ΙΟΝΙΟ ΠΑΝΕΠΙΣΤΗΜΙΟ

ΤΜΗΜΑ ΠΛΗΡΟΦΟΡΙΚΗΣ



Πτυχιακή Εργασία

“Ανάπτυξη ενός συστήματος συστάσεων αξιοποιώντας μοντέλα Deep Learning και το Tensorflow”

Βέττας Εμμανουήλ

Αριθμός Μητρώου: Π2019206

Επιβλέπουσα: Χριστοπούλου Ελένη

Κέρκυρα, Φεβρουάριος 2024

ΙΟΝΙΟ ΠΑΝΕΠΙΣΤΗΜΙΟ

ΤΜΗΜΑ ΠΛΗΡΟΦΟΡΙΚΗΣ



Πτυχιακή Εργασία

“Ανάπτυξη ενός συστήματος συστάσεων αξιοποιώντας μοντέλα Deep Learning και το Tensorflow”

Βέττας Εμμανουήλ

Αριθμός Μητρώου: Π2019206

Τριμελής Επιτροπή

Χριστοπούλου Ελένη
ΕΔΙΠ
Τμήμα Πληροφορικής,
Ιόνιο Πανεπιστήμιο

Ρίγγας Δημήτριος
ΕΔΙΠ
Τμήμα Πληροφορικής,
Ιόνιο Πανεπιστήμιο

Μιχαήλ Στεφανιδάκης,
Αναπληρωτής Καθηγητής
Τμήμα Πληροφορικής,
Ιόνιο Πανεπιστήμιο

Περίληψη

Η παρούσα μελέτη ασχολείται με την ανάπτυξη μιας καινοτόμου εφαρμογής που εστιάζει στην προσαρμογή των προτάσεων εικόνων στις προτιμήσεις των χρηστών, με έμφαση σε μοντέλα deep learning για την παραγωγή μεταδεδομένων. Η εφαρμογή αυτή, σχεδιασμένη να λειτουργεί παρόμοια με άλλα δημοφιλή κοινωνικά δίκτυα και στοχεύει στην παροχή μιας εξατομικευμένης εμπειρίας περιήγησης για τους χρήστες της.

Η αρχική φάση της ανάπτυξης εστίασε στη δημιουργία ενός μοντέλου κατηγοριοποίησης εικόνων. Αυτό το μοντέλο βασίστηκε σε μεταδεδομένα που προέκυψαν από τη χρήση deep learning αλγορίθμων, στοχεύοντας στην αυτόματη κατηγοριοποίηση νέων εικόνων που προστίθενται στην πλατφόρμα. Παρ' όλα αυτά, το μοντέλο αντιμετώπισε προκλήσεις σε σχέση με την απόδοση, κυρίως λόγω των περιορισμένων δεδομένων εκπαίδευσης.

Στη συνέχεια, με σκοπό την αντιμετώπιση των προβλημάτων που προέκυψαν, η εφαρμογή ενσωμάτωσε την προηγμένη τεχνολογία του GPT-4 για την κατηγοριοποίηση εικόνων. Η χρήση του GPT-4 έφερε σημαντικές βελτιώσεις, καθώς το μοντέλο αυτό είναι ικανό να αναλύει πιο περίπλοκα σετ δεδομένων και να προσφέρει μια πιο προσαρμοστική και ακριβή λειτουργία στην κατηγοριοποίηση και προσαρμογή του περιεχομένου.

Η εφαρμογή αυτή αντιπροσωπεύει ένα πρωτοποριακό βήμα στην ενσωμάτωση της τεχνητής νοημοσύνης στον τομέα των ψηφιακών πλατφορμών, υποδεικνύοντας τη σημασία της επιλογής των κατάλληλων τεχνολογικών λύσεων. Μέσω της ενσωμάτωσης του GPT-4, η πλατφόρμα έχει δείξει την ικανότητά της να προσφέρει πιο εξατομικευμένες και βελτιωμένες υπηρεσίες στους χρήστες της, ανοίγοντας παράλληλα νέες δυνατότητες για την περαιτέρω εξέλιξη και βελτίωση των ψηφιακών εμπειριών.

Λέξεις κλειδιά

Συστήματα συστάσεων, τεχνητή νοημοσύνη, tensorflow, Generative Pre-training Transformers (GPTs), μεταδεδομένα εικόνων, social media

Ευχαριστίες

Θα ήθελα να ευχαριστήσω την επιβλέπουσα καθηγήτριά μου κα. Ελένη Χριστοπούλου για την άψογη συνεργασία, επικοινωνία και εμπιστοσύνη που είχαμε καθ' όλη τη διάρκεια εκπόνησης της πτυχιακής μου εργασίας. Επίσης, θα ήθελα να ευχαριστώ την οικογένειά μου για την υποστήριξη που μου προσέφεραν σε όλο το διάστημα των σπουδών μου.

Περιεχόμενα

Κεφάλαιο 1 - Εισαγωγή.....7
1.2 Στόχος.....7
1.3 Συνεισφορά εργασίας.....7
1.4 Δομή εργασίας.....8
1.5 Τελικός Στόχος Εργασίας.....8
Κεφάλαιο 2 - Θεωρητικό Υπόβαθρο.....9
2.1 Τα συστήματα συστάσεων.....9
2.2 Τεχνητή νοημοσύνη.....10
2.3 Τεχνικές εκπαίδευσης μοντέλων τεχνητής νοημοσύνης.....11
2.4 Συστήματα συστάσεων και τεχνητή νοημοσύνη.....13
2.5 Μεγάλα μοντέλα τεχνητής νοημοσύνης γλωσσικού περιεχομένου (LLMs).....14
2.6 Γνωστές Μέθοδοι Δημιουργίας Συστημάτων Συστάσεων.....14
Content-based.....14
Συστήματα συνεργατικού φιλτραρίσματος.....14
Υβριδικά συστήματα.....15
2.7 Είδη δεδομένων και μετρικές αξιολόγησης των συστημάτων συστάσεων.....16
2.8 Προκλήσεις που αντιμετωπίζουν τα συστήματα συστάσεων.....17
2.9 Το πρόβλημα κρύας εκκίνησης (cold start problem).....18
2.10 Πλατφόρμες κοινωνικής δικτύωσης.....19
2.11 Κοινωνικά δίκτυα και συστήματα συστάσεων.....19
Κεφάλαιο 3 - Περιγραφή και υλοποίηση του συστήματος.....21
3.1 Δημιουργία πλατφόρμας κοινωνικού δικτύου.....22
3.1.1 Αναλυτική περιγραφή των εργαλείων:.....22
3.1.2 Τα μέρη και οι σελίδες της εφαρμογής.....24
Σελίδα εγγραφής/σύνδεσης χρηστών:.....24
Αρχική σελίδα.....26
Αλγόριθμος επιλογής εικόνων.....27
Αναλυτικά ο αλγόριθμος των εικόνων.....30
Ανάλυση του αλγορίθμου σε μορφή κώδικα (γλώσσα προγραμματισμού: Javascript).....31
Σελίδα προφίλ χρήστη.....34
Σελίδα εισαγωγής νέων εικόνων.....35
3.2 Ανάλυση των εικόνων.....35
3.2.1 Η αρχιτεκτονική του συστήματος παραγωγής μεταδεδομένων:.....36
3.2.2 Αναλυτική περιγραφή των εργαλείων.....37
3.2.3 Επιλογή των κατάλληλων αλγορίθμων και μοντέλων βαθιάς μάθησης.....41
3.2.4 Χρήση των μοντέλων.....42
Στάδιο 1: Αναγνώριση σκηνικού.....42
Αλγόριθμος παραγωγής μεταδεδομένων σκηνικού:.....47
Ορισμός του class activation map:.....50

Εκπαίδευση των μοντέλων που χρησιμοποιήθηκαν:.....50
Στάδιο 2: Αναγνώριση αντικειμένων.....51
Στάδιο 3: Κατηγοριοποίηση των εικόνων.....55
Διαδικασία δημιουργίας συνόλου δεδομένων.....56
Θεωρητικό υπόβαθρο των συναρτήσεων εκπαίδευσης:.....58
Χρήση και αξιολόγηση του μοντέλου:.....59
Χρήση LLM's για παραγωγή μεταδεδομένων.....62
Αξιολόγηση μοντέλου LLM:.....64
3.3 Κατηγοριοποίηση των χρηστών.....64
Κεφάλαιο 4 - Αξιολόγηση του εργαλείου.....65
4.1 Αξιολόγηση των αποτελεσμάτων της εφαρμογής.....65
Κεφάλαιο 5 - Επίλογος.....67
5.1 Συμπεράσματα που προκύπτουν.....67
5.2 Πιθανές μελλοντικές βελτιώσεις και αλλαγές.....68

Κεφάλαιο 1 - Εισαγωγή

Η τεχνολογία των συστημάτων συστάσεων έχει αναπτυχθεί σημαντικά τα τελευταία χρόνια και έχει εφαρμοστεί ευρέως σε διάφορους τομείς, όπως το ηλεκτρονικό εμπόριο, η ψυχαγωγία και τα κοινωνικά δίκτυα. Τα συστήματα συστάσεων αποσκοπούν στο να παρέχουν στους χρήστες προσαρμοσμένο και σχετικό περιεχόμενο, βελτιώνοντας έτσι την εμπειρία τους και προωθώντας την αλληλεπίδραση με την πλατφόρμα στην οποία βρίσκονται.

Στην προσπάθεια των ανθρώπων να αναπτύξουν μεγαλύτερα και αποδοτικότερα συστήματα συστάσεων, έχει αναδειχθεί αρκετά η ενσωμάτωση τεχνολογιών βαθειάς μάθησης (Deep Learning) η οποία έχει ανοίξει νέους δρόμους για ακόμα πιο προηγμένες και εξατομικευμένες προτάσεις. Χρησιμοποιώντας βαθιά νευρωνικά δίκτυα, τα σύγχρονα συστήματα μπορούν να αναλύσουν μεγάλες ποσότητες δεδομένων, να αναγνωρίσουν πολύπλοκα μοτίβα συμπεριφοράς των χρηστών και να προσφέρουν πιο ακριβείς και σχετικές συστάσεις.

1.2 Στόχος

Στο πλαίσιο αυτής της πτυχιακής εργασίας, θα εξετάσουμε τον τρόπο με τον οποίο το Deep Learning μπορεί να ενσωματωθεί σε συστήματα συστάσεων εικόνων για να βελτιώσει την ακρίβεια και την ποιότητα των συστάσεων σε μια πλατφόρμα ενός κοινωνικού δικτύου. Ειδικότερα, θα επικεντρωθούμε στη χρήση του TensorFlow, ενός ανοιχτού λογισμικού βιβλιοθήκης για μηχανική μάθηση, το οποίο προσφέρει εκτεταμένη υποστήριξη για Deep Learning. Η επιλογή του TensorFlow οφείλεται στην ευελιξία, τη δυνατότητα κλιμάκωσης και την εκτεταμένη κοινότητα υποστήριξης που προσφέρει.

1.3 Συνεισφορά εργασίας

Μέσω αυτής της μελέτης, θα εξερευνήσουμε τις διάφορες αρχιτεκτονικές του Deep Learning που εφαρμόζονται στα συστήματα συστάσεων, αναλύοντας παραδείγματα υλοποίησης και περιπτώσεις χρήσης. Στόχος είναι να κατανοήσουμε τον αντίκτυπο αυτών των τεχνολογιών στη βελτίωση της χρηστικής εμπειρίας και στην αποδοτικότητα των συστημάτων συστάσεων.

1.4 Δομή εργασίας

Η δομή της πτυχιακής εργασίας περιλαμβάνει τα ακόλουθα κεφάλαια:

- Στο **Κεφάλαιο 2** παρουσιάζεται μια εκτενής μελέτη της θεωρίας των συστημάτων συστάσεων, καλύπτοντας τις κατηγορίες, τις τεχνικές, τα είδη δεδομένων και τις μετρικές αξιολόγησης που χρησιμοποιούνται. Επίσης, περιγράφονται οι προκλήσεις που αντιμετωπίζουν τα συστήματα αυτά και μεθοδολογίες που μπορούν να λύσουν αυτές τις προκλήσεις.
- Το **Κεφάλαιο 3** συνδυάζει και περιγράφει συνοπτικά τη μεθοδολογία ανάπτυξης και την αρχιτεκτονική του συστήματος συστάσεων. Αρχικά, παρουσιάζεται η μεθοδολογία που ακολουθήθηκε για τη δημιουργία του συστήματος, περιλαμβάνοντας τον σχεδιασμό, την επιλογή των αλγορίθμων και τεχνικών, καθώς και τα εργαλεία και τις γλώσσες προγραμματισμού που χρησιμοποιήθηκαν. Στη συνέχεια, εξετάζεται η αρχιτεκτονική του συστήματος, αναλύοντας τις λειτουργίες και τις ροές δεδομένων μεταξύ των διαφόρων συστατικών.
- Το **Κεφάλαιο 4** περιλαμβάνει την αξιολόγηση του εργαλείου, με έμφαση στην απόδοση και την λειτουργικότητα του. Παρουσιάζονται τα θετικά και τα αρνητικά στοιχεία και γίνονται προτάσεις για περαιτέρω βελτιώσεις.
- Το **Κεφάλαιο 5** περιλαμβάνει τα συμπεράσματα που προκύπτουν από την εργασία, αναφέρονται πιθανές μελλοντικές επεκτάσεις και προτάσεις για μελλοντική έρευνα.
- Τέλος, στο τελευταίο μέρος της εργασίας περιλαμβάνονται οι βιβλιογραφικές αναφορές και άλλες πηγές που χρησιμοποιήθηκαν κατά τη διάρκεια της έρευνας και της ανάπτυξης του συστήματος.

1.5 Τελικός Στόχος Εργασίας

Τελικός στόχος αυτής της πτυχιακής εργασίας, είναι η προσφορά ενός πρωτότυπου συστήματος συστάσεων το οποίο βασίζεται σε εικόνες και χρησιμοποιεί τεχνολογίες τεχνητής νοημοσύνης για την ανάλυση των εικόνων αυτών και την παραγωγή μεταδεδομένων. Η εργασία αναπτύσσει τον σκοπό, τους στόχους και τη δομή του συστήματος, προσφέροντας ένα ολοκληρωμένο εργαλείο για την κατηγοριοποίηση των χρηστών και την παροχή εξατομικευμένων συστάσεων εικόνων. Η τελική υλοποίηση της εφαρμογής: https://ptuxiakhmanwlhs.web.app

Ο κώδικας της εφαρμογής: https://github.com/Mnlsvt/recom_system

Κεφάλαιο 2 - Θεωρητικό Υπόβαθρο

Στο κεφάλαιο αυτό, θα διερευνήσουμε αναλυτικά το θεωρητικό υπόβαθρο και τις βασικές αρχές που καθορίζουν τα συστήματα συστάσεων, μια τεχνολογία που έχει αλλάξει τον τρόπο που αλληλεπιδρούν οι χρήστες με τον ψηφιακό κόσμο. Τα συστήματα συστάσεων είναι συνθετικά λογισμικά που στοχεύουν στο να παρέχουν στους χρήστες προτάσεις για προϊόντα, υπηρεσίες, πληροφορίες και εμπειρίες, που είναι πιο πιθανό να είναι σχετικές και ενδιαφέρουσες για αυτούς. Η αποτελεσματικότητα αυτών των συστημάτων στηρίζεται στην ικανότητά τους να αναλύουν μεγάλες ποσότητες δεδομένων, καθώς και στην εφαρμογή προηγμένων αλγορίθμων για την παραγωγή προσωποποιημένων συστάσεων.

2.1 Τα συστήματα συστάσεων

Η ιδέα πίσω από τα συστήματα συστάσεων πρωτοεμφανίστηκε τη δεκαετία του 1990, και ο κύριος σκοπός τους ήταν να βοηθήσουν τους ανθρώπους να περιηγηθούν μέσα σε μια πληθώρα επιλογών και να ανακαλύψουν τα προϊόντα ή τις εμπειρίες που τους ταιριάζουν καλύτερα. Αυτός ο τύπος τεχνολογίας έχει βρει εφαρμογές σε μια μεγάλη ποικιλία από τομείς, από το ηλεκτρονικό εμπόριο και την ψυχαγωγία μέχρι τα κοινωνικά δίκτυα και τις ψηφιακές υπηρεσίες. Με την εξέλιξη του Διαδικτύου, τα συστήματα αυτά έχουν γίνει πιο πολύπλοκα και προηγμένα, παρέχοντας αυξημένη εξατομίκευση των συστάσεων τους. Τα συστήματα αυτά, έχουν φτάσει σε σημείο να θεωρούνται ζωτικής σημασίας για μια πλατφόρμα και αυτό συμβαίνει επειδή οι άνθρωποι συχνά βασίζονται στις απόψεις των γύρω τους για να αποφύγουν να είναι οι πρώτοι που δοκιμάζουν κάτι καινούριο. [1]

Το προβλήματα τα οποία σχεδιάστηκαν να λύνουν τα συστήματα συστάσεων, είναι κυρίως ψυχολογικού τύπου και έχουν σαν στόχο να κρατήσουν το ενδιαφέρον του χρήστη κάνοντας την εμπειρία του πιο ευχάριστη. Κατά την πάροδο του χρόνου, και την ανάπτυξη του παγκόσμιου ιστού, τα συστήματα αυτά έχουν παρεκκλίνει από τον αρχικό τους στόχο (την ευχρηστία της περιήγησης του χρήστη), παρόλα αυτά κρατούν ζωντανή τη βασική ιδέα του να διατηρούν το ενδιαφέρον του χρήστη ζωντανό.

Πλέον, τα συστήματα συστάσεων εξυπηρετούν κυρίως εμπορικούς σκοπούς και η εξέλιξη τους βασίζεται πάνω στις κινήσεις και προτιμήσεις των καταναλωτών απέναντι στο εκάστοτε προϊόν που προωθείται. Αυτό το προϊόν μπορεί να είναι κάτι υλικό και απτό, το οποίο μπορεί να πουληθεί σε κάποιο φυσικό / ηλεκτρονικό κατάστημα, ή μπορεί και να είναι κάποια υπηρεσία που παρέχεται από κάποιον πάροχο, όπως για παράδειγμα μια πλατφόρμα social media.

Σε αυτές τις περιπτώσεις, η πληροφορία που είναι διαθέσιμη στον χρήστη μπορεί πολλές φορές να είναι υπερβολική και μη διαχειρίσιμη. Για αυτόν τον λόγο, στις μέρες μας, κρίνεται απαραίτητη η χρήση ενός τέτοιου συστήματος το οποίο είναι εξατομικευμένο για τον κάθε τύπο χρήσης. Τα συστήματα συστάσεων λειτουργούν σαν μια μορφή οπτικοποίησης πληροφορίας για τον χρήστη, καθώς τον βοηθούν να βρει αυτό που ψάχνει γρηγορότερα και πολλές φορές χωρίς να χρειαστεί να ψάξει ο ίδιος το αντικείμενο που επιθυμεί. Επίσης, όταν ένα τέτοιο σύστημα εφαρμόζεται πάνω σε μια πλατφόρμα, εξυπηρετεί διπλό σκοπό, καθώς βοηθάει τον χρήστη που περιηγείται στην πλατφόρμα, ενώ ταυτόχρονα αυξάνει το κέρδος της εκάστοτε επιχείρησης ανεβάζοντας την επισκεψιμότητα χρηστών στην πλατφόρμα της. [2]

Για να επιτύχουν τους στόχους τους, τα συστήματα συστάσεων στηρίζονται σε διάφορες τεχνικές και μεθόδους. Στις πιο απλές μορφές τους, αυτά τα συστήματα μπορεί να βασίζονται σε αλγορίθμους φιλτραρίσματος που επικεντρώνονται στην αξιολόγηση και την κατηγοριοποίηση του περιεχομένου. Ωστόσο, με την εξέλιξη των τεχνολογιών και την αυξημένη διαθεσιμότητα των δεδομένων, τα συστήματα συστάσεων έχουν αρχίσει να ενσωματώνουν πιο προηγμένες τεχνικές όπως η μηχανική μάθηση (Machile Learning), η βαθιά μάθηση (Deep Learning) κ.α.

2.2 Τεχνητή νοημοσύνη

Ως τεχνητή νοημοσύνη, ορίζεται ένα υπολογιστικό σύστημα το οποίο έχει τη δυνατότητα να εξάγει πληροφορίες και συμπεράσματα μέσα από την αναγνώριση μοτίβων και ακολουθιών. Αυτά τα συστήματα, είναι σχεδιασμένα να μιμούνται τις ανθρώπινες διαδικασίες σκέψης και λήψης αποφάσεων. Η δυνατότητα αυτή προκύπτει μέσω της χρήσης αλγορίθμων, οι οποίοι επεξεργάζονται μεγάλες ποσότητες δεδομένων, αναλύουν τις πληροφορίες και μαθαίνουν από αυτές, βελτιώνοντας συνεχώς τις επιδόσεις τους. Η ικανότητα αυτή ονομάζεται **μηχανική μάθηση** ή **machine learning** και αποτελεί έναν από τους βασικούς πυλώνες της σύγχρονης τεχνητής νοημοσύνης.

Επιπλέον, τα σύγχρονα συστήματα τεχνητής νοημοσύνης μπορούν να ασχοληθούν με διάφορες εφαρμογές, όπως η αναγνώριση φωνής, η επεξεργασία φυσικής γλώσσας, η οπτική αναγνώριση, και η αυτόματη λήψη αποφάσεων. Αυτό επιτυγχάνεται μέσω προηγμένων τεχνικών όπως η **βαθιά μάθηση** ή **deep learning**, το οποίο χρησιμοποιεί νευρωνικά δίκτυα με πολλά επίπεδα για την αναγνώριση περίπλοκων μοτίβων στα δεδομένα. Αυτή είναι η μεθοδολογία που χρησιμοποιήθηκε για την ανάπτυξη αυτής της εργασίας η οποία εξηγείται αναλυτικά παρακάτω.

