



Τμήμα Ηλεκτρολόγων Μηχανικών  
& Μηχανικών Υπολογιστών

**ΠΑΝΕΠΙΣΤΗΜΙΟ  
ΠΕΛΟΠΟΝΝΗΣΟΥ**

**ΣΧΟΛΗ ΜΗΧΑΝΙΚΩΝ**  
**ΤΜΗΜΑ ΗΛΕΚΤΡΟΛΟΓΩΝ**  
**ΜΗΧΑΝΙΚΩΝ & ΜΗΧΑΝΙΚΩΝ**  
**ΥΠΟΛΟΓΙΣΤΩΝ**

Διπλωματική εργασία

**«ΚΑΤΗΓΟΡΙΟΠΟΙΗΣΗ ΜΟΥΣΙΚΗΣ ΒΑΣΕΙ  
ΣΥΝΑΙΣΘΗΜΑΤΩΝ ΜΕ ΧΡΗΣΗ ΤΕΧΝΙΚΩΝ  
ΜΗΧΑΝΙΚΗΣ ΜΑΘΗΣΗΣ»**

**Θεολογία Γουγούλα Α.Μ.2980**

Επιβλέπων Καθηγητής: Ιωάννης Τζήμας

**Πάτρα 2023**

Εγκρίθηκε από την τριμελή εξεταστική επιτροπή

Πάτρα, //2023

## ΕΠΙΤΡΟΠΗ ΑΞΙΟΛΟΓΗΣΗΣ

I.

II.

III.

### **Υπεύθυνη Δήλωση Φοιτητή**

*Βεβαιώνω ότι είμαι συγγραφέας αυτής της εργασίας και ότι κάθε βοήθεια την οποία είχα για την προετοιμασία της είναι πλήρως αναγνωρισμένη και αναφέρεται στην εργασία. Επίσης έχω αναφέρει τις όποιες πηγές από τις οποίες έκανα χρήση δεδομένων, ιδεών ή λέξεων, είτε αυτές αναφέρονται ακριβώς είτε παραφρασμένες. Επίσης βεβαιώνω ότι αυτή η εργασία προετοιμάστηκε από εμένα προσωπικά ειδικά για τη συγκεκριμένη εργασία.*

*Η έγκριση της διπλωματικής εργασίας από το Τμήμα Ηλεκτρολόγων Μηχανικών και Μηχανικών Υπολογιστών του Πανεπιστημίου Πελοποννήσου δεν υποδηλώνει απαραίτητως και αποδοχή των απόψεων του συγγραφέα εκ μέρους του Τμήματος.*

*Η παρούσα εργασία αποτελεί πνευματική ιδιοκτησία της φοιτήτριας Γουγούλα Θεολογίας που την εκπόνησε. Στο πλαίσιο της πολιτικής ανοικτής πρόσβασης ο συγγραφέας/δημιουργός εκχωρεί στο Πανεπιστήμιο Πελοποννήσου, μη αποκλειστική άδεια χρήσης του δικαιώματος αναπαραγωγής, προσαρμογής, δημόσιου δανεισμού, παρουσίασης στο κοινό και ψηφιακής διάχυσής τους διεθνώς, σε ηλεκτρονική μορφή και σε οποιοδήποτε μέσο, για διδακτικούς και ερευνητικούς σκοπούς, άνευ ανταλλάγματος και για όλο το χρόνο διάρκειας των δικαιωμάτων πνευματικής ιδιοκτησίας. Η ανοικτή πρόσβαση στο πλήρες κείμενο για μελέτη και ανάγνωση δεν σημαίνει καθ' οιονδήποτε τρόπο παραχώρηση δικαιωμάτων διανοητικής ιδιοκτησίας του συγγραφέα/δημιουργού ούτε επιτρέπει την αναπαραγωγή, αναδημοσίευση, αντιγραφή, αποθήκευση, πώληση, εμπορική χρήση, μετάδοση, διανομή, έκδοση, εκτέλεση, «μεταφόρτωση» (downloading), «ανάρτηση» (uploading), μετάφραση, τροποποίηση με οποιονδήποτε τρόπο, τμηματικά ή περιληπτικά της εργασίας, χωρίς τη ρητή προηγούμενη έγγραφη συναίνεση του συγγραφέα/δημιουργού. Ο συγγραφέας/δημιουργός διατηρεί το σύνολο των ηθικών και περιουσιακών του δικαιωμάτων.*

## I. Εισαγωγή

Για πολλούς η Μουσική είναι ένας από τους καλύτερους τρόπους έκφρασης συναισθημάτων. Η σχέση μουσικής και συναισθήματος έχει απασχολήσει αρκετούς ερευνητές και ως ερευνητικό πεδίο της επιστήμης της πληροφορικής ανήκει στον ευρύτερο κλάδο της Music Information Retrieval.

## II. Περίληψη

Η παρούσα εργασία θα καλύψει την ανάλυση της μουσικής με χρήση τεχνικών Μηχανικής Μάθησης με σκοπό τη συναισθηματική κατηγοριοποίηση της αφού πρώτα δώσει το θεωρητικό πλαίσιο. Η συναισθηματική αντίδραση στη μουσική είναι διαφορετική για κάθε άτομο οπότε η ανάλυση της είναι πιθανό να μη δώσει τέλεια αποτελέσματα. Η μέθοδος που ακολουθείται είναι η χρήση ορισμένων τραγουδιών που ενσωματώνουν μία κοινή διάθεση ως ground truth, και έπειτα η αντιστοίχιση άλλων τραγουδιών στις κατηγορίες αυτές. Η μέθοδος αυτή είναι χρήσιμη για την αυτόματη αναγνώριση μουσικών ειδών, την ταξινόμηση μεγάλου όγκου μουσικής η ακόμη και στη συμπεριφορική μελέτη ανθρώπων.

“Without music, life would be a mistake.”

