ανασχεδιασμό υπηρεσιών και υποδομών με κύρια στόχευση την εξοικονόμηση χρόνου και την ενίσχυση της αίσθησης ασφάλειας, ειδικά για νεότερες ηλικιακές ομάδες και ευάλωτα άτομα, χωρίς όμως να παραμελούνται οι περιβαλλοντικές πτυχές που προς το παρόν φαίνεται να υποεκτιμώνται.



Εικόνα 3.4: Βαθμός αίσθησης περιορισμού των πολιτών στα ταξίδια τους σε σχέση με την διαθεσιμότητα των Μ.Μ.Μ.

Στην Εικόνα 3.4 παρουσιάζεται η διαθεσιμότητα των μέσων μεταφοράς όπως την αντιλαμβάνονται οι διάφορες κοινωνικές ομάδες. Παρατηρούμε ότι οι ηλικιωμένοι («Elderly») και τα άτομα χαμηλού εισοδήματος («Low Income») δηλώνουν σε πολύ υψηλό ποσοστό («Very limited») ότι το σύστημα δεν είναι προσβάσιμο, ενώ οι φοιτητές Πανεπιστημίου και οι κάτοικοι αγροτικών περιοχών εμφανίζουν σχετικά καλύτερη εικόνα. Γενικότερα, οι πιο ευάλωτες ομάδες (ηλικιωμένοι, άτομα με γνωστικές ή σωματικές αναπηρίες) δίνουν τις χαμηλότερες βαθμολογίες διαθεσιμότητας.

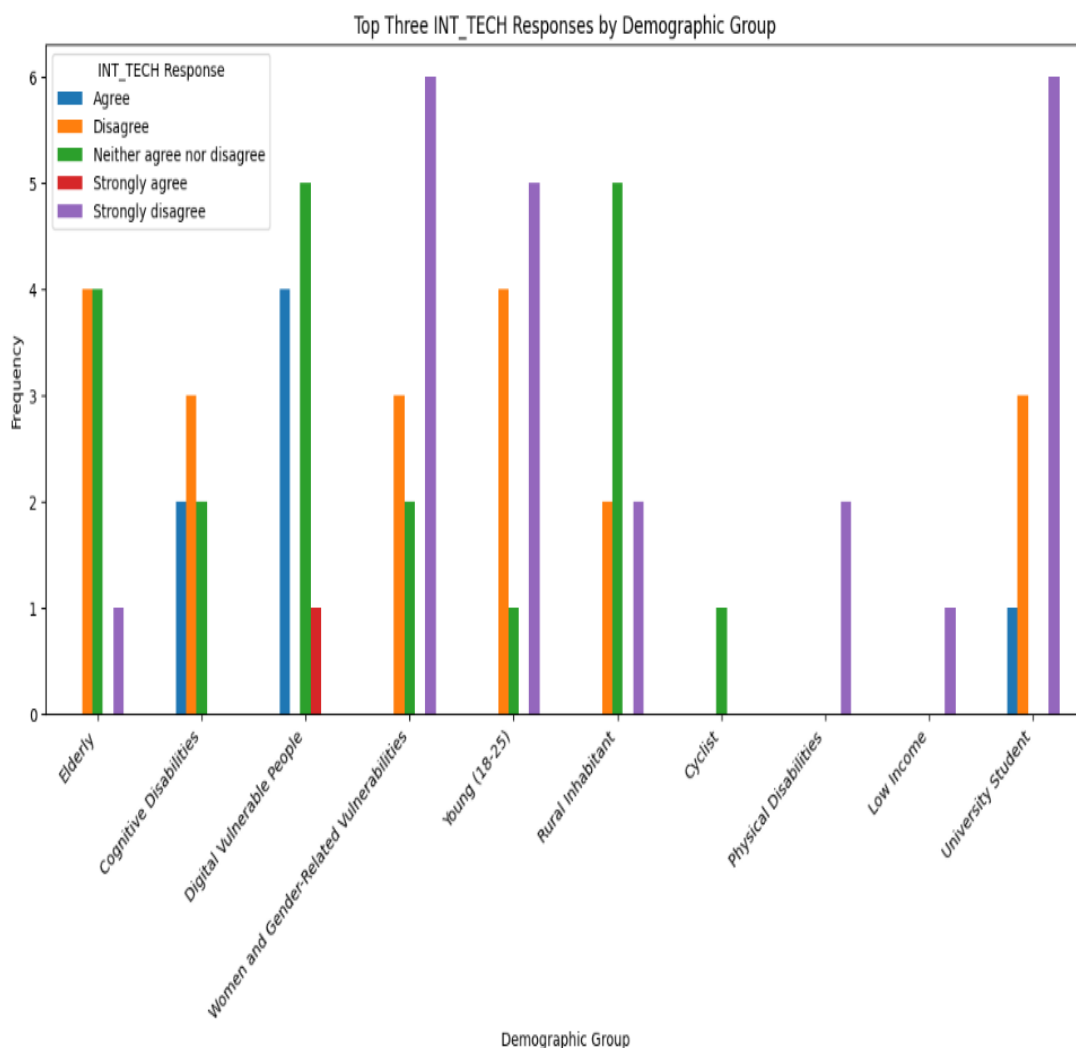
Προτείνεται η ενίσχυση της συχνότητας δρομολογίων και της προσβασιμότητας (π.χ. ράμπες, στάσεις χαμηλού ύψους) σε περιοχές με υψηλό ποσοστό «Very limited», ειδικά για ηλικιωμένους και χαμηλού εισοδήματος χρήστες, ώστε να μειωθούν τα εμπόδια στη μετακίνηση αυτών των ομάδων.



Εικόνα 3.5: Βαθμός συμφωνίας με την φράση “Αγοράζω καινούρια προϊόντα πριν τους φίλους μου και τους γείτονες μου”

Στην Εικόνα 3.5 οι απόψεις στην ερώτηση αν οι πολίτες θα αγόραζαν καινούρια προϊόντα πριν του φίλους ή τους γείτονες τους, ποικίλουν σημαντικά. Οι Rural Inhabitants και Low Income ομάδες τείνουν να διαφωνούν ή να είναι ουδέτερες, ενώ οι University Students και Cyclists εμφανίζουν υψηλότερη τάση να συμφωνούν. Οι ομάδες με ευάλωτες καταστάσεις (π.χ. Women and Gender-Related Vulnerabilities) εκφράζουν συχνά δυναμική διαφωνία Strongly Disagree, πιθανώς λόγω έλλειψης προσαρμογής του προϊόντος στις ανάγκες τους.

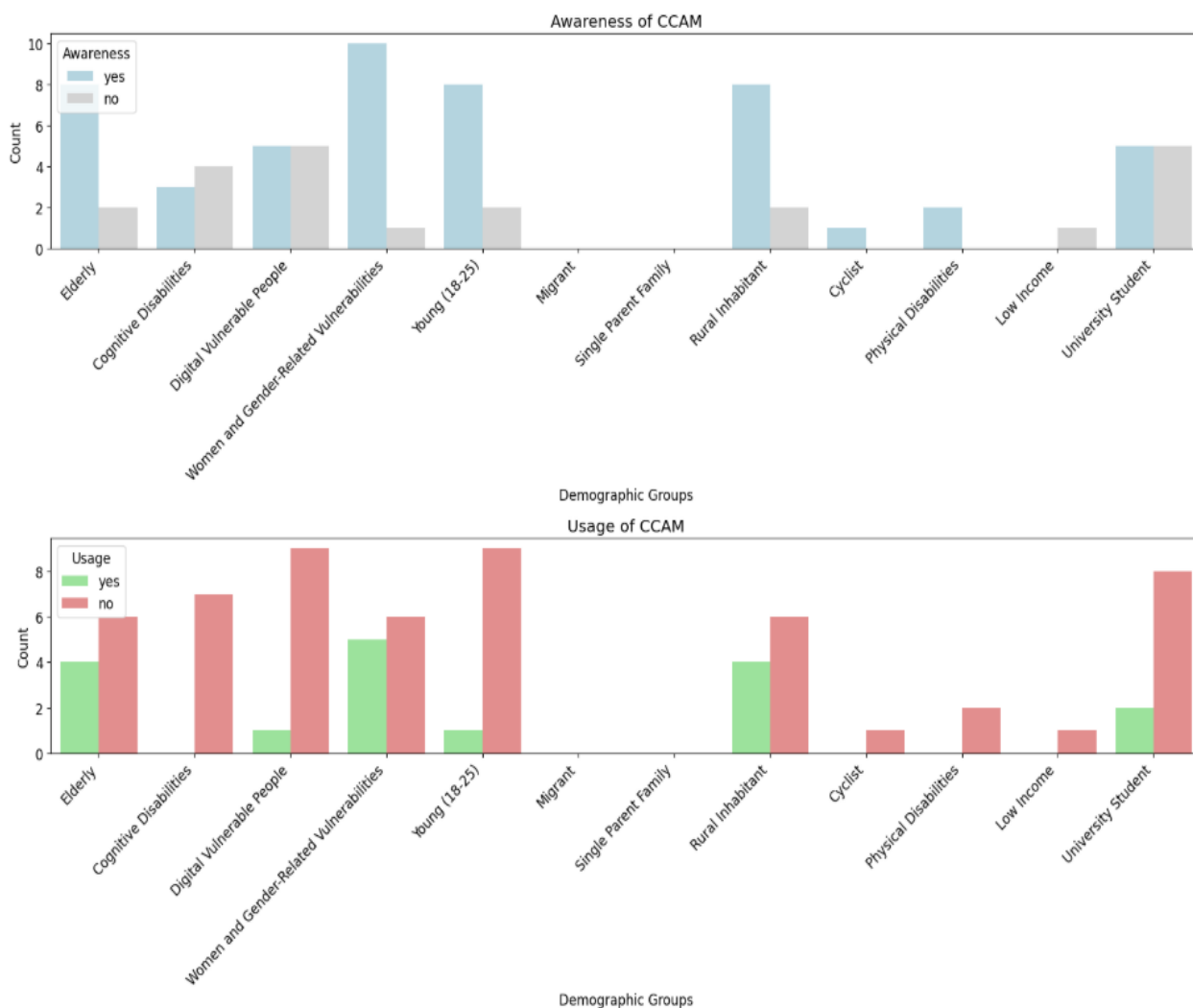
Προτείνεται η διεξαγωγή εστιασμένων ερευνών αγοράς για τις ομάδες που αντιδρούν αρνητικά, με στόχο την προσαρμογή του προϊόντος σε θέματα προσβασιμότητας και τιμολόγησης. Επίσης, η δημιουργία εκπαιδευτικών υλικών για Digital Vulnerable People θα βελτιώνει την αποδοχή του προϊόντος.



Εικόνα 3.6: Βαθμός συμφωνίας με τη φράση “Με ενδιαφέρει λίγο έως καθόλου η νέα τεχνολογία”

Στην Εικόνα 3.6 φαίνονται οι τρεις πιο συχνές απαντήσεις των ερωτηθέντων στην ερώτηση σχετικά με την αποδοχή νέων τεχνολογιών. Οι φοιτητές Πανεπιστημίου και οι νέοι (18–25) τείνουν να διαφωνούν ή να διαφωνούν έντονα («Disagree», «Strongly disagree»), ενώ οι κάτοικοι αγροτικών περιοχών και οι ψηφιακά ευάλωτες ομάδες εμφανίζουν μεγαλύτερη αμηχανία («Neither agree nor disagree»). Οι ηλικιωμένοι και τα άτομα με γνωστικές αναπηρίες δίνουν περισσότερα αρνητικά (disagree) παρά θετικά (agree).

