



ΕΘΝΙΚΟ ΜΕΤΣΟΒΙΟ ΠΟΛΥΤΕΧΝΕΙΟ

ΣΧΟΛΗ ΗΛΕΚΤΡΟΛΟΓΩΝ ΜΗΧΑΝΙΚΩΝ
ΚΑΙ ΜΗΧΑΝΙΚΩΝ ΥΠΟΛΟΓΙΣΤΩΝ

ΤΟΜΕΑΣ ΤΕΧΝΟΛΟΓΙΑΣ ΠΛΗΡΟΦΟΡΙΚΗΣ ΚΑΙ ΥΠΟΛΟΓΙΣΤΩΝ

Σύστημα Συστάσεων Μουσικής για κινητές συσκευές βασισμένο σε Αναγνώριση Συναισθήματος χρήστη

ΔΙΠΛΩΜΑΤΙΚΗ ΕΡΓΑΣΙΑ

Κωνσταντίνα Γ. Φρέρη

Επιβλέπων : Στέφανος Κόλλιας

Καθηγητής, ΕΜΠ

Συνεπιβλέπουσα : Παρασκευή Τζούβελη

ΕΔΙΠ, ΕΜΠ

Αθήνα, Οκτώβριος 2022



ΕΘΝΙΚΟ ΜΕΤΣΟΒΙΟ ΠΟΛΥΤΕΧΝΕΙΟ

ΣΧΟΛΗ ΗΛΕΚΤΡΟΛΟΓΩΝ ΜΗΧΑΝΙΚΩΝ
ΚΑΙ ΜΗΧΑΝΙΚΩΝ ΥΠΟΛΟΓΙΣΤΩΝ

ΤΟΜΕΑΣ ΤΕΧΝΟΛΟΓΙΑΣ ΠΛΗΡΟΦΟΡΙΚΗΣ ΚΑΙ ΥΠΟΛΟΓΙΣΤΩΝ

Σύστημα Συστάσεων Μουσικής για κινητές συσκευές βασισμένο σε Αναγνώριση Συναισθήματος χρήστη

ΔΙΠΛΩΜΑΤΙΚΗ ΕΡΓΑΣΙΑ

Κωνσταντίνα Γ. Φρέρη

Επιβλέπων : Στέφανος Κόλλιας

Καθηγητής, ΕΜΠ

Συνεπιβλέπουσα : Παρασκευή Τζούβελη

ΕΔΙΠ, ΕΜΠ

Εγκρίθηκε από την τριμελή εξεταστική επιτροπή την 31^η Σεπτεμβρίου 2022.

.....
Στέφανος Κόλλιας
Καθηγητής, ΕΜΠ

.....
Αθανάσιος Βουλόδημος
Επικουρος Καθηγητής, ΕΜΠ

.....
Γεώργιος Στάμου
Καθηγητής, ΕΜΠ

Αθήνα, Οκτώβριος 2022

.....

Κωνσταντίνα Γ. Φρέρη

Διπλωματούχος Ηλεκτρολόγος Μηχανικός και Μηχανικός Υπολογιστών Ε.Μ.Π.

Copyright © Κωνσταντίνα Γ. Φρέρη, 2022

Με επιφύλαξη παντός δικαιώματος. All rights reserved.

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

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

Περίληψη

Τα συστήματα συστάσεων μουσικής (MRS) είναι ένας αναδυόμενος τομέας εφαρμογών που βοηθά τους χρήστες να βρουν την αγαπημένη τους μουσική ανάμεσα σε ένα τεράστιο όγκο μουσικών κομματιών. Οι περισσότερες υπάρχουσες μέθοδοι προτάσεων μουσικής επικεντρώνονται στην εξερεύνηση των προφίλ των χρηστών, του ιστορικού ακρόασης και του ηχητικού σήματος της μουσικής για να προτείνουν τα πιο σχετικά στοιχεία στους χρήστες. Ωστόσο, τα συναισθήματα των χρηστών επηρεάζουν σε μεγάλο βαθμό τις μουσικές τους προτιμήσεις και διαδραματίζουν ζωτικό ρόλο στη διαδικασία συστάσεων μουσικής. Έτσι εμφανίζεται στο προσκήνιο το πεδίο της αναγνώρισης συναισθημάτων. Οι περισσότερες εφαρμογές αναγνώρισης συναισθημάτων βασίζονται στις εκφράσεις του προσώπου, καθώς αποτελούν ένα από τα κύρια κανάλια πληροφόρησης στη διαπροσωπική επικοινωνία. Η αναγνώριση συναισθημάτων προσώπου (FER) είναι ένα σημαντικό θέμα στους τομείς της υπολογιστικής όρασης, της αλληλεπίδρασης ανθρώπου-υπολογιστή και της τεχνητής νοημοσύνης, λόγω των σημαντικών ακαδημαϊκών και εμπορικών δυνατοτήτων της. Στην εργασία αυτή, φέρνουμε τα δύο πεδία μαζί κατασκευάζοντας ένα σύστημα προτάσεων μουσικής που βασίζεται στο συναίσθημα. Το σύστημά μας, που ονομάζεται "Soundays", είναι μια εφαρμογή Android που προτείνει στον χρήστη μια λίστα τραγουδιών με βάση το τρέχον συναίσθημά του. Αρχικά, καταγράφεται η εικόνα του προσώπου του χρήστη χρησιμοποιώντας την κάμερα και στη συνέχεια ανιχνεύεται το συναίσθημα και παράγεται μια λίστα μουσικών κομματιών που ταιριάζει στη διάθεσή του. Στο Soundays, η αναγνώριση συναισθημάτων εφαρμόζεται χρησιμοποιώντας το Py-Feat και η λίστα των τραγουδιών δημιουργείται με τη βοήθεια του Spotify API.

Λέξεις κλειδιά

Αναγνώριση συναισθήματος, Εφαρμογή, Συστήματα συστάσεων, Μουσική

Abstract

Music recommendation systems (MRS) are an emerging application that helps users find their favorite music across a wide range of tracks. Most existing music recommendation methods focus on exploring users' profiles, listening histories and audio signal of music to recommend the most relevant items to users. However, users' emotion greatly affects their music preference and plays a vital role in music recommendation. This is how the emotion recognition field comes into the picture. Most emotion recognition applications are based on facial expressions as they are one of the main information channels in interpersonal communication. Facial emotion recognition (FER) is an important topic in the fields of computer vision, human-computer interaction and artificial intelligence owing to its significant academic and commercial potential. In this paper, we bring the two fields together by building an emotion-based music recommendation system. Our system, called "Soundays", is an Android application that suggests the user a list of songs based on his current emotion. The system captures user's image using the camera. It then detects the emotion and produces a playlist that fits user's mood. In Soundays, emotion recognition is applied using Py-Feat and the list of songs is created with the help of Spotify API.

Key words

Emotion recognition, Mobile App, Recommendation system, Music

Η σελίδα αυτή είναι σκόπιμα λευκή.

Πίνακας περιεχομένων

1.	Εισαγωγή	10
2.	Θεωρητικό Μέρος	13
2.1	Αναγνώριση Συναισθημάτων	13
2.1.1	Μέθοδοι αναγνώρισης συναισθημάτων	14
2.1.2	Μοντελοποίηση συναισθημάτων	14
2.1.3	Αναγνώριση Συναισθημάτων από Εικόνες Προσώπου (FER).....	16
2.1.4	Διαδικασία FER.....	18
2.1.5	Σύνολα Δεδομένων για FER	24
2.1.6	Ανάλυση Δημοσιεύσεων για FER.....	28
2.1.7	Συναισθήματα και Μουσική.....	31
2.2	Συστήματα Συστάσεων	33
2.2.1	Κατηγορίες	34
2.2.2	Προκλήσεις.....	37
2.2.3	Μετρικές.....	38
2.2.4	Εφαρμογές	40
2.2.5	Συστήματα συστάσεων μουσικής (MRS).....	42
2.2.6	Ανάλυση Δημοσιεύσεων MRS.....	46
2.2.7	Συναισθήματα και MRS: Σχετικές έρευνες.....	48
2.2.8	Spotify	51
3.	Πρακτικό Μέρος.....	55
3.1	Flutter.....	55
3.2	Py-Feat	56
3.3	Spotify API	59
3.4	Παρουσίαση Εφαρμογής	61
4.	Επίλογος	78

1. Εισαγωγή

Η έκφραση συναισθημάτων είναι σημαντικό και αναπόσπαστο κομμάτι της ζωής του ανθρώπου αφού τα συναισθήματα επηρεάζουν τόσο την ψυχολογική όσο και τη φυσιολογική του κατάσταση. Τα θετικά συναισθήματα συμβάλουν στη βελτίωση της υγείας και μπορούν να οδηγήσουν σε χαλάρωση και ηρεμία. Από την άλλη, τα αρνητικά συναισθήματα είναι επιβαρυντικά για τον ανθρώπινο οργανισμό και πολλές φορές οδηγούν σε ακραία περιστατικά ενώ συνήθως εκφράζονται ως ένταση ή ανησυχία. [1] Πολλές φορές οι άνθρωποι αδυνατούν να εκφράσουν τα συναισθήματά τους λόγω σωματικών ή άλλων προβλημάτων υγείας. Μέσα από τα στοιχεία αυτά φαίνεται πως είναι σημαντικό να βρούμε τρόπους να αναγνωρίζουμε τα συναισθήματα.

Η αναγνώριση συναισθημάτων είναι ένας τομέας που έχει αποκτήσει ιδιαίτερο ενδιαφέρον τα τελευταία χρόνια λόγω της πληθώρας εφαρμογών του. Έχει δει μεγάλη ανάπτυξη μέσα από το μάρκετινγκ, τη ρομποτική, την υγεία και την εκπαίδευση, που είναι κάποιοι από τους πολλούς κλάδους στους οποίους έχει εφαρμοστεί. [2]–[5] Με βασικό μέσο μετάδοσης των συναισθημάτων να είναι οι εκφράσεις του προσώπου, πολλές έρευνες εστιάζουν στην αναγνώριση εκφράσεων προσώπου (Facial Emotion Recognition, FER).

Στον τομέα αυτό, δύο είναι οι βασικές προσεγγίσεις που εφαρμόζονται: οι κλασικές προσεγγίσεις και οι προσεγγίσεις που βασίζονται σε χρήση νευρωνικών δικτύων. [6] Παρόλο που σε βάθος χρόνου οι περισσότερες έρευνες χρησιμοποιούν τις κλασικές προσεγγίσεις, τελευταία η χρήση νευρωνικών δικτύων είναι πιο δημοφιλής. Η διαδικασία που ακολουθείται στο FER αποτελείται από τρία βασικά βήματα. Πρώτο βήμα είναι η αναγνώριση προσώπου, δηλαδή η εύρεση του προσώπου μέσα σε μια εικόνα, που είναι ένας επίσης πολύ μεγάλος τομέας έρευνας. Ο πιο γνωστός αλγόριθμος που χρησιμοποιείται, είναι ο αλγόριθμος Viola-Jones [7]. Δεύτερο βήμα, είναι η εξαγωγή χαρακτηριστικών, δηλαδή η αναγνώριση σημείων του προσώπου που ανήκουν στα μάτια, τη μύτη, το στόμα κτλ.. Τρίτον, είναι η ταξινόμηση χαρακτηριστικών, δηλαδή η διάκριση των εξαγόμενων χαρακτηριστικών ανάλογα με την κατηγορία εκφράσεων προσώπου στην οποία ανήκουν.

Μέσα από την παραπάνω διαδικασία καταλήγουμε σε κάποια συναισθήματα ανάλογα με την κάθε έρευνα και το μοντέλο συναισθημάτων που αξιοποίησε. Γενικά, τα μοντέλα συναισθημάτων διακρίνονται σε: διακριτά και συνεχή [8]. Στην πρώτη κατηγορία, έχουμε διακριτές κατηγορίες συναισθημάτων. Το βασικότερο μοντέλο εδώ είναι αυτό του Ekman [9] οποίος έκανε διάκριση σε έξι βασικά συναισθήματα: θυμός, αηδία, φόβος, χαρά, λύπη και έκπληξη. Στην δεύτερη κατηγορία, γίνεται χρήση πολλών διαστάσεων για την κατηγοριοποίηση των συναισθημάτων. Κύριο μοντέλο είναι αυτό του Russell [10], κατά το οποίο τα συναισθήματα κατανέμονται σε ένα δισδιάστατο κυκλικό χώρο: διάσταση «σθένους» (valence) και διάσταση «διέγερσης» (arousal).

Ένα μέσο που μπορεί να μεταδώσει απίστευτα συναισθήματα είναι η μουσική. Ο δρόμος ανάμεσα στη μουσική και το συναίσθημα είναι διπλής κατεύθυνσης. Από τη μία, κάποιος που ακούει μουσική διακατέχεται από συναισθήματα και από την άλλη κάποιος που κυριαρχείται από συναισθήματα μπορεί να τα εκφράσει μέσα από τη μουσική. Έτσι, οι μουσικές

πλατφόρμες έχουν πλέον πολλούς χρήστες και συνεχίζουν να επεκτείνονται σε όλο τον κόσμο. Η αναζήτηση ανάμεσα στα εκατομμύρια μουσικά κομμάτια είναι εξαιρετικά δύσκολη λόγω του όγκου των πληροφοριών. Για το λόγο αυτό τα συστήματα συστάσεων μουσικής έχουν γνωρίσει μεγάλη αύξηση στις μέρες μας. [11]

Συστήματα συστάσεων (RS) δεν εφαρμόζονται μόνο στον τομέα της μουσικής αλλά και σε άλλες υπηρεσίες ροής, στο ηλεκτρονικό εμπόριο, στους τομείς της υγείας, του τουρισμού και της εκπαίδευσης. [12] Τα συστήματα συστάσεων μπορούν να χωριστούν σε κάποιες βασικές κατηγορίες. Οι κατηγορίες αυτές είναι: Συνεργατικού φιλτραρίσματος (Collaborative filtering) όπου επιλέγονται στοιχεία με βάση τις ομοιότητες μεταξύ των προτιμήσεων διαφορετικών χρηστών, Φιλτραρίσματος βάσει περιεχομένου (Content-based filtering) όπου επιλέγονται αντικείμενα που βασίζονται στις ομοιότητες μεταξύ της περιγραφής του περιεχομένου ενός αντικείμενου και των προτιμήσεων του χρήστη, και Υβριδικού φιλτραρίσματος (Hybrid filtering) που συνδυάζει το συνεργατικό και το φιλτράρισμα βάσει περιεχομένου. [13] Κάθε σύστημα αντιμετωπίζει ορισμένα προβλήματα, έτσι και τα συστήματα συστάσεων πρέπει να αντιμετωπίσουν τις ακόλουθες προκλήσεις: ψυχρή εκκίνηση (cold-start problem), αραιότητα (sparsity), επεκτασιμότητα (scalability), συνωνυμία (synonymy), γκριζο πρόβατο (grey sheep), ανάλυση περιορισμένου περιεχομένου (limited content analysis) και εξειδίκευση περιεχομένου (overspecialization) [14].

Τα συστήματα συστάσεων μουσικής (MRS) αποτελούν μια υποκατηγορία των RS αλλά διαφέρουν από τα υπόλοιπα συστήματα συστάσεων καθώς τα μουσικά κομμάτια έχουν συγκεκριμένα χαρακτηριστικά, που τα κάνουν να ξεχωρίζουν από άλλα αντικείμενα/προϊόντα όπως ταινίες ή βιβλία. Οι ιδιαιτερότητες των μουσικών κομματιών περιλαμβάνουν: τη διάρκεια, το μέγεθος, τη διαδοχική κατανάλωση, την επαναληπτική σύσταση, την παθητική καταναλωτική συμπεριφορά και την πρόκληση δυνατών συναισθημάτων. [15] Όσον αφορά την αξιολόγηση των MRS, γίνεται χρήση συγκεκριμένων μετρικών και άλλων εξωτερικών στοιχείων.

Οι περισσότερες υπάρχουσες μέθοδοι προτάσεων μουσικής επικεντρώνονται στην εξερεύνηση των προφίλ των χρηστών και του ιστορικού ακρόασης για να προτείνουν σχετικά τραγούδια. Όμως, υπάρχουν και άλλα στοιχεία, όπως η προσωπικότητα, το περιβάλλον ακρόασης και το συναίσθημα που επηρεάζουν τις προτιμήσεις των χρηστών. Για το λόγο αυτό πρόσφατα φαίνεται να έχουν αρχίσει πολλές έρευνες για συστήματα συστάσεων με βάση το συναίσθημα (emotion-based music recommendation systems).

Σκοπός της παρούσας εργασίας, είναι να βρούμε και να αναλύσουμε τα βασικά στοιχεία των τομέων που αναφέρθηκαν παραπάνω με βάση μελέτες και έρευνες που έχουν γίνει στο παρελθόν, όπως επίσης και να δημιουργήσουμε μια εφαρμογή που προτείνει στο χρήστη τραγούδια με βάση το τρέχον συναίσθημά του. Στο πλαίσιο αυτό, η εργασία μας αποτελείται από ένα θεωρητικό μέρος και ένα πρακτικό μέρος. Στο θεωρητικό μέρος (κεφ. 2), περιγράφονται οι τομείς της αναγνώρισης συναισθήματος (κεφ. 2.1) και των συστημάτων συστάσεων (κεφ. 2.2). Για καθέναν από αυτούς δίνεται μια επεξήγηση, αναφέρονται τα εργαλεία και οι μέθοδοι που χρησιμοποιούνται και γίνεται μια μελέτη των σχετικών δημοσιεύσεων. Στο πρακτικό μέρος (κεφ. 3), περιγράφεται η διαδικασία ανάπτυξης της εφαρμογής, τα εργαλεία που χρησιμοποιήθηκαν και τα τελικά αποτελέσματα. Επίσης, γίνονται

κάποιες δοκιμές σχετικά με ακουστικά χαρακτηριστικά των μουσικών κομματιών που εκφράζουν συναίσθημα.

Η έρευνα που διεξάχθηκε για το θεωρητικό μέρος έγινε με χρήση του Google Scholar και οι βιβλιογραφικές αναφορές βρέθηκαν από πηγές όπως Science Direct, IEEE Xplore, SpringerLink, Research Gate κ.α. Οι αναφορές συλλέχθηκαν και προστέθηκαν στην εργασία με χρήση του εργαλείου Mendeley.

2. Θεωρητικό Μέρος

Στο θεωρητικό μέρος της εργασίας αυτής θα αναλυθούν δυο βασικές ενότητες: η αναγνώριση συναισθημάτων και τα συστήματα συστάσεων. Η εφαρμογή που θα αναλυθεί στο Πρακτικό Μέρος χρησιμοποιεί την αναγνώριση συναισθημάτων μέσω του εργαλείου Py-Feat για να αναγνωρίσει τη διάθεση του χρήστη, ενώ στη συνέχεια μέσω ενός συστήματος συστάσεων που παρέχεται από το Spotify, εμφανίζει στο χρήστη μια λίστα τραγουδιών που ταιριάζουν στη συναισθηματική του κατάσταση. Λαμβάνοντας αυτά υπόψιν, είναι απαραίτητο να αναλυθούν και οι δύο παραπάνω ενότητες για τη καλύτερη κατανόηση της λειτουργίας και της χρησιμότητας της εφαρμογής.

2.1 Αναγνώριση Συναισθημάτων

Η έκφραση συναισθημάτων είναι σημαντικό και αναπόσπαστο κομμάτι της ζωής ενός ανθρώπου. Τα συναισθήματα επηρεάζουν τόσο τη φυσιολογική όσο και την ψυχολογική κατάστασή μας. Τα θετικά συναισθήματα συμβάλλουν στη βελτίωση της ανθρώπινης υγείας και της αποδοτικότητας της εργασίας, ενώ τα αρνητικά συναισθήματα μπορεί να προκαλέσουν προβλήματα υγείας, να οδηγήσουν σε κατάθλιψη ή και σε ακόμα χειρότερες καταστάσεις. [1]

Η αυτόματη αναγνώριση συναισθημάτων (Emotion Recognition, ER) έχει αποκτήσει αυξημένο ενδιαφέρον τα τελευταία χρόνια με όλο και περισσότερα συστήματα να την ενσωματώνουν καθώς τα ενδυναμώνει και τα κάνει πιο θεμιτά προς τους χρήστες. [16] Η ER είναι ιδιαίτερα σημαντική σε τομείς όπως:

- ρομποτική – για την καλύτερη αλληλεπίδραση ανθρώπου υπολογιστή [17]
- μάρκετινγκ – για τη δημιουργία εξειδικευμένων διαφημίσεων, για εφαρμογές όπως μουσική ανάλογα με τη διάθεση [4]
- εκπαίδευση – για τη βελτίωση των μεθοδολογιών και τεχνικών που αξιοποιούνται κατά τη διαδικασία μάθησης, για αυτοματοποιημένα συστήματα εκμάθησης [5].

Ενώ η αναγνώριση με βάση τα φυσιολογικά σήματα έχει εφαρμοστεί και σε τομείς όπως:

- ασφαλής οδήγηση – ανιχνεύοντας κούραση, νύστα και άγχος και προειδοποιώντας τον οδηγό για την επικινδυνότητα της κατάστασης [2]
- υγεία – ιδιαίτερα στην έγκαιρη διάγνωση, παρακολούθηση και θεραπεία ψυχικών ασθενειών [3]
- κοινωνική ασφάλιση – για την ανίχνευση κρυμμένων πληροφοριών [1]. [8]

Πιο συγκεκριμένα, στον τομέα της υγείας εμφανίζονται πολλές περιπτώσεις αξιοποίησης του ER. Για παράδειγμα, η διαταραχή αυτιστικού φάσματος (ΔΑΦ) είναι μια αναπτυξιακή διαταραχή που χαρακτηρίζεται από ελλείμματα στην κοινωνική επικοινωνία και από την

παρουσία περιορισμένων, επαναλαμβανόμενων συμπεριφορών ή ενδιαφερόντων. Πολλές εφαρμογές έχουν αναπτυχθεί ώστε να βοηθήσουν τα παιδιά με αυτισμό εκπαιδεύοντάς τα στην αναγνώριση συναισθημάτων, ενισχύοντας έτσι τις κοινωνικές τους δεξιότητες [18]. Επιπλέον, η κατάθλιψη είναι μια διαταραχή της διάθεσης που προκαλεί ένα επίμονο αίσθημα θλίψης και απώλεια ενδιαφέροντος. Η αναγνώριση συναισθημάτων είναι ζωτικής σημασίας για την ανίχνευση και τη διάγνωση πιθανών ασθενών με ψυχικές διαταραχές όπως η κατάθλιψη [19]. Τέλος, η τεχνολογία αυτή μπορεί να αξιοποιηθεί γενικότερα για την υποστήριξη ατόμων με διαταραχές υγείας που μπορούν να μειώσουν τις επικοινωνιακές τους δεξιότητες (π.χ. μυϊκή απώλεια, εγκεφαλικό ή, πιο απλά, πόνο) προκειμένου να αναγνωρίσουμε συναισθήματα και να ανταποκρινόμαστε σε πραγματικό χρόνο. [20]

2.1.1 Μέθοδοι αναγνώρισης συναισθημάτων

Οι άνθρωποι συμπεραίνουν τις συναισθηματικές καταστάσεις άλλων ανθρώπων, όπως χαρά, λύπη και θυμό, από τις εκφράσεις του προσώπου, τα λόγια που χρησιμοποιούν και τον τόνο της φωνής. Σύμφωνα με διάφορες μελέτες, τα λεκτικά στοιχεία αποτελούν το ένα-τρίτο της ανθρώπινης επικοινωνίας και τα μη λεκτικά αποτελούν τα δύο-τρίτα. Μεταξύ των διαφόρων μη λεκτικών στοιχείων που φέρουν συναισθηματικό νόημα, είναι οι εκφράσεις του προσώπου, που αποτελούν ένα από τα βασικότερα μέσα πληροφόρησης στη διαπροσωπική επικοινωνία [6].

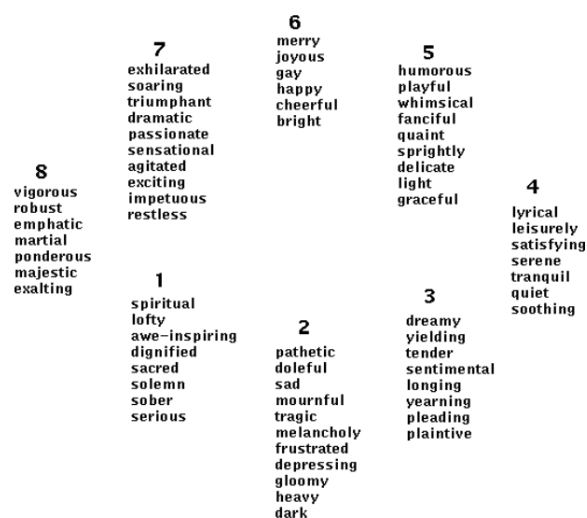
Για την αυτόματη αναγνώριση συναισθημάτων διάφορα στοιχεία μπορούν να αξιοποιηθούν. Αυτά μπορούν να ταξινομηθούν σε δύο μεγάλες κατηγορίες [1]. Στην πρώτη, εντάσσονται τα ανθρώπινα φυσικά σήματα όπως έκφραση προσώπου, ομιλία, χειρονομίες, στάση, κ.λπ., τα οποία έχουν το πλεονέκτημα της εύκολης συλλογής και έχουν μελετηθεί για χρόνια. Ωστόσο, η αξιοπιστία αυτών δεν μπορεί να εγγυηθεί αφού σε ορισμένες περιπτώσεις οι άνθρωποι μπορεί να συγκατατιούνται ή να μην μπορούν να εκφράσουν τα πραγματικά τους συναισθήματα. Έτσι, για να επιτευχθεί καλύτερη ακρίβεια στην αξιολόγηση των αληθινών συναισθημάτων, εισάγουμε τη δεύτερη κατηγορία. Η δεύτερη κατηγορία χρησιμοποιεί τα εσωτερικά σήματα ή φυσιολογικά σήματα, τα οποία περιλαμβάνουν το ηλεκτροεγκεφαλογράφημα (EEG), τη θερμοκρασία (T), το ηλεκτροκαρδιογράφημα (ΗΚΓ), το ηλεκτρομυογράφημα (EMG), τη γαλβανική απόκριση δέρματος (GSR), την αναπνοή (RSP), την αρτηριακή πίεση κ.λπ. [1][8] Βέβαια, πολύ συχνά γίνεται χρήση από τους μελετητές μιας μίξης των παραπάνω μεθόδων και συγχώνευση των πληροφοριών που αποκτήθηκαν (πχ. εκφράσεις προσώπου, κίνηση κεφαλιού και σώματος) για βελτίωση των αποτελεσμάτων.

2.1.2 Μοντελοποίηση συναισθημάτων

Τον 19ο αιώνα, η αξία της αναγνώρισης συναισθημάτων αναγνωρίστηκε όταν «Η Έκφραση Συναισθημάτων στον Άνθρωπο και τα Ζώα» γράφτηκε από τον Κάρολο Δαρβίνο. Αυτό το

βιβλίο ενέπνευσε σε μεγάλο βαθμό τη μελέτη των συναισθημάτων. [21] Διάφοροι ερευνητές απεικόνισαν την ταξινόμηση των συναισθημάτων με δύο μεθόδους: (i) διαχωρισμός των συναισθημάτων σε διακριτές κατηγορίες και (ii) χρήση πολλαπλών διαστάσεων για την κατηγοριοποίηση των συναισθημάτων. [8]

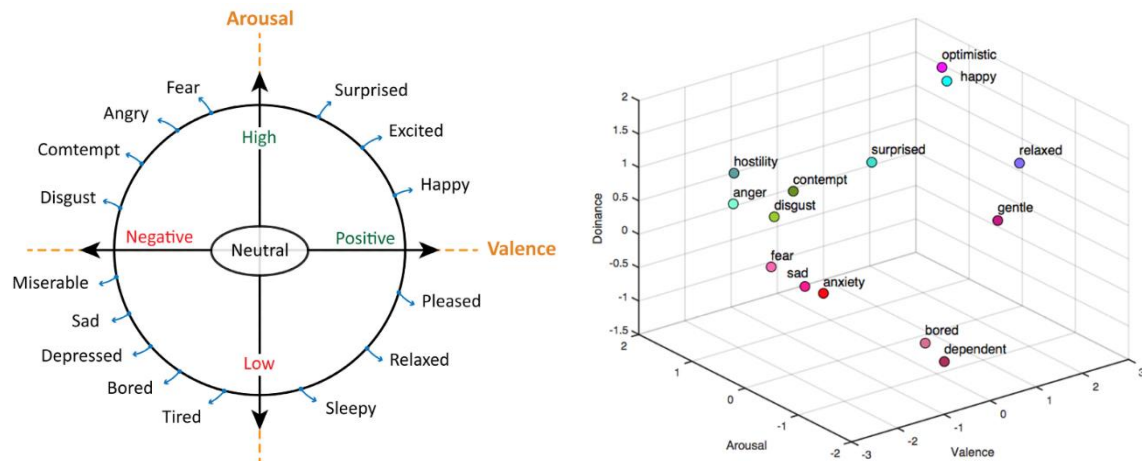
Στην πρώτη προσέγγιση (i) βρίσκεται το βασικό μοντέλο συναισθημάτων του Ekman (1969) [22] ο οποίος συμπέρανε ότι τα έξι βασικά συναισθήματα είναι: ο θυμός, η αηδία, ο φόβος, η χαρά, η λύπη και η έκπληξη. Ο Ekman και ο Friesen ανέπτυξαν επίσης το Facial Action Coding System (1978) [23], ένα σύστημα για την αναγνώριση των συναισθημάτων από εκφράσεις του προσώπου, όπως αναλύεται παρακάτω [19]. Ένα άλλο μοντέλο είναι αυτό του Hevner (1936) [24]. Ο κύκλος επιθέτων του Hevner (Hevner's adjective circle) είναι αναπαράσταση μουσικής-συναισθήματος, όπου οι μουσικές παράμετροι χαρτογραφούνται σε έναν κύκλο συναισθηματικών όρων. Ο κύκλος χωρίζεται σε οκτώ ομάδες στενά συνδεδεμένων επιθέτων όπως φαίνονται στην Εικόνα 1. Οι οκτώ καταστάσεις συναισθημάτων είναι: αξιοπρεπής (dignified), λυπημένη (sad), ονειρική (dreamy), καταπραΰντική (soothing), χαριτωμένη (graceful), χαρούμενη (joyous), συναρπαστική (exciting) και ζωνρή (vigorous). Αυτή η αναπαράσταση ταιριάζει σε μουσικές εφαρμογές, επειδή παρέχει έναν αριθμό μουσικών παραμέτρων για την περιγραφή κάθε διάθεσης. [25]



Εικόνα 1: Κύκλος επιθέτων του Hevner [25]

Στην δεύτερη προσέγγιση (ii) βρίσκεται ως το πιο αντιπροσωπευτικό δείγμα, το μοντέλο του Russell (1980) [10]. Το μοντέλο του Russell προτείνει ότι τα συναισθήματα κατανέμονται σε δισδιάστατο κυκλικό χώρο: διάσταση ‘σθένους’ (valence) και διάσταση ‘διέγερσης’ (arousal), όπως φαίνεται στην Εικόνα 2. Η διάσταση ‘σθένους’ υποδηλώνει πόσο ευχάριστο ή δυσάρεστο είναι ένα συναίσθημα. Η διάσταση ‘διέγερσης’ διαφοροποιεί τις καταστάσεις ανάλογα με το πόση ένταση προκαλεί στο άτομο. Από την άλλη, ο Thayer (1989) πρότεινε ότι οι δύο διαστάσεις του συναισθήματος ήταν δύο ξεχωριστές διαστάσεις διέγερσης: η ενεργητική διέγερση (energy arousal) και η τεταμένη διέγερση (tension arousal). Η ενέργεια (energy) αναφέρεται στην ένταση του ήχου στη μουσική και η ένταση (tension) αναφέρεται στην τονικότητα και το ρυθμό της μουσικής. Σε αυτή την δεύτερη προσέγγιση, βρίσκουμε

επίσης το μοντέλο του Mehrabian (1996) [26], ένα μοντέλο που βασίζεται σε μια τρισδιάστατη αναπαράσταση PAD (Pleasure – Arousal – Dominance), όπως φαίνεται στην Εικόνα 3. Η ‘ευχαρίστηση’ (pleasure) αντιπροσωπεύει το αν ένα συναίσθημα είναι θετικό ή αρνητικό. Η τρίτη διάσταση είναι η διάσταση ‘κυριαρχίας’ που δείχνει εάν το υποκείμενο αισθάνεται ότι ελέγχει την κατάσταση ή όχι. Για παράδειγμα, ενώ ο φόβος και ο θυμός είναι δυσάρεστα συναισθήματα, ο θυμός είναι ένα κυρίαρχο συναίσθημα και ο φόβος είναι ένα υποτακτικό [27], [28]



Εικόνες 2, 3: Μοντέλα συναισθημάτων. Στην αριστερή εικόνα εμφανίζεται το δισδιάστατο μοντέλο συναισθημάτων του Russell (valence-arousal plane) [8]. Στη δεξιά εικόνα εμφανίζεται το τρισδιάστατο μοντέλο του Mehrabian (valence-arousal-dominance plane) [1].

Οι περισσότερες υπολογιστικές τεχνικές βασίζονται σε διακριτές κατηγορίες συναισθημάτων λόγω της απλότητας και της οικειότητάς τους. Παρόλα αυτά, οι κατηγορίες είναι περιορισμένες και δε μπορούν πάντα να καλύψουν όλο το εύρος των συναισθημάτων. Αυτό είναι και το βασικό πλεονέκτημα της χρήσης πολλαπλών διαστάσεων. Από την άλλη, τα δισδιάστατα μοντέλα έχουν επικριθεί για την έλλειψη διαφοροποίησης μεταξύ συναισθημάτων που είναι στενά γειτονικά, όπως ο θυμός και ο φόβος. Παρατηρούμε, λοιπόν, ότι το μοντέλο που χρησιμοποιείται σε κάθε μελέτη εξαρτάται από το σύνολο των συναισθημάτων που τίθεται προς ανίχνευση και δεν είναι μοναδικό. [27]

2.1.3 Αναγνώριση Συναισθημάτων από Εικόνες Προσώπου (FER)

Στη συνέχεια της εργασίας εστιάζουμε στο πρόβλημα της αναγνώρισης συναισθημάτων σε εικόνες προσώπου (Facial Emotion Recognition, FER). Σε αυτό το κομμάτι μπορούμε να διακρίνουμε δύο τάσεις: (i) τις κλασικές προσεγγίσεις, που αποτελούνται από κλασική επεξεργασία εικόνας, αναγνώριση προτύπων και διάφορους ταξινομητές, πιο συγκεκριμένα που παρέχονται από μια χειροποίητη διαδικασία σχεδιασμού χαρακτηριστικών, και (ii) τις NNB (Neural Network Based) προσεγγίσεις (ή αλλιώς προσεγγίσεις βαθιάς μάθησης), όπου κλασικά νευρωνικά δίκτυα και το συνελκτικό αντίστοιχό τους εξετάζονται, τα

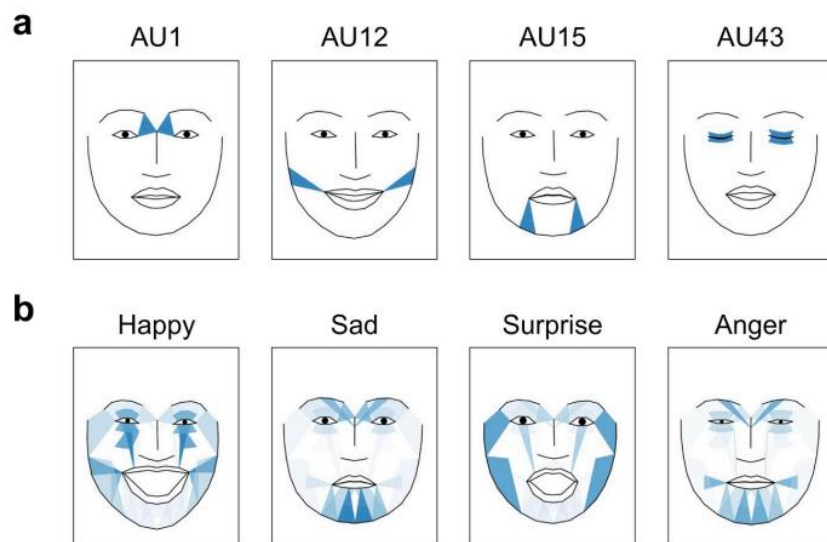
χαρακτηριστικά των οποίων μαθαίνονται από δεδομένα χρησιμοποιώντας εξαγωγές γενικών χαρακτηριστικών. [6], [29]

(i) Κλασικές προσεγγίσεις

Η βασική διαδικασία της κλασικής προσέγγισης του FER περιλαμβάνει τρία βασικά βήματα:

- Ανίχνευση προσώπου
- Εξαγωγή χαρακτηριστικών
- Ταξινόμηση χαρακτηριστικών

Αρχικά, σε μερικές μελέτες γίνεται κάποια προ-επεξεργασία της εικόνας, όπως αφαίρεση θορύβου και βελτίωση (ανάδειξη ορισμένων πληροφοριών μιας εικόνας, καθώς και αποδυνάμωση ή αφαίρεση οποιασδήποτε περιττής πληροφορίας) [4]. Έπειτα, ένα πρόσωπο ανιχνεύεται από μια εικόνα εισόδου και τα στοιχεία του προσώπου (π.χ. μάτια και μύτη) ή τα ορόσημα εντοπίζονται από την περιοχή του προσώπου. Στη συνέχεια, διάφορα χωρικά και χρονικά χαρακτηριστικά εξάγονται από τα στοιχεία του προσώπου. Τέλος, προ-εκπαιδευμένοι ταξινομητές, όπως η μηχανή διανύσματος υποστήριξης (Support Vector Machine, SVM), το AdaBoost και το τυχαίο δάσος, παράγουν τα αποτελέσματα αναγνώρισης βασισμένοι στα εξαγόμενα χαρακτηριστικά [6].



Εικόνα 4: Οπτικοποίηση των AUs. a) Εκφράσεις προσώπου που παράγονται ενεργοποιώντας ανεξάρτητα τα AUs 1 (εσωτερικό φρύδι), 12 (τράβηγμα γωνίας χειλιών), 15 (πιεστής γωνίας χειλιών) και 43 (κλειστά μάτια) b) Εκφράσεις προσώπου που παράγονται ενεργοποιώντας ένα συνδυασμό από AUs που έχουν εξαχθεί από εικόνας με ετικέτες τον αντίστοιχο τύπο συναισθήματος: χαρά, λύπη, ξάφνιασμα, θυμός [30]

FACS

Η ανίχνευση προσώπου σε εικόνα βασίζεται στο σύστημα FACS. Το Facial Action Coding System (FACS) είναι ένα ολοκληρωμένο σύστημα βασισμένο σε ανατομικά χαρακτηριστικά, για την περιγραφή όλων των οπτικά διακριτών κινήσεων του προσώπου. Όλα τα

συναισθήματα του προσώπου εξετάζονται και περιγράφονται από τη σύσπαση των μυών του προσώπου που θεωρούνται ως Action Units (AUs). Το σύστημα αυτό περιλαμβάνει 46 βασικά AUs, καθένα από τα οποία κωδικοποιείται από έναν ή περισσότερους μυς του προσώπου, ενώ έχουν εισαχθεί ακόμα 8 AUs που αφορούν την κίνηση του κεφαλιού και 4 AUs που αφορούν την κίνηση των ματιών.