— Friedrich Nietzsche, *Twilight of the Idols*

# Περιεχόμενα

Κεφάλαιο 1: Εισαγωγή .....	6
1.1. Τεχνητή νοημοσύνη - Artificial Intelligence (AI) .....	6
1.2. Μηχανική μάθηση - Machine learning (ML).....	6
1.2.1. Επιβλεπόμενη Μηχανική Μάθηση .....	7
1.2.2. Μη επιβλεπόμενη μηχανική μάθηση .....	8
1.2.3. Ημι-επιβλεπόμενη μηχανική μάθηση.....	9
1.3. Διαδικασία της επιβλεπόμενης ML .....	9
1.4. Τεχνητά Νευρωνικά Δίκτυα - Neural networks (ΤΝΔ) .....	10
1.5. Παραδείγματα χρήσης Μηχανικής Μάθησης.....	11
Κεφάλαιο 2: Αντικείμενο μελέτης.....	11
2.1. Ανάκτηση Πληροφορίας της Μουσικής - Music Information Retrieval (MIR).....	11
2.2. Αναγνώριση Συναισθήματος από Μουσική - Music Emotion Recognition (MER).....	12
2.3. Προκλήσεις - Προβλήματα στην MER.....	13
2.4. Θεωρία της συναισθηματικής ταξινόμησης.....	14
2.5. Λογιστική Παλινδρόμηση.....	17
2.5.1. Logistic regression as NN.....	17
2.5.2. Μαθηματικά πίσω από τη Λογιστική Παλινδρόμηση.....	18
2.6. Αλγόριθμος K-πλησιέστερων γειτόνων .....	18
2.7. Μετρικές .....	19
2.7.1. Recall .....	19
2.7.2. Confusion Matrix .....	20
2.7.3. Accuracy score.....	20
2.7.4. F1 score.....	21
2.8. State-of-the-art και μειονεκτήματα.....	21
2.8.1. State-of-the-art Τεχνικές .....	21
2.8.2. Δεδομένα.....	22
2.8.3. Μειονεκτήματα .....	23
Κεφάλαιο 3: Πειραματική διαδικασία .....	23
3.1. Συλλογή των δεδομένων .....	23
3.2. Ετικετοποίηση των δεδομένων .....	25
3.3. Train/Test Split Method.....	25

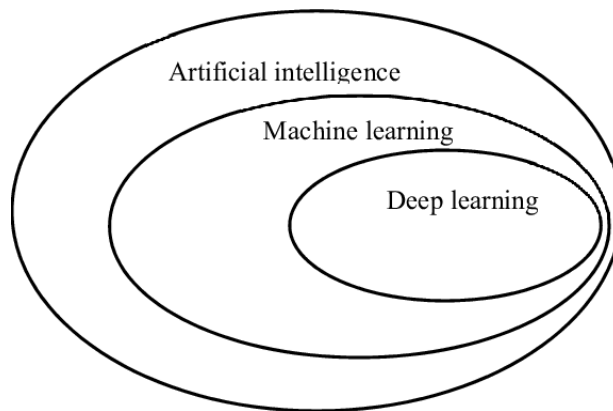
3.4. Τεχνολογίες που χρησιμοποιήθηκαν.....	26
3.4.1. Python 3 .....	26
3.4.2. Pandas .....	26
3.4.3. Plotly .....	26
3.4.4. Seaborn .....	26
3.4.5. Scikit-Learn.....	26
3.4.6. Visual Studio Code .....	27
3.4.7. Jupyter Notebook .....	27
3.4.9. Spotify Web API.....	27
3.4.10. Spotipy .....	27
Κεφάλαιο 6: Παράθεση κώδικα.....	27
6.1. Δημιουργία του Dataset .....	27
6.2. Κανονικοποίηση δεδομένων, εκπαίδευση μοντέλων & δημιουργία γραφημάτων.....	30
Κεφάλαιο 7: Παρατηρήσεις & Συμπεράσματα.....	42
7.1. Αξιολόγηση .....	42
7.1.1. Ακρίβεια.....	42
7.1.2. Πίνακας Σύγχυσης (Confusion matrix).....	43
7.2. Ευρήματα .....	44
7.3. Συμπέρασμα .....	45
Κεφάλαιο 8: Βιβλιογραφία .....	46



# Κεφάλαιο 1: Εισαγωγή

## 1.1. Τεχνητή νοημοσύνη - Artificial Intelligence (AI)

Ο κλάδος της τεχνητής νοημοσύνης προσπαθεί να κατανοήσει και να κατασκευάσει οντότητες με νοημοσύνη. Σύμφωνα με τους Bart και Feigenbaum [1] είναι ο σχεδιασμός συστημάτων που αναπαράγουν τις γνωστικές λειτουργίες ενός ανθρώπου και μιμούνται ό,τι σχετίζεται με τη νοημοσύνη στην ανθρώπινη συμπεριφορά. Ένα σύστημα τεχνητής νοημοσύνης μπορεί ως ένα βαθμό να είναι ικανό να προσαρμόσει τη συμπεριφορά του αναλύοντας τις συνέπειες προηγούμενων δράσεων, δίνοντας έτσι λύση σε προβλήματα αυτόνομα. Είναι κλάδος της επιστήμης των υπολογιστών, για την ανάπτυξη όμως συστημάτων που χρησιμοποιούν τεχνητή νοημοσύνη απαιτείται, πολλές φορές, η μελέτη επιπλέον επιστημών όπως της ιατρικής, της ψυχολογίας κ.α.



Σχήμα 1

## 1.2. Μηχανική μάθηση - Machine learning (ML)

Ορίζοντας στη μάθηση, βάσει των Written και Frank [2], κάποια οντότητα μαθαίνει όταν αλλάζει τη συμπεριφορά της έτσι ώστε να αποδίδει καλύτερα μελλοντικά. Η Μηχανική Μάθηση είναι η υπο-ενότητα της Τεχνητής Νοημοσύνης που ασχολείται με τον προγραμματισμό υπολογιστών μέσω δεδομένων. Είναι η δημιουργία μοντέλων και προτύπων από μια σειρά υπολογιστών με σκοπό τη μάθηση. Το μοντέλο αποτελεί μία απλουστευμένη μορφή του περιβάλλοντος και το πρότυπο ορίζεται ως μία δομή αποτελούμενη από παρόμοιες εμπειρίες. Ένας πιο τυπικός ορισμός [3]: Ένα πρόγραμμα θεωρούμε ότι μαθαίνει βάσει μίας εμπειρίας  $E$ , αν σε σχέση με έναν στόχο  $T$  και με μία μετρική απόδοσης  $P$ , η απόδοση  $P$  βελτιώνεται με τρόπο ανάλογο της εμπειρίας  $E$ .

Στην ουσία ένας αλγόριθμος μηχανικής μάθησης κατασκευάζει ένα μαθηματικό μοντέλο βάσει των δεδομένων εκπαίδευσης που του δίνονται. Το μοντέλο χρησιμοποιείται για την πρόβλεψη και τη λήψη αποφάσεων σχετικά με άλλα δεδομένα, τα οποία αποτελούν τα δεδομένα ελέγχου. Οι αλγόριθμοι μηχανικής μάθησης ταξινομούνται σε τρεις κατηγορίες, την **επιβλεπόμενη**

(Supervised) μάθηση, τη **μη επιβλεπόμενη** (Unsupervised) και την **ενισχυτική** (Reinforcement).