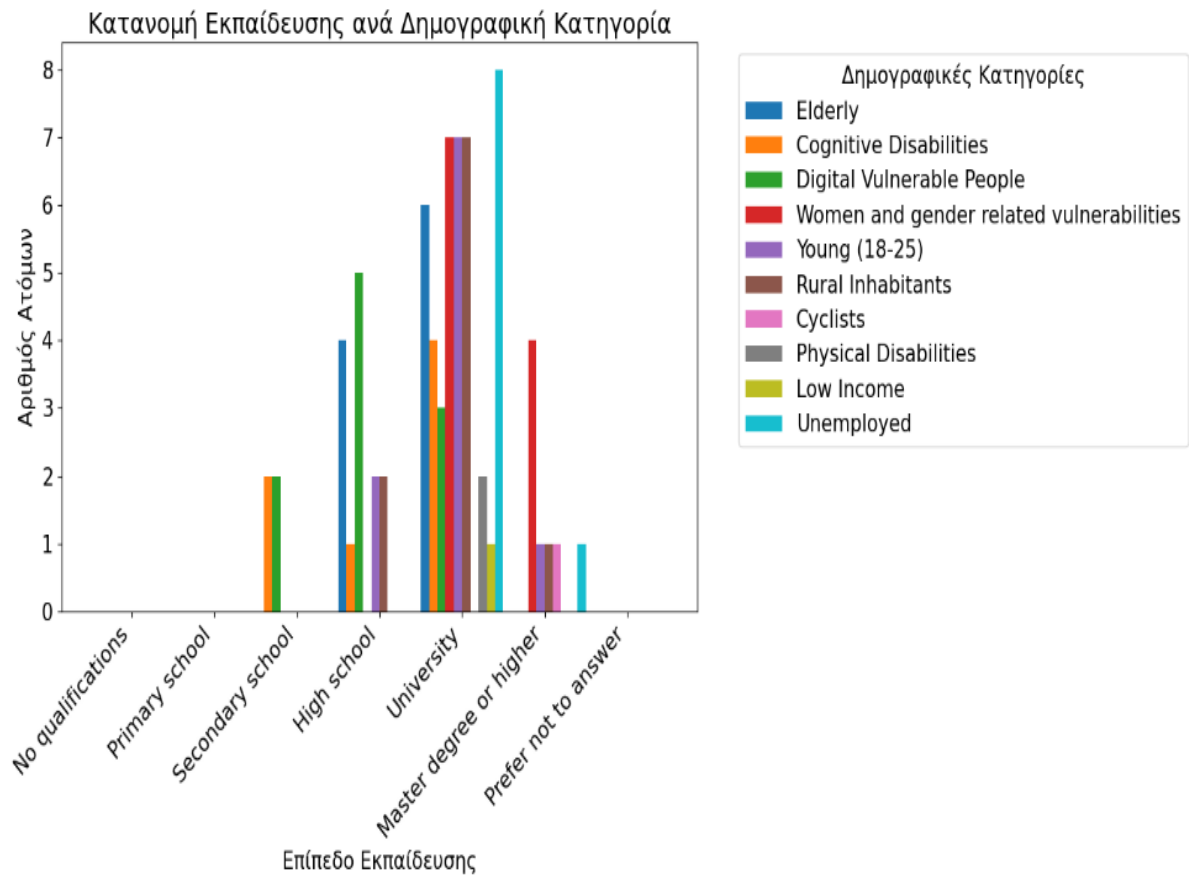
Συστήνεται η διοργάνωση στοχευμένων ενημερωτικών εκδηλώσεων και σεμιναρίων χρήσης νέων τεχνολογιών (π.χ. εφαρμογών κινητικότητας) για τους φοιτητές και νέους χρήστες, ώστε να αντιμετωπιστεί ο σκεπτικισμός και να ενισχυθεί η αποδοχή των καινοτόμων λύσεων.



Εικόνα 3.7: Χρήση ή γνώση της παρουσία του CCAM στο σύστημα κινητικότητας

Στην Εικόνα 3.7, το πρώτο διάγραμμα παρουσιάζει το επίπεδο ενημέρωσης διαφορετικών δημογραφικών ομάδων σχετικά με το CCAM, ενώ το δεύτερο διάγραμμα δείχνει την πραγματική χρήση του. Παρατηρείται ότι, ακόμα και σε ομάδες με σχετικά υψηλή ενημέρωση (όπως οι νέοι και οι αγροτικοί κάτοικοι), η χρήση του CCAM παραμένει χαμηλή. Αντίθετα, οι ευάλωτες ομάδες, όπως άτομα με αναπηρίες και άτομα με χαμηλό εισόδημα, εμφανίζουν ιδιαίτερα χαμηλή ενημέρωση και μηδαμινή χρήση.

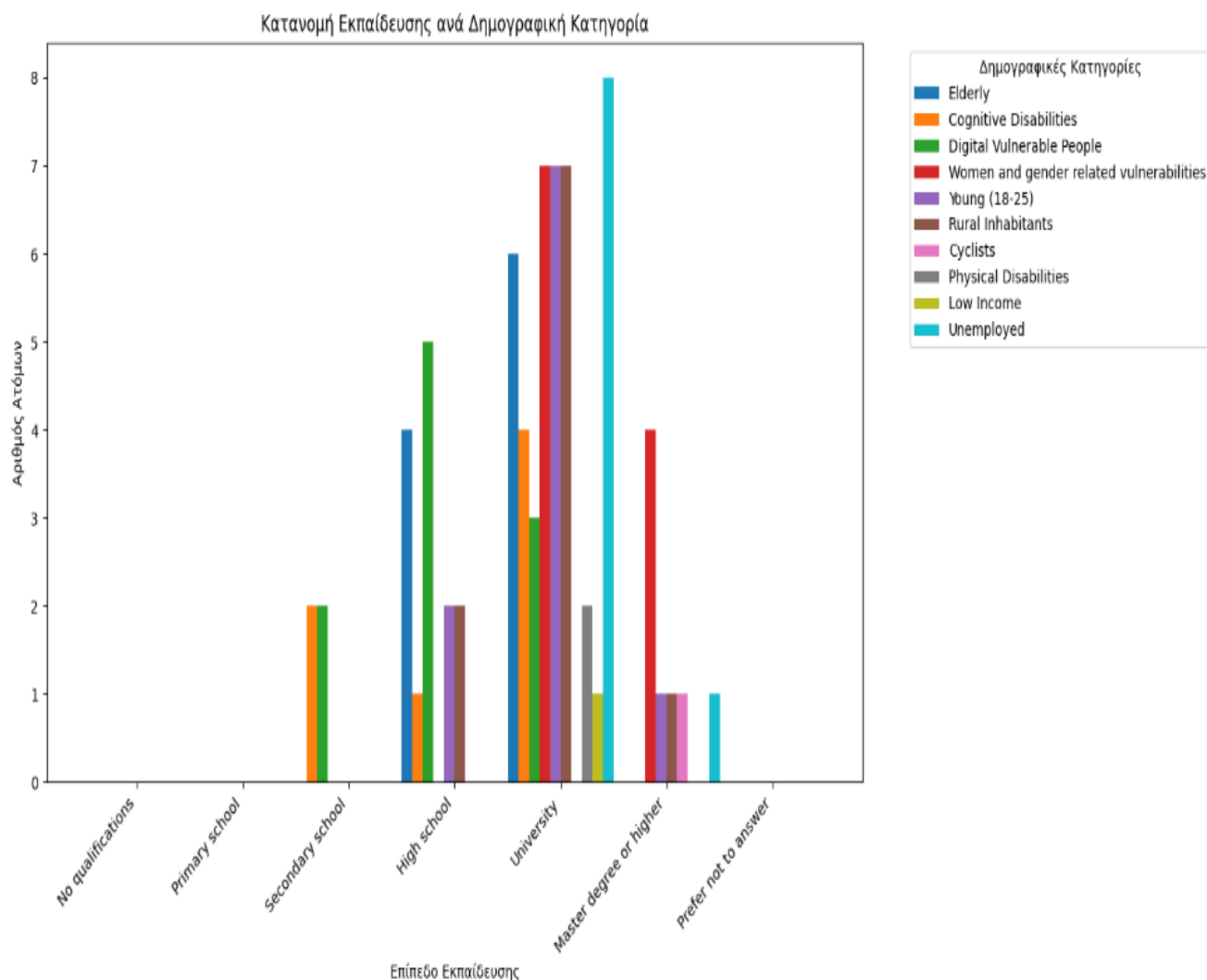
Προτείνεται η στοχευμένη υλοποίηση καμπανιών ευαισθητοποίησης και εκπαίδευσης σχετικά με το CCAM, ειδικά σε ευάλωτες ομάδες και νεαρούς ενήλικες, ώστε να γεφυρωθεί το χάσμα μεταξύ ενημέρωσης και πραγματικής χρήσης και να επιτευχθεί μεγαλύτερη ένταξη αυτών των πληθυσμών στις νέες τεχνολογίες κινητικότητας.



Εικόνα 3.8: Μέτρα ασφάλειας σε περίπτωση που δεν υπάρχει ανθρώπινο δυναμικό

Στην Εικόνα 3.8 απεικονίζονται οι τρεις σημαντικότερες επιλογές των ερωτηθέντων όσον αφορά τα χαρακτηριστικά ασφάλειας στα μέσα μεταφοράς: κουμπιά βοήθειας με απευθείας επικοινωνία σε κέντρο ελέγχου, κάμερες με κλειστό κύκλωμα τηλεόρασης CCTV και οπτική επαφή με χειριστή στο κέντρο ελέγχου. Οι ψηφιακά ευάλωτες ομάδες και οι ηλικιωμένοι δείχνουν πολύ ισχυρή προτίμηση στην παρουσία φυσικού ανθρώπου (monitoring) και στις κάμερες παρακολούθησης, ενώ οι νέοι και οι φοιτητές επιλέγουν περισσότερο τα κουμπιά βοήθειας.

Προτείνεται η εγκατάσταση συνδυαστικών λύσεων ασφάλειας — εξοπλισμός CCTV σε κρίσιμες στάσεις και οχήματα, μαζί με κουμπιά άμεσης επέμβασης — και η προώθηση πιλοτικών προγραμμάτων με ανθρώπινη επίβλεψη σε δωμάτιο ελέγχου (control-room), προκειμένου να αυξηθεί η εμπιστοσύνη των πιο ευάλωτων χρηστών.



Εικόνα 3.9: Επίπεδο εκπαίδευσης ανά κατηγορία πολιτών

Στην Εικόνα 3.9 η εκπαιδευτική κατανομή αποκαλύπτει ανισότητες. Οι ομάδες Cognitive Disabilities, Digital Vulnerable People, και Low Income εμφανίζουν υψηλό ποσοστό μόνο Δημοτικού ή Γυμνασίου επιπέδου. Αντίθετα, οι University Students και Young (18-25) έχουν συχνά πανεπιστημιακή εκπαίδευση. Η ομάδα Rural Inhabitants εμφανίζει μεγάλη διασπορά, με σημαντικό ποσοστό να δηλώνει Prefer not to answer.

Απαιτείται η ενίσχυση των εκπαιδευτικών προγραμμάτων για ενήλικες σε αγροτικές περιοχές και η δημιουργία προσβάσιμων πλατφορμών ψηφιακής εκπαίδευσης για ομάδες με γνωστικές ή οικονομικές προκλήσεις. Επίσης, η προώθηση υποτροφιών για ανώτερη εκπαίδευση σε άτομα με αναπηρίες θα μείωνε τις εκπαιδευτικές ανισότητες.

3.6 Συμπεράσματα και Use Cases

Η ανάλυση και η οπτικοποίηση των δεδομένων από τα ερωτηματολόγια ανέδειξε με σαφήνεια σημαντικές διαφορές στις ανάγκες και στις εμπειρίες των επιμέρους κοινωνικών ομάδων σε σχέση με τη μετακίνηση και τη χρήση τεχνολογιών κινητικότητας, όπως οι λύσεις CCAM. Μέσα από τα διαγράμματα που δημιουργήσαμε, παρατηρήσαμε μοτίβα τα οποία δεν ήταν άμεσα εμφανή με την απλή παρατήρηση των δεδομένων. Για παράδειγμα, οι ηλικιωμένοι και τα άτομα με σωματικές ή νοητικές αναπηρίες δήλωσαν σε μεγάλο ποσοστό προβληματισμούς σχετικούς με την ασφάλεια, ενώ οι ψηφιακά ευάλωτες ομάδες και οι μετανάστες φάνηκαν να αντιμετωπίζουν σημαντικές δυσκολίες στην πρόσβαση ή χρήση ψηφιακών εργαλείων πλοήγησης και ενημέρωσης.

Με βάση αυτή την κατανόηση, μπορούμε να οδηγηθούμε σε συγκεκριμένες περιπτώσεις χρήσης (use cases) για το εργαλείο που αναπτύσσουμε. Ένα τέτοιο παράδειγμα χρήσης φαίνεται και στην Εικόνα 3.9. Πιο συγκεκριμένα, η εφαρμογή μας μπορεί να δέχεται ως είσοδο βασικές παραμέτρους, όπως η χώρα διαμονής, η κοινωνική ομάδα στην οποία ανήκει ο χρήστης (π.χ. άτομο με αναπηρία, νέος, μετανάστης), και τον τύπο περιοχής (αστική, αγροτική, κ.λπ.). Ανάλογα με αυτά τα δεδομένα, η εφαρμογή μπορεί να εμφανίζει στο χρήστη προσωποποιημένες συστάσεις με βάση τα παραπάνω συμπεράσματα, όπως: βελτίωση της προσβασιμότητας και του κόστους των μεταφορών, ενίσχυση μέτρων ασφαλείας σε MMM, απλοποίηση των ψηφιακών εργαλείων, παροχή εκπαιδευτικών σεμιναρίων για τη χρήση ψηφιακών εφαρμογών και προώθηση της εμπιστοσύνης στις τεχνολογίες CCAM. Αυτά τα αποτελέσματα συνιστούν απτά παραδείγματα για το πώς η ανάλυση δεδομένων μπορεί να οδηγήσει σε πρακτικά και κοινωνικά επωφελή αποτελέσματα.

