



ΠΑΝΕΠΙΣΤΗΜΙΟ ΔΥΤΙΚΗΣ ΑΤΤΙΚΗΣ
ΤΜΗΜΑ ΜΗΧΑΝΙΚΩΝ ΠΛΗΡΟΦΟΡΙΚΗΣ ΚΑΙ ΥΠΟΛΟΓΙΣΤΩΝ

ΔΙΠΛΩΜΑΤΙΚΗ ΕΡΓΑΣΙΑ

**Σχεδιασμός και υλοποίηση μοντέλου και εφαρμογής που προτείνει
ενδιαφέροντα σε ταξιδιωτικούς προορισμούς με χρήση μηχανικής μάθησης**

Απόστολος Ι. Συριώδης

A.M. 46683

**Εισηγητές: Ευάγγελος Κοσμάτος, Πανεπιστημιακός Υπότροφος
Γεώργιος Πρεζεράκος, Καθηγητής**

Διπλωματική εργασία υποβληθείσα στο Τμήμα

ΑΙΓΑΛΕΩ, Οκτώβριος/2021



ΠΑΝΕΠΙΣΤΗΜΙΟ ΔΥΤΙΚΗΣ ΑΤΤΙΚΗΣ
ΤΜΗΜΑ ΜΗΧΑΝΙΚΩΝ ΠΛΗΡΟΦΟΡΙΚΗΣ ΚΑΙ ΥΠΟΛΟΓΙΣΤΩΝ

Η παρούσα διπλωματική εργασία παρουσιάστηκε από τον

Απόστολο Ι. Συριώδη

A.M. 46683

Εισηγητές:

Γεώργιος Πρεζεράκος, Καθηγητής

Ευάγγελος Κοσμάτος, Πανεπιστημιακός Υπότροφος

Εξεταστική Επιτροπή:

ΓΕΩΡΓΙΟΣ ΠΡΕΖΕΡΑΚΟΣ Καθηγητής	ΠΑΝΑΓΙΩΤΗΣ ΓΙΑΝΝΑΚΟΠΟΥΛΟΣ Καθηγητής
ΝΙΚΟΛΑΟΣ ΖΑΧΑΡΗΣ Καθηγητής	ΕΥΑΓΓΕΛΟΣ ΚΟΣΜΑΤΟΣ Πανεπιστημιακός Υπότροφος

Ημερομηνία Εξέτασης:

20/10/2021

Τμήμα Μηχανικών Πληροφορικής και Υπολογιστών

ii

ΔΗΛΩΣΗ ΣΥΓΓΡΑΦΕΑ ΔΙΠΛΩΜΑΤΙΚΗΣ ΕΡΓΑΣΙΑΣ

Ο κάτωθι υπογεγραμμένος Απόστολος Συριώδης του Ισίδωρου, με αριθμό μητρώου 46683 φοιτητής του Τμήματος Μηχανικών Πληροφορικής και Υπολογιστών της Σχολής Μηχανικών του Πανεπιστημίου Δυτικής Αττικής, δηλώνω ότι: «Είμαι συγγραφέας αυτής της διπλωματικής εργασίας και ότι κάθε βοήθεια την οποία είχα για την προετοιμασία της, είναι πλήρως αναγνωρισμένη και αναφέρεται στην εργασία. Επίσης, οι όποιες πηγές από τις οποίες έκανα χρήση δεδομένων, ιδεών ή λέξεων, είτε ακριβώς είτε παραφρασμένες, αναφέρονται στο σύνολό τους, με πλήρη αναφορά στους συγγραφείς, τον εκδοτικό οίκο ή το περιοδικό, συμπεριλαμβανομένων και των πηγών που ενδεχομένως χρησιμοποιήθηκαν από το διαδίκτυο. Επίσης, βεβαιώνω ότι αυτή η εργασία έχει συγγραφεί από μένα αποκλειστικά και αποτελεί προϊόν πνευματικής ιδιοκτησίας τόσο δικής μου, όσο και του Ιδρύματος. Παράβαση της ανωτέρω ακαδημαϊκής μου ευθύνης αποτελεί ουσιώδη λόγο για την ανάκληση του πτυχίου μου».

Ο Δηλών



ΠΕΡΙΛΗΨΗ

Σχεδιασμός και υλοποίηση μοντέλου και εφαρμογής που προτείνει ενδιαφέροντα σε ταξιδιωτικούς προορισμούς με χρήση μηχανικής μάθησης

Στην παρούσα διπλωματική εργασία, σχεδιάστηκε και υλοποιήθηκε μοντέλο, το οποίο με βάση το αγαπημένο τουριστικό ενδιαφέρον ενός ταξιδιώτη και τις κριτικές χρηστών στο διαδίκτυο ανάμεσα σε ορισμένες κατηγορίες στις χώρες της Ευρώπης και της Ανατολικής Ασίας, προτείνει άλλα δύο τουριστικά ενδιαφέροντα με χρήση μηχανικής μάθησης.

Παράλληλα, σχεδιάστηκε και υλοποιήθηκε εφαρμογή που χρησιμοποιεί το μοντέλο και δημιουργεί ταξιδιωτικά προγράμματα προσαρμοσμένα στον ταξιδιώτη.

Ο σχεδιασμός και η υλοποίηση του μοντέλου, έγινε με χρήση της γλώσσας προγραμματισμού Python, ενώ η εφαρμογή με χρήση των γλωσσών PHP, JavaScript, CSS και HTML. Τέλος, για την αποθήκευση των δεδομένων χρησιμοποιήθηκε το σύστημα διαχείρισης σχεσιακών βάσεων δεδομένων MySQL.

Λέξεις κλειδιά

Μηχανική μάθηση, μοντέλο, εφαρμογή, τουριστικά ενδιαφέροντα

ABSTRACT

Design and implementation of model and application which proposes interests in travel destinations using machine learning

Apostolos I. Syriodis

Design and implementation of model, which proposes interests in travel destinations across Europe and East Asia, depending on their favorite attractions and reviews from the web using machine learning. Design and implementation of application, which uses the model to create travel plans for the traveler.

The design and implementation of the model was done using the Python programming language, while the implementation was done using PHP, JavaScript, CSS and HTML languages. Finally, a MySQL relational database was used to store the data.

Keywords

Machine learning, model, application, travel destinations

ΚΑΤΑΛΟΓΟΣ ΕΙΚΟΝΩΝ

Εικόνα 2.4.2.1: Παράδειγμα χρήσης μεθόδου Elbow.....	9
Εικόνα 3.1: Αρχιτεκτονική εφαρμογής.....	11
Εικόνα 3.1.1: Η MVC αρχιτεκτονική.....	14
Εικόνα 3.1.2: Επίπεδα τυπικών τεχνολογιών ιστού.....	15
Εικόνα 4.1.1: Διάβασμα .csv αρχείου με χρήση της βιβλιοθήκης pandas.....	17
Εικόνα 4.1.2: Οπτικοποίηση δεδομένων.....	17
Εικόνα 4.1.3: Σύνδεση στη βάση δεδομένων MySQL.....	17
Εικόνα 4.1.4: Υλοποίηση αλγορίθμου NerarestNeighbors.....	18
Εικόνα 4.1.5: Υλοποίηση αλγορίθμου KMeans.....	18
Εικόνα 4.2.1: Δείγμα από τις κριτικές χρηστών για τις χώρες της Ευρώπης.....	19
Εικόνα 4.2.2: Δείγμα από τις κριτικές χρηστών για τις χώρες της Ανατολικής Ασίας.....	19
Εικόνα 4.2.3: Προτάσεις για τις χώρες της Ευρώπης.....	19
Εικόνα 4.2.4: Προτάσεις για τις χώρες της Ανατολικής Ασίας.....	19
Εικόνα 5.1.1: Μέσος όρος κάθε κατηγορίας για τις χώρες της Ανατολικής Ασίας.....	21
Εικόνα 5.1.2: Μέσος όρος κάθε κατηγορίας για τις χώρες της Ευρώπης.....	22
Εικόνα 5.2.1: Χρήση της μεθόδου Elbow στις χώρες της Ανατολικής Ασίας.....	23
Εικόνα 5.2.2: Συσταδοποίηση χρηστών στις χώρες της Ανατολικής Ασίας.....	23
Εικόνα 5.2.3: Χρήση της μεθόδου Elbow στις χώρες της Ευρώπης.....	24
Εικόνα 5.2.4: Συσταδοποίηση χρηστών στις χώρες της Ευρώπης.....	25
Εικόνα 5.3.1: Η αρχική σελίδα της ιστοσελίδας.....	26
Εικόνα 5.3.2: Πρόταση ενδιαφερόντων για τις χώρες της Ανατολικής Ασίας.....	27
Τμήμα Μηχανικών Πληροφορικής και Υπολογιστών	ix

Εικόνα 5.3.3: Πρόταση ενδιαφερόντων για τις χώρες της Ευρώπης.....	27
Εικόνα 6.1.1: Πρώτο σενάριο χρήσης.....	29
Εικόνα 6.1.2: Δεύτερο σενάριο χρήσης.....	29
Εικόνα 6.1.3: Τρίτο σενάριο χρήσης.....	30

ΠΕΡΙΕΧΟΜΕΝΑ

ΠΕΡΙΛΗΨΗ.....	V
ABSTRACT	VII
ΚΑΤΑΛΟΓΟΣ ΕΙΚΟΝΩΝ	IX
ΠΕΡΙΕΧΟΜΕΝΑ.....	XI
1. ΕΙΣΑΓΩΓΗ.....	3
1.1 ΠΕΡΙΓΡΑΦΗ ΤΟΥ ΑΝΤΙΚΕΙΜΕΝΟΥ ΤΗΣ ΔΙΠΛΩΜΑΤΙΚΗΣ ΕΡΓΑΣΙΑΣ	3
1.2 ΠΕΡΙΓΡΑΦΗ ΤΟΥ ΠΡΟΒΛΗΜΑΤΟΣ	3
ΚΕΦΑΛΑΙΟ 2	5
2.1 ΤΙ ΕΙΝΑΙ Η ΜΗΧΑΝΙΚΗ ΜΑΘΗΣΗ;	5
2.2 ΠΩΣ ΛΕΙΤΟΥΡΓΕΙ Η ΜΗΧΑΝΙΚΗ ΜΑΘΗΣΗ	5
2.3 ΜΕΘΟΔΟΙ ΜΗΧΑΝΙΚΗΣ ΜΑΘΗΣΗΣ	6
2.3.1 Εποπτευόμενη μάθηση (supervised machine learning).....	6
2.3.2 Μη εποπτευόμενη μάθηση (Unsupervised machine learning)	7
2.3.3 Ημί-εποπτευόμενη μάθηση (Semi-supervised machine learning).....	7
2.4 ΕΠΙΛΟΓΗ ΜΕΘΟΔΟΥ ΜΑΘΗΣΗΣ ΚΑΙ ΑΛΓΟΡΙΘΜΩΝ	8
2.4.1 Unsupervised Nearest Neighbors	8
2.4.2 KMeans.....	8
ΚΕΦΑΛΑΙΟ 3	11
3.1 ΑΡΧΙΚΤΕΚΤΟΝΙΚΗ ΕΦΑΡΜΟΓΗΣ	11
3.1.1 Η γλώσσα προγραμματισμού Python	12
3.1.2 Το σύστημα διαχείρισης βάσεων δεδομένων MySQL.....	13