Στην παρούσα εργασία αξίζει να εστιάσουμε στην επιβλεπόμενη μάθηση.

### 1.2.1. Επιβλεπόμενη Μηχανική Μάθηση

Η Επιβλεπόμενη Μηχανική Μάθηση είναι μια υποκατηγορία της Τεχνητής Νοημοσύνης και αφορά τη χρήση αλγορίθμων μηχανικής μάθησης για την εκμάθηση μοτίβων από ετικεταρισμένα δεδομένα. Συγκεκριμένα, αυτό σημαίνει ότι χρησιμοποιούνται δεδομένα που έχουν ετικέτες (labels) για την εκπαίδευση του συστήματος. Στόχος είναι να εκπαιδευτεί ένα μοντέλο που να μπορεί να κάνει ακριβείς προβλέψεις για νέα, απρόβλεπτα δεδομένα. Αυτή η μέθοδος χρησιμοποιείται σε πολλούς τομείς, όπως οι επιστήμες των δεδομένων, οι αγορές μετοχών, η ιατρική, η καταναλωτική ανάλυση και πολλοί άλλοι.

Στα προβλήματα μάθησης με επίβλεψη, υπάρχουν δεδομένα και οι αντίστοιχες ετικέτες τους. Τα δεδομένα χρησιμοποιούνται ως υλικό εκπαίδευσης στο μοντέλο, το οποίο μέσω της λειτουργίας του προσπαθεί να τα ταξινομήσει όσο το δυνατόν πιο κοντά στην βασική αλήθεια (ground truth). Ο απώτερος στόχος είναι να μπορεί να εφαρμόσει αυτή τη γνώση σε άγνωστα δεδομένα.

Πιο συγκεκριμένα, στην επιβλεπόμενη μάθηση ή μάθηση από παραδείγματα, το σύστημα μάθησης λαμβάνει ένα σύνολο δεδομένων εκπαίδευσης με τη μορφή  $(x_i, y_i)$ , όπου  $i = 1, 2, \dots$ , λαμβάνεται ένα σύνολο δεδομένων εκπαίδευσης που αποτελείται από ζεύγη της μορφής  $N$ . Κάθε ζεύγος αποτελείται από ένα διάνυσμα τιμών  $x_i$  των χαρακτηριστικών εισόδου και την αντίστοιχη τιμή της μεταβλητής απόφασης  $y_i$ . Με βάση αυτά τα δεδομένα εκπαίδευσης, δημιουργείται ένα μοντέλο ταξινόμησης για την πρόβλεψη της τιμής της μεταβλητής απόφασης  $y$  σε μελλοντικά δεδομένα  $x$ . Έτσι, η δημιουργία του μοντέλου είναι επαγωγική και η εφαρμογή του για την πρόβλεψη νέων τιμών της περίπτωσης είναι επαγωγική [4]. Γνωστοί αλγόριθμοι επιβλεπόμενης μάθησης περιλαμβάνουν τους K-nn, SVM, Naive Bayes, νευρωνικά δίκτυα, δέντρα αποφάσεων, λογιστική παλινδρόμηση, τυχαία δάση και γραμμική παλινδρόμηση.[4]

Η εκτίμηση της αποτελεσματικότητας - ακρίβειας του μοντέλου σε ένα πρόβλημα επιβλεπόμενης μάθησης γίνεται χρησιμοποιώντας ένα ξεχωριστό σύνολο δεδομένων, το οποίο ονομάζεται δείγμα ελέγχου (test set). Για το δείγμα ελέγχου είναι γνωστή η κατηγοριοποίηση των περιπτώσεων που περιέχει, όμως δεν χρησιμοποιείται κατά τη διάρκεια της εκπαίδευσης. Έτσι, η ορθότητα της κατηγοριοποίησης των περιπτώσεων του δείγματος ελέγχου αποτελεί μία καλή εκτίμηση για την αποτελεσματικότητα του μοντέλου.

Ανάλογα με τη μεταβλητή απόφασης  $y$ , τα προβλήματα εκπαίδευσης με επιβλεπόμενη μάθηση μπορούν να ομαδοποιηθούν σε προβλήματα κατηγοριοποίησης και παλινδρόμησης.[5]

(Σχήμα 2)

- **Κατηγοριοποίηση** (Classification): μία από τις βασικές εφαρμογές της μηχανικής μάθησης. Στην κατηγοριοποίηση, ένας αλγόριθμος μηχανικής μάθησης εκπαιδεύεται να αναθέτει σε κάθε είσοδο (δεδομένα) μια από πολλές προκαθορισμένες κατηγορίες (labels).

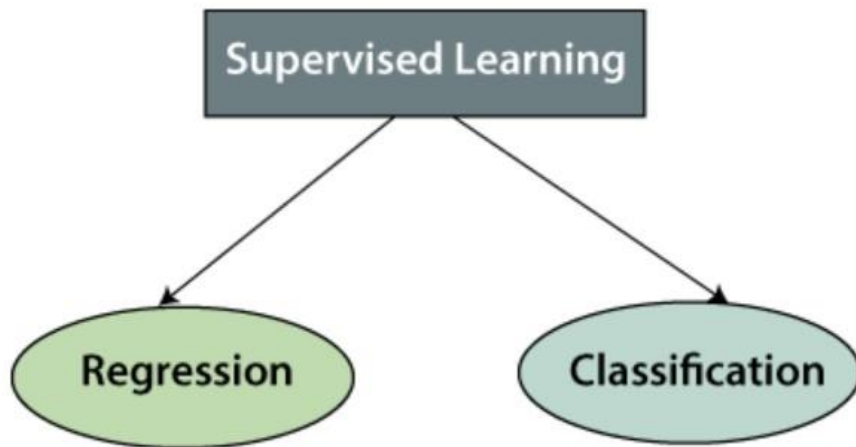
Για παράδειγμα, ένας αλγόριθμος κατηγοριοποίησης μπορεί να εκπαιδευτεί να αναγνωρίζει εικόνες και να τις κατατάσσει σε κατηγορίες, όπως "σκύλοι" ή "γάτες".



Αλλά η κατηγοριοποίηση μπορεί να χρησιμοποιηθεί σε πολλές άλλες εφαρμογές, όπως η αναγνώριση φωνής, η ταξινόμηση κειμένου σε κατηγορίες, η πρόβλεψη πελατών που θα αγοράσουν ένα προϊόν και πολλά άλλα.