(ii) NNB προσεγγίσεις

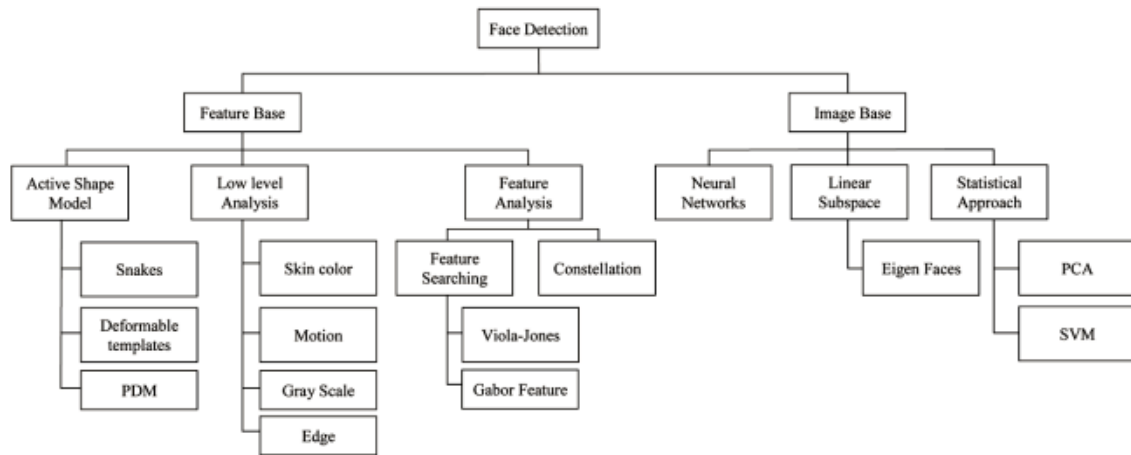
Οι προσεγγίσεις FER που βασίζονται στη βαθιά μάθηση μειώνουν σε μεγάλο βαθμό την εξάρτηση από μοντέλα που βασίζονται στα φυσικά χαρακτηριστικά του προσώπου και σε άλλες τεχνικές προεπεξεργασίας, επιτρέποντας έτσι την εκμάθηση απευθείας από τις εικόνες εισόδου. Αυτοί οι αλγόριθμοι βασισμένοι στη βαθιά μάθηση έχουν χρησιμοποιηθεί για εργασίες εξαγωγής, ταξινόμησης και αναγνώρισης χαρακτηριστικών. Ανάμεσα στα πολλά διαθέσιμα μοντέλα βαθιάς μάθησης, το συνελκτικό νευρωνικό δίκτυο (Convolutional Neural Network, CNN) είναι το πιο δημοφιλές μοντέλο δικτύου. Στις προσεγγίσεις που βασίζονται στο CNN, η εικόνα εισόδου περιπλέκεται μέσω μιας συλλογής φίλτρων στα επίπεδα συνέλιξης για να παραχθεί ένας χάρτης χαρακτηριστικών. Στη συνέχεια, κάθε χάρτης χαρακτηριστικών συνδυάζεται σε πλήρως συνδεδεμένα δίκτυα και η έκφραση του προσώπου αναγνωρίζεται ότι ανήκει σε μια συγκεκριμένη κατηγορία. Από την άλλη πλευρά, οι προσεγγίσεις FER που βασίζονται σε βαθιά μάθηση εξακολουθούν να έχουν ορισμένους περιορισμούς, συμπεριλαμβανομένης της ανάγκης για μεγάλης κλίμακας σύνολα δεδομένων, τεράστια υπολογιστική ισχύ και μεγάλες ποσότητες μνήμης, και είναι χρονοβόρες για τις φάσεις τόσο της εκπαίδευσης όσο και της δοκιμής. [6]

2.1.4 Διαδικασία FER

Στη συνέχεια επεξηγούμε περαιτέρω τη διαδικασία του FER αναλύοντας τα στάδια της ανίχνευσης προσώπου, εξαγωγής χαρακτηριστικών και ταξινόμησης χαρακτηριστικών.

2.1.4.1 Ανίχνευση προσώπου

Η ανίχνευση προσώπου είναι η ικανότητα του υπολογιστή να αναγνωρίζει ένα πρόσωπο μέσα σε μια εικόνα. Η ανίχνευση προσώπου μπορεί να διαχωριστεί σε δύο μεγάλες κατηγορίες: τις προσεγγίσεις που βασίζονται στα χαρακτηριστικά (feature-based) και τις προσεγγίσεις που βασίζονται στην εικόνα (image-based), όπως αναφέρεται εκτενώς στη μελέτη των Hasan et al. [31] Οι προσεγγίσεις που βασίζονται σε χαρακτηριστικά είναι ιδιαίτερα εφαρμόσιμες για ανίχνευση σε πραγματικό χρόνο, ενώ οι προσεγγίσεις που βασίζονται σε εικόνες αποδίδουν πολύ καλά με εικόνες σε κλίμακα του γκρι.



Εικόνα 5: Μέθοδοι αναγνώρισης προσώπου [32]

Από τις πιο γνωστές μεθόδους ανίχνευσης προσώπου, εντάσσεται στις feature-based προσεγγίσεις, είναι ο αλγόριθμος Viola-Jones. Ενώ τα τελευταία έτη ανάπτυξη εμφανίζουν και οι αλγόριθμοι νευρωνικών δικτύων (Neural Networks) που εντάσσονται στις image-based προσεγγίσεις, λόγω της πολύ καλής απόδοσης.

Viola-Jones Algorithm

Ο αλγόριθμος Viola Jones προτάθηκε από τους Paul Viola και Michael Jones το 2001 και ήταν το πρώτο πλαίσιο που έδωσε ανταγωνιστικά ποσοστά ανίχνευσης αντικειμένων. Μπορεί να χρησιμοποιηθεί για την ανίχνευση αντικειμένων σε πραγματικό χρόνο, αλλά εφαρμόζεται κυρίως για την ανίχνευση προσώπου. Ο αλγόριθμος περιλαμβάνει τέσσερα βασικά βήματα [7], [33], [34]:

- Επιλογή Haar χαρακτηριστικών

Τα περισσότερα ανθρώπινα πρόσωπα έχουν παρόμοια χαρακτηριστικά όπως:

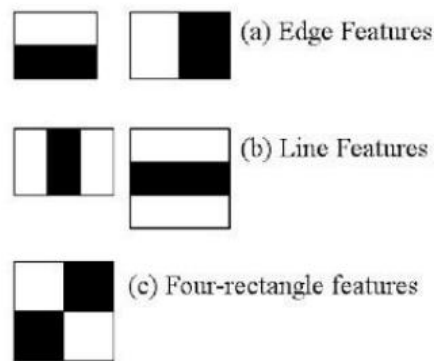
1. Η φωτεινή περιοχή της γέφυρας της μύτης είναι πιο ανοιχτόχρωμη σε σύγκριση με τη περιοχή του ματιού.
2. Η περιοχή των σκούρων ματιών είναι πιο σκούρα από τα άνω μάγουλα.
3. Ορισμένες περιοχές όπως η μύτη, το στόμα, τα μάτια.

Τα χαρακτηριστικά Haar λειτουργούν ως πυρήνες συνέλιξης. Μπορεί να είναι χαρακτηριστικά ακμών, χαρακτηριστικά γραμμών και χαρακτηριστικά τεσσάρων ορθογωνίων, δηλαδή ορθογώνια άσπρων και μαύρων εικονοστοιχείων. Κάθε χαρακτηριστικό εξάγει μια ενιαία τιμή η οποία λαμβάνεται από την αφαίρεση του αθροίσματος των λευκών εικονοστοιχείων με το άθροισμα των μαύρων εικονοστοιχείων.

- Ολοκλήρωμα εικόνας (Integral Image)

Η ολοκληρωμένη εικόνα, γνωστή και ως πίνακας αθροιστικής περιοχής, είναι ένας αλγόριθμος για τον γρήγορο και αποτελεσματικό υπολογισμό του αθροίσματος των τιμών σε

ένα ορθογώνιο υποσύνολο ενός πλέγματος. Οι Viola και Jones εφάρμοσαν την ολοκληρωμένη εικόνα για γρήγορο υπολογισμό των Haar χαρακτηριστικών.



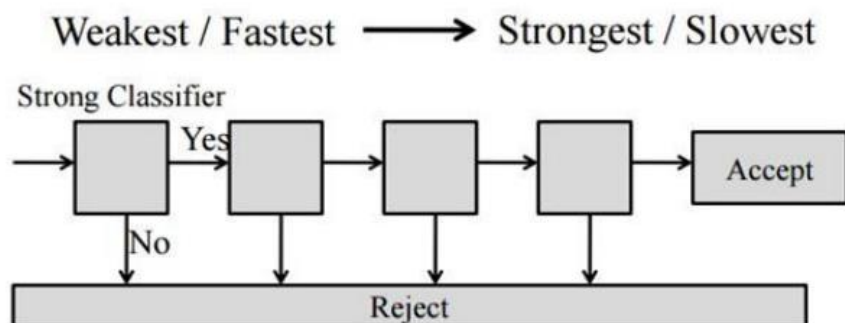
Εικόνα 6: Χαρακτηριστικά Haar. (a)Χαρακτηριστικά ακμών (b)Χαρακτηριστικά γραμμών (c)Χαρακτηριστικά τεσσάρων ορθογωνίων [7]

- Adaboost

Ο αλγόριθμος εκμάθησης Adaboost χρησιμοποιείται για τη δημιουργία του ταξινομητή που πρόκειται να εκπαιδευτεί. Αυτός ο αλγόριθμος βοηθά στην εύρεση μικρών κρίσιμων οπτικών χαρακτηριστικών μέσα από ένα μεγάλο σύνολο πιθανών χαρακτηριστικών. Με είσοδο το σύνολο των Haar χαρακτηριστικών που συλλέχθηκαν στο προηγούμενο βήμα, κατασκευάζεται ο τελικός ταξινομητής που θα χρησιμοποιηθεί στη συνέχεια.

- Χρήση Cascading Classifier

Ακολουθεί μια διαδικασία συνδυασμού των ταξινομητών που γρήγορα απορρίπτει τα παράθυρα που αφορούν το φόντο, έτσι ώστε να μπορούν να εκτελεστούν περισσότεροι υπολογισμοί σε περιοχές που μοιάζουν με πρόσωπο. Πιο αναλυτικά, σε μια εικόνα το μεγαλύτερο μέρος είναι περιοχές εκτός προσώπου. Επομένως, μπορούμε να ελέγξουμε εάν το παράθυρο βρίσκεται σε περιοχή του προσώπου και αν δεν είναι πρόσωπο τότε μπορεί να απορριφθεί. Η έννοια του Cascading Classifier δηλώνει ότι αντί να εφαρμόζονται 6000 χαρακτηριστικά σε ένα παράθυρο, θα πρέπει να γίνεται ομαδοποίηση των χαρακτηριστικών σε διαφορετικά στάδια. Εάν το παράθυρο αποτύχει στο αρχικό στάδιο απορρίπτεται, ενώ αν περάσει, εφαρμόζεται το δεύτερο στάδιο και η διαδικασία συνεχίζεται. Όταν ένα παράθυρο περάσει όλα τα στάδια τότε αφορά μια περιοχή προσώπου. Ο ανιχνευτής είχε 38 στάδια χαρακτηριστικών με 1, 10, 25 και 50 χαρακτηριστικά στα πρώτα πέντε στάδια.



Εικόνα 7: Cascading Classifiers [35]

Ο αλγόριθμος Viola-Jones έχει απλή αρχιτεκτονική, εμφανίζει υψηλή ακρίβεια ανίχνευσης προσώπου και γρήγορο χρόνο υπολογισμού άλλα δεν έχει καλά αποτελέσματα στα πλάγια πρόσωπα ή σε πρόσωπα που εν μέρη αποκρύπτονται.

Neural Networks

Οι αλγόριθμοι νευρωνικών δικτύων (NN) είναι εμπνευσμένοι από το βιολογικό νευρωνικό δίκτυο του ανθρώπινου εγκεφάλου. Τα νευρωνικά δίκτυα λαμβάνουν δεδομένα και εκπαιδεύονται να αναγνωρίζουν το μοτίβο (πχ. μοτίβο προσώπου). Στη συνέχεια, τα δίκτυα προβλέπουν την έξοδο για ένα νέο σύνολο παρόμοιων προσώπων. Τα νευρωνικά δίκτυα μπορούν να υποδιαιρεθούν σε τεχνητά νευρωνικά δίκτυα (Artificial Neural Networks, ANN), νευρωνικά δίκτυα που βασίζονται σε αποφάσεις (Decision-Based NN, DBNN) και ασαφή νευρωνικά δίκτυα (Fuzzy NN, FNN). [31]

Οι κύριες προκλήσεις που αντιμετωπίζουν όλοι οι αλγόριθμοι είναι η απόκρυψη μέρους του προσώπου, το σύνθετο φόντο, ο φωτισμός και ο προσανατολισμός. Οι περισσότεροι από τους αλγορίθμους για την ανίχνευση προσώπου λειτουργούν καλά υπό αυτές τις προκλήσεις και κυρίως οι αλγόριθμοι που πρόσφατα κέρδισαν δημοτικότητα, όπως οι NN. Κανένας συγκεκριμένος αλγόριθμος δεν είναι ο καλύτερος σε όλες τις περιπτώσεις, επομένως ο καλύτερος τρόπος για να επιλέξουμε αλγόριθμο είναι να γνωρίζουμε το πρόβλημα που πρέπει να αντιμετωπιστεί και τον αλγόριθμο που είναι ο καταλληλότερος για την επίλυση του κάθε προβλήματος.

2.1.4.2 Εξαγωγή χαρακτηριστικών

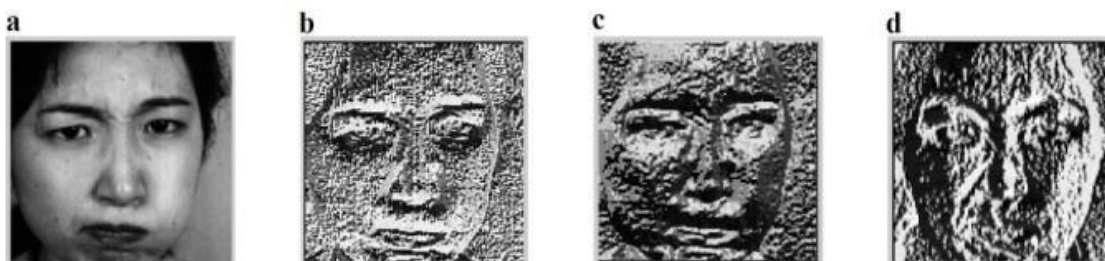
Η τρέχουσα αναγνώριση συναισθημάτων προσώπου βασίζεται κυρίως σε δισδιάστατα χαρακτηριστικά προσώπου χωρικής και χρονικής διάστασης. Γενικά, η εξαγωγή των χαρακτηριστικών του προσώπου μπορεί να χωριστεί σε δύο κατηγορίες: γεωμετρικά χαρακτηριστικά, που εντοπίζονται με τη μέθοδο παραμετροποίησης και χαρακτηριστικά εμφάνισης, που εντοπίζονται με τη μέθοδο αναγνώρισης.

Συνήθως, τα γεωμετρικά χαρακτηριστικά του προσώπου αντιπροσωπεύουν τις θέσεις και τα σχήματα των στοιχείων του προσώπου (μάτια, στόμα, κ.λπ.). Η μέθοδος παραμετροποίησης περιλαμβάνει τμηματοποίηση, εκχώρηση δυαδικών ετικετών σε κάθε εικονοστοιχείο και ανίχνευση όπου αποτυπώνεται ένα πλαίσιο όταν υπάρχει πρόσωπο στα δεδομένα. Το FACS είναι ένα παράδειγμα αυτής της μεθόδου. Και οι θέσεις και τα σχήματα συνήθως ορίζονται από ορόσημα του προσώπου. Από την άλλη, τα χαρακτηριστικά εμφάνισης αντιπροσωπεύουν την υφή του προσώπου, η οποία καταγράφει τις αλλαγές έντασης που σχετίζονται με διαφορετικές εκφράσεις, όπως ρυτίδες, εξογκώματα και αυλάκια. Για την μέθοδο αναγνώρισης λαμβάνονται υπόψη χαρακτηριστικά όπως εντάσεις, τιμές εικονοστοιχείων και ιστογράμματα.

Κατά τη γνώμη πολλών ερευνητών, η συγχώνευση των πληροφοριών από γεωμετρικά και χαρακτηριστικά εμφάνισης μπορεί να δώσει την καλύτερη απόδοση για ER. [2, 11]

Αξίζει να αναφερθεί ότι η εξαγωγή των χαρακτηριστικών μπορεί να δεχθεί και άλλες κατηγοριοποιήσεις όπως Action Unit (AU) και Non-AU based ή Local και Holistic [4].

Υπάρχουν διάφοροι μέθοδοι που μπορούν να αξιοποιηθούν για την εξαγωγή χαρακτηριστικών. Σε αυτές, τα χαρακτηριστικά εξάγονται θεωρώντας ολόκληρη την εικόνα ως ενιαία μονάδα και εφαρμόζοντας κάποιου είδους φίλτρα. Σύμφωνα με τη βιβλιογραφία, τα φίλτρα Gabor δείχνουν εξαιρετική απόδοση ανάλυσης προσώπου με μεγαλύτερο υπολογιστικό κόστος όσον αφορά το χρόνο και τη χρήση μνήμης. Μερικές από τις μεθόδους εξαγωγής χαρακτηριστικών είναι: Local Binary Pattern (LBP), Local Gradient Code (LGC), Local Directional Pattern (LDP), Histogram of Oriented Gradients (HOG) [4].



Εικόνα 8: a. Αρχική εικόνα, b. LBP εικόνα c. LDP εικόνα d. LGC εικόνα

2.1.4.3 Ταξινόμηση χαρακτηριστικών

Στο τελευταίο βήμα του FER γίνεται ταξινόμηση των χαρακτηριστικών στις αντίστοιχες κατηγορίες εκφράσεων προσώπου με βάση τις μεθόδους ταξινόμησης. Μερικές από τις πιο δημοφιλείς μεθόδους ταξινόμησης είναι η SVM (Support Vector Machine) και η NN (Nearest Neighbor) [4]. Ενώ πολλές μελέτες χρησιμοποιούν ακόμα ταξινομητές όπως CNN, Random Forest, Adaboost, MLP (MultiLayer Perceptron) κ.α. Στη συνέχεια αναλύονται μερικοί από τους παραπάνω ταξινομητές για να γίνει πιο κατανοητή η λειτουργία και η χρησιμότητά τους.

SVM

Οι μηχανές διανυσμάτων υποστήριξης ή ΜΔΥ (Support Vector Machines, SVM) προτάθηκαν από τον Vladimir Vapnik ως γραμμικοί ταξινομητές, το 1963. Απέκτησαν δημοσιότητα μετά το 1992, όταν ενισχύθηκαν από τον ίδιο και τους συνεργάτες του με το κόλπο του πυρήνα (kernel trick), που επέτρεψε τη χρήση τους και σε μη γραμμικώς διαχωρίσιμα προβλήματα. Θεωρώντας ένα πρόβλημα δυαδικής ταξινόμησης με θετικά και αρνητικά παραδείγματα, μια επιφάνεια που λειτουργεί ως σύνορο μεταξύ των κλάσεων χαρακτηρίζεται από το γεγονός ότι χωρίζει τα θετικά από τα αρνητικά παραδείγματα. Η μέθοδος SVM επιδιώκει να βρει το σύνορο που απέχει όσο το δυνατόν περισσότερο από τα παραδείγματα των κλάσεων που διαχωρίζει. Η υπερεπιφάνεια αυτή ονομάζεται υπερεπιφάνεια μεγίστου περιθωρίου (maximum margin hypersurface) και σε γραμμικώς διαχωρίσιμα προβλήματα ορίζεται από έναν πεπερασμένο αριθμό παραδειγμάτων του συνόλου εκπαίδευσης που ονομάζονται διανύσματα υποστήριξης (support vectors). [36]

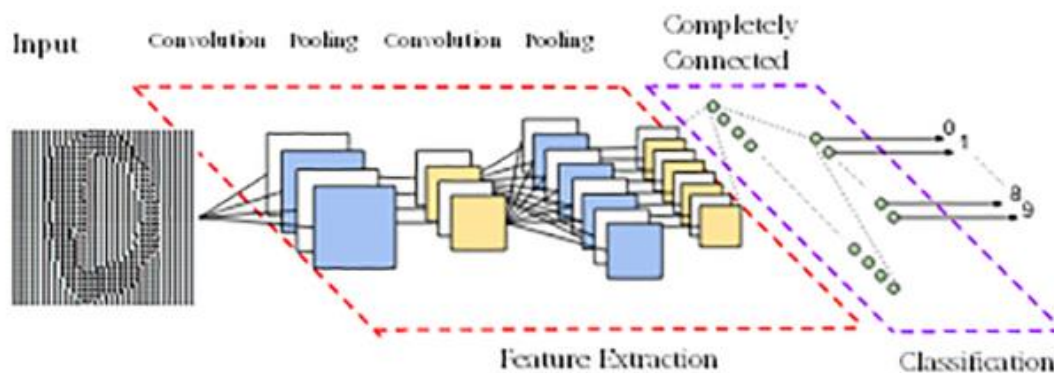


Fig. 11. Common CNN Architecture and its feature extraction convolutional pool.

Εικόνα 9: Βασική αρχιτεκτονική CNN. [29]

CNN

Ένα CNN αποτελείται από ένα ή περισσότερα επίπεδα συνέλιξης (convolutional layers) συχνά μαζί με ένα επίπεδο υποδειγματοληψίας (πχ pooling) ακολουθούμενο από ένα ή περισσότερα fully connected επίπεδα όπως συμβαίνει και σε ένα κλασικό πολύ-επίπεδο Νευρωνικό Δίκτυο. Ένα συνελκτικό επίπεδο (convolutional layer) είναι ουσιαστικά ένα σύνολο από νευρώνες που εκτελούν συνέλιξη των φίλτρων που έχουν προκαθοριστεί, με την εικόνα-διάγραμμα που δέχονται στην είσοδο. Τα επίπεδα συγκέντρωσης (pooling layers) στα CNNs, συνοψίζουν τις εξόδους γειτονικών γκρουπ νευρώνων εντός ενός παραθύρου (patch) με μια αντιπροσωπευτική τιμή, ενώ συνήθως τα γειτονικά παράθυρα δεν επικαλύπτονται. Πρόκειται ουσιαστικά για μια διαδικασία υπο-δειγματοληψίας των δεδομένων. Τέλος, στα fully connected επίπεδα κάθε νευρώνας συνδέεται με την έξοδο όλων των νευρώνων του προηγούμενου επιπέδου. Συγκριτικά με το κλασικό νευρωνικό δίκτυο, τα CNN είναι ευκολότερα στην εκπαίδευση και έχουν πολύ λιγότερες παραμέτρους, ενώ ακόμη είναι χρήσιμα σε περιπτώσεις με πολύ μεγάλη είσοδο. [37]

Το CNN καθιστά ευκολότερη και αποτελεσματική την ανάλυση των συναισθημάτων του προσώπου. Ως εργαλείο βαθιάς μάθησης, το CNN μαθαίνει τα μοτίβα στα δεδομένα. Μπορεί να είναι τα χρωματιστά, φωτεινά ή σκοτεινά σημεία. Ως αλγόριθμος πολλαπλών επιπέδων, το CNN αναγνωρίζει επίσης ορόσημα προσώπων. [38]

AdaBoost

Ο AdaBoost, συντομογραφία του Adaptive Boosting, ήταν ο πρώτος πραγματικά επιτυχημένος αλγόριθμος ενίσχυσης που αναπτύχθηκε με σκοπό τη δυαδική ταξινόμηση. Η ενίσχυση (boosting) είναι μια τεχνική μοντελοποίησης συνόλου που συνδυάζει πολλαπλούς αδύναμους ταξινομητές σε έναν μόνο ισχυρό ταξινομητή. Αυτό, γίνεται με την κατασκευή ενός μοντέλου με τη χρήση άλλων αδύναμων σε σειρά. Πρώτον, δημιουργείται ένα μοντέλο από τα δεδομένα εκπαίδευσης. Στη συνέχεια κατασκευάζεται ένα δεύτερο μοντέλο το οποίο προσπαθεί να διορθώσει τα σφάλματα που υπάρχουν στο πρώτο μοντέλο. Αυτή η διαδικασία συνεχίζεται και προστίθενται μοντέλα μέχρι να προβλεφθεί σωστά το πλήρες σύνολο δεδομένων εκπαίδευσης ή μέχρι να προστεθεί ο μέγιστος αριθμός μοντέλων. Ο AdaBoost

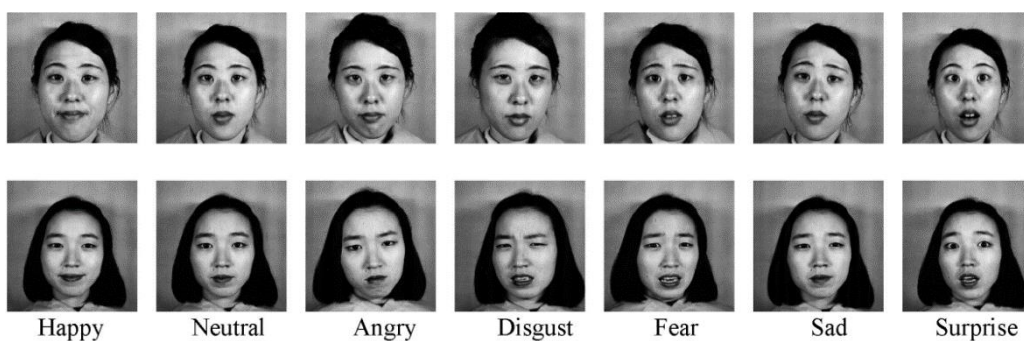
ονομάζεται προσαρμοστικός (adaptive) γιατί νέα βάρη αντιστοιχίζονται σε κάθε εμφάνιση, με υψηλότερα βάρη να εκχωρούνται σε εσφαλμένα ταξινομημένα στοιχεία. [39]

2.1.5 Σύνολα Δεδομένων για FER

Οι αλγόριθμοι αναγνώρισης συναισθημάτων προσώπου εξαρτώνται σε μεγάλο βαθμό από το σύνολο δεδομένων που χρησιμοποιείται. Τα σύνολα δεδομένων μπορεί να είναι στατικές εικόνες ή βίντεο ή συνδυασμός των δύο. Τα σύνολα δεδομένων μπορούν επίσης να ληφθούν σε ελεγχόμενο περιβάλλον, ενώ υπάρχουν και σύνολα που ελήφθησαν φυσικά υπό πραγματικές συνθήκες. Οι ακόλουθες υποενότητες περιγράφουν μερικά από τα τυπικά σύνολα δεδομένων. [38]

JAFPE

Στην Japanese Female Facial Expression (JaFFE) βάση δεδομένων, υπάρχουν 213 ασπρόμαυρες εικόνες που ταξινομούνται σε επτά εκφράσεις προσώπου. Το μέγεθος της κάθε εικόνας είναι 256 x 256 pixels. Τα δεδομένα απεικονίζουν 10 Γιαπωνέζες γυναίκες μοντέλα. Κάθε γυναίκα έχει δύο έως τέσσερα παραδείγματα για κάθε έκφραση. Κάθε εικόνα έχει αξιολογηθεί με 6 συναισθηματικά επίθετα από 60 Ιάπωνες. Η βάση δεδομένων σχεδιάστηκε και κατασκευάστηκε από τους Michael Lyons, Miyuki Kamachi και Jiros Gyoba. Τα συναισθήματα αποτελούνται, όπως φαίνεται και στην Εικ. 10, από χαρά, ουδέτερο, θυμό, αηδία, φόβο, λύπη και έκπληξη. [38], [40]



Εικόνα 10: Δείγματα από δυο εκφραστές που περιλαμβάνουν τα επτά διαφορετικά συναισθήματα στο σύνολο δεδομένων JAFPE.

Cohn-Kanade (CK) & Extended Cohn-Kanade (CK+)

Η Cohn-Kanade (CK) βάση δεδομένων περιλαμβάνει συνολικά 593 ακολουθίες εικόνων από 123 υποκείμενα και χωρίζεται σε δύο εκδόσεις. Η πρώτη έκδοση κυκλοφόρησε το 2000 και περιλαμβάνει 486 εικόνες από περίπου 97 διαφορετικούς ανθρώπους. Οι εικόνες λήφθηκαν σε ελεγχόμενο δωμάτιο όσον αφορά τα φώτα, το φόντο και τη θέση της κάμερας. Εκτός

αυτού, όλες οι εκφράσεις συλλέχθηκαν ζητώντας από τα υποκείμενα να τις εκτελέσουν, επομένως, μπορεί να διαφέρουν από τις αυθόρμητες εκφράσεις. Το σύνολο δεδομένων CK παρέχει δισδιάστατες εικόνες με ανάλυση 640x490 pixels ή 640x480 pixels και διάταξη εικονοστοιχείων 8-bit ασπρόμαυρα ή 24-bit έγχρωμα.

Η δεύτερη και τρέχουσα έκδοση της βάσης δεδομένων ονομάστηκε The Extended Cohn-Kanade Dataset (CK+) και κυκλοφόρησε το 2010. Για αυτήν την έκδοση, συμπεριλήφθηκαν 107 εικόνες επίδειξης συναισθημάτων από 26 ανθρώπους. Το 31% των εικόνων απεικονίζουν άνδρες και το 69% γυναίκες. Αυτή τη φορά τραβήχτηκαν αυθόρμητες εικόνες και κάθε εικόνα έλαβε μια ονομαστική ετικέτα συναισθήματος με βάση το FACS για καθεμία από τις 7 βασικές κατηγορίες συναισθημάτων. [29], [38]

MMI

Το MMI αποτελείται από περισσότερες από 2900 ακολουθίες βίντεο και ακίνητες εικόνες υψηλής ανάλυσης 75 υποκειμένων από διαφορετικές εθνότητες (Ευρώπης, Ασίας και Νότιας Αμερικής), με 44% γυναίκες και 56% άνδρες. Επισημαίνεται ευρέως για την παρουσία AU σε βίντεο και για την κωδικοποίηση σε επίπεδο καρέ (frame), αφού για κάθε καρέ υποδεικνύεται σε ποια φάση βρίσκεται η AU (ουδέτερη, έναρξης, κορυφής ή μετατόπισης). Οι συμμετέχοντες ζητήθηκαν να επιδείξουν όλα τα AUs και τα έξι βασικά συναισθήματα είτε σε ελεγχόμενο περιβάλλον είτε αυθόρμητα. Το αρχικό μέγεθος κάθε εικόνας προσώπου είναι 720 pixel × 576 pixel.[41], [42]

FER2013

Ένα από τα πιο ευρέως χρησιμοποιούμενα δεδομένα ονομάζεται FER που σημαίνει Facial Emotion Recognition (Αναγνώριση Συναισθημάτων Προσώπου). Αποτελείται από 35.887 εικόνες επτά διαφορετικών συναισθημάτων. Το FER2013 έχει 28.709 φωτογραφίες για εκπαίδευση (training), 3589 για επικύρωση και 3589 για δοκιμή (testing). Σε αυτό το σύνολο δεδομένων, τρεις στήλες ορίζουν κάθε φωτογραφία. Αυτές οι στήλες είναι πρώτον, ο τύπος συναισθήματος σε αριθμητική μορφή 0-6 που χαρακτηρίζεται αντίστοιχα ως θυμός, αηδία, φόβος, ευτυχία, λύπη, έκπληξη και ουδέτερος. Η δεύτερη στήλη είναι ένας πίνακας αριθμητικών τιμών που αντιπροσωπεύει τις φωτογραφίες. Η τελευταία στήλη υποδεικνύει τον τύπο της φωτογραφίας είτε πρόκειται για δεδομένα εκπαίδευσης είτε για δεδομένα δοκιμών. [38]



Εικόνα 11: Δείγματα εικόνων για τα βασικά συναισθήματα από το FER2013. [43]

BU-3DFE

Το Binghamton University 3D Facial Expression (BU-3DFE) σχεδιάστηκε για να καλύψει την ανάγκη για ένα σύνολο δεδομένων με 3D εικόνες προσώπου ταξινομημένες ανά συναίσθημα. Συνολικά 100 συμμετέχοντες υποβλήθηκαν σε σάρωση των εκφράσεών τους, 60% γυναίκες και 40% άνδρες, καταλήγοντας με συνολικά 2500 τρισδιάστατα μοντέλα. Όλοι οι συμμετέχοντες ήταν μέρος των τμημάτων ψυχολογίας, τεχνών και μηχανικής. Ζητήθηκε από αυτούς να δείξουν τα 7 συναισθήματα με 4 βαθμούς έντασης (χαμηλή, μέση, υψηλή και υψηλότερη). Το αρχικό μέγεθος κάθε εικόνας προσώπου είναι 1040 pixel × 1329 pixel. Πάνω σε αυτό το σύνολο δεδομένων, το BU-3DFE, υπάρχουν μεταγενέστερες μελέτες όπως το BU-4DFE στις οποίες προστέθηκε η διάσταση χρόνου, το BP4D-Spontaneous και το BP4D +. [6], [29]

RAF-DB

Η RAF-DB είναι μια δημόσια διαθέσιμη βάση δεδομένων εκφράσεων προσώπου πραγματικού κόσμου που δημιουργήθηκε μέσω μιας αξιόπιστης προσέγγισης crowdsourcing. Περιέχει περίπου 30000 εικόνες προσώπων από χιλιάδες άτομα. Κάθε εικόνα προσώπου έχει επισημανθεί περίπου 40 φορές από ξεχωριστά άτομα, μέσα από ένα σύνολο από 315 εκπαιδευμένους σχολιαστές (φοιτητές και στελέχη από πανεπιστήμια), με μια από τις 7 βασικές κατηγορίες συναισθημάτων. Στη συνέχεια, οι αναξιόπιστες ετικέτες φιλτράρονται από έναν αλγόριθμο αξιολόγησης αξιοπιστίας, μέσω του οποίου κάθε εικόνα μπορεί να αναπαρασταθεί από ένα διάνυσμα πιθανότητας συναισθημάτων 7 διαστάσεων. Αναλύοντας 1,2 εκατομμύρια ετικέτες 29672 εικόνων προσώπων μεγάλης ποικιλίας που έχουν ληφθεί από το διαδίκτυο, αυτά τα Συναισθηματικά Πρόσωπα του Πραγματικού Κόσμου (Real-world Affective Faces, RAF) κατηγοριοποιήθηκαν σε δύο τύπους: βασικές εκφράσεις με μονοτροπική κατανομή και σύνθετα συναισθήματα με διτροπική κατανομή. [44]

Aff-Wild

Η βάση δεδομένων Aff-Wild αποτελείται από 298 βίντεο, συνολικής διάρκειας άνω των 30 ωρών, που εμφανίζουν αντιδράσεις 200 ατόμων (εκ των οποίων 130 άνδρες και 70 γυναίκες). Ολόκληρη η βάση δεδομένων περιέχει περισσότερα από 1.180.000 καρέ (περίπου 900K για εκπαίδευση και 300K για δοκιμή). Στόχος ήταν η συλλογή αυθόρμητων συμπεριφορών προσώπου κάτω από αυθαίρετες συνθήκες καταγραφής. Για το σκοπό αυτό, έγινε συλλογή βίντεο χρησιμοποιώντας τον ιστότοπο κοινής χρήσης βίντεο στο Youtube. Όλα τα βίντεο έχουν σχολιαστεί ως προς δύο διαστάσεις συνεχούς συναισθήματος, το σθένος και τη διέγερση, από 6-8 απλούς ειδικούς. Χρησιμοποιήθηκε μια διαδικασία on-line σχολιασμού, σύμφωνα με την οποία οι σχολιαστές παρακολουθούσαν κάθε βίντεο και παρείχαν τους σχολιασμούς τους μέσω ενός joystick. Το σθένος και η διέγερση κυμαίνονταν συνεχώς σε [-1, +1]. Στην Εικ. 12 εμφανίζονται μερικά καρέ από τη βάση δεδομένων Aff-Wild. [45], [46]



Εικόνα 12 : Καρέ από τη βάση δεδομένων Aff-Wild που δείχνουν υποκείμενα σε διαφορετικές συναισθηματικές καταστάσεις, διαφορετικών εθνοτήτων, σε ποικίλες στάσεις κεφαλιού και συνθήκες φωτισμού. [46]

AffectNet

Το AffectNet είναι μια βάση δεδομένων εκφράσεων προσώπου που δημιουργήθηκε από το Διαδίκτυο μέσω ερωτημάτων σε διαφορετικές μηχανές αναζήτησης (Google, Bing και Yahoo) χρησιμοποιώντας 1250 ετικέτες που σχετίζονται με συναισθήματα σε έξι διαφορετικές γλώσσες. Περιέχει περισσότερες από ένα εκατομμύριο εικόνες με πρόσωπα και εξαγόμενα σημεία ορόσημων προσώπου. Δώδεκα ειδικοί στον άνθρωπο σχολίασαν με μη αυτόματο τρόπο 450.000 από αυτές τις εικόνες τόσο με διακριτά μοντέλα (με οκτώ βασικές ετικέτες έκφρασης, δηλ. τις 7 βασικές και την περιφρόνηση) όσο και με μοντέλα διαστάσεων (σθένος και διέγερση) και πρόσθεσαν ετικέτες στις εικόνες που έχουν οποιαδήποτε μερική απόκρυψη του προσώπου. Το AffectNet είναι η μεγαλύτερη βάση δεδομένων εκφράσεων προσώπου σε στατικές εικόνες που παρέχει ετικέτες τόσο για διακριτά μοντέλα συναισθημάτων όσο και για συνεχή μοντέλα (σθένους-διέγερσης). [47]

EmotioNet

Η βάση δεδομένων EmotioNet περιλαμβάνει ετικέτες για τα AUs, την ένταση των AUs και τις βασικές και σύνθετες κατηγορίες συναισθημάτων, σε ένα μεγάλο σύνολο (ένα εκατομμύριο) εικόνων εκφράσεων προσώπου 'in the wild' (δηλ. εικόνες από το διαδίκτυο). Για τη δημιουργία του EmotioNet συλλέχθηκαν 457 έννοιες, που προέρχονται από τη λέξη "αίσθημα" από το WordNet και χρησιμοποιήθηκαν για την αναζήτηση εικόνων προσώπων σε διάφορες δημοφιλείς μηχανές αναζήτησης, ως λέξεις-κλειδιά. Στη συνέχεια, ένας ανιχνευτής προσώπου χρησιμοποιήθηκε για την εξάλειψη των εικόνων που δεν περιείχαν πρόσωπο και ακολούθησε η επισήμανση των εικόνων με ετικέτες. Για το σκοπό αυτό χρησιμοποιήθηκαν τρεις διαθέσιμες βάσεις δεδομένων με χειροκίνητα σχολιασμένα AUs και εντάσεις AUs για την εκπαίδευση ταξινομητών. Έτσι κάθε εικόνα εντάσσεται σε μία από τις 23 βασικές ή σύνθετες κατηγορίες συναισθημάτων που δημιουργούνται με την ενεργοποίηση ενός μοναδικού μοτίβου από AUs. [48]

2.1.6 Ανάλυση Δημοσιεύσεων για FER

Στο σημείο αυτό θα παρουσιάσουμε κάποιες πρόσφατες μελέτες που έχουν γίνει στον τομέα της αναγνώρισης συναισθήματος από εκφράσεις του προσώπου.