Ωστόσο, για να μπορέσουμε να διαχειριστούμε αποτελεσματικά τις εισόδους αυτές και να παράγουμε τις αντίστοιχες εξατομικευμένες εξόδους, απαιτείται η χρήση ενός κατάλληλου μηχανισμού απόφασης στην πλευρά του διακομιστή (back-end) της εφαρμογής μας. Μία κατάλληλη προσέγγιση είναι η ενσωμάτωση ενός rule-based συστήματος, το οποίο θα βασίζεται σε συνθήκες τύπου if-else. Το σύστημα αυτό θα μπορεί να επεξεργάζεται τις εισόδους του χρήστη και να αντιστοιχεί τις κατάλληλες προτεινόμενες ενέργειες, αξιοποιώντας τόσο τα δεδομένα που προέκυψαν από τα ερωτηματολόγια όσο και κάποια οντολογία που θα ορίζει τις σχέσεις και τις ιδιαιτερότητες κάθε τρόπου μετακίνησης. Έτσι, η εφαρμογή θα λειτουργεί σαν ένας “έξυπνος οδηγός”, ο οποίος προτείνει λύσεις με βάση τις πραγματικές ανάγκες των πολιτών.

Συμπερασματικά, η οπτικοποίηση των δεδομένων δεν αποτελεί απλώς έναν τρόπο παρουσίασης, αλλά ένα βασικό εργαλείο κατανόησης και εξαγωγής νοήματος από σύνθετες πληροφορίες. Μέσα από αυτή τη διαδικασία, εντοπίζουμε συγκεκριμένες ανάγκες και μπορούμε να τις μεταφράσουμε σε στοχευμένες ενέργειες. Η υλοποίηση ενός rule-based μηχανισμού αποτελεί το επόμενο βήμα για να γεφυρωθεί η ανάλυση με την πρακτική εφαρμογή, οδηγώντας σε ένα πραγματικά χρήσιμο και προσαρμοστικό εργαλείο υποστήριξης αποφάσεων που απαντά στις ανάγκες διαφορετικών κοινωνικών ομάδων.

Use Case 1	
Stakeholder – End user	CCAM Industry actors
	technology developers & suppliers, vehicles' manufacturers, service providers
What wants to achieve using the KME	Receive guidance on designing CCAM which meets the needs of different types of users and a checklist of considerations to be considered.
Questions to be answered by the user of KME (User profile)	Question 1 (prompt): Which Country Are you interested in? Options: Greece, Germany, Holland, Great Britain
	Question 2 (prompt): What type of user(s) are you interested in? Options: Elderly, Physical Disabilities etc.
	Questions 3 (prompt): What type of area is the site of interest? Options: Rural, Urban, Suburban
	Question 4 (prompt): Select the income level Options: High income, Medium income, Low income
Recommendations List (output)	• Improve the availability and affordability of transport services. • Improve safety measures and security on public transit to alleviate concerns. • Ensure travel planning tools are more user-friendly and accessible for those less comfortable with digital tools. • Offer training or simplified guidance on using mobility-related apps and digital platforms to enhance the end user's comfort and trust in using digital tools. • Increase public engagement and informational sessions on CCAM technologies, emphasizing safety, reliability, and potential cost-saving benefits, to reduce skepticism. • Equip autonomous vehicles with safety features, such as help buttons and CCTV cameras to address security concerns. Hint: These recommendations are provided to the user in case he selects: Greece, Elderly, Urban and Medium Income.

Εικόνα 3.10: Παράδειγμα Περίπτωσης Χρήσης

Κεφάλαιο 4

Συστήματα που βασίζονται σε κανόνες (Rule Based System)



Εικόνα 4.1: Σύστημα που βασίζεται σε κανόνες

Τα συστήματα που βασίζονται σε κανόνες (rule-based systems) είναι ένα είδος συστήματος τεχνητής νοημοσύνης που λειτουργεί βάσει ενός συνόλου προκαθορισμένων κανόνων (if-then statements) για τη λήψη αποφάσεων ή την επίλυση προβλημάτων. Αυτά τα συστήματα ακολουθούν λογικές διαδικασίες για να καταλήξουν σε ένα αποτέλεσμα ή μια δράση, με βάση την εισερχόμενη πληροφορία και τους προκαθορισμένους κανόνες που έχουν οριστεί από τους προγραμματιστές. [3]

Βασικά χαρακτηριστικά:

1. Κανόνες (Rules): Οι κανόνες είναι οι βασικές μονάδες που καθορίζουν τη συμπεριφορά του συστήματος. Κάθε κανόνας συνήθως έχει τη μορφή "if-then", δηλαδή αν ισχύει μια συνθήκη, τότε εκτελείται μια συγκεκριμένη ενέργεια.
2. Γνώση (Knowledge base): Αποθηκεύει την πληροφορία και τους κανόνες που χρησιμοποιούνται για τη λήψη αποφάσεων.
3. Inference engine (μηχανισμός συμπερασμού): Αυτό το στοιχείο επεξεργάζεται τους κανόνες και την εισερχόμενη πληροφορία για να καταλήξει σε συμπεράσματα ή ενέργειες. Ελέγχει ποιοι κανόνες ισχύουν και ενεργοποιούνται σε κάθε δεδομένη κατάσταση.

Τα rule-based systems είναι ιδιαίτερα χρήσιμα γιατί η λογική του συστήματος είναι πλήρως κατανοητή, ελέγξιμη και επεκτάσιμη. Σε εφαρμογές όπου οι αποφάσεις βασίζονται σε συγκεκριμένα, ξεκάθαρα κριτήρια — όπως η ανάλυση δεδομένων από ερωτηματολόγια ή η παραγωγή προσωποποιημένων συστάσεων — τα rule-based συστήματα προσφέρουν αξιοπιστία, διαφάνεια και δυνατότητα εύκολης προσαρμογής σε νέα δεδομένα ή συνθήκες.

Επιπλέον, τέτοια συστήματα είναι ιδανικά όταν η γνώση του πεδίου μπορεί να περιγραφεί με σαφείς κανόνες και όταν η ανάγκη για ερμηνεύσιμες αποφάσεις είναι υψηλή. Σε σύγκριση με "μαύρα κουτιά" όπως τα νευρωνικά δίκτυα, τα rule-based συστήματα μπορούν να εξηγήσουν γιατί πάρθηκε μια συγκεκριμένη απόφαση, κάτι κρίσιμο σε τομείς όπως η υγεία ή οι μεταφορές[29]. Έτσι, προσφέρουν έναν ισχυρό μηχανισμό υποστήριξης αποφάσεων σε περιβάλλοντα που απαιτούν αξιοπιστία.

4.1 Το εργαλείο Drools

Το Drools[3] είναι ένα από τα πιο ισχυρά και δημοφιλή rule-based συστήματα ανοιχτού κώδικα. Προσφέρει υψηλή ευελιξία και μπορεί να χρησιμοποιηθεί για την ανάπτυξη πολύπλοκων συστημάτων που βασίζονται σε κανόνες. Είναι ιδανικό για την κατασκευή recommendation engines, καθώς μπορείς να ορίσεις διαφορετικούς κανόνες για διαφορετικούς τύπους χρηστών και συνθήκες.

Διαθέτει έναν ισχυρό μηχανισμό συμπερασμάτων (inference engine) και μπορεί να ενσωματωθεί εύκολα με την χρήση της γλώσσας προγραμματισμού Java στην εργασία μας. Το πιο σημαντικό χαρακτηριστικό του εργαλείου Drools που αξιοποιήσαμε και στην δική μας διπλωματική εργασία για να υλοποιήσουμε το Recommendation Engine, είναι ότι μπορούμε να διαχειριστούμε κανόνες για να παρέχουμε διαφορετικές προτάσεις στους πολίτες με βάση τα δεδομένα των χρηστών

και τις προτιμήσεις τους. Έχει εξαιρετική απόδοση για μεγάλα σύνολα δεδομένων και μεγάλη ευελιξία στους κανόνες, χαρακτηριστικά που το κάνουν ιδανικό για χρήση σε εφαρμογές που απαιτούν πολύπλοκη λογική.

Η αρχιτεκτονική του Drools φαίνεται στην Εικόνα 4.2. Ο μηχανισμός κανόνων του Drools, αποθηκεύει, επεξεργάζεται και αξιολογεί δεδομένα για να εκτελέσει τους επιχειρηματικούς κανόνες ή τα μοντέλα αποφάσεων που ορίζουμε. Η βασική λειτουργία του μηχανισμού κανόνων του Drools είναι να ταιριάζει τα εισερχόμενα δεδομένα, ή γεγονότα, με τις συνθήκες των κανόνων και να καθορίζει αν και με ποιον τρόπο θα εκτελεστούν οι κανόνες.

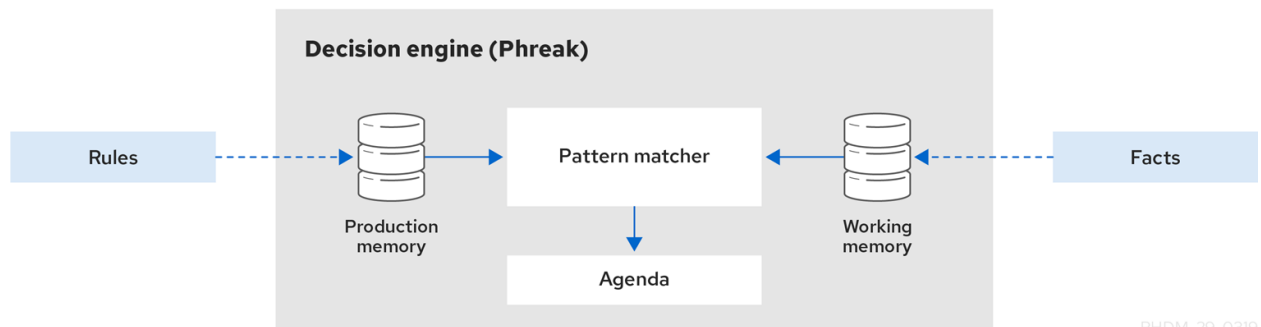
Ο μηχανισμός κανόνων του Drools λειτουργεί χρησιμοποιώντας τα ακόλουθα βασικά συστατικά :

- **Κανόνες:** Επιχειρηματικοί κανόνες που ορίζουμε. Όλοι οι κανόνες πρέπει να περιέχουν τουλάχιστον τις συνθήκες που ενεργοποιούν τον κανόνα και τις ενέργειες που ορίζει ο κανόνας.
- **Γεγονότα:** Δεδομένα που εισέρχονται ή μεταβάλλονται στη μηχανή κανόνων Drools, τα οποία ταυτίζονται με τις συνθήκες των εκάστοτε κανόνων για την εκτέλεση τους.
- **Μνήμη παραγωγής:** Ο χώρος όπου αποθηκεύονται οι κανόνες στη μηχανή κανόνων Drools.
- **Μνήμη εργασίας:** Ο χώρος όπου αποθηκεύονται τα γεγονότα στη μηχανή κανόνων Drools.
- **Agenda (Ημερήσια διάταξη):** Ο χώρος όπου καταχωρούνται και (σε περίπτωση που υπάρχουν) ταξινομούνται οι ενεργοποιημένοι κανόνες, ώστε να εκτελεστούν.

Όταν ένας χρήστης ή ένα αυτοματοποιημένο σύστημα προσθέτει ή ενημερώνει πληροφορίες σχετικά με τους κανόνες στο Drools, αυτές οι πληροφορίες εισάγονται στη μνήμη εργασίας της μηχανής σε μορφή ενός ή περισσότερων γεγονότων. Ο μηχανισμός κανόνων του Drools ταιριάζει αυτά τα γεγονότα με τις συνθήκες των κανόνων που είναι αποθηκευμένοι στη μνήμη παραγωγής για να καθορίσει ποιες εκτελέσεις κανόνων είναι επιλέξιμες. Όταν πληρούνται οι συνθήκες ενός κανόνα, ο μηχανισμός του Drools ενεργοποιεί και καταχωρεί τους κανόνες στην agenda, όπου στη συνέχεια ταξινομεί αυτούς που έχουν προτεραιότητα ή αντικρουόμενους κανόνες σε προετοιμασία για εκτέλεση.