- **Παλινδρόμηση (Regression):** αφορά τη διαδικασία όπου ένας αλγόριθμος μηχανικής μάθησης εκπαιδεύεται να εκτιμά μια συνεχή τιμή εξόδου (output) βάσει μιας ή περισσότερων εισόδων (inputs).

Για παράδειγμα, ένας αλγόριθμος παλινδρόμησης μπορεί να εκπαιδευτεί να προβλέπει την τιμή ενός ακινήτου βάσει διαφόρων χαρακτηριστικών, όπως ο αριθμός των δωματίων, η τοποθεσία και η έτη κατασκευής. Η παλινδρόμηση μπορεί να χρησιμοποιηθεί σε πολλές άλλες εφαρμογές, όπως η πρόβλεψη της τιμής των μετοχών ή η πρόβλεψη του καταναλωτικού πατρών ενός πελάτη βάσει διαφόρων χαρακτηριστικών.



Σχήμα 2

### 1.2.2. Μη επιβλεπόμενη μηχανική μάθηση

Η μη επιβλεπόμενη μηχανική μάθηση είναι μια κατηγορία της μηχανικής μάθησης όπου ο στόχος είναι να ανακαλυφθούν δομικά μοτίβα και σχέσεις σε ένα σύνολο δεδομένων χωρίς καμία προηγούμενη γνώση ή ετικέτα για το κάθε δείγμα. Αντίθετα με την επιβλεπόμενη μάθηση, όπου τα δεδομένα έχουν ετικέτες ή σωστές απαντήσεις για κάθε δείγμα, η μη επιβλεπόμενη μάθηση αποσκοπεί στην ανακάλυψη δομικών παραμέτρων που μπορούν να βοηθήσουν στην κατηγοριοποίηση ή στην επεξεργασία των δεδομένων. Αυτό μπορεί να γίνει μέσω αλγορίθμων clustering, που αναγνωρίζουν ομάδες δεδομένων με κοινά χαρακτηριστικά, ή μέσω διάφορων μεθόδων ανάλυσης δεδομένων όπως οικογενειακή ανάλυση, ανάλυση κυρίαρχων συνιστωσών και άλλες.

Η μη επιβλεπόμενη μάθηση είναι χρήσιμη όταν δεν υπάρχει προηγούμενη γνώση για τα δεδομένα ή όταν ο στόχος είναι να ανακαλυφθούν κρυφά μοτίβα και σχέσεις στα δεδομένα. Η ικανότητά του να βρίσκει ομοιότητες και διαφορές στις πληροφορίες το καθιστά ιδανική λύση

για διερευνητική ανάλυση δεδομένων, στρατηγικές cross-selling, τμηματοποίηση πελατών και αναγνώριση εικόνας.

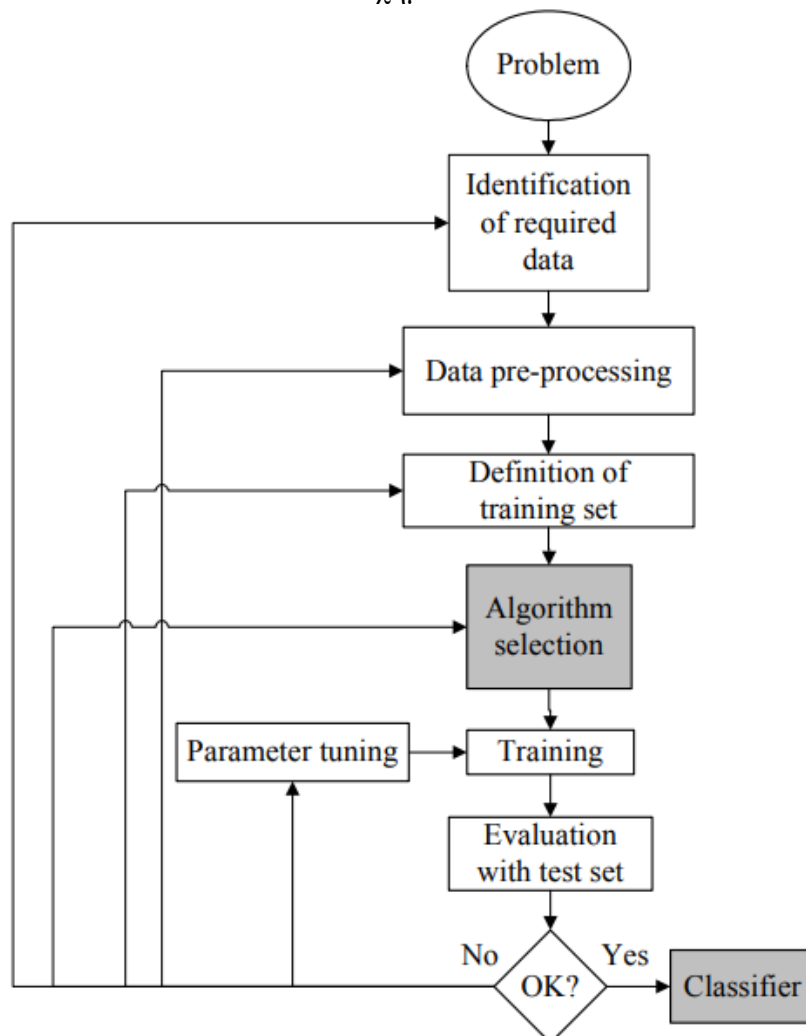
### 1.2.3. Ημι-επιβλεπόμενη μηχανική μάθηση

Η ημι-εποπτευόμενη μάθηση είναι μια τεχνική μηχανικής μάθησης στην οποία το μοντέλο μαθαίνει από έναν συνδυασμό εποπτευόμενης και μη εποπτευόμενης μάθησης. Η εποπτευόμενη μάθηση είναι όταν ένας αλγόριθμος χρησιμοποιεί ένα σύνολο δεδομένων που έχουν επισημανθεί ή ταξινομηθεί, προκειμένου να μάθει πώς να κάνει προβλέψεις για νέα δεδομένα. Η μάθηση χωρίς επίβλεψη είναι όταν ο αλγόριθμος αναζητά μοτίβα και δομές στα δεδομένα χωρίς ετικέτα. Η ημι-εποπτευόμενη μάθηση συνδυάζει αυτές τις δύο προσεγγίσεις για να βελτιώσει την απόδοση του αλγορίθμου και να μάθει από δεδομένα χωρίς ετικέτα. Αυτή η τεχνική χρησιμοποιείται συχνά σε περιπτώσεις όπου τα δεδομένα με ετικέτα είναι σπάνια και το μοντέλο πρέπει να μάθει από δεδομένα χωρίς ετικέτα.