Το κύριο πλεονέκτημα της τεχνητής νοημοσύνης είναι η ικανότητά της να επεξεργάζεται και να αναλύει τεράστιες ποσότητες δεδομένων πολύ πιο γρήγορα και αποτελεσματικά από τον ανθρώπινο εγκέφαλο. Αυτό επιτρέπει στα συστήματα AI να εντοπίζουν τάσεις και να προβλέπουν συμπεριφορές που θα ήταν δύσκολο ή αδύνατο για τους ανθρώπους να αναγνωρίσουν. Σημαντικό είναι επίσης το θέμα της επιλογής και διαχείρισης αλγορίθμων και αρχιτεκτονικών μοντέλων για συγκεκριμένες εφαρμογές, καθώς και η συνεχής προσαρμογή και ενημέρωση των μοντέλων για να παραμείνουν ακριβή και αποτελεσματικά στον πραγματικό κόσμο. [3]

Ωστόσο, παρά τις προόδους στον τομέα αυτό, η τεχνητή νοημοσύνη αντιμετωπίζει ακόμα σημαντικές προκλήσεις. Μία από αυτές είναι η ανάγκη για μεγάλες ποσότητες δεδομένων για την εκπαίδευση των αλγορίθμων, κάτι που μπορεί να δημιουργήσει ζητήματα προσωπικής απορρήτου και ασφάλειας δεδομένων. Επιπλέον, η τεχνητή νοημοσύνη μπορεί να είναι προκατειλημμένη, εφόσον οι αλγόριθμοι της μπορεί να μαθαίνουν από δεδομένα που αντανακλούν υπάρχουσες προκαταλήψεις στην κοινωνία. Αυτό το πρόβλημα είναι πολύ σύνηθες και η καλύτερη λύση που υπάρχει είναι η προσαρμογή και επιλογή των κατάλληλων δεδομένων έτσι ώστε να αποφευχθεί όσο το δυνατόν περισσότερο.

Τέλος, ενώ η τεχνητή νοημοσύνη έχει εκπληκτικές δυνατότητες, είναι σημαντικό να αναγνωρίζουμε τα όρια της και να εργαζόμαστε προς την ανάπτυξη συστημάτων που είναι όχι μόνο έξυπνα, αλλά και ασφαλή, δίκαια και διαφανή.

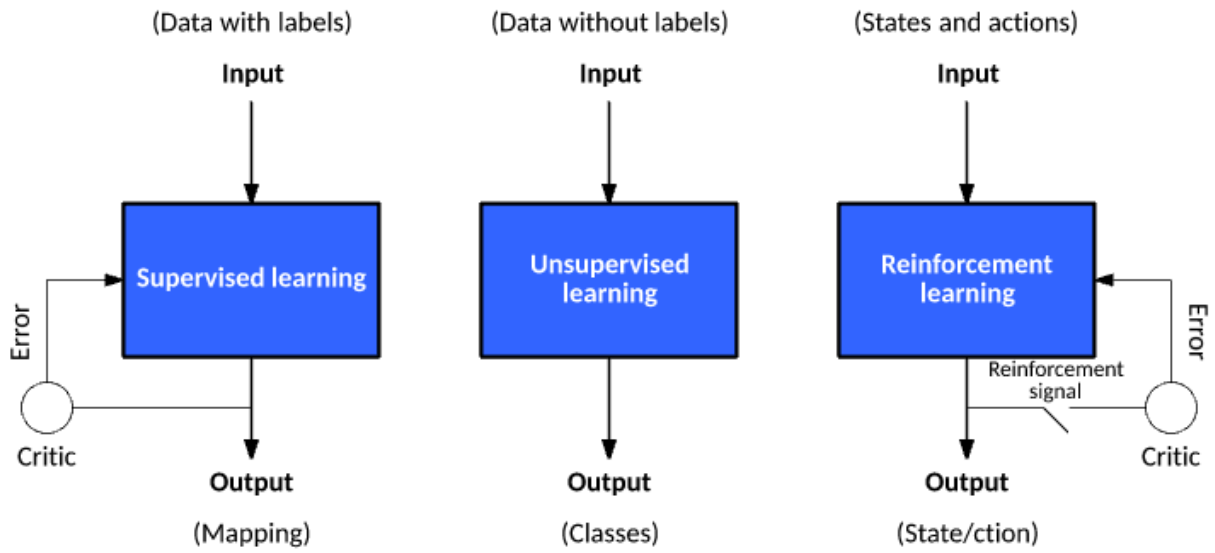
2.3 Τεχνικές εκπαίδευσης μοντέλων τεχνητής νοημοσύνης

Οι τρεις βασικές τεχνικές εκπαίδευσης μοντέλων στην τεχνητή νοημοσύνη είναι η εποπτευόμενη μάθηση (supervised learning), η μη εποπτευόμενη μάθηση (unsupervised learning) και η ενισχυτική μάθηση (reinforcement learning). Κάθε μία από αυτές τις τεχνικές χρησιμοποιεί διαφορετικές μεθόδους και είναι κατάλληλη για διάφορους τύπους προβλημάτων.

Εποπτευόμενη Μάθηση: Αυτή η τεχνική περιλαμβάνει την εκπαίδευση του μοντέλου σε ένα σύνολο δεδομένων με ετικέτες, όπου κάθε παράδειγμα στο σύνολο δεδομένων συνδέεται με τη σωστή έξοδο. Το μοντέλο μαθαίνει να κάνει προβλέψεις ή αποφάσεις με βάση αυτά τα δεδομένα. Η συγκεκριμένη τεχνική, είναι ιδανική για εφαρμογές όπου τα ιστορικά δεδομένα προβλέπουν μελλοντικά γεγονότα και είναι η τεχνική που χρησιμοποιήθηκε για την εφαρμογή αυτής της εργασίας.

Μη Εποπτευόμενη Μάθηση: Σε αντίθεση με την εποπτευόμενη μάθηση, η μη εποπτευόμενη μάθηση περιλαμβάνει την εκπαίδευση του μοντέλου σε δεδομένα χωρίς ετικέτες. Ο στόχος είναι να εξερευνήσει τη δομή των δεδομένων και να αναγνωρίσει πρότυπα ή ενσωματωμένα μοτίβα μέσα σε αυτά. Κοινές εφαρμογές περιλαμβάνουν τη συσταδοποίηση, την ανίχνευση ανωμαλιών και τη συσχέτιση.

Ενισχυτική Μάθηση: Αυτή η τεχνική επικεντρώνεται γύρω από την ιδέα των agents (πρακτόρων, υπολογιστικά προγράμματα) που μαθαίνουν να λαμβάνουν αποφάσεις μέσω της αλληλεπίδρασης με το περιβάλλον τους. Η διαδικασία μάθησης καθοδηγείται από ανταμοιβές και ποινές, παρόμοια με τη μάθηση μέσω δοκιμής και σφάλματος (trial & error technique). Χρησιμοποιείται ευρέως σε σενάρια όπου το μοντέλο χρειάζεται να κάνει μια σειρά αποφάσεων, και τα αποτελέσματα αυτών ενημερώνουν τη διαδικασία μάθησης. Η ενισχυτική μάθηση είναι ιδιαίτερα αποτελεσματική σε δυναμικά περιβάλλοντα όπου το μοντέλο χρειάζεται να προσαρμόζει τη στρατηγική του με βάση τις ανατροφοδοτήσεις.



Οι μέθοδοι εκπαίδευσης οπτικοποιημένες (visualized)

Πηγή: AI Wiki

2.4 Συστήματα συστάσεων και τεχνητή νοημοσύνη

Η μηχανική μάθηση έχει επιτρέψει στα συστήματα αυτά να αναγνωρίζουν σύνθετα μοτίβα στα δεδομένα των χρηστών και να κάνουν πιο εξατομικευμένες τις προτάσεις που δίνουν. Από την άλλη πλευρά, το Deep Learning, μια υποκατηγορία της μηχανικής μάθησης, χρησιμοποιεί βαθιά νευρωνικά δίκτυα για να αναλύσει μεγάλες ποσότητες δεδομένων, διερευνώντας πιθανές συσχετίσεις και μοτίβα που μπορεί να μην είναι εμφανή με πιο παραδοσιακές προσεγγίσεις.

Επιπλέον, η χρήση τεχνολογιών όπως το TensorFlow, που χρησιμοποιήθηκε σε αυτή την έρευνα, έχει απλοποιήσει την υλοποίηση σύνθετων αλγορίθμων βαθιάς μάθησης, επιτρέποντας στους ερευνητές και τους μηχανικούς να εξελίξουν τα συστήματα συστάσεων με μεγαλύτερη ευκολία και ακρίβεια. Η δυνατότητα ενσωμάτωσης πολύπλοκων μοντέλων σε εφαρμογές πραγματικού κόσμου έχει ανοίξει νέες διαστάσεις για την προσωποποίηση και την ακρίβεια των συστημάτων συστάσεων.

Η επιτυχία των σύγχρονων συστημάτων συστάσεων, ωστόσο, δεν είναι απαλλαγμένη από προκλήσεις. Θέματα όπως η ιδιωτικότητα των δεδομένων, η διαφάνεια των αλγορίθμων, και οι προκλήσεις στην επεξεργασία και διαχείριση τεράστιων ποσοτήτων δεδομένων αποτελούν σημαντικά ζητήματα που πρέπει να ληφθούν υπόψη. Παράλληλα, η ικανότητα των συστημάτων να παρέχουν ουδέτερες και δίκαιες συστάσεις, αποφεύγοντας τις προκαταλήψεις, είναι μια ακόμη πηγή ανησυχίας και ερευνητικού ενδιαφέροντος.

Στα επόμενα τμήματα αυτού του κεφαλαίου, θα εξετάσουμε πιο λεπτομερώς τις διάφορες τεχνολογίες και τεχνικές που χρησιμοποιούνται στα σύγχρονα συστήματα συστάσεων, με έμφαση στην εφαρμογή της βαθιάς μάθησης και του TensorFlow. Επιπρόσθετα, θα διερευνήσουμε τις προκλήσεις και τις ηθικές διαστάσεις που συνδέονται με αυτές τις τεχνολογίες, καθώς και τις προοπτικές για το μέλλον των συστημάτων συστάσεων.

2.5 Μεγάλα μοντέλα τεχνητής νοημοσύνης γλωσσικού περιεχομένου (LLMs)

Τα Large Language Models (LLMs), όπως το GPT-4, είναι προηγμένα συστήματα τεχνητής νοημοσύνης που εκπαιδεύονται να κατανοούν και να παράγουν φυσική γλώσσα. Αυτά τα μοντέλα διαθέτουν εκατομμύρια ή ακόμα και δισεκατομμύρια παραμέτρους, κάτι που τους επιτρέπει να αναλύουν και να παράγουν περίπλοκα κείμενα με εκπληκτική ακρίβεια. Τα LLMs εκπαιδεύονται σε τεράστια σύνολα δεδομένων, που περιλαμβάνουν βιβλία, άρθρα, ιστοσελίδες και άλλες πηγές γραπτού λόγου, επιτρέποντας στο μοντέλο να μάθει πώς να μιμείται τον ανθρώπινο τρόπο γραφής και σκέψης. Τα LLMs χρησιμοποιούνται σε πληθώρα εφαρμογών, όπως η αυτόματη δημιουργία κειμένων, η μετάφραση γλωσσών, η ανάλυση συναισθημάτων και η απάντηση σε ερωτήσεις, αλλάζοντας τον τρόπο που αλληλεπιδρούμε με τις μηχανές και διευρύνοντας τις δυνατότητες της τεχνητής νοημοσύνης.

2.6 Γνωστές Μέθοδοι Δημιουργίας Συστημάτων Συστάσεων

Στον τομέα των συστημάτων συστάσεων, υπάρχουν τρεις κύριες κατηγορίες μεθόδων τα content-based, τα συστήματα συνεργατικού φιλτραρίσματος, και τα υβριδικά συστήματα.

Content-based

Αυτά τα συστήματα εστιάζουν στα χαρακτηριστικά των αντικειμένων. Για παράδειγμα, σε ένα σύστημα συστάσεων βιβλίων, τα χαρακτηριστικά μπορεί να περιλαμβάνουν τον συγγραφέα, το είδος, ή το θέμα του βιβλίου. Όταν ένας χρήστης εκφράζει θετική προτίμηση για ένα αντικείμενο, το σύστημα προτείνει άλλα αντικείμενα με παρόμοια χαρακτηριστικά.

Συστήματα συνεργατικού φιλτραρίσματος

Βασίζονται στην ιδέα ότι χρήστες που συμφώνησαν στο παρελθόν θα συμφωνήσουν και στο μέλλον. Για παράδειγμα, αν δύο χρήστες έχουν παρόμοιες προτιμήσεις για συγκεκριμένες ταινίες, τότε μια ταινία που αρέσει στον έναν πιθανότατα θα αρέσει και στον άλλον.

Υβριδικά συστήματα

Τα υβριδικά συστήματα είναι ένας συνδυασμός των content-based και των συνεργατικών συστημάτων και καταφέρνουν να εκμεταλλεύονται τα πλεονεκτήματα και των δύο. Μπορούν να προσαρμόζουν τις προτάσεις τους με βάση το περιεχόμενο των αντικειμένων και τις προτιμήσεις άλλων χρηστών, προσφέροντας πιο διευρυμένες και ποικίλες επιλογές στους χρήστες.

Αυτές οι μέθοδοι αντιμετωπίζουν προκλήσεις όπως η σπανιότητα των δεδομένων, η δυσκολία κλιμάκωσης (scalability issue) και το πρόβλημα κρύας εκκίνησης (cold-start problem), επιδιώκοντας να προσφέρουν πιο αποτελεσματικές και εξατομικευμένες συστάσεις στους χρήστες σε διάφορους τομείς εφαρμογών. [4]

Οι παραπάνω προσεγγίσεις αν και αρκετά υποσχόμενες, έχουν αρχίσει να μοιάζουν παρωχημένες με τα σημερινά δεδομένα. Με την ανάπτυξη των τεχνολογιών και τη διαρκή αύξηση του αριθμού των δεδομένων που είναι διαθέσιμα προς τους χρήστες, είναι πλέον απαραίτητη η εύρεση καινούριων προσεγγίσεων οι οποίες θα μπορούν να κλιμακωθούν και να προσφέρουν λύσεις για μεγάλους όγκους δεδομένων. Αυτός είναι ο λόγος που οι πλέον σύγχρονες προσεγγίσεις στα συστήματα συστάσεων επικεντρώνονται στην ενσωμάτωση τεχνολογιών βαθιάς μάθησης και υβριδικών μεθόδων. Αυτές οι τεχνικές επιτρέπουν μια πιο δυναμική ανάλυση δεδομένων και προσφέρουν βελτιωμένη προσαρμογή στις ανάγκες και τις προτιμήσεις των χρηστών. Η αναγνώριση πολύπλοκων μοτίβων συμπεριφοράς και η προσφορά πιο εξατομικευμένων συστάσεων είναι κεντρικά στοιχεία αυτών των προσεγγίσεων. Παράλληλα, η συνεχής ανάπτυξη και εξέλιξη των συστημάτων αυτών είναι κρίσιμη για την αντιμετώπιση νέων προκλήσεων και την εκμετάλλευση ευκαιριών στον ταχύτατα εξελισσόμενο τεχνολογικό κόσμο. **[5]**

2.7 Είδη δεδομένων και μετρικές αξιολόγησης των συστημάτων συστάσεων

Στα συστήματα συστάσεων, η επιλογή και η χρήση των δεδομένων, καθώς και οι μετρικές αξιολόγησης που εφαρμόζονται, είναι κρίσιμες για την αποτελεσματικότητα και την ακρίβεια των συστημάτων. Πιο αναλυτικά:

Είδη Δεδομένων:

1. Δεδομένα Προτιμήσεων: Το πιο κοινό είδος δεδομένων στα συστήματα συστάσεων είναι τα δεδομένα προτιμήσεων των χρηστών. Αυτά περιλαμβάνουν συχνά βαθμολογίες προϊόντων ή υπηρεσιών, όπως αξιολογήσεις με αστέρια, ποσοστά ικανοποίησης, μηχανισμός εκδήλωσης αρέσκειας (liking mechanism) κ.α.

2. Δεδομένα Συμπεριφοράς: Περιλαμβάνουν πληροφορίες για την αλληλεπίδραση των χρηστών με διάφορα στοιχεία, όπως ιστορικό αγορών, προβολές προϊόντων, ή ακόμα και χρόνος παραμονής σε μια σελίδα.

3. Δεδομένα Περιεχομένου: Τα δεδομένα αυτά αφορούν στοιχεία όπως περιγραφές προϊόντων, κατηγορίες, ετικέτες και άλλα χαρακτηριστικά που μπορούν να βοηθήσουν το σύστημα να κατανοήσει και να κατατάξει τα προϊόντα. Πολλές φορές αποκαλούνται και μεταδεδομένα των προϊόντων καθώς δεν αναφέρονται στο ίδιο το προϊόν αλλά σε χαρακτηριστικά αυτού.

Μετρικές Αξιολόγησης:

1. Ακρίβεια Προβλέψεων: Μετράει πόσο κοντά στις πραγματικές προτιμήσεις των χρηστών είναι οι προβλέψεις του συστήματος. Μετρικές όπως το Mean Absolute Error (MAE) και το Root Mean Square Error (RMSE) χρησιμοποιούνται συχνά για την αξιολόγηση της ακρίβειας.

2. Ανάκληση και Ακρίβεια (Recall and Precision): Αυτές οι μετρικές εξετάζουν το πόσο καλά το σύστημα μπορεί να βρει όλα τα σχετικά αντικείμενα (ανάκληση) και πόσο από τα προτεινόμενα αντικείμενα είναι πραγματικά σχετικά (ακρίβεια).

3. F1 Score: Είναι η αρμονική μέση της ακρίβειας και της ανάκλησης και χρησιμοποιείται για να παρέχει μια ισορροπημένη αξιολόγηση ανάμεσα στις δύο αυτές μετρικές.

4. Διαφορετικότητα και Νεωτερικότητα: Αυτές οι μετρικές αξιολογούν την ικανότητα του συστήματος να προτείνει νέα ή ασυνήθιστα αντικείμενα, αντί για αντικείμενα που ο χρήστης ήδη γνωρίζει ή έχει δει.

5. Χρόνος Απόκρισης: Ειδικά σε online συστήματα, ο χρόνος απόκρισης είναι κρίσιμος για την εμπειρία του χρήστη. Μετρά την ταχύτητα με την οποία το σύστημα μπορεί να παράγει συστάσεις.

6. Αντανάκλαση Πραγματικής Εμπειρίας: Το feedback των χρηστών αντανακλά την πραγματική τους εμπειρία και ικανοποίηση από τις προτάσεις του συστήματος. Αυτό μπορεί να παρέχει πιο άμεση και συγκεκριμένη ανατροφοδότηση σε σχέση με απλές μετρικές όπως η ακρίβεια ή η ανάκληση.

Καθώς τα συστήματα συστάσεων εξελίσσονται, η σημασία της αποτελεσματικής διαχείρισης δεδομένων και η επιλογή των κατάλληλων μετρικών αξιολόγησης γίνονται όλο και πιο κρίσιμες για την επίτευξη υψηλής απόδοσης και ποιότητας συστάσεων. [6]

2.8 Προκλήσεις που αντιμετωπίζουν τα συστήματα συστάσεων

Η πρώτη μεγάλη πρόκληση είναι η μεγάλη ποσότητα δεδομένων που υπάρχει διαθέσιμη. Στην εποχή της μαζικής δημιουργίας δεδομένων, τα συστήματα συστάσεων πρέπει να είναι σε θέση να διαχειριστούν και να αναλύσουν τεράστιες ποσότητες πληροφοριών για να είναι ακριβή και σχετικά. Αυτό απαιτεί πολύπλοκους αλγορίθμους και μεγάλες υπολογιστικές υποδομές.

Μια άλλη σημαντική πρόκληση είναι η "ποικιλότητα και προσωποποίηση". Παρά την αυξημένη ικανότητα των συστημάτων συστάσεων να αναλύουν δεδομένα, η δημιουργία προσαρμοσμένων και ποικίλων συστάσεων παραμένει δύσκολη. Υπάρχει ο κίνδυνος τα συστήματα να παράγουν φυσαλίδες φίλτρου (Filter Bubbles) όπου οι χρήστες λαμβάνουν συστάσεις που ενισχύουν τις υπάρχουσες προτιμήσεις τους, αντί να τους εκθέτουν σε νέο περιεχόμενο. Αυτό το πρόβλημα είναι επίσης γνωστό ως πρόβλημα υπερεξειδίκευσης (overspecification problem) και είναι ένα από τα σημαντικότερα προβλήματα των συστημάτων αυτών, καθώς σε μεγάλες και κρίσιμες εφαρμογές μπορεί να αποκλείσει τον χρήστη από σημαντικό περιεχόμενο.

Επίσης, η προστασία της ιδιωτικότητας και η ασφάλεια δεδομένων, παίζει έναν πολύ σημαντικό ρόλο στην ανάπτυξη και συντήρηση ενός τέτοιου συστήματος. Με την αυξανόμενη ευαισθησία γύρω από τα προσωπικά δεδομένα, τα συστήματα συστάσεων πρέπει να είναι σε θέση να προσφέρουν υψηλού επιπέδου ασφάλεια και να σέβονται την ιδιωτικότητα των χρηστών, ενώ ταυτόχρονα παρέχουν σχετικές και προσωπικές συστάσεις.

Τέλος, ένα ακόμη σημαντικό ζήτημα των συστημάτων συστάσεων είναι η "πρόκληση της διαφάνειας και της εξηγησιμότητας". Οι χρήστες και οι ρυθμιστικές αρχές απαιτούν ολοένα και περισσότερο να κατανοούν πώς λειτουργούν αυτά τα συστήματα και πώς λαμβάνονται οι αποφάσεις σχετικά με τις προτάσεις που προσφέρουν. Η δημιουργία διαφανών αλγορίθμων που μπορούν να εξηγηθούν είναι ένας σημαντικός τομέας εργασίας για τους ερευνητές και τους αναπτυξιακούς μηχανικούς.