3.1.3	Η γλώσσα προγραμματισμού PHP.....	13
3.1.4	Η γλώσσα προγραμματισμού JavaScript	15
3.1.5	Η γλώσσα σήμανσης HTML.....	16
3.1.6	Η γλώσσα κανόνων στυλ CSS	16
ΚΕΦΑΛΑΙΟ 4		17
4.1	ΥΛΟΠΟΙΗΣΗ ΕΦΑΡΜΟΓΗΣ	17
4.2	ΒΑΣΕΙΣ ΔΕΛΟΜΕΝΩΝ	19
ΚΕΦΑΛΑΙΟ 5		21
5.1	ΚΑΤΗΓΟΡΙΕΣ ΚΑΙ ΠΡΟΤΑΣΗ ΕΝΔΙΑΦΕΡΟΝΤΩΝ.....	21
5.1.1	Μέσος όρος κάθε κατηγορίας ενδιαφέροντος για τις χώρες της Ανατολικής Ασίας.....	21
5.1.2	Μέσος όρος κάθε κατηγορίας ενδιαφέροντος για τις χώρες της Ευρώπης	22
5.2	ΣΥΣΤΑΔΟΠΟΙΗΣΗ ΧΡΗΣΤΩΝ.....	22
5.2.1	Συσταδοποίηση χρηστών στις χώρες της Ανατολικής Ασίας.....	23
5.2.2	Συσταδοποίηση χρηστών στις χώρες της Ευρώπης.....	24
5.3	ΠΡΟΤΑΣΗ ΕΝΔΙΑΦΕΡΟΝΤΩΝ.....	25
ΚΕΦΑΛΑΙΟ 6		29
6.1	ΣΕΝΑΡΙΑ ΧΡΗΣΗΣ.....	29
ΚΕΦΑΛΑΙΟ 7		31
7.1	ΕΠΙΛΟΓΟΣ.....	31
ΒΙΒΛΙΟΓΡΑΦΙΑ.....		33

ΠΡΟΛΟΓΟΣ

Η παρούσα διπλωματική εργασία με θέμα «Σχεδιασμός και υλοποίηση μοντέλου και εφαρμογής που προτείνει ενδιαφέροντα σε ταξιδιωτικούς προορισμούς με χρήση μηχανικής μάθησης» πραγματοποιήθηκε στο πλαίσιο ολοκλήρωσης των σπουδών μου στο Τμήμα Μηχανικών Πληροφορικής και Υπολογιστών του Πανεπιστημίου Δυτικής Αττικής.

Στο σημείο αυτό, αισθάνομαι την ανάγκη να εκφράσω τις ειλικρινείς και θερμές ευχαριστίες μου στην οικογένεια μου, που συνέβαλε τα μέγιστα στην εκπλήρωση του ονείρου μου, καθώς και στον κ. Ευάγγελο Κοσμάτο για την επίβλεψη και τη στήριξη που μου παρείχε.

1. ΕΙΣΑΓΩΓΗ

1.1 ΠΕΡΙΓΡΑΦΗ ΤΟΥ ΑΝΤΙΚΕΙΜΕΝΟΥ ΤΗΣ ΔΙΠΛΩΜΑΤΙΚΗΣ ΕΡΓΑΣΙΑΣ

Σκοπός της παρούσας διπλωματικής εργασίας είναι η σχεδίαση και η υλοποίηση ενός μοντέλου, έτσι ώστε με βάση τα αγαπημένα ενδιαφέροντα ενός ταξιδιώτη, και με βάση τις κριτικές χρηστών από το διαδίκτυο για ορισμένες κατηγορίες τουριστικών ενδιαφερόντων στις χώρες της Ευρώπης και της Ανατολικής Ασίας, να γίνεται πρόταση και άλλων τουριστικών ενδιαφερόντων που θα ικανοποιήσουν τους ταξιδιώτες.

Για την επίτευξη αυτού του στόχου, θα δημιουργήσουμε μία εφαρμογή. Αρχικά, τα δεδομένα θα αποθηκεύονται σε μία βάση δεδομένων, έπειτα θα χρησιμοποιούνται αλγόριθμοι μηχανικής μάθησης για την ανάλυση των δεδομένων, η γνώση που θα λαμβάνουμε θα αποθηκεύεται εκ νέου στη βάση δεδομένων και τέλος μέσω μίας διαδραστικής ιστοσελίδας, θα γίνεται διαθέσιμη προς το κοινό η εφαρμογή.

1.2 ΠΕΡΙΓΡΑΦΗ ΤΟΥ ΠΡΟΒΛΗΜΑΤΟΣ

Ο κλάδος του τουρισμού συνεχώς εξελίσσεται, γίνεται πιο κερδοφόρος και πιο ελκυστικός για νέες επενδύσεις. Συνεχώς δημιουργούνται νέοι ταξιδιωτικοί οργανισμοί, οι οποίοι υπόσχονται όλο και καλύτερη εμπειρία για τους ταξιδιώτες.

Οι ταξιδιώτες με τη σειρά τους, έχουν ακόμα μεγαλύτερες απαιτήσεις, συνεπώς ο καλύτερος δυνατός τρόπος για την ικανοποίησή τους, είναι η δημιουργία πακέτων διακοπών ανάλογα με τις προτιμήσεις του εκάστοτε ταξιδιώτη.

Ακολουθώντας την κατάλληλη διαδικασία, θα αντλείται νέα γνώση, η οποία χρησιμοποιούμενη ορθά, θα συμβάλλει στην ικανοποίηση των ταξιδιωτών και κατ' επέκταση την επιθυμία τους, να ανακαλύψουν νέους προορισμούς, υποστηρίζοντας κατ' αυτόν τον τρόπο τον κλάδο του τουρισμού.

ΚΕΦΑΛΑΙΟ 2

2.1 ΤΙ ΕΙΝΑΙ Η ΜΗΧΑΝΙΚΗ ΜΑΘΗΣΗ;

Η μηχανική μάθηση είναι ένας κλάδος της τεχνητής νοημοσύνης και της επιστήμης των υπολογιστών, που επικεντρώνεται στη χρήση δεδομένων και αλγορίθμων, ώστε να μιμηθεί τον τρόπο με τον οποίο μαθαίνουν οι άνθρωποι, βελτιώνοντας σταδιακά την ακρίβεια του.

Η μηχανική μάθηση είναι επίσης, ένα σημαντικό συστατικό του συνεχώς αναπτυσσόμενου πεδίου της επιστήμης των δεδομένων. Μέσω της χρήσης στατιστικών μεθόδων, οι αλγόριθμοι εκπαιδεύονται να κάνουν ταξινομήσεις ή προβλέψεις, αποκαλύπτοντας βασικές γνώσεις στα έργα εξόρυξης δεδομένων. Αυτές οι γνώσεις στη συνέχεια οδηγούν στη λήψη αποφάσεων σε επιχειρήσεις και εφαρμογές. [3]

2.2 ΠΩΣ ΛΕΙΤΟΥΡΓΕΙ Η ΜΗΧΑΝΙΚΗ ΜΑΘΗΣΗ

Το Πανεπιστήμιο Μπέρκλεϋ της Καλιφόρνιας [1], χωρίζει το σύστημα εκμάθησης ενός αλγορίθμου μηχανικής μάθησης σε τρία κύρια μέρη.

1. Διαδικασία απόφασης: Γενικά, οι αλγόριθμοι μηχανικής μάθησης χρησιμοποιούνται για μία πρόβλεψη ή μία ταξινόμηση. Με βάση ορισμένα δεδομένα εισόδου, τα οποία μπορούν να επισημανθούν (labeled data) ή να μην επισημανθούν (unlabeled data), ο αλγόριθμος παράγει μία εκτίμηση σχετικά με ένα μοτίβο που εμφανίζεται στα δεδομένα.
2. Συνάρτηση σφάλματος: Μία συνάρτηση σφάλματος χρησιμεύει για την αξιολόγηση της πρόβλεψης του μοντέλου. Αν υπάρχουν γνωστά παραδείγματα, μία συνάρτηση σφάλματος μπορεί να κάνει μία σύγκριση για να εκτιμήσει την ακρίβεια του μοντέλου.

3. Διαδικασία βελτιστοποίησης του μοντέλου: Αν το μοντέλο μπορεί να ταιριάζει καλύτερα στα σημεία δεδομένων στο training set, τότε τα βάρη προσαρμόζονται, ώστε να μειωθεί η απόκλιση μεταξύ του γνωστού παραδείγματος και της εκτίμησης του μοντέλου. Ο αλγόριθμος επαναλαμβάνει αυτήν τη διαδικασία αξιολόγησης και βελτιστοποίησης, ενημερώνοντας αυτόνομα τα βάρη μέχρι να επιτευχθεί ένα όριο ακριβείας.

2.3 ΜΕΘΟΔΟΙ ΜΗΧΑΝΙΚΗΣ ΜΑΘΗΣΗΣ

Υπάρχουν τρεις διαφορετικές μέθοδοι μηχανικής μάθησης.

2.3.1 Εποπτευόμενη μάθηση (supervised machine learning)

Η εποπτευόμενη μηχανική μάθηση ορίζεται από τη χρήση συνόλων δεδομένων με ετικέτες (labeled data) για την εκπαίδευση αλγορίθμων (training) που ταξινομούν δεδομένα ή προβλέπουν με ακρίβεια αποτελέσματα. Καθώς τα δεδομένα εισόδου τροφοδοτούνται στο μοντέλο, προσαρμόζουν τα βάρη του, έως ότου το μοντέλο φτάσει στο επιθυμητό σημείο. Αυτό συμβαίνει ως μέρος της διαδικασίας διασταυρούμενης επικύρωσης, διασφαλίζοντας ότι το μοντέλο αποφεύγει την υπερβολική προσαρμογή (overfitting) ή την υπολειτουργία (underfitting).

Η εποπτευόμενη μηχανική μάθηση βοηθά τους οργανισμούς να επιλύουν διάφορα προβλήματα πραγματικού κόσμου σε κλίμακα, όπως για παράδειγμα η ταξινόμηση των ανεπιθύμητων μηνυμάτων σε ξεχωριστό φάκελο από τα εισερχόμενα.