### 1.3. Διαδικασία της επιβλεπόμενης ML

Η τυπική διαδικασία της επιβλεπόμενης μηχανικής μάθησης φαίνεται στο παρακάτω σχήμα [6]

Σχήμα 3



- Καθορισμός του προβλήματος.
- Συλλογή των δεδομένων του προβλήματος.
- Προ-επεξεργασία των δεδομένων.
- Καθορισμός του συνόλου εκπαίδευσης και του συνόλου ελέγχου.
- Επιλογή του κατάλληλου ταξινομητή (classifier).
- Εκπαίδευση του ταξινομητή με τη χρήση των δεδομένων του συνόλου εκπαίδευσης για την κατασκευή ενός μοντέλου κατηγοριοποίησης.
- Αξιολόγηση των αποτελεσμάτων με το σύνολο ελέγχου.

#### 1.4. Τεχνητά Νευρωνικά Δίκτυα - Neural networks (ΤΝΔ)

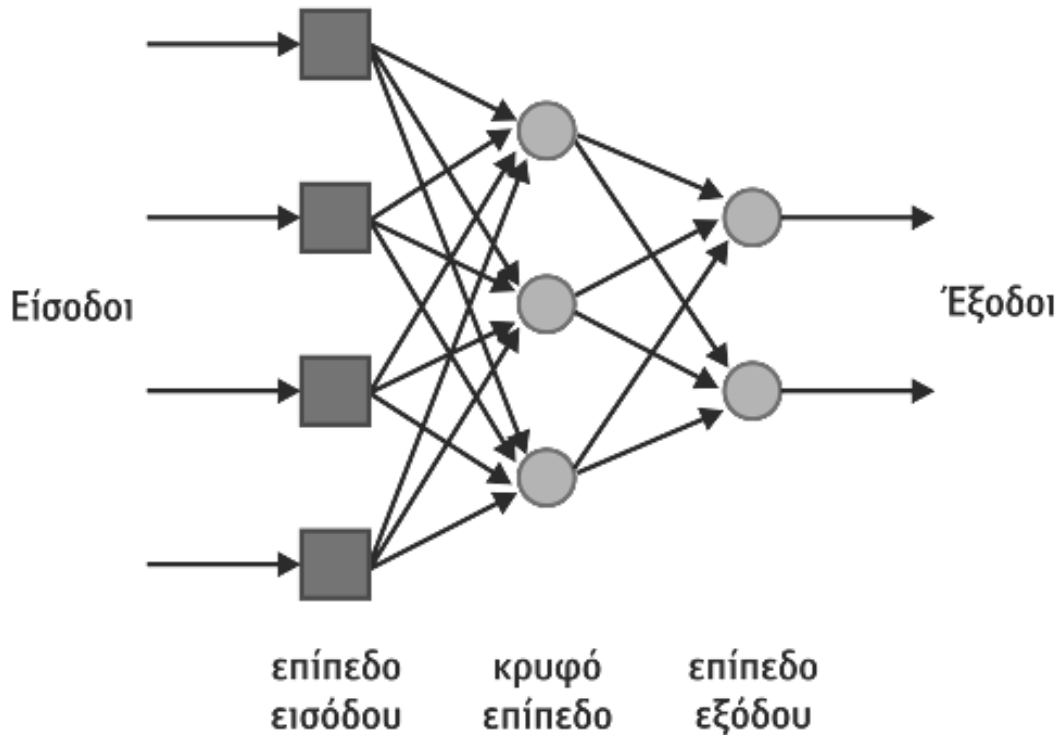
Τα Τεχνητά Νευρωνικά Δίκτυα που χρησιμοποιούνται στη μηχανική μάθηση βασίζονται σε μεγάλο βαθμό, στο βιολογικό νευρικό σύστημα. αποτελούν ένα σύστημα επεξεργασίας δεδομένων με ορισμένα κοινά χαρακτηριστικά απόδοσης με τα βιολογικά νευρωνικά δίκτυα. συνθέτουν δηλαδή ένα μοντέλο παρόμοιο τον βιολογικό, δεν προσπαθούν όμως να το αντιγράψουν. Το κύριο στοιχείο σε ένα τεχνητό νευρωνικό δίκτυο είναι ο νευρώνας. Ένα ΤΝΔ είναι αποτέλεσμα της διασύνδεσης ενός μεγάλου αριθμού νευρώνων οι οποίοι επεξεργάζονται δεδομένα για την επίλυση συγκεκριμένων προβλημάτων.

Τα σήματα μεταδίδονται μεταξύ των νευρώνων μέσω συνδέσμων. Κάθε σύνδεσμος έχει ένα σχετικό βάρος, το οποίο σε ένα τυπικό νευρωνικό δίκτυο πολλαπλασιάζει το μεταδιδόμενο σήμα. Κάθε νευρώνας εφαρμόζει μια μη γραμμική συνάρτηση ενεργοποίησης στην αρχική είσοδο για να καθορίσει το σήμα εξόδου.[7]

Ένα νευρωνικό δίκτυο αποτελείται από τρεις βασικές κατηγορίες νευρώνων:

1. Τους νευρώνες εισόδου (input neurons), οι οποίοι δέχονται τις πληροφορίες που θα επεξεργαστούν.
2. Τους νευρώνες εξόδου (output neurons), στους οποίους καταλήγουν τα αποτελέσματα της πιο πάνω επεξεργασίας.
3. Τους ενδιάμεσους νευρώνες, οι οποίοι βρίσκονται μεταξύ των νευρώνων εισόδου και εξόδου, οι οποίοι ονομάζονται κρυφοί νευρώνες (hidden neurons).