2.9 Το πρόβλημα κρύας εκκίνησης (cold start problem)

Το πρόβλημα της κρύας εκκίνησης (cold start problem) στα συστήματα συστάσεων αναφέρεται στη δυσκολία που αντιμετωπίζουν αυτά τα συστήματα όταν πρέπει να παρέχουν αξιόπιστες και προσωποποιημένες συστάσεις σε νέους χρήστες ή για νέα προϊόντα, για τα οποία δεν υπάρχουν αρκετά δεδομένα. Για τους νέους χρήστες, το σύστημα δεν έχει αρκετές πληροφορίες για τις προτιμήσεις τους, καθώς δεν έχουν αλληλεπιδράσει ακόμα με το σύστημα. Αντίστοιχα, για τα νέα προϊόντα, λείπουν ιστορικά δεδομένα αλληλεπιδράσεων χρηστών που θα βοηθούσαν το σύστημα να κατανοήσει και να εντοπίσει τις πιθανές προτιμήσεις συγκεκριμένων χρηστών για αυτά τα προϊόντα. Αυτό το πρόβλημα αποτελεί μια σημαντική πρόκληση, καθώς η έλλειψη δεδομένων μπορεί να οδηγήσει σε λιγότερο ακριβείς ή λιγότερο σχετικές συστάσεις, επηρεάζοντας την εμπειρία του χρήστη και την αποδοτικότητα του συστήματος. [15]

2.10 Πλατφόρμες κοινωνικής δικτύωσης

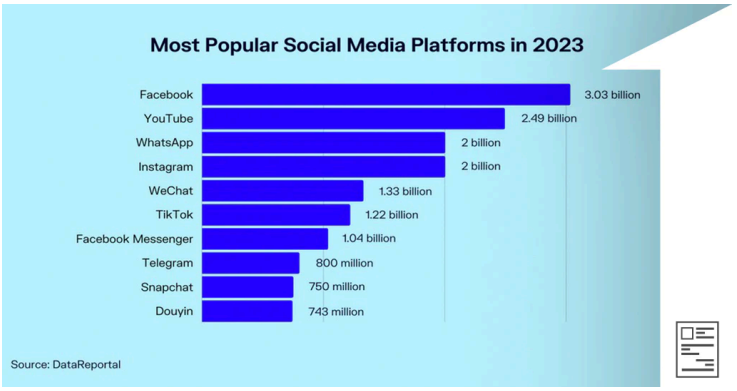
Οι πλατφόρμες κοινωνικής δικτύωσης αποτελούν μια σύγχρονη φαινομενική εξέλιξη στον τομέα της επικοινωνίας και της αλληλεπίδρασης. Αυτές οι πλατφόρμες, όπως το Facebook, το Instagram, το X, το LinkedIn και το Pinterest, έχουν μεταμορφωθεί από απλά μέσα για προσωπική επικοινωνία και δικτύωση σε πολυδιάστατες πλατφόρμες που προσφέρουν ευκαιρίες για εμπορική δραστηριότητα, διαφήμιση, πολιτική κινητοποίηση, κοινωνική αλλαγή κ.α. Χρησιμοποιούνται όχι μόνο για να διατηρούν οι άνθρωποι τις κοινωνικές τους επαφές, αλλά και για να ενημερώνονται για ειδήσεις και τάσεις, να μοιράζονται και να αναζητούν πληροφορίες, να προωθούν επαγγελματικές ευκαιρίες, και να συμμετέχουν σε κοινότητες με κοινά ενδιαφέροντα.

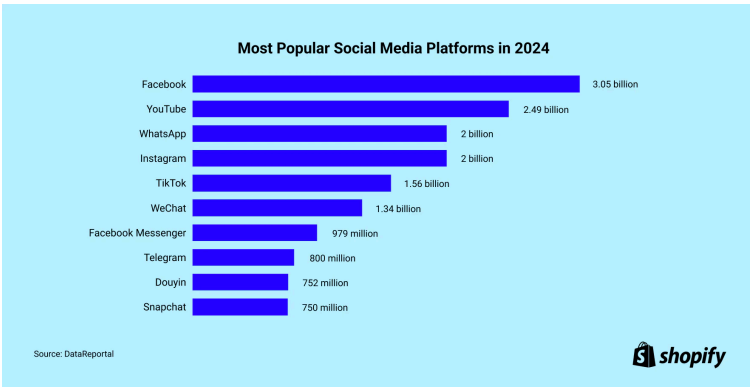
Η λειτουργικότητα των πλατφορμών κοινωνικής δικτύωσης είναι εξαιρετικά ποικίλη. Χρηστες μπορούν να δημιουργήσουν και να διαχειριστούν προσωπικά προφίλ, να αναρτούν φωτογραφίες και βίντεο, να κοινοποιούν σκέψεις και άρθρα, να ακολουθούν άλλους χρήστες, να σχολιάζουν και να αντιδρούν σε δημοσιεύσεις, και να συμμετέχουν σε ομάδες και φόρουμ συζητήσεων. Από την άλλη πλευρά, για τις επιχειρήσεις και τους επαγγελματίες, αποτελούν ένα ισχυρό εργαλείο για την προβολή προϊόντων και υπηρεσιών, τη δημιουργία εμπορικών σχέσεων και την ενίσχυση της επωνυμίας τους.

2.11 Κοινωνικά δίκτυα και συστήματα συστάσεων

Στο πλαίσιο αυτό της μεγάλης ποικιλίας χρήσεων και επιδράσεων, τα συστήματα συστάσεων στις πλατφόρμες κοινωνικής δικτύωσης παίζουν έναν κρίσιμο και πολυεπίπεδο ρόλο. Χρησιμοποιώντας σύνθετους αλγορίθμους και τεχνολογίες ανάλυσης δεδομένων, αυτά τα συστήματα μπορούν να ελέγξουν και να αναλύσουν τη συμπεριφορά των χρηστών, τις αλληλεπιδράσεις τους με διάφορα είδη περιεχομένου και τις προτιμήσεις τους, για να παρέχουν προσωποποιημένες προτάσεις. Αυτές οι προτάσεις μπορεί να περιλαμβάνουν την προβολή σχετικών δημοσιεύσεων, την πρόταση νέων φίλων ή επαγγελματικών επαφών, ομάδων με κοινά ενδιαφέροντα, ή ακόμα και εκδηλώσεων που μπορεί να ενδιαφέρουν τον χρήστη. Στην ουσία, αυτά τα συστήματα εργάζονται για να κάνουν την εμπειρία του χρήστη πιο ευχάριστη, προσαρμοσμένη και αποδοτική, εξοικονομώντας χρόνο και παρέχοντας του πιο σχετικές πληροφορίες και διασκεδάσεις.

Εκτός από τη βελτίωση της εμπειρίας του χρήστη, τα συστήματα συστάσεων είναι ζωτικής σημασίας για την επιτυχία των στρατηγικών ψηφιακού μάρκετινγκ. Προσφέρουν προσαρμοσμένη προβολή διαφημίσεων, βοηθώντας έτσι τις επιχειρήσεις να στοχεύσουν στο κατάλληλο κοινό με το σωστό μήνυμα. Αυτή η στοχευμένη διαφήμιση είναι όχι μόνο πιο αποδοτική από την παραδοσιακή διαφήμιση, αλλά βοηθά επίσης τους χρήστες να βρουν προϊόντα και υπηρεσίες που πραγματικά τους ενδιαφέρουν.





Στατιστικά χρήσης των κοινωνικών δικτύων τα τελευταία χρόνια

Πηγή: Shopify Blog

Σε γενικές γραμμές, οι πλατφόρμες κοινωνικής δικτύωσης και τα ενσωματωμένα συστήματα συστάσεων αποτελούν μια πολύπλοκη και δυναμική οικοσυστημική δομή. Μέσω της παροχής προσωποποιημένων περιεχομένων και διαφημίσεων, αυτά τα συστήματα βοηθούν στη διαμόρφωση της ψηφιακής εμπειρίας, ενισχύουν την επικοινωνία και την αλληλεπίδραση μεταξύ των χρηστών, και διευκολύνουν την πρόσβαση σε πληροφορίες και προϊόντα που είναι σημαντικά και ενδιαφέροντα για αυτούς, αναδεικνύοντας έτσι την τεράστια σημασία τους στον σύγχρονο ψηφιακό κόσμο.

Κεφάλαιο 3 - Περιγραφή και υλοποίηση του συστήματος

Για τον σκοπό της έρευνας αυτής της εργασίας, υλοποιήθηκε μια πλατφόρμα ενός κοινωνικού δικτύου βασισμένου σε εικόνες. Παρόμοιες εφαρμογές υπάρχουν ήδη διαθέσιμες για το κοινό, κάποιες από τις οποίες είναι οι εξής:

- **Pinterest, We heart it, Dribbble, FoodGawker**

και άλλες. Στη δική μας μελέτη περίπτωσης, το μοντέλο που ακολουθήσαμε για τη δημιουργία του συστήματος συστάσεων, μοιάζει περισσότερο με αυτό του **Pinterest** με τη διαφορά ότι τα μεταδεδομένα των εικόνων παράγονται από μοντέλα τεχνητής νοημοσύνης με έμφαση στο πεδίο του Deep Learning.

Πιο αναλυτικά, στην εφαρμογή που υλοποιήθηκε, οι χρήστες μπορούν να δημιουργήσουν και να συνδεθούν με τον λογαριασμό τους, να αλληλεπιδράσουν με εικόνες άλλων χρηστών και να ανεβάσουν τις δικές τους εικόνες με τις οποίες μπορούν έπειτα να αλληλεπιδράσουν άλλοι χρήστες. Όταν κάποιος χρήστης ανεβάσει μια εικόνα στη εφαρμογή, αυτή η εικόνα θα περάσει από μια διαδικασία ανάλυσης που θα πραγματοποιήσουν AI μοντέλα έτσι ώστε να παραχθούν τα μεταδεδομένα της εικόνας αυτής. Τα μεταδεδομένα αυτά είναι **context based**, δηλαδή το συμπέρασμα θα βγαίνει μέσα από την ίδια την εικόνα και αυτό που αναπαρίσταται σε αυτή και όχι από εξωτερικούς παράγοντες όπως τα γνωστά **hashtags (#)** που είναι η πιο γνωστή μορφή συλλογής τέτοιου είδους δεδομένων. Το αποτέλεσμα αυτών των δεδομένων θα είναι η αντιστοίχιση των εικόνων σε μια λίστα από κατηγορίες οι οποίες χρησιμοποιούνται αργότερα για την κατηγοριοποίηση των χρηστών και τη δημιουργία νέων real-time συστάσεων κατά τη χρήση της εφαρμογής.

3.1 Δημιουργία πλατφόρμας κοινωνικού δικτύου

Για την υλοποίηση της τελικής εφαρμογής, ήταν απαραίτητη η δημιουργία μιας πλατφόρμας στην οποία οι χρήστες θα μπορούν να συνδεθούν, να αλληλεπιδράσουν με τις εικόνες της εφαρμογής και να ανεβάσουν νέο υλικό. Για τον σκοπό αυτό επιλέχθηκαν οι παρακάτω τεχνολογίες:

- **React** - Για τη δημιουργία του front-end (μπροστά μέρους) της εφαρμογής και την επικοινωνία με το back-end (πίσω μέρος).
- **Firebase** - Για τη δημιουργία της βάσης δεδομένων, τη φιλοξενία (hosting) της εφαρμογής και την αυθεντικοποίηση των χρηστών.

3.1.1 Αναλυτική περιγραφή των εργαλείων:



- Το React είναι ένα δημοφιλές, ανοιχτού κώδικα, JavaScript framework το οποίο αναπτύχθηκε από το Facebook για τη δημιουργία διεπαφών χρήστη σε web εφαρμογές. Είναι ειδικά σχεδιασμένο για την ανάπτυξη μονοσέλιδων εφαρμογών (single-page applications, SPAs) και χαρακτηρίζεται από την αποδοτικότητα και την ευελιξία του. Ένα από τα κύρια χαρακτηριστικά του React είναι η χρήση των στοιχείων (components), τα οποία είναι ανεξάρτητα και επαναχρησιμοποιήσιμα κομμάτια κώδικα που αντιπροσωπεύουν μέρη της ίδιας της διεπαφής. Αυτή η προσέγγιση επιτρέπει στους προγραμματιστές να διαχειρίζονται πιο αποτελεσματικά την κατάσταση και τη λειτουργικότητα της εφαρμογής, καθώς και να δημιουργούν πιο σύνθετες διεπαφές με λιγότερα βήματα και μικρότερη πολυπλοκότητα.

Ένα άλλο σημαντικό χαρακτηριστικό του React framework είναι η χρήση του Virtual DOM, μια ελαφριά αναπαράσταση του πραγματικού DOM (Document Object Model) της σελίδας. Αυτή η δυνατότητα βοηθά στη βελτιστοποίηση της απόδοσης της εφαρμογής, καθώς οι αλλαγές στη διεπαφή χρήστη είναι πρώτα εφαρμοσμένες στο Virtual DOM και στη συνέχεια, μέσω ενός αποτελεσματικού μηχανισμού σύγκρισης, ανανεώνονται στο πραγματικό DOM. Αυτό μειώνει τον

αριθμό των αλλαγών που πρέπει να γίνουν στο DOM, οι οποίες είναι συχνά χρονοβόρες και βελτιώνει την απόδοση της εφαρμογής, ειδικά σε περίπτωση σύνθετων διεπαφών.

Το React είναι επίσης γνωστό για τη δυνατότητά του να δουλεύει σε συνδυασμό με άλλες βιβλιοθήκες ή frameworks, όπως το Redux για τη διαχείριση κατάστασης ή το React Router για τη διαχείριση διαδρομών (routing) σε πολυσέλιδες εφαρμογές. Η ευελιξία και η αποδοτικότητα του React framework το κάνουν μια ιδανική επιλογή για την ανάπτυξη σύγχρονων εφαρμογών web.



- Το Firebase είναι μια πλατφόρμα ανάπτυξης εφαρμογών από την Google που παρέχει μια σειρά από εργαλεία και υπηρεσίες για να βοηθήσει τους προγραμματιστές να δημιουργούν, να διαχειρίζονται και να βελτιώνουν τις εφαρμογές τους πιο αποτελεσματικά. Αυτή η πλατφόρμα παρέχει λύσεις για βάσεις δεδομένων, αυθεντικοποίηση χρηστών, online hosting, αποθήκευση αρχείων, data analytics κ.α. Ένα από τα κύρια χαρακτηριστικά του Firebase είναι η Realtime Database, μια βάση δεδομένων NoSQL που επιτρέπει την συγχρονισμένη αποθήκευση και συγχρονισμό δεδομένων σε πραγματικό χρόνο μεταξύ των χρηστών της εφαρμογής. Αυτό καθιστά το Firebase ιδανικό για την ανάπτυξη διαδραστικών εφαρμογών που απαιτούν σταθερή ενημέρωση δεδομένων, όπως online παιχνίδια, chat εφαρμογές, ζωντανές υπηρεσίες ειδήσεων και πλατφόρμες κοινωνικών δικτύων όπως αυτή που υλοποιήθηκε για τον σκοπό αυτής της έρευνας.

Εκτός από τις δυνατότητες βάσης δεδομένων, το Firebase προσφέρει επίσης υπηρεσίες όπως η Firebase Authentication, που διευκολύνει τη διαδικασία ελέγχου ταυτότητας για τους χρήστες, και η Firebase Cloud Messaging, που επιτρέπει στους προγραμματιστές να στέλνουν ειδοποιήσεις και μηνύματα στις συσκευές των χρηστών. Η πλατφόρμα είναι σχεδιασμένη για να είναι ευέλικτη και κλιμακούμενη, προσφέροντας λύσεις για εφαρμογές κάθε μεγέθους, από μικρές startups μέχρι μεγάλες επιχειρήσεις. Το Firebase Analytics παρέχει σημαντικές πληροφορίες και δεδομένα για την απόδοση και τη συμπεριφορά των χρηστών στις εφαρμογές, βοηθώντας τους προγραμματιστές να κατανοήσουν καλύτερα το κοινό τους και να βελτιώσουν την εμπειρία των χρηστών τους. Με την ευρεία γκάμα των διαθέσιμων εργαλείων και την ενσωμάτωσή τους με άλλες υπηρεσίες της Google, το Firebase αποτελεί μια ισχυρή και ευέλικτη λύση για τη σύγχρονη ανάπτυξη εφαρμογών.

3.1.2 Τα μέρη και οι σελίδες της εφαρμογής

Η εφαρμογή αποτελείται από τέσσερις κύριες σελίδες:

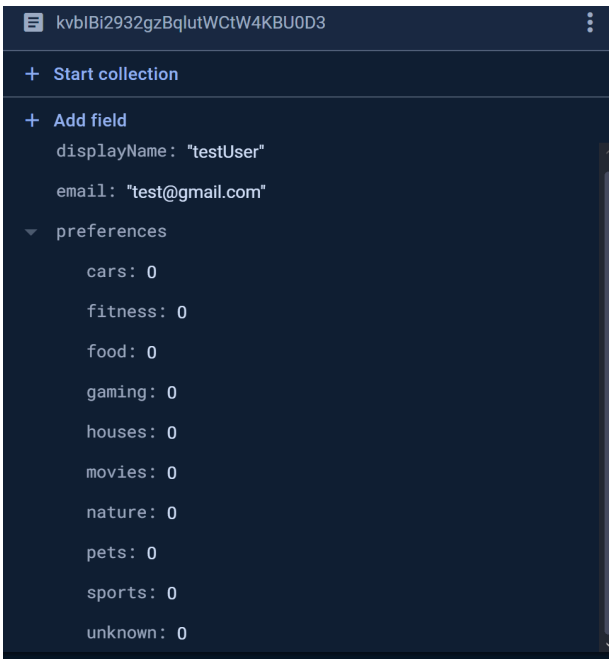
- **Σελίδα εγγραφής/σύνδεσης χρηστών**
- **Αρχική σελίδα**
- **Σελίδα προφίλ χρήστη**
- **Σελίδα εισαγωγής νέων εικόνων.**

Αναλυτικά για την κάθε σελίδα:

Σελίδα εγγραφής/σύνδεσης χρηστών:

Σε αυτή τη σελίδα οι χρήστες μπορούν να πραγματοποιήσουν εγγραφή στην εφαρμογή μέσω email και κωδικού πρόσβασης. Η διαχείριση των στοιχείων αυτών γίνεται αυτόματα από το firebase και οι χρήστες μπορούν να συνδέονται κανονικά στην εφαρμογή με τα στοιχεία αυτά. Επίσης υπάρχει η εναλλακτική επιλογή σύνδεσης του χρήστη με τον google λογαριασμό του, πράγμα που κάνει τη σύνδεση πιο γρήγορη και εύκολη.

Κατά την εγγραφή ενός νέου χρήστη στην εφαρμογή, δημιουργείται ένα νέο document στο “users” collection της βάσης το οποίο περιέχει τα στοιχεία του χρήστη και τις κατηγορίες με τις οποίες είχε τη μεγαλύτερη αλληλεπίδραση (το τελευταίο κομμάτι θα αναλυθεί αργότερα).

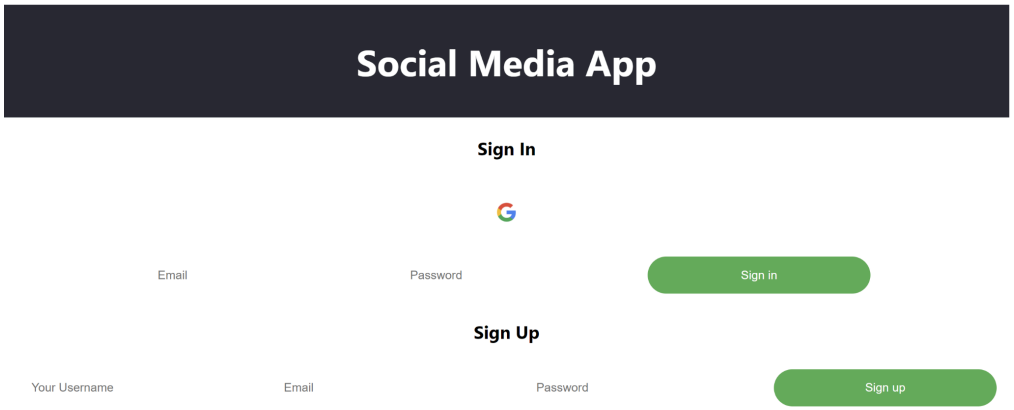


Στο πεδίο “Preferences” φαίνονται οι κατηγορίες στις οποίες έχουν αντιστοιχηθεί οι εικόνες της εφαρμογής. Οι κατηγορίες αυτές είναι οι εξής:

cars, fitness, food, gaming, houses, movies, nature, pets, sports, unknown

Αυτές οι κατηγορίες χρησιμοποιούνται για την εκτίμηση των προτιμήσεων των χρηστών έτσι ώστε το σύστημα να δημιουργεί νέες προτάσεις εικόνων αργότερα βασισμένο σε αυτή την εκτίμηση.

Η σελίδα:



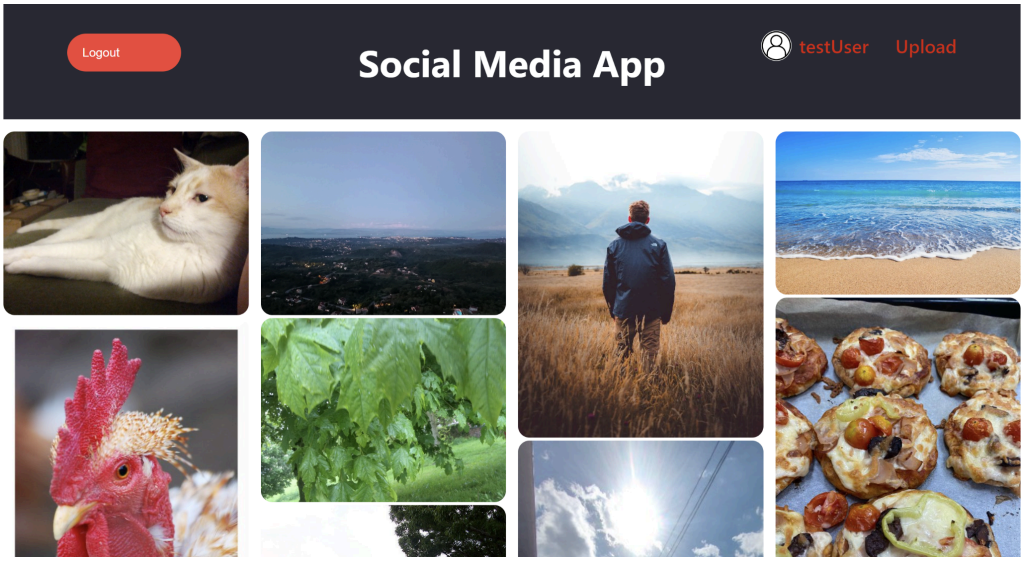
Το πρωτόκολλο ασφαλείας του firebase είναι αρκετά αυστηρό, κάνοντας την εφαρμογή να μην επιτρέπει αδύναμους κωδικούς, emails που έχουν ήδη εγγραφεί κ.α.

Αρχική σελίδα

Κατά την πρώτη είσοδο ενός χρήστη σε αυτή τη σελίδα, η εφαρμογή φέρνει 10 τυχαίες εικόνες από τη βάση. Αν ο χρήστης αλληλεπιδράσει με κάποια από τις εικόνες αυτές τότε θα γίνει μια αυτόματη ενημέρωση στο προφίλ του βάσει της κατηγορίας στην οποία ανήκει η εικόνα αυτή. Η αλληλεπίδραση του χρήστη - εικόνας γίνεται με την χρήση των **likes** (μου αρέσει) ενώ οι προτάσεις του συστήματος που αναπτύχθηκε, είναι με τη μορφή νέων εικόνων κατά τη χρήση της εφαρμογής. Όταν ο χρήστης φτάσει στο τέλος της σελίδας ή όταν επαναφορτώσει τη σελίδα, η εφαρμογή θα φέρει εικόνες οι οποίες είναι πιο κοντά σε αυτές με τις οποίες έχει αλληλεπιδράσει. Η περιγραφή του αλγορίθμου θα αναλυθεί αργότερα.