Ορισμένες μέθοδοι και αλγόριθμοι που χρησιμοποιούνται στην εποπτευόμενη μηχανική μάθηση περιλαμβάνουν νευρωνικά δίκτυα, naïve bayes, linear regression, logistic regression, random forest, SVM, K-Nearest Neighbors κ.α.

2.3.2 Μη εποπτευόμενη μάθηση (Unsupervised machine learning)

Η μη εποπτευόμενη μηχανική μάθηση χρησιμοποιεί αλγόριθμους για την ανάλυση και την ομαδοποίηση συνόλων δεδομένων χωρίς ετικέτα (unlabeled data). Αυτοί οι αλγόριθμοι ανακαλύπτουν μοτίβα ή ομαδοποιήσεις δεδομένων χωρίς την ανάγκη ανθρώπινης παρέμβασης.

Η ικανότητα αυτής της μεθόδου της μηχανικής μάθησης να ανακαλύπτει ομοιότητες και διαφορές, την καθιστά ως ιδανική λύση για διερευνητικές αναλύσεις δεδομένων, στρατηγικές διασταυρούμενων πωλήσεων, τμηματοποίηση πελατών, αναγνώριση εικόνας και προτύπων.

Μερικοί από τους αλγόριθμους που χρησιμοποιούνται είναι ο αλγόριθμος KMeans clustering, hierarchical clustering κ.α.

2.3.3 Ημί-εποπτευόμενη μάθηση (Semi-supervised machine learning)

Η ημί-εποπτευόμενη μηχανική μάθηση προσφέρει ένα μέσο μεταξύ της εποπτευόμενης και της μη εποπτευόμενης μηχανικής μάθησης. Κατά τη διάρκεια της εκπαίδευσης, χρησιμοποιεί ένα μικρότερο σύνολο δεδομένων με ετικέτα, για να καθοδηγήσει την ταξινόμηση και την εξαγωγή χαρακτηριστικών από ένα μεγαλύτερο σύνολο χωρίς ετικέτα.

Η ημί-εποπτευόμενη μηχανική μάθηση, μπορεί να λύσει το πρόβλημα τη μη ύπαρξης αρκετών δεδομένων με ετικέτα για την εκπαίδευση ενός εποπτευόμενου αλγορίθμου μηχανικής μάθησης.

2.4 ΕΠΙΛΟΓΗ ΜΕΘΟΔΟΥ ΜΑΘΗΣΗΣ ΚΑΙ ΑΛΓΟΡΙΘΜΩΝ

Καθώς τα δεδομένα που χρησιμοποιούμε είναι χωρίς ετικέτα, θα προχωρήσουμε με χρήση της μεθόδου της μη-εποπτευόμενης μηχανικής μάθησης. Οι αλγόριθμοι που θα χρησιμοποιήσουμε είναι δύο, ο Unsupervised Nearest Neighbors και ο KMeans.

2.4.1 Unsupervised Nearest Neighbors

Είναι μία παραλλαγή του αλγορίθμου K-Nearest Neighbors (KNN), για δεδομένα χωρίς ετικέτα. Αυτός ο αλγόριθμος υποθέτει ότι παρόμοια δεδομένα υπάρχουν σε κοντινή απόσταση μεταξύ τους. Η απόσταση μπορεί να είναι οποιοδήποτε μετρικό μέτρο, αν και η πιο συνήθης είναι η ευκλείδεια. [5]

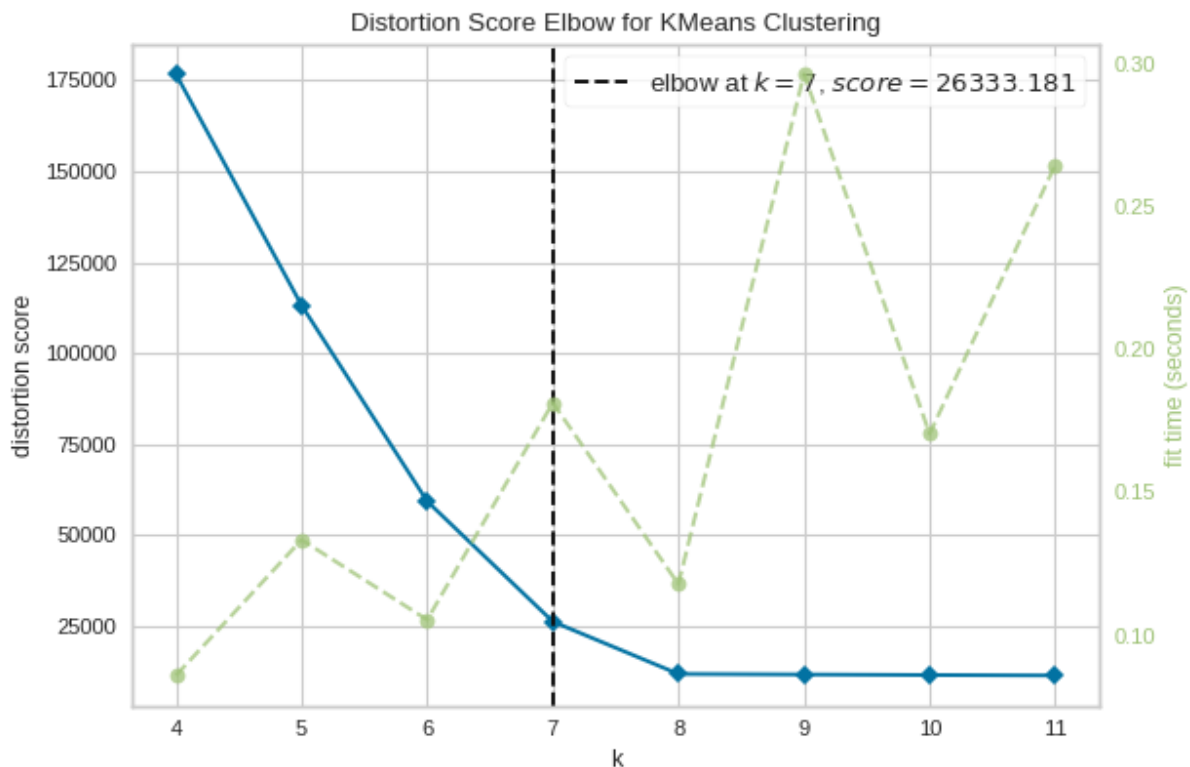
2.4.2 KMeans

Σε αυτόν τον αλγόριθμο γίνεται ομαδοποίηση παρόμοιων δεδομένων, με βάση κάποια μοτίβα που ακολουθούν αυτά. Για να επιτευχθεί ο στόχος αυτός, χρειάζεται να προκαθορίσουμε τον αριθμό των συστάδων (clusters) k που επιθυμούμε. Για την επεξεργασία των δεδομένων, ο αλγόριθμος αυτός ξεκινά με μία πρώτη ομάδα τυχαία επιλεγμένων κέντρων (centroids), τα οποία χρησιμοποιούνται ως σημεία έναρξης για κάθε συστάδα και στη συνέχεια εκτελεί επαναλαμβανόμενους υπολογισμούς για τη βελτιστοποίηση των θέσεων των κέντρων. Διακόπτει

τη λειτουργία του όταν τα κέντρα έχουν σταθεροποιηθεί ή ο προκαθορισμένος αριθμός των επαναλήψεων έχει επιτευχθεί. [4] [6]

2.4.2.1 Μέθοδος Elbow

Στη συσταδοποίηση, η μέθοδος Elbow (αγκώνα) είναι μία ευρετική μέθοδος που χρησιμοποιείται για τον προσδιορισμό του αριθμού k συστάδων σε ένα σύνολο δεδομένων. Το γράφημα που προκύπτει μοιάζει με βραχίονα και το σημείο που εμφανίζει ο «αγκώνας» (το σημείο καμπής στην καμπύλη) είναι μία καλή ένδειξη ότι το μοντέλο ταιριάζει καλύτερο σε εκείνο το σημείο, όπως φαίνεται και στο παρακάτω παράδειγμα. [18]

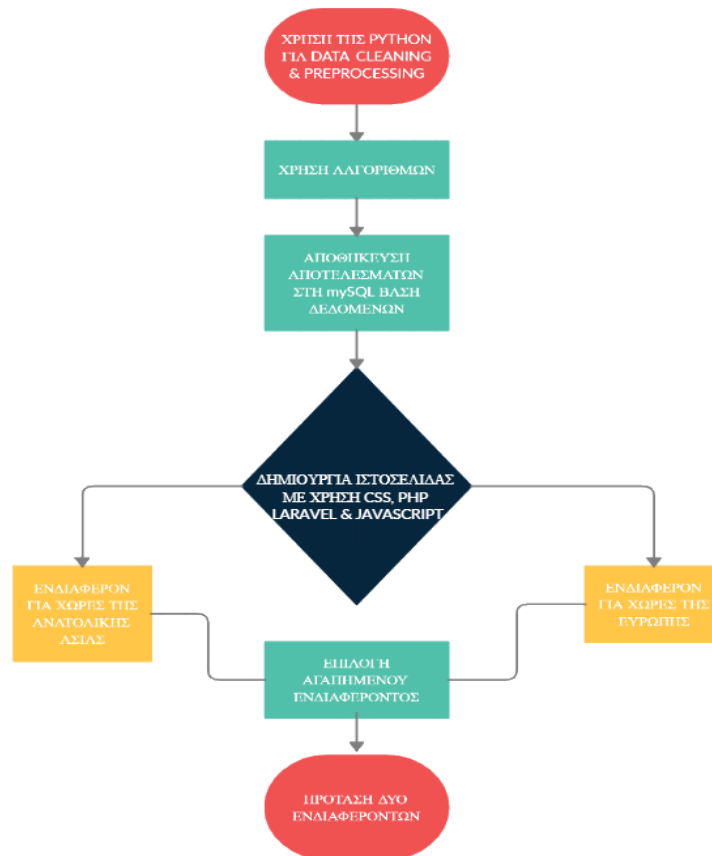


Εικόνα 2.4.2.1. Παράδειγμα μεθόδου Elbow

ΚΕΦΑΛΑΙΟ 3

3.1 ΑΡΧΙΚΤΕΚΤΟΝΙΚΗ ΕΦΑΡΜΟΓΗΣ

Η αρχιτεκτονική της εφαρμογής, όπως εμφανίζεται και στην παρακάτω εικόνα, χρησιμοποιεί τη γλώσσα προγραμματισμού Python για την ανάλυση των δεδομένων μέσω κατάλληλων αλγορίθμων. Τα νέα δεδομένα που προκύπτουν, αποθηκεύονται σε μία MySQL βάση δεδομένων. Στη συνέχεια, χρησιμοποιείται μία ιστοσελίδα, η οποία αποσκοπεί στην πρόταση ενδιαφερόντων, με βάση το αγαπημένο τουριστικό ενδιαφέρον των ταξιδιωτών. Η ιστοσελίδα έχει δημιουργηθεί με χρήση των γλωσσών προγραμματισμού PHP και JavaScript, καθώς και των γλωσσών σήμανσης & κανόνων στυλ HTML και CSS αντίστοιχα.