Γενικά, ένα αντιπροσωπευτικό σχήμα για την απεικόνιση ενός νευρωνικού δικτύου είναι το παρακάτω:



Σχήμα 4

### 1.5. Παραδείγματα χρήσης Μηχανικής Μάθησης

Οι παραπάνω τεχνολογίες έχουν εφαρμογή σε διάφορους επιστημονικούς τομείς, αλλά και στην καθημερινότητα. Μερικά παραδείγματα είναι τα εξής:

- Επεξεργασία εικόνας (Image processing)
- Βιοπληροφορική (Bioinformatics)
- Επεξεργασία γλώσσας και μετάφραση (Language processing and translation)
- Ανίχνευση διαδρομών (Route detection)
- Αναγνώριση ομιλίας (Speech recognition)
- Πρόβλεψη (Forecasting)

## Κεφάλαιο 2: Αντικείμενο μελέτης

### 2.1. Ανάκτηση Πληροφορίας της Μουσικής - Music Information Retrieval (MIR)

Σύμφωνα με τη βιβλιογραφία, η θεμελίωση του όρου MIR και του ίδιου του κλάδου έγινε τη δεκαετία του 60' από τους ερευνητές Kessler & Lincoln, οι οποίοι όρισαν την MIR ως τη διαδικασία εξαγωγής, από ένα μεγάλο όγκο δεδομένων, τμήματα πληροφορίας που επαληθεύουν κάποια έννοια της Μουσικολογίας. Ως Μουσικολογία ορίζεται η θεωρητική επιστήμη της μελέτης της μουσικής. Είναι λογικό ο στόχος της MIR να έχει αλλάξει από τη δεκαετία του 60' έως τώρα για αυτό δίνεται ένας πιο σύγχρονος ορισμός της MIR είναι το ερευνητικό πεδίο της πληροφορικής που ασχολείται με την εξαγωγή χρήσιμων χαρακτηριστικών από μουσικά κομμάτια, κυρίως από κυματομορφές, την κατηγοριοποίηση της μουσικής βάσει των χαρακτηριστικών αυτών και την ανάπτυξη συστημάτων αναζήτησης και ανάκτησης. [8] Στον κλάδο αυτό γίνεται ευρεία χρήση της Μηχανικής Μάθησης, αλλά αντλεί γνώση και από άλλες επιστήμες και κλάδους όπως η ψυχολογία, η επεξεργασία σημάτων και η μουσικολογία.

Σήμερα, είναι καθημερινή επαφή μας με μουσικό περιεχόμενο μέσω του διαδικτύου. Επιπλέον, η ποσότητα του περιεχομένου αυτού αυξάνεται εκθετικό τρόπο και είναι πολύ εύκολα προσβάσιμο από όλους μέσω διαφόρων πλατφορμών (Spotify, YouTube, Apple music κ.α.) Ωστόσο, αυτή η έκρηξη της μουσικής πληροφορίας την καθιστά μη διαχειρίσιμη και οδηγεί στην ανάγκη αυτόματων τεχνικών για την διαχείριση και ανάκτηση πληροφορίας από τη μουσική.

Η κοινότητα MIR παρακινείται από την ανάγκη να αντιμετωπιστούν οι τεράστιοι όγκοι μουσικών δεδομένων με αυτόματες μεθόδους κατηγοριοποίησης και ανάκτησης. Η έρευνα MIR στοχεύει στη δυνατότητα δημιουργίας εργαλείων για την αυτόματη ανάλυση ήχου και παρτιτούρας. Ο τομέας ασχολείται με μια σειρά προβλημάτων: μεταγραφή ήχου σε παρτιτούρα, ανίχνευση συγχορδιών, διαχωρισμός πηγών, ο διαχωρισμός φωνητικών και ενόργανου μέρους, καθώς και προβλήματα που δεν έχουν σαφώς καθορισμένες απαντήσεις και μπορούν να επιλυθούν μόνο μέσω επιβλεπόμενης μηχανικής μάθησης, όπως η μουσική ομοιότητα και η ταξινόμηση ανά είδος ή συναίσθημα. [8]

Οι τεχνικές που έχουν χρησιμοποιηθεί στον κλάδο της MIR είναι αρκετές και βασίζονται κατά κύριο λόγο στα δύο παρακάτω βασικά χαρακτηριστικά ενός μουσικού κομματιού:

- Τα μουσικά χαρακτηριστικά (audio features) που προκύπτουν από την επεξεργασία του ηχητικού σήματος ενός κομματιού.
- Τα μεταδεδομένα (tags) που χαρακτηρίζουν και συνοδεύουν ένα κομμάτι.

## 2.2. Αναγνώριση Συναισθήματος από Μουσική - Music Emotion Recognition (MER)

Η αναγνώριση συναισθήματος από τη μουσική, είναι ένα από τα υπο-πεδία της ανάκτησης μουσικής πληροφορίας (MIR), που ασχολείται με την ταξινόμηση της μουσικής και αναπτύσσει μετρικές ομοιότητας βάσει συναισθημάτων, χρησιμοποιώντας τεχνικές επεξεργασίας σήματος & μηχανικής μάθησης. Ως προς τη μεθοδολογία και τις εφαρμογές της η MER, είναι παρόμοια με ένα άλλο εμβληματικό υπο-πεδίο της MIR – την αναγνώριση μουσικών ειδών (η οποία είναι

επίσης αρκετά ασαφής και διφορούμενη) ωστόσο η οντολογία συναισθημάτων στη μουσική είναι ακόμα λιγότερο εδραιωμένη. [9]

Το συναίσθημα μπορεί να εννοιολογηθεί μέσω διαφόρων μοντέλων κατηγορικών (categorical), διαστατικών (dimensional) ή ειδικών για τον τομέα (domain specific).

Το συναίσθημα μπορεί να αναπαρασταθεί με μία μόνο ετικέτα (single label), πολλαπλές ετικέτες (multiple label) η κατανομές πιθανοτήτων (probability distributions) στην οποία μπορεί να εφαρμοστεί μια ετικέτα που κυμαίνεται από μισό δευτερόλεπτο έως ολόκληρο μουσικό κομμάτι.[10]

Ο βασικός σκοπός της MER και γενικότερα της MIR είναι να διευκολύνει την ανάκτηση και τη διαχείριση της μουσικής στην καθημερινή ακρόαση.

Συνήθως, η τεχνική της μηχανικής μάθησης που χρησιμοποιείται για την εκπαίδευση ενός μοντέλου αυτόματης αναγνώρισης περιλαμβάνει τα παρακάτω βήματα: Εξάγεται ένας αριθμός χαρακτηριστικών για την αναπαράσταση του μουσικού σήματος, συλλέγονται οι ετικέτες ή οι τιμές συναισθήματος (ground truth) από ανθρώπους - σχολιαστές και εφαρμόζεται ένας αλγόριθμος μάθησης για την εκμάθηση των σχέσεων μεταξύ των μουσικών χαρακτηριστικών και των ετικετών/τιμών συναισθήματος. Το υπολογιστικό μοντέλο που προκύπτει εφαρμόζεται στη συνέχεια για την πρόβλεψη του συναισθήματος οποιουδήποτε άλλου τραγουδιού δοθεί σαν είσοδος.