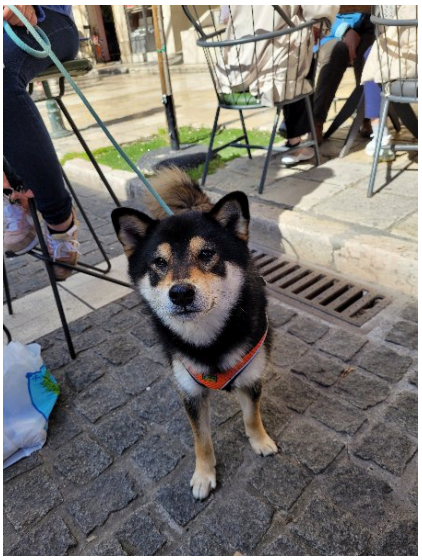
Παράδειγμα λειτουργικότητας της εφαρμογής:

Έστω ότι ένας χρήστης συνδέεται στην εφαρμογη για πρώτη φορά. Αφού κάνει login και επισκεφθεί την αρχική σελίδα, θα δει κάτι σαν αυτό:



Κάθε μια από αυτές τις εικόνες ανήκει σε μια από τις εννέα κατηγορίες οι οποίες προαναφέρθηκαν.

Παραδείγματα εικόνων και των κατηγοριών στις οποίες ανήκουν:

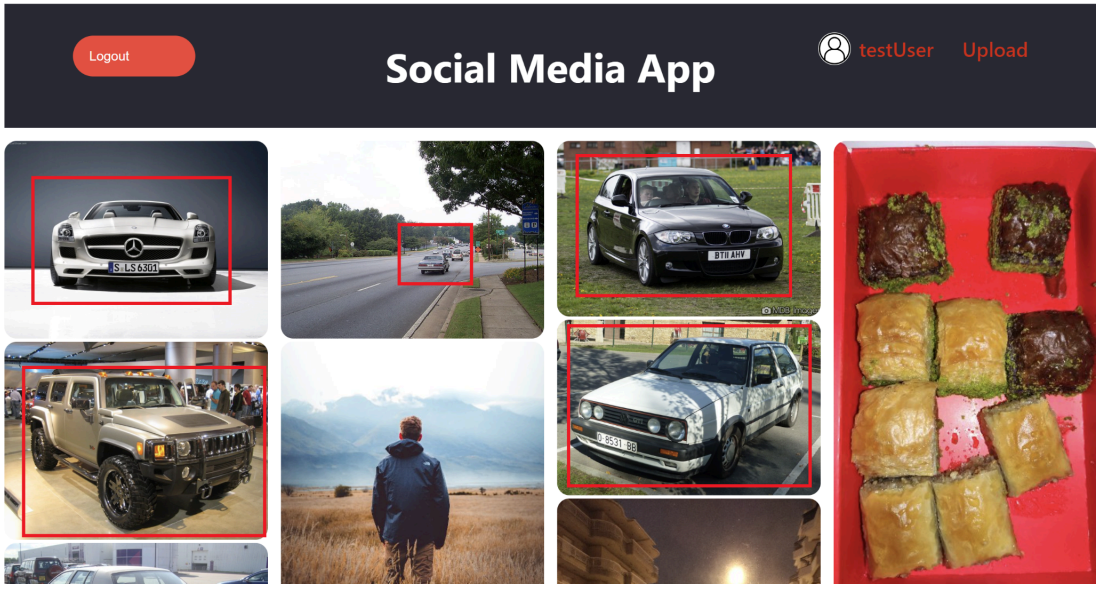






pets **food** **cars**

Αν ο χρήστης, κάνει like σε μια εικόνα η οποία βρίσκεται στην κατηγορία **cars**, τότε η εφαρμογή θα δώσει μεγαλύτερο βάρος στην κατηγορία αυτή με αποτέλεσμα να εμφανίζονται στον χρήστη περισσότερες εικόνες όπου απεικονίζονται αυτοκίνητα.

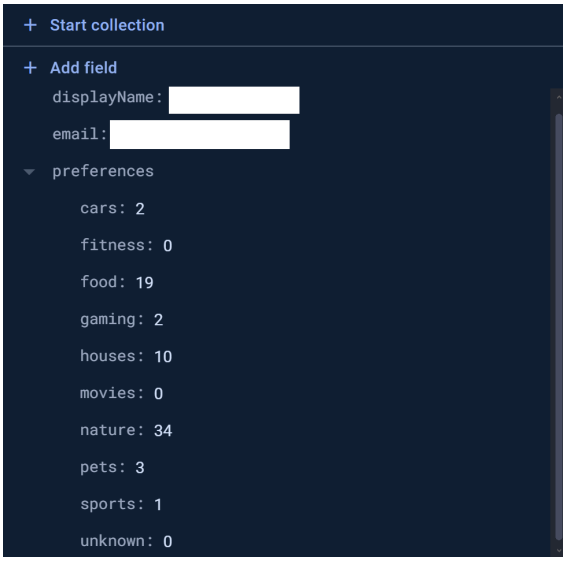


Στην εικόνα φαίνεται το αποτέλεσμα όταν ο χρήστης προτιμά την κατηγορία “cars” (αυτοκίνητα).

Αλγόριθμος επιλογής εικόνων

Για να γίνει μια νέα πρόταση εικόνων ήταν απαραίτητη η δημιουργία ενός συστήματος παρακολούθησης των κατηγοριών με τις οποίες ο χρήστης αλληλεπιδρά περισσότερο. Κατά τη δημιουργία ενός νέου προφίλ χρήστη, δημιουργείται ένα νέο αρχείο στη βάση δεδομένων το οποίο περιέχει όλες τις κατηγορίες μαζί με έναν αριθμό δίπλα από κάθε κατηγορία. Αυτός ο αριθμός, είναι το σύνολο των εικόνων αυτής της κατηγορίας που ο χρήστης έχει αλληλεπιδράσει (έχει κάνει like). Η εφαρμογή ελέγχει τη βαθμολογία αυτή, επιλέγει τις τρεις κατηγορίες με τη μεγαλύτερη βαθμολογία και τους αναθέτει ένα βάρος (weight). Τα βάρη είναι **0.5** για την πρώτη κατηγορία, **0.3** για τη δεύτερη και **0.2** για την τρίτη. Τα νούμερα αυτά προέκυψαν μετά από μια σειρά πειραμάτων με πραγματικούς χρήστες για να δούμε (μέσω της εμπειρίας τους) αν τα αποτελέσματα που λαμβάνουν συνάδουν με την ιδεολογία της εφαρμογής ή αν ο αλγόριθμος χρειάζεται βελτιστοποίηση.

Παράδειγμα:



Στην περίπτωση αυτού του χρήστη οι τρεις κατηγορίες με τη μεγαλύτερη βαθμολογία είναι οι εξής:

nature: 34 -> weight 0.5

food: 19 -> weight 0.3

houses: 10 -> weight 0.2

Κατά τη φόρτωση της σελίδας, η εφαρμογή θα φορτώσει 10 νέες εικόνες από τη βάση. Το 70% των εικόνων αυτών θα αντιστοιχίζεται στις προτιμήσεις του χρήστη ενώ το υπόλοιπο 30% θα είναι τυχαίες εικόνες έτσι ώστε ο χρήστης να έχει την επιλογή να αλλάξει τις προτιμήσεις του μελλονικά.

Το 70% των εικόνων που βασίζεται στις προτιμήσεις του χρήστη, μοιράζεται στις 3 πρώτες κατηγορίες που προαναφέρθηκαν ανάλογα με τον βαθμό των βαρών τους. Έτσι το αποτέλεσμα για τον συγκεκριμένο χρήστη θα είναι το εξής:

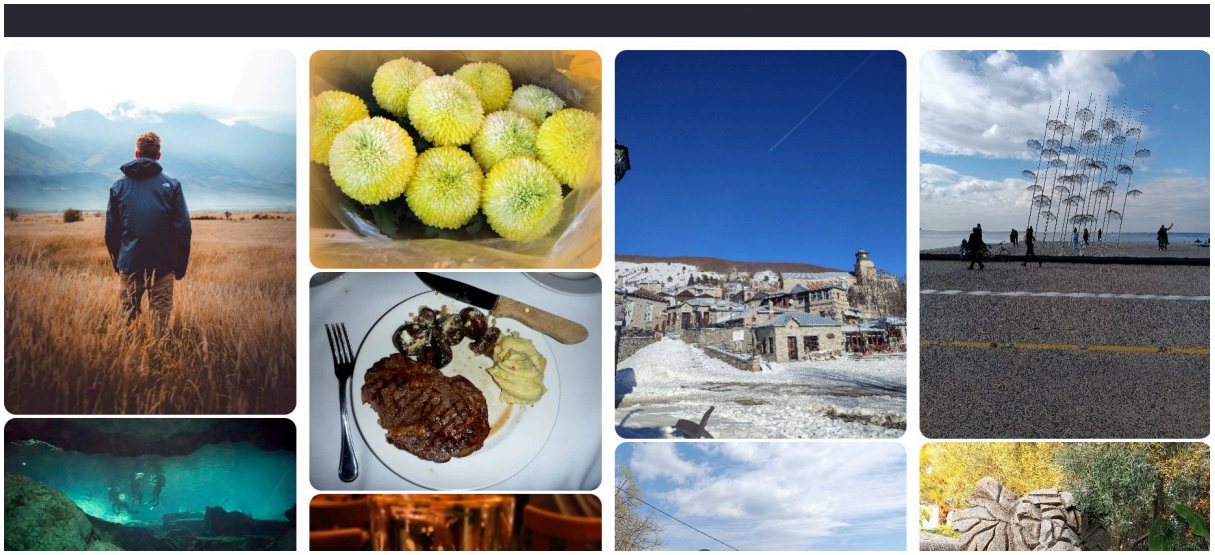


Σημείωση: Ο λόγος που τα τυχαία αποτελέσματα είναι 3xnature, είναι επειδή η συνάρτηση τυχαιότητας λαμβάνει σαν seed (αρχικοποίηση της γεννήτριας τυχαίων αριθμών) την ημερομηνία

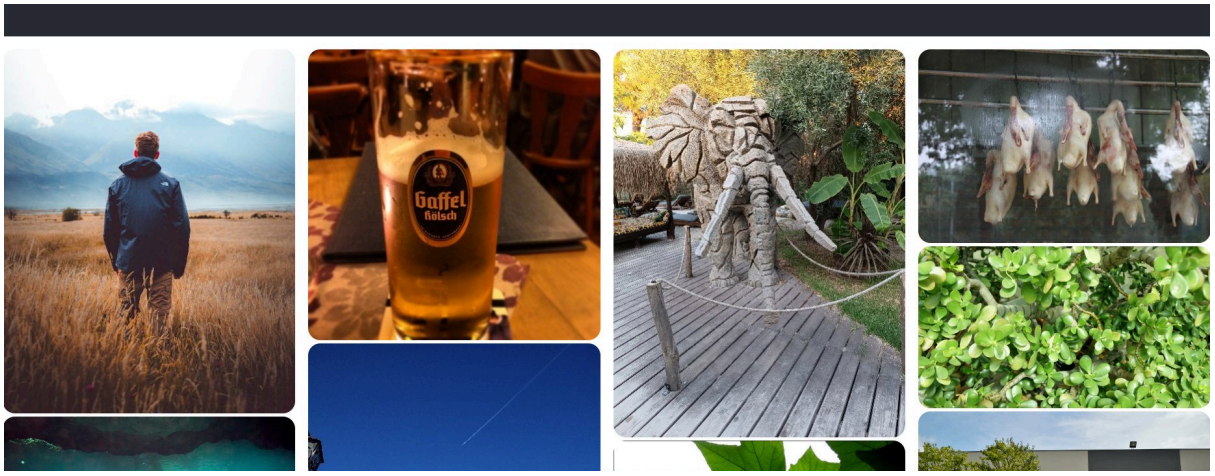
των προστέθηκαν οι εικόνες στην εφαρμογή και θεωρείται ντετερμινιστική. Με την προσθήκη νέων εικόνων κατά τη χρήση της εφαρμογής, οι τυχαίες εικόνες θα αλλάζουν.

Αρχική σελίδα χρήστη του παραδείγματος

Παράδειγμα 1:



Παράδειγμα 2:



Αναλυτικά ο αλγόριθμος των εικόνων

Ορισμοί

1. **Προτιμήσεις Χρήστη (U)**:
   - Ένα διάνυσμα που αντιπροσωπεύει τον αριθμό των "μου αρέσει" για κάθε κλάση από τον χρήστη.
   - $U = [u_1, u_2, ..., u_9]$ όπου u_i είναι ο αριθμός των "μου αρέσει" για την κλάση i.
2. **Συνολικές Εικόνες προς Ανάκτηση (N)**:
   - Μια σταθερά που αντιπροσωπεύει το συνολικό αριθμό εικόνων που πρέπει να ανακτηθούν.
   - N
3. **Βάρη (W)**:
   - Ένα σύνολο βαρών για τις τρεις κορυφαίες κλάσεις.
   - $W = \{w_1 = 0.5, w_2 = 0.3, w_3 = 0.2\}$
4. **Κατάταξη Κλάσης (R)**:
   - Μια αντιστοίχιση των κλάσεων με την κατάταξή τους βάσει των προτιμήσεων του χρήστη.
   - $R(i)$ είναι η κατάταξη της κλάσης i βάσει του U.

Εξισώσεις

1. **Αριθμός Εικόνων για τις Κορυφαίες Κλάσεις**:
   - Για κάθε μία από τις τρεις κορυφαίες κλάσεις i, ο αριθμός των εικόνων προς ανάκτηση δίνεται από την εξίσωση:
     $n_i = \text{στρογγυλοποίηση}(w_{R(i)} \times 0.7 \times N)$
   - Όπου $\text{στρογγυλοποίηση}()$ είναι μια συνάρτηση που στρογγυλοποιεί στον πλησιέστερο ακέραιο.
2. **Αριθμός Τυχαίων Εικόνων**:
   - Ο αριθμός των τυχαίων εικόνων προς ανάκτηση είναι:
     $n_{\text{τυχαίο}} = N - \sum_{i=1}^{3} n_i$
3. **Συνολικές Εικόνες που Ανακτήθηκαν**:
   - Ο συνολικός αριθμός των εικόνων που πρέπει να ανακτηθούν θα πρέπει ιδανικά να είναι:
     $N = \sum_{i=1}^{3} n_i + n_{\text{τυχαίο}}$
   - Λόγω της στρογγυλοποίησης στο n_i, μπορεί να υπάρχει μικρή απόκλιση, η οποία πρέπει να διορθωθεί κατά την υλοποίηση.

Ανάλυση του αλγορίθμου σε μορφή κώδικα (γλώσσα προγραμματισμού: Javascript)

Αρχικοποίηση της συνάρτησης "**fetchUserPreferences**" η οποία επιστρέφει τις κατηγορίες και τις βαθμολογίες που έχει ο εκάστοτε χρήστης στο προφίλ του. []

```
const fetchUserPreferences = async (userId, db) => {
    // Reference to the user's document in the 'users' collection
    const userRef = db.collection('users').doc(userId);

    try {
        const doc = await userRef.get();
        if (doc.exists) {
            const preferences = doc.data().preferences;
            return preferences; // This should be an object like { nature: 3, cars: 5, movies: 1 }
        } else {
            console.log('No such user document!');
            return {}; // Return an empty object if the user document doesn't exist
        }
    } catch (error) {
        console.error('Error fetching user preferences:', error);
        throw error; // Rethrow the error for calling function to handle
    }
};
```

Αρχικοποίηση της συνάρτησης του αλγορίθμου και των απαραίτητων μεταβλητών "**fetchImagesForUser**"

```
export const fetchImagesForUser = async (userId, db, startAfter, totalImagesToFetch, fetchedIds) => {
    const PREFERENCE_RATIO = 0.7; // 70% images based on preferences
    const WEIGHTS = [0.5, 0.3, 0.2]; // Weights for top 3 categories
    const preferences = await fetchUserPreferences(userId, db);
    const sortedPreferences = Object.entries(preferences).sort((a, b) => b[1] - a[1]); // Sort preferences based on score

    let lastDoc = startAfter; // This is used by the main app to avoid fetching the same images after each scroll
    let images = [];
    let currentFetchedIds = new Set(fetchedIds); // This is used by this function to avoid fetching the same images after each scroll
```

Συλλογή των εικόνων βάσει των προτιμήσεων του χρήστη και των βαρών τους

```
// Preference based image fetching
for (let i = 0; i < 3; i++) { // Fetch images for the top 3 categories
    const [category, _] = sortedPreferences[i];
    let categoryImagesToFetch = totalImagesToFetch * PREFERENCE_RATIO * WEIGHTS[i];
    if (categoryImagesToFetch) { // round the weights based on the decimal part
        const decimalPart = categoryImagesToFetch % 1;
        if (decimalPart >= 0.5) {
            categoryImagesToFetch = Math.ceil(categoryImagesToFetch);
        } else {
            categoryImagesToFetch = Math.floor(categoryImagesToFetch);
        }
    }
    let query = db.collection('images').where('metadata.predicted_class', '==', category);
    if (lastDoc) {
        query = query.startAfter(lastDoc);
    }
    query = query.limit(categoryImagesToFetch); // Limit to the number that is calculated using weights

    const snapshot = await query.get();
    console.log(`Fetching ${categoryImagesToFetch} images from category: ${category}`);

    snapshot.forEach((doc) => {
        if (!currentFetchedIds.has(doc.id)) {
            images.push({ ...doc.data(), id: doc.id });
            lastDoc = doc;
            currentFetchedIds.add(doc.id);
        }
    });
}
```

Συλλογή επιπλέον εικόνων βάσει της συνάρτησης τυχαιότητας

```
// Fetch the rest of the images without category preference
let randomImagesToFetch = totalImagesToFetch - images.length;
if (randomImagesToFetch > 0) {
    let query = db.collection('images').orderBy('metadata.timestamp');
    if (lastDoc) {
        query = query.startAfter(lastDoc);
    }
    query = query.limit(randomImagesToFetch);

    const snapshot = await query.get();
    snapshot.forEach((doc) => {
        if (!currentFetchedIds.has(doc.id)) {
            images.push({ ...doc.data(), id: doc.id });
            lastDoc = doc;
            console.log(`Fetched random image from category: ${doc.data().metadata.predicted_class}`);
            currentFetchedIds.add(doc.id);
        }
    });
}

console.log(`Total images fetched: ${images.length}`);
console.log('Fetched image IDs:', images.map(img => img.id));
return { images, lastDoc };
```

Ο παραπάνω κώδικας στο παρακάτω github repository:

https://github.com/Mnlsvt/recom_system/blob/main/Pint_UI/src/functions/user_classification/user_image_recommendation.js

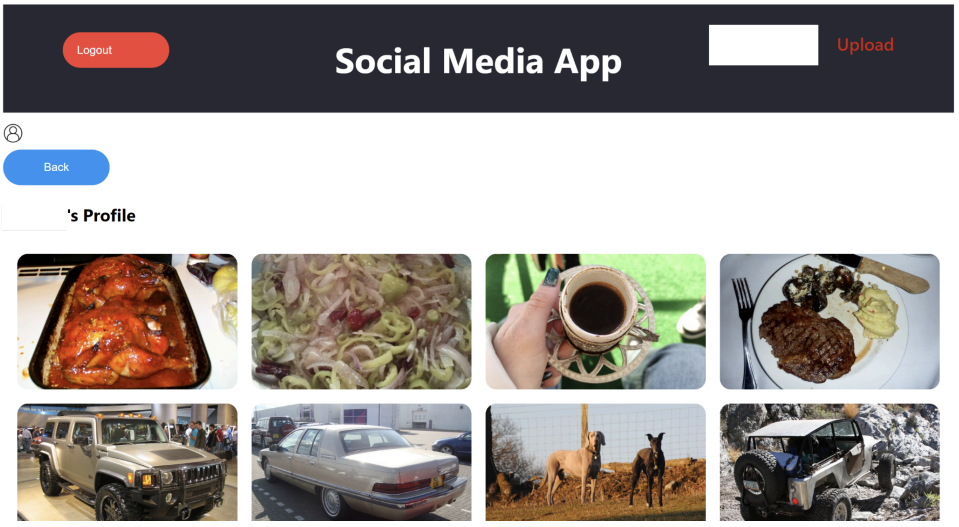
Το πρόβλημα της κρύας εκκίνησης στην πλατφόρμα

Κατά την είσοδο νέων χρηστών στην πλατφόρμα, υπάρχει το πρόβλημα της έλλειψης απαραίτητων δεδομένων για τη δημιουργία συστάσεων. Αυτό είναι ουσιαστικά το πρόβλημα της κρύας εκκίνησης (cold start problem) το οποίο περιγράφεται αναλυτικά στην ενότητα **2.9**. Για την αντιμετώπιση αυτού του προβλήματος, κατά την είσοδο νέων χρηστών, η πλατφόρμα προσφέρει τυχαία δεδομένα από διάφορες κλάσεις/κατηγορίες έτσι ώστε ο χρήστης να έχει μια ποικιλία επιλογών και να μπορεί εύκολα να δημιουργήσει το προφίλ των προτιμήσεων του. Μέσα από τα πειράματα που διεξήχθησαν, παρατηρήθηκε πως ο χρόνος που χρειάστηκε στους χρήστες της εφαρμογής να δημιουργήσουν το προφίλ των προτιμήσεων τους, είναι κατά μέσο όρο η 1 ώρα. Αυτό βέβαια δε σημαίνει πως ο ίδιος χρόνος προσαρμογής θα ισχύει και σε μια εφαρμογή με χιλιάδες χρήστες και εκατοντάδες κατηγορίες καθώς όσο κλιμακώνεται η εφαρμογή τόσο περισσότερο χρόνο θα παίρνει σε έναν χρήστη να δημιουργήσει ένα εξατομικευμένο προφίλ. **[1]**

Σελίδα προφίλ χρήστη

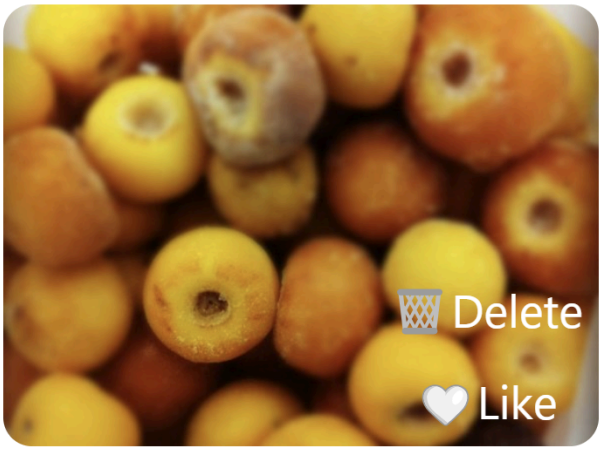
Στη σελίδα αυτή, ο χρήστος μπορεί να δει τις εικόνες που έχει ανεβάσει, να κάνει like σε αυτές ή και να τις διαγράψει.