Εικόνα 3.1. Αρχιτεκτονική εφαρμογής

3.1.1 Η γλώσσα προγραμματισμού Python

Η Python είναι μια υψηλού επιπέδου γλώσσα προγραμματισμού, η οποία δημιουργήθηκε από το 1990 από τον Ολλανδό Guido van Rossum. Ο κύριος στόχος της είναι η αναγνωσιμότητα του κώδικα της, η ευκολία χρήσης της και το συντακτικό της, επιτρέποντας στους προγραμματιστές να εκφράσουν έννοιες σε λιγότερες γραμμές κώδικα, απ' ό,τι θα ήταν δυνατόν σε γλώσσες όπως η C++ και η Java.

Διακρίνεται λόγω των πολλών βιβλιοθηκών που διευκολύνουν ιδιαίτερα αρκετές συνηθισμένες εργασίες και για την ταχύτητα εκμάθησης.

Οι διερμηνείς της Python είναι διαθέσιμοι για εγκατάσταση σε πολλά λειτουργικά συστήματα επιτρέποντας στην Python την εκτέλεση κώδικα σε ευρεία γκάμα λειτουργικών συστημάτων. [7]

Η Python 2.0 κυκλοφόρησε στις 16 Οκτωβρίου του 2000. Στις 3 Δεκεμβρίου 2008 κυκλοφόρησε η έκδοση 3.0. Πολλά από τα καινούργια χαρακτηριστικά αυτής της έκδοσης έχουν μεταφερθεί στις εκδόσεις 2.6 και 2.7 που είναι προς τα πίσω συμβατές. Η Python 3 είναι ιστορικά η πρώτη γλώσσα προγραμματισμού που σπάει την προς τα πίσω συμβατότητα με προηγούμενες εκδόσεις, ώστε να διορθωθούν κάποια λάθη που υπήρχαν σε προγενέστερες εκδόσεις και να καταστεί με σαφήνεια ο απλός τρόπος που μπορούν να γίνουν κάποια πράγματα.

Η έκδοση που χρησιμοποιήθηκε στην παρούσα διπλωματική εργασία, είναι η Python 3.0.

3.1.2 Το σύστημα διαχείρισης βάσεων δεδομένων MySQL

Η MySQL είναι το πιο δημοφιλές σύστημα διαχείρισης βάσεων δεδομένων SQL ανοικτού κώδικα. Αναπτύσσεται, διανέμεται και υποστηρίζεται από την Oracle Corporation. Η αρχική έκδοση έγινε διαθέσιμη στις 23 Μαΐου 1995.

Ως βάση δεδομένων ορίζεται μία δομημένη συλλογή δεδομένων. Οι βάσεις δεδομένων MySQL είναι σχεσιακές, δηλαδή ορίζονται από κανόνες που διέπουν τις σχέσεις μεταξύ διαφορετικών πεδίων δεδομένων, όπως ένα προς ένα, ένα προς πολλά, μοναδικά, απαιτούμενα ή προαιρετικά και «δείκτες» μεταξύ διαφορετικών πινάκων. Η βάση δεδομένων επιβάλλει αυτούς τους κανόνες, έτσι ώστε με μια καλά σχεδιασμένη βάση δεδομένων, η εφαρμογή να μη χρησιμοποιεί ποτέ ασυνεπή, διπλά, ορφανά, ξεπερασμένα ή ελλείποντα δεδομένα. [12]

3.1.3 Η γλώσσα προγραμματισμού PHP

Η PHP είναι η πιο δημοφιλής server-side scripting γλώσσα προγραμματισμού. Σχεδιάστηκε το 1994 από τον Δανικής – Καναδικής καταγωγής Rasmus Lerdorf. Το ακρωνύμιο PHP σημαίνει HyperText Preprocessor, αλλάζοντας από το αρχικό όνομα Personal Home Page.

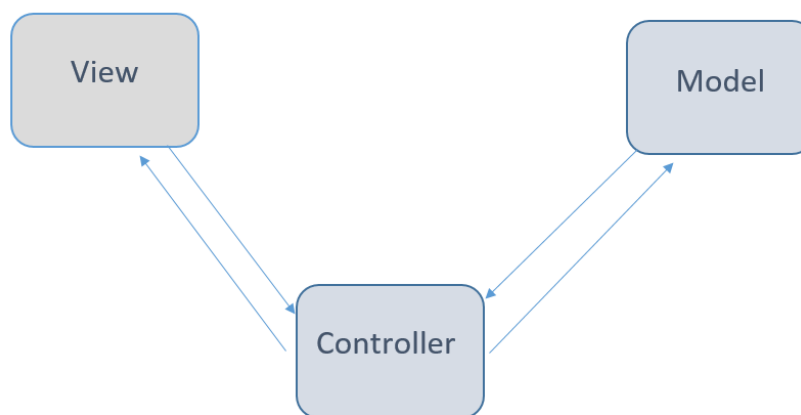
Η PHP 1.0 κυκλοφόρησε στις 8 Ιουνίου 1995. Αρχικά δεν προοριζόταν να είναι γλώσσα προγραμματισμού, αλλά λόγω της δημοφιλίας της, δημιουργήθηκε μία νέα ομάδα ανάπτυξης στα μέσα του 1997, ώστε να κινηθεί προς αυτήν την κατεύθυνση. Το Νοέμβριο του 1997, κυκλοφόρησε μία λειτουργική έκδοσης της γλώσσας με όνομα PHP/FI 2. [17]

Η έκδοση που χρησιμοποιήθηκε στην παρούσα διπλωματική εργασία, είναι η PHP 8.0.

3.1.3.1 PHP Laravel

Πιο συγκεκριμένα, χρησιμοποιήθηκε το ανοικτού κώδικα PHP web framework Laravel. Δημιουργήθηκε από τον Taylor Otwell στις 9 Ιουνίου 2011 για την ανάπτυξη διαδικτυακών εφαρμογών, ακολουθώντας την αρχιτεκτονική model-view-controller (MVC). [8] [9]

Η αρχιτεκτονική model-view-controller είναι ένα σχέδιο σχεδίασης λογισμικού, που χρησιμοποιείται συνήθως για την ανάπτυξη διεπαφών χρήστη που χωρίζουν τη σχετική λογική του προγράμματος σε τρία διασυνδεδεμένα μέρη, όπως παρουσιάζεται στην εικόνα 3.2. [10] [11]

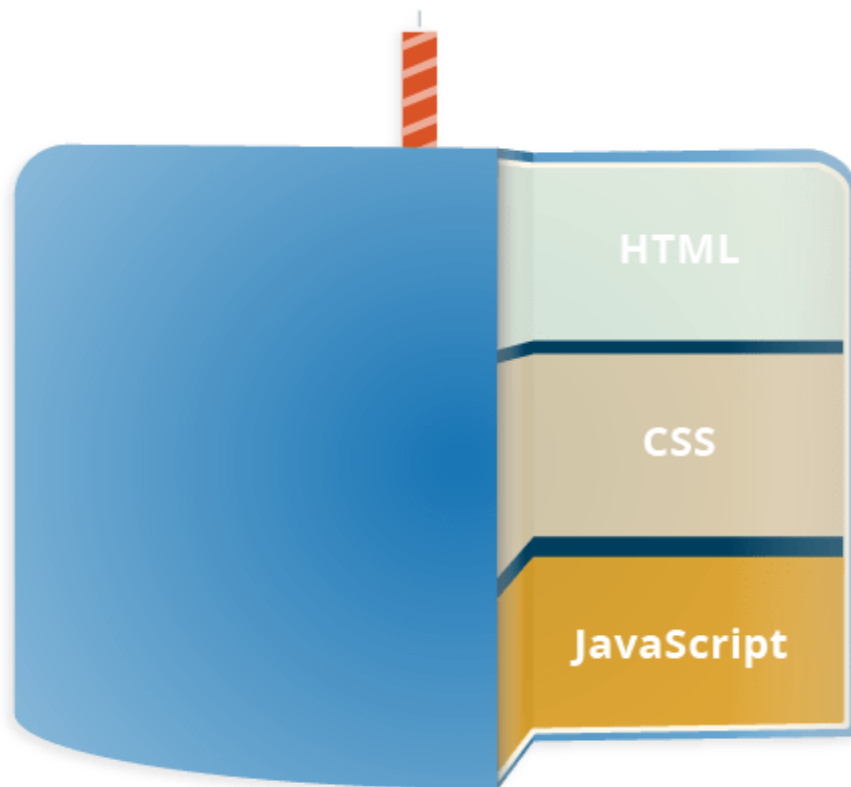


Εικόνα 3.1.1. Η MVC αρχιτεκτονική

3.1.4 Η γλώσσα προγραμματισμού JavaScript

Η JavaScript είναι μία γλώσσα προγραμματισμού που βασίζεται στα πρωτότυπα (prototype-based), είναι δυναμική, με ασθενείς τύπους και έχει συναρτήσεις ως αντικείμενα πρώτης τάξης. Είναι βασισμένη σε διαφορετικά προγραμματιστικά παραδείγματα, υποστηρίζοντας αντικειμενοστραφές, προστακτικό και συναρτησιακό στυλ προγραμματισμού. Το πρότυπο της γλώσσας κατά τον οργανισμό τυποποίησης ECMA ονομάζεται ECMAScript [2]. Έκανε την εμφάνιση της στις 4 Δεκεμβρίου 1995.