Το Drools χρησιμοποιεί έναν εξειδικευμένο αλγόριθμο αντιστοίχισης γεγονότων και κανόνων, τον Rete Algorithm[30]. Ο Rete Algorithm δημιουργεί ένα δίκτυο από κόμβους (rete network), όπου κάθε κόμβος αναπαριστά είτε ένα απλό είτε ένα σύνθετο μοτίβο αντιστοίχισης όρων. Όταν ένα νέο γεγονός προστίθεται ή ένα υπάρχον γεγονός ενημερώνεται, ο αλγόριθμος αποθηκεύει τα ενδιάμεσα αποτελέσματα και επανεξετάζει μόνο τα στοιχεία που έχουν αλλάξει. Με αυτόν τον τρόπο, αποφεύγει την ανάγκη επανεκτίμησης όλων των κανόνων από την αρχή, επιτυγχάνοντας έτσι σημαντική μείωση του υπολογιστικού κόστους. Επιπλέον, το δίκτυο Rete υποστηρίζει την παρακολούθηση αντικρουόμενων κανόνων και τον καθορισμό προτεραιοτήτων μεταξύ τους, διασφαλίζοντας αξιόπιστη και συνεπή εκτέλεση της λογικής της εφαρμογής.

Εν κατακλείδι, το Drools είναι μια εξαιρετική επιλογή, με υψηλές βαθμολογίες σε κατηγορίες όπως η απόδοση, η υποστήριξη από την κοινότητα των χρηστών, η τεκμηρίωση και η μελλοντική ανάπτυξη. Είναι ένα από τα πιο ευρέως χρησιμοποιούμενα rule engines και μπορεί να διαχειριστεί περίπλοκες επιχειρησιακές λογικές, κάτι που είναι κρίσιμο για το recommendation engine που κατασκευάζουμε. Με βάση τις απαιτήσεις για ένα recommendation engine για CCAM solutions, το Drools είναι μια εξαιρετική επιλογή για την ευελιξία του και την ικανότητά του να διαχειρίζεται περίπλοκους κανόνες για πολλούς τύπους χρηστών.

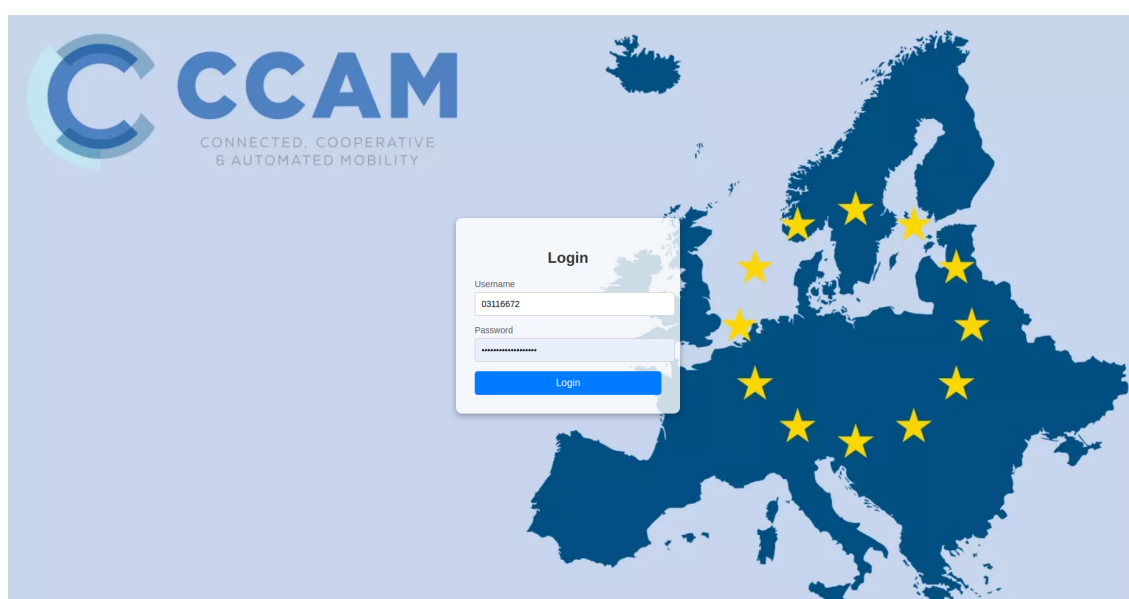


RHDM_29_0319

Εικόνα 4.2: Η αρχιτεκτονική του εργαλείου Drools

4.2 Ενσωμάτωση στο Έργο

Η ενσωμάτωση του Drools στην διπλωματική εργασία, υλοποιήθηκε με σκοπό την αυτοματοποίηση της διαδικασίας λήψης αποφάσεων βάσει των χαρακτηριστικών κάθε χρήστη. Στην εφαρμογή Java Maven που αναπτύχθηκε (Εικόνα 4.3), ο χρήστης αφού συνδεθεί με τα στοιχεία του, καλείται να συμπληρώσει μία φόρμα με τα βασικά δημογραφικά και κοινωνικά χαρακτηριστικά του. Συγκεκριμένα, επιλέγει τη χώρα διαμονής του, τον τύπο χρήστη στον οποίο ανήκει (π.χ. ηλικιωμένος, φοιτητής, άτομο με αναπηρία κ.λπ.), τον τύπο της περιοχής στην οποία ζει (αστική, αγροτική, ημιαστική) και το επίπεδο εισοδήματός του (υψηλό, μεσαίο, χαμηλό) (Εικόνα 4.4).



Εικόνα 4.3: Σελίδα εισόδου στην εφαρμογή που δημιουργήσαμε

CCAM Application

Select Country:

Select User Type:

Select Area:

Select Income Level:

Εικόνα 4.4: Φόρμα συμπλήρωσης στοιχείων χρηστών της εφαρμογής

Αυτά τα δεδομένα μετατρέπονται σε γεγονότα (facts) και εισάγονται στο rule engine του Drools, το οποίο τα καταχωρεί στη λεγόμενη working memory. Παράλληλα, οι κανόνες που έχουν θεσπιστεί για κάθε πιθανό συνδυασμό επιλογών έχουν προαποθηκευτεί στη production memory. Το Drools, χρησιμοποιώντας τη διαδικασία pattern matching, συγκρίνει τις εισαγόμενες τιμές με τις συνθήκες που ορίζονται στους κανόνες, και μόλις εντοπίσει κάποιον κανόνα του οποίου οι προϋποθέσεις πληρούνται, τον ενεργοποιεί και τον καταχωρεί στην agenda για εκτέλεση.

Κάθε ενεργοποιημένος κανόνας επιστρέφει τρεις συστάσεις (recommendation) σχετικές με τις ανάγκες του χρήστη. Οι προτεινόμενες ενέργειες μπορεί να αφορούν την ενίσχυση των μεταφορικών υπηρεσιών, την παροχή εκπαιδευτικών εργαλείων για την εξοικείωση με την τεχνολογία, ή την ενημέρωση για λύσεις κινητικότητας φιλικές προς τις ανάγκες του χρήστη. Η απόκριση είναι άμεση και εξατομικευμένη, με βάση τον προκαθορισμένο κανόνα που έχει σχεδιαστεί για κάθε στοχευμένη κοινωνική ομάδα. Σε περίπτωση που δεν προκύπτουν recommendations, εμφανίζεται στην έξοδο ένα μήνυμα προς τον χρήστη όπου αναφέρει ότι δεν υπάρχουν αρκετά δεδομένα.

Είναι σημαντικό να αναφέρουμε ότι στην δική μας περίπτωση δεν προκύπτουν αντικρουόμενοι κανόνες (Conflicting rules), δεδομένου ότι το σύστημα μας παράγει προκαθορισμένες συστάσεις ανάλογα τα χαρακτηριστικά του εκάστοτε πολίτη που εισέρχεται στην εφαρμογή.

Η χρήση του Drools καθιστά το σύστημα ευέλικτο και επεκτάσιμο. Επιπλέον κανόνες μπορούν να προστεθούν χωρίς να απαιτούνται σημαντικές αλλαγές στον υπόλοιπο κώδικα, γεγονός που διευκολύνει τη συντήρηση και την εξέλιξη της εφαρμογής. Έτσι, επιτυγχάνεται μία πλήρως παραμετροποιήσιμη και επεκτάσιμη λογική λήψης αποφάσεων, η οποία ενισχύει την αξιοπιστία και την ακρίβεια των συστάσεων που παρέχονται στον τελικό χρήστη.

Για την ενσωμάτωση του Drools στην εφαρμογή Java Maven, βασικό βήμα είναι η δημιουργία και η φόρτωση του KIE Container, ο οποίος αποτελεί το σημείο εκκίνησης για την αλληλεπίδραση με τον rule engine. Το KIE (Knowledge Is Everything) API είναι το νέο σύστημα, διαρθρωμένο σε ενότητες (modular), διαχείρισης γνώσης του Drools, μέσω του οποίου γίνεται η φόρτωση και εκτέλεση των κανόνων.

Η αρχική ρύθμιση ξεκινάει με τη δημιουργία ενός αρχείου kmodule.xml, το οποίο περιγράφει ποια rules packages θα χρησιμοποιηθούν. Αυτό το αρχείο τοποθετείται στον φάκελο resources/META-INF/. Στη συνέχεια, μέσα στον Java κώδικα,

φορτώνεται το KIE Container.

Ο KIE Container διαβάζει τους κανόνες από το classpath, όπως έχουν καθοριστεί στο kmodule.xml, και δημιουργεί ένα KieSession, το οποίο είναι υπεύθυνο για την εισαγωγή των facts (δεδομένων του χρήστη), την ενεργοποίηση των κανόνων και την εκτέλεση της λογικής. Στον Κώδικα 4.1 που παραθέτουμε παρακάτω παρουσιάζεται ο κώδικας που υλοποιήσαμε με βάση όσα αναφέραμε παραπάνω.

Κώδικας 4.1: Kie session code

```
1 package com.ccam_solutions.drools_web.service;
2 import com.ccam_solutions.drools_web.model.UserSelection;
3 import org.kie.api.runtime.KieContainer;
4 import org.kie.api.runtime.KieSession;
5 import org.springframework.beans.factory.annotation.Autowired;
6 import org.springframework.stereotype.Service;
7 import java.util.ArrayList;
8 import java.util.List;
9
10 @Service
11
12 public class RdfService {
13
14     @Autowired
15     private KieContainer kieContainer;
16
17     public List<String> processRdfAndGenerateRecommendations \
18     (String country, String userType, String area, \
19     String incomeLevel) {
20         List<String> recommendations = new ArrayList<>();
21
22         UserSelection selection = new UserSelection(country, \
23         userType, area, incomeLevel);
24         KieSession kieSession = kieContainer.newKieSession();
25         kieSession.setGlobal("results", recommendations);
26         kieSession.insert(selection);
27         kieSession.fireAllRules();
28         kieSession.dispose();
29         return recommendations;
30     }
31 }
```

Κατά την εκτέλεση, τα δεδομένα του χρήστη (όπως χώρα, τύπος χρήστη, περιοχή, εισόδημα) εισάγονται στο `kieSession` ως Java αντικείμενα. Μετά την εκτέλεση, η εφαρμογή συλλέγει τις συστάσεις που παρήγαγαν οι κανόνες και τις εμφανίζει στον χρήστη. Αυτή η προσέγγιση διαχωρίζει τη λογική αποφάσεων από τον υπόλοιπο κώδικα και επιτρέπει στο σύστημα να είναι δυναμικό, ευέλικτο και επεκτάσιμο, καθώς μπορεί να δεχτεί νέους κανόνες χωρίς επανεκκίνηση ή επανασυγγραφή του backend κώδικα.

Στην συνέχεια παραθέτουμε ένα μέρος του κώδικα που υλοποιήσαμε για να παράξουμε τις συστάσεις με βάση τον μηχανισμό του Drools. Όπως βλέπουμε στον Κώδικα 4.2, η συγγραφή των κανόνων Drools είναι σε απλοποιημένη γλώσσα κανόνων Drools, εντελώς ξεκομμένη από τα υπόλοιπα κομμάτια κώδικα και από το υπόλοιπο σύστημα.