Η σελίδα:



Επίσης ο χρήστης έχει τη δυνατότητα να αλλάξει την εικόνα προφίλ του κάνοντας κλικ πάνω της.

Η διαδικασία των likes και της διαγραφής των εικόνων γίνεται με εμφάνιση κουμπιών πάνω στις εικόνες κάθε φορα που ο κέρσορας περνάει από πάνω.



Σελίδα εισαγωγής νέων εικόνων

Σε αυτή τη σελίδα, οι χρήστες έχουν τη δυνατότητα να ανεβάσουν είτε μια είτε πολλαπλές εικόνες μέσω του αντίστοιχου πεδίου. Αυτές οι εικόνες ανεβαίνουν απευθείας στη βάση δεδομένων της εφαρμογής και είναι άμεσα συνδεδεμένες με το προφίλ του χρήστη που τις ανέβασε. Έτσι διαφορετικοί χρήστες μπορούν να δουν και να αλληλεπιδράσουν με την εικόνα αλλά μόνο ο αρχικός χρήστης μπορεί να τη διαγράψει.

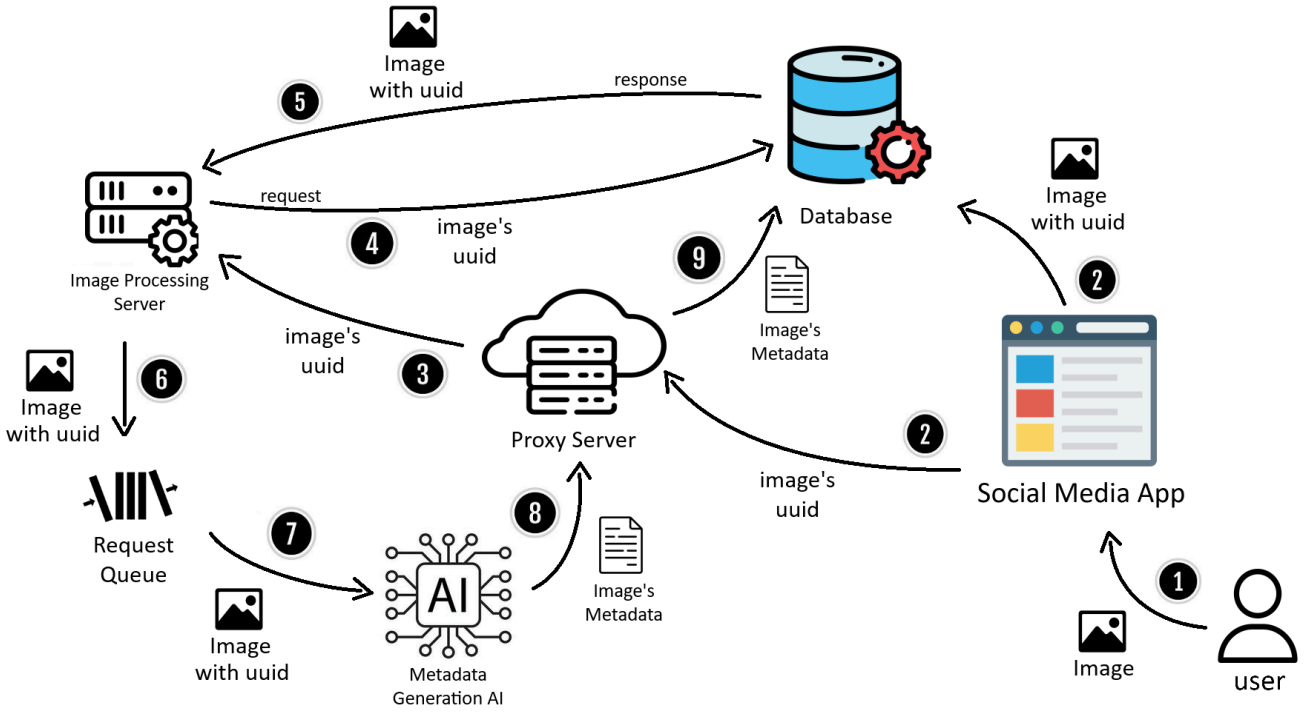
Η σελίδα:



3.2 Ανάλυση των εικόνων

Για να επιτευχθεί ο στόχος της εφαρμογής, είναι απαραίτητη η χρήση ορισμένων μοντέλων βαθιάς μάθησης για την ανάλυση των εικόνων και την παραγωγή των μεταδεδομένων τους. Για τη σωστή λειτουργία αυτών των μοντέλων και τη συμβατότητα τους με τη λογική αυτής της εργασίας, χρειάστηκε να δημιουργηθεί ένα σύστημα στο οποίο θα τροφοδοτούνται οι εικόνες και θα παράγονται τα απαραίτητα μεταδεδομένα.

3.2.1 Η αρχιτεκτονική του συστήματος παραγωγής μεταδεδομένων:



1. Ο χρήστης ανεβάζει μια εικόνα στην εφαρμογή.
2. Η εφαρμογή παράγει έναν μοναδικό αριθμό αναγνώρισης (unique identifier - uuid) και τον δίνει στην εικόνα. Έπειτα στέλνει την εικόνα μαζί με το uuid της στη βάση ενώ ταυτόχρονα στένει το uuid στον ενδιάμεσο διακομιστή (proxy server).
3. Ο proxy server μεταφέρει το uuid της εικόνας στον διακομιστή που είναι υπεύθυνος για την παραγωγή των μεταδεδομένων (Image Processing Server).

4. Ο Image Processing Server στέλνει ένα αίτημα (request) προς τη βάση χρησιμοποιώντας το uuid της εικόνας.
5. Η βάση δέχεται το αίτημα του Image Processing Server, διαβάζει το uuid που έλαβε, το αναζητά στη βάση και μόλις βρει την εικόνα στην οποία αντιστοιχεί, απαντά στον Image Processing Server (response) με την εικόνα που βρήκε.
6. Ο Image Processing Server στέλνει την εικόνα που έλαβε ως απάντηση σε μια ουρά από αιτήματα έτσι ώστε να τα πάρει ο μηχανισμός τεχνητής νοημοσύνης (Metadata Generation AI) και να τα επεξεργαστεί με τη σειρά (για να αποφευχθεί συμφόρηση και απώλεια πόρων).
7. Ο μηχανισμός τεχνητής νοημοσύνης, λαμβάνει την εικόνα από την ουρά και την επεξεργάζεται κατάλληλα έτσι ώστε να παράξει τα απαραίτητα μεταδεδομένα.
8. Αφού τελειώσει η επεξεργασία της εικόνας και παραχθούν τα μεταδεδομένα, ο μηχανισμός τεχνητής νοημοσύνης τα στέλνει στον proxy server.
9. Ο proxy server δρομολογεί τα μεταδεδομένα στη βάση δεδομένων όπου και συσχετίζονται με την εικόνα μέσω του uuid.

3.2.2 Αναλυτική περιγραφή των εργαλείων



- Το TensorFlow είναι μια ανοιχτού κώδικα βιβλιοθήκη μηχανικής μάθησης που αναπτύχθηκε από την Google για να επιτρέπει την εύκολη δημιουργία και εκπαίδευση νευρωνικών δικτύων και μοντέλων βαθιάς μάθησης. Ένας από τους κύριους στόχους του TensorFlow είναι να διευκολύνει την πρόσβαση σε σύνθετες δυνατότητες μηχανικής μάθησης για ερευνητές και για ανθρώπους που φτιάχνουν εφαρμογές πραγματικού κόσμου. Χρησιμοποιείται σε μια ευρεία γκάμα από τομείς, από την αναγνώριση ομιλίας και εικόνων μέχρι την προηγμένη ανάλυση δεδομένων και την δημιουργία

προβλέψεων. Το TensorFlow προσφέρει ευελιξία στον τρόπο σχεδίασης και εκπαίδευσης μοντέλων, υποστηρίζοντας πολλές διαφορετικές αρχιτεκτονικές νευρωνικών δικτύων.

Η πλατφόρμα είναι γνωστή για την ευελιξία και τη δυνατότητά της να λειτουργεί σε διάφορα λειτουργικά συστήματα και πλατφόρμες. Αυτό επιτρέπει στους χρήστες να εκπαιδεύουν και να αναπτύσσουν μοντέλα σε διάφορα περιβάλλοντα, από τοπικούς υπολογιστές μέχρι cloud-based υποδομές και μονάδες επεξεργασίας γραφικών (GPUs). Η βιβλιοθήκη παρέχει επίσης ένα σύνολο εργαλείων για την οπτικοποίηση των δεδομένων και των μοντέλων, όπως το TensorBoard, το οποίο βοηθά τους χρήστες να κατανοήσουν και να βελτιστοποιήσουν τις διαδικασίες εκπαίδευσης και ανάλυσης.



- Η Python είναι μια δημοφιλής, υψηλού επιπέδου, γλώσσα προγραμματισμού με δυναμική σημασιολογία. Αναπτύχθηκε από τον Guido van Rossum και πρώτο κυκλοφόρησε το 1991. Η Python είναι σχεδιασμένη με τη φιλοσοφία να είναι ευανάγνωστη και συντομότερη από άλλες γλώσσες προγραμματισμού, με έμφαση στην καθαρότητα και την απλότητα του κώδικα. Αυτό επιτρέπει στους προγραμματιστές να εκφράζουν ιδέες σε λιγότερες γραμμές κώδικα από ό,τι θα ήταν δυνατόν σε γλώσσες όπως C++ ή Java.

Η Python υποστηρίζει διάφορα παραδείγματα προγραμματισμού, όπως τον αντικειμενοστραφή, τον διαδικαστικό και τον λειτουργικό προγραμματισμό. Αυτό την καθιστά πολύ ευέλικτη και κατάλληλη για μια ευρεία γκάμα εφαρμογών. Είναι επίσης γνωστή για την ισχυρή της τυπική βιβλιοθήκη και για την ευρεία υποστήριξη τρίτων βιβλιοθηκών, παρέχοντας εργαλεία για διάφορες εργασίες όπως την ανάλυση δεδομένων, τη μηχανική μάθηση, την ανάπτυξη web, την αυτοματοποίηση, και πολλά άλλα.

Η Python είναι επίσης πολύ δημοφιλής στην επιστημονική κοινότητα, με πολλές βιβλιοθήκες που παρέχουν λειτουργικότητες για επιστημονικούς υπολογισμούς και ανάλυση δεδομένων, όπως NumPy, Pandas, και Matplotlib. Η ευκολία χρήσης της Python και η ικανότητά της να αντιμετωπίζει μεγάλο εύρος προβλημάτων την καθιστούν εξαιρετική επιλογή για αρχάριους

προγραμματιστές, ενώ παράλληλα παραμένει ισχυρή και ευπροσάρμοστη για σύνθετες εφαρμογές. Η ευρεία χρήση και η ενεργή κοινότητα της Python σημαίνουν ότι υπάρχει ένα πλούσιο οικοσύστημα υποστήριξης και πόρων, καθιστώντας την μια από τις πιο προσβάσιμες και κοινοτικά-υποστηριζόμενες γλώσσες προγραμματισμού σήμερα.

Επιπλέον Βιβλιοθήκες της Python που χρησιμοποιήθηκαν

- torch, torchvision, Pillow, OpenCV, scikit-learn, celery



- Το Flask είναι ένα ελαφρύ, εύχρηστο microframework για την Python, σχεδιασμένο για την ανάπτυξη web εφαρμογών. Αναπτύχθηκε από τον Armin Ronacher και κυκλοφόρησε το 2010. Το Flask διακρίνεται για την απλότητα και την ευελιξία του, παρέχοντας το απαραίτητο πλαίσιο για την ανάπτυξη web εφαρμογών χωρίς να επιβάλλει περιορισμούς στη δομή της εφαρμογής.

Ένας από τους βασικούς λόγους για την δημοτικότητα του Flask είναι η "φιλοσοφία του minimal", η οποία επιτρέπει στους προγραμματιστές να ξεκινήσουν με ένα πολύ βασικό περιβάλλον και να προσθέσουν μόνο τα απαραίτητα επιπλέον εξαρτήματα και βιβλιοθήκες καθώς αναπτύσσεται η εφαρμογή. Αυτό δίνει στους χρήστες την ευελιξία να προσαρμόσουν το πλαίσιο στις ανάγκες τους, αντί να πρέπει να προσαρμοστούν σε ένα πιο σταθερό πλαίσιο με πολλές ενσωματωμένες λειτουργίες.

Το Flask παρέχει τα βασικά εργαλεία και τις βιβλιοθήκες για τη διαχείριση αιτημάτων (requests), αποκρίσεων (responds), συνόδων (sessions) και cookies, καθώς και υποστήριξη για το routing και τα πρότυπα (templates). Η απλότητα του Flask το καθιστά ιδανικό για μικρά έργα ή για την ανάπτυξη πρωτοτύπων, αλλά είναι αρκετά ευέλικτο για να χρησιμοποιηθεί και σε μεγαλύτερες, πιο πολύπλοκες εφαρμογές. Είναι επίσης δημοφιλές στην κοινότητα της Python λόγω της ικανότητάς του να ενσωματώνεται εύκολα με άλλες βιβλιοθήκες Python, όπως οι SQLAlchemy για τη διαχείριση βάσεων δεδομένων και η Marshmallow για τη σειριοποίηση.



- Το ChatGPT είναι ένα προηγμένο σύστημα γλωσσικού μοντέλου από την OpenAI, βασισμένο στην αρχιτεκτονική GPT (Generative Pre-trained Transformer). Πρόκειται για μια τεχνολογία που χρησιμοποιεί τεχνητή νοημοσύνη και μηχανική μάθηση για να αναλύει, κατανοήσει και να παράγει φυσική γλώσσα. Το ChatGPT έχει εκπαιδευτεί σε ένα ευρύ φάσμα κειμένων, από λογοτεχνία και επιστημονικά άρθρα μέχρι πληροφορίες από το διαδίκτυο, προκειμένου να μπορεί να απαντά σε ερωτήσεις, να συνθέτει κείμενα και να παρέχει πληροφορίες σε ποικίλα θέματα.

Το ChatGPT διακρίνεται για την ικανότητά του να διεξάγει συνεπείς και συνεκτικές συνομιλίες, προσομοιώνοντας τον τρόπο που ένας άνθρωπος θα ανταποκρινόταν σε παρόμοιες καταστάσεις. Είναι ικανό να αντιλαμβάνεται περίπλοκες εννοιολογικές συνδέσεις και να παράγει απαντήσεις που δείχνουν βαθιά κατανόηση του θέματος συζήτησης. Αυτό το καθιστά ιδιαίτερα χρήσιμο για εφαρμογές όπως η αυτόματη παραγωγή κειμένου, η εξυπηρέτηση πελατών μέσω chatbots, η δημιουργία εκπαιδευτικού περιεχομένου, και η βοήθεια στη σύνθεση λογοτεχνικών έργων.

Ένας από τους πυλώνες της λειτουργικότητας του ChatGPT είναι η συνεχής εκμάθηση και βελτίωση. Το μοντέλο ενημερώνεται τακτικά με νέα δεδομένα και ανατροφοδότηση χρηστών, βελτιώνοντας συνεχώς την ακρίβεια και την ποικιλία των απαντήσεων του. Επιπλέον, η αρχιτεκτονική του μοντέλου επιτρέπει την ευέλικτη προσαρμογή σε διάφορες γλώσσες και πολιτισμικά πλαίσια, κάνοντας το ChatGPT ένα παγκόσμια εφαρμόσιμο εργαλείο.

3.2.3 Επιλογή των κατάλληλων αλγορίθμων και μοντέλων βαθιάς μάθησης

Στην επιλογή των αλγορίθμων και μοντέλων βαθιάς μάθησης για την ανάλυση εικόνων, η βασική προτεραιότητα είναι η αποδοτικότητα και η ακρίβεια στην παραγωγή μεταδεδομένων. Στο πλαίσιο αυτό, η επιλογή μοντέλων όπως τα συνελικτικά νευρωνικά δίκτυα (CNNs) ή τα αναδρομικά νευρωνικά δίκτυα (RNNs) γίνεται με βάση την ικανότητά τους να αντιμετωπίζουν τις συγκεκριμένες προκλήσεις του έργου μας. Για την εφαρμογή αυτή, επιλέχθηκαν τα εξής μοντέλα:ResNet18, ResNet50, AlexNet, WideResNet18, Faster R-CNN ResNet50 FPN καθένα με μοναδικές ιδιότητες και ικανότητες.

Αναλυτικότερα, το ResNet18 και το ResNet50 χρησιμοποιούνται για την ικανότητά τους να αναγνωρίζουν περίπλοκα μοτίβα στις εικόνες, με το ResNet50 να προσφέρει βαθύτερη ανάλυση λόγω του μεγαλύτερου αριθμού στρωμάτων. Το AlexNet, ένα από τα πρώτα CNN μοντέλα που επέδειξαν εντυπωσιακή απόδοση σε διαγωνισμούς αναγνώρισης εικόνας, χρησιμοποιήθηκε για την αποδεδειγμένη αποτελεσματικότητά του στην ανίχνευση χαρακτηριστικών. Το WideResNet18 παρέχει μια εναλλακτική με βελτιωμένη ικανότητα ανάλυσης σε σύγκριση με το ResNet18, ενώ διατηρεί μικρότερο μέγεθος μοντέλου.

Το Faster R-CNN ResNet50 FPN από την άλλη, είναι ένα προηγμένο μοντέλο για ανίχνευση αντικειμένων που συνδυάζει τις ικανότητες του ResNet50 με την τεχνολογία Faster R-CNN και FPN (Feature Pyramid Networks), παρέχοντας υψηλή ακρίβεια και ταχύτητα στην εντοπισμό και ταξινόμηση αντικειμένων στις εικόνες. Αυτή η διαφοροποιημένη προσέγγιση στην επιλογή των μοντέλων επιτρέπει την αντιμετώπιση ενός ευρέος φάσματος αναγκών και προκλήσεων στην ανάλυση εικόνων, εξασφαλίζοντας ταυτόχρονα υψηλή απόδοση και ακρίβεια.

Επιπρόσθετα, το Mask R-CNN με ResNet50 και FPN είναι μια πολύ καλή επιλογή για εφαρμογές που απαιτούν όχι μόνο ανίχνευση αντικειμένων, αλλά και ακριβή τοποθέτηση μάσκας σε αυτά. Αυτό το μοντέλο είναι μια εξέλιξη του Faster R-CNN, ενσωματώνοντας την αρχιτεκτονική ResNet50 για βαθύτερη και πιο αποτελεσματική εξαγωγή χαρακτηριστικών. Το Feature Pyramid Network (FPN) προστίθεται για να ενισχύσει την ανίχνευση αντικειμένων σε διαφορετικές κλίμακες, επιτρέποντας το μοντέλο να αναγνωρίζει αντικείμενα μικρού μεγέθους με την ίδια ακρίβεια με τα μεγαλύτερα. Η ικανότητα του Mask R-CNN να δημιουργεί μάσκες για κάθε αντικείμενο στην εικόνα το καθιστά ιδανικό για εργασίες στις οποίες είναι απαραίτητη η λεπτομερής σημασιολογική ανάλυση, όπως η ανίχνευση και διαχωρισμός επιμέρους αντικειμένων με ακριβή συνοριογράμματα.

Παρόλο που το κάθε μοντέλο παίζει έναν διαφορετικό ρόλο στην εφαρμογή, το συνολικό αποτέλεσμα την επεξεργασίας των εικόνων που παράγεται είναι συλλογικό και η χρήση πολλαπλών μοντέλων είναι ιδιαίτερα σημαντική. Αυτό συμβαίνει επειδή ο βαθμός εμπιστοσύνης ενός μοντέλου

τέτοιου βελησνεκούς, δεν μπορεί να θεωρηθεί αρκετός για την επίτευξη της λειτουργικότητας μιας τέτοιας εφαρμογής.

3.2.4 Χρήση των μοντέλων

Όπως προαναφέρθηκε, τα μοντέλα που χρησιμοποιούνται σε αυτή την εφαρμογή, εξυπηρετούν διαφορετικούς σκοπούς για την παραγωγή των τελικών μεταδεδομένων. Τελικός στόχος αυτής της διαδικασίας παραγωγής μεταδεδομένων, είναι η κατάταξη των εικόνων σε μια λίστα από κατηγορίες. Για τη σωστή κατάταξη των εικόνων αυτών, είναι πολύ σημαντική η κατανόηση του περιεχομένου (context) που προβάλλεται στην εικόνα.

Έτσι η διαδικασία παραγωγής των μεταδεδομένων χωρίστηκε σε τρία στάδια στα οποία έχουν αντιστοιχηθεί κατάλληλα τα μοντέλα βαθιάς μάθησης και συμβάλλουν με διαφορετικό τρόπο στην κατανόηση του περιεχομένου αυτού.

Στάδιο 1: Αναγνώριση σκηνικού

Στην αναγνώριση περιεχομένου μιας εικόνας πραγματικού κόσμου, η αναγνώριση σκηνικού παίζει έναν ιδιαίτερα σημαντικό ρόλο. Το υλικό το οποίο προβάλλεται σε μια εικόνα x περιεχομένου, μπορεί αντίστοιχα να προβάλλεται σε μια εικόνα y περιεχομένου όπου τα x, y αποτελούν δύο τελείως διαφορετικές εννοιολογικές σημασίες.

Παράδειγμα:



Εικόνα 1



Εικόνα 2

- Στην εικόνα 1, απεικονίζεται ένας σκύλος που δαγκώνει ένα ξύλο ραβδοειδούς σχήματος ενώ παίζει και τρέχει σε έναν εξωτερικό χώρο.

- Στην εικόνα 2, απεικονίζεται ένας σκύλος που δαγκώνει σκυλοτροφή η οποία έχει ραβδοειδές σχήμα και του τη δίνει ένας άνθρωπος.