Οι Khairuddin και Chen [49] πέτυχαν την υψηλότερη ακρίβεια ταξινόμησης ενός δικτύου στο σύνολο δεδομένων FER2013. Στο έργο τους υιοθέτησαν την αρχιτεκτονική VGGNet, προσαρμόζοντας αυστηρά τις υπερπαραμέτρους της και πειραματίστηκαν με διάφορες μεθόδους βελτιστοποίησης. Την περίοδο που έγινε η έρευνα, το μοντέλο τους πέτυχε 73,28%, την υψηλότερη ακρίβεια ενός δικτύου στο FER2013 που δεν χρησιμοποιεί επιπλέον δεδομένα εκπαίδευσης.

Στο άρθρο των Alreshidi και Ullah [50], προτάθηκε ένα μοντέλο για την αναγνώριση των συναισθημάτων του ανθρώπινου προσώπου. Το μοντέλο αυτό αποτελείται από δύο αλγόριθμους μηχανικής μάθησης (για ανίχνευση και ταξινόμηση) που θα μπορούσαν να εκπαιδευτούν εκτός σύνδεσης για εφαρμογές σε πραγματικό χρόνο. Αρχικά, ανιχνεύονται πρόσωπα στις εικόνες μέσω του ταξινομητή καταρράκτη AdaBoost. Στη συνέχεια εξάγονται τα Neighborhood Difference Features (NDF), τα οποία αντιπροσωπεύουν τα χαρακτηριστικά ενός προσώπου με βάση τοπικές πληροφορίες εμφάνισης. Το άρθρο αυτό επικεντρώνεται στις επτά εκφράσεις του προσώπου που χρησιμοποιούνται πιο πολύ. Ωστόσο, λόγω του σχεδιασμού του μοντέλου, μπορεί να επεκταθεί ώστε να ταξινομήσει έναν N αριθμό εκφράσεων του προσώπου. Για την ταξινόμηση των εκφράσεων του προσώπου, εκπαιδεύτηκε ένας Random Forest ταξινομητής. Η προτεινόμενη μέθοδος αξιολογήθηκε σε διαφορετικά σύνολα δεδομένων και συγκρίθηκε με πέντε μεθόδους με αποτέλεσμα να δώσει καλύτερα αποτελέσματα από τις μεθόδους αναφοράς.

Μια μέθοδος που χρησιμοποιεί υβριδικά γεωμετρικά χαρακτηριστικά και χαρακτηριστικά εμφάνισης προτείνεται για την αναγνώριση της έκφρασης του προσώπου στην εργασία των Chen et al. [51]. Τα υβριδικά χαρακτηριστικά περιλαμβάνουν τις μετατοπίσεις των σημείων των χαρακτηριστικών του προσώπου και τις τοπικές διαφορές υψής μεταξύ των εικόνων ουδέτερης και εκφραστικής (χαρά, λύπη, θυμός κλπ) έκφρασης του προσώπου. Η προτεινόμενη μέθοδος πέτυχε μέση ακρίβεια 95% στην εκτεταμένη βάση δεδομένων Cohn-Kanade (CK+) με χρήση της μεθόδου ταξινόμησης SVM.

Οι Yuan et al. [52] κατασκεύασαν έναν αλγόριθμο αναγνώρισης συναισθημάτων των μαθητών για χρήση σε αίθουσες διδασκαλίας, ταξινομώντας τα οπτικά συναισθήματα, για τη βελτίωση της ποιότητας της διδασκαλίας στην τάξη. Για το σκοπό αυτό, συλλέχθηκαν βίντεο των εκφράσεων του προσώπου των μαθητών της τάξης. Για την ανίχνευση προσώπου και την τμηματοποίηση εικόνων χρησιμοποιήθηκε ένα συνελκτικό νευρωνικό δίκτυο πολλαπλών εργασιών (Multi-task CNN, MTCNN) και τα βίντεο με την καλύτερη μορφολογία χαρακτηριστικών επιλέχθηκαν για τη δημιουργία μιας τυπικής βάσης δεδομένων. Για να επικυρωθεί η αποτελεσματικότητα του σχεδιασμένου μοντέλου MTCNN, δύο κύρια δίκτυα ταξινόμησης, το VGG16 και το ResNet18, δοκιμάστηκαν και συγκρίθηκαν με το MTCNN και η τελική ακρίβεια μετά την εκπαίδευση ήταν 78,26% και 75,03% για το ResNet18 και το VGG16, αντίστοιχα, ενώ το μοντέλο MTCNN είχε ακρίβεια 93,53%.

Το μοντέλο που προτάθηκε από τους Depnath et al. ,το "ConvNet", ανιχνεύει τα επτά βασικά συναισθήματα από δεδομένα εικόνας. Για την εκπαίδευση του μοντέλου CNN, χρησιμοποιήθηκε η βάση δεδομένων FER2013 και δοκιμάστηκε η επιτυχία του συστήματος εντοπίζοντας τις εκφράσεις του προσώπου σε πραγματικό χρόνο. Το ConvNet αποτελείται από τέσσερα επίπεδα συνέλιξης μαζί με δύο πλήρως συνδεδεμένα επίπεδα. Τα πειραματικά αποτελέσματα έδειξαν ότι το ConvNet είναι σε θέση να επιτύχει 96% ακρίβεια εκπαίδευσης, πολύ καλύτερη από άλλα μοντέλα. Το ConvNet πέτυχε επίσης ακρίβεια επικύρωσης από 65% έως 70% (λαμβάνοντας υπόψη διαφορετικά σύνολα δεδομένων που χρησιμοποιούνται για πειράματα), με αποτέλεσμα υψηλότερη ακρίβεια ταξινόμησης σε σύγκριση με άλλα υπάρχοντα μοντέλα. [43]

Στην εργασία των Jung et al. [53], υιοθετείται μια τεχνική βαθιάς μάθησης, ως εργαλείο αυτόματης εξαγωγής χρήσιμων χαρακτηριστικών από ακατέργαστα δεδομένα. Το βαθύ δίκτυο βασίζεται σε δύο διαφορετικά μοντέλα. Το πρώτο deep network εξάγει χαρακτηριστικά χρονικής εμφάνισης από ακολουθίες εικόνων, ενώ το άλλο εξάγει χαρακτηριστικά χρονικής γεωμετρίας από χρονικά ορόσημα του προσώπου. Αυτά τα δύο μοντέλα συνδυάζονται προκειμένου να ενισχυθεί η απόδοση της αναγνώρισης εκφράσεων προσώπου. Ως αποτέλεσμα, επιτυγχάνεται απόδοση καλύτερη από άλλες μεθόδους αιχμής στις βάσεις δεδομένων CK+ και Oulu-CASIA.

Μερικοί ερευνητές προσπάθησαν να αναγνωρίσουν τα συναισθήματα του προσώπου χρησιμοποιώντας υπέρυθρες εικόνες αντί για εικόνα ορατού φάσματος φωτός (Visible Light Spectrum, VIS), επειδή η εικόνα ορατού φωτός (VIS) μεταβάλλεται ανάλογα με την κατάσταση του φωτισμού. [6] Στην εργασία των Shen et al. [54], προτείνεται μια μέθοδος αναγνώρισης εκφράσεων προσώπου με χρήση υπέρυθρων θερμικών βίντεο. Αρχικά, τα χαρακτηριστικά της ακολουθίας εξάγονται από τις υπέρυθρες θερμικές οριζόντιες και κάθετες αλληλουχίες διαφοράς θερμοκρασίας διαφορετικών υποπεριοχών του προσώπου. Στη συνέχεια, ένα υποσύνολο χαρακτηριστικών επιλέγεται σύμφωνα με τις F-values. Τέλος, ο αλγόριθμος Adaboost, με τους ασθενείς ταξινομητές του k-Nearest Neighbor (kNN), χρησιμοποιείται για την ταξινόμηση των εκφράσεων του προσώπου στις διαστάσεις διέγερσης (arousal) και σθένους (valence).

Στην εργασία των Maalej et al. [55] εξετάζεται το πρόβλημα της αναγνώρισης τρισδιάστατων εκφράσεων προσώπου. Προτείνεται μια τοπική ανάλυση γεωμετρικών σχημάτων των επιφανειών του προσώπου σε συνδυασμό με τεχνικές μηχανικής μάθησης για ταξινόμηση εκφράσεων. Τα πειραματικά αποτελέσματα καταδεικνύουν την αποτελεσματικότητα της προτεινόμενης προσέγγισης. Χρησιμοποιώντας τους ταξινομητές Multi-boosting and Support Vector Machines (SVM), επιτεύχθηκαν αντίστοιχα μέσοι ρυθμοί αναγνώρισης 98,81% και 97,75%, για την αναγνώριση των έξι πρωτότυπων εκφράσεων προσώπου στη βάση δεδομένων BU3DFE.

Οι Sujono και Gunawan [56] χρησιμοποίησαν τον αισθητήρα κίνησης Kinect που μπορεί να παρακολουθεί το πρόσωπο που έχει εντοπιστεί χρησιμοποιώντας το Active Appearance Model (AAM). Το AAM είναι η μέθοδος που προσαρμόζει το μοντέλο σχήματος και υφής σε ένα νέο πρόσωπο, όταν υπάρχει διακύμανση σχήματος και υφής σε σύγκριση με το αποτέλεσμα της προπόνησης. Ο στόχος αυτής της έρευνας είναι να ανιχνεύσει την έκφραση

του προσώπου παρατηρώντας την αλλαγή των βασικών χαρακτηριστικών στο AAM χρησιμοποιώντας τη Fuzzy Logic. Το Fuzzy Logic χρησιμοποιείται για τον προσδιορισμό των τρεχόντων συναισθημάτων με βάση προηγούμενες γνώσεις που προέρχονται από το Facial Action Coding System (FACS).

Οι Kollias et al. [57] στην εργασία τους, χτίζουν το πρώτο ολιστικό πλαίσιο για ανάλυση προσώπου μεγάλης κλίμακας, το FaceBehaviorNet, με μελέτες περιπτώσεων στον συναισθηματικό υπολογισμό και στην αναγνώριση προσώπου. Στην πρώτη μελέτη περίπτωσης, το FaceBehaviorNet εκπαιδεύεται για κοινή αναγνώριση βασικής έκφρασης, ανίχνευση μονάδων δράσης και εκτίμηση διέγερσης σθένους. Χρησιμοποιείται η βάση δεδομένων Aff-Wild2, η οποία είναι η πρώτη in-the-wild βάση δεδομένων που περιέχει σχολιασμούς και για τις τρεις εργασίες συμπεριφοράς για να συμπεραθεί η συνάφεια εργασιών. Αρχικά εκπαιδεύουν ένα δίκτυο για ανίχνευση AU στην ένωση των βάσεων δεδομένων Aff-Wild2 και GFT και χρησιμοποιούν αυτό το δίκτυο για τον αυτόματο σχολιασμό του Aff-Wild2 με AU. Στην δεύτερη μελέτη περίπτωσης, το FaceBehaviorNet εκπαιδεύεται για την ανίχνευση των κοινών χαρακτηριστικών προσώπου και την αναγνώριση προσώπου. Χρησιμοποιείται η βάση δεδομένων CelebA, η οποία περιέχει επικαλυπτόμενους σχολιασμούς για χαρακτηριστικά και ταυτότητες προσώπου και υπολογίζεται η κατανομή των χαρακτηριστικών για κάθε ταυτότητα. Εκτελείται μια εκτεταμένη πειραματική μελέτη - σε 10 βάσεις δεδομένων- που συγκρίνει την απόδοση του ολιστικού (πολλαπλών εργασιών, πολλών τομέων, πολλαπλών ετικετών) FaceBehaviorNet (εκπαιδευμένο με και χωρίς να λαμβάνει υπόψη τη συσχέτιση εργασιών) με την απόδοση των δικτύων που εκτελούν μοναδική εργασία, καθώς και με την απόδοση των καλύτερων δικτύων. Το FaceBehaviorNet ξεπέρασε, με μεγάλο περιθώριο, όλα τα δίκτυα σε όλες τις βάσεις δεδομένων, και στις δύο μελέτες περιπτώσεων.

Στο μελέτη τους, οι Kollias et al. [46], παρουσιάζουν το Aff-Wild, μια νέα, μεγάλη in-the-wild βάση δεδομένων που αποτελείται από 298 βίντεο με 200 υποκείμενα, συνολικής διάρκειας άνω των 30 ωρών. Αναφέρουν τα αποτελέσματα του Aff-Wild Challenge, καθώς και τις παγίδες και τις προκλήσεις όσον αφορά την πρόβλεψη in-the-wild του σθένους και της διέγερσης. Επιπλέον, σχεδιάζουν και εκπαιδεύουν εκτενώς μια αρχιτεκτονική βαθύ νευρωνικού δικτύου που εκτελεί πρόβλεψη συνεχών διαστάσεων συναισθημάτων με βάση οπτικές ενδείξεις. Η προτεινόμενη αρχιτεκτονική βαθιάς μάθησης, AffWildNet, περιλαμβάνει συνελκτικά και επαναλαμβανόμενα επίπεδα νευρωνικών δικτύων (CNN-RNN layers), που εκμεταλλεύονται τις αμετάβλητες ιδιότητες των συνελκτικών χαρακτηριστικών, ενώ παράλληλα μοντελοποιούν χρονικές δυναμικές που προκύπτουν στην ανθρώπινη συμπεριφορά μέσω των επαναλαμβανόμενων επιπέδων. Το AffWildNet έδωσε εξαιρετικά αποτελέσματα στο Aff-Wild Challenge. Επίσης, στη μελέτη αυτή παρατηρείται ότι το AffWildNet μπορεί να παράγει επιδόσεις αιχμής, όχι μόνο για συνεχής διαστάσεις, αλλά και για διακριτή κατηγοριοποίηση των συναισθημάτων. Για να επιτευχθεί αυτό, το AffWildNet εκπαιδεύτηκε αποτελεσματικά με τη μεγαλύτερη υπάρχουσα, in-the-wild, βάση δεδομένων για συνεχή αναγνώριση σθένους-διέγερσης (regression analysis problem) και στη συνέχεια χρησιμοποιήθηκε για την αντιμετώπιση του προβλήματος της αναγνώρισης (ταξινόμηση) των επτά βασικών συναισθημάτων.

2.1.7 Συναισθήματα και Μουσική

Η μουσική είναι ένα μεγάλο κομμάτι της ζωής των ανθρώπων. Πολλοί είναι αυτοί που ακούνε μουσική σε καθημερινή βάση ενώ ακόμη και αν κάποιος δεν ακούει ο ίδιος μουσική, συντροφεύεται από αυτήν στο μεγαλύτερο μέρος της ζωής του. Για παράδειγμα μέσα από εκπομπές ή διαφημίσεις στη τηλεόραση ή στο ραδιόφωνο, μέσα από ταινίες ή σειρές που παρακολουθεί, ή ακόμα και στο σουπερμάρκετ, σε εμπορικά καταστήματα και σε καφετέριες. Η μουσική είναι τόσο ένα μέσο αυτοέκφρασης όσο και ένα επικοινωνιακό μέσο. Είναι επίσης ένα πολύ ισχυρό εργαλείο, κυρίως λόγω της δυνατότητας που έχει να προκαλεί έντονα συναισθήματα.

Πολλές έρευνες έχουν γίνει πάνω στη μουσική. Έρευνες στην ψυχολογία τόνισαν ότι η μουσική όχι μόνο βελτιώνει τη διάθεση, αυξάνει την ενεργοποίηση, διεγείρει οπτικές και ακουστικές εικόνες, αλλά επίσης ανακαλεί σχετικές ταινίες ή μουσικά βίντεο και ανακουφίζει από το άγχος. Από εμπειρικά πειράματα στο χώρο του αθλητισμού, όπου προτιμάται συνήθως η πιο έντονη, ενεργητική και ρυθμική μουσική, φαίνεται ότι η ακρόαση μουσικών κομματιών βοηθά στην βελτίωση της απόδοσης, επιβραδύνει το αίσθημα κόπωσης και αποσυνδέει από δυσάρεστα συναισθήματα. Γενικότερα, η γρήγορη, μουσική έντονου ρυθμού παράγει διεγερτικά αποτελέσματα ενώ η αργή, απαλή μουσική παράγει ηρεμιστικά αποτελέσματα. Σημαντικό είναι επίσης, με βάση τους ψυχολόγους, το γεγονός ότι η προτίμηση των χρηστών στη μουσική συνδέεται με την προσωπικότητά τους. Όλα τα παραπάνω υπογραμμίζουν την χρησιμότητα των συστημάτων συστάσεων μουσικής που θα αναλυθούν στο επόμενο κεφάλαιο. [58]

Κάθε μουσικό κομμάτι περιγράφεται από ορισμένα ακουστικά χαρακτηριστικά. Διάφορες μελέτες έχουν προσπαθήσει να κατηγοριοποιήσουν τα ακουστικά χαρακτηριστικά με βάση τα συναισθήματα ερευνώντας τη σχέση μεταξύ τους [59], [60] αλλά παρόλα αυτά δεν έχει δοθεί ακόμα κανένας ακριβής ορισμός που να ξεκαθαρίζει την αντιστοίχιση αυτών των δύο. Η σχέση μουσικών χαρακτηριστικών και συναισθημάτων αναλύεται από μια ερευνητική περιοχή που ονομάζεται αναγνώριση συναισθημάτων (MER) μουσικής που με τη σειρά της αποτελεί μέρος του ερευνητικού τομέα της ανάκτησης πληροφοριών μουσικής (MIR).

Η ανάκτηση πληροφοριών μουσικής (Music Information Retrieval, MIR) είναι ένας αναδυόμενος ερευνητικός τομέας που αντιμετωπίζει το πρόβλημα της αναζήτησης και της ανάκτησης ορισμένων τύπων μουσικής από ένα μεγάλο σύνολο δεδομένων μουσικής. Η ταξινόμηση είναι ένα θεμελιώδες πρόβλημα στο MIR. Πολλές εργασίες στο MIR μπορούν φυσικά να μετατραπούν σε προβλήματα ταξινόμησης, όπως ταξινόμηση είδους, ταξινόμηση διάθεσης, αναγνώριση καλλιτέχνη, αναγνώριση μουσικών οργάνων κ.λπ. Ο σχολιασμός μουσικής (music annotation), μια νέα ερευνητική περιοχή στο MIR που έχει προσελκύσει μεγάλη προσοχή τα τελευταία χρόνια, είναι επίσης με μια γενική έννοια, ένα πρόβλημα ταξινόμησης. [61]

Μελέτες ψυχολογίας στη μουσική έχουν οδηγήσει σε μια σειρά από θεμελιώδεις γενικεύσεις που μπορούν να ωφελήσουν την έρευνα MIR. Πρώτον, δεν είναι εξίσου πιθανό να διεγείρονται όλες οι διαθέσεις ακούγοντας μουσική, δηλαδή η χαρά και η λύπη είναι τα πιο

συχνά συναισθήματα που εκφράζονται από τη μουσική αλλά η αηδία είναι πολύ σπάνιο. Δεύτερον, υπάρχουν ομοιόμορφες επιδράσεις στη διάθεση μεταξύ διαφορετικών ανθρώπων, οπότε μια πλειοψηφία των ανθρώπων θα δηλώσει ότι ένα συγκεκριμένο συναίσθημα εκφράζεται μέσα από ένα μουσικό κομμάτι. Τρίτον, δεν έχουν όλοι οι τύποι διαθέσεων το ίδιο επίπεδο συμφωνίας μεταξύ των ακροατών. Οπότε ορισμένες κατηγορίες διάθεσης θα ήταν πιο δύσκολο να ταξινομηθούν σε σχέση με άλλες. Τέλος, υπάρχει κάποια αντιστοιχία μεταξύ των κρίσεων των ακροατών σχετικά με τη διάθεση και τις μουσικές παραμέτρους όπως ο ρυθμός, η δυναμική, ο ρυθμός, η χροιά, η άρθρωση, ο τόνος, ο τρόπος, οι επιθέσεις τόνου και η αρμονία. Αυτή η αντιστοιχία έχει πολύ σημαντικές συνέπειες για το σχεδιασμό και την ανάπτυξη αλγορίθμων ταξινόμησης της μουσικής διάθεσης. [62]

Η ερευνητική περιοχή που διερευνά υπολογιστικά μοντέλα για την ανίχνευση των συναισθημάτων των τραγουδιών είναι γνωστή ως αναγνώριση συναισθημάτων μουσικής (Music Emotion Recognition, MER). Αποτελεί μια διαδικασία που περιλαμβάνει τη χρήση υπολογιστών για την εξαγωγή και ανάλυση μουσικών χαρακτηριστικών, τη διαμόρφωση των σχέσεων συσχέτισης των μουσικών χαρακτηριστικών και του χώρου συναισθημάτων και την αναγνώριση του συναισθήματος που εκφράζει η μουσική. Σκοπός είναι οι μουσικές βάσεις δεδομένων να μπορούν να οργανωθούν και να διαχειρίζονται με βάση το συναίσθημα. [63]

Παρόλο που το MER μπορεί να χρησιμοποιηθεί για να μπουν ετικέτες συναισθημάτων σε μουσικά κομμάτια, η ενσωμάτωση αυτών σε ένα σύστημα συστάσεων μουσικής δεν είναι εύκολη υπόθεση για διάφορους λόγους. Αρχικά, το μουσικό περιεχόμενο εξακολουθεί να περιέχει διάφορα χαρακτηριστικά που επηρεάζουν τη συναισθηματική κατάσταση του ακροατή, όπως οι στίχοι, ο ρυθμός και η αρμονία, και ο τρόπος με τον οποίο επηρεάζεται ο ακροατής είναι εξαιρετικά υποκειμενικός. Επίσης, η προτίμηση για ένα συγκεκριμένο είδος μουσικής εξαρτάται από το αν ο χρήστης θέλει να βελτιώσει ή να διαμορφώσει τη διάθεσή του. Τρίτον, μέσα στο ίδιο μουσικό κομμάτι συμβαίνουν συχνά συναισθηματικές αλλαγές, ενώ συνήθως ετικέτες εξάγονται για ολόκληρο το κομμάτι. [11]

Ένα παράδειγμα έρευνας στον τομέα του MER είναι η έρευνα των Pyrovolakis et al. [64] που προσέγγισαν το έργο της ανίχνευσης της μουσικής διάθεσης, αναλύοντας στίχους και ακουστικά σήματα, εφαρμόζοντας τις αρχές και τη λογική της εποπτευόμενης μάθησης. Στην έρευνα αυτή εισάγονται εφαρμογές συστημάτων βαθιάς μάθησης σε συνδυασμό με ορισμένες μεθόδους αναπαράστασης δεδομένων που βασίζονται στην επεξεργασία φυσικής γλώσσας και σε τεχνικές επεξεργασίας ψηφιακού σήματος. Ενώ η διαδικασία ολοκληρώνεται με την εκπαίδευση και την αξιολόγηση των τριών προτεινόμενων συστημάτων - μόνο στίχοι, μόνο ήχος και πολυτροπικό. Συμπεραίνεται, ότι το ηχητικό μέρος της μουσικής έχει μεγαλύτερη επιρροή όσον αφορά το συναίσθημα. Ωστόσο, όταν συνδυάζονται τόσο το λυρικό όσο και το ηχητικό μέρος ενός τραγουδιού, τα αποτελέσματα είναι εκπληκτικά (σχεδόν 24% βελτίωση στην ακρίβεια).

Όσον αφορά τη σημασία των στίχων στο μουσικό συναίσθημα, η έρευνα των Nasser et al. [65] υποστηρίζει ότι η σημασία των στίχων για την πρόβλεψη της διάθεσης σε σύγκριση με την ακουστική εξαρτάται από τη συγκεκριμένη διάθεση. Οι στίχοι είναι πιο σημαντικοί για την αντίληψη των θλιβερών συναισθημάτων από την ακουστική, αλλά η ακουστική είναι πρωταρχικής σημασίας για την αντίληψη των χαρούμενων συναισθημάτων.

Σε σχέση με τα μοντέλα συναισθημάτων που αναλύθηκαν παραπάνω θα πρέπει να αναφέρουμε ότι: Τα έξι βασικά συναισθήματα που ορίζονται από τον Ekman είναι πολύ γνωστά στην ψυχολογία. Ωστόσο, δεδομένου ότι σχεδιάστηκαν για την κωδικοποίηση των εκφράσεων του προσώπου, ορισμένα από αυτά μπορεί να μην είναι κατάλληλα για μουσική (π.χ. αηδία) και επίσης λείπουν κάποιες κοινές μουσικές διαθέσεις (π.χ. ηρεμία). Στη μουσική ψυχολογία, η πιο πρώιμη και ακόμα πιο γνωστή συστηματική προσπάθεια δημιουργίας μιας ταξινόμησης της μουσικής διάθεσης ήταν αυτή του Hevner. [62]

2.2 Συστήματα Συστάσεων

Το διαδίκτυο, οι ηλεκτρονικές υπηρεσίες και οι βιβλιοθήκες μουσικής, βίντεο και εικόνων, έχουν αυξηθεί τις τελευταίες δεκαετίες με αποτέλεσμα ένα πλεόνασμα πληροφοριών να είναι πλέον διαθέσιμο σε όλους. Ο χρήστης μέσα σε αυτό το περιβάλλον της υπερφόρτωσης πληροφορίας αντιμετωπίζει προβλήματα όσον αφορά τον χρόνο που αφιερώνει στην αναζήτησή του και τον όγκο των πληροφοριών που ανακτώνται από αυτήν. [66] Για το λόγο αυτό θα πρέπει να υπάρχουν βολικοί τρόποι εύρεσης σχετικών με την αναζήτηση πληροφοριών σε σύντομο χρονικό διάστημα. Πολλές εταιρείες ηλεκτρονικού εμπορίου προτείνουν προϊόντα στους χρήστες τους, πουλώντας εκατομμύρια προϊόντα σε μία πλατφόρμα. Για έναν καθημερινό χρήστη, η περιήγηση σε όλες τις δυνατότητες είναι εξαντλητική και δυσχεραίνει την αναζήτησή του.

Τα συστήματα συστάσεων στοχεύουν στην επίλυση αυτού του προβλήματος υπερφόρτωσης πληροφορίας. Είναι συστήματα φιλτραρίσματος πληροφοριών που συλλέγουν τις πληροφορίες που πραγματικά χρειάζεται ο χρήστης μέσα από ένα μεγάλο όγκο δυναμικά παραγόμενων πληροφοριών, εξατομικεύοντας παράλληλα την εμπειρία του χρήστη παρέχοντας ακριβείς συστάσεις για είδη/προϊόντα σε αυτόν σύμφωνα με τις προτιμήσεις του. Ένα σύστημα συστάσεων (RS) στοχεύει να προβλέψει εάν ένα προϊόν θα ήταν χρήσιμο σε έναν χρήστη με βάση δεδομένες πληροφορίες (πχ. το προφίλ του χρήστη). [67] , [12]

Σημαντικό είναι επίσης το γεγονός ότι τα συστήματα συστάσεων είναι ωφέλιμα τόσο για τους παρόχους υπηρεσιών όσο και για τους χρήστες. Μειώνουν το κόστος συναλλαγών για την εύρεση και την επιλογή αντικειμένων σε περιβάλλον διαδικτυακών αγορών. Τα συστήματα συστάσεων έχουν επίσης αποδειχθεί ότι βελτιώνουν τη διαδικασία λήψης αποφάσεων και την ποιότητα. Έτσι, η ανάγκη χρήσης αποτελεσματικών και ακριβών τεχνικών συστάσεων σε ένα σύστημα που θα παρέχει σχετικές και αξιόπιστες συστάσεις στους χρήστες δεν μπορεί να τονιστεί αρκετά. [14]

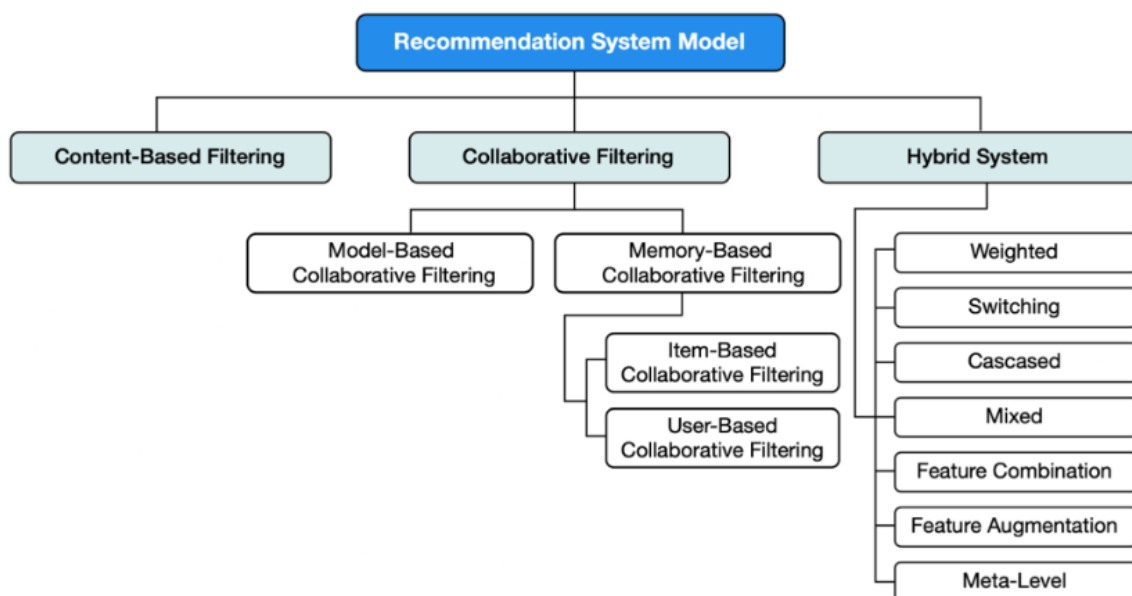
Η ύπαρξη των συστημάτων συστάσεων έχει εντοπιστεί στα τέλη της δεκαετίας του 1970, οπότε πιο συγκεκριμένα το 1979 εμφανίστηκε η πρώτη εφαρμογή της ιδέας μέσα από το σύστημα που ονομάστηκε Grundy [68]. Το Grundy ήταν ένα σύστημα που προσομοιάζει τον βιβλιοθηκάριο, που παρείχε προτάσεις στον χρήστη για το τι βιβλία να διαβάσει. Κατά τους Raghuwanshi και Pateriya [69] το πρώτο σύστημα συστάσεων που βασίζεται σε υπολογιστή αναπτύχθηκε το 1992 από τους Goldberg et al. [70] και ονομάστηκε Tapestry. Το Tapestry είναι ένα σύστημα συστάσεων αλληλογραφίας, ένα πειραματικό σύστημα φιλτραρίσματος

πληροφοριών που διαχειρίζεται την τεράστια εισερχόμενη ροή εγγράφων όπως e-mail, ειδήσεις και άρθρα. Το σύστημα αυτό βασίστηκε στην πεποίθηση ότι το φιλτράρισμα πληροφοριών μπορεί να είναι πιο αποτελεσματικό όταν οι άνθρωποι εμπλέκονται στη διαδικασία φιλτραρίσματος.

Από τότε πολλοί ερευνητές έχουν προτείνει διάφορες προσεγγίσεις για την ανάπτυξη αποτελεσματικών συστημάτων συστάσεων. Τα συστήματα συστάσεων αρχικά χρησιμοποιούσαν φιλτράρισμα περιεχομένου (content filtering), κατά το οποίο μία ομάδα ειδικών ταξινομεί τα προϊόντα σε κατηγορίες, ενώ οι χρήστες επιλέγουν τις κατηγορίες που προτιμούν και αντιστοιχίζονται με βάση τις προτιμήσεις τους. [71] Το πεδίο στη συνέχεια εξελίχθηκε και χρησιμοποιήθηκε το συνεργατικό φιλτράρισμα (collaborative filtering), όπου οι προτάσεις βασίζονται σε συμπεριφορές του παρελθόντος των χρηστών, όπως για παράδειγμα προηγούμενες αξιολογήσεις που δόθηκαν σε προϊόντα. [71] Αργότερα αναπτύχθηκαν με επιτυχία και άλλες μέθοδοι όπως θα αναλυθούν και στη συνέχεια.

2.2.1 Κατηγορίες

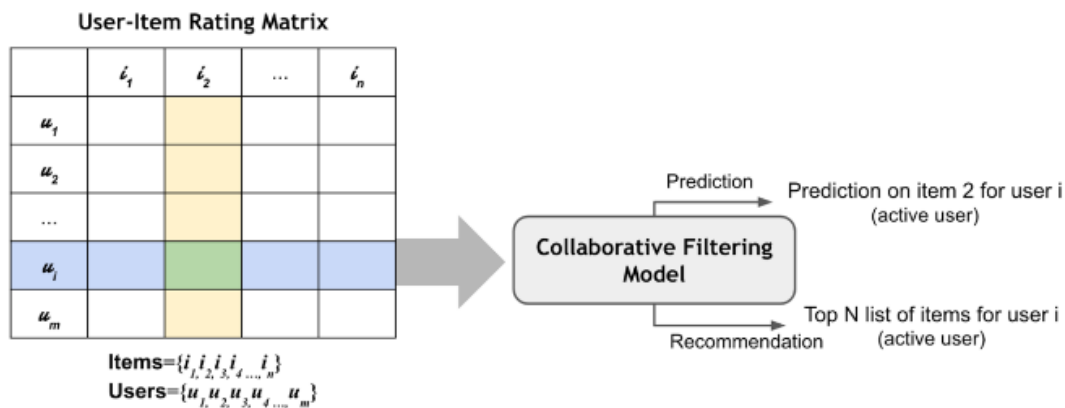
Πολλοί ερευνητές έχουν ορίσει διαφορετικές κατηγοριοποιήσεις των συστημάτων συστάσεων. Ένας κλασικός τρόπος κατηγοριοποίησης των διαφορετικών τύπων RS δόθηκε από τον Burke, ο οποίος διέκρινε μεταξύ έξι διαφορετικών κατηγοριών προσεγγίσεων συστάσεων: συνεργατικό φιλτράρισμα (Collaborative Filtering, CF), φιλτράρισμα περιεχομένου (Content-Based Filtering, CBF), φιλτράρισμα βάσει γνώσης (knowledge-based filtering), φιλτράρισμα βάσει περιβάλλοντος (context-based filtering), δημογραφικό φιλτράρισμα (demographic filtering) και υβριδικό φιλτράρισμα (hybrid filtering, HF). Τα RS που χρησιμοποιούνται πιο συχνά είναι τα CF, CBF και HF τα οποία χωρίζονται και σε περαιτέρω κατηγορίες, όπως φαίνεται στην Εικόνα 9.



Εικόνα 13: Επισκόπηση των βασικών μοντέλων συστημάτων συστάσεων [13]

Collaborative filtering

Το Συνεργατικό Φιλτράρισμα αξιολογεί προϊόντα χρησιμοποιώντας τις αξιολογήσεις των χρηστών (ρητές ή άρρητες, explicit or implicit) από ιστορικά δεδομένα. Η λειτουργία του στηρίζεται στη δημιουργία μιας βάσης δεδομένων (user-item matrix) που περιέχει τις προτιμήσεις του χρήστη για αντικείμενα. Στη συνέχεια, αντιστοιχίζει χρήστες με σχετικά ενδιαφέροντα και προτιμήσεις, υπολογίζοντας τις ομοιότητες μεταξύ των προφίλ τους για να κάνει συστάσεις. Τέτοιοι χρήστες δημιουργούν μια ομάδα που ονομάζεται γειτονιά. Ένας χρήστης λαμβάνει συστάσεις για εκείνα τα αντικείμενα που δεν έχει αξιολογήσει στο παρελθόν, αλλά τα οποία έχουν ήδη αξιολογηθεί θετικά από χρήστες στη γειτονιά του. [12], [14]



Εικόνα 14: Αρχή συστάσεων του μοντέλου συνεργατικού φιλτραρίσματος [13]

Η τεχνική του συνεργατικού φιλτραρίσματος μπορεί να χωριστεί σε δύο κατηγορίες: βάσει μνήμης (memory-based) και βάσει μοντέλου (model-based). [14]

Βάσει μνήμης: Τα στοιχεία που είχαν ήδη βαθμολογηθεί από τον χρήστη στο παρελθόν παίζουν σημαντικό ρόλο στην αναζήτηση ενός γείτονα που μοιράζεται κοινά ενδιαφέροντα μαζί του. Μόλις βρεθεί ένας γείτονας ενός χρήστη, διάφοροι αλγόριθμοι μπορούν να χρησιμοποιηθούν για να συνδυαστούν οι προτιμήσεις των γειτόνων για τη παραγωγή προτάσεων. Το CF που βασίζεται στη μνήμη μπορεί να επιτευχθεί με δύο τρόπους, μέσω τεχνικών βάσει χρηστών και τεχνικών βάσει στοιχείων.

Η τεχνική συνεργατικού φιλτραρίσματος **βάσει χρήστη** υπολογίζει την ομοιότητα μεταξύ των χρηστών συγκρίνοντας τις αξιολογήσεις τους για το ίδιο αντικείμενο και στη συνέχεια υπολογίζει την προβλεπόμενη βαθμολογία για ένα στοιχείο από τον ενεργό χρήστη, ως σταθμισμένο μέσο όρο των αξιολογήσεων του στοιχείου από χρήστες παρόμοιους με τον ενεργό χρήστη. Τα βάρη είναι οι ομοιότητες αυτών των χρηστών με το αντικείμενο-στόχο. Από την άλλη, η τεχνική CF **βάσει στοιχείων** υπολογίζει τις προβλέψεις χρησιμοποιώντας την ομοιότητα μεταξύ των στοιχείων/αντικειμένων και όχι την ομοιότητα μεταξύ των χρηστών. Δημιουργεί ένα μοντέλο ομοιοτήτων αντικειμένων ανακτώντας όλα τα στοιχεία που έχουν βαθμολογηθεί από έναν ενεργό χρήστη από τον πίνακα στοιχείου-χρήστη (τη βάση δεδομένων), καθορίζει πόσο παρόμοια είναι τα ανακτημένα στοιχεία με το αντικείμενο-

στόχος και, στη συνέχεια, επιλέγει τα k πιο παρόμοια στοιχεία και υπολογίζονται οι αντίστοιχες ομοιότητές τους.