Κώδικας 4.2: Drool's rules

```
1 package rules;
2 import com.ccam_solutions.drools_web.model.UserSelection;
3 import java.util.List;
4 global List results;
5
6 // Greece
7
8 rule "Elderly Greece Rural Low income"
9
10 when
11     $sel : UserSelection(country == "Greece", userType \
12         == "Elderly", area == "Rural", incomeLevel == "Low income")
13 then
14     results.add("Create free/subsidized transport routes for \
15         medical visits.");
16     results.add("Install sheltered bus stops with benches at \
17         central locations.");
18     results.add("Enhance accessibility with accessible vehicles \
19         (ramps, low-floor buses).");
20 end
21
22 rule "Cognitive Disabilities Greece Urban Medium income"
```

```

23
24 when
25     $sel : UserSelection(country == "Greece", userType == \
26         "Cognitive Disabilities", area == "Urban", \
27         incomeLevel == "Medium income")
28 then
29     results.add("Train drivers on communication with individuals \
30         with cognitive disabilities.");
31     results.add("Use visual alerts and simplified maps in public \
32         transport.");
33     results.add("Implement companion system for daily commuting \
34         assistance.");
35 end
36
37 rule "Digital Vulnerable People Greece Suburban High income"
38
39 when
40     $sel : UserSelection(country == "Greece", userType \
41         == "Digital Vulnerable People", area == "Suburban", \
42         incomeLevel == "High income")
43 then
44     results.add("Offer telephone booking service for \
45         non-digital users.");
46     results.add("Install physical information panels at major \
47         transport stations.");
48     results.add("Provide free digital literacy workshops \
49         for basic mobility apps.");
50 end
51
52 rule "Women Greece Urban Low income"
53
54 when
55     $sel : UserSelection(country == "Greece", userType == "Women \
56         and gender related vulnerabilities", area == "Urban", \
57         incomeLevel == "Low income")
58 then
59     results.add("Increase lighting and install security cameras \
60         at stops and on buses.");
61     results.add("Offer free night-time transport service \
62         for women.");
63     results.add("Create women-only zones on selected lines.");

```

```
64 end
65
66 rule "Young (18-25) Greece Urban Medium income"
67
68 when
69     $sel : UserSelection(country == "Greece", userType == \
70         "Young (18-25)", area == "Urban", \
71         incomeLevel == "Medium income")
72 then
73     results.add("Provide special monthly pass prices for \
74         students and young people.");
75     results.add("Extend night-time routes to entertainment \
76         areas.");
77     results.add("Integrate mobility apps with real-time info \
78         and youth discounts.");
79 end
80
81 //(...
```

4.3 Αποτελέσματα

Όπως προαναφέραμε και παραπάνω, το σύστημα μας, βασισμένο στους κανόνες του Drools, αποκτά σημαντικά πλεονεκτήματα σε υψηλό επίπεδο (high level). Πιο συγκεκριμένα:

- Το σύστημα μας γίνεται πιο γρήγορο και πιο αποδοτικό, αφού σε μεγάλο όγκο δεδομένων αποκρίνεται άμεσα.
- Εξασφαλίζεται η εύκολη προσθήκη/αφαίρεση κανόνων χωρίς αλλαγές στον πυρήνα του κώδικα.
- Έχει την καλύτερη δυνατή οργάνωση της λογικής των αποφάσεων.
- Η χρήση κανόνων επέτρεψε στο σύστημα να παρέχει πιο στοχευμένες και σχετικές συστάσεις, αυξάνοντας την ικανοποίηση των χρηστών.

Ένα πρώτο επίπεδο επικύρωσης (validation) των κανόνων που υλοποιήθηκαν στο Drools πραγματοποιήθηκε μέσω ελέγχων τύπου unit test. Συγκεκριμένα,

δοκιμάστηκαν επιμέρους εκτελέσεις με προκαθορισμένα σενάρια εισόδου ώστε να επιβεβαιωθεί η ορθότητα και η σταθερότητα της λογικής που εφαρμόζεται σε κάθε κανόνα, σύμφωνα με τις βασικές αρχές ελέγχου μονάδων (unit testing) [31]. Οι έλεγχοι αυτοί διασφαλίζουν ότι κάθε τμήμα του συστήματος λειτουργεί όπως προβλέπεται, εντοπίζοντας έγκαιρα πιθανές λογικές ασυμβατότητες.

Παρόλα αυτά, η πλήρης αξιολόγηση (evaluation) και η συγκριτική ανάλυση (benchmarking) της αποδοτικότητας του συστήματος με τη χρήση ποσοτικών κριτηρίων αποτελούν αντικείμενο μελλοντικής εργασίας. Στο πλαίσιο των επόμενων φάσεων του έργου, προβλέπεται η ανάπτυξη πιο εκτεταμένων μεθόδων αξιολόγησης, ώστε να μετρηθεί αντικειμενικά η αποτελεσματικότητα των κανόνων και η συνεισφορά τους στην υποστήριξη των CCAM λύσεων.

Στις Εικόνες 4.5, 4.6, 4.7 παρουσιάζουμε τρία παραδείγματα από το unit test με εξόδους του συστήματος μας σε τυχαίες εισόδους. Όπως φαίνεται και στις εικόνες, το σύστημα βγάζει ως έξοδο τρεις συστάσεις κάθε φορά, οι οποίες ανταποκρίνονται πλήρως στις αντίστοιχες εισόδους (ηλικιωμένο άτομο από την Ελλάδα από αγροτική περιοχή με υψηλό εισόδημα - ποδηλάτης από την Γερμανία και αστική περιοχή με υψηλό εισόδημα - άτομο με ειδικές ανάγκες από την Ολλανδία από ημιαστική περιοχή με χαμηλό εισόδημα).

CCAM Application

Select Country:
Greece

Select User Type:
Elderly

Select Area:
Rural

Select Income Level:
High income

Submit

Generated Recommendations:

Εκπαίδευση οδηγών σε επικοινωνία με άτομα με γνωστικές διαταραχές.
Χρήση οπτικών ειδοποιήσεων και απλοποιημένων χαρτών στα MMM.
Δημιουργία συστήματος συνοδού για καθημερινές μετακινήσεις.

Εικόνα 4.5: 1η Έξοδος της εφαρμογής με μία τυχαία είσοδο

CCAM Application

Select Country:

Germany

Select User Type:

Cyclist

Select Area:

Urban

Select Income Level:

High income

Submit

Generated Recommendations:

Αποκλειστικοί ποδηλατοδρόμοι χωρίς παρεμβολές από πεζούς/οχήματα.
Σταθμοί φόρτισης ηλεκτρικών ποδηλάτων με ηλιακή ενέργεια.
Σύστημα Bike & Ride με έκπτωση 30% για συνδυασμό ποδηλάτου/MMM.

Εικόνα 4.6: 2η Έξοδος της εφαρμογής με μία τυχαία είσοδο

CCAM Application

Select Country:

Holland

Select User Type:

Physical Disabilities

Select Area:

Suburban

Select Income Level:

Low income

Submit

Generated Recommendations:

Αναβάθμιση όλων των λεωφορείων με αυτόματες ράμπες και ηχοϋπερσύνδεση.
Σύστημα προτεραιότητας θέσεων για αναπηρικά αμαξίδια με real-time ειδοποιήσεις.
Εβδομαδιαία δρομολόγια για πρόσβαση σε κέντρα αποκατάστασης και φαρμακεία.

Εικόνα 4.7: 3η Έξοδος της εφαρμογής με μία τυχαία είσοδο

4.4 Συμπεράσματα

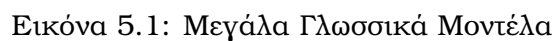
Η χρήση του Drools στην εφαρμογή μας αποτέλεσε μια εξαιρετική επιλογή για την υλοποίηση ενός συστήματος συστάσεων μετακίνησης προσαρμοσμένο στις ανάγκες διαφορετικών ομάδων πολιτών. Το Drools είναι ένα ευέλικτο και ισχυρό rule-based σύστημα, το οποίο μας επέτρεψε να καθορίσουμε με σαφήνεια κανόνες βασισμένους σε παραμέτρους όπως η χώρα, η ηλικιακή ομάδα, ο τύπος χρήστη, η περιοχή κατοικίας και το επίπεδο εισοδήματος. Μέσω των κανόνων αυτών, η εφαρμογή μπορεί να προτείνει στοχευμένες λύσεις και παρεμβάσεις, οι οποίες απαντούν στις μεταφορικές ανάγκες ευάλωτων πληθυσμιακών ομάδων.

Ωστόσο, είναι σημαντικό να τονίσουμε ότι το σύστημα, στην παρούσα του μορφή, δεν είναι πλήρως αξιόπιστο ή ολοκληρωμένο. Οι συστάσεις που παρέχει βασίζονται αποκλειστικά στα διαθέσιμα δεδομένα και στους προϋπάρχοντες κανόνες. Τα δεδομένα που έχουμε στη διάθεσή μας είναι περιορισμένα και, ως εκ τούτου, το επίπεδο εξατομίκευσης ή προσαρμογής των συστάσεων έχει φυσικά όρια.

Μια ενδιαφέρουσα προοπτική για μελλοντική ανάπτυξη του συστήματος είναι η ενσωμάτωση μιας οντολογίας. Η χρήση του εργαλείου Protégé[9] μπορεί να μας βοηθήσει να μοντελοποιήσουμε τη γνώση γύρω από το θέμα της προσβασιμότητας, της μετακίνησης και των διάφορων μεταφορικών μέσων. Μέσω του Protégé, είναι εφικτό να δημιουργηθεί μια οντολογία από την οποία προκύπτει ένα αρχείο RDF. Το Resource Description Framework (RDF) αρχείο λειτουργεί σαν βάση δεδομένων της οντολογίας και επιτρέπει την ανάκτηση πληροφορίας με μεγαλύτερη ευελιξία και ακρίβεια, κάνοντας ερωτήματα μέσω της SPARQL, μίας γλώσσας που ειδικεύεται στην άντληση πληροφοριών από μη σχεσιακές βάσεις δεδομένων. Μια τέτοια βάση θα μπορούσε να βελτιώσει αισθητά την ποιότητα και την αξιοπιστία των συστάσεων, συνδέοντας πολλαπλές πηγές πληροφορίας και εμπλουτίζοντας το σύστημα με εννοιολογική γνώση.

Παρόλο που η ενσωμάτωση του RDF μοντέλου και της οντολογίας με το υπάρχον Drools- σύστημα είναι τεχνικά δυνατή, στην πράξη παρουσιάζει σημαντικές δυσκολίες. Η σύνδεση αυτών των τεχνολογιών, ειδικά μέσα από μια Java Maven εφαρμογή, απαιτεί πολύπλοκη διασύνδεση μεταξύ του semantic web και ενός rule engine. Η διαδικασία αυτή δεν είναι απλή και αποτελεί αντικείμενο μελλοντικής έρευνας. Η διερεύνηση αυτής της κατεύθυνσης, σε συνδυασμό με τον εμπλουτισμό των δεδομένων και την πιθανή χρήση τεχνικών machine learning που θα αναπτύξουμε και στο επόμενο κεφάλαιο, μπορεί να οδηγήσει σε ένα ακόμη πιο ευφύες και προσαρμόσιμο σύστημα συστάσεων στο μέλλον.

Μεγάλα Γλωσσικά Μοντέλα (Large Language Models)



73

κατανόηση και παραγωγή κειμένου από μηχανές αποτελεί κρίσιμο πεδίο έρευνας και ανάπτυξης[32].

Σε αυτό το πλαίσιο αναπτύχθηκαν τα Μεγάλα Γλωσσικά Μοντέλα (Large Language Models - LLMs), τα οποία είναι προηγμένα μοντέλα τεχνητής νοημοσύνης τα οποία έχουν εκπαιδευτεί σε τεράστιες ποσότητες δεδομένων κειμένου με σκοπό να κατανοούν και να παράγουν φυσική γλώσσα. Τα μοντέλα αυτά, όπως το GPT[4] της OpenAI, βασίζονται κυρίως σε αρχιτεκτονικές νευρωνικών δικτύων τύπου Transformer και είναι ικανά να εκτελούν σύνθετες γλωσσικές διεργασίες, όπως μετάφραση, σύνοψη, δημιουργία περιεχομένου, συνομιλία και πολλά άλλα [33].

Η σημασία των LLMs στη σύγχρονη εποχή είναι τεράστια. Αποτελούν την καρδιά πολλών τεχνολογιών που συναντάμε καθημερινά — από έξυπνους ψηφιακούς βοηθούς και chatbots εξυπηρέτησης πελατών, έως εργαλεία δημιουργικής γραφής και υποβοήθησης προγραμματισμού. Επιτρέπουν την αυτοματοποίηση διαδικασιών, την ανάλυση μεγάλου όγκου πληροφοριών, καθώς και την παροχή προσωποποιημένων οδηγιών στους χρήστες.

Συμπερασματικά, τα LLMs έχουν αναδειχθεί ως ένα από τα πιο κρίσιμα εργαλεία για την επόμενη γενιά εφαρμογών τεχνητής νοημοσύνης. Παρόλο που υπάρχουν ακόμη προκλήσεις — όπως η διαχείριση της παραπληροφόρησης και η ηθική χρήση — η συμβολή τους στην έρευνα, στην εκπαίδευση και στην τεχνολογική καινοτομία είναι ανεκτίμητη. Οι δυνατότητές τους συνεχώς εξελίσσονται, και με κατάλληλη αξιοποίηση μπορούν να προσφέρουν τεράστια οφέλη στην κοινωνία[34].