Είναι το τρίτο στρώμα των επιπέδων των τυπικών τεχνολογιών ιστού, κάτω από HTML και CSS αντίστοιχα. [13] [14]



Εικόνα 3.1.2. Επίπεδα τυπικών τεχνολογιών ιστού

3.1.4.1 Η Vue.js

Χρησιμοποιήθηκε το ανοικτού κώδικα JavaScript web framework Vue.js. Δημοσιεύθηκε από τον Evan You το Φεβρουάριο του 2014 για τη δημιουργία διεπαφών χρήστη (UI) και εφαρμογών μίας σελίδας. Αυτό που ξεχωρίζει τη Vue.js από άλλα JavaScript web frameworks, είναι η ευκολία στην επέκταση των λειτουργικοτήτων. [19]

3.1.5 Η γλώσσα σήμανσης HTML

Η HTML είναι η γλώσσα σήμανσης που χρησιμοποιείται για να δομήσουμε και να δώσουμε νόημα στο περιεχόμενο του ιστού μας, για παράδειγμα καθορίζοντας παραγράφους, επικεφαλίδες, εικόνες κ.α. Η πρώτη έκδοση της κυκλοφόρησε το 1993 από τον Τιμ Μπέρνερς-Λι. [21]

3.1.6 Η γλώσσα κανόνων στυλ CSS

Η CSS είναι μια γλώσσα κανόνων στυλ, που χρησιμοποιούμε για να εφαρμόζουμε στυλ στο περιεχόμενο HTML μας, ορίζοντας χρώματα, γραμματοσειρές φόντου, εκθέτοντας το περιεχόμενό μας σε πολλές στήλες κ.α. Η πρώτη έκδοση της κυκλοφόρησε στις 17 Δεκεμβρίου 1996 από την W3C. [20]

ΚΕΦΑΛΑΙΟ 4

4.1 ΥΛΟΠΟΙΗΣΗ ΕΦΑΡΜΟΓΗΣ

Αρχικά, χρησιμοποιήσαμε τη βιβλιοθήκη pandas [22] για το διάβασμα των δεδομένων από τα .csv αρχεία που αντλήσαμε από το UCI Machine Learning Repository [15][16] και για τη δημιουργία DataFrames. Στιγμιότυπο του πηγαίου κώδικα που υλοποιεί τα παραπάνω, απεικονίζεται στην εικόνα 4.1.1.

```
df = pd.read_csv('/home/apsyriodis/PADA/Diplomatiki/csv/europe_reviews.csv')
```

Εικόνα 4.1.1 Διάβασμα .csv αρχείου με χρήση της βιβλιοθήκης pandas

Για τη διαχείριση των DataFrames χρησιμοποιήσαμε τη βιβλιοθήκη NumPy [23], καθώς επίσης και τη βιβλιοθήκη matplotlib.pyplot [24] για τη οπτικοποίηση γραφικών παραστάσεων.

```
plt.figure(figsize=(10,7))
plt.barh(np.arange(len(column_names[1:])), avg_rating.values, align='center', alpha=0.5)
plt.yticks(np.arange(len(column_names[1:])), avg_rating.index)
plt.tick_params(axis='x')
plt.tick_params(axis='y')
plt.xlabel('Average Rating')
plt.title('Average Rating Per Category')
```

Εικόνα 4.1.2 Οπτικοποίηση δεδομένων

Επίσης, για την σύνδεση μεταξύ της Python και της βάσης δεδομένων MySQL έγινε χρήση της βιβλιοθήκης mysql.connector [25].

```
connection = mysql.connector.connect(
    host="localhost",
    user="root",
    password="XXXXXXXXXX",
    database="python_ml",
    auth_plugin='mysql_native_password'
)
```

```
cursor = connection.cursor()
```

Εικόνα 4.1.3 Σύνδεση στη βάση δεδομένων MySQL

Στη συνέχεια, για την υλοποίηση του Συστήματος Προτάσεων με χρήση του αλγορίθμου NearestNeighbors, έγινε χρήση της βιβλιοθήκης `scipy.sparse.csr_matrix` [26] για τη μετατροπή του DataFrame από πυκνό σε αραιό και της βιβλιοθήκης `sklearn.neighbors.NearestNeighbors` [27] για την αναζήτηση πλησιέστερων γειτόνων. Ο αλγόριθμος αναζήτησης που χρησιμοποιήσαμε για τον υπολογισμό των πλησιέστερων γειτόνων ήταν ο «brute», δηλαδή «ωμής βίας».

```
df_pivot_matrix = csr_matrix(df_pivot.values)
model = NearestNeighbors(metric = 'cosine', algorithm = 'brute')
model.fit(df_pivot_matrix)
```

Εικόνα 4.1.4 Υλοποίηση αλγορίθμου NearestNeighbors

Για στη συσταδοποίηση των χρηστών, υλοποιήθηκε ο αλγόριθμος KMeans με χρήση της βιβλιοθήκης `sklearn.cluster.KMeans` [29]. Για να κανονικοποιήσουμε τα δεδομένα αφαιρώντας το μέσο όρο και κλιμακώνοντας της διακύμανση, χρησιμοποιήσαμε τη βιβλιοθήκη `sklearn.preprocessing.StandardScaler` [28]. Έπειτα, προχωρήσαμε στη χρήση της μεθόδου Ανάλυσης Κύριων Συνιστωσών (PCA), που αποτελεί μία γραμμική μέθοδο συμπίεσης δεδομένων. Η βιβλιοθήκη που χρησιμοποιήσαμε σε αυτήν την περίπτωση ήταν η `sklearn.decomposition.PCA` [30]. Τέλος, για την οπτικοποίηση δεδομένων χρησιμοποιήσαμε επίσης τις βιβλιοθήκες `matplotlib.ticker` [31] και `seaborn` [32].

```
n_clusters=4

pca = PCA(n_components = 2,random_state=42)
df_PCA_scaled = pca.fit_transform(df)

model = KMeans(n_clusters,random_state=42)
cluster_column = model.fit_predict(df_PCA_scaled)
new_column = model.fit_predict(df_PCA_scaled)
cluster = model.labels_
```

Εικόνα 4.1.5 Υλοποίηση αλγορίθμου KMeans

4.2 ΒΑΣΕΙΣ ΔΕΔΟΜΕΝΩΝ

Οι βάσεις δεδομένων που δημιουργήθηκαν είναι τέσσερις, δύο για κάθε ήπειρο. Δύο βάσεις δεδομένων που περιέχουν τις κριτικές των χρηστών και δύο βάσεις δεδομένων που περιέχουν τις προτάσεις για κάθε κατηγορία ταξιδιωτικού ενδιαφέροντος, όπως παρουσιάζονται στις παρακάτω εικόνες. Για τη δημιουργία των βάσεων δεδομένων που αφορούν τις προτάσεις, βλ. εικόνες 4.2.3 και 4.2.4, πρέπει να ανοίξουμε ένα τερματικό στο φάκελο `fullstack_website` και να πληκτρολογήσουμε την εντολή `php artisan migrate`.

user_id	churches	resorts	beaches	parks	theatres	museums	malls	zoo	restaurants	pubs_bars	local_services	burger_pizza_shops	hotels_other_lodgings	juice_bars	art_galleries	dance_clubs	swimming_pool	gyms	bakeries	beauty_spas	cafes	view_points	monuments	gardens
User 1	0	0	3.63	3.65	5	2.92	5	2.35	2.33	2.64	1.7	1.69	1.7	1.72	1.74	0.59	0.5	0	0.5	0	0	0	0	0
User 2	0	0	3.63	3.65	5	2.92	5	2.64	2.33	2.65	1.7	1.69	1.7	1.72	1.74	0.59	0.5	0	0.5	0	0	0	0	0
User 3	0	0	3.63	3.63	5	2.92	5	2.64	2.33	2.64	1.7	1.69	1.7	1.72	1.74	0.59	0.5	0	0.5	0	0	0	0	0
User 4	0	0.5	3.63	3.63	5	2.92	5	2.35	2.33	2.64	1.73	1.69	1.7	1.72	1.74	0.59	0.5	0	0.5	0	0	0	0	0
User 5	0	0	3.63	3.63	5	2.92	5	2.64	2.33	2.64	1.7	1.69	1.7	1.72	1.74	0.59	0.5	0	0.5	0	0	0	0	0
User 6	0	0	3.63	3.63	5	2.92	5	2.63	2.33	2.65	1.71	1.69	1.69	1.72	1.74	0.59	0.5	0	0.5	0	0	0	0	0
User 7	0	5	3.63	3.63	5	2.92	3.03	2.35	2.33	2.64	1.73	1.68	1.69	1.71	1.75	0.59	0.5	0	0.5	0	0	0	0	0
User 8	0	5	3.63	3.63	5	2.92	5	2.63	2.33	2.64	1.7	1.68	1.69	1.71	1.74	0.6	0.5	0	0.5	0	0	0	0	0
User 9	0	5	3.64	3.64	5	2.92	3.03	2.62	2.32	2.63	1.71	1.67	1.68	1.7	0.75	0.6	0	0	0.5	0	0	0	0	0
User 10	0	5	3.64	3.64	5	2.92	5	2.35	2.32	2.63	1.69	1.67	1.67	1.7	0.74	0.59	0	0	0	0	0	0	0	0

Εικόνα 4.2.1 Δείγμα από τις κριτικές χρηστών για τις χώρες της Ευρώπης

user_id	art_galleries	dance_clubs	juice_bars	restaurants	museums	resorts	parks_picnic_spots	beaches	theaters	religious_institutions
User 1	0.93	1.8	2.29	0.62	0.8	2.42	3.19	2.79	1.82	2.42
User 2	1.02	2.2	2.66	0.64	1.42	3.18	3.21	2.63	1.86	2.32
User 3	1.22	0.8	0.54	0.53	0.24	1.54	3.18	2.8	1.31	2.5
User 4	0.45	1.8	0.29	0.57	0.46	1.52	3.18	2.96	1.57	2.86
User 5	0.51	1.2	1.18	0.57	1.54	2.02	3.18	2.78	1.18	2.54
User 6	0.99	1.28	0.72	0.27	0.74	1.26	3.17	2.89	1.66	3.66
User 7	0.9	1.36	0.26	0.32	0.86	1.58	3.17	2.66	1.22	3.22
User 8	0.74	1.4	0.22	0.41	0.82	1.5	3.17	2.81	1.54	2.88
User 9	1.12	1.76	1.04	0.64	0.82	2.14	3.18	2.79	1.41	2.54
User 10	0.7	1.36	0.22	0.26	1.5	1.54	3.17	2.82	2.24	3.12

Εικόνα 4.2.2 Δείγμα από τις κριτικές χρηστών για τις χώρες της Ανατολικής Ασίας

id	choice	suggestion_1	suggestion_2
1	art_galleries	juice_bars	restaurants
2	bakeries	gyms	churches
3	beaches	parks	theatres
4	beauty_spas	churches	cafes
5	burger_pizza_shops	hotels_other_lodgings	local_services
6	cafes	churches	view_points
7	churches	gardens	beaches
8	dance_clubs	swimming_pools	parks
9	gardens	churches	monuments
10	gyms	swimming_pools	bakeries
11	hotels_other_lodgings	burger_pizza_shops	juice_bars
12	juice_bars	hotels_other_lodgings	burger_pizza_shops
13	local_services	pubs_bars	zoo
14	malls	restaurants	zoo
15	monuments	churches	gardens
16	museums	theatres	malls
17	parks	theatres	beaches
18	pubs_bars	restaurants	zoo
19	resorts	beaches	theatres
20	restaurants	pubs_bars	zoo
21	swimming_pools	gyms	dance_clubs
22	theatres	parks	museums
23	view_points	monuments	churches
24	zoo	restaurants	pubs_bars