Βάσει μοντέλου: Αυτή η τεχνική χρησιμοποιεί τις προηγούμενες αξιολογήσεις του χρήστη για την εκμάθηση ενός μοντέλου προκειμένου να βελτιώσει την απόδοση της Τεχνικής Συνεργατικού φιλτραρίσματος. Η διαδικασία κατασκευής μοντέλων μπορεί να γίνει χρησιμοποιώντας τεχνικές μηχανικής μάθησης ή εξόρυξης δεδομένων. Αυτές οι τεχνικές μπορούν γρήγορα να προτείνουν ένα σύνολο στοιχείων διότι χρησιμοποιούν προ-υπολογισμένο μοντέλο και έχουν αποδειχθεί ότι παράγουν αποτελέσματα συστάσεων που είναι παρόμοια με τις τεχνικές συστάσεων που βασίζονται στη γειτονιά. Οι τεχνικές που βασίζονται σε μοντέλα αναλύουν τον πίνακα χρήστη-στοιχείου για να προσδιορίσουν τις σχέσεις μεταξύ των στοιχείων. Έπειτα, χρησιμοποιούν αυτές τις σχέσεις για να συγκρίνουν τη λίστα των N κορυφαίων συστάσεων. Οι τεχνικές που βασίζονται σε μοντέλα επιλύουν τα προβλήματα αραιότητας (sparsity problems) που σχετίζονται με τα συστήματα συστάσεων.

Content-based filtering

Το φιλτράρισμα βάσει περιεχομένου δίνει έμφαση περισσότερο στην ανάλυση των χαρακτηριστικών των αντικειμένων, προκειμένου να δημιουργηθούν οι προβλέψεις. Η σύσταση γίνεται με βάση το προφίλ του χρήστη χρησιμοποιώντας χαρακτηριστικά που εξάγονται από το περιεχόμενο των στοιχείων που αυτός έχει αξιολογήσει στο παρελθόν. Όταν προτείνονται έγγραφα όπως ιστοσελίδες, δημοσιεύσεις και ειδήσεις, η τεχνική φιλτραρίσματος βάσει περιεχομένου είναι η πιο επιτυχημένη. [14]

Η τεχνική φιλτραρίσματος βάσει περιεχομένου δεν χρειάζεται το προφίλ άλλων χρηστών, καθώς δεν επηρεάζουν το αποτέλεσμα των προτάσεων. Επίσης, εάν αλλάξει το προφίλ του χρήστη, η τεχνική CBF εξακολουθεί να έχει τη δυνατότητα να προσαρμόσει τις συστάσεις της σε πολύ σύντομο χρονικό διάστημα. Το σημαντικότερο μειονέκτημα αυτής της τεχνικής είναι η ανάγκη να έχουμε μια εις βάθος γνώση και περιγραφή των χαρακτηριστικών του κάθε στοιχείου. [14] Επίσης, η τεχνική αυτή συνιστά μόνο δεδομένα που σχετίζονται με στοιχεία που έχουν στενή σχέση με άλλα στοιχεία που είχαν προηγουμένως αξιολογηθεί από τον χρήστη, επομένως το σύστημα είναι γνωστό για τον περιορισμό του ότι δεν μπορεί να προτείνει νέα στοιχεία. [13]

Το μοντέλο φιλτραρίσματος βάσει περιεχομένου είναι το πιο βασικό μοντέλο στα συστήματα συστάσεων και έχει χρησιμοποιηθεί κυρίως σε πρώιμα συστήματα συστάσεων.

Hybrid filtering

Η μέθοδος υβριδικού φιλτραρίσματος συνδυάζει δύο ή περισσότερες από τις διαφορετικές προσεγγίσεις συστάσεων που αναφέρθηκαν στην αρχή, με σκοπό την βελτίωση της απόδοσης και την επίλυση προβλημάτων ή περιορισμών που προκύπτουν μέσω της χρήσης των επί μέρους προσεγγίσεων. Το HF χωρίζεται σε επτά κατηγορίες: Weighted, Switching, Cascaded, Mixed, Feature-Combination, Feature-Augmentation και Meta-level, σύμφωνα με τη μέθοδο που συνδυάζει τις τεχνικές φιλτραρίσματος στη μελέτη του Burke [72].

Στη συνέχεια θα αναφέρουμε πολύ περιληπτικά τι είναι η κάθε μία από τις παραπάνω κατηγορίες [12]–[14].

- Ο σταθμισμένος υβριδισμός (weighted) συνδυάζει τα αποτελέσματα διαφορετικών συστάσεων για να δημιουργήσει μια λίστα συστάσεων ή πρόβλεψη ενσωματώνοντας τις βαθμολογίες από καθεμία από τις τεχνικές που χρησιμοποιούνται με έναν γραμμικό τύπο.
- Με την εναλλαγή (switching) επιλέγεται κάθε φορά το μοντέλο συστάσεων που προσφέρει τα καλύτερα αποτελέσματα.
- Ο μικτός υβριδισμός (mixed) συνδυάζει τα αποτελέσματα διαφορετικών τεχνικών συστάσεων που γίνονται ταυτόχρονα αντί να δίνει μόνο μία σύσταση ανά είδος. Σε αυτή τη μέθοδο, οι επιμέρους επιδόσεις δεν επηρεάζουν πάντα τη γενική απόδοση μιας τοπικής περιοχής.
- Στον υβριδισμό καταρράκτη (cascaded) μετά τη χρήση ενός από τα μοντέλα συστήματος συστάσεων για τη δημιουργία ενός υποψηφίου σετ με παρόμοιο γούστο με το χρήστη, η μέθοδος συνδυάζει το μοντέλο συστήματος συστάσεων που χρησιμοποιήθηκε προηγουμένως με ένα άλλο μοντέλο για να ταξινομήσει το υποψήφιο σύνολο με τη σειρά των αντικειμένων που ταιριάζουν περισσότερο στις προτιμήσεις του χρήστη.
- Κατά τον υβριδισμό συνδυασμού χαρακτηριστικών (feature-combination) τα χαρακτηριστικά που παράγονται από μια συγκεκριμένη τεχνική συστάσεων δίνονται ως είσοδος σε κάποια άλλη τεχνική.
- Η επαύξηση χαρακτηριστικών (feature augmentation) εφαρμόζεται για την παραγωγή της βαθμολογίας ενός αντικειμένου και, στη συνέχεια, την ενσωμάτωση αυτών των πληροφοριών στην διαδικασία επεξεργασίας της επόμενης τεχνικής σύστασης.
- Στον μετα-επίπεδο υβριδισμό (meta-level) ολόκληρο το μοντέλο ενός συστήματος συστάσεων δίνεται ως είσοδος σε ένα άλλο σύστημα συστάσεων.

2.2.2 Προκλήσεις

Είναι δύσκολο να μετρηθεί η απόδοση των RS λόγω της μεταβαλλόμενης ζήτησης του οργανισμού στη χρήση και την ανάπτυξή τους. Γενικά, το πιο ενδεικτικό μέτρο είναι η ικανοποίηση των χρηστών. Παρόλο που δεν είναι δυνατός ο υπολογισμός της ικανοποίησης των χρηστών, μπορούμε να μετρήσουμε την απόδοση των RS με βάση το πόσο καλά μπορούν να χειριστούν κοινά ζητήματα. [12] Κάθε μια από τις κατηγορίες συστημάτων συστάσεων εμφανίζει προβλήματα ή περιορισμούς ως προς τη χρήση της. Στη συνέχεια δίνεται μια σύντομη επεξήγηση για το κάθε πρόβλημα.

Όσον αφορά το συνεργατικό φιλτράρισμα, εάν το μοντέλο δεν περιλαμβάνει αρκετά δεδομένα, μπορεί να προκύψουν τρία προβλήματα: αραιότητα (sparsity), ψυχρή εκκίνηση (cold-start) και γκριζό πρόβατο (grey sheep). Πρώτον, το πρόβλημα της αραιότητας είναι ένα

πρόβλημα που παρουσιάζεται όταν δεν υπάρχουν αρκετά διαθέσιμα δεδομένα για σύσταση. Ομοίως, το πρόβλημα ψυχρής εκκίνησης εμφανίζεται όταν δεν υπάρχουν καθόλου δεδομένα αξιολόγησης, δηλαδή ο πρώτος βαθμολογητής λόγω της εισροής νέων χρηστών στην αρχή μιας υπηρεσίας. Τρίτον, το γκρίζο πρόβατο είναι ένα πρόβλημα στο οποίο προκύπτουν δυσκολίες σύστασης όταν το σύνολο των χρηστών των οποίων τα δεδομένα αξιολόγησης είναι παρόμοια με αυτά του μεμονωμένου χρήστη είναι πολύ μικρό. [13] Επιπρόσθετα, το συνεργατικό φιλτράρισμα μπορεί να έχει και άλλα προβλήματα όπως επεκτασιμότητα (scalability) και συνωνυμία (synonymy). Το πρόβλημα της επεκτασιμότητας εμφανίζεται επειδή ο υπολογισμός συνήθως αυξάνεται γραμμικά με τον αριθμό των χρηστών και των στοιχείων. Μια τεχνική συστάσεων που είναι αποτελεσματική όταν ο αριθμός των δεδομένων είναι περιορισμένος μπορεί να μην είναι σε θέση να δημιουργήσει ικανοποιητικό αριθμό προτάσεων όταν ο όγκος του συνόλου δεδομένων αυξάνεται. Τέλος, συνωνυμία είναι η τάση πολύ όμοιων αντικειμένων να έχουν διαφορετικά ονόματα ή καταχωρήσεις. Σε αυτή την περίπτωση, τα περισσότερα συστήματα συστάσεων δυσκολεύονται να κάνουν διάκριση μεταξύ στενά συνδεδεμένων στοιχείων. [14]

Οι τεχνικές φιλτραρίσματος βάσει περιεχομένου υποφέρουν και αυτές από διάφορα προβλήματα. Το βασικό πρόβλημα είναι ότι εξαρτώνται από τα μεταδεδομένα των στοιχείων. Δηλαδή, απαιτούν πλούσια περιγραφή των αντικειμένων και πολύ καλά οργανωμένο προφίλ χρήστη πριν γίνει σύσταση στους χρήστες. Αυτό ονομάζεται ανάλυση περιορισμένου περιεχομένου (limited content analysis). Έτσι, η αποτελεσματικότητα του CBF εξαρτάται από τη διαθεσιμότητα περιγραφικών δεδομένων. Παρομοίως, μπορούμε να διακρίνουμε το πρόβλημα του νέου χρήστη, επίσης γνωστό ως πρόβλημα έλλειψης πληροφοριών. Όταν ο νέος χρήστης εισάγεται στο σύστημα, το σύστημα δεν έχει επαρκείς πληροφορίες για το προφίλ χρήστη και την προτίμησή του για συγκεκριμένα προϊόντα, επομένως δεν μπορεί να δημιουργηθεί το κατάλληλο προφίλ αντικειμένου με βάση αυτό. Ένα άλλο σοβαρό πρόβλημα της τεχνικής CBF είναι η εξειδίκευση του περιεχομένου (overspecialization). Με άλλα λόγια, οι χρήστες περιορίζονται στο να λαμβάνουν προτάσεις παρόμοιες με στοιχεία που έχουν ήδη οριστεί στα προφίλ τους. [14], [73]

Το υβριδικό μοντέλο φιλτραρίσματος έχει εφαρμοστεί ώστε να ξεπεράσει τα μειονεκτήματα των παραπάνω και να βελτιώσει την απόδοση του συστήματος συστάσεων. Έτσι, γίνεται χρήση και CB και CF μεθόδων με τον κατάλληλο συνδυασμό ώστε να επιτευχθεί η μέγιστη δυνατή απόδοση.

2.2.3 Μετρικές

Η ποιότητα του συστήματος συστάσεων μετριέται μέσω διαφόρων τύπων μετρικών αξιολόγησης με βάση την ακρίβεια της πρόβλεψης. Η επιλογή της μετρικής εξαρτάται από την τεχνική φιλτραρίσματος, τα χαρακτηριστικά του συνόλου δεδομένων και την εργασία του συστήματος συστάσεων. Κάποιες τυπικές μετρικές για την αξιολόγηση των συστημάτων συστάσεων είναι οι ακόλουθες και χωρίζονται σε δύο κατηγορίες [14], [69], [73].

Η πρώτη κατηγορία περιλαμβάνει τις μετρικές για την ακρίβεια της πρόβλεψης (στατιστικές μετρικές).

- $$\text{Precision} = \frac{TP}{TP+FP}$$

Precision (ακρίβεια) είναι το κλάσμα των προτεινόμενων αντικειμένων που είναι πραγματικά σχετικό με τον χρήστη. (# Σωστά προτεινόμενα στοιχεία / # όλα τα προτεινόμενα στοιχεία)

- $$\text{Recall} = \frac{TP}{TP+FN}$$

Recall (ανάκληση) μπορεί να οριστεί ως το κλάσμα των σχετικών αντικειμένων που αποτελούν επίσης μέρος του συνόλου των προτεινόμενων αντικειμένων. (# Σωστά προτεινόμενα στοιχεία / # όλα τα χρήσιμα προτεινόμενα στοιχεία)

Το Recall αναφέρεται μερικές φορές και ως Accuracy.

- $$\text{F-measure} = \frac{2 * \text{Precision} * \text{Recall}}{\text{Precision} + \text{Recall}}$$

Η F-μετρική υπολογίζει το σταθμισμένο αρμονικό μέσο των Precision και Recall. Η τιμή που προκύπτει κάνει τη σύγκριση μεταξύ αλγορίθμων και μεταξύ των συνόλων δεδομένων πολύ απλή και άμεση.

- ROC (Receiver operating characteristic) curve

Οι καμπύλες ROC χρησιμοποιούνται για να περιγράψουν οπτικά την σχέση μεταξύ ανάκλησης και ακρίβειας όταν αλλάζει το κατώφλι που μας βοηθά να ταξινομήσουμε ένα είδος ως "προτείνεται" και "δεν συνιστάται". Είναι πολύ επιτυχημένες στην αξιολόγηση της απόδοσης συγκεκριμένων αλγορίθμων. Χρησιμοποιείται εναλλακτικά των precision/recall.

Η δεύτερη κατηγορία περιλαμβάνει τις μετρικές για την ακρίβεια της ταξινόμησης (μετρικές υποστήριξης αποφάσεων).

- MAE: Mean Absolute Error

$$MAE = \frac{1}{N} \sum_{u,i} |p_{u,i} - r_{u,i}|$$

όπου $p_{u,i}$ είναι η προβλεπόμενη αξιολόγηση του χρήστη u για το αντικείμενο i , $r_{u,i}$ είναι η πραγματική αξιολόγηση και N ο αριθμός όλων των αξιολογήσεων για το αντικείμενο.

Το MAE είναι το πιο δημοφιλές και ευρέως χρησιμοποιούμενο. Αντιπροσωπεύει ένα μέτρο της απόκλισης της προτεινόμενης τιμής από τη συγκεκριμένη τιμή του χρήστη. Όσο χαμηλότερο είναι το MAE, τόσο καλύτερη η ακρίβεια της πρότασης.

- RMSE: Root Mean Square Error

$$RMSE = \sqrt{\frac{1}{n} \sum_{u,i} (p_{u,i} - r_{u,i})^2}$$

Το RMSE δίνει μεγαλύτερη έμφαση στο μεγαλύτερο απόλυτο σφάλμα. Όσο χαμηλότερο είναι το RMSE, τόσο καλύτερη είναι η ακρίβεια της σύστασης.

2.2.4 Εφαρμογές

Τα συστήματα συστάσεων ήταν κάποτε μια καινοτόμος τεχνική που χρησιμοποιούνταν από πολύ λίγους ιστότοπους ηλεκτρονικού εμπορίου. Τώρα, τα RS έχουν επεκταθεί και χρησιμοποιούνται σε πολλούς τομείς υπηρεσιών. Οι μελέτες που συλλέχθηκαν για αυτή την εργασία ανέφεραν διάφορες κατηγοριοποιήσεις των συστημάτων συστάσεων ανάλογα με τους τομείς εφαρμογών τους, όπως: ηλεκτρονικό εμπόριο (e-commerce), ψυχαγωγία/ υπηρεσίες ροής (streaming), τομέας υγείας, τουρισμός, γεωργία, μεταφορές, εκπαίδευση. Στη συνέχεια παρουσιάζονται μερικοί από αυτούς [12], [13].

Ηλεκτρονικό εμπόριο

Πολλές από τις μεγαλύτερες επιχειρήσεις ηλεκτρονικού εμπορίου χρησιμοποιούν RS για να βοηθήσουν τους χρήστες να προσδιορίσουν τι θέλουν, μετριάζοντας το πρόβλημα υπερφόρτωσης πληροφοριών.

Τα τελευταία χρόνια, με την ανάπτυξη των ψηφιακών πλατφορμών, η μορφή κατανάλωσης ειδών έχει αλλάξει σε αυτή του ηλεκτρονικού εμπορίου. Το ηλεκτρονικό εμπόριο προσφέρει στους καταναλωτές πολλά είδη και διάφορες επιλογές και παρέχει στους πωλητές έναν εύκολο τρόπο πώλησης. Ειδικότερα στην εποχή του COVID-19, οι καταναλωτές δεν μπόρεσαν να βγουν έξω λόγω των μέτρων καραντίνας, με αποτέλεσμα να μην μπορούν να χρησιμοποιήσουν τα φυσικά καταστήματα και να επιδοθούν στη χρήση ψηφιακών πλατφορμών, η οποία αυξήθηκε εκθετικά.

Το πιο σημαντικό χαρακτηριστικό της υπηρεσίας ηλεκτρονικού εμπορίου είναι ότι οι καταναλωτές συνήθως εμφανίζουν ένα μοτίβο δαπανών που συμπληρώνει τα είδη που έχουν προτιμήσει στο παρελθόν ή έχουν ήδη αγοράσει. Επομένως, η εύρεση αντικειμένων παρόμοιων με αυτά που είχε αγοράσει προηγουμένως ο χρήστης είναι χρήσιμη για την σύσταση άλλων αντικειμένων κατάλληλων για αυτόν. Για το λόγο αυτό, στην υπηρεσία χρησιμοποιούνται κυρίως το μοντέλο συστάσεων για συνεργατικό φιλτράρισμα και το μοντέλο υβριδικού φιλτραρίσματος.

Επιπρόσθετα, οι ψηφιακές πλατφόρμες συλλέγουν πληροφορίες για τους χρήστες τους με τον εντοπισμό των ενεργειών του ποντικιού και του πληκτρολογίου όταν ο χρήστης χρησιμοποιεί την υπηρεσία. Τα δεδομένα αυτά μπορούν να αναλύσουν τις προτιμήσεις των χρηστών παρακολουθώντας τις αλληλεπιδράσεις τους με προγράμματα περιήγησης και εφαρμογές για να καθορίσουν τις προθέσεις αγοράς. Τέλος, είναι δυνατό ένα προϊόν να προτείνεται σε έναν

χρήστη χρησιμοποιώντας τις πληροφορίες προτιμήσεων μιας ομάδας άλλων χρηστών με παρόμοιες προτιμήσεις με τον χρήστη.

Ιδιαίτερα δημοφιλείς πλατφόρμες ηλεκτρονικού εμπορίου είναι το Amazon, το eBay και η Alibaba. Πιο συγκεκριμένα, το Amazon έχει γίνει αντιληπτό από ερευνητές λόγω του εξαιρετικά εξατομικευμένου ιστότοπου του και της χρήσης του συνεργατικού φιλτραρίσματος βάσει στοιχείων (item-based CF), ενός από τους πιο συχνά χρησιμοποιούμενους τύπους RS.

Ψυχαγωγία / Υπηρεσίες ροής

Στις μέρες μας, ένας σημαντικός όγκος των βίντεο, όπως ταινίες ή σειρές, καταναλώνεται μέσω πλατφορμών ροής όπως το Netflix και το YouTube και ξεφεύγει από την τηλεόραση και τις κινηματογραφικές αίθουσες. Το ηχητικό περιεχόμενο αλλάζει επίσης από τη λήψη και την κατανάλωση αρχείων στην τοπική συσκευή ενός χρήστη, στην κατανάλωση περιεχομένου μέσω πλατφορμών ροής όπως το Spotify. Οι υπηρεσίες ροής που σχετίζονται με περιεχόμενο πολυμέσων έχουν αναπτυχθεί ταυτόχρονα με τα συστήματα συστάσεων, αφού είναι απαραίτητο να περιοριστούν οι ανησυχίες των χρηστών σχετικά με την επιλογή στοιχείων μέσα από ένα μεγάλο όγκο περιεχομένων και να παρέχεται περιεχόμενο προσαρμοσμένο σε κάθε χρήστη.

Γενικά, στο πεδίο της υπηρεσίας ροής, τα δεδομένα προτιμήσεων χρήστη συλλέγονται κυρίως μέσα από το ιστορικό χρήσης της υπηρεσίας. Μετά την αντιστοίχιση της προτίμησης του χρήστη με όλο το περιεχόμενο που ανήκει στην υπηρεσία ροής, δημιουργούνται προτάσεις που του εμφανίζονται με τη σειρά που ταιριάζει καλύτερα στις προτιμήσεις.

Σε αυτό το κλίμα ανάγκης προσωποποιημένων προτάσεων, έχουν γίνει μελέτες για CB μοντέλο προτάσεων, λαμβάνοντας υπόψη δεδομένα χρήστη και δεδομένα περιεχομένου ήχου, όπως και για HF μοντέλο που συνδυάζει ένα CF μοντέλο που λαμβάνει υπόψη δεδομένα αξιολόγησης άλλων χρηστών παρόμοια με τον χρήστη. Το YouTube και το Netflix έχουν χρησιμοποιήσει συνεργατικό φιλτράρισμα για προτάσεις βίντεο και ταινιών. Από την άλλη πλευρά, στον τομέα της υπηρεσίας ροής μουσικής, έχουν πραγματοποιηθεί κυρίως μελέτες με CB μοντέλα, λαμβάνοντας υπόψη χαρακτηριστικά ήχου όπως το είδος, η μελωδία και ο ρυθμός που αντικατοπτρίζουν το γούστο του χρήστη.

Αξίζει να αναφερθεί ότι πολλές πληροφορίες για τους χρήστες συλλέγονται από τα προφίλ και τις αναρτήσεις τους στα μέσα κοινωνικής δικτύωσης, όπως facebook, Instagram, twitter, tiktok ή ακόμα και από τα σχόλια και τις προτιμήσεις που δηλώνουν κατά την αλληλεπίδρασή τους με τις παραπάνω εφαρμογές.

Τομέας υγείας

Καθώς το ενδιαφέρον για την υγεία αυξάνεται, ο αριθμός των ανθρώπων που χρησιμοποιούν έξυπνες φορητές συσκευές έχει επίσης αυξηθεί λόγω και της ευχρηστίας των συσκευών αυτών. Οι φορητές συσκευές συλλέγουν έναν τεράστιο όγκο βιομετρικών δεδομένων για να βοηθήσουν στην έρευνα που σχετίζεται με ασθένειες, στην κατάλληλη διάγνωση ή και στην

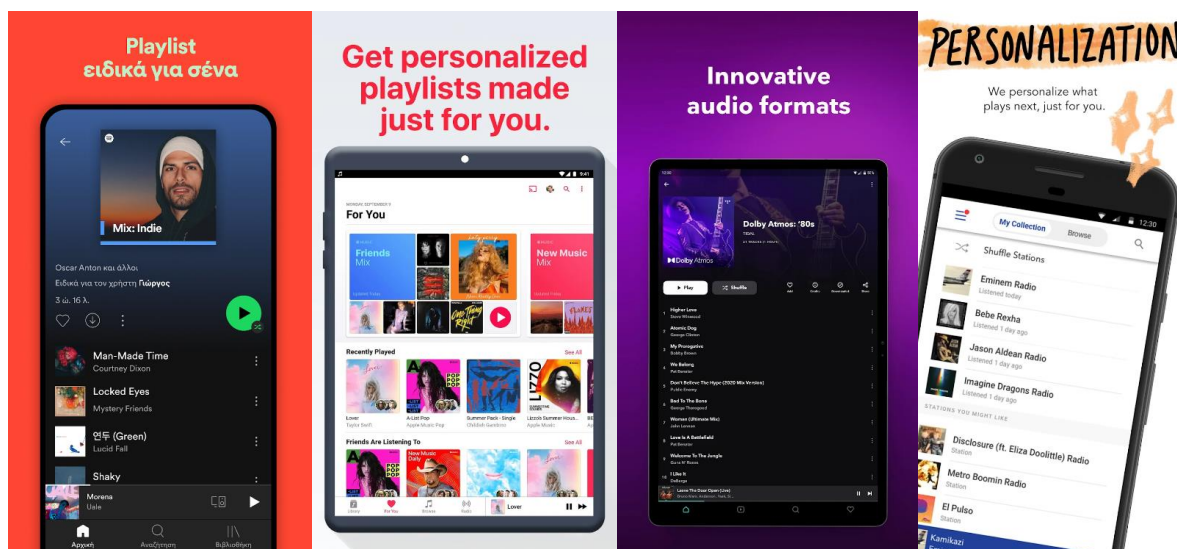
έρευνα που συνιστά θεραπεία. Μια αντιπροσωπευτική φορητή συσκευή είναι το Smart Watch που μπορεί να μετρήσει τους καρδιακούς παλμούς, τα επίπεδα οξυγόνου στο αίμα, τις θερμίδες που καταναλώθηκαν, την ποιότητα του ύπνου κ.α.

Στον τομέα των υπηρεσιών ηλεκτρονικής υγείας, εξατομικευμένες πληροφορίες υγείας χρησιμοποιούνται για να βοηθήσουν τον χρήστη να ελέγξει ανεξάρτητα τα δεδομένα υγείας του. Επίσης, δίνεται στους χρήστες πρόσβαση σε καθημερινές πληροφορίες υγείας ανά πάσα στιγμή και οπουδήποτε. Οι μελέτες συστημάτων συστάσεων που εντοπίζονται στον τομέα της υγείας, αναλύουν τη σχέση μεταξύ των προτύπων συμπτωμάτων των ασθενών και των ασθενειών για να παρέχουν στους χρήστες πληροφορίες για καλύτερες επιλογές θεραπείας.

Το CB μοντέλο συστάσεων χρησιμοποιείται συχνά επειδή είναι απαραίτητο να αναλυθούν οι πληροφορίες των ασθενών και τα χαρακτηριστικά της νόσου των ασθενών, ενώ το μοντέλο CF χρησιμοποιείται συχνά σε μελέτες συστημάτων συστάσεων υγείας που προτείνουν άλλες ομάδες ασθενών με συμπτώματα παρόμοια με αυτά του ασθενούς.

2.2.5 Συστήματα συστάσεων μουσικής (MRS)

Τα συστήματα συστάσεων μουσικής (Music Recommendation Systems, MRS) γνώρισαν έκρηξη τα τελευταία χρόνια, χάρη στην εμφάνιση και την επιτυχία των διαδικτυακών υπηρεσιών ροής, όπως το Spotify, η Pandora, η Apple Music, η Amazon Music, το YouTube Music ή το TIDAL, που σήμερα καθιστούν διαθέσιμη στα χέρια του χρήστη σχεδόν όλη τη μουσική του κόσμου. [11]



Εικόνα 15: Υπηρεσίες ροής μουσικής. Spotify, Apple Music, Tidal, Pandora [74]

Η πρόκληση των συστημάτων σύστασης μουσικής είναι ότι τα ανθρώπινα όντα είναι απρόβλεπτα και μπορεί να τους αρέσει μια μουσική σήμερα, αλλά μπορεί να τους αρέσει μια εντελώς διαφορετική μουσική αύριο. Είναι αδύνατο να εντοπιστούν οι αντιδράσεις των

χρηστών σε πραγματικό χρόνο. [75] Η ικανοποίηση των αναγκών μουσικής ψυχαγωγίας των χρηστών απαιτεί τη συνεκτίμηση εσωτερικών (implicit), εξωτερικών (explicit) και συμφραζόμενων (contextual) πτυχών των ακροατών, καθώς και άλλων πληροφοριών αλληλεπίδρασης. Για παράδειγμα, η προσωπικότητα και η συναισθηματική κατάσταση των ακροατών (εσωτερική) καθώς και η δραστηριότητά τους (εξωγενής) είναι γνωστό ότι επηρεάζουν τις μουσικές προτιμήσεις και ανάγκες. Το ίδιο ισχύει και για τα συμφραζόμενα στοιχεία των χρηστών, όπως οι καιρικές συνθήκες, το κοινωνικό περιβάλλον ή τα μέρη ενδιαφέροντος. [11] Έτσι, έχουν γίνει διάφορες μελέτες όπως αυτή των Schedl et al. [11] όπου εξετάζονται θέματα όπως η ψυχολογικά (προσωπικότητα και τα συναισθήματα) εμπνευσμένη μουσική, η μουσική με επίγνωση της κατάστασης (τοποθεσία, ώρα της ημέρας, κοινωνικό πλαίσιο, καιρός και δραστηριότητα) και η μουσική με επίγνωση του πολιτισμού.

Κύριες προκλήσεις

Οι πιο πιεστικές τρέχουσες προκλήσεις στην έρευνα των MRS είναι: η ψυχρή εκκίνηση, η αυτόματη συνέχιση της λίστας αναπαραγωγής και η αξιολόγηση των MRS. [11] Η ψυχρή εκκίνηση (cold-start), όπως έχει αναφερθεί και παραπάνω (κεφ. 2.2.3), αφορά το πρόβλημα που εμφανίζεται όταν εισάγεται ένα νέο τραγούδι ή ένας νέος χρήστης οπότε το σύστημα δεν έχει επαρκή στοιχεία για να γίνουν οι κατάλληλες προτάσεις μουσικών τραγουδιών στο χρήστη.

Η δεύτερη κύρια πρόκληση αφορά την αυτόματη συνέχιση της λίστας αναπαραγωγής (APC) ή αλλιώς την σύσταση διαδοχικών μουσικών κομματιών. Η εργασία αυτή συνίσταται στην προσθήκη ενός ή περισσότερων κομματιών σε μια λίστα αναπαραγωγής με τρόπο που ταιριάζει στα χαρακτηριστικά της αρχικής λίστας αναπαραγωγής. Η επίλυση του APC προσφέρει πλεονεκτήματα τόσο στην ακρόαση όσο και στη δημιουργία λιστών αναπαραγωγής: οι χρήστες μπορούν να απολαμβάνουν την ακρόαση συνεχών περιόδων σύνδεσης, ενώ επίσης, είναι πιο εύκολο να δημιουργήσουν μεγαλύτερες, πιο συναρπαστικές λίστες αναπαραγωγής χωρίς να χρειάζεται να έχουν μεγάλη εξοικείωση με τη μουσική.

Μια ακόμα βασική πρόκληση είναι η αξιολόγηση των MRS. Εκτός από τις μετρικές που εφαρμόζονται γενικότερα στα συστήματα συστάσεων, όπως precision, recall, accuracy (κεφ. 2.2.4), τα τελευταία χρόνια έχουν εμφανιστεί νέες μετρικές που ονομάζονται πέρα-από-την-ακρίβεια μετρικές. Αυτές αντιμετωπίζουν τις ιδιαιτερότητες των συστημάτων συστάσεων και μετρούν, για παράδειγμα, τη χρησιμότητα (utility), την καινοτομία (novelty), το serendipity ή την ποικιλομορφία (diversity) ενός αντικειμένου. Η καινοτομία μετρά την ικανότητα ενός συστήματος συστάσεων να προτείνει νέα στοιχεία για τα οποία ο χρήστης δεν γνώριζε πριν. Το serendipity στοχεύει στην αξιολόγηση του MRS με βάση σχετικές συστάσεις και συστάσεις που προκαλούν έκπληξη. Βέβαια, το ερώτημα του πώς να μετρηθεί ο βαθμός του serendipity για μια λίστα συστάσεων είναι αμφιλεγόμενο, αφού εξαρτάται από το εάν ο παράγοντας της έκπληξης συνεπάγεται ότι τα αντικείμενα πρέπει να είναι νέα για τον χρήστη. Τέλος, η ποικιλομορφία μετράει τον βαθμό στον οποίο τα προτεινόμενα στοιχεία είναι διαφορετικά μεταξύ τους, όπου η διαφορά μπορεί να σχετίζεται με πτυχές, όπως μουσικό στυλ, καλλιτέχνη, στίχους ή όργανα.

Ιδιαιτερότητες του MRS

Σε σύγκριση με άλλους τομείς στους οποίους χρησιμοποιούνται συστήματα συστάσεων, όπως προϊόντα, ταινίες ή ξενοδοχεία, η σύσταση στον τομέα της μουσικής έχει ορισμένα συγκεκριμένα χαρακτηριστικά που πρέπει να λαμβάνονται υπόψη κατά τη δημιουργία ενός MRS. [15], [11] Οι ιδιαιτερότητες της σύστασης μουσικής περιλαμβάνουν επιγραμματικά: τη διάρκεια, το μέγεθος, τη διαδοχική κατανάλωση, την επαναληπτική σύσταση (σύσταση προηγούμενων συστάσεων), την παθητική καταναλωτική συμπεριφορά και την πρόκληση δυνατών συναισθημάτων.

Αρχικά, η διάρκεια ενός μουσικού κομματιού είναι πολύ μικρότερη από το χρόνο που απαιτείται για την προβολή μιας ταινίας ή το διάβασμα ενός βιβλίου. Επίσης, ο αριθμός των μουσικών κομματιών σε έναν εμπορικό κατάλογο είναι της τάξης των δέκα εκατομμυρίων κομματιών, πολύ μεγαλύτερος από άλλα προϊόντα, πχ. για ταινίες είναι της τάξης των εξήντα χιλιάδες. Εξαιτίας αυτών των δύο χαρακτηριστικών ένα μουσικό είδος μπορεί να θεωρηθεί αναλώσιμο και η σύσταση ενός μουσικού κομματιού που δεν ανταποκρίνεται απόλυτα στις ανάγκες του χρήστη περνάει σχετικά απαρατήρητη.

Τα μουσικά κομμάτια καταναλώνονται συχνά διαδοχικά, συνήθως ως λίστες αναπαραγωγής ή συνεδρίες ακρόασης. Επομένως, η σύσταση μιας ουσιαστικής ακολουθίας τραγουδιών είναι πολύ σημαντική στον τομέα της μουσικής, σε αντίθεση με τον τομέα των ταινιών. Επιπλέον, η επαναληπτική σύσταση, δηλαδή η πρόταση ενός κομματιού που έχει ήδη ακούσει ο χρήστης σε δεύτερο χρόνο, είναι πολλές φορές θεμιτή και εκτιμάται από αυτόν, σε αντίθεση με τις συστάσεων ταινιών που έχει δει ή προϊόντων που έχει ήδη αγοράσει.

Ταυτόχρονα, η καταναλωτική συμπεριφορά του ακροατή είναι καθοριστική, αφού τις περισσότερες φορές η μουσική καταναλώνεται παθητικά, δηλαδή παίζει στο παρασκήνιο ενώ ο χρήστης ασχολείται με κάποια άλλη δραστηριότητα οπότε και δεν δίνει ιδιαίτερη προσοχή στη μουσική. Για παράδειγμα η μουσική που παίζει στο σουπερμάρκετ ή σε άλλα καταστήματα και η μουσική που βάζουν κάποιοι όσο κάνουν δουλειές στο σπίτι. Αυτή η συμπεριφορά είναι κρίσιμη όταν λαμβάνεται έμμεση ανατροφοδότηση (implicit feedback) πχ. ένα τραγούδι να παίζει ολόκληρο χωρίς παράλειψη, καθώς εσφαλμένα θα θεωρηθεί ως θετική ανταπόκριση στο μουσικό κομμάτι. Η ενσωμάτωση πρόσθετων πληροφοριών με βάση τα συμφραζόμενα (context), όπως η δραστηριότητα του χρήστη, είναι μια λύση για την αντιμετώπιση αυτού του προβλήματος.

Τέλος, η μουσική έχει την ιδιαιτερότητα να προκαλεί εξαιρετικά δυνατά συναισθήματα. Από την άλλη πλευρά και τα συναισθήματα του χρήστη επηρεάζουν τις μουσικές του προτιμήσεις.

Λόγω αυτής της ισχυρής σχέσης μεταξύ μουσικής και συναισθημάτων, το πρόβλημα της αυτόματης περιγραφής της μουσικής με όρους συναισθημάτων είναι ένας ενεργός ερευνητικός τομέας, που συνήθως αναφέρεται ως αναγνώριση συναισθημάτων μουσικής (Music Emotion Recognition, MER). Αυτός ο τομέας αναλύεται παραπάνω στο κεφ. 2.1.7.

Μοντέλα συστάσεων που χρησιμοποιούνται στα MRS

Ακολουθώντας, διερευνούμε τις μεθόδους συστημάτων συστάσεων που χρησιμοποιούνται στον τομέα της μουσικής. Διαφορετικές μελέτες χρησιμοποιούν διαφορετικές μεθόδους, συμπεριλαμβανομένων των Collaborative filtering, Content-Based filtering, Hybrid filtering ενώ συχνά αναφέρεται επίσης και το Context-Based filtering.

Το συνεργατικό φιλτράρισμα (CF), όπως έχει ήδη αναφερθεί, παίρνει άλλα άτομα στην περιοχή ή τον κύκλο του χρήστη και κάνει προτάσεις σε αυτόν ανάλογα με το ενδιαφέρον των γειτόνων του. Για την παροχή συστάσεων αξιοποιεί τον αλγόριθμο kNN (k nearest neighbours). Το CF χρησιμοποιείται ευρέως σε εργασίες συστάσεων μεγάλης κλίμακας και έχει σχετικά καλή απόδοση. Ωστόσο, η προτίμηση στη μουσική είναι υποκειμενική, επομένως η υπόθεση πίσω από τη μέθοδο CF, ότι οι χρήστες με παρόμοια ακουστική συμπεριφορά έχουν παρόμοια γούστα στη μουσική, είναι ευάλωτη. Μεγάλο μειονέκτημα αυτής της μεθόδου φιλτραρίσματος είναι το πρόβλημα της ψυχρής εκκίνησης. [76], [75]

Από την άλλη πλευρά, το φιλτράρισμα βάσει περιεχομένου (CBF) εξερευνά την ομοιότητα μεταξύ χαρακτηριστικών των μουσικών κομματιών και προτείνει μουσική σύμφωνα με το ιστορικό ακρόασης. [76] Ακουστικά χαρακτηριστικά του τραγουδιού όπως η ένταση, ο ρυθμός και η χροιά αναλύονται για να προταθούν τραγούδια. Οι πιο συνηθισμένες μέθοδοι για τον υπολογισμό της ομοιότητας μεταξύ κομματιών είναι: η ομαδοποίηση K-μέσων (K-means clustering) και ο αλγόριθμος προσδοκίας-μεγιστοποίησης (expectation-maximization) με δειγματοληψία Monte Carlo. Το κύριο μειονέκτημα του μοντέλου βάσει περιεχομένου είναι ότι βασίζεται στην ορθότητα των πληροφοριών του κάθε στοιχείου. Ένα άλλο σημαντικό μειονέκτημα είναι ότι αυτή η τεχνική αποτυγχάνει να διαφοροποιήσει σημαντικές διαφορές μεταξύ παρόμοιων κατά τα άλλα τραγουδιών. [75]

Σύμφωνα με τους Deldjoo et al. [77] τα CB χαρακτηριστικά των μουσικών κομματιών μπορούν να κατηγοριοποιηθούν ευρέως σε: (1) χαρακτηριστικά υψηλού επιπέδου / μεταδεδομένα (π.χ. είδος, καλλιτέχνης, ετικέτες, Wikidata) και (2) χαρακτηριστικά χαμηλού επιπέδου, τα οποία εξάγονται από το βασικό ηχητικό σήμα (π.χ. τόνος, ηχώχρωμα ή τέμπο) χρησιμοποιώντας τεχνικές επεξεργασίας σήματος ήχου/μουσικής και μηχανικής εκμάθησης. Η δημιουργία ενός MRS που αξιοποιεί τα διαθέσιμα μεταδεδομένα, όπως ο καλλιτέχνης, το άλμπουμ και το έτος κυκλοφορίας είναι ίσως η πιο απλή προσέγγιση για τη δημιουργία ενός MRS. Ωστόσο, η αποκλειστική βάση σε τέτοια μεταδεδομένα θα οδηγήσει σε προβλέψιμες συστάσεις. Αντίθετα, η αξιοποίηση της εκφραστικής δύναμης των περιγραφικών χαρακτηριστικών που εξάγονται από τα ηχητικά και άλλα σήματα πολυμέσων (π.χ. μουσικά βίντεο κλιπ) μπορεί να δημιουργήσει πιο ενημερωμένες και λιγότερο ασήμαντες συστάσεις.