### 2.3. Προκλήσεις - Προβλήματα στην MER

Η αναγνώριση συναισθημάτων από το μουσικό σήμα είναι ένα εγγενώς δύσκολο έργο για τους ακόλουθους λόγους. Πρώτον, η αντίληψη των συναισθημάτων είναι υποκειμενική διεργασία και οι άνθρωποι μπορούν να αντιληφθούν διαφορετικά συναισθήματα για το ίδιο τραγούδι. Αυτό το ζήτημα της υποκειμενικότητας καθιστά την αξιολόγηση της απόδοσης ενός συστήματος MER θεμελιωδώς δύσκολη, καθώς είναι δύσκολο να επιτευχθεί μια κοινή συμφωνία για το αποτέλεσμα της ταξινόμησης. Δεύτερον, δεν είναι εύκολο να περιγραφεί το συναίσθημα με καθολικό τρόπο επειδή τα επίθετα που χρησιμοποιούνται για την περιγραφή των συναισθημάτων μπορεί να είναι διφορούμενα και η χρήση επιθέτων για να εκφράσουν το ίδιο συναίσθημα μπορεί να διαφέρει από άνθρωπο σε άνθρωπο. Τρίτον, εξακολουθεί να είναι ανεξήγητο πώς η μουσική προκαλεί συναισθήματα στον άνθρωπο. Δεν έχει ακόμη κατανοηθεί επαρκώς ποιο στοιχείο της μουσικής, αν υπάρχει, δημιουργεί μια συγκεκριμένη συναισθηματική αντίδραση στον ακροατή.

Ευτυχώς, αρκετές θεμελιώδεις ερευνητικές εργασίες σχετικά με τις βασικές διαστάσεις των συναισθημάτων παρέχουν τη βάση για την υπολογιστική μοντελοποίηση των μουσικών διαθέσεων. Στη παρούσα προσέγγιση, υιοθετείται το μοντέλο διάθεσης του Thayer [11] ως βάση ταξινόμησης της διάθεσης και ανίχνευσης της διάθεσης.

Ορισμένα χαρακτηριστικά που αναπτύχθηκαν για την ανάλυση μουσικής στην τρέχουσα βιβλιογραφία, όπως οι συντελεστές συχνότητας (mel-frequency cepstral coefficients - MFCC), η ενέργεια μικρού χρόνου (short-time energy - STE) και ο ρυθμός μηδενικής διασταύρωσης (zero-crossing rate - ZCR) , προτάθηκαν αρχικά για την ανάλυση ομιλίας και ήχου. Ορισμένα χαρακτηριστικά που χρησιμοποιούνται για την ανίχνευση συναισθημάτων ομιλίας, όπως το τονικό ύψος, δεν είναι εφικτά για την αναπαράσταση της μουσικής διάθεσης. Ορισμένα ειδικά για τη μουσική χαρακτηριστικά, όπως η υφή του ηχοχρώματος και το περιεχόμενο του ρυθμού που προτάθηκαν από τους Tzanetakis et al. [12], χρησιμοποιούνται για έναν διαφορετικό στόχο, την ταξινόμηση μουσικών ειδών.

## 2.4. Θεωρία της συναισθηματικής ταξινόμησης

Μέχρι σήμερα, οι ψυχολόγοι έχουν προτείνει διάφορα κλασικά συναισθηματικά ψυχολογικά μοντέλα μελετώντας τη σχέση μεταξύ μουσικής και συναισθήματος. Συνήθως χρησιμοποιούνται τα διακριτά μοντέλα συναισθημάτων και τα συνεχή μοντέλα συναισθημάτων. Η μέθοδος του διακριτού μοντέλου συναισθήματος χωρίζει αρκετές τυπικές διαφορετικές κατηγορίες συναισθημάτων και κάθε κατηγορία μπορεί να περιέχει μια ομάδα παρόμοιων προσδιορισμών των συναισθημάτων. Το μοντέλο συνεχούς διάστασης αναπαριστά την ανθρώπινη συναισθηματική κατάσταση ως σημεία σε ένα δισδιάστατο ή τρισδιάστατο συνεχή χώρο, το οποίο είναι μια συνεχής μέθοδος ψυχολογικής περιγραφής συναισθημάτων. Στο μοντέλο διακριτών συναισθημάτων, χρησιμοποιούνται διαφορετικά επίθετα για την περιγραφή ορισμένων βασικών συναισθημάτων. Μεταξύ αυτών, το πιο διάσημο είναι το μοντέλο συναισθηματικού δακτυλίου του Hevner [13] το οποίο παρουσιάζεται στο σχήμα 5. Το μοντέλο Hevner αποτελείται από 8 ομάδες 67 συναισθηματικών επιθέτων τοποθετημένων σε κύκλους. Μεταξύ αυτών, τα επίθετα σε κάθε μεγάλη ομάδα εκφράζουν παρόμοια συναισθήματα, τα συναισθήματα που εκφράζονται από τις γειτονικές μεγάλες ομάδες είναι πιο κοντά και προοδεύουν σταδιακά, ενώ τα συναισθήματα που εκφράζονται από τις μεγάλες ομάδες σε σχετική θέση είναι ακριβώς το αντίθετο.

Ο Russell [14] πρότεινε ένα μοντέλο που βασίζεται σε δύο διπολικές διαστάσεις αντί πολλών μονοπολικών καταστάσεων. Οι δύο αυτές διαστάσεις ονομάζονται valence (ευχάριστο / δυσάρεστο) και arousal (ενεργητικό / ήρεμο). Ο όρος Valence στην ψυχολογία είναι γνωστός και ως “inducing force,” το οποίο είναι μέτρο για το αν κάποιος νιώθει χαρούμενος ή όχι. Έτσι, κάθε συναίσθημα θα μπορούσε να οριστεί ως κάποιος συνδυασμός των συνιστωσών valence και arousal. (Σχήμα 6.1 & 6.2)

Αργότερα, στα τέλη του 1990, ο Thayer [8] ακολουθώντας το μοντέλο του Russell πρότεινε ένα δισδιάστατο μοντέλο που χρησιμοποιεί μεμονωμένα επίθετα τα οποία συλλογικά σχηματίζουν ένα ολοκληρωμένο μοτίβο διάθεσης. Αυτή η διαστατική προσέγγιση υιοθετεί τη θεωρία ότι η διάθεση απορρέει από δύο παράγοντες: Άγχος (χαρούμενος/αγχώδης) και Ενέργεια (ήρεμος/ενεργητικός), και διαιρεί τη μουσική διάθεση σε τέσσερις ομάδες: Ικανοποίηση (Contentment), Κατάθλιψη (Depression), Πληθωρικότητα (Exuberance) και Αγχώδης/Φρενήρης (Anxious/Frantic).