Στις δύο παραπάνω εικόνες, το αντικείμενο που απεικονίζεται είναι το ίδιο και μάλιστα σε κοινή θέση, στάση και μορφή. Αυτό που διαφέρει, είναι ο τρόπος με τον οποίο μεταφράζεται η εικόνα στον ανθρώπινο νου λαμβάνοντας υπόψη τρίτους παράγοντες που αντιλαμβάνεται από την εικόνα. Αυτοί οι παράγοντες είναι:

- Ο εξωτερικός χώρος στον οποίο τρέχει ο σκύλος της εικόνας 1.
- Η μορφή της ράβδου στις εικόνες 1 και 2.
- Ο άνθρωπος που ταΐζει τον σκύλο στην εικόνα 2.

Έτσι, ένας άνθρωπος είναι σε θέση να καταλάβει ότι η εικόνα 1 έχει να κάνει με παιχνίδι ενώ η δεύτερη έχει να κάνει με φαγητό.

Η αναγνώριση σκηνών είναι σημαντική επειδή περιλαμβάνει την αναγνώριση ακολουθιών οπτικών αντικειμένων, όπως το κείμενο σε φυσική σκηνή, τα οποία συνήθως δεν εμφανίζονται απομονωμένα αλλά ως μέρος μιας ακολουθίας. Σε αντίθεση με τη γενική αναγνώριση αντικειμένων, η οποία συνήθως περιλαμβάνει την πρόβλεψη μιας μεμονωμένης ετικέτας για μια εικόνα, η αναγνώριση σκηνών απαιτεί από το σύστημα να προβλέψει μια σειρά από ετικέτες αντικειμένων, αντανακλώντας την ακολουθιακή φύση των αντικειμένων μέσα στη σκηνή. Αυτή η πολυπλοκότητα εντείνεται από το μεταβαλλόμενο μήκος αυτών των ακολουθιών - για παράδειγμα, οι αγγλικές λέξεις μπορούν να διαφέρουν σημαντικά σε μήκος. [7]

Η παραγωγή δεδομένων σκηνικού (scene freezing) είναι μια τεχνική που χρησιμοποιείται σε εφαρμογές όπως η κατηγοριοποίηση εικόνας. Αυτή η τεχνική περιλαμβάνει τη σταθεροποίηση ή "πάγωμα" του φόντου ή του περιβάλλοντος χώρου μιας εικόνας, ενώ επικεντρώνεται στον ανίχνευση και ανάλυση των κύριων αντικειμένων ή θεμάτων.

Η σημασία της παραγωγής δεδομένων σκηνικού στην κατηγοριοποίηση εικόνας είναι σημαντική για πολλούς λόγους:

1. Βελτίωση Ακρίβειας: Με την απομόνωση του κύριου αντικειμένου ή θέματος από το περιβάλλον, τα συστήματα μπορούν να εστιάσουν καλύτερα στα χαρακτηριστικά που είναι πιο σημαντικά για την κατηγοριοποίηση.
2. Μείωση Θορύβου: Στα δεδομένα εικόνας, το φόντο μπορεί να περιέχει θόρυβο ή ασήμαντα στοιχεία που αποσπούν την προσοχή από το κύριο αντικείμενο. Η παραγωγή αφαιρεί ή μειώνει αυτό τον θόρυβο.
3. Προσαρμογή σε Διαφορετικά Περιβάλλοντα: Σε εφαρμογές όπου τα αντικείμενα μπορεί να εμφανίζονται σε διάφορα περιβάλλοντα, η παραγωγή δεδομένων σκηνικού βοηθά στην ανίχνευση του αντικειμένου ανεξάρτητα από το περιβάλλον.
4. Βελτιστοποίηση Υπολογιστικών Πόρων: Με τον περιορισμό της επεξεργασίας στα κύρια στοιχεία μιας εικόνας, η απαίτηση για υπολογιστικούς πόρους μειώνεται, κάνοντας την επεξεργασία πιο αποδοτική.
5. Ευελιξία σε Διάφορες Εφαρμογές: Η παραγωγή δεδομένων σκηνικού είναι χρήσιμη σε διάφορες εφαρμογές, όπως στην αναγνώριση προσώπων, ανίχνευση αντικειμένων, και στην αυτόματη οδήγηση, όπου το κύριο αντικείμενο πρέπει να είναι ευδιάκριτο από το περιβάλλον.

Συνολικά, η παραγωγή δεδομένων σκηνικού είναι μια σημαντική τεχνική που βοηθά στη βελτίωση της απόδοσης και της ακρίβειας στην κατηγοριοποίηση εικόνας, καθώς και σε άλλες σχετικές εφαρμογές. [5]

Στην εφαρμογή αυτής της εργασίας, η αναγνώριση σκηνικού γίνεται με τη χρήση τεσσάρων από τα μοντέλα που προαναφέρθηκαν. Αυτά είναι τα:

alexnet, resnet18, resnet50, wideresnet

Τα μοντέλα αυτά είναι όλα εκπαιδευμένα πάνω στην αναγνώριση σκηνικού αλλά με διαφορές όπως ο αριθμός νευρωνικών επιπέδων που διαθέτουν (neuron layers), τα βάρη που λαμβάνουν υπόψη

κ.α. Κάθε φορά που ένα μοντέλο επεξεργάζεται την εικόνα, επιστρέφει τα δεδομένα σκηνικού που εντόπισε και έναν ποσοστό βεβαιότητας για την απάντηση του. **[8, 9, 10, 11]**

Παράδειγμα 1:



Στην παραπάνω εικόνα τα αποτελέσματα του **resnet18** είναι τα εξής:

```
0.817 -> highway
0.076 -> residential_neighborhood
0.038 -> field/wild
```

Αυτό σημαίνει πως το μοντέλο προέβλεψε ότι το σκηνικό της εικόνας είναι highway (λεωφόρο) με βεβαιότητα της τάξης του 81.7%, residential_neighborhood (κατοικημένη γειτονιά) με βεβαιότητα της τάξης του 7.6% και field/wild (χωράφι/εξοχή) με βεβαιότητα της τάξης του 3.8%. Τα αποτελέσματα αυτά είναι αρκετά κοντά στο σωστό αποτέλεσμα παρόλο που τα 2/3 από αυτά έχουν χαμηλό δίκτη βεβαιότητας. Ωστόσο, δεν ισχύει το ίδιο για όλες τις εικόνες.

Παράδειγμα 2:



Στην παραπάνω εικόνα τα αποτελέσματα του **resnet18** είναι τα εξής:

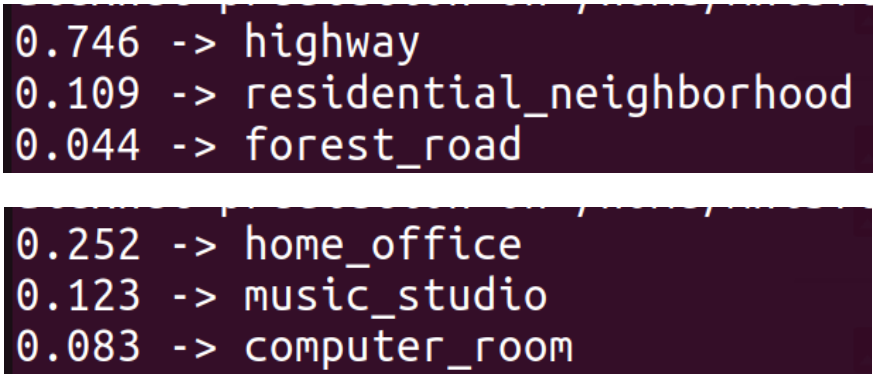
```
0.126 -> science_museum
0.109 -> home_office
0.088 -> staircase
```

Στην περίπτωση αυτή, το μοντέλο προέβλεψε πως το σκηνικό της εικόνας είναι science_museum (μουσείο επιστημών) με βεβαιότητα της τάξης του 12.6%, home_office (γραφείο σπιτιού) με βεβαιότητα της τάξης του 10.9% και staircase (σκάλα) με βεβαιότητα της τάξης του 8.8%. Είναι προφανές πως τα αποτελέσματα του μοντέλου είναι αρκετά μακριά από το πραγματικό περιεχόμενο της εικόνας. Επίσης παρατηρείται σημαντική μείωση στον βαθμό βεβαιότητας.

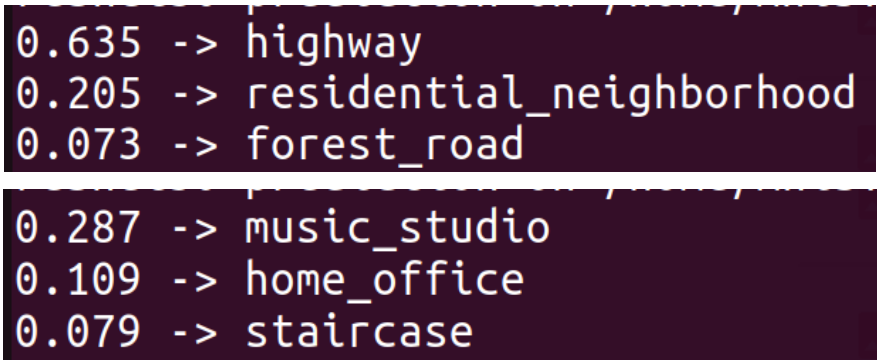
Μέσα από τα παραπάνω παραδείγματα, επιβεβαιώνεται η προαναφερθείσα θεωρία πως ένα μοντέλο αυτού του βελγενεκούς, δεν μπορεί να υποστηρίξει μια τέτοια εφαρμογή από μόνο του. Έτσι, στην παρούσα εφαρμογή. χρησιμοποιούνται τα τέσσερα μοντέλα που αναφέρθηκαν προηγουμένως για να παραχθεί μια συλλογική απόφαση.

Τα αποτελέσματα των μοντέλων alexnet και resnet50 για τις εικόνες του παραδείγματος:

Alexnet:



Resnet50:



Το wideresnet δεν αναλύεται σε αυτό το σημείο καθώς κατέχει μια επιπλέον ιδιότητα που περιγράφεται αναλυτικά παρακάτω.

Αλγόριθμος παραγωγής μεταδεδομένων σκηνικού:

Για τη συλλογική παραγωγή των μεταδεδομένων σκηνικού, η ίδια εικόνα τροφοδοτείται στα τέσσερα μοντέλα σκηνικού διαδοχικά. Κάθε φορά που ένα μοντέλο παράγει τα δεδομένα σκηνικού που προέβλεψε, τα δεδομένα αυτά φιλτράρονται έτσι ώστε να περνούν στο επόμενο στάδιο μόνο οι προβλέψεις οι οποίες έχουν ποσοστό βεβαιότητας μεγαλύτερο της τάξεως του 20%. Οι προβλέψεις που βρίσκονται κάτω από αυτό το όριο απορρίπτονται. Αυτή η τιμή, προέκυψε από πολλαπλές

δοκιμές πάνω σε έναν μεγάλο αριθμό εικόνων. Μέσα από αυτές τις δοκιμές, παρατηρήθηκε ότι οι προβλέψεις που κατά μέσο όρο βρίσκονται πάνω από το κατώφλι του 20%, τείνουν να επιστρέφουν αποτελέσματα τα οποία είναι πιο κοντά στο πραγματικό περιεχόμενο της εικόνας σε σχέση με τα αποτελέσματα που βρίσκονται κάτω από το κατώφλι.

Έπειτα, γίνεται καταμέτρηση των αποτελεσμάτων που επέστρεψε το κάθε μοντέλο. Στην περίπτωση που το ίδιο αποτέλεσμα έχει εμφανιστεί πάνω από δύο φορές, πράγμα που σημαίνει ότι πάνω από δύο μοντέλα επέστρεψαν το ίδιο αποτέλεσμα, τότε η πρόβλεψη αυτή γίνεται αποδεκτή, διαφορετικά απορρίπτεται. Με αυτή τη μέθοδο φιλτραρίσματος μικραίνουν οι πιθανότητες να υπάρξουν έκτοπες τιμές (outliers).

Αφού λοιπόν τελειώσει η διαδικασία παραγωγής των δεδομένων πρόβλεψης σκηνικού, χρησιμοποιείται μια επιπλέον ιδιότητα του μοντέλου wideresnet. Η ιδιότητα αυτή, είναι η παραγωγή επιπλέον μεταδεδομένων σκηνικού τα οποία “εξηγούν” πως προέκυψαν οι προβλέψεις σκηνικού που επέστρεψε καθώς και ένας χάρτης ένδειξης των σημείων που το νευρωνικό δίκτυο χρησιμοποίησε για να καταλήξει στο εκάστοτε αποτέλεσμα (CAM - class activation mapping).

Παράδειγμα επιπλέον δεδομένων του wideresnet:



εικόνα 1



εικόνα 2

Τα αποτελέσματα για την εικόνα 1:

```
--TYPE OF ENVIRONMENT: outdoor
--SCENE CATEGORIES:
0.313 -> residential_neighborhood
0.267 -> highway
0.054 -> forest_road
0.053 -> street
0.043 -> field_road
--SCENE ATTRIBUTES:
natural light, open area, driving, man-made, asphalt, biking, trees, foliage, pavement
```

Τα αποτελέσματα για την εικόνα 2:

```
--TYPE OF ENVIRONMENT: indoor
--SCENE CATEGORIES:
0.080 -> elevator_shaft
0.057 -> beauty_salon
0.038 -> television_studio
0.034 -> natural_history_museum
0.032 -> staircase
--SCENE ATTRIBUTES:
no horizon, enclosed area, man-made, wood, indoor lighting, cloth, working, glass, metal
```

Χάρτης ενδείξεων (CAM):



εικόνα 1



εικόνα 2

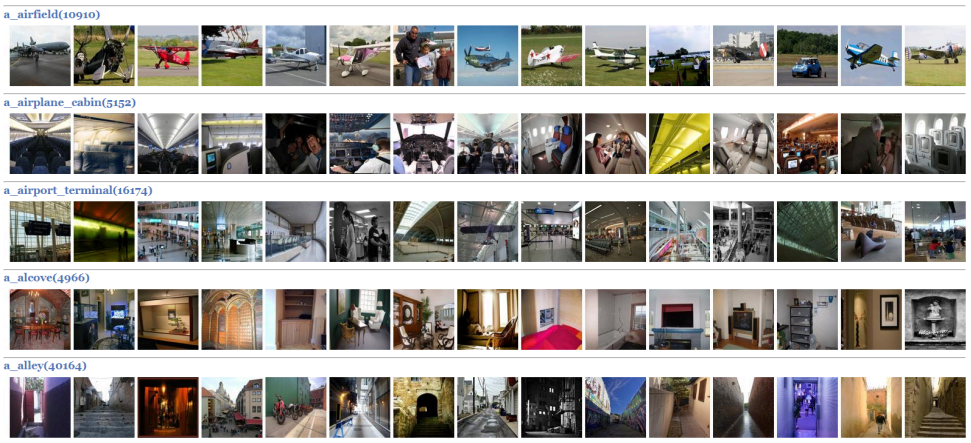
Ορισμός του class activation map:

Το Class Activation Mapping (CAM) είναι μια τεχνική οπτικοποίησης που χρησιμοποιείται σε συνελικτικά νευρωνικά δίκτυα (CNNs) για να εντοπίσει τις περιοχές μιας εικόνας που έχουν ενεργοποιήσει σημαντικά το νευρωνικό δίκτυο για μια συγκεκριμένη κλάση εξόδου. Σε άλλα λόγια, το CAM δείχνει πού κοίταξε το μοντέλο για να πάρει την απόφασή του σχετικά με την ταξινόμηση μιας εικόνας. Το μοντέλο WideResNet είναι μια παραλλαγή του Residual Network (ResNet), το οποίο είναι ένα βαθύ νευρωνικό δίκτυο που χρησιμοποιείται για εργασίες όπως η ταξινόμηση εικόνων, και το CAM που παράγεται μπορεί να αποκαλύψει ποιες περιοχές της εικόνας επηρέασαν περισσότερο την τελική του κρίση.

Εκπαίδευση των μοντέλων που χρησιμοποιήθηκαν:

Τα μοντέλα που χρησιμοποιήθηκαν είναι όλα εκπαιδευμένα πάνω στο ίδιο σύνολο δεδομένων (dataset) το οποίο ονομάζεται **Places365**. Το **Places365** είναι ένα μεγάλο σύνολο δεδομένων στον τομέα της όρασης των υπολογιστών (computer vision), το οποίο αναπτύχθηκε για την εκπαίδευση και τη δοκιμή αλγορίθμων αναγνώρισης σκηνικών. Αυτό το σύνολο δεδομένων περιλαμβάνει πάνω από δέκα εκατομμύρια εικόνες, κατανεμημένες σε 365 κατηγορίες σκηνών, όπως παραλίες, βουνά, δάση, αστικά τοπία κ.α. Οι εικόνες αυτές συλλέχθηκαν από διάφορες πηγές στο διαδίκτυο και είναι ετικετοποιημένες (labeled) με την αντίστοιχη κατηγορία σκηνής, παρέχοντας ένα μεγάλο σύνολο δεδομένων για την εκπαίδευση νευρωνικών δικτύων. Το Places365 χρησιμοποιείται ευρέως στην έρευνα και σε εφαρμογές που απαιτούν αναγνώριση ή κατανόηση του σκηνικού, όπως σε αυτόνομα οχήματα, αυτοματοποιημένα συστήματα πλοήγησης, και στην ανάλυση εικόνων και βίντεο. [17]

Το dataset:



Στάδιο 2: Αναγνώριση αντικειμένων

Η αναγνώριση των αντικειμένων που απεικονίζονται σε μια εικόνα, είναι ίσως το σημαντικότερο βήμα στην αναγνώριση του περιεχομένου και στην κατηγοριοποίηση της ίδιας της εικόνας. Αυτή η διαδικασία επιτρέπει στα συστήματα υπολογιστικής όρασης (computer vision) να "κατανοήσουν" και να ερμηνεύσουν τι ακριβώς βλέπουν στην εικόνα, αναγνωρίζοντας και ταξινομώντας διάφορα αντικείμενα όπως ανθρώπους, ζώα, αυτοκίνητα κ.α. Αυτό δεν είναι μόνο βασικό για την ακριβή κατηγοριοποίηση της εικόνας αλλά και για την κατανόηση του πλαισίου και της σημασίας της σκηνής. Επιπλέον, η αναγνώριση αντικειμένων παίζει έναν κρίσιμο ρόλο σε εφαρμογές όπως η αυτόνομη πλοήγηση, η ασφάλεια, η ανάλυση ιατρικών εικόνων και στη βελτίωση της αλληλεπίδρασης μεταξύ ανθρώπου και υπολογιστή.

Ένα από τα μεγαλύτερα σύνολα δεδομένων (datasets) που χρησιμοποιούνται για την εκπαίδευση μοντέλων αναγνώρισης αντικειμένων, είναι το COCO της Microsoft. Το COCO Dataset (Common Objects in Context) είναι ένα από τα πιο σημαντικά και εκτεταμένα σύνολα δεδομένων που χρησιμοποιούνται σε αυτόν τον τομέα, ενώ οι κυριότερες χρήσεις του είναι οι ταξινόμηση εικόνων, ανίχνευση αντικειμένων και η σημειοποίηση και κατανόηση σκηνών. Το σύνολο αυτό περιέχει πάνω από 330.000 εικόνες με πάνω από 2,5 εκατομμύρια ετικετοποιημένα αντικείμενα σε 80 κατηγορίες. Το COCO, διακρίνεται για την ποικιλία των σκηνών και των αντικειμένων που περιλαμβάνει, καθώς και για τη λεπτομερή σημειοποίηση του κάθε αντικειμένου μέσα στην εικόνα, πράγμα που το καθιστά ιδανικό για εκπαίδευση και επικύρωση αλγορίθμων βαθιάς μάθησης. [18]

Μερικές εικόνες από το COCO:



Αντικείμενα που βρέθηκαν στην εικόνα:

- άνθρωπος, αυτοκίνητο, αεροπλάνο, φορτηγό



Αντικείμενα που βρέθηκαν στην εικόνα:

- παγκάκι, μήλο, καρέκλα, καναπές τραπέζι, τηλεόραση

Χρήση αναγνώρισης αντικειμένων στην εφαρμογή

Στην εφαρμογή της εργασίας, τα μοντέλα αναγνώρισης αντικειμένων που χρησιμοποιούνται, όπως προαναφέρθηκαν, είναι τα Faster R-CNN και Mask R-CNN. Και τα δύο αυτά μοντέλα είναι εκπαιδευμένα πάνω στο COCO dataset και τα αποτελέσματα που επιστρέφουν έχουν σημαντικό βαθμό επιτυχίας. Παρακάτω είναι μερικά παραδείγματα της χρήσης των μοντέλων και των αποτελεσμάτων που επιστρέφουν. **[12, 13, 14]**

Παράδειγμα 1



Αποτελέσματα Faster R-CNN μοντέλου:

- [car, vehicle], [stop sign, outdoor], [truck, vehicle]

Αποτελέσματα Mask R-CNN μοντέλου:

- [car, vehicle]

Στο παράδειγμα φαίνεται πως το μοντέλο Faster R-CNN εντόπισε περισσότερα αντικείμενα αλλά όπως είναι προφανές στην εικόνα, τα 2/3 αυτών είναι λανθασμένα.

Παράδειγμα 2



Αποτελέσματα Faster R-CNN μοντέλου:

- [chair, furniture], [cell phone, electronic], [book, indoor], [handbag, accessory], [bottle, kitchen], [person, person], [tv, electronic]

Αποτελέσματα Mask R-CNN μοντέλου:

- [chair, furniture], [person, person], [bottle, kitchen], [book, indoor], [cell phone, electronic], [tv, electronic], [tie, accessory]

Στο παράδειγμα φαίνεται πως τα δύο μοντέλα εντόπισαν τον ίδιο αριθμό αντικειμένων με εξαιρετική ακρίβεια αλλά και τα δύο έκαναν από ένα λάθος.

Μέσα από αυτά τα παραδείγματα, συμπεραίνουμε πως τα μοντέλα αυτά από μόνα τους, δεν είναι ικανά να υποστηρίξουν την επιθυμητή λειτουργικότητα της εφαρμογής αυτής της εργασίας. Σε μια προσπάθεια να αποφύγουμε τυχόντα λάθη, στην υλοποίηση της εφαρμογής κρατάμε μόνο τα κοινά αντικείμενα που εντόπισαν τα δύο μοντέλα.