Στον εμπορικό κόσμο, οι ιστότοποι ροής μουσικής όπως το Spotify, η Apple Music και η Pandora χρησιμοποιούν διαφορετικές τεχνικές συστάσεων για να προτείνουν μουσική στους χρήστες τους. Σε γενικές γραμμές, χρησιμοποιείται το υβριδικό μοντέλο σύστασης και παραλλαγές αυτού. Όταν ένας χρήστης εγγράφεται για πρώτη φορά με τη διεύθυνση email του, το σύστημα προτάσεων λειτουργεί ζητώντας από τον χρήστη να εισάγει κάποιους καλλιτέχνες που μπορεί να του αρέσουν. Έτσι, με αυτήν την επιλογή του χρήστη και με βάση τα δεδομένα τοποθεσίας προτείνονται τα τραγούδια στον χρήστη. Στη συνέχεια, με την

πάροδο του χρόνου, καθώς ο χρήστης ακούει όλο και περισσότερα τραγούδια, το σύστημα προτάσεων χρησιμοποιεί το ιστορικό ακρόασης του χρήστη για να προτείνει τραγούδια. [75]

2.2.6 Ανάλυση Δημοσιεύσεων MRS

Στο σημείο αυτό θα παρουσιάσουμε κάποιες πρόσφατες μελέτες που έχουν γίνει αναφορικά με τα συστήματα συστάσεων μουσικής.

Στην έρευνα των Niyazon et al. [78] αναπτύχθηκε ένα CB σύστημα σύστασης μουσικής που βασίζεται στην ακουστική ομοιότητα των μουσικών κομματιών. Εξετάζονται δύο προσεγγίσεις: η πρώτη είναι μια κοινή προσέγγιση που χρησιμοποιεί ανάλυση ακουστικών χαρακτηριστικών και η δεύτερη περιλαμβάνει μεθόδους βαθιάς μάθησης και υπολογιστικής όρασης, με στόχο τη βελτίωση των αποτελεσμάτων του συστήματος συστάσεων. Στην πρώτη προσέγγιση, γίνεται αρχικά εξαγωγή των ακουστικών χαρακτηριστικών, στη συνέχεια δημιουργείται ένα διάνυσμα χαρακτηριστικών και τέλος υπολογίζεται η ομοιότητα μεταξύ των διανυσμάτων για να παραχθεί η λίστα των μουσικών συστάσεων. Στην δεύτερη προσέγγιση, χρησιμοποιούνται φασματογράμματα (spectrograms) ακουστικού σήματος ως γραφικές αναπαραστάσεις του ηχητικού σήματος που στη συνέχεια δίνονται ως είσοδος σε τεχνητό νευρωνικό δίκτυο και προσδιορίζεται η ομοιότητα των μουσικών συνθέσεων. Η χρήση νευρωνικού δικτύου βελτιώνει την απόδοση του συστήματος συστάσεων.

Στην έρευνα των Chen et al. [79] αρχικά ερευνάται το ιστορικό ακρόασης των χρηστών και τα δημογραφικά στοιχεία για να δημιουργηθεί μια ενσωμάτωση χρήστη (user embedding) που αντιπροσωπεύει τη μουσική προτίμηση του χρήστη. Με την ενσωμάτωση χρήστη και τα δεδομένα ήχου από τα κομμάτια που αρέσουν και τα δεν αρέσουν στον χρήστη, μπορεί να επιτευχθεί μια ενσωμάτωση ήχου για κάθε κομμάτι, χρησιμοποιώντας μετρική εκμάθησης με σιαμαία δίκτυα (Siamese Network, SM). Για ένα νέο κομμάτι, αποφασίζεται η καλύτερη ομάδα χρηστών που θα προταθεί υπολογίζοντας την ομοιότητα μεταξύ της ενσωμάτωσης ήχου του κομματιού και των διαφορετικών ενσωματώσεων χρηστών. Το προτεινόμενο σύστημα αποδίδει κορυφαίες επιδόσεις σε συστάσεις μουσικής βάσει περιεχομένου, δοκιμασμένο με εκατομμύρια χρήστες και κομμάτια.

Οι Zarzour et al. [80] στην εργασία τους, προτείνουν ένα νέο αλγόριθμο συστάσεων συνεργατικού φιλτραρίσματος που βασίζεται σε τεχνικές μείωσης διαστάσεων και ομαδοποίησης. Το προτεινόμενο σύστημα συστάσεων είναι δύο σταδίων. Το πρώτο στάδιο ονομάζεται δημιουργία μοντέλου εκτός σύνδεσης (offline model creation). Σε αυτό το στάδιο, το μοντέλο σύστασης δημιουργείται ομαδοποιώντας τις αξιολογήσεις των χρηστών ακολουθώντας τις προτιμήσεις τους, μειώνοντας τις διαστάσεις των δεδομένων και στη συνέχεια υπολογίζοντας τις ομοιότητες. Ο αλγόριθμος k-μέσων και η τεχνική Singular Value Decomposition (SVD) χρησιμοποιούνται και οι δύο σε αυτό το στάδιο για την ομαδοποίηση παρόμοιων χρηστών και τη μείωση των διαστάσεων, αντίστοιχα. Το δεύτερο στάδιο είναι η χρήση διαδικτυακού μοντέλου (online model utilization) όπου το μοντέλο που δημιουργήθηκε χρησιμοποιείται για την παραγωγή συστάσεων ακριβείας για έναν δεδομένο

ενεργό χρήστη. Τα πειραματικά αποτελέσματα δείχνουν ότι αυτή η νέα μέθοδος βελτιώνει σημαντικά την απόδοση των συστημάτων συστάσεων.

Ο Wang Wenzhen [81] έναν εξατομικευμένο αλγόριθμο προτάσεων μουσικής που βασίζεται σε υβριδική τεχνολογία συνεργατικού φιλτραρίσματος. Στο CF σύστημα προστίθενται κανόνες συσχέτισης και γονίδια μουσικής (όπως όνομα, καλλιτέχνης, γλώσσα αλλά και μελωδία, ρυθμός, ταχύτητα). Αναλύοντας τα ενδιαφέροντα και τις προτιμήσεις των χρηστών για διαφορετικά χαρακτηριστικά γονιδίων μουσικής, ο αλγόριθμος αναλύει διεξοδικά τη συμπεριφορά των χρηστών και χρησιμοποιεί την ομοιότητα των ενδιαφερόντων μεταξύ διαφορετικών χρηστών για να κατασκευάσει τη σχέση γειτονιάς μεταξύ τους. Η αξιολόγηση βασίστηκε στις μετρικές ακρίβειας και ανάκλησης και έγινε σύγκριση των αποτελεσμάτων με αυτά των απλών CF συστημάτων.

Η μελέτη των Girsang et al. [82] εστιάζει στην προσέγγιση συνεργατικού φιλτραρίσματος (CF) σε συνδυασμό με νευρωνικό δίκτυο, που ονομάζεται και νευρωνικό συνεργατικό φιλτράρισμα (neural CF, NCF). Σε αυτήν τη μελέτη, χρησιμοποιούνται 20.000 χρήστες, 6.000 τραγούδια και 470.000 αξιολογήσεις και στη συνέχεια προβλέπεται η βαθμολογία χρησιμοποιώντας τις CF και NCF προσεγγίσεις. Ο στόχος αυτής της μελέτης είναι να διαφοροποιήσει τα συστήματα συστάσεων με χρήση μόνο του CF από αυτά που κάνουν χρήση του NCF. Μέσω αυτής της έρευνας, διαπιστώθηκε ότι το NCF είναι καλύτερο από το user-based συνεργατικό φιλτράρισμα στη συγκέντρωση των playlist που πραγματικά θέλουν να ακούσουν οι χρήστες, αλλά απαιτεί περισσότερο χρόνο για τη δημιουργία του.

Ορισμένες από τις μελέτες που διερευνήθηκαν εξετάζουν συγκεκριμένες προκλήσεις που αντιμετωπίζουν τα συστήματα συστάσεων μουσικής. Στην εργασία των Gatzoura et al. [83] προτείνεται το HybA, ένα υβριδικό σύστημα συστάσεων για αυτόματη συνέχιση της λίστας αναπαραγωγής, που συνδυάζει λανθάνουσα κατανομή Dirichlet (Latent Dirichlet Allocation) και συλλογιστική βάσει περιπτώσεων (case-based reasoning). Το HybA προσδιορίζει τα σημασιολογικά χαρακτηριστικά μιας αρχικής λίστας αναπαραγωγή και επαναχρησιμοποιεί κάποια από τα παρόμοια χαρακτηριστικά, για να προτείνει σχετικές συνεχίσεις της λίστας αναπαραγωγής. Για να ξεπεραστεί το σημασιολογικό χάσμα μεταξύ των περιγραφών μουσικής και των προτιμήσεων των χρηστών, να προσδιοριστούν οι δομές της λίστας αναπαραγωγής και να αποτυπωθεί η ομοιότητα των τραγουδιών, χρησιμοποιείται ένα μοντέλο γραφήματος.

Οι Cao et al. [84] προτείνουν ένα σύστημα για την βελτίωση του προβλήματος της ψυχρής εκκίνησης στα MRS που βασίζεται σε αλγόριθμο ανίχνευσης κοινότητας. Με την προβολή του διμερούς δικτύου, υπολογίζοντας την ομοιότητα μεταξύ χρήστη και αντικειμένου, ο αλγόριθμος Lounvain χρησιμοποιείται για την ανίχνευση κοινότητας στο προβαλλόμενο δίκτυο, έτσι ώστε η νέα εγγραφή να ενημερωθεί στην αρχική κοινότητα και στη συνέχεια προτείνεται στο χρήστη η ομάδα που ανιχνεύθηκε. Με υψηλότερη ακρίβεια, σταθερή σταθερότητα και μικρότερο χρόνο λειτουργίας (σε σύγκριση με άλλους αλγόριθμους ψυχρής εκκίνησης), ο αλγόριθμος αυτός μπορεί να εφαρμοστεί με ασφάλεια στο σύστημα προτάσεων μουσικής ή στο πεδίο πρόβλεψης μουσικής.

Η έρευνά των Adiyansjah et al. [85] αφορά την ανάπτυξη ενός συστήματος προτάσεων μουσικής που μπορεί να δώσει συστάσεις με βάση την ομοιότητα των χαρακτηριστικών του ηχητικού

σήματος. Αυτή η έρευνα χρησιμοποιεί ένα συνελκτικό επαναλαμβανόμενο νευρωνικό δίκτυο (Convolutional Recurrent Neural Network, CRNN) για εξαγωγή χαρακτηριστικών και μια απόσταση ομοιότητας για να βρεθεί η ομοιότητα μεταξύ των χαρακτηριστικών. Τα αποτελέσματα δείχνουν ότι οι χρήστες προτιμούν προτάσεις που λαμβάνουν υπόψη τα είδη μουσικής σε σύγκριση με προτάσεις που βασίζονται αποκλειστικά στην ομοιότητα.

2.2.7 Συναισθήματα και MRS: Σχετικές έρευνες

Πλήθος ερευνών έχουν γίνει στα συστήματα συστάσεων μουσικής με βάση το συναίσθημα και αρκετές από αυτές χρησιμοποιούν ως μέσο συναισθημάτων τις εκφράσεις του προσώπου. Παρακάτω, αναφέρονται περιληπτικά έρευνες που έχουν γίνει πάνω σε αυτά τα συστήματα.

Οι James et al. [86] ανέπτυξαν ένα λογισμικό για την αναγνώριση των συναισθημάτων του χρήστη με βάση την έκφραση του προσώπου χρησιμοποιώντας Python και στη συνέχεια ενσωμάτωσαν τον κώδικα python στην υπηρεσία web, ώστε να παίζει μουσική με βάση το συναίσθημα του χρήστη. Το προτεινόμενο σύστημα χρησιμοποιεί κάμερα για να λάβει ως είσοδο την εικόνα από το ανθρώπινο πρόσωπο. Η διαδικασία ανίχνευσης προσώπου περιλαμβάνει τα ακόλουθα βήματα: Η προεπεξεργασία πραγματοποιείται χρησιμοποιώντας την πυραμίδα Gauss, στη συνέχεια χρησιμοποιείται Ιστόγραμμα προσανατολισμένων διαβαθμίσεων (Histogram of Oriented Gradients, HOG) για να περιγράψει το πρόσωπο μέσα στην εικόνα και τέλος ένας γραμμικός ταξινομητής κάνει την ταξινόμηση έτσι ώστε να απαιτείται λιγότερος υπολογιστικός χρόνος για την ανίχνευση προσώπου. Μετά από αυτό, μια συνάρτηση «πρόβλεψης» χρησιμοποιείται για την εξαγωγή των 68 σημείων ορόσημων προσώπου και ένας ταξινομητής SVM ελέγχει σε ποια κατηγορία συναισθήματος ανήκουν τα δεδομένα που εισάγονται (χαρούμενος, θυμωμένος, λυπημένος, έκπληκτος). Η απόδοση αυτού του ταξινομητή είναι περίπου 90-95%. Ολοκληρώνοντας, παίζονται στο χρήστη τα τραγούδια που έχουν ανατεθεί για το συγκεκριμένο συναίσθημα.

Η έρευνα των Iyer et al. [35] προτείνει ένα σύστημα που ονομάζεται «EmoPlayer», μια εφαρμογή Android, που προτείνει στον χρήστη μια λίστα τραγουδιών με βάση το τρέχον συναίσθημά του. Το σύστημα καταγράφει την εικόνα του χρήστη χρησιμοποιώντας κάμερα και εντοπίζει το πρόσωπό του. Στη συνέχεια, ανιχνεύει το συναίσθημα και κάνει μια λίστα με τραγούδια που θα βελτιώσουν τη διάθεσή του καθώς τα τραγούδια συνεχίζουν να παίζουν. Το EmoPlayer χρησιμοποιεί τον αλγόριθμο Viola Jones για την ανίχνευση προσώπου και τον ταξινομητή Fisherfaces για την ταξινόμηση συναισθημάτων (θυμωμένος, χαρούμενος, λυπημένος). Ο ταξινομητής συναισθημάτων εκπαιδεύτηκε χρησιμοποιώντας εικόνες από τη βάση δεδομένων CK+.

Οι Alrihaili et al. [87] στην έρευνά τους, προτείνουν ένα σύστημα συστάσεων για την αναγνώριση συναισθημάτων που είναι ικανό να ανιχνεύει τα συναισθήματα του χρήστη (μεταξύ χαρούμενου, λυπημένου, φυσικού ή έκπληκτου) και να προτείνει μια λίστα με τα κατάλληλα τραγούδια που μπορούν να βελτιώσουν τη διάθεσή του. Η εφαρμογή του προτεινόμενου συστήματος συστάσεων έχει ως εξής: Πρώτον, μια web κάμερα

χρησιμοποιείται για τη λήψη των εικόνων που θα χρησιμοποιηθούν ως είσοδος στο προτεινόμενο σύστημα. Στη συνέχεια, το πρόσωπο ανιχνεύεται μέσα στην εικόνα χρησιμοποιώντας τον αλγόριθμο Viola-Jones. Ακολούθως απαιτείται να εντοπιστεί το συναίσθημα του χρήστη, επομένως η μέθοδος ανάλυσης κύριου στοιχείου (Principal Component Analysis, PCA) χρησιμοποιείται για την κατασκευή του χώρου προσώπου και την επιλογή των ιδιοδιανυσμάτων που έχουν τις υψηλότερες ιδιοτιμές. Τέλος, η λίστα αναπαραγωγής μουσικής που αντιστοιχεί στο συναίσθημα που εντοπίστηκε παρουσιάζεται στον χρήστη.

Οι Babanne et al. [88] προτείνουν ένα σύστημα συστάσεων που βασίζεται σε συναισθήματα και λειτουργεί σε πραγματικό χρόνο. Σε αυτό το σύστημα, εξάγονται τα χαρακτηριστικά του προσώπου μέσω εισόδων βίντεο από τους χρήστες. Τα χαρακτηριστικά του προσώπου υπολογίζονται μέσω των αλγορίθμων LBP και Haar Cascade. Οι εκφράσεις του προσώπου (χαρούμενος, θυμωμένος, λυπημένος, φοβισμένος, αδιασμένος, ουδέτερος, έκπληκτος) παράγονται από τις κινήσεις των μυών του προσώπου με τρόπο τέτοιο που θα μπορούσε να αλλάξει τη θέση και το σχήμα των χαρακτηριστικών του προσώπου όπως τα μάτια, μύτη, χείλια. Μέσω αυτών ανιχνεύονται τα συναισθήματα του χρήστη. Αυτό το σύστημα ξεχωρίζει καθώς εντοπίζει επίσης την ηλικία και το φύλο του ατόμου και χρησιμοποιεί αυτές τις πληροφορίες για τη σύσταση.

Οι Helmholtz et al. [89] ανέπτυξαν το σύστημα "Moosic" ως μέρος του ερευνητικού τους έργου. Η εφαρμογή χρησιμοποιεί το Spotify API για να δημιουργήσει μια λίστα αναπαραγωγής με βάση τα είδη μουσικής που επιλέγει ο χρήστης και τη συναισθηματική του κατάσταση. Πιο συγκεκριμένα, ο χρήστης επιλέγει τη συναισθηματική του κατάσταση από μια χρωματικά κωδικοποιημένη περιοχή εισόδου με βάση το μοντέλο συναισθημάτων του Russell ή του Thayer και επιλέγει επίσης τα είδη που προτιμά και τη δημοτικότητα των τραγουδιών. Στη συνέχεια, ελάχιστες και μέγιστες τιμές για το σθένος (valence) και την ενέργεια (energy) ορίζονται σύμφωνα με το επιλεγμένο συναίσθημα του χρήστη και συνδυάζονται με τα επιλεγμένα είδη σε ένα αίτημα που στέλνεται στο Spotify API. Το Spotify επιστρέφει μια λίστα αναπαραγωγής με 20 τραγούδια που ταιριάζουν στο αίτημα, η οποία παίζεται τυχαία.

Στο άρθρο των Gilda et al. [90] παρουσιάζεται ένα πρόγραμμα αναπαραγωγής μουσικής πολλαπλών πλατφορμών με βάση το συναίσθημα, το EMP, το οποίο προτείνει μουσική με βάση τη διάθεση του χρήστη σε πραγματικό χρόνο. Το EMP περιέχει τρεις ενότητες: τη μονάδα συναισθημάτων, τη μονάδα ταξινόμησης μουσικής και τη μονάδα συστάσεων. Η μονάδα συναισθημάτων παίρνει μια εικόνα του προσώπου του χρήστη ως είσοδο και χρησιμοποιεί αλγόριθμους βαθιάς μάθησης για να προσδιορίσει τη διάθεσή του. Πιο συγκεκριμένα, ένα CNN εφαρμόζεται για την κατασκευή υπολογιστικού μοντέλου που ταξινομεί τα συναισθήματα σε 4 κατηγορίες (χαρά, λύπη, θυμός, ουδέτερη) με ακρίβεια 90,23%. Η μονάδα ταξινόμησης μουσικής αντιστοιχίζει κάθε μουσικό κομμάτι σε μία από τις κατηγορίες συναισθημάτων που ορίστηκαν. Για το σκοπό αυτό χρησιμοποιούνται αλγόριθμοι εξαγωγής ακουστικών χαρακτηριστικών, όπως LibROSA και audiopitch. Με βάση τα εξαγόμενα χαρακτηριστικά εκπαιδεύεται ένα ANN που ταξινομεί τα τραγούδια στις 4 κατηγορίες διάθεσης επιτυγχάνοντας αποτέλεσμα 97,69%. Η μονάδα συστάσεων, τέλος, παράγει μια λίστα αναπαραγωγής με τραγούδια σχετικά με τη διάθεση του χρήστη.

Στην εργασία των Moscato et al. [91], περιγράφεται μια νέα τεχνική σύστασης μουσικής που βασίζεται στον προσδιορισμό της προσωπικότητας, των διαθέσεων και των συναισθημάτων του χρήστη, με βάση ψυχολογικές παρατηρήσεις για την ανάλυση της συμπεριφοράς των χρηστών σε ένα κοινωνικό περιβάλλον. Συγκεκριμένα, η προσωπικότητα και η διάθεση των χρηστών έχουν ενσωματωθεί σε μια προσέγγιση φιλτραρίσματος βάσει περιεχομένου. Το προτεινόμενο σύστημα λειτουργεί ακολουθώντας τρία βήματα. Πρώτον, η αναγνώριση της προσωπικότητας του χρήστη: η προσωπικότητα του χρήστη υπολογίζεται με βάση τα Big Five components (OCEAN), λαμβάνοντας υπόψη τη συμπεριφορά του χρήστη στα κοινωνικά δίκτυα. Τα χαρακτηριστικά της προσωπικότητας του χρήστη χαρτογραφούνται στη συνέχεια στον χώρο συναισθημάτων του Mehrabian (PAD). Δεύτερο βήμα, η ανίχνευση διάθεσης: τα αντικείμενα που έχουν προσπελαστεί τελευταία αναλύονται, για να εντοπιστεί η τρέχουσα διάθεση του χρήστη. Τρίτον, η σύσταση βάσει περιεχομένου: τα μουσικά κομμάτια προτείνονται στους χρήστες με βάση το περιεχόμενό τους και τη σχετική ομοιότητα. Χαρτογραφούνται σε συγκεκριμένα σημεία του χώρου PAD, χρησιμοποιώντας το Mel-Spectrogram του τραγουδιού που δίνεται ως είσοδος σε ένα VGG CNN, για τον υπολογισμό των τιμών διέγερσης και σθένους. Τέλος, τα στοιχεία προτείνονται στους χρήστες με βάση ένα κριτήριο ελάχιστης Ευκλείδειας απόστασης ως προς τη διάθεση του χρήστη.

Στη μελέτη των Jazi et al., [92] προτείνεται ένα σύστημα σύστασης μουσικής που προτείνει μουσική με βάση τα πατήματα πλήκτρων των χρηστών και τα μοτίβα των κλικ του ποντικιού. Το προτεινόμενο σύστημα αντιστοιχίζει αυτά τα μοτίβα απευθείας στην αγαπημένη μουσική του χρήστη, χωρίς να επισημαίνει το τρέχον συναίσθημά του. Τα αποτελέσματα δείχνουν ότι παρόλο που αυτό το σύστημα δεν χρησιμοποιεί καμία πρόσθετη συσκευή, είναι πολύ ακριβές σε σύγκριση με άλλες μεθόδους. Η αρχιτεκτονική του συστήματος αυτού, EMA-based EMotion-Aware (EMAEMA), είναι η εξής: Τα έμμεσα δεδομένα εισαγωγής αποτελούνται από τα πλήκτρα και τα μοτίβα κλικ του ποντικιού χρήστη. Η μέθοδος EMA-EMA βασίζεται στο συνεργατικό φιλτράρισμα, όπου η ομοιότητα των χρηστών μετράται με βάση τον τρόπο με τον οποίο αλληλεπιδρούν με το ποντίκι και το πληκτρολόγιο. Με αυτόν τον τρόπο δημιουργούνται διανύσματα αλληλεπίδρασης για κάθε χρήστη, τα οποία ανανεώνονται με την πάροδο του χρόνου. Για τη σύσταση των τραγουδιών, το πιο πρόσφατο διάνυσμα αλληλεπίδρασης του χρήστη θεωρείται ο δείκτης της τρέχουσας συναισθηματικής του κατάστασης.

Οι Wang et al. [93] στην εργασία τους, προτείνουν μια προσέγγιση για τη βελτίωση της πρόβλεψης της προτίμησης ενός χρήστη που ενσωματώνει τα στοιχεία του περιβάλλοντος (context) και τα συναισθήματα και στοχεύει στο να παρέχει στους χρήστες μια διαισθητική και ευχάριστη εμπειρία ακρόασης. Αρχικά, συλλέγονται το συναίσθημα, η τοποθεσία ακρόασης και ο χρόνος (η στιγμή της μέρας) ακρόασης των τραγουδιών ως βάση του μηχανισμού σύστασης. Το σύστημα υιοθετεί τη μέθοδο ταξινόμησης μουσικής βάσει συναισθηματικού περιεχομένου που προτείνεται από τον Meyers (η μέθοδος Meyers αντιστοιχίζει το συναίσθημα ενός τραγουδιού στον κύκλο του Hevner - Hevner's adjective cycle) για να καθορίσει τις σχέσεις μεταξύ των συναισθημάτων και των μουσικών χαρακτηριστικών ενός τραγουδιού. Τέλος, συνιστώνται λίστες αναπαραγωγής που ταιριάζουν με τα συναισθήματα, τις πληροφορίες και τις προτιμήσεις των χρηστών, εκτελώντας υπολογισμό βάρους, φιλτράρισμα μουσικής και μέτρηση ομοιότητας.

2.2.8 Spotify

Το Spotify [94] είναι μια ψηφιακή υπηρεσία μουσικής, podcast και βίντεο που δίνει πρόσβαση σε εκατομμύρια τραγούδια και σε άλλο περιεχόμενο από δημιουργούς/καλλιτέχνες σε όλο τον κόσμο. Οι βασικές λειτουργίες όπως η αναπαραγωγή μουσικής είναι εντελώς δωρεάν, αλλά μπορεί επίσης να γίνει αναβάθμιση σε Spotify Premium όπου ο χρήστης μπορεί να ακούσει μουσική χωρίς διαφημίσεις, με βελτιωμένη ποιότητα ήχου και να αποθηκεύσει μουσική για ακρόαση εκτός σύνδεσης. Ανάμεσα σε πολλά πράγματα που μπορεί να κάνει ο χρήστης στο Spotify, είναι να λαμβάνει προτάσεις με βάση το γούστο του και να δημιουργεί συλλογές μουσικής και podcast. Το Spotify ξεκίνησε να λειτουργεί το 2008 και εν έτη 2022 έχει πάνω από 430 εκ. ενεργούς χρήστες, εκ των οποίων 188 εκ. συνδρομητές στο Premium και προσφέρει πάνω από 80 εκ. μουσικά κομμάτια, συμπεριλαμβανομένων πάνω από 4 εκ. podcasts [95].



Αυτός ο τεράστιος όγκος δεδομένων απαιτεί καλύτερους μηχανισμούς αναζήτησης και ανακάλυψης από την απλή αναζήτηση ανά καλλιτέχνη, τίτλο ή είδος. Για το λόγο αυτό το Spotify έχει δημιουργήσει ένα ιδιαίτερο σύστημα συστάσεων που βασίζεται σε μεγάλο βαθμό στα δεδομένα και αξιοποιεί αρκετά μοντέλα μηχανικής μάθησης και αλγορίθμους για να παράγει τις καλύτερες προτάσεις. Δυστυχώς, δεν έχει δημοσιοποιηθεί κάποιο πλήρες documentation που να περιγράφει το πως λειτουργεί το σύστημα συστάσεων, αλλά υπάρχουν τα εκτεταμένα δημόσια αρχεία Spotify R&D [96] της εταιρείας, το API της και αρκετές λεπτομέρειες είναι γνωστές ώστε να μπορούμε να εξάγουμε συμπεράσματα.

BaRT

Το πρώτο πράγμα που παρατηρεί κανείς στο Spotify είναι η αρχική οθόνη, που είναι γεμάτη με πολλές προσωποποιημένες προτάσεις και λίστες αναπαραγωγής, όπως “Δημιουργήθηκε για σένα”, “Τα κορυφαία σου mix”, “Άκουσες πρόσφατα”, “Οι μεγαλύτερες επιτυχίες του σήμερα” κ.α. Η αρχική οθόνη δημιουργείται και επιμελείται από ένα σύστημα AI που ονομάζεται Bandits for Recommendations as Treatments ή απλά γνωστό ως BaRT. Το BaRT αποτελείται από δύο βασικές ιδέες: Εκμετάλλευση (exploitation) και Εξερεύνηση (exploration). [97]

Εκμετάλλευση σημαίνει παροχή συστάσεων που βασίζεται σε προηγούμενες συνήθειες ακρόασης (ιστορικό) αλλά και άλλες πληροφορίες που το σύστημα έχει συλλέξει από το χρήστη, όπως παραλείψεις και αγαπημένα τραγούδια. Είναι η συνήθης λειτουργία των συστημάτων που βασίζονται στο συνεργατικό φιλτράρισμα. Ένα πρόβλημα των συστημάτων

που βασίζονται μόνο στην εκμετάλλευση είναι η συνάφεια του στοιχείου. Εάν υπάρχουν ελάχιστα δεδομένα για έναν χρήστη ή ένα αντικείμενο (τραγούδι), τότε το σύστημα είναι αβέβαιο για το αν θα το προτείνει (εκμετάλλευση) ή όχι (αγνοεί).

Από την άλλη, η εξερεύνηση μελετά τον υπόλοιπο κόσμο και προτείνει τραγούδια με βάση εξωτερικές πληροφορίες, όπως τι άκουσαν άλλοι χρήστες, τι λίστες αναπαραγωγής δημιούργησαν, ποια είναι η τάση κλπ. Για να λύσει το πρόβλημα της εκμετάλλευσης, η εξερεύνηση συνιστά περιεχόμενο που είναι αβέβαιη προβλεπόμενη επιλογή των χρηστών με σκοπό τη συλλογή περισσότερων πληροφοριών. Η σημασία της εξερεύνησης έχει αναγνωριστεί τα τελευταία χρόνια, ιδιαίτερα σε συστήματα με νέους χρήστες, νέα στοιχεία, μη σταθερές προτιμήσεις και χαρακτηριστικά.

Με άλλα λόγια, εάν ένα νέο τραγούδι δεν έχει παιχτεί αρκετά ακόμα, χρειάζονται δεδομένα για να επαληθευτεί εάν το τραγούδι έχει δυνατότητες. Έτσι, το BaRT του Spotify προτείνει αυτά τα τραγούδια μέσω εξερεύνησης και συλλέγει πληροφορίες για το νέο τραγούδι. Ένα τραγούδι θεωρείται θετική σύσταση μετά από 30 δευτερόλεπτα (30-sec rule). Αυτό σημαίνει ότι αν ο χρήστης ακούσει ένα τραγούδι για λιγότερο από μισό λεπτό, μετράει ως αρνητικό αλλιώς θεωρείται θετικό για τη σύσταση.

Το μοντέλο BaRT είναι σε θέση να μαθαίνει και να προβλέπει την ικανοποίηση. Βασίζεται στην Ενισχυτική μάθηση (Reinforcement learning) και προσπαθεί να λάβει σχόλια για να μεγιστοποιήσει την ικανοποίηση των χρηστών και να διορθώσει τις προβλεπόμενες συστάσεις.

Δεδομένα

Το Spotify έχει στο επίκεντρο του ενδιαφέροντος δύο κατηγορίες: τους δημιουργούς (artists) και τους χρήστες (users). Από αυτούς μαζεύει πληροφορίες και δεδομένα που χρειάζεται ώστε να δημιουργήσει τις κατάλληλες συστάσεις. Όσον αφορά τους δημιουργούς, το Spotify αποθηκεύει όλα τα δεδομένα που έχουν εισάγει: ονόματα τραγουδιών, περιγραφή, είδος, εικόνες, στίχους και αρχεία τραγουδιών. Για τους χρήστες, το Spotify συλλέγει και παρακολουθεί τα ακόλουθα δεδομένα: το ιστορικό ακρόασης, τις παραλήψεις, τη συχνότητα αναπαραγωγής ενός τραγουδιού, τις λίστες αναπαραγωγής που αποθηκεύτηκαν, τη λήψη μουσικής, τις κοινωνικές αλληλεπιδράσεις όπως κοινόχρηστες λίστες αναπαραγωγής ή κοινή χρήση μουσικής και άλλα. Δίπλα σε αυτές τις δύο πηγές «εσωτερικών» δεδομένων, το Spotify πιθανότατα χρησιμοποιεί και εξωτερικά δεδομένα, όπως άρθρα, αναρτήσεις σε blog ή άλλα δεδομένα κειμένου σχετικά με τραγούδια ή καλλιτέχνες. [97]

Για τη δημιουργία του profile χρήστη το Spotify λαμβάνει δεδομένα από το feedback των χρηστών. Το feedback μπορεί να χωριστεί σε δύο κύριες κατηγορίες: Ρητά (explicit) σχόλια όπως: αποθηκεύσεις βιβλιοθήκης, προσθήκες playlist, κοινοποιήσεις, παραλείψεις, επίσκεψη στη σελίδα καλλιτέχνη/άλμπουμ, καλλιτέχνες που ακολουθεί. Και άρρητο (implicit) feedback όπως: διάρκεια συνεδριών ακρόασης, playthrough κομματιού και επαναλαμβανόμενες ακρόασεις. [98]

Τα συστήματα συστάσεων μπορούν να χωριστούν σε δύο διαφορετικές κατηγορίες: συνεργατικό φιλτράρισμα (CF) και φιλτράρισμα βάσει περιεχομένου (CBF). Το Spotify

χρησιμοποιεί και τους δύο αυτούς αλγόριθμους, για να δώσει γνώριμες αλλά ακόμα φρέσκοις λίστες playlist. Το Spotify χρησιμοποιεί επίσης την Επεξεργασία Φυσικής Γλώσσας (NLP) για την ανάλυση ειδήσεων, άρθρων και blog που γράφτηκαν στον ιστό σχετικά με συγκεκριμένα τραγούδια ή καλλιτέχνες. [99]

Collaborative Filtering (CF)

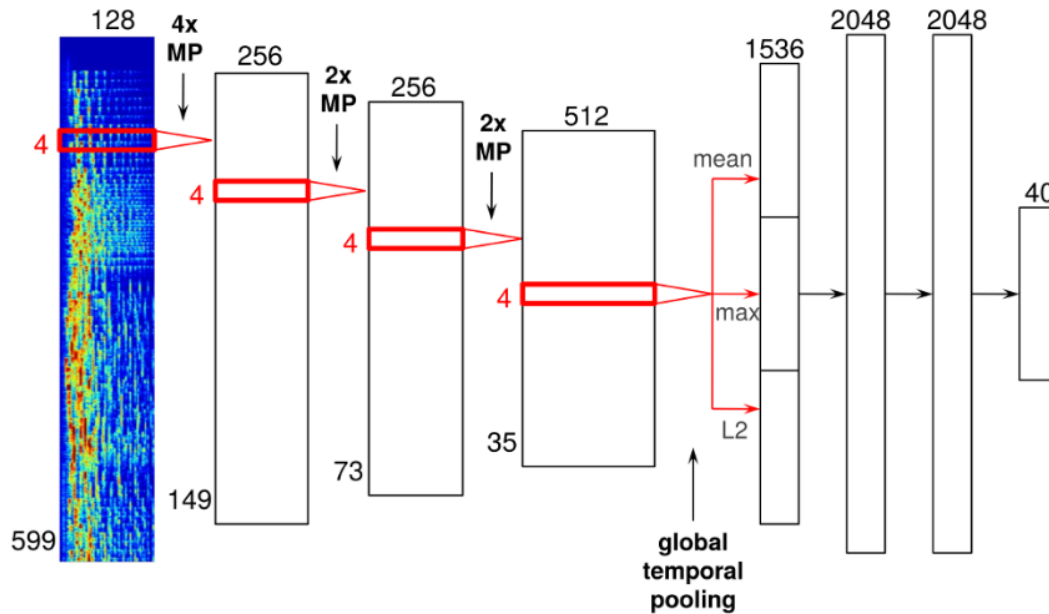
Η λογική του συνεργατικού φιλτραρίσματος είναι ότι μπορούμε να κατανοήσουμε τα τραγούδια που πρέπει να προτείνουμε σε έναν χρήστη κοιτάζοντας τι ακούνε άλλοι χρήστες με παρόμοια γούστα. Για το σκοπό αυτό δημιουργείται ένας τεράστιος πίνακας χρήστη-τραγουδιού που περιέχει το πλήθος των φορών που ο χρήστης έχει ακούσει το κάθε τραγούδι. Αυτός ο πίνακας, που ονομάζεται πίνακας αξιολόγησης (R) μετασχηματίζεται σε δύο πίνακες, τον πίνακα προτιμήσεων (P) και τον πίνακα εμπιστοσύνης (C). Ο πίνακας R παραγοντοποιείται στο γινόμενο δύο πινάκων (X και Y) χρησιμοποιώντας τον αλγόριθμο παραγοντοποίησης σταθμισμένου πίνακα (Weighted Matrix Factorization, WMF). Οι συστάσεις για κάθε χρήστη γίνονται βρίσκοντας τα K πλησιέστερα διανύσματα τραγουδιού για κάθε διάνυσμα χρήστη, χρησιμοποιώντας τον αλγόριθμο του πλησιέστερου γείτονα. Ομοίως, παρόμοια τραγούδια μπορούν να βρεθούν χρησιμοποιώντας τον ίδιο αλγόριθμο σε κάθε διάνυσμα τραγουδιού.

Σε αυτή τη μέθοδο υπάρχει πρόβλημα όταν το τραγούδι δεν έχει ακουστεί ξανά (cold-start problem). Επίσης, επειδή βασίζεται μόνο στα δεδομένα ακρόασης, τα δημοφιλή τραγούδια τείνουν να μονοπωλούν τις συστάσεις. Οπότε, αυτό το σύστημα συστάσεων αποτυγχάνει για νέα και μη δημοφιλή κομμάτια.

Content-Based Filtering (CBF)

Το φιλτράρισμα βάσει περιεχομένου μπορεί να λύσει αυτό το πρόβλημα μέσω της ανάλυσης ακατέργαστων σημάτων ήχου. Το raw audio είναι δύσκολο να αναλυθεί, επομένως το Spotify επεξεργάζεται αυτόν τον ακατέργαστο ήχο μετατρέποντάς τον σε φασματογράφημα mel (mel spectrogram) και περνώντας τον μέσω ενός συνελκτικού νευρωνικού δικτύου (CNN). Ένα φασματογράφημα mel είναι μια αναπαράσταση χρόνου-συχνότητας του ήχου, όπου οι συχνότητες μετατρέπονται στην κλίμακα mel, μια μη γραμμική κλίμακα που προσέγγιζε πολύ την ανθρώπινη ακουστική απόκριση. Το νευρωνικό δίκτυο που χρησιμοποιείται δημιουργήθηκε από έναν υπάλληλο του Spotify, τον Sander Dieleman, το 2014. Το δίκτυο Dieleman αποτελείται από 4 συνελκτικά επίπεδα και 3 βαθιά επίπεδα όπως φαίνεται και στην Εικ. 16.