5.1 Ενσωμάτωση LLM στο έργο μας

Μια εναλλακτική και ιδιαίτερα σύγχρονη προσέγγιση στη δημιουργία συστημάτων συστάσεων Recommendation Engine είναι η ενσωμάτωση Large Language Models (LLMs) που έχουν εκπαιδευτεί να παράγουν απαντήσεις με βάση τα δεδομένα που τους παρέχονται. Τα μοντέλα αυτά, τα οποία βασίζονται σε τεχνικές βαθιάς γνώσης (deep learning), μπορούν να προσφέρουν απαντήσεις όχι μόνο σε πλαίσια σύστασης για πολίτες, αλλά και σε πιο απλές αλλά επιστημονικές εφαρμογές, όπως για παράδειγμα η άντληση και σύνοψη πληροφορίας από συλλογές άρθρων, έρευνες ή βάσεις δεδομένων με αρχεία κειμένου. Η ευελιξία των LLMs να λειτουργούν με φυσική γλώσσα τα καθιστά ιδιαίτερα αποτελεσματικά για χρήση σε εργαλεία που στοχεύουν στη διευκόλυνση της πρόσβασης σε πληροφορία.

Στα πλαίσια της παρούσας διπλωματικής εργασίας, η ενσωμάτωση ενός LLM

κρίνεται ως μια ιδιαίτερα υποσχόμενη λύση - μελλοντική επέκταση. Ωστόσο, λόγω των περιορισμών στους διαθέσιμους υπολογιστικούς πόρους, της δυσκολίας στη διαχείριση μεγάλου όγκου δεδομένων (rdf αρχείο), αλλά και της δυνατότητας να δημιουργήσουμε ένα εργαλείο που θα μπορεί να χρησιμοποιηθεί με διάφορους τρόπους (όπως αναφέρουμε και παραπάνω), επιλέξαμε να προχωρήσουμε σε μια πιλοτική υλοποίηση φτιάχνοντας ένα εργαλείο που απαντάει σε ερωτήσεις με βάση τα αρχεία κειμένου (σε μορφή pdf, docx, txt) που του δίνονται ως είσοδο. Με αυτόν τον τρόπο, αξιολογείται η τεχνική βιωσιμότητα της ιδέας και δημιουργείται η βάση ώστε το σύστημα να εξελιχθεί και να ενσωματωθεί μελλοντικά πλήρως είτε στο παρόν έργο είτε σε αντίστοιχες ερευνητικές εργασίες με παρόμοιο αντικείμενο.

Παρακάτω παρουσιάζουμε δύο πολύ σημαντικά εργαλεία για την υλοποίηση του LLM και τον αντίστοιχο κώδικα που δημιουργήσαμε για να υλοποιήσουμε το chatbox.

5.2 Το μοντέλο OpenAI

Η πλατφόρμα OpenAI παρέχει πρόσβαση σε ισχυρά γλωσσικά μοντέλα όπως το GPT-4, καθώς και σε εργαλεία δημιουργίας ενσωμάτωσης(embeddings), τα οποία είναι κρίσιμα για τη μετατροπή φυσικού κειμένου σε πολυδιάστατες αναπαραστάσεις - διανύσματα(vectors) που μπορούμε να τις επεξεργαστούμε υπολογιστικά. Τα embeddings είναι διανύσματα πραγματικών αριθμών (σε αυτήν την περίπτωση 1536 διαστάσεων), που αποτυπώνουν το σημασιολογικό περιεχόμενο ενός κειμένου, επιτρέποντας την ποσοτικοποίηση της ομοιότητας μεταξύ διαφορετικών κειμένων.

Στην εφαρμογή που αναπτύχθηκε, το μοντέλο text-embedding-ada-002 της OpenAI χρησιμοποιείται για την εξαγωγή embeddings από έγγραφα διαφόρων μορφών όπως PDF, DOCX, TXT κ.ά. Αυτά τα διανύσματα στη συνέχεια αποθηκεύονται και αναζητούνται μέσω του Pinecone (που εξηγείται παρακάτω), ώστε να παρέχεται στο GPT-4 το σχετικό συμπραζόμενο περιεχόμενο (context) όταν απαντά σε μία ερώτηση του χρήστη.

Η διαδικασία αυτή καθιστά δυνατή την έξυπνη ανάκτηση πληροφορίας και την παροχή στοχευμένων απαντήσεων, βασισμένων στα αρχεία που έχει προσδιορίσει ο χρήστης. Με αυτόν τον τρόπο, το GPT-4 απαντάει με βάση το περιεχόμενο του τοπικού knowledge base, ενισχύοντας την αξιοπιστία της απόκρισης.

```
1 from openai import OpenAI
2 api_key = "MY OPENAI KEY"
3 client = OpenAI(api_key=api_key) # Create client for OpenAI
4 def ask_gpt4(question, context):
5     prompt = f"Use following documents to answer the \
6     question:\n\n{context}\n\nQuestion: {question}\nAnswer:"
7     response = client.chat.completions.create(
8         model="gpt-4",
9         messages=[
10             {"role": "system", "content": \
11             "You are an assistant who answer from produced \
12             documents."},
13             {"role": "user", "content": prompt}
14         ]
15     )
16     return response.choices[0].message.content
```

Κώδικας 5.1: OpenAI code

5.3 Το εργαλείο Pinecone

Το Pinecone[5] είναι μια βασισμένη σε cloud πλατφόρμα που επιτρέπει την αποθήκευση, διαχείριση και γρήγορη αναζήτηση embeddings χρησιμοποιώντας τεχνικές αναζήτησης ομοιότητας διανυσμάτων (vector similarity search). Η αναζήτηση ομοιότητας διανυσμάτων είναι μια τεχνική που επιτρέπει την εύρεση αντικειμένων που είναι παρόμοια με ένα δεδομένο ερώτημα, συγκρίνοντας τις αριθμητικές αναπαραστάσεις (διανύσματα) τους[35]. Είναι ιδιαίτερα χρήσιμο σε περιπτώσεις όπου απαιτείται αντιστοίχιση φυσικού λόγου με αποθηκευμένα δεδομένα, όπως συμβαίνει και στην δική μας περίπτωση, όπου θέλουμε μέσω ερωτήσεων να αντλούμε απαντήσεις από αποθηκευμένα δεδομένα κειμένου.

Η ροή υλοποιείται ως εξής: Αρχικά, κάθε έγγραφο μετατρέπεται σε embedding μέσω του OpenAI και κατόπιν αποθηκεύεται μαζί με το αρχικό του κείμενο στον δείκτη Pinecone. Κάθε φορά που ένας χρήστης υποβάλλει ένα ερώτημα, αυτό μετατρέπεται και πάλι σε embedding, και με βάση αυτό γίνεται αναζήτηση στους υπάρχοντες δείκτες για να βρεθούν τα πιο συναφή κείμενα με τη βοήθεια του μέτρου συσχέτισης cosine similarity [36].

Το cosine similarity μετρά το συνημίτονο της γωνίας μεταξύ δύο διανυσμάτων και ορίζεται ως εξής:

$$\text{Cosine Similarity} = \cos(\theta) = \frac{\vec{A} \cdot \vec{B}}{\|\vec{A}\| \times \|\vec{B}\|} \quad (5.1)$$

όπου $\vec{A}$ και $\vec{B}$ είναι τα διανύσματα των embeddings, το $\cdot$ συμβολίζει το εσωτερικό γινόμενο και $\|\vec{A}\|$, $\|\vec{B}\|$ το μέτρο (norm) των διανυσμάτων.

Αυτή η δομή επιτρέπει την κατασκευή ενός συστήματος αναζήτησης που είναι ταυτόχρονα γρήγορο και σημασιολογικά ακριβές. Η χρήση του Pinecone απαλλάσσει την εφαρμογή από την ανάγκη τοπικής αποθήκευσης και υλοποίησης περίπλοκων δομών αναζήτησης, επιτρέποντας την ευελιξία της λύσης. Ακόμα είναι πολύ σημαντικό το γεγονός ότι δεν χρειάζεται κάθε φορά που γίνεται μία ερώτηση ή κάθε φορά που τρέχουμε την εφαρμογή μας, να δημιουργούνται και πάλι τα embeddings για έγγραφα που έχουμε ξανά χρησιμοποιήσει στο παρελθόν. Το Pinecone παρέχει μηχανισμό αποθήκευσης των αντίστοιχων embeddings με αποτέλεσμα να βελτιώνει σημαντικά την απόδοση του συστήματος μας.

Τέλος, το γεγονός ότι κάθε embedding συνοδεύεται από τα αρχικά μεταδεδομένα, όπως το πλήρες κείμενο του εγγράφου, μας επιτρέπει να ανακτήσουμε ακριβώς τα κείμενα που σχετίζονται με ένα ερώτημα. Αυτά τα σχετικά κείμενα μπορούν στη συνέχεια να χρησιμοποιηθούν ως συμπραζόμενο υλικό (context) για την παραγωγή απαντήσεων μέσω του LLM που υλοποιήσαμε. Με αυτόν τον τρόπο, υλοποιούμε έναν μηχανισμό Retrieval-Augmented Generation (RAG), όπου το σύστημα πρώτα ανακτά πληροφορίες και μετά χρησιμοποιεί αυτές τις πληροφορίες για να παράγει απαντήσεις. Η χρήση του RAG μοντέλου αυξάνει την ακρίβεια και τη συνάφεια των απαντήσεων, ακόμη και όταν οι διαθέσιμοι υπολογιστικοί πόροι είναι περιορισμένοι.

```
1 from pinecone import Pinecone, ServerlessSpec
2
3 # Index Pinecone and Create Embeddings
4 def generate_embedding(text):
5     response = client.embeddings.create(
6         input=text,
7         model="text-embedding-ada-002"
8     )
9     return response.data[0].embedding
10
11 # Create index Pinecone
12
```

```
13 index_name = "quickstart"
14 pc = Pinecone(api_key="MY KEY")
15 if index_name in pc.list_indexes().names():
16     pc.delete_index(index_name)
17     print(f"Index '{index_name}' removed and \
18         I will creat it again.")
19 if index_name not in pc.list_indexes().names():
20     pc.create_index(name=index_name,
21                     dimension=1536,
22                     metric="cosine",
23                     spec=ServerlessSpec(cloud="aws", region="us-east-1"))
24
25 index = pc.Index(index_name)
26
27 for i, embedding in enumerate(document_embeddings):
28     index.upsert(vectors=[(str(i), embedding, \
29                           {"text": documents[i]})])
30 # Def for searching in Pinecone
31 def search_pinecone(query, index):
32     query_embedding = generate_embedding(query)
33     results = index.query(vector=query_embedding, top_k=1, \
34                           include_metadata=True)
35     return results['matches']
```

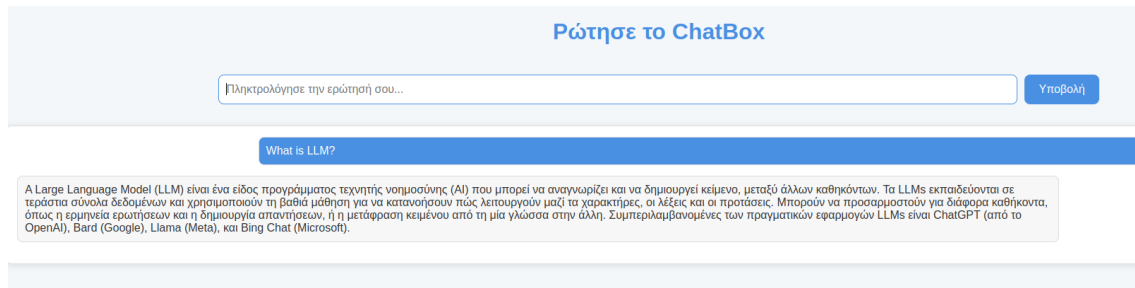
Κώδικας 5.2: Pinecone Code

5.4 Αποτελέσματα

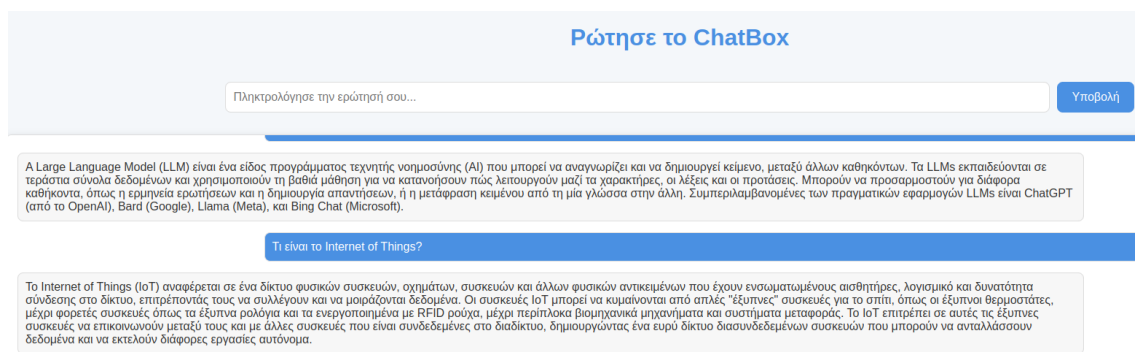
Έλεγχος μοντέλου με δύο διαφορετικές περιπτώσεις δεδομένων μέσω ενός Web App

Μέσω του Google Colab, κάναμε εγκατάσταση των απαραίτητων πόρων και τρέξαμε τον κώδικα που παραθέσαμε και πιο πάνω. Στην πρώτη περίπτωση χρησιμοποιήσαμε ως δεδομένα πέντε αρχεία (τρία σε docx μορφή και δύο σε pdf). Τα κείμενα που αποθηκεύσαμε σε όλα τα αρχεία αφορούν άρθρα τεχνολογικής φύσεως από το διαδίκτυο που αφορούν την τεχνητή νοημοσύνη, το Internet of Things, τα μεγάλα γλωσσικά μοντέλα, την μηχανική μάθηση και τα συστήματα βάσεων δεδομένων. Αφού τρέξαμε την εφαρμογή μας μέσω του flask (βιβλιοθήκη της Python για ανάπτυξη διαδικτυακών εφαρμογών), κάναμε κάποιες ερωτήσεις στο ChatBox για να δούμε ότι αποκρίνεται σωστά και να αποτιμήσουμε την απόδοση

του LLM. Οι απαντήσεις που λάβαμε φαίνονται στις εικόνες 5.2 και 5.3.