Εικόνα 4.2.4 Προτάσεις για τις χώρες της Ανατολικής Ασίας

Εικόνα 4.2.3 Προτάσεις για τις χώρες της Ευρώπης

ΚΕΦΑΛΑΙΟ 5

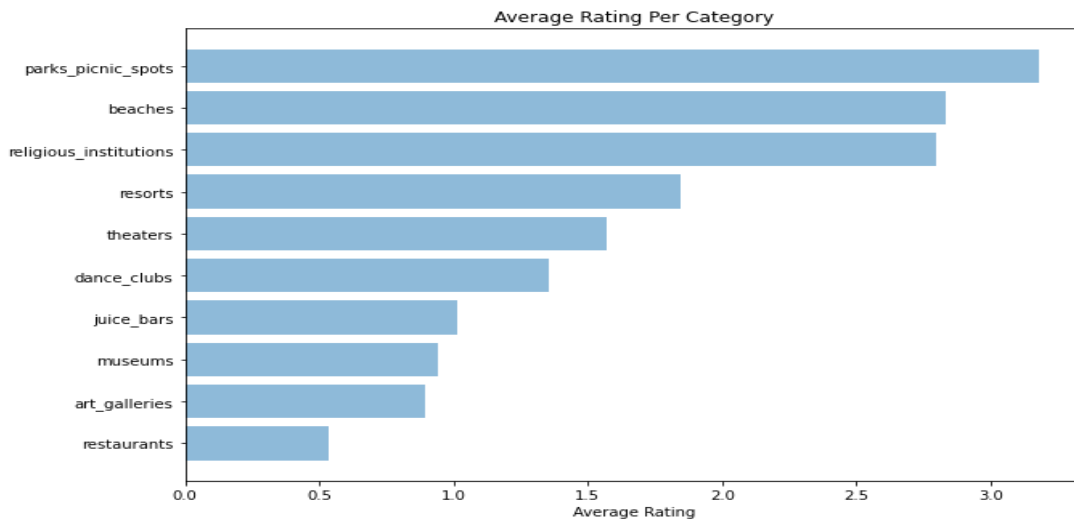
5.1 ΚΑΤΗΓΟΡΙΕΣ ΚΑΙ ΠΡΟΤΑΣΗ ΕΝΔΙΑΦΕΡΟΝΤΩΝ

Στα δεδομένα που χρησιμοποιούνται στην παρούσα διπλωματική εργασία, οι κατηγορίες των ενδιαφερόντων για τις χώρες της Ανατολικής Ασίας είναι 10, ενώ για τις χώρες της Ευρώπης είναι 24.

Ο αριθμός των κριτικών στην πρώτη περίπτωση είναι 980 και κάθε κριτική βρίσκεται στο εύρος 0 (απαίσιο) έως 4 (εξάισιο). Στη δεύτερη περίπτωση ο αριθμός των κριτικών είναι 5456 και κάθε κριτική βρίσκεται στο εύρος (1) απαίσιο έως 5 (εξάισιο).

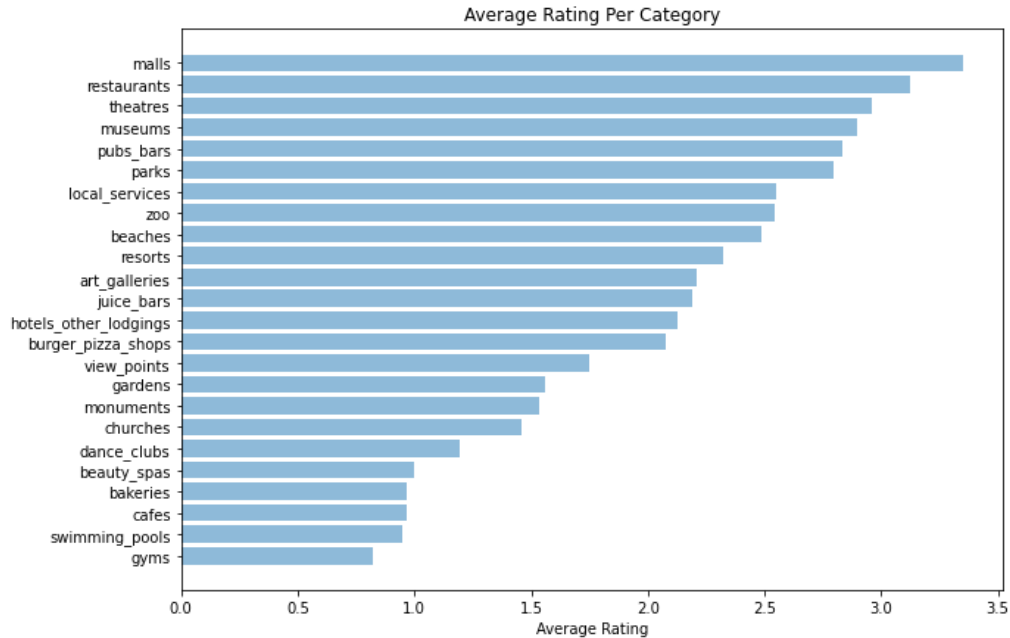
Στις παρακάτω εικόνες που δημιουργήθηκαν με χρήση της Python, παρουσιάζονται τα ονόματα των κατηγοριών, καθώς και ο μέσος όρος της κάθε μίας.

5.1.1 Μέσος όρος κάθε κατηγορίας ενδιαφέροντος για τις χώρες της Ανατολικής Ασίας



Εικόνα 5.1.1. Μέσος όρος κάθε κατηγορίας ενδιαφέροντος για τις χώρες της Ανατολικής Ασίας

5.1.2 Μέσος όρος κάθε κατηγορίας ενδιαφέροντος για τις χώρες της Ευρώπης



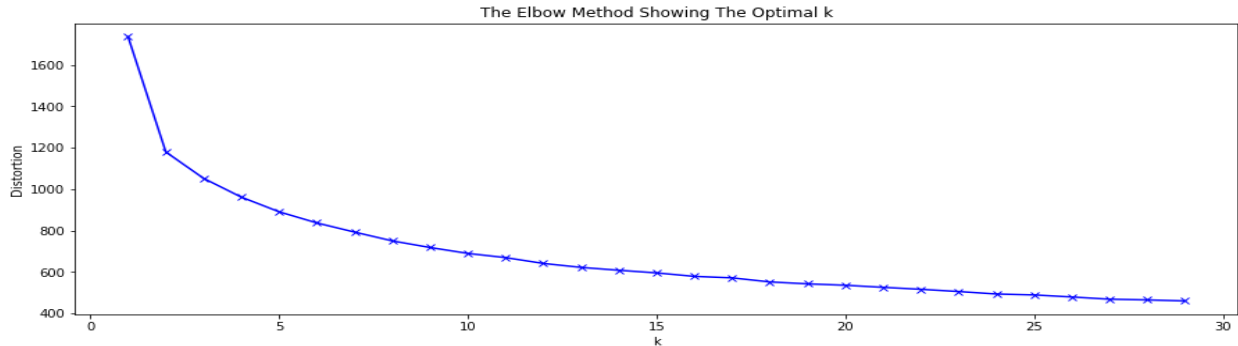
Εικόνα 5.1.2. Μέσος όρος κάθε κατηγορίας ενδιαφέροντος για τις χώρες της Ευρώπης

5.2 ΣΥΣΤΑΔΟΠΟΙΗΣΗ ΧΡΗΣΤΩΝ

Για την καλύτερη κατανόηση των συστάδων στις οποίες είναι χωρισμένοι οι χρήστες, προχωρήσαμε σε συσταδοποίηση κάθε χρήστη με χρήση της Python. Για να βρούμε τον κατάλληλο αριθμό συστάδων, χρησιμοποιήσαμε τη μέθοδο Elbow, που αναφέραμε νωρίτερα.

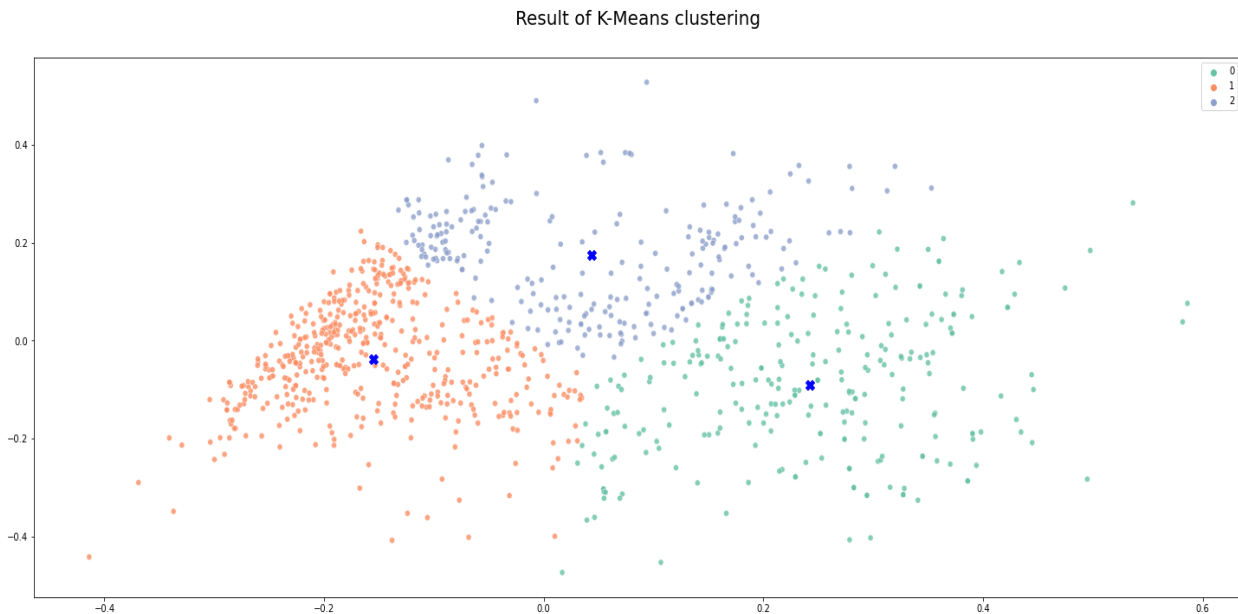
5.2.1 Συσταδοποίηση χρηστών στις χώρες της Ανατολικής Ασίας

Ο αριθμός των συστάδων που επιλέξαμε με βάση τη μέθοδο Elbow σε αυτό το σύνολο δεδομένων είναι 2.