Αφού το φασματογράφημα περάσει μέσα από αυτό το δίκτυο, δίνει στην έξοδο μια ερμηνεία του τραγουδιού, που περιλαμβάνει χαρακτηριστικά όπως η υπογραφή εκτιμώμενου χρόνου, το πλήκτρο (key), η λειτουργία (mode), ο ρυθμός (tempo) και η ένταση (loudness). Έτσι, όταν διαπιστωθεί ότι ένα νέο τραγούδι έχει παρόμοιες παραμέτρους με άλλα τραγούδια που αρέσουν στο χρήστη, το Spotify το προσθέτει στη λίστα αναπαραγωγής του.



Εικόνα 16 : Δίκτυο Dieleman [97]

Natural Language Processing (NLP)

Η επεξεργασία φυσικής γλώσσας χρησιμοποιείται για την εξαγωγή σημασιολογικών πληροφοριών που περιγράφουν το κομμάτι/καλλιτέχνη από περιεχόμενο κειμένου που σχετίζεται με τη μουσική. Εφαρμόζονται σε: ανάλυση στίχων, ιστοσελίδες, μέσα κοινωνικής δικτύωσης, λίστες αναπαραγωγής που δημιουργούνται από χρήστες (όνομα και περιγραφή) κ.α. Το NLP αναλύει τα παραπάνω για να συντάξει μια λίστα με τους πιο συχνά χρησιμοποιούμενους περιγραφικούς όρους για ένα συγκεκριμένο τραγούδι ή καλλιτέχνη. Καθένας από αυτούς τους όρους (που ονομάζονται cultural vectors), σχετίζεται με ένα βάρος που ποσοτικοποιεί τη σχετική σημασία του για ένα δεδομένο τραγούδι ή καλλιτέχνη.

Για να πετύχει όλα τα παραπάνω το Spotify έχει αγοράσει διάφορα εξωτερικά συστήματα. Το 2014 το Spotify εξαγόρασε την Echonest, μια start-up εταιρία για τη μουσική ευφυΐα. Δίπλα στο Echonest, το Spotify εξαγόρασε το 2015 τη Seed Scientific, μια εταιρεία συμβούλων επιστήμης δεδομένων και ανάλυσης για να αποκτήσει γνώση και εξειδίκευση χωρίς εξωτερικές παρεμβάσεις. Το 2017 το Spotify ενσωμάτωσε τη Sonalytic, μια start-up ανίχνευσης ήχου (η ανίχνευση ήχου αφορά λειτουργίες όπως αυτή του Shazam). Επιπλέον, τα συστήματα ανίχνευσης ήχου μπορούν να χρησιμοποιηθούν για την πρόληψη παραβιάσεων πνευματικών δικαιωμάτων. Το Spotify χρησιμοποιεί ανίχνευση ήχου για να βελτιώσει την εξατομίκευση σε λίστες αναπαραγωγής και σε τραγούδια, για να "ταιριάσει τραγούδια με συνθέσεις και να βελτιώσει το σύστημα δεδομένων δημοσίευσής του". [97]

Συμπερασματικά, το σύστημα συστάσεων του Spotify είναι ένα εξαιρετικά περίπλοκο σύστημα, με δεκάδες (αν όχι εκατοντάδες) αλγόριθμους και μοντέλα ML που χρησιμοποιούνται σε διάφορα επίπεδα, όλα μαζί για να δημιουργήσουν μια από τις πιο προηγμένες εμπειρίες συστάσεων στην αγορά ροής μουσικής. [98]

3. Πρακτικό Μέρος

Στο πρακτικό μέρος της εργασίας αυτής, θα γίνει παρουσίαση της εφαρμογής που αναπτύχθηκε με σκοπό την αναγνώριση συναισθήματος από εικόνες προσώπου και την σύσταση μουσικών κομματιών ανάλογα με τη διάθεση του χρήστη. Προτού περάσουμε στην παρουσίαση της εφαρμογής όμως θα πρέπει να διερευνήσουμε τη λειτουργία και τη χρησιμότητα των εργαλείων που χρησιμοποιήθηκαν για την ανάπτυξή της. Τα εργαλεία αυτά περιλαμβάνουν: το Flutter για την ανάπτυξη του κώδικα της εφαρμογής, το Py-Feat για την αναγνώριση συναισθημάτων από εικόνες προσώπου και το Spotify API για τις συστάσεις μουσικών κομματιών.

3.1 Flutter

Το Flutter [100] είναι ένα εργαλείο ανάπτυξης λογισμικού UI ανοιχτού κώδικα που δημιουργήθηκε από την Google. Χρησιμοποιείται για την δημιουργία mobile, web και desktop εφαρμογών από μια ενιαία βάση κώδικα. Οι εφαρμογές που δημιουργούνται στο Flutter χρησιμοποιούν τη γλώσσα προγραμματισμού Dart [101] της Google.



Μερικά ακόμα στοιχεία που είναι χρήσιμο να γνωρίζουμε σε σχέση με το Flutter είναι τα ακόλουθα [102]:

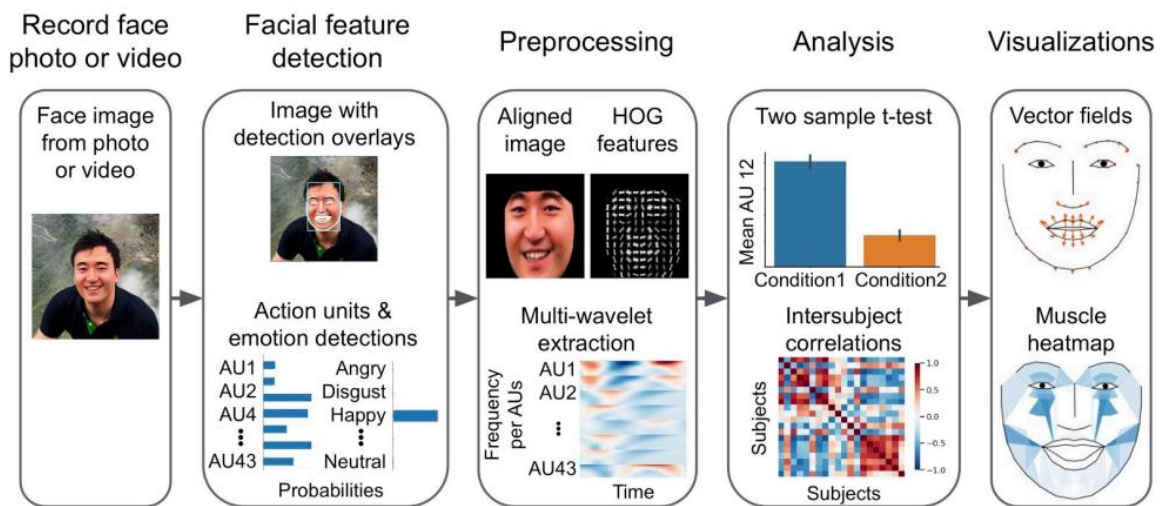
- Μπορεί να λειτουργήσει σε διάφορους editor ενώ παρέχει plugins για Android Studio, IntelliJ IDEA και VS Code.
- Εμπεριέχει ένα σετ υψηλής ποιότητας Material Design και Cupertino (σε στυλ iOS) widgets, layouts και themes. Φυσικά, αυτά τα γραφικά στοιχεία είναι μόνο για αρχή, αφού το Flutter έχει σχεδιαστεί για να διευκολύνει τη δημιουργία καινούργιων widgets από τους developers ή την προσαρμογή των υπαρχόντων.
- Υποστηρίζει πλήρως το Material Design.
- Παρέχει επίσης API για τη σύνταξη unit και integration tests.
- Συνοδεύεται από το Flutter DevTools (ονομάζεται επίσης Dart DevTools) για εντοπισμό σφαλμάτων (debugging).

Περισσότερες πληροφορίες για την εγκατάσταση του Flutter και την δημιουργία του πρώτου project βρίσκονται στο Παρ. 2 ή και στη σελίδα Get Started [103] του Flutter documentation.

3.2 Py-Feat

Το Python Facial Expression Analysis Toolbox (Py-Feat) [104] είναι μια εργαλειοθήκη Python ανοιχτού κώδικα που παρέχει ένα ολοκληρωμένο σύνολο εργαλείων και μοντέλων για την εύκολη ανίχνευση εκφράσεων προσώπου (Action Units, συναισθήματα, ορόσημα προσώπου) από εικόνες και βίντεο, την προεπεξεργασία και ανάλυση δεδομένων εκφράσεων προσώπου και την οπτικοποίηση δεδομένων εκφράσεων προσώπου.

Υπάρχουν και άλλα εργαλεία και αλγόριθμοι που εξάγουν Action Units, facial landmarks και συναισθήματα από εκφράσεις προσώπου με μεγάλη ταχύτητα και ακρίβεια, όπως OpenFace, iMotions, Affectiva, ή Noldus FaceReader. Παρόλα αυτά, το Py-Feat αναπτύχθηκε για να δημιουργηθεί ένα δωρεάν, ανοιχτού κώδικα και εύκολο στη χρήση εργαλείο για εργασία με δεδομένα εκφράσεων προσώπου.



Εικόνα 17: Διαδικασία ανάλυσης χαρακτηριστικών προσώπου με χρήση του Py-Feat. [30]

Το Py-Feat περιλαμβάνει δύο βασικά modules [30]:

Detector module

Το Py-Feat περιλαμβάνει μια μονάδα ανιχνευτή (Detector module) για την ανίχνευση χαρακτηριστικών έκφρασης προσώπου (δηλαδή, ανίχνευση προσώπων, ορόσημα προσώπου, ενεργοποιήσεις AU, συναισθηματικές εκφράσεις) από φωτογραφίες και βίντεο προσώπων.

Η μονάδα ανιχνευτή προσφέρει πολλά μοντέλα για την ανίχνευση καθενός από τα ακόλουθα χαρακτηριστικά του προσώπου: (α) εύρεση προσώπου σε εικόνα ή σε ένα καρέ από βίντεο, (β) εντοπισμός ορόσημων (landmarks) προσώπου, (γ) ανίχνευση ενεργοποιήσεων των AUs των μυών του προσώπου, (δ) ανίχνευση επιδείξεων κανονικών εκφράσεων συναισθημάτων και (ε) εκτίμηση πόζας προσώπου/κεφαλιού. Οι χρήστες είναι σε θέση να αποφασίζουν

ποιους αλγόριθμους θα χρησιμοποιήσουν για κάθε εργασία ανίχνευσης με βάση τις ανάγκες τους για ακρίβεια και ταχύτητα.

(α) Για την ανίχνευση προσώπου, το Py-Feat περιλαμβάνει τρεις δημοφιλείς ανιχνευτές προσώπου, συμπεριλαμβανομένων των Faceboxes, του Multi-task Convolutional Neural Network (MTCNN) και του RetinaFace. Αυτοί οι ανιχνευτές χρησιμοποιούνται ευρέως σε άλλα λογισμικά ανοιχτού κώδικα και είναι γνωστό ότι επιτυγχάνουν γρήγορα και ακριβή αποτελέσματα ανίχνευσης προσώπου ακόμη και για μερικώς κρυμμένα ή μη μετωπιαία πρόσωπα. Τα αποτελέσματα ανίχνευσης προσώπου αναφέρονται ως ένα bounding box του προσώπου, συμπεριλαμβανομένων βαθμολογιών εμπιστοσύνης (confidence scores).

(β) Τα ορόσημα προσώπου είναι σημεία που προσδιορίζονται στο χώρο της εικόνας και περιγράφουν το σαγόνι, το στόμα, τη μύτη, τα μάτια και τα φρύδια ενός προσώπου. Η απόσταση και οι γωνιακές σχέσεις μεταξύ των ορόσημων μπορούν να χρησιμοποιηθούν για να συναχθεί η συναισθηματική κατάσταση. Το Py-Feat προσφέρει τρεις ανιχνευτές ορόσημων προσώπου, συμπεριλαμβανομένων των αλγορίθμων Practical Facial Landmark Detector (PFLD), MobileNets και MobileFaceNets. Ο ανιχνευτής ορόσημων προσώπου λαμβάνει το bounding box του προσώπου από τον ανιχνευτή προσώπου ως είσοδο για να εκτιμήσει τις συντεταγμένες x και y των 68 facial landmarks.

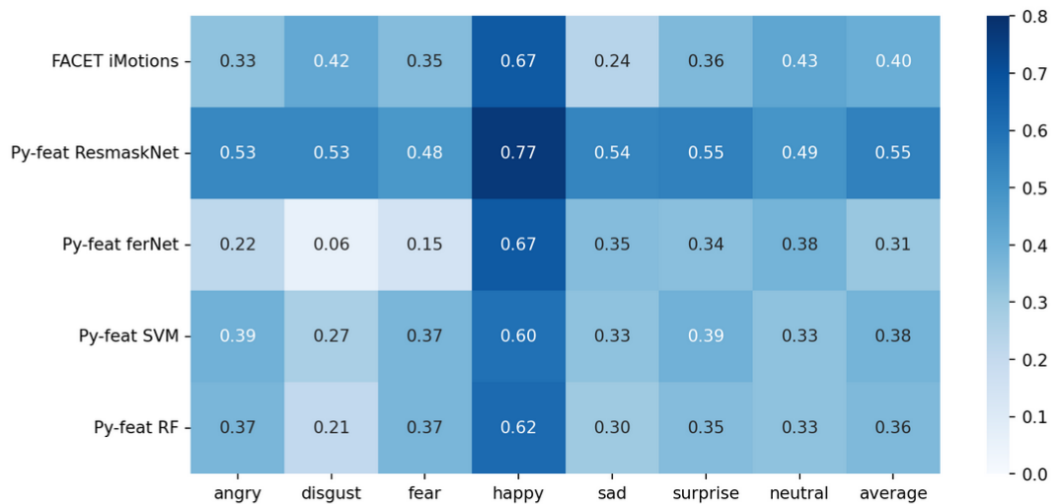
(γ) Οι μονάδες δράσης (AUs) είναι τα δομικά στοιχεία των εκφράσεων του προσώπου όπου κάθε αριθμός AU αντιστοιχεί σε μια συγκεκριμένη κίνηση των μυών του προσώπου. Ο συνδυασμός των μονάδων δράσης μπορεί να υποδηλώνει μια συναισθηματική έκφραση. Το Py-Feat παρέχει τέσσερις ανιχνευτές AU, το JAA-Net neural network model, ένα Random Forest classifier (Feat-RF), έναν γραμμικό Support Vector Machines classifier (Feat-SVM), και ένα logistic regression classifier (Feat-Logistic).

(δ) Οι ανιχνευτές συναισθημάτων εκπαιδεύονται σε στημένες και φυσικές εικόνες εκφράσεων προσώπου που επιτρέπουν στους ανιχνευτές να ταξινομούν νέες εικόνες με βάση το πόσο μοιάζει ένα πρόσωπο να έχει μια κανονική συναισθηματική έκφραση. Το Py-Feat παρέχει τέσσερις ανιχνευτές συναισθημάτων ικανούς να ανιχνεύουν τις επτά βασικές κατηγορίες συναισθημάτων: θυμό, αηδία, φόβο, ευτυχία, λύπη, έκπληξη και ουδέτερο. Οι ανιχνευτές είναι: το Residual Masking Network (ResMaskNet), το FerNet και δύο στατιστικά μοντέλα εκμάθησης, τα Random Forest και Linear SVM.

(ε) Για την εκτίμηση της πόζας του προσώπου/κεφαλιού, το Py-Feat παρέχει τρεις μεθόδους οι οποίες είναι img2pose, img2pose-c (constrained) και Perspective-n-Point.

Fex Module

Το Py-Feat περιλαμβάνει επίσης μια κλάση δεδομένων Fex που περιλαμβάνει μεθόδους προεπεξεργασίας, ανάλυσης και οπτικοποίησης δεδομένων έκφρασης προσώπου. Η κλάση Fex είναι μια επέκταση του ευρέως χρησιμοποιούμενου Pandas DataFrame.



Εικόνα 18: F1-scores για την πρόβλεψη συναισθημάτων σε νέα δεδομένα [104]

Για κάθε ανιχνευτή υπάρχουν κάποια προεπιλεγμένα μοντέλα αφού αξιολογήθηκαν ως τα καλύτερα τη στιγμή της δημοσίευσης. Τα default μοντέλα είναι:

- RetinaFace - μοντέλο ανίχνευσης προσώπου
- MobileFaceNet - μοντέλο ανίχνευσης ορόσημων
- SVM - μοντέλο ανίχνευσης μονάδας δράσης
- Residual Masking Network - μοντέλο ανίχνευσης συναισθημάτων
- Img2Pose – μοντέλο εκτίμησης πόζας

```
from feat import Detector

detector = Detector(
    face_model="retinaface",
    landmark_model="mobilefacenet",
    au_model='svm',
    emotion_model="resmasknet",
    facepose_model="img2pose",
)

detector
```

Εικόνα 19: Τα προεπιλεγμένα μοντέλα του Detector module του Py-Feat. [105]

Για παράδειγμα, αν στον παραπάνω detector εφαρμόσουμε τη συνάρτηση detect_image() που λαμβάνει ως είσοδο το path στο οποίο είναι αποθηκευμένη η εικόνα που θέλουμε να αναλύσουμε, τότε μπορούμε να δούμε διάφορα αποτελέσματα όπως facebox (τις συντεταγμένες του bounding box του προσώπου), AUs (τα Action Units που ενεργοποιήθηκαν – binary τιμές), emotions (το ποσοστό στο οποίο ανιχνεύθηκε καθένα από τα βασικά συναισθήματα) και facepose (την πόζα του προσώπου).

```
single_face_prediction = detector.detect_image(single_face_img_path)

# Show results
single_face_prediction
```

single_face_prediction.facebox					
	FaceRectX	FaceRectY	FaceRectWidth	FaceRectHeight	FaceScore
0	196.976837	140.997742	173.810486	257.639343	0.999681

```
single_face_prediction.ous
```

	AU01	AU02	AU04	AU05	AU06	AU07	AU09	AU10	AU11	AU12	AU14	AU15
0	0	0	0	0	0	1	0	1	1	1	1	0

```
single_face_prediction.emotions
```

	anger	disgust	fear	happiness	sadness	surprise	neutral
0	0.000369	0.000026	0.000485	0.986996	0.000046	0.01201	0.000068

```
single_face_prediction.facepose # (in degrees)
```

	Pitch	Roll	Yaw
0	0.832747	-3.809027	6.605721

Εικόνα 20: Αποτελέσματα της `detector.detect_image()` του `Py-Feat`. [105]

Για την εγκατάσταση του `Py-Feat` εκτελούμε την εντολή: `pip install py-feat`.

3.3 Spotify API

Το Spotify API [106] είναι ένα RESTful API. Αποτελεί μια διεπαφή που μπορούν να χρησιμοποιήσουν τα προγράμματα/εφαρμογές για την ανάκτηση και τη διαχείριση δεδομένων του Spotify μέσω του διαδικτύου. Το API είναι σχεδιασμένο για πρόσβαση μέσω προγραμματισμού και επιστρέφει δεδομένα σε μορφή JSON (JavaScript Object Notation).

Παρέχει ένα σύνολο από endpoints (περίπου 80 εν έτη 2022) μέσω των οποίων οι προγραμματιστές μπορούν να ανακτήσουν δεδομένα καταλόγου Spotify σχετικά με καλλιτέχνες, κομμάτια, λίστες αναπαραγωγής, επεισόδια και εκπομπές podcasts. Περιλαμβάνει επίσης το OAuth (ένα ανοιχτό πρότυπο για εξουσιοδότηση) για την ανάκτηση προσωπικών δεδομένων σχετικά με έναν χρήστη του Spotify και τη διαχείριση αυτών των δεδομένων (πχ. προσθήκη μουσικού κομματιού σε playlist που έχει φτιάξει ο χρήστης). [107]

```
{
  "danceability": 0.199,
  "energy": 0.156,
  "key": 7,
  "loudness": -16.916,
  "mode": 1,
  "speechiness": 0.0388,
  "acousticness": 0.813,
  "instrumentalness": 0.862,
  "liveness": 0.348,
  "valence": 0.18,
  "tempo": 137.234,
  "type": "audio_features",
  "id": "02P60AdunANMp5sHoZVps4",
  "uri": "spotify:track:02P60AdunANMp5sHoZVps4",
  "track_href":
    "https://api.spotify.com/v1/tracks/02P60AdunANMp5sHoZVps4",
  "analysis_url": "https://api.spotify.com/v1/audio-
    analysis/02P60AdunANMp5sHoZVps4",
  "duration_ms": 58920,
  "time_signature": 3
}
```

Εικόνα 21: Audio features για ένα Spotify track. [108]

Για την έρευνά μας, τα endpoints από τα οποία έχουμε πάρει πληροφορίες είναι τα εξής:

- Get Track's Audio Features - που επιστρέφει τα ακουστικά χαρακτηριστικά ενός κομματιού με βάση το Spotify ID του (track id)

Η απάντηση του Spotify φαίνεται στην Εικ. 21, όπου εμφανίζονται τα ακουστικά χαρακτηριστικά που είναι διαθέσιμα για κάθε μουσικό κομμάτι μέσω του Spotify API. Αυτά είναι πιο αναλυτικά τα ακόλουθα:

- *danceability* - πόσο κατάλληλο είναι ένα κομμάτι για χορό βάσει ενός συνδυασμού μουσικών στοιχείων όπως tempo, rhythm stability, beat strength, και overall regularity
- *energy* - ένα μέτρο της έντασης και της δραστηριότητας (συνήθως τα ενεργητικά κομμάτια είναι γρήγορα, δυνατά και θορυβώδη)
- *key* - το κλειδί στο οποίο βρίσκεται το κομμάτι σύμφωνα με το standard Pitch Class notation (-1 έως 11),
- *loudness* - η συνολική ένταση ενός κομματιού σε ντεσιμπέλ (συνήθως -60 έως 0 dB)
- *mode* – ο τρόπος (major 1 ή minor 0) ενός κομματιού,
- *speechiness* - παρουσία προφορικών λέξεων σε ένα κομμάτι
- *acousticness* - εάν το κομμάτι είναι ακουστικό
- *instrumentalness* - εάν ένα κομμάτι δεν περιέχει φωνητικά
- *liveness* - η παρουσία ενός κοινού στην ηχογράφηση

- *valence* - η μουσική θετικότητα που μεταδίδεται από ένα κομμάτι
- *tempo* - συνολικός εκτιμώμενος ρυθμός ενός κομματιού σε κτύπους ανά λεπτό (BPM)
- *time signature* - εκτιμώμενη συνολική υπογραφή χρόνου (μέτρο) ενός κομματιού

Για τα χαρακτηριστικά που δεν αναγράφεται εύρος τιμών, οι τιμές είναι από 0 έως 1.

- **Get Recommendations** - που επιστρέφει μια λίστα από προτεινόμενα μουσικά κομμάτια και παίρνει ως είσοδο ένα συνδυασμό από genres, artists ή tracks. Προαιρετικά μπορούν να δοθούν και μέγιστες/ελάχιστες τιμές και τιμές-στόχοι των 12 ακουστικών χαρακτηριστικών. Επιπλέον, μπορεί να καθοριστεί μέσω του limit, ο αριθμός των κομματιών που θα επιστρέφονται (από 1 έως 100).

Το Spotify API σε αυτό το ερώτημα επιστρέφει, μια λίστα τραγουδιών για καθένα από τα οποία αναφέρονται βασικές πληροφορίες όπως όνομα, διάρκεια, δημοτικότητα κ.α. καθώς επίσης και σε ποιο άλμπουμ ανήκει και τη λίστα των καλλιτεχνών που το ερμηνεύουν. Για το άλμπουμ δίνονται πληροφορίες όπως όνομα, ημερομηνία έκδοσης, το σύνολο των κομματιών που περιέχει, η εικόνα του άλμπουμ κ.α.

Έτσι, για τις ανάγκες της εφαρμογής χρησιμοποιούμε το endpoint Get Recommendations δίνοντας στην είσοδο έως 5 μουσικά είδη που επιλέγει ο χρήστης και τιμές για τα valence και energy που αντιστοιχίζονται στο συναίσθημα του χρήστη με τρόπο που θα παρουσιαστεί στο πειραματικό μέρος της παρουσίασης της εφαρμογής.

3.4 Παρουσίαση Εφαρμογής

Η εφαρμογή που αναπτύχθηκε κατά τη διάρκεια αυτής της εργασίας, αφορά την αναγνώριση συναισθήματος από εικόνες προσώπου και την πρόταση μουσικών κομματιών με βάση τα συναισθήματα του χρήστη. Στο μέρος αυτό θα αναλύσουμε βήμα-βήμα τη διαδικασία που ακολουθήθηκε για την ανάπτυξη αυτής της εφαρμογής. Από το στήσιμο και το πως δημιουργήθηκαν όλα τα επιμέρους κομμάτια, μέχρι και το τελικό αποτέλεσμα.

Για τη δημιουργία της εφαρμογής, αρχικά έγινε εγκατάσταση του Flutter, ένα εργαλείο ανάπτυξης λογισμικού ανοιχτού κώδικα που βοηθά στη δημιουργία εφαρμογών πολλαπλών πλατφορμών μέσα από μια ενιαία βάση κώδικα. Αρχικά, φτιάχτηκαν δυο-τρία project ώστε να γίνει η πρώτη επαφή με τη Dart (γλώσσα προγραμματισμού του Flutter) και να κατανοήσουμε πως λειτουργεί και στη συνέχεια μελετήθηκαν τα Flutter packages ώστε να μπορέσει να γίνει χρήση της κάμερας, χρήση http κλπ. Ταυτόχρονα, δημιουργήθηκε ένας οδηγός εγκατάστασης του Flutter (όπως φαίνεται στο Παρ.2). Καθώς δεν υπήρχε διαθέσιμη συσκευή iOS κατά την ανάπτυξη της εφαρμογής, αυτή υλοποιήθηκε αποκλειστικά για android συσκευές.

Η βασική ιδέα της εφαρμογής ήταν η Αναγνώριση Συναισθημάτων, για το σκοπό αυτό μελετήθηκε και χρησιμοποιήθηκε το Py-Feat, μια εργαλειοθήκη Python ανοιχτού κώδικα που περιέχει εργαλεία για την αναγνώριση εκφράσεων προσώπου από εικόνα και βίντεο. Για την ενσωμάτωση της Python στο Flutter, αποφασίστηκε να δημιουργηθεί ένα Python Flask API με ένα endpoint. Μέσα από το endpoint (POST request) γίνεται χρήση του Detector του Py-Feat που παίρνει ως είσοδο το path μιας εικόνας και μετά από επεξεργασία, επιστρέφει ως αποτέλεσμα ένα σύνολο από AUs, ποσοστά συναισθημάτων και άλλα στοιχεία που περιγράφουν την εικόνα.

Η εφαρμογή χρησιμοποιεί την κάμερα για να συλλέξει την εικόνα του προσώπου του χρήστη. Στη συνέχεια, η εικόνα αυτή στέλνεται στο Python API που δημιουργήθηκε και ο Detector του Py-Feat επιστρέφει τα αποτελέσματα από τα οποία κρατάμε μόνο τα συναισθήματα που ανιχνεύθηκαν στην εικόνα. Αυτά περιλαμβάνουν: χαρά, λύπη, θυμό, έκπληξη, αηδία, φόβο και ουδέτερη.

Στο δεύτερο μέρος της εφαρμογής, αποφασίσαμε να εξελίξουμε το ER σύστημα που φτιάξαμε, σε ένα Emotion Based Music Recommendation σύστημα. Για το σκοπό αυτό, αξιοποιήθηκε το Spotify API και το endpoint Get Recommendations που επιστρέφει μια λίστα από τραγούδια ανάλογα με κάποια δεδομένα εισόδου. Για να μπορέσει να λειτουργήσει το endpoint είναι απαραίτητο να δώσουμε ως είσοδο κάποια στοιχεία για αγαπημένα είδη μουσικής, για καλλιτέχνες ή αγαπημένα μουσικά κομμάτια. Καθώς θέλαμε να κρατήσουμε το σύστημα συστάσεων πιο γενικό αποφασίσαμε να διαλέξουμε τα είδη (genres). Έτσι, για να ξεκινήσει η διαδικασία των συστάσεων, η εφαρμογή ζητάει από το χρήστη να επιλέξει κάποια (από 1 έως 5) είδη μουσικής που προτιμάει. Κάθε φορά που ξεκινάει πάλι το σύστημα συστάσεων, ο χρήστης μπορεί να αλλάξει τις προτιμήσεις του.

Στοιχεία που μπορούμε να δώσουμε ως είσοδο στο Get Recommendations είναι και οι min, max, target τιμές όλων των ακουστικών χαρακτηριστικών που περιγράφουν τα μουσικά κομμάτια (όπως φαίνονται μέσα από το endpoint Get Track's Audio Features). Η λογική που ακολουθήσαμε για να δώσουμε τιμές σε αυτά τα χαρακτηριστικά είναι η ακόλουθη:

- Γνωρίζουμε ότι υπάρχει ένα δισδιάστατο μοντέλο συναισθημάτων, το μοντέλο του Russell, που έχει ως άξονες τα valence και arousal. Ανάλογα με τις τιμές αυτών κατατάσσονται τα συναισθήματα πάνω στο επίπεδο.
- Το μοντέλο του Russell είναι το πιο ευρέως αποδεκτό μοντέλο που δείχνει τα valence και arousal και πώς σχετίζονται με τη συναισθηματική κατάσταση. [109]
- Τα valence και energy είναι ακουστικά χαρακτηριστικά που περιλαμβάνονται μέσα στο Spotify API για το κάθε μουσικό κομμάτι.

Θεωρούμε πως τα arousal και energy ταυτίζονται. Στην πραγματικότητα, η ενέργεια (energy) δεν είναι ακριβώς διέγερση (arousal), αλλά χρησιμεύει ως υποκατάστατό της. Ως αποτέλεσμα, μπορούμε να συνδέσουμε τη συναισθηματική κατάσταση του χρήστη με τα μουσικά κομμάτια του Spotify και να κάνουμε τις κατάλληλες προτάσεις.

Εφόσον το Spotify API μας δίνει τη δυνατότητα να ρυθμίσουμε και άλλα ακουστικά χαρακτηριστικά, αποφασίζουμε να προσθέσουμε κάποια ακόμα για να δούμε αν βελτιώνονται οι συστάσεις. Στην έρευνα των Panda et al. [110] τρία από τα 12 χαρακτηριστικά του Spotify API αναγνωρίστηκαν ως εξαιρετικά σχετικά με την ταξινόμηση συναισθημάτων: τα energy, valence και acousticness. Παρατηρήθηκε πως το χαμηλό acousticness, που υποδηλώνει κυρίως ηλεκτρικούς ήχους, είναι πιο διαδεδομένο στα χαρούμενα τραγούδια και ειδικά σε τραγούδια με ένταση/θυμό. Επίσης, ως αποτέλεσμα του πρώτου, βρέθηκε υψηλή (αρνητική) συσχέτιση μεταξύ energy και acousticness. Διαισθητικά, θεωρούμε ότι και το danceability έχει άμεση σχέση με τα συναισθήματα που εκφράζει το τραγούδι. Αν και αυτό έρχεται σε αντίθεση με την παραπάνω έρευνα που υποστηρίζει ότι δεν υπάρχει κάποια αξιοσημείωτη σχέση μεταξύ danceability και MER. Όλα τα παραπάνω θα εξεταστούν και στο πειραματικό μέρος της παρουσιάσής μας.

Για την αξιοποίηση του Spotify API αρχικά έπρεπε να γίνει εγγραφή της εφαρμογής μας στο Spotify for Developers. Οπότε, κάναμε login στη σελίδα αυτή με τα στοιχεία του λογαριασμού μας στο Spotify, δημιουργήσαμε την εφαρμογή εισάγοντας ένα όνομα και πήραμε τα απαραίτητα client credentials (client ID + secret) για να μπορέσουμε να επικοινωνήσουμε με το Spotify API.

Επιστρέφοντας στη λειτουργία της εφαρμογής, η διαδικασία των μουσικών συστάσεων είναι η ακόλουθη. Αρχικά, όπως αναφέρθηκε, η εφαρμογή ζητάει από το χρήστη να επιλέξει μουσικά είδη. Στη συνέχεια, ξεκινάει η διαδικασία της αναγνώρισης συναισθημάτων. Ανοίγει η κάμερα, τραβάμε μια φωτογραφία του προσώπου του χρήστη, αναγνωρίζεται το κυρίαρχο συναίσθημά του και στην επόμενη οθόνη της εφαρμογής εμφανίζεται η λίστα με τα μουσικά κομμάτια που προτείνουμε στο χρήστη ανάλογα με το συναίσθημα της εικόνας. Ο χρήστης μπορεί να δει όλα τα μουσικά κομμάτια και να ακούσει ένα preview (δυστυχώς το Spotify API δεν δίνει πρόσβαση στο πλήρες κομμάτι) ή να επιλέξει να ακούσει ολόκληρο το κομμάτι μέσα από την εφαρμογή του Spotify.

Τέλος, μετά την ολοκλήρωση της εφαρμογής διερευνήθηκαν οι τρόποι με τους οποίους μπορεί να γίνει η αξιολόγηση του συστήματος. Οι δύο επιλογές είναι: offline αξιολόγηση (με χρήση μετρικών) και online αξιολόγηση (μέσω της ψήφου των χρηστών). Δυστυχώς, δεν μπορεί εύκολα να γίνει χρήση μετρικών διότι το Spotify API δεν παρέχει όλη την πληροφορία που χρειαζόμαστε μέσα σε ένα endpoint. Θα έπρεπε, λοιπόν, να δημιουργηθεί ένα dataset που να συνδυάζει τα json αποτελέσματα των endpoints του Spotify, ώστε να γίνει ακολούθως χρήση μετρικών. Στα πλαίσια της εργασίας μας, επιλέξαμε την online αξιολόγηση του συστήματος, ώστε να δούμε κατά πόσο η playlist που παράγεται ανταποκρίνεται στο συναίσθημα που ανιχνεύεται από την εικόνα. Έτσι, ο κάθε χρήστης μπορεί να αξιολογήσει την λίστα αναπαραγωγής σε μια κλίμακα από 1 έως 5. Για το σκοπό αυτό, προσθέσαμε ένα endpoint στο Python API, που λαμβάνει την αξιολόγηση και το αντίστοιχο συναίσθημα του χρήστη και τα αποθηκεύει σε ένα φάκελο (τοπικά), που μπορούμε στη συνέχεια να αξιοποιήσουμε για την εξαγωγή συμπερασμάτων.

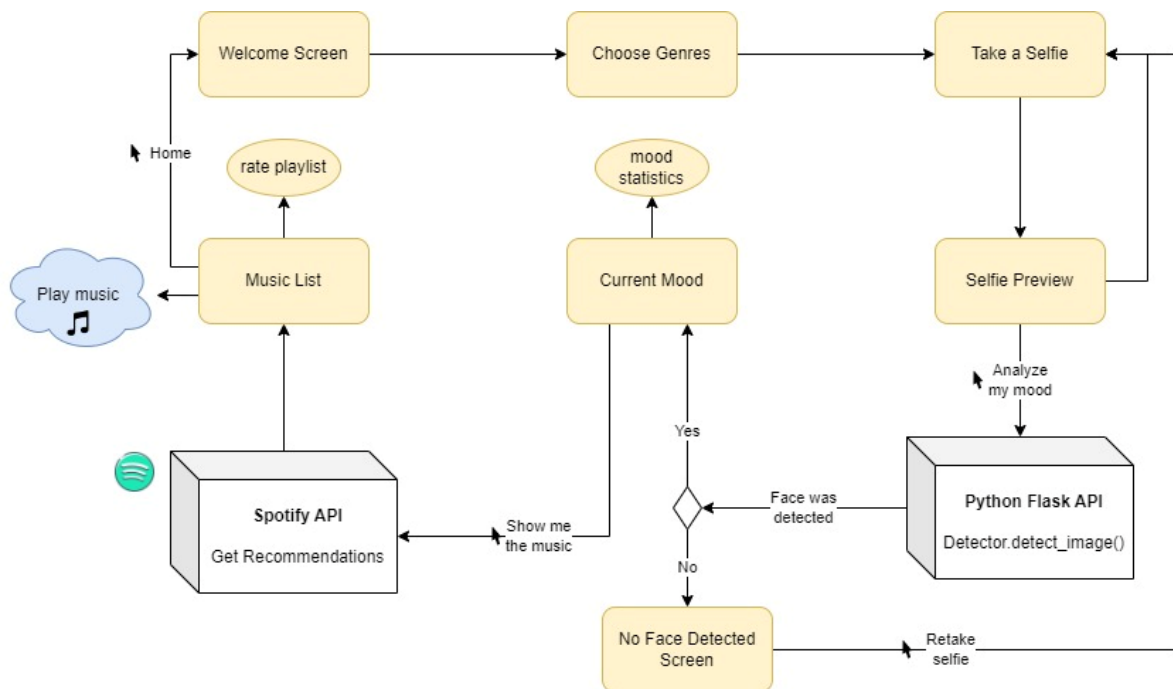
Στον πίνακα που ακολουθεί σημειώσαμε το μέσο όρο των τιμών των αξιολογήσεων που συλλέχθηκαν για το κάθε συναίσθημα.