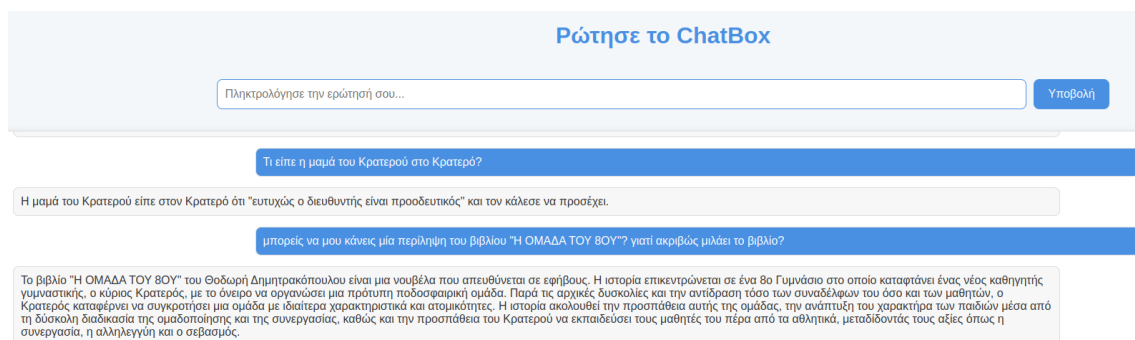


Εικόνα 5.2: ChatBox:Ερώτηση 1



Εικόνα 5.3: ChatBox:Ερώτηση 2

Στην δεύτερη περίπτωση βάλαμε ως είσοδο ένα αρχείο σε docx μορφή που έχει σαν περιεχόμενο μία νουβέλα για παιδιά. Αυτή την φορά κάναμε μία πιο σύνθετη ερώτηση στο ChatBox. Η απάντηση φαίνεται στην εικόνα 5.4.



Εικόνα 5.4: ChatBox:Ερώτηση 3

Μετρικές και Έλεγχος Απόδοσης

Παρακάτω παρουσιάζουμε τις μετρικές που συλλάξαμε για τα δύο παραδείγματα που αναλύουμε πιο πάνω. Στην κάθε περίπτωση φαίνεται το ποσοστό με το οποίο οι ερωτήσεις που κάνουμε ταιριάζουμε με τα embeddings των απαντήσεων. Παρατηρούμε ότι στην πρώτη περίπτωση το αποτέλεσμα είναι πιο επιτυχημένο και πιο ακριβές σε σχέση με την δεύτερη.

1	ID: 3 text: "What is machine learning?" Machine learning (ML) is a branch of artificial intelligence (AI) focused on enabling computers and machines to imitate the way that humans learn, to perform ta...	***
SCORE		
0.9993		
2	ID: 2 text: "What is AI?" Artificial intelligence (AI) is technology that enables computers and machines to simulate human learning, comprehension, problem solving, decision making, creativity and aut...	***
SCORE		
0.9238		
3	ID: 4 text: "What is data science?" Data science combines math and statistics, specialized programming, advanced analytics, artificial intelligence (AI) and machine learning with specific subject m...	***
SCORE		
0.8619		
4	ID: 1 text: "What is a large language model (LLM)?" A large language model (LLM) is a type of artificial intelligence (AI) program that can recognize and generate text, among other tasks. LLMs are trai...	***
SCORE		
0.8584		

Εικόνα 5.5: ChatBox: Έλεγχος Απόδοσης Πρώτη Περίπτωση

5	ID: 0 text: "Θοδωρής Δημητρακόπουλος Η ΟΜΑΔΑ ΤΟΥ 8ΟΥ Η ΝΟΥΒΕΛΑ ΓΙΑ ΕΦΗΒΟΥΣ ΜΗ ΒΡΑΒΕΥΘΕΙΣΑ ΣΤΟ ΔΙΑΓΩΝΙΣΜΟ ΤΗΣ ΓΥΝΑΙΚΕΙΑΣ ΛΟΓΟΤΕΧΝΙΚΗΣ ΣΥΝΤΡΟΦΙΑΣ ISBN: 978-618-8682..."	***
SCORE		
0.6836		

Search used 6 RUs

Εικόνα 5.6: ChatBox: Έλεγχος Απόδοσης Δεύτερη Περίπτωση

5.5 Συμπεράσματα

Συνολικά, η ενσωμάτωση του LLM στην διπλωματική μας εργασία αποδείχθηκε καθοριστική για τη βελτιστοποίηση της διαδικασίας λήψης αποφάσεων και παροχής εξατομικευμένων συστάσεων εάν μελλοντικά δοθούν σαν είσοδο δεδομένα στον τομέα της μετακίνησης με την μορφή κειμένου. Μέσω της ικανότητας των LLMs να αναλύουν τεράστια σύνολα δεδομένων και να δημιουργούν embeddings που αποτυπώνουν το σημασιολογικό περιεχόμενο του κειμένου, καταφέραμε να διαμορφώσουμε ένα σύστημα που αναγνωρίζει και αντιστοιχεί τις ανάγκες των χρηστών με υψηλή ακρίβεια. Το μοντέλο μας, με σκορ περίπου 0.9, υποδεικνύει

ότι οι διανυσματικές αναπαραστάσεις είναι πολύ κοντά στις ιδανικές, προσφέροντας απαντήσεις που είναι συνεπείς, πλήρεις και εξαιρετικά σχετικές με τα εισερχόμενα δεδομένα.

Επιπρόσθετα, η χρήση του LLM επέτρεψε την υλοποίηση ενός μηχανισμού Retrieval-Augmented Generation (RAG) που συνδυάζει την ανάκτηση πληροφορίας από τα αποθηκευμένα έγγραφα με την παραγωγή απαντήσεων υψηλής ποιότητας. Αυτό οδήγησε σε μια πιο διαδραστική και δυναμική προσέγγιση, όπου το σύστημα όχι μόνο απαντά σε ερωτήσεις, αλλά το κάνει με βάση ένα εξειδικευμένο σύνολο εγγράφων που περιέχει κρίσιμες πληροφορίες και συμβουλές, στην προκειμένη περίπτωση για την τεχνητή νοημοσύνη. Με τον τρόπο αυτό, το LLM απέδειξε τη χρησιμότητά του ως εργαλείο που μπορεί να αναβαθμίσει την ακρίβεια και την αξιοπιστία των συστάσεων, συμβάλλοντας ουσιαστικά τόσο στην διπλωματική μας εργασία όσο και σε μελλοντικά projects που μπορεί να ενσωματωθεί.

Ακόμα, όσον αφορά την αξιολόγηση της απόδοσης ενός LLM, απαιτείται η χρήση ποικίλων μετρικών και ποιοτικών παραμέτρων, ώστε να διαπιστωθεί η ακρίβεια, η συνεκτικότητα και η συνολική χρησιμότητά του. Οι μετρικές όπως η cosine similarity στα embeddings μας επιτρέπουν να εξετάσουμε το βαθμό ομοιότητας μεταξύ του ερωτήματος και των απαντήσεων που παράγει το μοντέλο, δίνοντάς μας μια ιδέα για το πόσο καλά το μοντέλο κατανοεί και επεξεργάζεται το φυσικό κείμενο.

Εκτός από τις καθαρά ποσοτικές μετρικές, κρίσιμη είναι η ποιοτική αξιολόγηση της συνοχής και της ακριβείας των παραγόμενων απαντήσεων. Η αξιολόγηση αυτή περιλαμβάνει την ικανότητα του μοντέλου να αντιμετωπίζει ασαφή ή πολύπλοκα ερωτήματα. Πρακτικά, αυτό σημαίνει ότι το μοντέλο πρέπει να ανταποκρίνεται με απαντήσεις που δεν είναι μόνο σωστές αλλά και επαρκώς λεπτομερείς, δίνοντας έμφαση στην ακρίβεια της πληροφορίας. Σημαντικό ρόλο παίζει επίσης η contextual understanding, δηλαδή η δυνατότητα του μοντέλου να κατανοεί το πλαίσιο στο οποίο τίθεται το ερώτημα.

Το γεγονός ότι το δικό μας σύστημα καταφέρνει ένα σκορ κοντά στο 0.9 αποτελεί ένδειξη ότι τα embeddings που παράγονται και η διαδικασία pattern matching είναι σε μεγάλο βαθμό ακριβείς και συμφωνούν με τα επιθυμητά πρότυπα. Ένα σκορ 0.9 δείχνει πως οι απαντήσεις που παράγει το μοντέλο είναι εξαιρετικά προσαρμοσμένες στο περιεχόμενο της ερώτησης, υποδηλώνοντας ότι οι διανυσματικές αναπαραστάσεις του κειμένου μας φτάνουν να συλλαμβάνουν τις ουσιώδεις σημασιολογικές πληροφορίες του. Αυτό οδηγεί σε απαντήσεις που παρουσιάζουν υψηλό βαθμό συνοχής και συμβατότητας με τα εισερχόμενα δεδομένα, πράγμα που είναι καθοριστικό για την αξιοπιστία του συστήματος.

Βέβαια πρέπει να πάρουμε υπόψη ότι οι ερωτήσεις που κάνουμε είναι πολύ απλές και ουσιαστικά υπάρχουν σχεδόν αυτούσιες αναφορές αυτών, στα έγγραφα που βάζουμε ως είσοδο, ζήτημα που διευκολύνει το σύστημα μας να απαντήσει σωστά αφού το μέτρο συσχέτισης(cosine similarity) αυξάνεται πολύ. Την ίδια στιγμή βλέπουμε ότι μια πιο σύνθετη ερώτηση σαν αυτή που κάναμε στο παράδειγμα με το βιβλίο, αν και η απάντηση είναι ικανοποιητική, το σκορ που επιτυγχάνει είναι πολύ μικρότερο, κοντά στο 0.7. Είναι αναγκαίο βέβαια να αναφέρουμε ότι πετυχαίνουμε ικανοποιητικά αποτελέσματα για την περίπτωση χρήσης στα πλαίσια του έργου SINFONICA, με μειωμένους υπολογιστικούς πόρους. Αυτό έχει ως συνέπεια το μοντέλο μας να είναι κλιμακώσιμο (scalable), εύκολα ενσωματώσιμο και άμεσα χρησιμοποιήσιμο (plug-n-play) ακόμα και σε περιβάλλοντα με περιορισμένη υπολογιστική δυνατότητα (capacity), χωρίς να απαιτείται η ύπαρξη ενός συνόλου διακομιστών (clusters of servers), όπως για παράδειγμα σε ένα κέντρο δεδομένων (data center) ενός Πανεπιστημίου.

Κεφάλαιο 6

Συμπεράσματα



Εικόνα 6.1: Συμπεράσματα

Η παρούσα διπλωματική εργασία ανέδειξε τη σημασία του σχεδιασμού εξατομικευμένων εργαλείων υποστήριξης αποφάσεων στον τομέα της μετακίνησης, με στόχο την ενίσχυση της συμπερίληψης και την κάλυψη των αναγκών ευάλωτων κοινωνικών ομάδων. Μέσα από μία ολοκληρωμένη προσέγγιση που συνδύασε ανάλυση δεδομένων, μηχανισμούς βασισμένους σε κανόνες και τεχνολογίες τεχνητής νοημοσύνης, παρουσιάστηκε ένα πρακτικό υπόδειγμα υλοποίησης εργαλείου συστάσεων το οποίο μπορεί να χρησιμοποιηθεί σε περιβάλλοντα CCAM,

συμβάλλοντας στη βελτίωση της ποιότητας ζωής και της προσβασιμότητας στις μεταφορές.