Τα κοινά αντικείμενα του **παραδείγματος 1**:

- [car, vehicle]

Τα κοινά αντικείμενα του **παραδείγματος 2**:

- [chair, furniture], [cell phone, electronic], [book, indoor], [bottle, kitchen], [person, person], [tv, electronic]

Αυτή η μέθοδος αξιολογήθηκε μέσα από μια σειρά δοκιμών και αποδείχθηκε πως με τη χρήση της, τα λάθη των μοντέλων μειώνονται σημαντικά, πράγμα το οποίο είναι υψίστης σημασίας για το επόμενο στάδιο που είναι η κατηγοριοποίηση των εικόνων.

Στάδιο 3: Κατηγοριοποίηση των εικόνων

Σε αυτό το στάδιο έχουν παραχθεί πλέον τα περισσότερα από τα μεταδεδομένα που είναι απαραίτητα για τη λειτουργία του συστήματος της εφαρμογής. Αυτό που λείπει είναι η κατάταξη των εικόνων σε μια κατηγορία έτσι ώστε να μπορεί η εφαρμογή να κάνει τις κατάλληλες προτάσεις στους χρήστες ανάλογα με τις προτιμήσεις τους.

Για να επιτευχθεί η κατηγοριοποίηση των εικόνων, δημιουργήθηκε ένα νέο μοντέλο κατηγοριοποίησης (classification model) με τη χρήση των τεχνολογιών tensorflow, keras και scikit-learn το οποίο εκπαιδεύτηκε πάνω σε ένα σύνολο δεδομένων που επίσης δημιουργήθηκε για τον σκοπό αυτής της εργασίας.

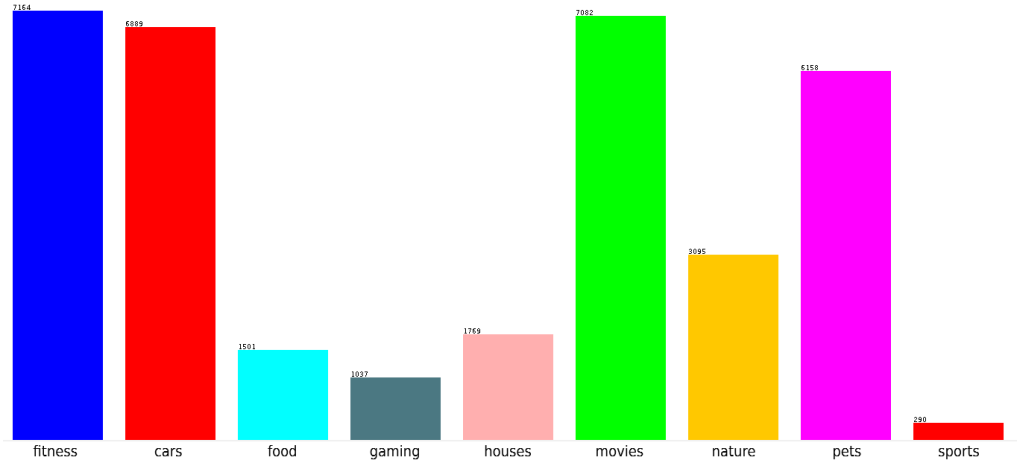
Διαδικασία δημιουργίας συνόλου δεδομένων

Για να δημιουργηθεί το σύνολο των δεδομένων πάνω στο οποίο εκπαιδεύτηκε το μοντέλο αυτής της εργασίας, αρχικά χρειάστηκε να επιλεχθούν οι κατηγορίες πάνω στις οποίες θα γινόταν η εκπαίδευση. Αυτό ήταν απαραίτητο καθώς για την εκπαίδευση του μοντέλου επιλέχθηκε η μέθοδος εποπτευόμενης εκμάθησης (Supervised learning). Οι κατηγορίες που επιλέχθηκαν ήταν οι εξής:

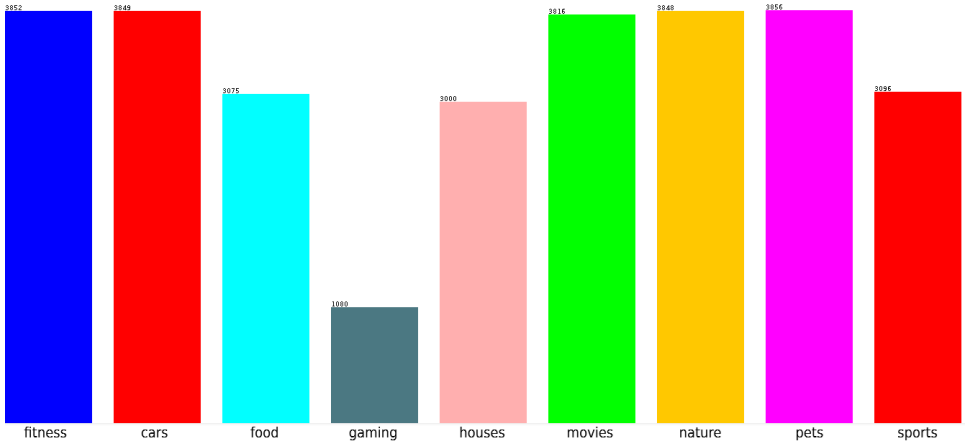
- **Cars (αυτοκίνητα**
- **Fitness (γυμναστική και ευεξία)**
- **Food (φαγητό)**
- **Gaming (παιχνίδια)**
- **Houses (σπίτια)**
- **Movies (Ταινίες)**
- **Nature (Φύση)**
- **Pets (Κατοικίδια)**
- **Sports (Σπορ και δραστηριότητες)**

Αυτές οι κατηγορίες είναι ένα υποσύνολο των πιο γενικών κατηγοριών εικόνων που υπάρχουν στις πιο γνωστές πλατφόρμες κοινωνικών δικτύων. [20]

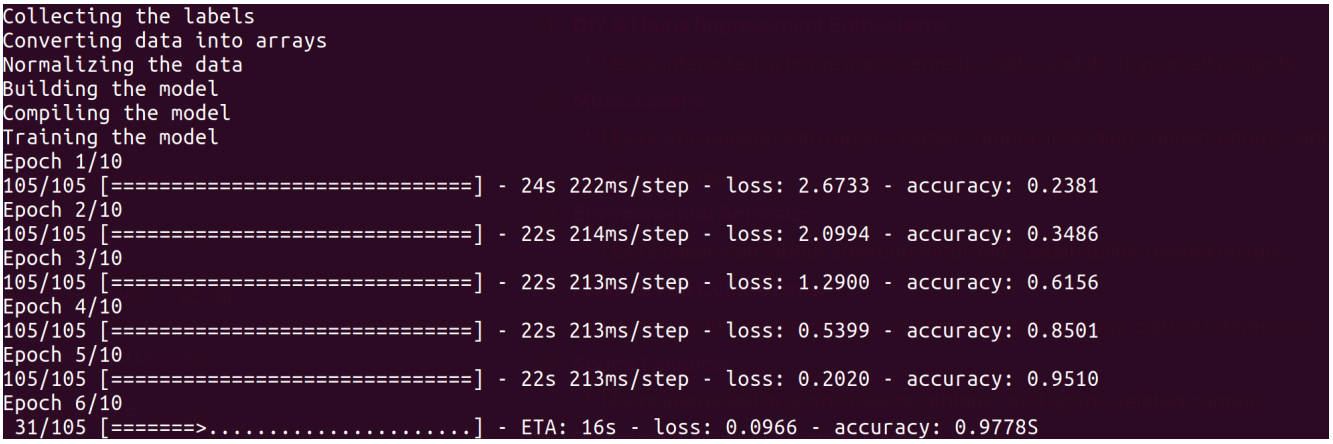
Στην επιβλεπόμενη εκμάθηση, η εκπαίδευση του μοντέλου για την αναγνώριση κατηγοριών μέσα από μεταδεδομένα προαπαιτεί έναν μεγάλο όγκο τέτοιου είδους μεταδεδομένων για να μπορεί το μοντέλο να αναγνωρίσει μοτίβα και να δημιουργήσει το νευρωνικό δίκτυο που είναι απαραίτητο. Έτσι, για κάθε μια από τις κατηγορίες που επιλέχθηκαν, χρησιμοποιήθηκε ένα dataset εικόνων για να παραχθούν τα απαραίτητα μεταδεδομένα της εκπαίδευσης (training data). Για την παραγωγή αυτών των δεδομένων, εφαρμόστηκε η μεθοδολογία που περιγράφηκε αναλυτικά προηγουμένως. Το αποτέλεσμα ήταν το εξής:



Στην παραπάνω εικόνα φαίνεται η μεγάλη ανισορροπία των κλάσεων (class imbalance) που υπήρξε μετά την παραγωγή των δεδομένων. Αυτό συνέβη επειδή τα σύνολα εικόνων που επιλέχθηκαν δεν είχαν τον ίδιο συνολικό αριθμό εικόνων. Για να αντιμετωπιστεί αυτό το πρόβλημα, χρειάστηκε να μειώσουμε τις εικόνες από κάποιες υπερπληθείς κλάσεις έτσι ώστε ο αριθμός των εικόνων τους να είναι πιο κοντά σε εκείνων με τις λιγότερες εικόνες. Επίσης, ήταν απαραίτητο να διευρυνθούν οι κλάσεις που είχαν πολύ μικρό αριθμό εικόνων όπως οι κατηγορίες food, sports και houses. Το αποτέλεσμα αυτής της προεπεξεργασίας των δεδομένων (data preprocessing) είναι το εξής:



Όπως φαίνεται στο παραπάνω ιστόγραμμα, οι κλάσεις πλέον έχουν έρθει σε μια μορφή ισορροπίας η οποία ακόμη χρήζει αλλαγών. Πιο συγκεκριμένα, η κατηγορία gaming, παρουσιάζει σημαντικά μικρότερο αριθμό δεδομένων από τις άλλες. Παρόλα αυτά, για τον σκοπό αυτής της εργασίας δεν είναι πλήρως απαραίτητη η απόλυτη ισορροπία των κατηγοριών/κλάσεων και έτσι λοιπόν αυτό ήταν το τελικό σύνολο δεδομένων πάνω στο οποίο εκπαιδεύτηκε το μοντέλο της εφαρμογής.



Για την εκπαίδευση του μοντέλου χρησιμοποιήθηκε ο αλγόριθμος ενεργοποίησης νευρώνων '**Adam**' και η συνάρτηση '**sparse_categorical_crossentropy**'

Θεωρητικό υπόβαθρο των συναρτήσεων εκπαίδευσης:

Adam:

Ο αλγόριθμος ενεργοποίησης νευρώνων Adam (Adaptive Moment Estimation) είναι μία μέθοδος βελτιστοποίησης που χρησιμοποιείται στην εκπαίδευση νευρωνικών δικτύων. Αυτή η τεχνική είναι γνωστή για την αποδοτικότητά της στην επίτευξη ταχείας και ακριβούς σύγκλισης κατά τη διαδικασία εκμάθησης. Ο Adam συνδυάζει δύο κρίσιμες τεχνικές βελτιστοποίησης: την Εκθετική Κινητή Μέση των διαδοχικών κλιμακωτών βαθμών της συνάρτησης κόστους (παρόμοια με τον αλγόριθμο momentum) και την Εκθετική Κινητή Μέση των τετραγώνων αυτών των βαθμών (παρόμοια με τον αλγόριθμο RMSprop). Αυτός ο συνδυασμός καθιστά τον Adam ιδιαίτερα αποτελεσματικό σε προβλήματα με μεγάλο όγκο δεδομένων και παραμέτρων, δίνοντας μια καλή ισορροπία μεταξύ ταχύτητας εκπαίδευσης και ακριβείας των προβλέψεων.

Sparse Categorical Crossentropy:

Η Sparse Categorical Crossentropy είναι μια συνάρτηση απώλειας (loss function) που χρησιμοποιείται συχνά στην ταξινόμηση πολλαπλών κλάσεων, όπου οι κλάσεις είναι αμοιβαία αποκλειστικές. Σε αντίθεση με την Categorical Crossentropy, η οποία απαιτεί τη χρήση τη χρήση της μεθόδου one-hot encoding για τις ετικέτες των κλάσεων (δηλαδή, κάθε ετικέτα κλάσης εκπροσωπείται ως ένα διάνυσμα με μηδενικά σε όλες τις θέσεις εκτός από έναν, όπου είναι ένα), η Sparse Categorical Crossentropy επιτρέπει τη χρήση απλών ακεραίων για την ετικέτα κάθε κλάσης. Αυτό την καθιστά πιο πρακτική και αποδοτική σε καταστάσεις όπου ο αριθμός των κλάσεων είναι μεγάλος, καθώς μειώνει τον απαιτούμενο όγκο των δεδομένων εκπαίδευσης. Η συνάρτηση αυτή είναι σημαντική για την εκπαίδευση νευρωνικών δικτύων σε σύγχρονες εφαρμογές ταξινόμησης, καθώς βοηθά στην αξιολόγηση της απόδοσης του μοντέλου σε προβλήματα με πολλαπλές κλάσεις.

Χρήση και αξιολόγηση του μοντέλου:

Το μοντέλο δοκιμάστηκε στην εφαρμογή από πραγματικούς χρήστες από τους οποίους συλλέχθηκαν πολύτιμα δεδομένα προς μελέτη. Τα δεδομένα αυτά, μαζί με δεδομένα ανατροφοδότησης που συλλέχθηκαν από τους ίδιους τους χρήστες, χρησιμοποιήθηκαν για την αξιολόγηση του μοντέλου και της εφαρμογής.

Κατά τη χρήση της εφαρμογής παρατηρήθηκε πως το μοντέλο αδυνατούσε να κατηγοριοποιήσει όλες τις νέες εικόνες που λάμβανε σε μια από τις διαθέσιμες κατηγορίες. Το αίτιο αυτής της συμπεριφοράς, είναι ο όγκος των δεδομένων εκπαίδευσης ο οποίος ήταν πολύ μικρός για ένα μοντέλο υπολογιστικής όρασης. Αυτό είναι ένα από τα μεγαλύτερα προβλήματα που αντιμετωπίζουν καθημερινά οι επιστήμονες της πληροφορικής, καθώς για να γίνει επιτυχημένη εκπαίδευση ενός μοντέλου (δηλαδή να έχει μεγάλο ποσοστό επιτυχίας το αποτέλεσμα), πρέπει να διατίθεται πολύ μεγάλος ο όγκος δεδομένων πράγμα αυξάνει και τη ζήτηση για υπολογιστικούς πόρους.

Στην περίπτωση της εφαρμογής αυτής της εργασίας, το μοντέλο έτρεξε με επιτάχυνση υλικού (hardware acceleration) και πιο συγκεκριμένα επιτάχυνση μονάδας γραφικών (GPU acceleration). Σε περίπτωση που δε γινόταν χρήση τέτοιων τεχνολογιών, η εφαρμογή της παραγωγής των μεταδεδομένων δε θα ήταν δυνατό να τρέξει σε ταχύτητες απαραίτητες για τη χρήση της σε πλατφόρμες κοινωνικών δικτύων. Επίσης, για να επιλυθεί το πρόβλημα των μη επαρκών κατηγοριών, προστέθηκε μια ακόμη κατηγορία με όνομα '**unknown**' έτσι ώστε να μπορούν να αγνοηθούν οι εικόνες που δεν ανήκουν σε κάποια από τις κατηγορίες που δεν είναι αντικείμενο μελέτης αυτής της εργασίας.

Η παραγωγή των μεταδεδομένων και χρήση τεχνολογιών τεχνητής νοημοσύνης είναι μια ακριβή και χρονοβόρα διαδικασία. Εκτός από τη χρήση τεχνολογιών όπως το GPU acceleration, ήταν απαραίτητη η χρήση ενός συστήματος ουρών το οποίο λαμβάνει τα αιτήματα για παραγωγή μεταδεδομένων από τις εικόνες των χρηστών και τα τροφοδοτεί με σειριακή αλληλουχία στα μοντέλα τεχνητής νοημοσύνης. Για τη δημιουργία αυτού του συστήματος, επιλέχθηκε η βιβλιοθήκη **Celery**. Το Celery είναι ένα ασύγχρονο σύστημα εργασιών κατανεμημένης ουράς για την Python. Χρησιμοποιείται κυρίως για την εκτέλεση υποβάθρου και περιοδικών εργασιών, όπως την αποστολή email ή την ενημέρωση βάσεων δεδομένων. Αυτό επιτρέπει στις εφαρμογές να ανταποκρίνονται γρηγορότερα στα αιτήματα των χρηστών καθώς οι βαριές ή χρονοβόρες διαδικασίες μεταφέρονται στο παρασκήνιο ή μπαίνουν σε μια ουρά εκτέλεσης.

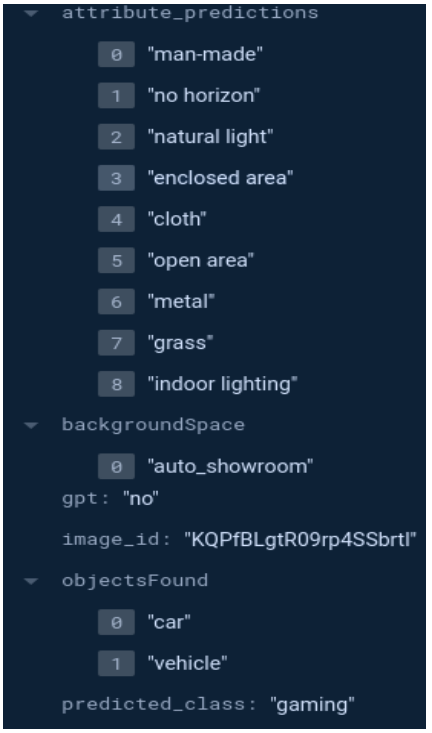
Οι χρόνοι που παρατηρήθηκαν στην παραγωγή μεταδεδομένων μιας εικόνας με GPU acceleration στην κάρτα γραφικών **Gigabyte RTX 3070 ti** (κάρτα γραφικών οικιακής χρήσης) με δυνατότητα υπολογισμού **8.6** (βαθμολογία δυνατοτήτων υπολογισμού από το **CUDA** - Πλατφόρμα παράλληλου υπολογισμού), ήταν κατά μέσο όρο 10 δευτερόλεπτα ανα εικόνα ενώ η ουρά μπορεί να έφτανε και τα 100 αιτήματα (με δεδομένα μιας μικρής ομάδας χρηστών). Αυτό δείχνει ότι για την κλιμάκωση ενός τέτοιου συστήματος είναι απαραίτητη η χρήση υπέρογκων υπολογιστικών συστημάτων των οποίων το κόστος είναι υψηλό.

Επίσης, παρατηρήθηκε από τους χρήστες πως κατά τη χρήση της εφαρμογής, οι νέες προτάσεις εικόνων που ερχόταν δεν ήταν τόσο σχετικές με το προφίλ του χρήστη. Αυτό συνέβαινε επειδή το μοντέλο έτεινε να αντιστοιχίζει έναν μεγάλο αριθμό εικόνων σε λανθασμένες κατηγορίες.

Μερικά παραδείγματα μεταδεδομένων που παρήχθησαν από το μοντέλο:

Παράδειγμα 1:





Σε αυτό το παράδειγμα φαίνεται πως παρόλο που τα μεταδεδομένα της εικόνας φαίνονται σωστά, η κατηγορία που αντιστοιχήθηκε (**predicted_class**) είναι λανθασμένη. Η σωστή κατηγορία για τη συγκεκριμένη εικόνα είναι η '**cars**'. Τέτοιου είδους προβλήματα προκύπτουν από τον μικρό όγκο δεδομένων που έχουν χρησιμοποιηθεί κατά την εκπαίδευση όπως αναφέρθηκε και προηγουμένως.

Παράδειγμα 2:



attribute_predictions
0 "man-made"
1 "no horizon"
2 "natural light"
3 "enclosed area"
4 "cloth"
5 "open area"
6 "metal"
7 "grass"
8 "indoor lighting"
backgroundSpace
gpt: "no"
image_id: "aqE5QRF7OxtlaTa9VyL9"
objectsFound
0 "cat"
1 "dog"
2 "animal"
predicted_class: "unknown"

Και πάλι σε αυτό το παράδειγμα, τα μοντέλα δείχνουν πολύ υψηλό ποσοστό επιτυχίας ενώ αυτή τη φορά το μοντέλο έπεσε στην περίπτωση αδυναμίας κατηγοριοποίησης της εικόνας παρόλο που έχει εκπαιδευτεί πάνω σε μεταδεδομένα εικόνων με παρόμοια χαρακτηριστικά.

Χρήση LLM's για παραγωγή μεταδεδομένων

Μετά από τις παρατηρήσεις που έγιναν και την ανατροφοδότηση των χρηστών της εφαρμογής, διεξάχθηκε ένα νέο πείραμα κατά το οποίο χρησιμοποιείται η κατηγοριοποίηση των εικόνων με τη χρήση μοντέλων LLM (Large Language Models). Πιο συγκεκριμένα, χρησιμοποιήθηκε το μοντέλο **gpt-4-vision-preview** της **OpenAI** το οποίο διαθέτει και οπτικές δυνατότητες για την αναγνώριση περιεχομένου εικόνων. Το θέμα που υπήρξε με το GPT-4 είναι πως κάνει απευθείας κατηγοριοποίηση των εικόνων που λαμβάνει και δε βασίζεται σε μεταδεδομένα όπως το προηγούμενο μοντέλο της εφαρμογής. Για αυτόν τον λόγο, ζητήθηκε από το LLM να παράξει μεταδεδομένα ίδιου τύπου με τα υπόλοιπα μοντέλα έτσι ώστε να μπορούμε να συγκρίνουμε τις διαφορές στην κατηγοριοποίηση εικόνων με παρόμοια μεταδεδομένα και διαφορετικές κλάσεις και να δούμε τι έκανε το μοντέλο να καταλήξει στο συμπέρασμα του.

Το κείμενο (prompt) που δόθηκε στο GPT μαζί με την εικόνα:

Give me the metadata of this image. The format is {'image_id': '%s', 'attribute_predictions': [], 'backgroundSpace': [], 'objectsFound': [], 'predicted_class': '', 'gpt': '%s'}

where 'image_id' is the id of the image, 'attribute_predictions' is a list of of 1-2 word strings with information about how the prediction happened

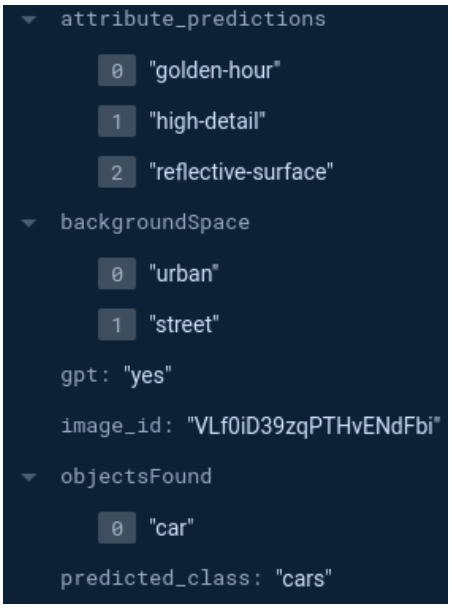
(e.g. natural-light etc), 'backgroundSpace' is a list of strings that tells what is the bg of the image (e.g. coffee shop), 'objectsFound' is a list of the detected objects, and 'predicted_class' is a the final classification.