	Χαρά	Λύπη	Έκπληξη	Θυμός	Φόβος	Αηδία	Ουδέτερο
Αριθμός αξιολογήσεων	9	3	4	2	2	2	11
Μέσος όρος αξιολογήσεων	4.78	4.34	4.25	4.0	2.0	3.0	4.82

Πίνακας 1: Αξιολόγηση της προτεινόμενης λίστας αναπαραγωγής από τους χρήστες

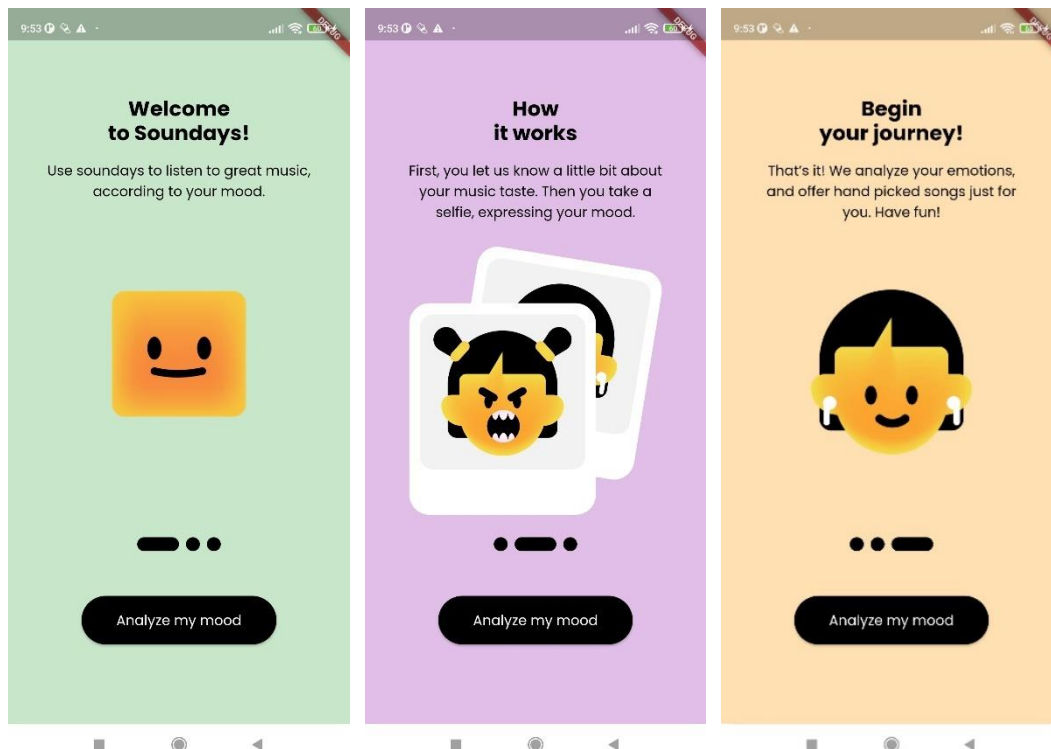
Οι χρήστες αξιολόγησαν τις playlist για ουδέτερο συναίσθημα με την πιο υψηλή αξιολόγηση (4.82 στα 5), ενώ φαίνεται να ακολουθεί το συναίσθημα της χαράς (με 4.78). Από την άλλη, το συναίσθημα που έλαβε την μικρότερη τιμή είναι αυτό του φόβου (2.0). Μέσα από τον αριθμό των αξιολογήσεων για το κάθε συναίσθημα διαπιστώνουμε ότι ελάχιστοι χρήστες απέδωσαν συναισθήματα όπως ο θυμός, ο φόβος και η αηδία.

Η συνολική αρχιτεκτονική της εφαρμογής περιγράφεται και διαγραμματικά μέσα από την Εικ. 22, όπου τα ορθογώνια αντιπροσωπεύουν τις οθόνες της εφαρμογής και τα κουτιά αντιπροσωπεύουν τα εξωτερικά συστήματα που χρησιμοποιήθηκαν.

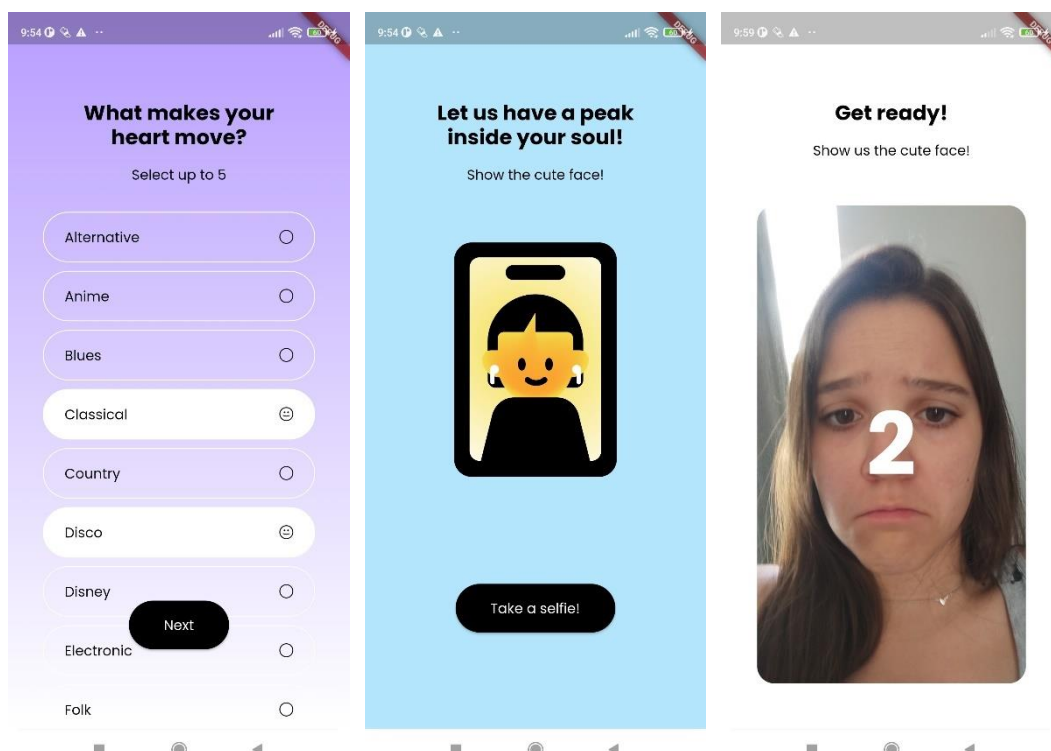


Εικόνα 22: Αρχιτεκτονική της εφαρμογής μας

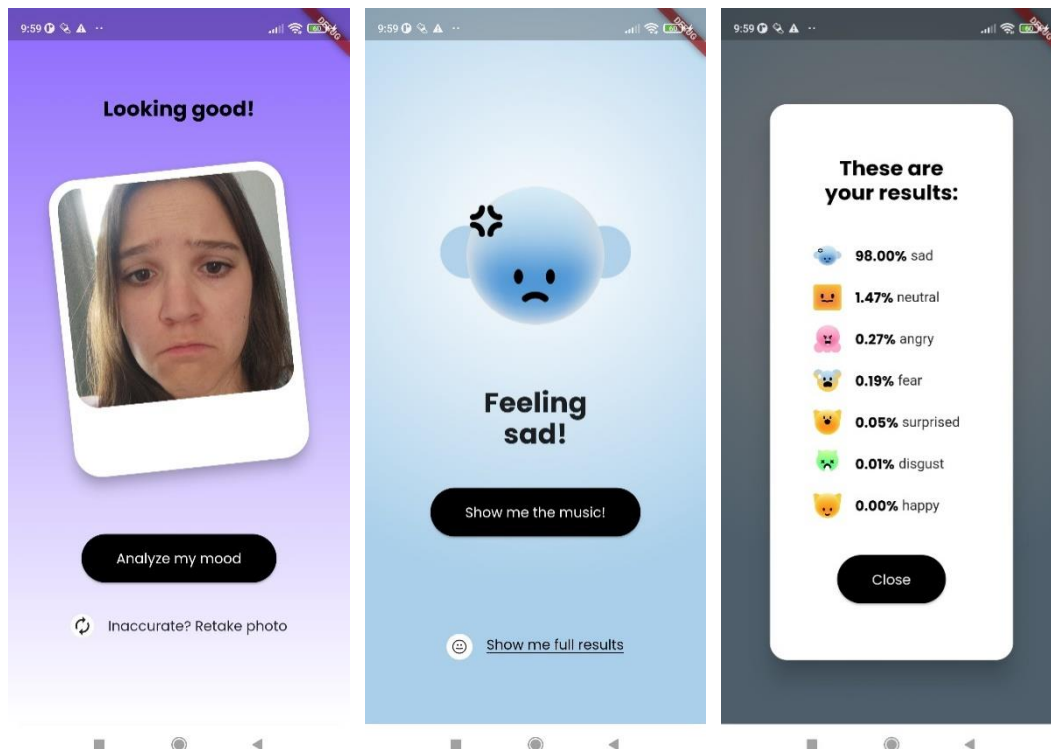
Για να βρεθεί το όνομα της εφαρμογής χρησιμοποιήθηκε ένας online AI generator και δίνοντας κάποιες λέξεις κλειδιά (όπως music, AI κ.α.), επιλέχθηκε η πρόταση που μας άρεσε περισσότερο: “Soundays”. Έτσι, ξεκίνησε και η διαδικασία βελτίωσης του UI ώστε να γίνει πιο όμορφο και φιλικό προς το χρήστη. Η τελική μορφή της εφαρμογής, φαίνεται μέσα από τα screenshots στις εικόνες των επόμενων σελίδων.



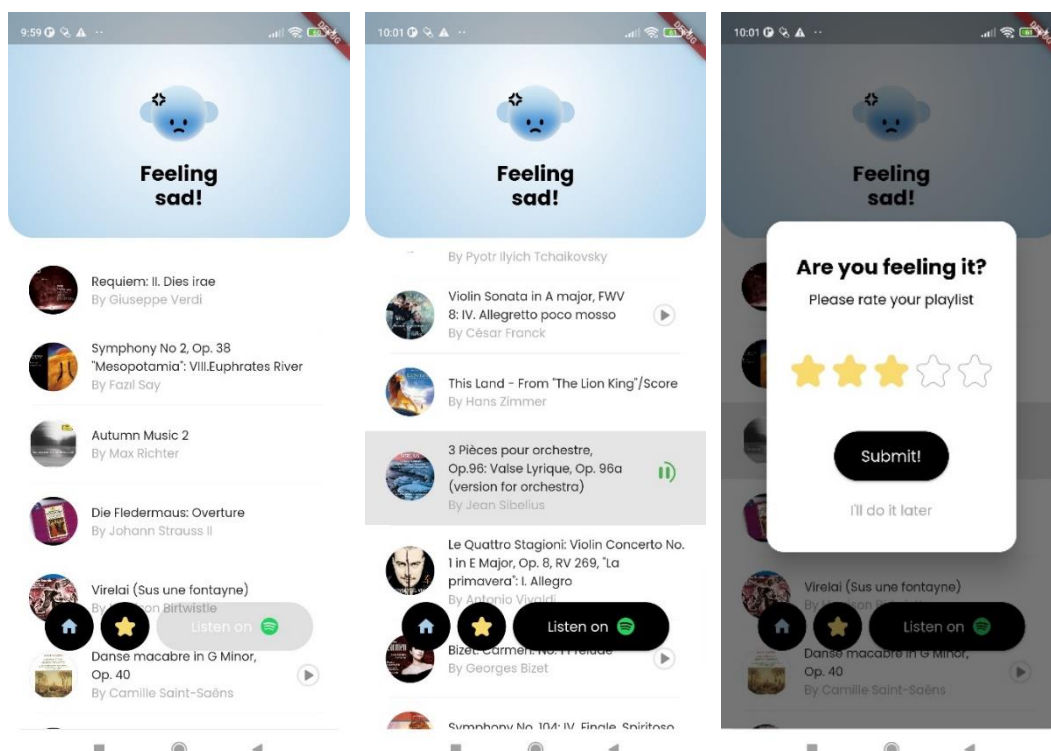
Εικόνες 23-25: Welcome screen με carousel για την επεξήγηση της δημιουργίας της εφαρμογής.



Εικόνα 26-28: Επιλογή μουσικών ειδών. Προετοιμασία για την λήψη φωτογραφίας. Λήψη φωτογραφίας με αντίστροφη μέτρηση.



Εικόνα 29-31: Προεπισκόπηση φωτογραφίας. Κυρίαρχο συναίσθημα που ανιχνεύθηκε. Πληροφορίες για τα ποσοστά των συναισθημάτων.



Εικόνα 32-34: Προτεινόμενη λίστα αναπαραγωγής με audio player. Επιλογή μουσικού κομματιού – κατάσταση play. Αξιολόγηση λίστας με 5-star rating.

Πειραματικό μέρος

Στο πειραματικό κομμάτι της εφαρμογής «παίξαμε» με τα ακουστικά χαρακτηριστικά των μουσικών κομματιών που μπορούν να χρησιμοποιηθούν για την κατηγοριοποίησή τους με βάση τα συναισθήματα. Τα πειράματα έγιναν επιλέγοντας κάθε φορά ως μοναδικό είδος μουσικής το “Pop” καθώς αυτό εκφράζει τη μεγαλύτερη κλίμακα συναισθημάτων με αρκετά εμφανή τρόπο σε ένα χρήστη που δεν είναι σχετικός με τη μουσική.

Αρχικά εξετάζουμε μόνο τα valence και energy. Αν παρατηρήσουμε το δισδιάστατο χώρο συναισθημάτων του Russell μπορούμε να δούμε ότι τα συναισθήματα που μας αφορούν (αυτά που ανιχνεύονται από το σύστημά μας) βρίσκονται στα πρώτα τρία τεταρτημόρια. Η χαρά και η έκπληξη βρίσκονται στο Q1, ο θυμός, ο φόβος και η αηδία βρίσκονται στο Q2 και η λύπη βρίσκεται στο Q3. Καθώς πολλά συναισθήματα εντάσσονται στο ίδιο τεταρτημόριο είναι δύσκολο να διακρίνουμε ξεχωριστές τιμές για το καθένα οπότε είναι πολύ πιθανό να συμπίπτουν. Οι τιμές για τα valence και energy κυμαίνονται από 0 έως 1 στο Spotify API. Έτσι, κάνουμε την ακόλουθη αντιστοίχιση:

	Energy		Valence	
Συναίσθημα	(min, max)	target	(min, max)	target
Χαρά	0.7 - 1.0	0.8	0.7 - 1.0	0.9
Λύπη	0.0 - 0.4	0.2	0.0 - 0.4	0.2
Έκπληξη	0.7 - 1.0	0.8	0.7 - 1.0	0.8
Αηδία	0.6 - 0.8	0.7	0.0 - 0.2	0.1
Θυμός	0.6 - 1.0	0.9	0.0 - 0.4	0.2
Φόβος	0.6 - 1.0	0.9	0.0 - 0.4	0.3
Ουδέτερη	0.4 - 0.6	0.5	0.4 - 0.6	0.5

Πίνακας 2: Μια πρώτη αντιστοίχιση των energy και valence με τα 7 βασικά συναισθήματα, μέσα από απλή παρατήρηση του μοντέλου του Russell

Για την κατάταξη του ουδέτερου συναισθήματος θεωρούμε ότι βρίσκεται στο κέντρο του επιπέδου, δηλ. στο (0, 0) καθώς όπως σημειώνεται και στην έρευνα των Sharma et al. [109] το ουδέτερο (βασικό) συναίσθημα των περισσότερων ανθρώπων βρίσκεται κοντά στο σημείο (0, 0), την αρχή των αξόνων, στο μοντέλο του Russell.


Μέσα από τις δοκιμές που έγιναν καταλήξαμε στις εξής τιμές:

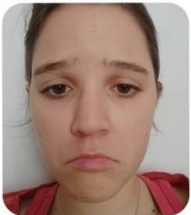
	Energy		Valence	
Συναίσθημα	(min, max)	target	(min, max)	target
Χαρά	0.7 - 0.9	0.8	0.7 - 1.0	0.8
Λύπη	0.0 - 0.4	0.2	0.0 - 0.3	0.2
Έκπληξη	0.9 - 1.0	0.95	0.6 - 0.8	0.7
Αηδία	0.6 - 0.8	0.7	0.0 - 0.2	0.1
Θυμός	0.7 - 1.0	0.9	0.0 - 0.5	0.3
Φόβος	0.8 - 1.0	0.95	0.0 - 0.4	0.3
Ουδέτερη	0.4 - 0.6	0.5	0.4 - 0.6	0.5

Πίνακας 3: Τελικές τιμές των *energy* και *valence* για κάθε συναίσθημα.

Εδώ θα πρέπει να αναφέρουμε ότι τα συναισθήματα έκπληξης, αηδίας και φόβου είναι λίγο πιο δύσκολο να αξιολογηθούν σε σχέση με τα υπόλοιπα οπότε δεν περιμένουμε οι τιμές τους να είναι τόσο στοχευμένες. Ειδικότερα για την αηδία, όπως αναφέρθηκε και στο θεωρητικό μέρος, φαίνεται ότι σπάνια γίνεται αντιληπτή μέσα από τα μουσικά κομμάτια.

Στη συνέχεια παρουσιάζεται ένας συγκεντρωτικός πίνακας που περιλαμβάνει: την εικόνα που δίνεται ως είσοδος στο σύστημα, τα ποσοστά που υπολογίζονται για το κάθε συναίσθημα όπως εντοπίζονται στην εικόνα, η λίστα των τραγουδιών που επιστρέφονται με βάση την αναζήτηση στο Spotify (που περιλαμβάνει τα επιλεγμένα genres και τις τιμές των valence και energy όπως αναγράφονται στον Πίν. 3) και τέλος τα energy και valence για το κάθε τραγούδι.

		Τιμές τραγουδιού		Εξαγόμενα ποσοστά συναισθημάτων (%)						
Συναίσθημα		energy	valence	χαρά	λύπη	έκπληξη	αηδία	θυμός	φόβος	ουδέτερη
<div>Χαρά</div> 	S1	0.795	0.788	87.92	0.04	11.69	0.002	0.008	0.08	0.26
	S2	0.806	0.787							
	S3	0.816	0.816							
	S4	0.817	0.782							
	S5	0.785	0.825							

	S6	0.791	0.756							
	S7	0.791	0.755							
	S8	0.819	0.766							
	S9	0.774	0.763							
	S10	0.843	0.778							
	S11	0.819	0.747							
Λύπη 	S1	0.174	0.231	0.0017	96.01	0.31	0.0024	0.07	0.26	3.35
	S2	0.265	0.204							
	S3	0.262	0.228							
	S4	0.275	0.18							
	S5	0.242	0.257							
	S6	0.289	0.213							
	S7	0.111	0.219							
	S8	0.303	0.17							
	S9	0.303	0.15							
	S10	0.356	0.189							
	S11	0.296	0.12							

Πίνακας 4: Συγκεντρωτικός πίνακας που περιλαμβάνει: το συναίσθημα και την εικόνα εισόδου, τις τιμές των valence και energy για τα μουσικά κομμάτια που επιστράφηκαν, τα ποσοστά συναισθημάτων όπως ανιχνεύθηκαν από το σύστημα.

Για τη χαρά, τα μουσικά κομμάτια του Πίν. 4 είναι:

- S1: No Problem (by Chance the Rapper)
- S2: Poker Face (by Lady Gaga)
- S3: Despacito (by Luis Fonsi)
- S4: Swalla (by Jason Derulo)

- S5: Toothbrush (by DNCE)
- S6: Classic (by MKTO)
- S7: All In My Head (by Fifth Harmony)
- S8: Safe and Sound (by Capital Cities)
- S9: Gone (by Afrojack)
- S10: If It Ain't Love (by Jason Derulo)
- S11: Treat You Better (by Shawn Mendes)

Για τη λύπη, τα μουσικά κομμάτια του Πίν. 4 είναι:

- S1: Dancing On My Own (by Calum Scott)
- S2: Angels (by Khalid)
- S3: Sweet Creature (by Harry Styles)
- S4: I hate u, I love u (by gnash)
- S5: Supermarket Flowers (by Ed Sheeran)
- S6: Consequences (by Camila Cabello)
- S7: Not About Angels (by Birdy)
- S8: Sober (by Demi Lovato)
- S9: The One (by The Chainsmokers)
- S10: Don't (by Bryson Tiller)
- S11: lovely (by Billie Eilish)

Μέσα από τις λίστες τραγουδιών που μας επιστράφηκαν από τις δοκιμές για τον Πίν. 3, κάναμε κάποιες παρατηρήσεις όπως φαίνονται στον πίνακα που ακολουθεί:

Συναίσθημα	Αποτελέσματα	Energy	Valence	Παρατηρήσεις
Χαρά	No Problem (by Chance the Rapper)	0.795	0.788	Τα αποτελέσματα φαίνεται να είναι αρκετά ικανοποιητικά.
	Poker Face (by Lady Gaga)	0.806	0.787	
	Despacito (by Luis Fonsi)	0.816	0.816	
	Swalla (by Jason Derulo)	0.817	0.782	
	Toothbrush (by DNCE)	0.785	0.825	

	Classic (by MKTO)	0.791	0.756	
	All In My Head (by Fifth Harmony)	0.791	0.755	
	Safe and Sound (by Capital Cities)	0.819	0.766	
	Gone (by Afrojack)	0.774	0.763	
	If It Ain't Love (by Jason Derulo)	0.843	0.778	
	Treat You Better (by Shawn Mendes)	0.819	0.747	
Έκπληξη	Whistle (by Flo Rida)	0.937	0.739	<p>Παρατηρούμε ότι υπάρχουν ορισμένες αστοχίες, (πχ. Still into You (by Paramore)) όπου ενώ λόγω των μουσικών οργάνων το τραγούδι φαίνεται να έχει ένταση που θα ήταν απαραίτητη για το συναίσθημα που εξετάζουμε, η φωνή και η μελωδία των στίχων είναι πιο απαλή οπότε δεν προκύπτει το επιθυμητό συναίσθημα.</p> <p>Επιπλέον, πολύ καλή επιλογή φαίνεται να είναι το ιδιαίτερα ανεβασμένο energy.</p>
	High Hopes (by Panic! At The Disco)	0.904	0.681	
	Just Got Paid (by Sigala)	0.921	0.65	
	Everybody (by Logic)	0.94	0.773	
	Still into You (by Paramore)	0.923	0.765	
	Ain't Giving Up (by Craig David)	0.909	0.643	
	Love on Me (by Galantis)	0.91	0.649	
	Timber (by Pitbull)	0.963	0.788	
	No Money (by Galantis)	0.915	0.78	
Λύπη	Dancing On My Own (by Calum Scott)	0.174	0.231	<p>Ανάμεσα στο τραγούδια εμφανίζονται και κάποια πιο απαλά/ήρεμα κομμάτια, όπως Supermarket Flowers (by Ed Sheeran), Sweet Creature (by Harry Styles) και The One (Chainsmokers) που δεν είναι απαραίτητα λυπητερά και τραγούδια όπως το Don't</p>
	Angels (by Khalid)	0.265	0.204	
	Sweet Creature (by Harry Styles)	0.262	0.228	
	I hate u, I love u (by gnash)	0.275	0.18	

	Supermarket Flowers (by Ed Sheeran)	0.242	0.257	(Bryson Tiller) που έχουν πιο «μπιταρισμένο» ήχο. Αλλά γενικά τα αποτελέσματα ανταποκρίνονται στη “λύπη”.
	Consequences (by Camila Cabello)	0.289	0.213	
	Not About Angels (by Birdy)	0.111	0.219	
	Sober (by Demi Lovato)	0.303	0.17	
	The One (by The Chainsmokers)	0.303	0.15	
	Don’t (by Bryson Tiller)	0.356	0.189	
	lovely (by Billie Eilish)	0.296	0.12	
Αηδία	Darkside (by Alan Walker)	0.642	0.112	Δεν μπορεί να γίνει κάποια παρατήρηση – δεν μπορούμε να διακρίνουμε το συναίσθημα της αηδίας μέσα σε κάποιο μουσικό κομμάτι.
	Girl on Fire (by Alicia Keys)	0.706	0.144	
	A Sky full of Stars (Coldplay)	0.675	0.162	
	Sex (Cheat Codes)	0.691	0.186	
	Let’s Get Lost (G-Easy)	0.672	0.186	
	Faded (by Alan Walker)	0.651	0.166	
	Turn Down for What (DJ Snake)	0.799	0.0815	
	Pray For Me (by The Weekend)	0.677	0.188	
	Unsteady (by X Ambassadors)	0.665	0.199	
Θυμός	Purple Lamborghini (by Skrillex)	0.899	0.298	Πολλά από τα αποτελέσματα δεν ταιριάζουν στο θυμό πχ. All My Love, Lost In Japan, I Got You. Βέβαια τα remix κομμάτια (σε σχέση με τα original) και όσα περιέχουν
	All My Love (by cash cash)	0.871	0.304	
	Cruel (By Snakehips)	0.879	0.318	

	Lost In Japan – remix (by Shawn Mendes),	0.855	0.293	ηλεκτρονικούς ήχους, εκφράζουν πολλές φορές και θυμό. Επίσης, εμφανίστηκαν μουσικά κομμάτια που ίσως θα ταίριαζαν πιο πολύ στο συναίσθημα του φόβου, που είναι λογικό μιας και βρίσκονται πολύ κοντά μέσα στο ίδιο τεταρτημόριο Q2.
	I Got You (by Bebe Rexha)	0.887	0.349	
	Thunder (by Imagine Dragons)	0.822	0.288	
	I Lived (by OneRepublic)	0.851	0.344	
	Wolves (by Selena Gomez)	0.804	0.306	
	Man Of The Year (by ScHoolboy Q)	0.865	0.371	
Φόβος	Them Bones (by Alice in Chains)	0.946	0.309	Εδώ αλλάξαμε genre, επιλέξαμε 'rock' και 'alternative' γιατί τα αποτελέσματα στο είδος pop ήταν άσχετα με το συναίσθημα που ζητήθηκε. Αρκετά από τα αποτελέσματα φαίνονται να είναι πιο κοντά στο συναίσθημα του φόβου, αλλά και πάλι δεν βρέθηκαν τιμές για τα valence και arousal που να ανταποκρίνονται πλήρως στο ζητούμενο. Δύσκολο και σπάνιο να περάσεις το συναίσθημα φόβου μέσα από τη μουσική.
	Kill Your Heroes (by AWOLNATION)	0.951	0.314	
	Chop Suey! (by System Of A Down)	0.934	0.287	
	Welcome To The Jungle (by Guns N' Roses)	0.954	0.331	
	Riot (by Three Days Grace)	0.948	0.316	
	My Hero (by Foo Fighters)	0.925	0.301	
	Sound of Madness (by Shinedown)	0.939	0.329	
	The Kill (by Thirty Seconds To Mars)	0.912	0.289	
	Even Flow (by Pearl Jam)	0.944	0.265	
Ουδέτερη	The Cure (by Lady Gaga)	0.506	0.516	Στην κατηγορία αυτή, παρατηρούμε ότι τα περισσότερα τραγούδια που εμφανίστηκαν δεν προκαλούν πολύ έντονα
	It's a Vibe (by 2 Chainz)	0.502	0.525	
	Mirrors (by Justin Timberlake)	0.512	0.512	

	Let Me Explain (by Bryson Tiller)	0.504	0.525	συναισθήματα, αλλά σίγουρα δεν θα μπορούσαμε να τα χαρακτηρίσουμε και ως ουδέτερα, αφού η μουσική αποτελεί το κατεξοχήν μέσο μεταφοράς συναισθημάτων.
	Chanel (by Frank Ocean)	0.503	0.473	
	Some Kind Of Drug (by G-Easy)	0.525	0.49	
	Hands To Myself (by Selena Gomez)	0.499	0.466	
	Another Sad Love Song (by Khaalid)	0.472	0.517	
	Strip that down (by Liam Payne)	0.497	0.542	
	Promise (by Kid Ink)	0.469	0.481	
	Say You Won't Let Go (by James Arthur)	0.557	0.494	

Πίνακας 5: Αποτελέσματα και παρατηρήσεις μετά από δοκιμές για τις τιμές των *valence* και *energy*

Στη συνέχεια, δοκιμάζουμε να προσθέσουμε και το *acousticness*, που με βάση τα προαναφερθέντα, οι χαμηλές τιμές (<0.5 , $\text{target} = 0.25$) φαίνεται να ταιριάζουν στα συναισθήματα που ανήκουν στα Q1 (χαρά), Q2 (ένταση/θυμός) ενώ οι υψηλές τιμές (>0.5 , $\text{target} = 0.75$) υποδηλώνουν λύπη ή χαλάρωση, οπότε Q3 και Q4. Έτσι προκύπτει ο ακόλουθος πίνακας:

Συναίσθημα	Αποτελέσματα	Acousticness	Παρατηρήσεις
Χαρά	Bang Bang (by Jessie J)	0.26	Το καινούργιο χαρακτηριστικό δεν φαίνεται να επηρεάζει συναισθηματικά τα αποτελέσματα, αλλά φαίνεται απλά να αλλάζει λίγο το στυλ των τραγουδιών που επιστρέφονται (για χαμηλές τιμές – πιο ηλεκτρονικοί ήχοι) ή το πλήθος των τραγουδιών που εμφανίζονται (για υψηλές τιμές επιστράφηκαν μόνο δυο μουσικά κομμάτια).
	Sucker for Pain (by Lil Wayne)	0.225	
	Gone (by Afrojack)	0.193	
	No Problem (by Chance the Rapper)	0.156	
	Shotgun (by George Ezra)	0.286	
	Woman Like Me (by Little Mix)	0.173	
	On My Mind (by Ellie Goulding)	0.258	
	1, 2, 3 (by Sofia Reyes)	0.165	

	Mala Mia (by Maluma)	0.31	
	Lash Life (by Zara Larson)	0.132	
Λύπη	Dancing On My Own (by Calum Scott)	0.837	Τα αποτελέσματα για υψηλές τιμές, παρόλο που είναι πολύ λίγα σε σχέση με πριν, φαίνεται να έχουν βελτιωθεί (δηλαδή, λείπουν τα τραγούδια όπως The One και Don't). Ενώ για χαμηλές τιμές, τα αποτελέσματα είναι ίδια.
	Angels (by Khalid)	0.959	
	Sweet Creature (by Harry Styles)	0.746	
	I hate u, I love u (by gnash)	0.687	
	Supermarket Flowers (by Ed Sheeran)	0.914	
	Consequences (by Camila Cabello)	0.645	
	Not About Angels (by Birdy)	0.964	
	The One (by The Chainsmokers)	0.238	
	Sober (by Demi Lovato)	0.798	
	Don't (by Bryson Tiller)	0.223	
	lovely (by Billie Eilish)	0.934	
Έκπληξη	High Hopes (by Panic! At The Disco)	0.193	Τα αποτελέσματα φαίνονται να είναι ίδια. Για υψηλές τιμές πάλι δεν εμφανίζονται καθόλου αποτελέσματα. Αυτό είναι λογικό αν σκεφτούμε ότι έχουμε ορίσει ήδη τιμές με περιορισμένο εύρος για τα valence και energy και ότι το acousticness σχετίζεται αρνητικά με το energy.
	Love on Me (by Galantis)	0.113	
	Whistle (by Flo Rida)	0.0208	
	Just Got Paid (by Sigala)	0.0332	
	Ain't Giving Up (by Craig David)	0.0981	
	No Money (by Galantis)	0.0282	
	Still into You (by Paramore)	0.0098	
	Everybody (by Logic)	0.161	
	Timber (by Pitbull)	0.0295	
Αηδία	-	-	Δεν εξετάστηκε
Θυμός	All My Love (by cash cash)	0.148	

	Silence (by Marshmello)	0.256	Για χαμηλές τιμές τα αποτελέσματα είναι τα ίδια, ενώ υπάρχουν και νέα κομμάτια που είναι άσχετα με το συναίσθημα που ζητήθηκε (πχ Silence, Let you love me This is what you came for). Για υψηλές τιμές δεν δίνει αποτελέσματα.
	Tie Me Down (by Gryffin)	0.135	
	Lost In Japan – remix (by Shawn Mendes)	0.0911	
	Polaroid (by Jonas Blue)	0.293	
	Alone (by Alan Walker)	0.186	
	Let You Love Me (by Rita Ora)	0.288	
	Purple Lamborghini (by Skrillex)	0.0019	
	Cruel (by Snakehips)	0.0184	
	I lived (by One Republic)	0.0683	
	This Is What You Came For (by Calvin Harris)	0.199	
Φόβος	Shuffle (by Bombay Bicycle Club)	0.316	Για χαμηλές τιμές, έχουμε παρόμοια αποτελέσματα χωρίς πάλι να ικανοποιείται το ζητούμενο. Για υψηλές τιμές δεν δίνει αποτελέσματα.
	Don't Look Back in Anger (by Oasis)	0.0708	
	Be Yourself (by Audioslave)	0.143	
	Them Bones (by Alice In Chains)	0.0235	
	Kill Your Heroes (by AWOLNATION)	0.00768	
	Chop Suey! (by System Of A Down)	0.000278	
	Welcome To The Jungle (by Guns N' Roses)	0.0235	
	Roots (by In This Moment)	0.0319	
	Riot (by Three Days Grace)	0.0104	
	The Doomed (by A Perfect Circle)	0.052	
	Out In The Fields (by Gary Moore)	0.0275	
Ουδέτερη	I'm not the only one (by Sam Smith)	0.529	Δεν εμφανίζεται κάποια σημαντική διαφορά ως προς το συναίσθημα. Οι τιμές που
	Fast Car (by Jonas Blue)	0.484	

	I Like Me Better (by Lauv)	0.535	εφαρμόστηκαν είναι 0.4-0.6 με target = 0.5.
	Girls Like You (by Maroon 5)	0.568	
	Body (by Dreezy)	0.554	
	Psycho (by Post Malone)	0.546	
	Issues (by Julia Michaels)	0.401	
	Like I'm Gonna Lose You (by Meghan Trainor)	0.4	

Πίνακας 6 : Αποτελέσματα και παρατηρήσεις μετά από δοκιμές για τις τιμές του acoustiness

Εναλλακτικά με το acoustiness, δοκιμάσαμε να προσθέσουμε το danceability. Διαισθητικά, μπορούμε να διαπιστώσουμε ότι τραγούδια που εμφανίζουν υψηλές τιμές θα είναι χαρούμενα (Q1), ενώ στις χαμηλές τιμές θα είναι η λύπη και η ηρεμία (Q4). Η προσθήκη όμως αυτού του χαρακτηριστικού δεν προσθέτει κάποια βελτίωση, όλα τα αποτελέσματα ήταν παρόμοια.

Συμπεράσματα

Συμπερασματικά, οι τιμές των valence και energy δίνουν μια πολύ καλή προσέγγιση του συστήματος που θέλαμε να δημιουργήσουμε, η οποία εμφανίζεται μέσα από τις μετρήσεις στις οποίες καταλήξαμε. Παρόλα αυτά, δεν μπορούμε να φτιάξουμε ένα ιδανικό σύστημα προτάσεων βασισμένοι αποκλειστικά στο μοντέλο του Russell. Η προσθήκη άλλων χαρακτηριστικών δεν φαίνεται να άλλαξε το συναισθηματικό αποτύπωμα της λίστας των τραγουδιών που επιστρέφονται. Επιβεβαιώνεται, λοιπόν, ότι αυτά τα δύο ακουστικά χαρακτηριστικά παίζουν το μεγαλύτερο ρόλο (σε σχέση με τα υπόλοιπα χαρακτηριστικά του Spotify) έκφρασης συναισθήματος στη μουσική. Θα πρέπει να έχουμε στο μυαλό μας ότι πολλά συναισθήματα μπορεί να εκφράζονται μέσα σε κάθε μουσικό κομμάτι. Η μουσική είναι μια μορφή τέχνης και ποτέ δε θα μπορεί να κριθεί με έναν μόνο τρόπο, με ένα μοναδικό συναίσθημα, αλλά σίγουρα υπάρχουν πολλές μέθοδοι που βελτιώνουν τα αποτελέσματα συστάσεων.

















4. Επίλογος

Τα συστήματα συστάσεων χρησιμοποιούνται ευρέως για να παρέχουν στους χρήστες προτάσεις αντικειμένων με βάση τις προτιμήσεις τους. Με τον συνεχώς αυξανόμενο όγκο πληροφοριών στο διαδίκτυο, τα συστήματα συστάσεων είναι ένα χρήσιμο εργαλείο για την αντιμετώπιση της υπερφόρτωσης πληροφοριών. Διάφορες εφαρμογές έχουν ήδη υιοθετήσει συστήματα συστάσεων, όπως το ηλεκτρονικό εμπόριο, η υγεία, οι μεταφορές, ο τουρισμός και η εκπαίδευση. Η χρήση αποτελεσματικών και ακριβών τεχνικών συστάσεων είναι πολύ σημαντική για ένα σύστημα που θα παρέχει καλές και χρήσιμες συστάσεις στους μεμονωμένους χρήστες του. Έτσι, και ο τομέας της μουσικής έχει υιοθετήσει ιδιαίτερα συστήματα συστάσεων αλλά έχει δυνατότητες για περαιτέρω βελτιώσεις αν για παράδειγμα λάβει υπόψη τα συναισθήματα του χρήστη. Πολλές έρευνες γίνονται τα τελευταία χρόνια σε συστήματα συστάσεων μουσικής με βάση το συναίσθημα και η αναγνώριση συναισθήματος είναι ένας τεράστιος τομέας με μεγάλη χρησιμότητα στην ανάπτυξη αυτών των ερευνών.

Μέσα από την εργασία μας, δώσαμε ένα παράδειγμα μιας εφαρμογής που μπορεί να δημιουργηθεί χρησιμοποιώντας την αναγνώριση συναισθήματος μέσα από εικόνες του προσώπου του χρήστη και προτείνοντας τραγούδια με βάση το συναίσθημά του. Βελτιώσεις θα μπορούσαν να γίνουν, όπως να δίνεται η δυνατότητα στο χρήστη να φτιάχνει προφίλ, να διαλέγει αγαπημένους καλλιτέχνες και τραγούδια και να επιλέγει αν θέλει να βελτιώσει τη διάθεση του ή όχι. Στις μελλοντικές επεκτάσεις εντάσσεται επίσης η offline αξιολόγηση του συστήματος μέσα από διάφορες μετρικές. Όλα τα παραπάνω θα οδηγήσουν σε ένα πιο ολοκληρωμένο σύστημα συστάσεων μουσικής με βάση το συναίσθημα. Στο μέλλον, ελπίζουμε να επεκταθεί περαιτέρω αυτός ο τομέας και τα MRS να λαμβάνουν υπόψη όχι μόνο το ιστορικό και τις προτιμήσεις του χρήστη αλλά και τα συναισθήματα, την προσωπικότητά του και το περιβάλλον ακρόασης.

Παράρτημα 1: Facial Action Coding System

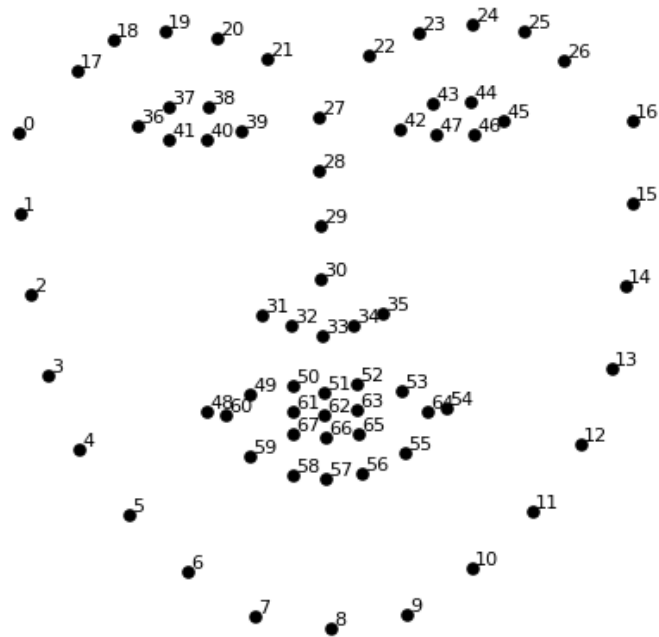
Το FACS είναι ένα σύστημα μέτρησης των κινήσεων του ανθρώπινου προσώπου. Για να κατανοήσουμε λίγο καλύτερα πως λειτουργεί αυτό το σύστημα κωδικοποίησης και να διευρύνουμε τις γνώσεις μας, παραθέτουμε τις ακόλουθες εικόνες που θεωρήθηκαν σημαντικές για την ανάλυσή μας σχετικά με τις εκφράσεις του προσώπου.