Εικόνα 5.2.1. Χρήση της μεθόδου Elbow στις χώρες της Ανατολικής Ασίας

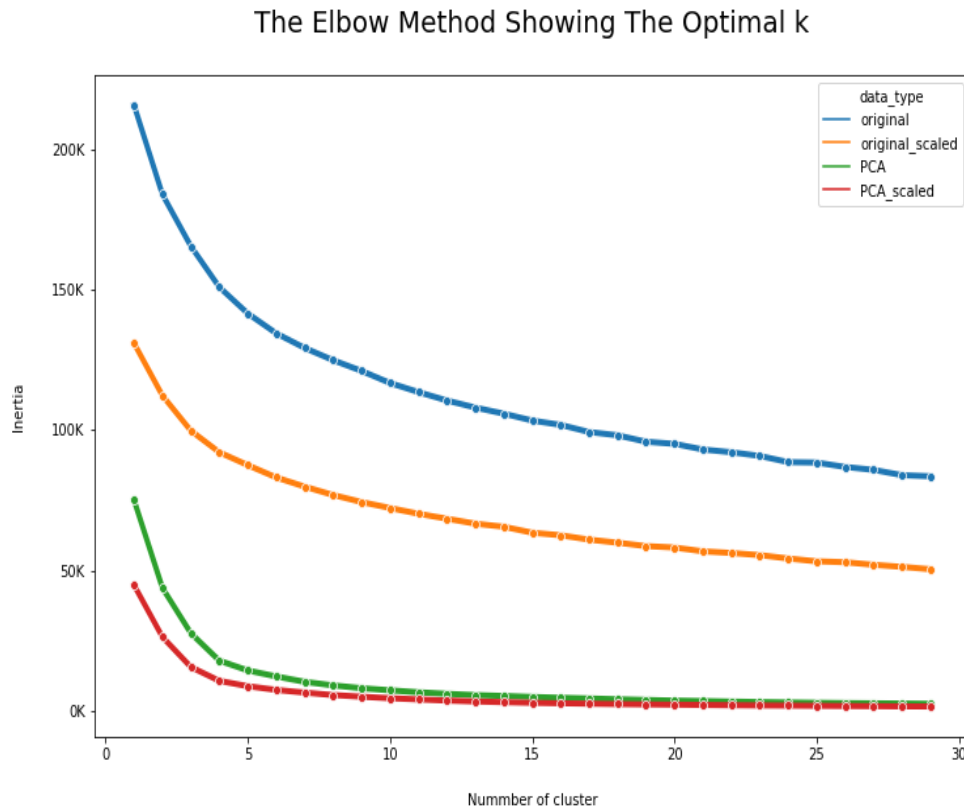
Αυτός είναι ο τρόπος οπτικοποίησης των δύο συστάδων με χρήση του αλγορίθμου KMeans.



Εικόνα 5.2.2. Συσταδοποίηση χρηστών στις χώρες της Ανατολικής Ασίας

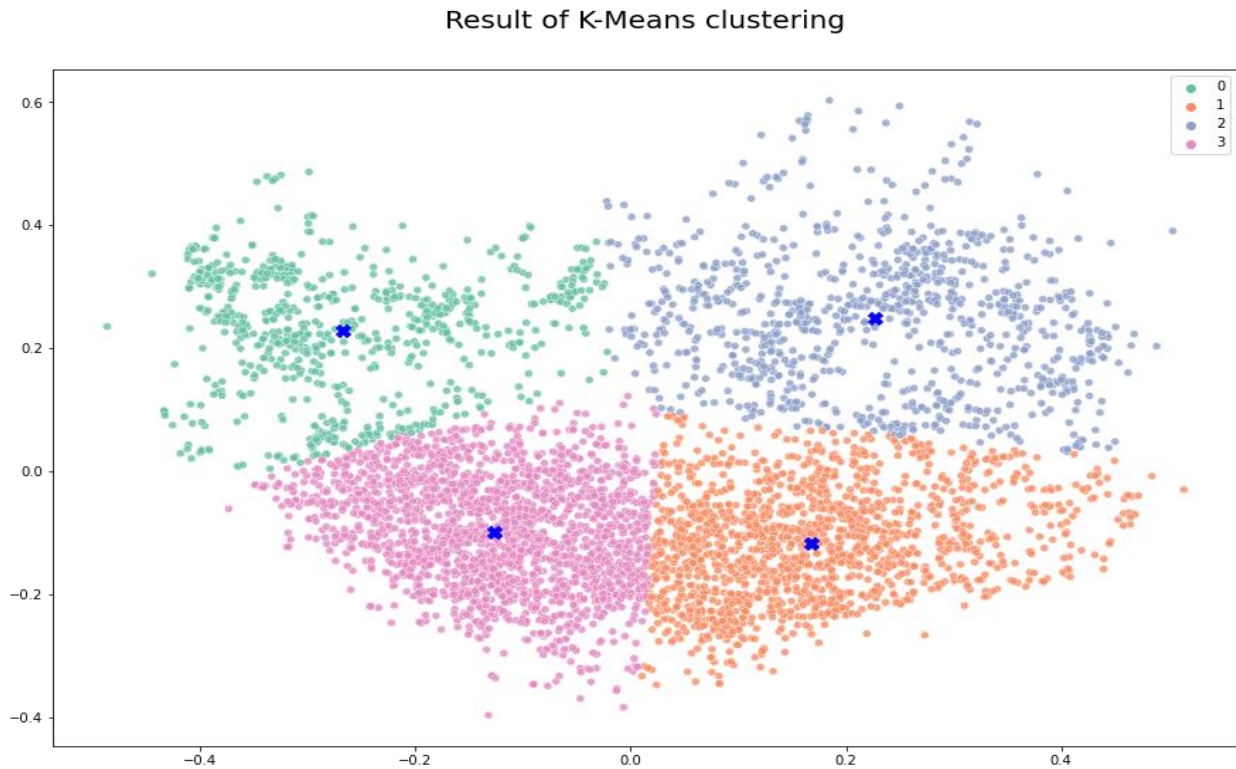
5.2.2 Συσταδοποίηση χρηστών στις χώρες της Ευρώπης

Ο αριθμός των συστάδων που επιλέξαμε με βάση τη μέθοδο Elbow σε αυτό το σύνολο δεδομένων είναι 4.



Εικόνα 5.2.3. Χρήση της μεθόδου Elbow χρηστών στις χώρες της Ευρώπης

Αυτός είναι ο τρόπος οπτικοποίησης των τεσσάρων συστάδων με χρήση του αλγορίθμου KMeans.

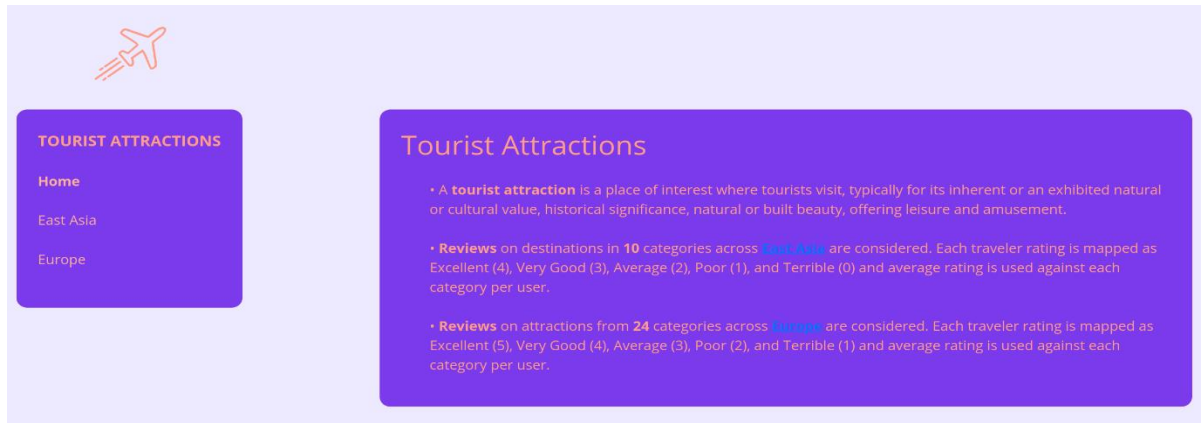


Εικόνα 5.2.4. Συσταδοποίηση χρηστών στις χώρες της Ευρώπης

5.3 ΠΡΟΤΑΣΗ ΕΝΔΙΑΦΕΡΟΝΤΩΝ

Για να έχουμε πρόσβαση στην ιστοσελίδα τοπικά, πρέπει να ανοίξουμε ένα τερματικό στο φάκελο `fullstack_website` και να πληκτρολογήσουμε την εντολή `php artisan serve`.

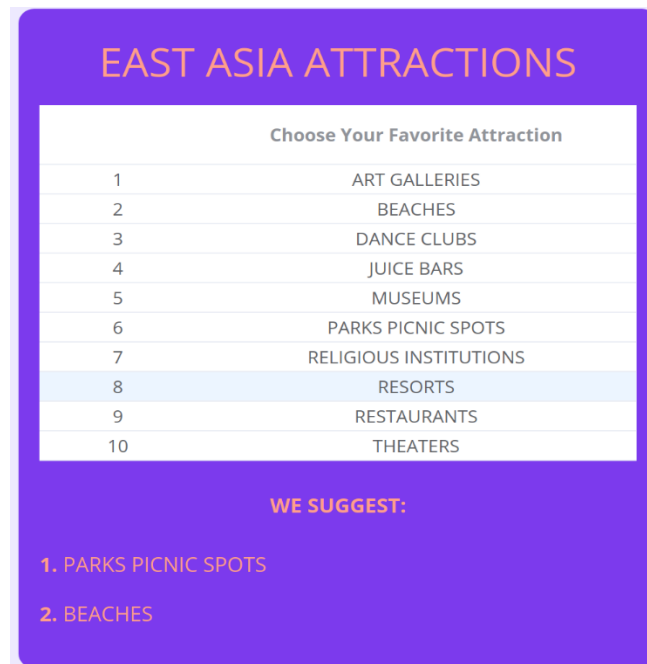
Η ιστοσελίδα αποτελείται από ένα απλό μενού, που έχει ως επιλογές την αρχική σελίδα, τη σελίδα που αφορά τις κατηγορίες ενδιαφερόντων της Ανατολικής Ασίας και την αντίστοιχη σελίδα για τις χώρες της Ευρώπης. Στην αρχική σελίδα γίνεται μία απλή αναφορά, σχετικά με το τι ορίζεται κατηγορία τουριστικού ενδιαφέροντος, καθώς και λίγα λόγια για κάθε σύνολο δεδομένων.