6.1 Η χρησιμότητα των CCAM solutions στον τομέα των μεταφορών

Οι CCAM solutions αποτελούν ένα από τα πιο ελπιδοφόρα εργαλεία για τη διαμόρφωση του μέλλοντος των μεταφορών. Η εφαρμογή τους μπορεί να οδηγήσει σε ασφαλέστερες, αποδοτικότερες και πιο προσβάσιμες μετακινήσεις, ειδικά για πολίτες που αντιμετωπίζουν κοινωνικές, οικονομικές ή φυσικές προκλήσεις. Οι λύσεις αυτές ενισχύουν τη συνεργασία μεταξύ υποδομών, οχημάτων και ανθρώπων και επιτρέπουν τη χρήση τεχνολογιών για την εξυπηρέτηση διαφορετικών πληθυσμιακών ομάδων. Ειδικότερα, η στόχευση των αναγκών ευάλωτων ομάδων, όπως ηλικιωμένοι, άτομα με αναπηρίες, μετανάστες ή φοιτητές, ενισχύει τη διασφάλιση της ισότιμης συμμετοχής όλων στις καθημερινές δραστηριότητες. Η διπλωματική εργασία τοποθετείται σε αυτό το πλαίσιο, υλοποιώντας ένα εργαλείο που υποστηρίζει τον σχεδιασμό και την αξιολόγηση CCAM λύσεων μέσα από την ανάλυση πραγματικών δεδομένων και τη χρήση τεχνολογικών συστημάτων υποστήριξης αποφάσεων.

6.2 Οι λύσεις που προσφέρει το Recommendation Engine Tool

Το εργαλείο που αναπτύχθηκε στο πλαίσιο της διπλωματικής εργασίας αξιοποιεί δεδομένα από ερωτηματολόγια, προκειμένου να κατανοήσει σε βάθος τις ανάγκες των πολιτών ως προς την κινητικότητα. Η ανάλυση και η οπτικοποίηση των δεδομένων ανέδειξε κρίσιμα μοτίβα και διαφοροποιήσεις στις ανάγκες των ομάδων στόχου, κάτι που οδήγησε στον ορισμό περιπτώσεων χρήσης (use cases) για την εφαρμογή. Στη συνέχεια, το εργαλείο επεκτάθηκε με τη χρήση ενός rule-based συστήματος (Drools) το οποίο επιτρέπει την αντιστοίχιση εισόδων (όπως χώρα, τύπος χρήστη, περιοχή, εισόδημα) με προτεινόμενες παρεμβάσεις και συστάσεις. Οι κανόνες αυτοί στηρίζονται τόσο σε ερευνητικά δεδομένα όσο και σε επιμέρους ευρήματα από την οπτικοποίηση των αποτελεσμάτων.

Συμπληρωματικά, εξερευνήθηκε και η ενσωμάτωση μοντέλων τεχνητής νοημοσύνης τύπου LLM, τα οποία επέτρεψαν την υλοποίηση ενός μηχανισμού ερωταποκρίσεων με βάση αποθηκευμένο υλικό. Το σύστημα χρησιμοποιεί μεθοδολογία

Retrieval-Augmented Generation (RAG) και έχει τη δυνατότητα να απαντά σε ερωτήσεις φυσικής γλώσσας, προσφέροντας εξατομικευμένες απαντήσεις με βάση το περιεχόμενο των σχετικών εγγράφων. Η υψηλή απόδοση του μοντέλου (με cosine similarity σκορ κοντά στο 0.9 σε απλές ερωτήσεις) επιβεβαιώνει ότι πρόκειται για μια πολλά υποσχόμενη προσέγγιση, η οποία μπορεί να ενισχύσει περαιτέρω την ευφυΐα και τη χρηστικότητα του εργαλείου.

6.3 Περαιτέρω Έρευνα

Αν και μέσα από την διπλωματική εργασία γίνεται προσπάθεια να δοθούν απαντήσεις σε όλα τα ζητήματα που αφορούν την δημιουργία ενός engine recommendation εργαλείου, προφανώς υπάρχουν αρκετά σημεία που μπορούν να ενισχυθούν μέσω περαιτέρω έρευνας και ανάπτυξης. Ένα βασικό ζήτημα είναι η διεύρυνση της βάσης δεδομένων, ώστε το σύστημα να λαμβάνει πιο πλούσια και ποικιλόμορφα δεδομένα, κάτι που θα επιτρέψει μεγαλύτερη ακρίβεια και εξατομίκευση στις συστάσεις. Επιπλέον, η ενσωμάτωση οντολογιών μέσω εργαλείων όπως το Protégé, μπορεί να προσφέρει έναν ισχυρό μηχανισμό γνώσης, ο οποίος θα ενισχύσει τη σημασιολογική ερμηνεία των εισόδων και θα επιτρέψει πιο ευέλικτη και δυναμική παραγωγή συστάσεων.

Ένα σημαντικό πεδίο για μελλοντική διερεύνηση είναι η τεχνική διασύνδεση μεταξύ συστημάτων βασισμένων σε οντολογίες (RDF) και κανόνων (Drools), αλλά και η ενσωμάτωσή τους με μοντέλα LLM. Παρότι τεχνικά εφικτή, η προσέγγιση αυτή απαιτεί σύνθετους μηχανισμούς διαχείρισης γνώσης και δεδομένων, οι οποίοι ενδεχομένως ξεπερνούν τα διαθέσιμα μέσα και τις δυνατότητες του προπτυχιακού επιπέδου. Ωστόσο, τα αποτελέσματα της παρούσας εργασίας αποδεικνύουν ότι οι τεχνολογίες αυτές μπορούν να συνδυαστούν αρμονικά και να δημιουργήσουν ένα ευφυές, επεκτάσιμο και κοινωνικά ευαίσθητο σύστημα συστάσεων μετακίνησης, το οποίο αξίζει να εξελιχθεί στα πλαίσια και άλλων ερευνητικών ή πιλοτικών έργων.

Βιβλιογραφία

- [1] SINFONICA Project. *SINFONICA Project*. Tech. rep. Horizon Europe Programme, 2022. URL: <https://sinfonica.eu/>.
- [2] CEN-CENELEC. *Standards for CCAM - Cooperative, Connected and Automated Mobility: European framework and initiatives*. Tech. rep. 2023. URL: <https://www.ccam.eu/wp-content/uploads/2023/11/CCAM-SRIA-Update-2023.pdf>.
- [3] Michal Bali. *Drools JBoss Rules 5.0 Developer's Guide*. Packt Publishing, 2009. URL: <https://bista.sites.dmi.unipg.it/didattica/CSP/drools/books/Drools%20Jboss%20Rules%205.0%20Developer's%20Guide.pdf>.
- [4] OpenAI. *GPT-4 Technical Report*. 2023. URL: <https://arxiv.org/abs/2303.08774>.
- [5] CA © Pinecone Systems Inc. | San Francisco. *Pinecone*. URL: <https://www.pinecone.io/>.
- [6] European Commission. *Horizon Europe*. URL: https://research-and-innovation.ec.europa.eu/funding/funding-opportunities/funding-programmes-and-open-calls/horizon-europe_en.
- [7] SINFONICA Project. *D1.1 - Mobility needs and requirements of European citizens*. Tech. rep. Horizon Europe Programme, 2022. URL: <https://sinfonica.eu/download/>.
- [8] SINFONICA Project. *D4.1 - Knowledge Map Creation and System Architecture Specification*. Tech. rep. Grant Agreement 101064988. Horizon Europe Programme, 2022. URL: <https://sinfonica.eu/download/>.
- [9] Stanford University. *Protege*. URL: <https://protege.stanford.edu/>.
- [10] Jake VanderPlas. *Python Data Science Handbook: Essential Tools for Working with Data*. O'Reilly Media, 2016. URL: <https://jakevdp.github.io/PythonDataScienceHandbook/>.
- [11] European Union's Horizon 2020 research. *suaave*. URL: <https://www.suaave.eu/>.

- [12] European Union’s Horizon 2020 research. *show*. URL: <https://show-project.eu/>.
- [13] SINFONICA Consortium. *GUEST Methodology for Innovation Management in Mobility Projects*. Tech. rep. Unpublished internal report. SINFONICA, 2024.
- [14] Eric Ries. *The Lean Startup: How Today’s Entrepreneurs Use Continuous Innovation to Create Radically Successful Businesses*. Crown Business, 2011.
- [15] Richard J. Schönberger. *World Class Manufacturing: The Lessons of Simplicity Applied*. Free Press, 1986.
- [16] Kent Beck et al. *Manifesto for Agile Software Development*. <https://agilemanifesto.org/>. Accessed: 2025-04-27. 2001.
- [17] TInnGO Consortium. *Transport Innovation Gender Observatory (TInnGO)*. <https://www.tinngo.eu/>. Accessed: 2025-04-28. 2019.
- [18] SUMPs-Up Consortium. *SUMPs-Up - Supporting Sustainable Urban Mobility Plans*. <https://sumps-up.eu/>. Accessed: 2025-04-28. 2017.
- [19] DIGNITY Consortium. *DIGNITY - Digital Transport In and For Society*. <https://www.dignity-project.eu/>. Accessed: 2025-04-28. 2020.
- [20] KPMG LLP. *The Long-Awaited Data Revolution Has Begun*. Tech. rep. Accessed: 2025-04-29. KPMG, 2025. URL: <https://kpmg.com/kpmg-us/content/dam/kpmg/pdf/2025/data-revolution-tl-final.pdf>.
- [21] Simon Kemp. *Digital 2024 Global Overview Report*. <https://datareportal.com/reports/digital-2024-global-overview-report>. Accessed: 2025-04-29. 2024.
- [22] DataCamp. *Data Trends and Predictions 2024 Report*. <https://www.datacamp.com/report/data-trends-predictions-2024>. Accessed: 2025-04-29. 2024.
- [23] Foster Provost and Tom Fawcett. *Data Science for Business: What You Need to Know about Data Mining and Data-Analytic Thinking*. O’Reilly Media, 2013. URL: <https://www.amazon.com/Data-Science-Business-Data-Analytic-Thinking/dp/1449361323>.
- [24] Michael L. Waskom. “Seaborn: Statistical data visualization”. In: *Journal of Open Source Software* 6.60 (2021), p. 3021. doi: 10.21105/joss.03021.
- [25] Matplotlib Development Team. *Matplotlib: Visualization with Python*. <https://matplotlib.org/>. 2024.
- [26] The MathWorks, Inc. *MATLAB Release R2023a*. The MathWorks, Inc. 2023. URL: <https://www.mathworks.com/products/matlab.html>.

- [27] Travis E. Oliphant. *A Guide to NumPy*. Trelgol Publishing, 2006. URL: <https://web.mit.edu/dvp/Public/numpybook.pdf>.
- [28] The pandas development team. *pandas-dev/pandas: Pandas 2.0.0*. GitHub. <https://github.com/pandas-dev/pandas>. 2023.
- [29] Oshani Seneviratne, Barbara Capuzzo, and William Van Woensel. “Explainability-Driven Quality Assessment for Rule-Based Systems”. In: *arXiv preprint arXiv:2502.01253* (2025). URL: <https://arxiv.org/abs/2502.01253>.
- [30] Charles L Forgy. “Rete: A fast algorithm for the many pattern/many object pattern match problem”. In: *Artificial Intelligence* 19.1 (1982), pp. 17–37.
- [31] SmartBear Software. *What is Unit Testing?* <https://smartbear.com/learn/automated-testing/what-is-unit-testing/>. Accessed: 2025-04-28. 2024.
- [32] Kevin P. Murphy. *Machine Learning: A Probabilistic Perspective*. MIT Press, 2012.
- [33] Alec Radford et al. *Improving Language Understanding by Generative Pre-Training*. Tech. rep. OpenAI, 2018. URL: https://cdn.openai.com/research-covers/language-unsupervised/language_understanding_paper.pdf.
- [34] Lizhou Fan et al. “A Bibliometric Review of Large Language Models Research from 2017 to 2023”. In: *arXiv preprint arXiv:2304.02020* (2023).
- [35] Pinecone. *Vector Similarity Explained*. <https://www.pinecone.io/learn/vector-similarity/>. Accessed: 2025-04-28. 2023.
- [36] Jiawei Han, Micheline Kamber, and Jian Pei. “Data Mining: Concepts and Techniques”. In: *Elsevier* (2011).