Classify the image in one of these classes: cars, fitness, food, gaming, houses, movies, nature, pets, sports, unknown.

Return only this json and nothing else.

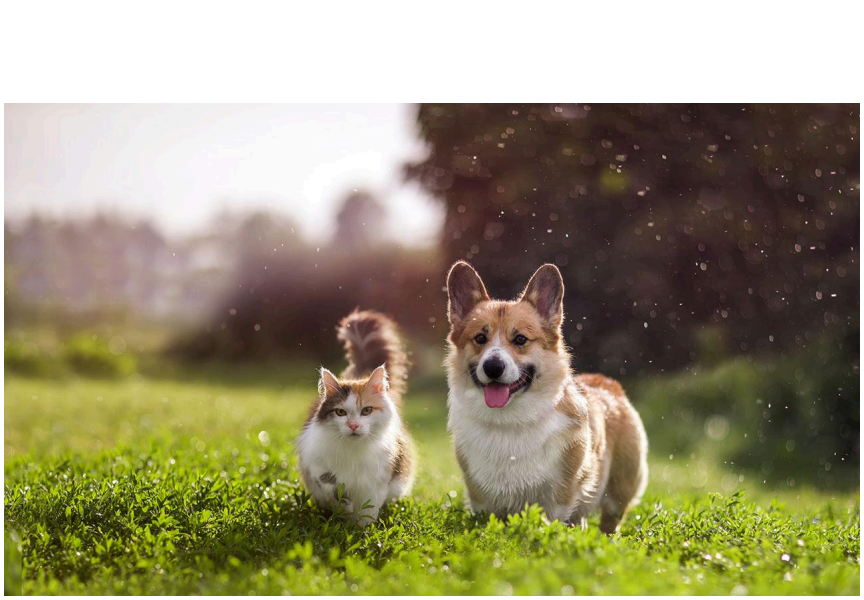
Παράδειγμα 1:





Σε αυτό το παράδειγμα φαίνεται πως τα μεταδεδομένα είναι λιγότερα αλλά πιο στοχευμένα από τα προηγούμενα παραδείγματα ενώ επίσης είναι προφανές πως η εικόνα κατηγοριοποιήθηκε σωστά στην κατηγορία '**cars**'.

Παράδειγμα 2:




```
attribute_predictions
  0 "natural-light"
  1 "focused-subjects"
backgroundSpace
  0 "outdoor"
  1 "grass"
  2 "park"
gpt: "yes"
image_id: "6gyTOfeX3kNIBKlij4HH"
objectsFound
  0 "dog"
  1 "cat"
predicted_class: "pets"
```


Και σε αυτό το παράδειγμα, διαπιστώνεται πως η εικόνα έχει κατηγοριοποιηθεί σωστά και τα δεδομένα που εξήχθησαν από το GPT-4 περιγράφουν σωστά το περιεχόμενο της εικόνας.

Αξιολόγηση μοντέλου LLM:

Τα αποτελέσματα που επέστρεψε το μοντέλο παρουσίασαν εξαιρετικό ποσοστό ακρίβειας στην αναγνώριση και κατηγοριοποίηση των εικόνων. Οι χρήστες παρατήρησαν πως κατά τη χρήση της εφαρμογής, υπήρξε μεγάλη συσχέτιση ανάμεσα στις προτιμήσεις τους και στο περιεχόμενο που λάμβαναν. Το αποτέλεσμα αυτό είναι αναμενόμενο καθώς τα LLMs, και ειδικότερα το GPT-4, είναι τεράστια μοντέλα που έχουν εκπαιδευτεί σε υπέρογκα σύνολα δεδομένων σε αντίθεση με το μοντέλο που εκπαιδεύτηκε για αυτή την εφαρμογή.

Σημείωση: Πρέπει να σημειωθεί πως για τους σκοπούς του πειράματος, το GPT-4 περιορίστηκε στις κατηγορίες που εκπαιδεύτηκε και το μοντέλο της εφαρμογής έτσι ώστε να υπάρχει ένα ορθό μέτρο σύγκρισης.

3.3 Κατηγοριοποίηση των χρηστών

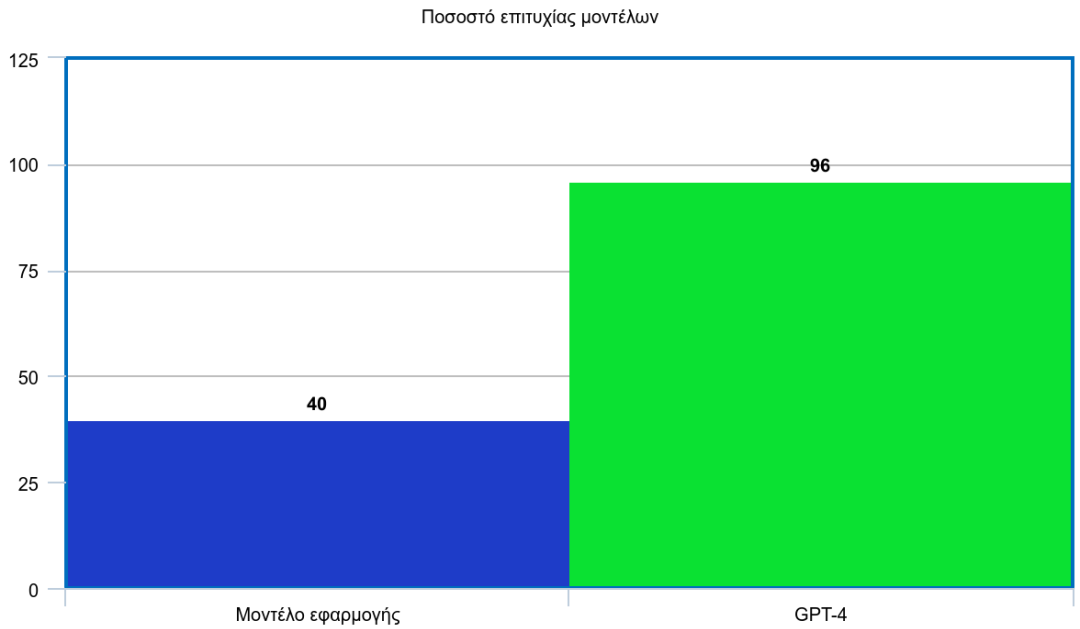
Μετά την επιτυχή συλλογή και καταγραφή των δεδομένων εικόνων στις προαναφερθείσες κατηγορίες, το επόμενο στάδιο αφορά τη δημιουργία του τελικού συστήματος συστάσεων. Αυτό το σύστημα, που αποτελεί το επίκεντρο της παρούσας εργασίας, ενσωματώθηκε εντός της δομής της πλατφόρμας του κοινωνικού δικτύου και η λειτουργικότητά του έχει περιγραφεί αναλυτικά στην ενότητα **3.1.2**. Χρησιμοποιώντας αυτό το σύστημα, επιτεύχθηκε η άμεση και αυτοματοποιημένη κατηγοριοποίηση των χρηστών με βάση τις προτιμήσεις τους, πράγμα που γίνεται της ανάλυσης των δραστηριοτήτων των χρηστών στην πλατφόρμα. Έτσι, το σύστημα είναι σε θέση να προσδιορίσει τα προσωπικά ενδιαφέροντα και τις τάσεις του κάθε χρήστη, επιτρέποντας την παροχή πιο στοχευμένων και προσωποποιημένων προτάσεων.

Κεφάλαιο 4 - Αξιολόγηση του εργαλείου

4.1 Αξιολόγηση των αποτελεσμάτων της εφαρμογής

Η ανάπτυξη της εφαρμογής μας, η οποία εστιάζει στην παροχή προσαρμοσμένων εικονικών προτάσεων στους χρήστες, αποτελεί ένα ενδιαφέρον παράδειγμα της εφαρμογής τεχνολογιών βασισμένων στην τεχνητή νοημοσύνη στον τομέα των ψηφιακών πλατφορμών. Για τον έλεγχο της λειτουργικότητας του συστήματος, ήταν απαραίτητη η αρχική εισαγωγή ενός μεγάλου αριθμού εικόνων. Για αυτόν τον λόγο χρησιμοποιήθηκε ένα σύνολο εικόνων που έχει δημιουργηθεί από την Google με σκόπο την έρευνα και εξέλιξη της τεχνολογίας βαθειάς μάθησης (deep learning) [19]. Η αρχική προσέγγιση, η οποία συνδύαζε την εκπαίδευση ενός μοντέλου κατηγοριοποίησης με βάση μεταδεδομένα από deep learning μοντέλα, αντιμετώπισε προκλήσεις σχετικά με την απόδοση και την ακρίβεια. Πιο συγκεκριμένα, τα δεδομένα εκπαίδευσης δεν ήταν αρκετά πλούσια ή διαφοροποιημένα, γεγονός που οδήγησε σε μειωμένη απόδοση του μοντέλου.

Η εισαγωγή του GPT-4 ως εργαλείου για την κατηγοριοποίηση εικόνων αποτέλεσε μια σημαντική βελτίωση στην αποδοτικότητα της εφαρμογής. Το GPT-4, με τις προηγμένες του δυνατότητες στην ανάλυση και επεξεργασία δεδομένων, προσφέρει μια πιο ολοκληρωμένη και ακριβή προσέγγιση στην κατηγοριοποίηση των εικόνων. Αυτό επιτρέπει στην εφαρμογή να παρουσιάζει πιο σχετικό περιεχόμενο στους χρήστες, βασιζόμενη στις ατομικές τους προτιμήσεις, κάτι που αποτελεί κεντρικό στόχο για πλατφόρμες προτάσεων περιεχομένου.



Όπως φαίνεται στην εικόνα, τα μοντέλα έδειξαν ποσοστό επιτυχίας 40% και 96% αντίστοιχα στη σωστή κατηγοριοποίηση των εικόνων. Το αποτέλεσμα αυτό, υπογραμμίζει τη σημασία της επιλογής των κατάλληλων εργαλείων και τεχνολογιών για την επίτευξη των στόχων μιας τέτοιας εφαρμογής. Η απόδοση του αρχικού μοντέλου κατηγοριοποίησης, υπέδειξε έναν περιορισμό στην ικανότητα επεξεργασίας και ανάλυσης των δεδομένων που είχε στη διάθεσή του. Αντίθετα, η υιοθέτηση του GPT-4 αποτελεί μια ενδεικτική κίνηση προς την αξιοποίηση πιο προηγμένων τεχνολογικών λύσεων, που μπορούν να χειριστούν πιο αποτελεσματικά την πολυπλοκότητα και τον όγκο των δεδομένων.

Κεφάλαιο 5 - Επίλογος

5.1 Συμπεράσματα που προκύπτουν

Συνοψίζοντας, η εφαρμογή που αναπτύχθηκε παρέχει πολύτιμες πληροφορίες σχετικά με την ανάγκη για ευέλικτη προσέγγιση στην ανάπτυξη και βελτίωση παρόμοιων συστημάτων για προτάσεις υλικού. Η αποδοτική ενσωμάτωση του GPT-4, η σύγκριση με ένα μικρότερο μοντέλο και η συνεχής προσαρμογή των τεχνολογιών στις ανάγκες της εφαρμογής και των χρηστών είναι κρίσιμα στοιχεία που συνέβαλαν στα συμπεράσματα αυτής της εργασίας. Η αξιοποίηση της τεχνητής νοημοσύνης ως καταλύτη για τη βελτίωση της απόδοσης και της εμπειρίας του χρήστη ανοίγει νέους δρόμους για την ανάπτυξη και εξέλιξη των ψηφιακών εφαρμογών στο μέλλον.

Η ανάλυση και αξιολόγηση της παρούσας εφαρμογής αποκαλύπτει σημαντικά διδάγματα στον τομέα της τεχνητής νοημοσύνης και της αυτοματοποιημένης επεξεργασίας εικόνων. Η επιτυχία του GPT-4 στην κατηγοριοποίηση και προσαρμογή του περιεχομένου αναδεικνύει την αξία των προηγμένων αλγορίθμων τεχνητής νοημοσύνης στην παροχή εξατομικευμένων υπηρεσιών. Αυτό επιβεβαιώνει ότι η επένδυση σε καινοτόμες τεχνολογίες μπορεί να οδηγήσει σε σημαντικές βελτιώσεις στην ποιότητα και την απόδοση των ψηφιακών πλατφορμών.

Παράλληλα, το πρόβλημα που παρουσιάστηκε με το αρχικό μοντέλο κατηγοριοποίησης υπογραμμίζει τη σημασία της διαθέσιμης ποσότητας και ποιότητας των δεδομένων εκπαίδευσης. Αυτό σηματοδοτεί την ανάγκη για συνεχή επιμέλεια και βελτίωση των συνόλων δεδομένων (datasets), καθώς και την αξιοποίηση τεχνικών ενίσχυσης δεδομένων για την κάλυψη ελλείψεων και την αύξηση της ποικιλίας.

Επιπλέον, η μελέτη αυτή αναδεικνύει την ανάγκη για προσαρμοστικότητα και ευελιξία στην ανάπτυξη και βελτίωση τεχνολογικών λύσεων. Η δυνατότητα του GPT-4 να χειρίζεται διαφορετικούς τύπους δεδομένων και να προσαρμόζεται σε διαφορετικά σενάρια χρήσης επιβεβαιώνει την αξία της ευελιξίας στις ψηφιακές τεχνολογίες.

Τέλος, τα συμπεράσματα αυτής της εργασίας καταδεικνύουν ότι η ενσωμάτωση της τεχνητής νοημοσύνης σε ψηφιακές πλατφόρμες μπορεί να προσφέρει βαθμιαίες και σημαντικές βελτιώσεις στην εμπειρία του χρήστη. Η συνεχής εξέλιξη και επέκταση των δυνατοτήτων των AI συστημάτων ανοίγει νέους διαδρόμους για την παροχή πιο προσωπικών και εξειδικευμένων υπηρεσιών, διασφαλίζοντας έτσι μια πιο άμεση και ευχάριστη αλληλεπίδραση μεταξύ της εφαρμογής και των χρηστών της.

5.2 Πιθανές μελλοντικές βελτιώσεις και αλλαγές

1. Ενίσχυση του συνόλου δεδομένων εκπαίδευσης: Μια από τις πιθανές αλλαγές για τη βελτίωση και εξέλιξη αυτής της έρευνας είναι η ενίσχυση του συνόλου δεδομένων είτε με μεγαλύτερο όγκο εικόνων για κάθε κατηγορία είτε αύξηση των κατηγοριών και διεύρυνση του φάσματος κατηγοριοποίησης του μοντέλου.

2. Προσθήκη επιπλέον χαρακτηριστικών συλλογής δεδομένων στην εφαρμογή: Ακόμη μια αλλαγή θα μπορούσε να είναι η ενίσχυση του αλγορίθμου συλλογής δεδομένων με χαρακτηριστικά όπως η επιλογή σχολιασμού σε εικόνες ή η δυνατότητα αποθήκευσης και επαναδημοσίευσης (share) μιας εικόνας αυξάνοντας έτσι τους τρόπους με τους οποίους ένας χρήστης μπορεί να αλληλεπιδράσει με τις εικόνες. Με αυτή την αλλαγή μπορεί να επιτευχθεί μεγαλύτερη και ορθότερη συλλογή των προτιμήσεων του χρήστη.

3. Εξέταση των αποτελεσμάτων περισσότερων μοντέλων: Για τη βελτίωση της εφαρμογής μια ενδιαφέρουσα αλλαγή θα μπορούσε να αποτελέσει η εξέταση διαφορετικών μοντέλων τεχνητής νοημοσύνης για την παραγωγή των μεταδεδομένων των εικόνων. Αυτή η αλλαγή, θα συμβάλει στην έρευνα εύρεσης του καταλληλότερου μοντέλου του οποίου τα αποτελέσματα μεγιστοποιούν της αποδοτικότητα της εφαρμογής.

Βιβλιογραφία

[1] Sharma, R., & Singh, R. (2016). Evolution of recommender systems from ancient times to modern era: a survey. *Indian Journal of Science and Technology*, *9*(20), 1-12.

[2] Kumar, B., & Sharma, N. (2016). Approaches, issues and challenges in recommender systems: a systematic review. *Indian J. Sci. Technol*, *9*(47), 1-12.

[3] Janiesch, C., Zschech, P., & Heinrich, K. (2021). Machine learning and deep learning. *Electronic Markets*, *31*(3), 685-695.

[4] Roy, D., & Dutta, M. (2022). A systematic review and research perspective on recommender systems. *Journal of Big Data*, *9*(1), 59.

[5] Chen, R., Hua, Q., Chang, Y. S., Wang, B., Zhang, L., & Kong, X. (2018). A survey of collaborative filtering-based recommender systems: From traditional methods to hybrid methods based on social networks. *IEEE Access*, *6*, 64301-64320.

[6] Isinkaye, F. O., Folajimi, Y. O., & Ojokoh, B. A. (2015). Recommendation systems: Principles, methods and evaluation. *Egyptian informatics journal*, *16*(3), 261-273.

[7] Shi, B., Bai, X., & Yao, C. (2016). An end-to-end trainable neural network for image-based sequence recognition and its application to scene text recognition. *IEEE transactions on pattern analysis and machine intelligence*, *39*(11), 2298-2304.

[8] Sai Abhishek, Allena Venkata. (2022). Resnet18 Model With Sequential Layer For Computing Accuracy On Image Classification Dataset. 10. 2320-2882

[9] He, K., Zhang, X., Ren, S., & Sun, J. (2016). Deep residual learning for image recognition. In *Proceedings of the IEEE conference on computer vision and pattern recognition* (pp. 770-778).

[10] Krizhevsky, A., Sutskever, I., & Hinton, G. E. (2012). Imagenet classification with deep convolutional neural networks. *Advances in neural information processing systems*, *25*.

[11] Zagoruyko, S., & Komodakis, N. (2016). Wide residual networks. *arXiv preprint arXiv:1605.07146*.

[12] Ren, S., He, K., Girshick, R., & Sun, J. (2015). Faster r-cnn: Towards real-time object detection with region proposal networks. *Advances in neural information processing systems, 28*.

[13] Lin, T. Y., Dollár, P., Girshick, R., He, K., Hariharan, B., & Belongie, S. (2017). Feature pyramid networks for object detection. In *Proceedings of the IEEE conference on computer vision and pattern recognition* (pp. 2117-2125).

[14] He, K., Gkioxari, G., Dollár, P., & Girshick, R. (2017). Mask r-cnn. In *Proceedings of the IEEE international conference on computer vision* (pp. 2961-2969).

[15] Lika, B., Kolomvatsos, K., & Hadjiefthymiades, S. (2014). Facing the cold start problem in recommender systems. *Expert systems with applications*, *41*(4), 2065-2073.

[16] Vettas, E. (2024). AI metadata based Social Media Recommendation System (Version 1.0.0) [Computer software]. https://github.com/Mnlsvt/recom_system

[17] Zhou, B., Lapedriza, A., Khosla, A., Oliva, A., & Torralba, A. (2017). Places: A 10 million image database for scene recognition. IEEE transactions on pattern analysis and machine intelligence, 40(6), 1452-1464.

- Places 365 code: https://github.com/CSAILVision/places365

[18] Lin, T. Y., Maire, M., Belongie, S., Hays, J., Perona, P., Ramanan, D., ... & Zitnick, C. L. (2014). Microsoft coco: Common objects in context. In Computer Vision–ECCV 2014: 13th European Conference, Zurich, Switzerland, September 6-12, 2014, Proceedings, Part V 13 (pp. 740-755). Springer International Publishing.

[19] Το dataset εικόνων που χρησιμοποιήθηκε για να γεμίσει η βάση δεδομένων:

Google. (October, 2022). Open Images V7 [Data set]. Retrieved 20-2-2024 from https://storage.googleapis.com/openimages/web/download_v7.html

[20] Τα Datasets των κατηγοριών που χρησιμοποιήθηκαν για την εκπαίδευση του μοντέλου:

1. Nature Dataset: "ARNAUD ROUGETET. Landscape Pictures, Retrieved 3-1-2023 from https://www.kaggle.com/datasets/arnaud58/landscape-pictures/code".
2. Sports Dataset: "GERRY. 100 Sports Image Classification, Retrieved 11-1-2024 from https://www.kaggle.com/datasets/gpiosenka/sports-classification".
3. Cars Dataset: "Jessica Li. Stanford Cars Dataset, 10-01-2024 from https://www.kaggle.com/datasets/jessicali9530/stanford-cars-dataset/data".
4. Gaming Dataset: "The_AlphaLaser. Gameplay Images, Retrieved 11-1-2024 from https://www.kaggle.com/datasets/aditmagotra/gameplay-images".
5. Houses Dataset: "mbadal1996. Subset of House Prices and Images (SoCal), Retrieved 11-1-2023 from https://www.kaggle.com/datasets/mbadal1996/subset-of-house-prices-and-images-socal".
6. Movies Dataset: "raman. Movie Classifier, Retrieved 8-1-2024 from https://www.kaggle.com/datasets/raman77768/movie-classifier".
7. Food Dataset: "Aleksandr Antonov. Food11 Image Dataset, Retrieved 22-12-2023 from https://www.kaggle.com/datasets/trolukovich/food11-image-dataset?select=training".
8. Fitness Dataset: "Hasyim Abdillah. Workout/Exercises Images, Retrieved 11-1-2023 from https://www.kaggle.com/datasets/hasyimabdillah/workoutexercises-images".
9. Pets Dataset: "ilovescience. The Oxford-IIIT Pet Dataset, Retrieved 11-1-2023 from https://www.kaggle.com/datasets/tanlikesmath/the-oxfordiiit-pet-dataset".

[21] Οι πηγές των εικόνων που εμφανίζονται στο κείμενο:

1. Εικόνες σκύλων - σελίδα 42: https://cornercafebc.com
2. Δρόμος - σελίδα 53: https://en.wiktionary.org/wiki/street
3. Αυτοκίνητο - σελίδα 61: https://playground.com
4. Κατοικίδια - σελίδα 63: https://twitter.com/everpixapp/status/1528988749467635714