Upper Face Action Units					
AU 1	AU 2	AU 4	AU 5	AU 6	AU 7
					
Inner Brow Raiser	Outer Brow Raiser	Brow Lowerer	Upper Lid Raiser	Cheek Raiser	Lid Tightener
*AU 41	*AU 42	*AU 43	AU 44	AU 45	AU 46
					
Lid Droop	Slit	Eyes Closed	Squint	Blink	Wink
Lower Face Action Units					
AU 9	AU 10	AU 11	AU 12	AU 13	AU 14
					
Nose Wrinkler	Upper Lip Raiser	Nasolabial Deepener	Lip Corner Puller	Cheek Puffer	Dimpler
AU 15	AU 16	AU 17	AU 18	AU 20	AU 22
					
Lip Corner Depressor	Lower Lip Depressor	Chin Raiser	Lip Puckerer	Lip Stretcher	Lip Funneler
AU 23	AU 24	*AU 25	*AU 26	*AU 27	AU 28
					
Lip Tightener	Lip Pressor	Lips Part	Jaw Drop	Mouth Stretch	Lip Suck

Εικόνα 35: Βασικά Action Units – οι εκφράσεις χωρίζονται σε άνω (upper face) και κάτω (lower face) προσώπου. [29]

Category	AUs	Category	AUs
Happy	12, 25	Sadly disgusted	4, 10
Sad	4, 15	Fearfully angry	4, 20, 25
Fearful	1, 4, 20, 25	Fearfully surprised	1, 2, 5, 20, 25
Angry	4, 7, 24	Fearfully disgusted	1, 4, 10, 20, 25
Surprised	1, 2, 25, 26	Angrily surprised	4, 25, 26
Disgusted	9, 10, 17	Disgusted surprised	1, 2, 5, 10
Happily sad	4, 6, 12, 25	Happily fearful	1, 2, 12, 25, 26
Happily surprised	1, 2, 12, 25	Angrily disgusted	4, 10, 17
Happily disgusted	10, 12, 25	Awed	1, 2, 5, 25
Sadly fearful	1, 4, 15, 25	Appalled	4, 9, 10
Sadly angry	4, 7, 15	Hatred	4, 7, 10
Sadly surprised	1, 4, 25, 26	-	-

Εικόνα 36: Τα AUs που παρατηρούνται σε κάθε κατηγορία βασικών και σύνθετων συναισθημάτων [6]



Εικόνα 37: 68 Facial Landmarks – τα σημεία κλειδιά του προσώπου [111]

Παράρτημα 2: Οδηγός εγκατάστασης του Flutter

Ο οδηγός εγκατάστασης που δημιουργήθηκε αφορά λειτουργικό σύστημα Windows και δίνει συγκεκριμένες οδηγίες για χρήση Android Studio ως editor.

Πώς ξεκινάμε:

- με χρήση android studio (ή vscode) ως επιλογή για editor
- για να κατεβάσεις το Flutter πρέπει να έχεις προεγκατεστημένα τα ακόλουθα:

Windows PowerShell 5.0 ή νεότερο (είναι ήδη εγκατεστημένο στα Windows 10)

Git για Windows 2.x, με την επιλογή Use Git από το Windows Command Prompt

Για την εγκατάσταση του Git:

- πλοηγήσου στη σελίδα <https://git-scm.com/download/>

- διάλεξε το λειτουργικό σύστημα που διαθέτεις

- επέλεξε κατέβασμα και εγκατάσταση του git

* Για windows θυμήσου να ορίσεις τις κατάλληλες μεταβλητές περιβάλλοντος (environment variables) αν αυτό είναι απαραίτητο.

Για την εγκατάσταση του Flutter:

- πλοηγήσου στη σελίδα <https://docs.flutter.dev/get-started/install>

- διάλεξε το λειτουργικό σύστημα που διαθέτεις

- κατέβασε το installation bundle ή πάρε το source code με την εντολή:

```
C:\src> git clone https://github.com/flutter/flutter.git -b stable
```

- άνοιξε το φάκελο flutter στη διεύθυνση C:\src\flutter

- άνοιξε το flutter console

* για να λειτουργήσει αυτό το βήμα βεβαιώσου ότι όρισες τις κατάλληλες μεταβλητές περιβάλλοντος Windows, όπως: powershell, git/bin, git/cmd, flutter/bin

- τρέξε την εντολή 'flutter doctor'

Στην οθόνη που βλέπεις, αν δεν υπάρχει κάποιο πρόβλημα, θα εμφανιστούν τα ακόλουθα:

```
C:\Users\frena>flutter doctor
Doctor summary (to see all details, run flutter doctor -v):
[✓] Flutter (Channel stable, 2.10.4, on Microsoft Windows [Version 10.0.19043.1586], locale el-GR)
[✓] Android toolchain - develop for Android devices (Android SDK version 30.0.3)
[✓] Chrome - develop for the web
[✓] Visual Studio - develop for Windows (Visual Studio Build Tools 2019 16.11.10)
[✓] Android Studio (version 4.2)
[✓] VS Code (version 1.66.0)
[✓] Connected device (3 available)
[✓] HTTP Host Availability

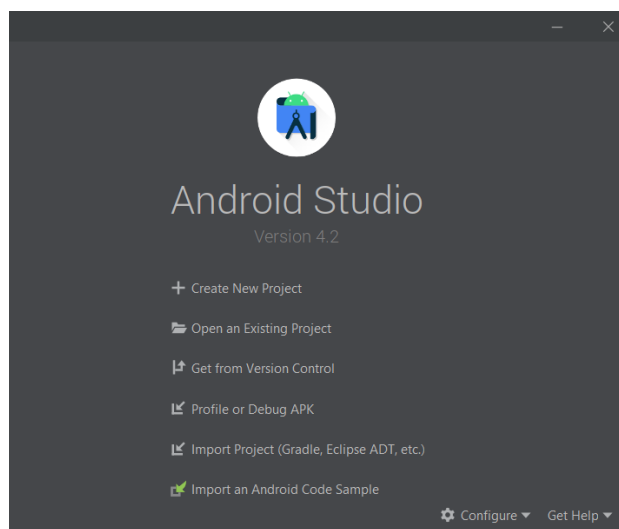
• No issues found!
```

Μας ενδιαφέρουν κυρίως τα Flutter και Android Studio.

Σημείωση: Αν υπάρχει πρόβλημα με το Android toolchain ακολουθήσε τα παρακάτω βήματα:

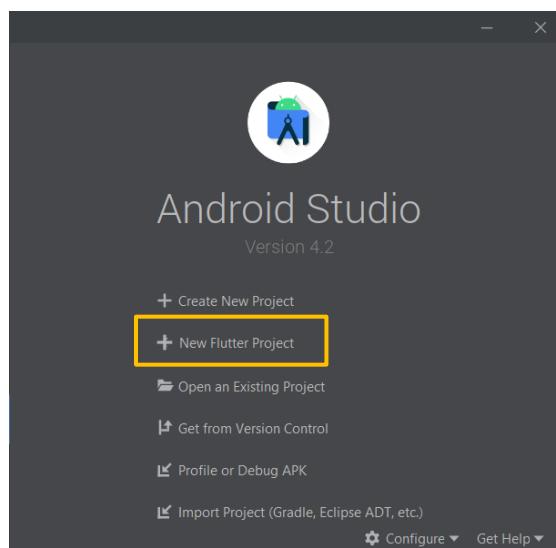
- Άνοιξε το Android Studio (κλείσε όποιο project είναι ανοιχτό). Στη συνέχεια πήγαινε στο Configure → SDK Manager → Android SDK.
- Πήγαινε στην καρτέλα SDK tools.
- Επίλεξε το Android SDK Command-line Tools (latest) και πάτα Apply.
- Τρέξε την εντολή 'flutter doctor --android-licenses' και πάτα αποδοχή σε όλα τα licenses.

Στη συνέχεια, εφόσον έχει ολοκληρωθεί η παραπάνω εγκατάσταση ανοίγουμε το Android Studio:



Και όπως φαίνεται στην παραπάνω εικόνα, επέλεξε: Configure → Plugins → στην καρτέλα Marketplace στο search bar γράφουμε flutter → κάνουμε install το Flutter plugin → click Restart IDE.

Στην οθόνη έναρξης του Android Studio θα υπάρχει πλέον η επιλογή New Flutter Project!



Πριν ξεκινήσουμε το καινούργιο project θα χρειαστούμε ένα virtual device. Για να φτιάξεις ένα virtual device:

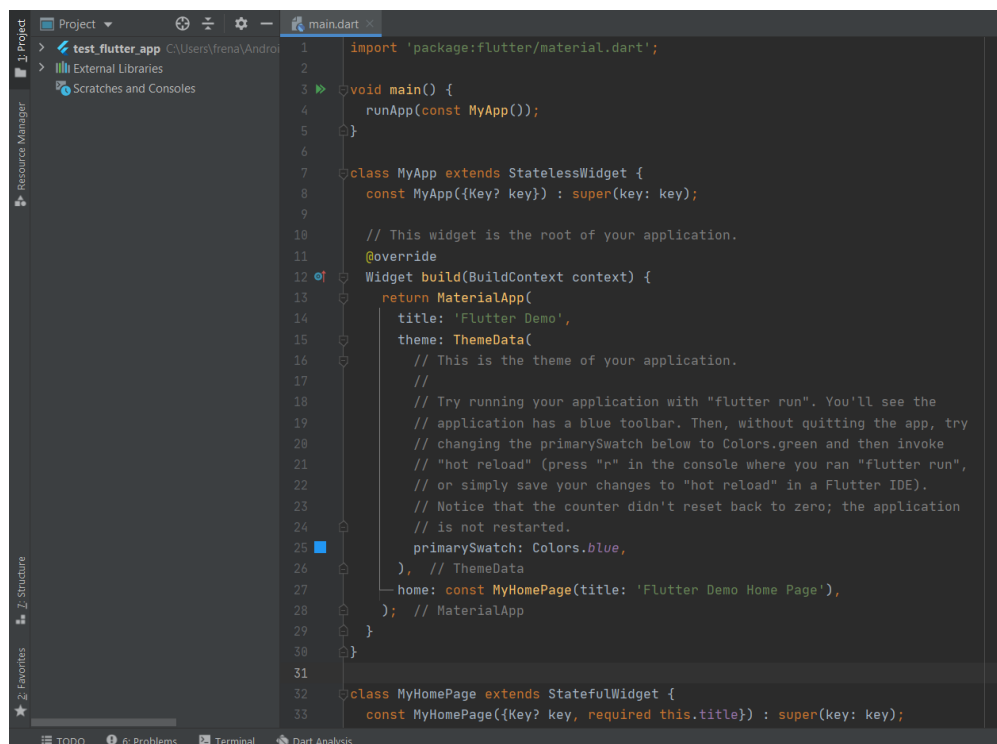
- πήγαινε στο Configure → AVD Manager → Create Virtual Device
- επέλεξε τα χαρακτηριστικά που θέλεις πχ. Nexus 6 → επιλογή Next → Pie → επιλογή Next → στο Graphics βάζουμε Hardware-GLES2.0 → Finish

Η συσκευή αυτή θα εμφανίζεται πλέον στα Connected devices του Android Studio.

Πώς φτιάχνουμε το πρώτο μας project:

- επέλεξε New flutter project από το Android Studio menu
- είσαγε το flutter SDK path πχ. C:\src\flutter
- είσαγε όνομα για το project

Μετά από αυτά έχουμε πλέον μπροστά μας τον κώδικα για το dummy app (main.dart) με το οποίο ξεκινάει κάθε flutter project.



```
1 import 'package:flutter/material.dart';
2
3 void main() {
4   runApp(const MyApp());
5 }
6
7 class MyApp extends StatelessWidget {
8   const MyApp({Key? key}) : super(key: key);
9
10  // This widget is the root of your application.
11  @override
12  Widget build(BuildContext context) {
13    return MaterialApp(
14      title: 'Flutter Demo',
15      theme: ThemeData(
16        // This is the theme of your application.
17        //
18        // Try running your application with "flutter run". You'll see the
19        // application has a blue toolbar. Then, without quitting the app, try
20        // changing the primarySwatch below to Colors.green and then invoke
21        // "hot reload" (press "r" in the console where you ran "flutter run",
22        // or simply save your changes to "hot reload" in a Flutter IDE).
23        // Notice that the counter didn't reset back to zero; the application
24        // is not restarted.
25        primarySwatch: Colors.blue,
26      ), // ThemeData
27      home: const MyHomePage(title: 'Flutter Demo Home Page'),
28    ); // MaterialApp
29  }
30 }
31
32 class MyHomePage extends StatefulWidget {
33   const MyHomePage({Key? key, required this.title}) : super(key: key);
```

Πώς χρησιμοποιούμε τα packages του Flutter:

Για να ενσωματώσουμε κάποιες πιο περίπλοκες λειτουργίες στην εφαρμογή μας μπορούμε να χρησιμοποιήσουμε τα Flutter Packages. Για να δεις όλα τα διαθέσιμα πακέτα πλοηγήσου στη σελίδα <https://pub.dev/> (official package repository for Dart and Flutter apps). Για παράδειγμα, για να χρησιμοποιήσουμε την κάμερα του κινητού μπορούμε να διαλέξουμε το camera package. Για να γίνει αυτό πηγαίνουμε στη σελίδα <https://pub.dev/packages/camera> και ακολουθούμε τις οδηγίες όπως αναφέρονται και παρακάτω:

- Πρόσθεσε τη γραμμή: `camera: ^0.10.0+1` στο αρχείο `pubspec.yaml` στην ενότητα των `dependencies`.
- Πάτα `Get dependencies` για να ενημερωθεί το project ή εκτέλεσε την εντολή `flutter pub get`.
- Στο αρχείο που θέλεις να χρησιμοποιήσεις την κάμερα πρόσθεσε το ακόλουθο:

```
import 'package:camera/camera.dart';
```

Βιβλιογραφία

- [1] L. Shu *et al.*, “A Review of Emotion Recognition Using Physiological Signals,” doi: 10.3390/s18072074.
- [2] S. De Nadai *et al.*, “Enhancing safety of transport by road by on-line monitoring of driver emotions,” *2016 11th Syst. Syst. Eng. Conf. SoSE 2016*, 2016, doi: 10.1109/SYSOSE.2016.7542941.
- [3] R. Guo, S. Li, L. He, W. Gao, H. Qi, and G. Owens, “Pervasive and unobtrusive emotion sensing for human mental health,” *Proc. 2013 7th Int. Conf. Pervasive Comput. Technol. Healthc. Work. PervasiveHealth 2013*, pp. 436–439, 2013, doi: 10.4108/icst.pervasivehealth.2013.252133.
- [4] J. Kumari, R. Rajesh, and K. M. Pooja, “Facial Expression Recognition: A Survey,” *Procedia Comput. Sci.*, vol. 58, pp. 486–491, 2015, doi: 10.1016/j.procs.2015.08.011.
- [5] A. Dziedzickis, A. Uras Kaklauskas, and V. Bucinskas, “Human Emotion Recognition: Review of Sensors and Methods,” doi: 10.3390/s20030592.
- [6] B. Chul and K. Id, “A Brief Review of Facial Emotion Recognition Based on Visual Information,” doi: 10.3390/s18020401.
- [7] A. Jadhav, S. Lone, S. Matey, T. Madamwar, and S. Jakhete, “Survey on Face Detection Algorithms,” *Int. J. Innov. Sci. Res. Technol.*, vol. 6, no. 2, pp. 291–297, 2021, [Online]. Available: www.ijisrt.com.
- [8] M. Maithri *et al.*, “Automated emotion recognition: Current trends and future perspectives,” *Comput. Methods Programs Biomed.*, vol. 215, Mar. 2022, doi: 10.1016/J.CMPB.2022.106646.
- [9] P. Ekman, “Basic Emotions,” *University of California, San Francisco, CA, USA*. 1999, doi: 10.1007/978-3-319-28099-8_495-1.
- [10] J. A. Russell, “A circumplex model of emotion,” *J. Pers. Soc. Psychol.*, vol. 39, no. 6, pp. 1161–1178, 1980.
- [11] M. Schedl, H. Zamani, C.-W. Chen, Y. Deldjoo, · Mehdi Elahi, and M. Elahi, “Current challenges and visions in music recommender systems research,” *Int. J. Multimed. Inf. Retr.*, vol. 7, pp. 95–116, 2018, doi: 10.1007/s13735-018-0154-2.
- [12] Z. Fayyaz, M. Ebrahimian, D. Nawara, A. Ibrahim, and R. Kashef, “Recommendation Systems: Algorithms, Challenges, Metrics, and Business Opportunities,” doi: 10.3390/app10217748.
- [13] H. Ko, S. Lee, Y. Park, and A. Choi, “A Survey of Recommendation Systems: Recommendation Models, Techniques, and Application Fields,” 2022, doi: 10.3390/electronics11010141.
- [14] F. O. Isinkaye, Y. O. Folajimi, and B. A. Ojokoh, “Recommendation systems: Principles, methods and evaluation,” *Egypt. Informatics J.*, vol. 16, no. 3, pp. 261–273, 2015, doi: 10.1016/j.eij.2015.06.005.
- [15] M. A. Riegler, J. Wang, L. Su, and M. Schedl, “Deep Learning in Music Recommendation Systems,” *Front. Appl. Math. Stat. / www.frontiersin.org*, vol. 1, p. 44, 2019, doi: 10.3389/fams.2019.00044.
- [16] M. Xiaoxi, L. Weisi, H. Dongyan, D. Minghui, and H. Li, “Facial emotion

recognition,” *2017 IEEE 2nd Int. Conf. Signal Image Process. ICSIP 2017*, vol. 2017-Janua, pp. 77–81, 2017, doi: 10.1109/SIPROCESS.2017.8124509.

- [17] S. Sosnowski, A. Bittermann, K. Kühnlenz, and M. Buss, “Design and evaluation of emotion-display EDDIE,” *IEEE Int. Conf. Intell. Robot. Syst.*, pp. 3113–3118, 2006, doi: 10.1109/IROS.2006.282330.
- [18] H. Kalantarian, K. Jedoui, P. Washington, and D. P. Wall, “A Mobile Game for Automatic Emotion-Labeling of Images,” *IEEE Trans. Games*, vol. 12, no. 2, pp. 213–218, 2020, doi: 10.1109/TG.2018.2877325.
- [19] C. Xiaotong, W. Xiaoxia, O. Tante, and F. Zhengzhi, “Advances in Emotion Recognition: Link to Depressive Disorder.” .
- [20] O. Gervasi, V. Franzoni, M. Riganelli, and S. Tasso, “Automating facial emotion recognition,” *Web Intell.*, vol. 17, no. 1, pp. 17–27, 2019, doi: 10.3233/WEB-190397.
- [21] D. Mehta, M. Faridul, H. Siddiqui, and A. Y. Javaid, “Facial Emotion Recognition: A Survey and Real-World User Experiences in Mixed Reality,” doi: 10.3390/s18020416.
- [22] P. Ekman, E. R. Sorenson, and W. V. Friesen, “Pan-culture Element in Facial Display of Emotion.pdf,” *Science*, vol. 164, no. 3875, pp. 86–88, 1969.
- [23] E. P. and F. W. V., “Facial Action Coding System,” 1978.
<https://psycnet.apa.org/doiLanding?doi=10.1037%2Ft27734-000> (accessed Sep. 23, 2022).
- [24] K. Hevner, “Experimental Studies of the Elements of Expression in Music,” *Am. J. Psychol.*, vol. 48, no. 2, p. 246, 1936, doi: 10.2307/1415746.
- [25] O. Meyers, “Mysoundtrack: A commonsense playlist generator,” *Massachusetts Inst. Technol.*, pp. 1–3, 2005, [Online]. Available: <http://web.media.mit.edu/~meyers/mysoundtrack.pdf>.
- [26] A. Mehrabian, “Pleasure-Arousal-Dominance: A general framework for describing and measuring individual differences in temperament,” *Curr. Psychol.*, vol. 14, no. 4, pp. 261–292, 1996, doi: 10.1007/bf02686918.
- [27] K. Sailunaz, M. Dhaliwal, J. Rokne, and R. Alhajj, “Emotion detection from text and speech: a survey,” *Soc. Netw. Anal. Min.*, vol. 8, no. 1, pp. 1–8, 2018, doi: 10.1007/s13278-018-0505-2.
- [28] T. Eerola and J. K. Vuoskoski, “A comparison of the discrete and dimensional models of emotion in music,” *Psychol. Music*, vol. 39, no. 1, pp. 18–49, 2011, doi: 10.1177/0305735610362821.
- [29] F. Z. Canal *et al.*, “A survey on facial emotion recognition techniques: A state-of-the-art literature review,” *Inf. Sci. (Ny)*, vol. 582, pp. 593–617, Jan. 2022, doi: 10.1016/J.INS.2021.10.005.
- [30] J. H. Cheong, T. Xie, S. Byrne, and L. J. Chang, “Py-Feat: Python Facial Expression Analysis Toolbox,” 2021, [Online]. Available: <http://arxiv.org/abs/2104.03509>.
- [31] M. K. Hasan, M. S. Ahsan, Abdullah-Al-Mamun, S. H. S. Newaz, and G. M. Lee, “Human face detection techniques: A comprehensive review and future research directions,” *Electronics (Switzerland)*, vol. 10, no. 19. 2021, doi: 10.3390/electronics10192354.

- [32] H. Hatem, Z. Beiji, and R. Majeed, "A Survey of Feature Base Methods for Human Face Detection," *Int. J. Control Autom.*, vol. 8, no. 5, pp. 61–78, 2015, doi: 10.14257/ijca.2015.8.5.07.
- [33] K. Dang and S. Sharma, "Review and comparison of face detection algorithms," *Proc. 7th Int. Conf. Conflu. 2017 Cloud Comput. Data Sci. Eng.*, pp. 629–633, 2017, doi: 10.1109/CONFLUENCE.2017.7943228.
- [34] A. Zimman, J. A. Berliner, and T. G. Graeber, "A Survey of Recent Advances in Face Detection," *Methods Mol. Biol.*, vol. 1000, no. June, pp. 53–69, 2013, doi: 10.1007/978-1-62703-405-0_5.
- [35] A. V. Iyer, V. Pasad, S. R. Sankhe, and K. Prajapati, "Emotion based mood enhancing music recommendation," *RTEICT 2017 - 2nd IEEE Int. Conf. Recent Trends Electron. Inf. Commun. Technol. Proc.*, vol. 2018-Janua, pp. 1573–1577, 2017, doi: 10.1109/RTEICT.2017.8256863.
- [36] S. V. Machines, "18.9 Μηχανές Διανυσμάτων Υποστήριξης," pp. 485–487, 2000.
- [37] A. Papadopoulos and University of Patras, "Convolutional Neural Networks and their usage on computer vision," 2016.
- [38] P. Naga, S. Das Marri, and R. Borreo, "Facial emotion recognition methods, datasets and technologies: A literature survey," *Mater. Today Proc.*, no. xxxx, 2021, doi: 10.1016/j.matpr.2021.07.046.
- [39] "AdaBoost Algorithm." <https://www.mygreatlearning.com/blog/adaboost-algorithm/> (accessed Aug. 19, 2022).
- [40] F. Y. Shih, C. F. Chuang, and P. S. P. Wang, "Performance comparisons of facial expression recognition in JAFFE database," *Int. J. Pattern Recognit. Artif. Intell.*, vol. 22, no. 3, pp. 445–459, 2008, doi: 10.1142/S0218001408006284.
- [41] S. Busemann *et al.*, "The Workshop Programme."
- [42] "MMI Facial Expression Database." <https://mmifacedb.eu/> (accessed Aug. 18, 2022).
- [43] T. Debnath, M. M. Reza, A. Rahman, A. Beheshti, S. S. Band, and H. Alinejad-Rokny, "Four-layer ConvNet to facial emotion recognition with minimal epochs and the significance of data diversity," *Sci. Rep.*, vol. 12, no. 1, pp. 1–13, 2022, doi: 10.1038/s41598-022-11173-0.
- [44] S. Li, W. Deng, and J. Du, "Reliable Crowdsourcing and Deep Locality-Preserving Learning for Expression Recognition in the Wild," pp. 2852–2861.
- [45] S. Zafeiriou and I. Kotsia, "Aff-Wild : Valence and Arousal ‘ in-the-wild ’ Challenge," pp. 34–41.
- [46] D. Kollias *et al.*, "Deep Affect Prediction in-the-Wild: Aff-Wild Database and Challenge, Deep Architectures, and Beyond Keywords Deep · Convolutional · Recurrent · Aff-Wild · Database · Challenge · In-the-wild · Facial · Dimensional · Categorical · Emotion · Recognition · Valence · Arousal · AffWildNet · RECOLA · AFEW · AFEW-VA · EmotiW," *Int. J. Comput. Vis.*, vol. 127, no. 3, pp. 907–929, 2019, doi: 10.1007/s11263-019-01158-4.
- [47] A. Mollahosseini, B. Hasani, and M. H. Mahoor, "AffectNet: A Database for Facial Expression, Valence, and Arousal Computing in the Wild," *IEEE Trans. Affect. Comput.*, vol. 10, no. 1, pp. 18–31, 2019, doi: 10.1109/TAFFC.2017.2740923.

- [48] C. F. Benitez-Quiroz, R. Srinivasan, and A. M. Martinez, "EmotioNet: An accurate, real-time algorithm for the automatic annotation of a million facial expressions in the wild," *Proc. IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognit.*, vol. 2016-December, pp. 5562–5570, 2016, doi: 10.1109/CVPR.2016.600.
- [49] Y. Khairuddin and Z. Chen, "Facial Emotion Recognition: State of the Art Performance on FER2013."
- [50] A. Alreshidi and M. Ullah, "informatics Facial Emotion Recognition Using Hybrid Features," doi: 10.3390/informatics7010006.
- [51] J. Chen, D. Chen, Y. Gong, M. Yu, K. Zhang, and L. Wang, "Facial expression recognition using geometric and appearance features," *ACM Int. Conf. Proceeding Ser.*, pp. 29–33, 2012, doi: 10.1145/2382336.2382345.
- [52] Q. Yuan, "Research on Classroom Emotion Recognition Algorithm Based on Visual Emotion Classification," vol. 2022, 2022.
- [53] H. Jung, S. Lee, J. Yim, S. Park, and J. Kim, "Joint fine-tuning in deep neural networks for facial expression recognition," *Proc. IEEE Int. Conf. Comput. Vis.*, vol. 2015 Inter, pp. 2983–2991, 2015, doi: 10.1109/ICCV.2015.341.
- [54] P. Shen, S. Wang, and Z. Liu, "from Infrared Thermal Videos," pp. 323–333, 2013.
- [55] A. Maalej, B. Ben Amor, M. Daoudi, A. Srivastava, and S. Berretti, "Shape analysis of local facial patches for 3D facial expression recognition," *Pattern Recognit.*, vol. 44, no. 8, pp. 1581–1589, 2011, doi: 10.1016/j.patcog.2011.02.012.
- [56] Sujonoa and Alexander A S Gunawan, "Face Expression Detection on Kinect using Active Appearance Model and Fuzzy Logic." .
- [57] D. Kollias, V. Sharmanska, and S. Zafeiriou, "IEEE TRANSACTIONS ON PATTERN ANALYSIS AND MACHINE INTELLIGENCE 1 Distribution Matching for Heterogeneous Multi-Task Learning: a Large-scale Face Study." [Online]. Available: <https://ibug.doc.ic.ac.uk/resources/fg-2020-competition-affective->.
- [58] Y. Song, S. Dixon, and M. T. Pearce, "A survey of music recommendation systems and future perspectives," *Proc. 9th Int. Symp. Comput. Music Model. Retr.*, no. June, pp. 395–410, 2012.
- [59] R. Panda, R. M. Malheiro, and R. P. Paiva, "Audio Features for Music Emotion Recognition: a Survey," *IEEE Trans. Affect. Comput.*, vol. 3045, no. c, pp. 1–1, 2020, doi: 10.1109/taffc.2020.3032373.
- [60] T. Petri, "Exploring relationships between audio features and emotion in music," *Front. Hum. Neurosci.*, vol. 3, no. Escom, pp. 260–264, 2009, doi: 10.3389/conf.neuro.09.2009.02.033.
- [61] Z. Fu, G. Lu, K. M. Ting, and D. Zhang, "A survey of audio-based music classification and annotation," *IEEE Trans. Multimed.*, vol. 13, no. 2, pp. 303–319, 2011, doi: 10.1109/TMM.2010.2098858.
- [62] X. Hu, "Music and mood: Where theory and reality meet," *Proc. iConference*, pp. 1–8, 2010, [Online]. Available: <http://www.ideals.illinois.edu/handle/2142/14956>.
- [63] X. Yang, Y. Dong, and J. Li, "Review of data features-based music emotion recognition methods," *Multimed. Syst.*, vol. 24, no. 4, pp. 365–389, 2018, doi: 10.1007/s00530-017-0559-4.

- [64] K. Pyrovolakis, P. Tzouveli, and G. Stamou, "Multi-Modal Song Mood Detection with Deep Learning[†]," *Sensors*, vol. 22, no. 3. 2022, doi: 10.3390/s22031065.
- [65] S. Naseri, S. Reddy, J. Correia, J. Karlgren, and R. Jones, "The Contribution of Lyrics and Acoustics to Collaborative Understanding of Mood," *Proc. Int. AAAI Conf. Web Soc. Media*, vol. 16, no. 1, pp. 687–698, 2022, [Online]. Available: <https://ojs.aaai.org/index.php/ICWSM/article/view/19326>.
- [66] M. Cora Urdaneta-Ponte, A. Mendez-Zorrilla, I. Oleagordia-Ruiz, G. Kostopoulos, and S. Kotsiantis, "electronics Recommendation Systems for Education: Systematic Review," 2021, doi: 10.3390/electronics10141611.
- [67] A. Singhal, P. Sinha, and R. Pant, "Use of Deep Learning in Modern Recommendation System: A Summary of Recent Works," 2017.
- [68] E. Rich, "User Modeling via Stereotypes*," *Cogn. Sci.*, vol. 3, no. 4, pp. 329–354, 1979, doi: 10.1207/s15516709cog0304_3.
- [69] S. K. Raghuwanshi and R. K. Pateriya, *Recommendation Systems: Techniques, Challenges, Application, and Evaluation*, vol. 2, no. January. Springer Singapore, 2019.
- [70] D. Goldberg, D. Nichols, B. M. Oki, and D. Terry, "Using collaborative filtering to Weave an Information tapestry," *Commun. ACM*, vol. 35, no. 12, pp. 61–70, 1992, doi: 10.1145/138859.138867.
- [71] M. Naumov *et al.*, "Deep Learning Recommendation Model for Personalization and Recommendation Systems."
- [72] R. Burke, "Hybrid Recommender Systems: Survey and Experiments," *User Model. User-adapt. Interact.*, vol. 12, no. 4, pp. 331–370, 2002, [Online]. Available: <http://www.springerlink.com/openurl.asp?id=doi:10.1023/A:1021240730564%5Cnpapers2://publication/doi/10.1023/A:1021240730564>.
- [73] L. Shah, H. Gaudani, and P. Balani, "Survey on Recommendation System," *Int. J. Comput. Appl.*, vol. 137, no. 7, pp. 43–49, 2016, doi: 10.5120/ijca2016908821.
- [74] "Google Play." https://play.google.com/store/games?hl=en_US&gl=US (accessed Sep. 01, 2022).
- [75] D. Paul and S. Kundu, *A Survey of Music Recommendation Systems with a Proposed Music Recommendation System*, vol. 937. Springer Singapore, 2020.
- [76] A. Patel and R. Wadhvani, "A Comparative Study of Music Recommendation Systems," *2018 IEEE Int. Students' Conf. Electr. Electron. Comput. Sci. SCEECS 2018*, pp. 1–4, 2018, doi: 10.1109/SCEECS.2018.8546852.
- [77] Y. Deldjoo, "Content-driven Music Recommendation: Evolution, State of the Art, and Challenges," 2021. [Online]. Available: <https://www.dbpedia.org>.
- [78] A. Niyazov, E. Mikhailova, and O. Egorova, "Content-based Music Recommendation System," *Conf. Open Innov. Assoc. Fruct*, vol. 2021-May, pp. 274–279, 2021, doi: 10.23919/FRUCT52173.2021.9435533.
- [79] K. Chen, B. Liang, X. Ma, and M. Gu, "Learning audio embeddings with user listening data for content-based music recommendation," *ICASSP, IEEE Int. Conf. Acoust. Speech Signal Process. - Proc.*, vol. 2021-June, pp. 3015–3019, 2021, doi: 10.1109/ICASSP39728.2021.9414458.

- [80] H. Zarzour, Z. Al-Sharif, M. Al-Ayyoub, and Y. Jararweh, "A new collaborative filtering recommendation algorithm based on dimensionality reduction and clustering techniques," *2018 9th Int. Conf. Inf. Commun. Syst. ICICS 2018*, vol. 2018-Janua, pp. 102–106, 2018, doi: 10.1109/IACS.2018.8355449.
- [81] W. Wenzhen, "Personalized music recommendation algorithm based on hybrid collaborative filtering technology," *Proc. - 2019 Int. Conf. Smart Grid Electr. Autom. ICSGEA 2019*, pp. 280–283, 2019, doi: 10.1109/ICSGEA.2019.00071.
- [82] A. S. Girsang, A. Wibowo, Jason, and Roslynlia, "Neural Collaborative For Music Recommendation System," *IOP Conf. Ser. Mater. Sci. Eng.*, vol. 1071, no. 1, p. 012021, 2021, doi: 10.1088/1757-899x/1071/1/012021.
- [83] A. Gatzoura, J. Vinagre, A. M. Jorge, and M. Sanchez-Marre, "A Hybrid Recommender System for Improving Automatic Playlist Continuation," *IEEE Trans. Knowl. Data Eng.*, vol. 33, no. 5, pp. 1819–1830, 2021, doi: 10.1109/TKDE.2019.2952099.
- [84] K. Y. Cao, Y. Liu, and H. X. Zhang, "Improving the Cold Start Problem in Music Recommender Systems," *J. Phys. Conf. Ser.*, vol. 1651, no. 1, 2020, doi: 10.1088/1742-6596/1651/1/012067.
- [85] Adiyansjah and D. Gunawan, Alexander A S Suhartono, "Music Recommender System Based on Genre using Convolutional Recurrent Neural Networks." .
- [86] H. I. James, J. J. A. Arnold, J. M. M. Ruban, M. Tamilarasan, and R. Saranya, "IRJET- EMOTION BASED MUSIC RECOMMENDATION SYSTEM."
- [87] A. Alrihaili, A. Alsaedi, K. Albalawi, and L. Syed, "Music recommender system for users based on emotion detection through facial features," *Proc. - Int. Conf. Dev. eSystems Eng. DeSE*, vol. October-20, pp. 1014–1019, 2019, doi: 10.1109/DeSE.2019.00188.
- [88] V. Babanne, M. Borgaonkar, M. Katta, P. Kudale, and V. Deshpande, "EMOTION based PERSONALIZED RECOMMENDATION SYSTEM," *Int. Res. J. Eng. Technol.*, pp. 701–705, 2020, [Online]. Available: www.irjet.net.
- [89] P. Helmholtz, M. Meyer, and S. Robra-Bissantz, "Feel the moosic: Emotion-based music selection and recommendation," *32nd Bled eConference Humaniz. Technol. a Sustain. Soc. BLED 2019 - Conf. Proc.*, no. June, pp. 203–221, 2020, doi: 10.18690/978-961-286-280-0.11.
- [90] S. Gilda, H. Zafar, C. Soni, and K. Waghurdekar, "Smart music player integrating facial emotion recognition and music mood recommendation," *Proc. 2017 Int. Conf. Wirel. Commun. Signal Process. Networking, WiSPNET 2017*, vol. 2018-Janua, pp. 154–158, 2018, doi: 10.1109/WiSPNET.8299738.
- [91] V. Moscato, A. Picariello, and G. Sperli, "An Emotional Recommender System for Music," *IEEE Intell. Syst.*, vol. 36, no. 5, pp. 57–68, 2021, doi: 10.1109/MIS.2020.3026000.
- [92] S. Yousefian Jazi, M. Kaedi, and A. Fatemi, "An emotion-aware music recommender system: bridging the user's interaction and music recommendation," *Multimed. Tools Appl.*, vol. 80, no. 9, pp. 13559–13574, 2021, doi: 10.1007/s11042-020-10386-7.
- [93] C. Y. Wang, Y. C. Wang, and S. C. T. Chou, "A context and emotion aware system for personalized music recommendation," *J. Internet Technol.*, vol. 19, no. 3, pp. 765–779, 2018, doi: 10.3966/160792642018051903013.

- [94] “Spotify.” <https://open.spotify.com/> (accessed Sep. 09, 2022).
- [95] “Spotify company info.” <https://newsroom.spotify.com/company-info/> (accessed Sep. 09, 2022).
- [96] “Spotify Research.” <https://research.atspotify.com/> (accessed Sep. 11, 2022).
- [97] “Uncovering how the Spotify algorithm works.” <https://towardsdatascience.com/uncovering-how-the-spotify-algorithm-works-4d3c021ebc0> (accessed Sep. 11, 2022).
- [98] “Inside Spotify’s Recommender System: A Complete Guide to Spotify Recommendation Algorithms.” <https://www.music-tomorrow.com/blog/how-spotify-recommendation-system-works-a-complete-guide-2022> (accessed Sep. 11, 2022).
- [99] “How does Spotify’s recommendation system work.” <https://www.univ.ai/post/spotify-recommendations> (accessed Sep. 11, 2022).
- [100] “Flutter.” <https://flutter.dev/> (accessed Sep. 08, 2022).
- [101] “Dart.” <https://dart.dev/> (accessed Sep. 08, 2022).
- [102] “FAQ Flutter.” <https://docs.flutter.dev/resources/faq> (accessed Sep. 08, 2022).
- [103] “Get started with Flutter.” <https://docs.flutter.dev/get-started/install> (accessed Sep. 08, 2022).
- [104] “Py-Feat.” <https://py-feat.org/pages/intro.html> (accessed Sep. 08, 2022).
- [105] “Py-Feat Detecting single face.” https://py-feat.org/basic_tutorials/02_detector_imgs.html#detecting-a-single-face-from-a-single-image (accessed Sep. 08, 2022).
- [106] “Spotify Web API.” <https://developer.spotify.com/documentation/web-api/> (accessed Sep. 11, 2022).
- [107] “Understanding Spotify Web API.” <https://engineering.atspotify.com/2015/03/understanding-spotify-web-api/> (accessed Sep. 11, 2022).
- [108] “Spotify for Developers - Get track’s audio features.” <https://developer.spotify.com/console/get-audio-features-track/> (accessed Sep. 09, 2022).
- [109] H. Sharma, S. Gupta, Y. Sharma, and A. Purwar, “A New Model for Emotion Prediction in Music,” *2020 6th Int. Conf. Signal Process. Commun. ICSC 2020*, pp. 156–161, 2020, doi: 10.1109/ICSC48311.2020.9182745.
- [110] R. Panda, H. Redinho, C. Gonçalves, R. Malheiro, and R. P. Paiva, “HOW DOES the SPOTIFY API COMPARE to the MUSIC EMOTION RECOGNITION STATE-OF-THE-ART?,” *Proc. Sound Music Comput. Conf.*, vol. 2021-June, pp. 238–245, 2021.
- [111] “Face landmark detection using python.” <https://towardsdatascience.com/face-landmark-detection-using-python-1964cb620837> (accessed Sep. 13, 2022).