Εικόνα 5.3.1. Η αρχική σελίδα της ιστοσελίδας

Ο τρόπος πρότασης ενδιαφερόντων, είναι αρκετά απλός. Ο χρήστης πρέπει απλώς να επιλέξει την αγαπημένη του και τότε εμφανίζονται δύο προτάσεις.

Στο παρακάτω παράδειγμα, ο χρήστης επιλέγει ως αγαπημένη κατηγορία τα θέρετρα, και τότε του προτείνονται τα σημεία για picnic σε πάρκα και οι παραλίες.



Εικόνα 5.3.2. Πρόταση ενδιαφερόντων για τις χώρες της Ανατολικής Ασίας

Παρομοίως για τις χώρες της Ευρώπης, ο χρήστης επιλέγει και πάλι ως αγαπημένη κατηγορία τα θέρετρα, και τότε η πρόταση αφορά παραλίες και θέατρα.



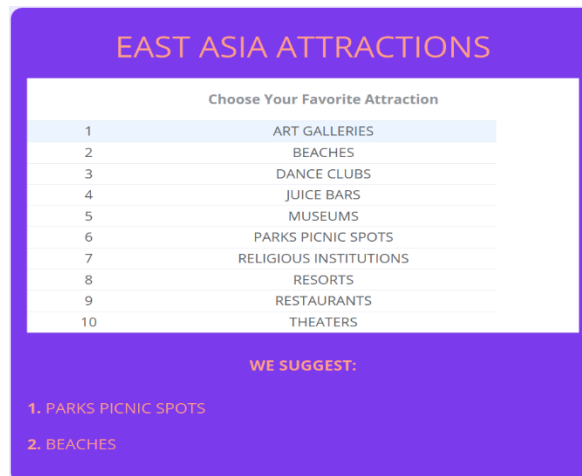
Εικόνα 5.3.3. Πρόταση ενδιαφερόντων για τις χώρες της Ευρώπης

ΚΕΦΑΛΑΙΟ 6

6.1 ΣΕΝΑΡΙΑ ΧΡΗΣΗΣ



Εικόνα 6.1.1. Πρώτο σενάριο χρήσης



Εικόνα 6.1.2. Δεύτερο σενάριο χρήσης

Ένας ταξιδιωτικός οργανισμός αναλαμβάνει να υλοποιήσει ένα ταξιδιωτικό πακέτο για έναν πελάτη που θέλει να ταξιδέψει στην Ισπανία. Επιθυμεί να συμπεριλάβει ορισμένα ταξιδιωτικά ενδιαφέροντα, με βάση το αγαπημένο ταξιδιωτικό ενδιαφέρον του πελάτη, το οποίο είναι τα μουσεία. Για να το πετύχει αυτό, θα επιλέξει από το μενού την επιλογή *Europe* και στη συνέχεια θα κάνει κλικ στην επιλογή *beaches*. Η εφαρμογή του προτείνει δύο ακόμα.

Το αγαπημένο ενδιαφέρον ενός ταξιδιώτη που θέλει να ταξιδέψει στην Κίνα είναι οι Γκαλερί Τέχνης. Επιθυμεί να του προταθούν δύο ακόμα ταξιδιωτικά ενδιαφέροντα. Για να επιτευχθεί αυτό, θα επιλέξει από το μενού την επιλογή *East Asia* και στη συνέχεια θα κάνει κλικ στην επιλογή *art_galleries*. Η εφαρμογή του προτείνει δύο ακόμα.



Εικόνα 6.1.3. Τρίτο σενάριο χρήσης

Το αγαπημένο ενδιαφέρον ενός ταξιδιώτη που θέλει να ταξιδέψει στη Γερμανία είναι οι ζωολογικοί κήποι και επιθυμεί να του προταθούν δύο ακόμα ταξιδιωτικά ενδιαφέροντα. Για να επιτευχθεί αυτό, θα επιλέξει από το μενού την επιλογή *Europe* και στη συνέχεια θα κάνει κλικ στην επιλογή *zoo*. Η εφαρμογή του προτείνει δύο ακόμα.

ΚΕΦΑΛΑΙΟ 7

7.1 ΕΠΙΛΟΓΟΣ

Στην παρούσα διπλωματική εργασία, σχεδιάστηκαν και υλοποιήθηκαν το μοντέλο και η εφαρμογή, που υλοποιούν την ιδέα για πρόταση ταξιδιωτικών ενδιαφερόντων με χρήση μηχανικής μάθησης και με βάση το αγαπημένο ενδιαφέρον ενός ταξιδιώτη.

Μετά από δοκιμές που έγιναν, καταλήξαμε ότι ο κατάλληλος αριθμός προτάσεων είναι δύο, καθώς τα σύνολα δεδομένων δε λογίζονται ως μεγάλα και οι αποστάσεις που υπολογίζονταν με τον αλγόριθμο NearestNeighbors, μεταξύ του αγαπημένου ενδιαφέροντος και του τρίτου κατά σειρά προτεινόμενου, ήταν αρκετά μεγάλες.

Για την επέκταση αυτής της υλοποίησης, υπάρχει η δυνατότητα κατηγοριοποίησης κάθε ταξιδιώτη. Κάθε κατηγορία θα αντιστοιχίζεται με συγκεκριμένες κατηγορίες. Με αυτόν τον τρόπο, θα λύνεται το πρόβλημα που ενδέχεται να δημιουργείται, σε περιπτώσεις που ο ταξιδιώτης δεν έχει κάποιο αγαπημένο ταξιδιωτικό ενδιαφέρον, αλλά θεωρεί ότι ανήκει σε μία ευρύτερη κατηγορία.

Τέλος, δύναται η δυνατότητα δημιουργίας mobile εφαρμογής, για καλύτερη εμπειρία πρόσβασης των χρηστών μέσω κινητών συσκευών ή tablet.

ΒΙΒΛΙΟΓΡΑΦΙΑ

1. University of Berkeley, What Is Machine Learning, (2020, June 26), retrieved from <https://ischoolonline.berkeley.edu/blog/what-is-machine-learning/>
2. ECMA International, ECMAScript language specification, (2021, June), retrieved from <https://tc39.es/ecma262/#sec-overview>
3. IBM, Machine Learning, (2020, July 15), retrieved from <https://www.ibm.com/cloud/learn/machine-learning>
4. Towards Data Science, Understanding K-Means Clustering in Machine Learning, (2018, September 13), retrieved from <https://towardsdatascience.com/understanding-k-means-clustering-in-machine-learning-6a6e67336aa1>
5. Towards Data Science, Machine Learning Basics with the K-Nearest Neighbors Algorithm, (2018, September 10), retrieved from <https://towardsdatascience.com/machine-learning-basics-with-the-k-nearest-neighbors-algorithm-6a6e71d01761>
6. Analytics Vidhya, A Simple Explanation of K-Means Clustering, (2020, October 4), retrieved from <https://www.analyticsvidhya.com/blog/2020/10/a-simple-explanation-of-k-means-clustering>
7. Business Insider, What is Python? The popular, scalable programming language, explained, (2021, June 26), retrieved from <https://www.businessinsider.com/what-is-python>
8. Sitepoint, Goodbye CodeIgniter, Hello Laravel, (2023, May 6), retrieved from <https://www.sitepoint.com/goodbye-codeigniter-hello-laravel/>
9. Laravel, Laravel Documentation, retrieved from <https://laravel.com/docs/8.x>

10. Artima, The DCI Architecture: A New Vision of Object-Oriented Programming, (2009, March 20), retrieved from <https://www.artima.com/articles/the-dci-architecture-a-new-vision-of-object-oriented-programming>
11. Mozilla Developer, MVC, (2021, January 13), retrieved from <https://developer.mozilla.org/en-US/docs/Glossary/MVC>
12. MySQL, What is MySQL?, retrieved from <https://dev.mysql.com/doc/refman/8.0/en/what-is-mysql.html>
13. Crockford, The Little JavaScripter, (2019, May 27), retrieved from <https://www.crockford.com/little.html>
14. Mozilla Developer, What is JavaScript?, (2021, June 18), retrieved from https://developer.mozilla.org/en-US/docs/Learn/JavaScript/First_steps/What_is_JavaScript
15. UCI Machine Learning Repository, Travel Review Ratings Data Set, (2018, December 19), retrieved from <https://archive.ics.uci.edu/ml/datasets/Tarvel+Review+Ratings>
16. UCI Machine Learning Repository, Travel Review Data Set, (2018, December 19), retrieved from <https://archive.ics.uci.edu/ml/datasets/Travel+Reviews>
17. Learn Code, PHP, (2020, December 14), retrieved from <https://www.learncode.gr/2020/12/php-basic-ref.html>
18. Scikit, Elbow Method, retrieved from <https://www.scikit-yb.org/en/latest/api/cluster/elbow.html>
19. Avantica, What is Vue.js and How do we Use It?, (2019, November 26), retrieved from <https://www.avantica.com/blog/what-is-vue.js-and-how-do-we-use-it>
20. W3, 20 Years of CSS, retrieved from <https://www.w3.org/Style/CSS20/>

21. Mozilla Developer, HTML Basics, (2021, September 3), retrieved from https://developer.mozilla.org/en-US/docs/Learn/Getting_started_with_the_web/HTML_basics
22. Pandas, pandas library, retrieved from <https://pandas.pydata.org/>
23. Numpy, NumPy library, retrieved from <https://numpy.org/>
24. Matplotlib, matplotlib.pyplot, retrieved from https://matplotlib.org/stable/api/_as_gen/matplotlib.pyplot.html
25. MySQL Dev, MySQL Connector/Python, retrieved from <https://dev.mysql.com/downloads/connector/python/>
26. Scipy, scipy.sparse.csr_matrix, retrieved from https://docs.scipy.org/doc/scipy/reference/generated/scipy.sparse.csr_matrix.html
27. Scikit Learn, sklearn.neighbors.NearestNeighbors, retrieved from <https://scikit-learn.org/stable/modules/generated/sklearn.neighbors.NearestNeighbors.html>
28. Scikit Learn, sklearn.neighbors.StandardScaler, retrieved from <https://scikit-learn.org/stable/modules/generated/sklearn.neighbors.NearestNeighbors.html>
29. Scikit Learn, sklearn.neighbors.KMeans, retrieved from <https://scikit-learn.org/stable/modules/generated/sklearn.cluster.KMeans.html>
30. Scikit Learn, sklearn.neighbors.PCA, retrieved from <https://scikit-learn.org/stable/modules/generated/sklearn.decomposition.PCA.html>
31. Matplotlib, matplotlib.ticker, retrieved from https://matplotlib.org/stable/api/ticker_api.html
32. Seaborn, seaborn: statistical data visualization, retrieved from <https://seaborn.pydata.org>