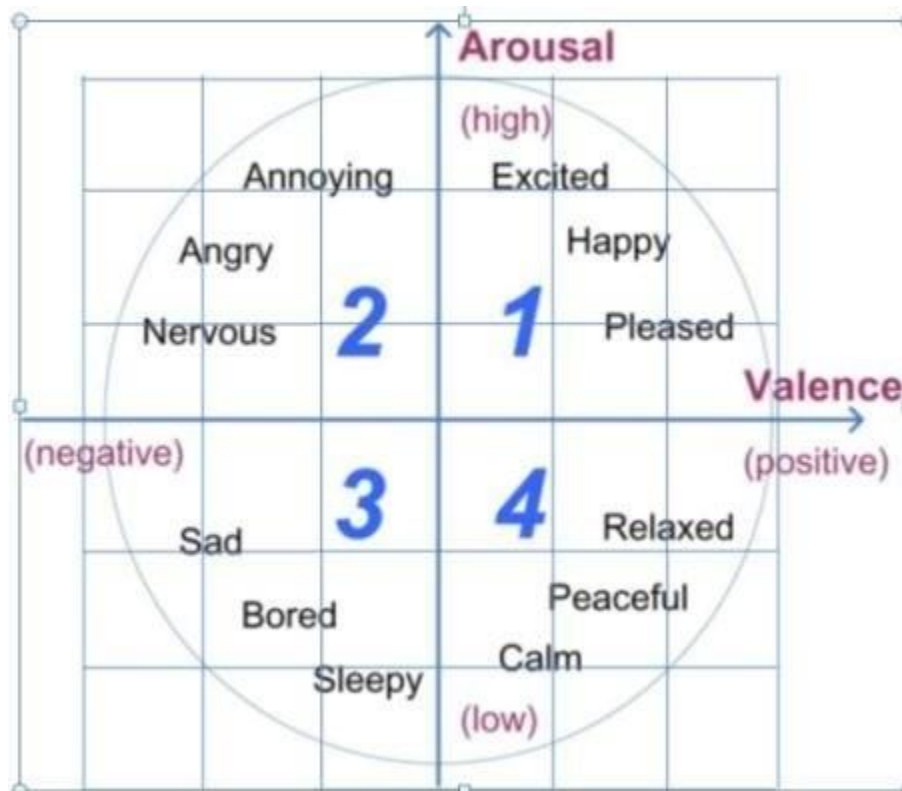


	7	6	5	
	exhilarated	metry	humorous	
	soaring	joyous	playful	
	triumphant	gay	whimsical	
	dramatic	happy	fanciful	
	passionate	cheerful	quaint	4
	sensational	bright	sprightly	lyrical
	agitated		delicate	leisurely
	exciting		light	satisfying
	impetuous		graceful	serene
	restless			tranquil
				quiet
				soothing
8	1	2	3	
vigorous	spiritual	pathetic	dreamy	
robust	lofty	doleful	yielding	
emphatic	awe-inspiring	sad	tender	
martial	dignified	mournful	sentimental	
ponderous	sacred	tragic	longing	
majestic	solemn	melancholy	yearning	
exalting	sober	frustrated	pleading	
	serious	depressing	plaintive	
		gloomy		
		heavy		
		dark		

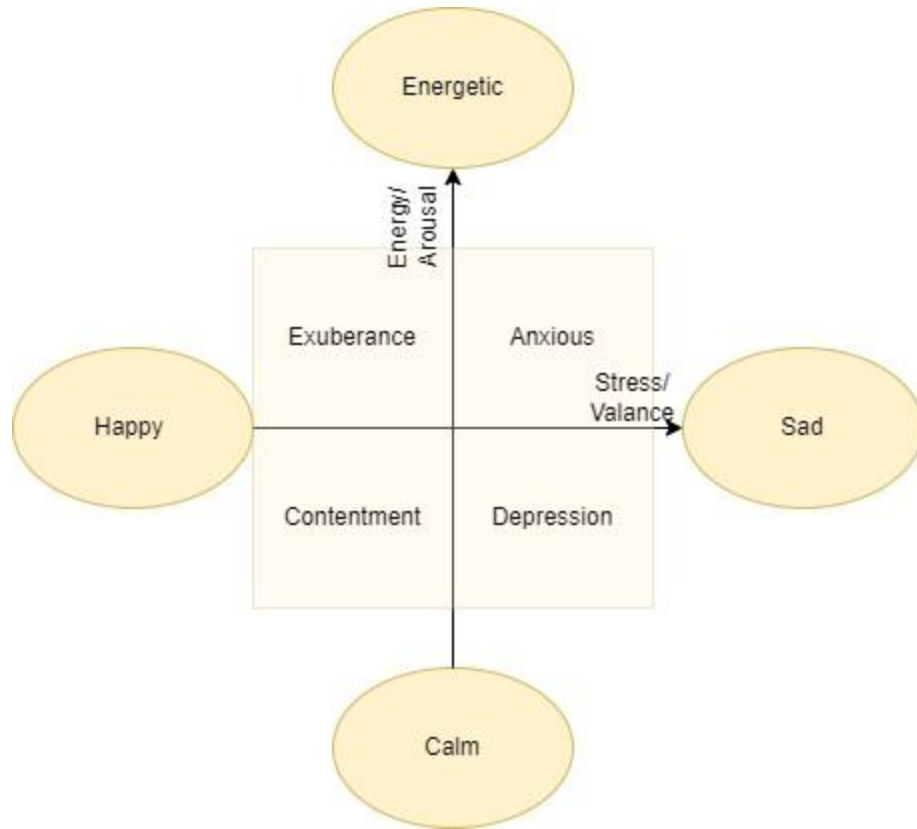
Σχήμα 5

Το επίπεδο συναισθημάτων του Thayer υιοθετείται στη παρούσα εργασία για την αποφυγή της ασάφειας των επιθέτων. Καθορίζει τις κλάσεις συναισθημάτων διαστασιακά ως προς τη διέγερση (πόσο διεγερτική/ηρεμιστική) και την αξία (πόσο θετική/αρνητική). Για παράδειγμα, οι κλάσεις συναισθημάτων μπορούν να χωριστούν στα τέσσερα τεταρτημόρια στο επίπεδο συναισθημάτων, όπως φαίνεται στο Σχήμα 7.

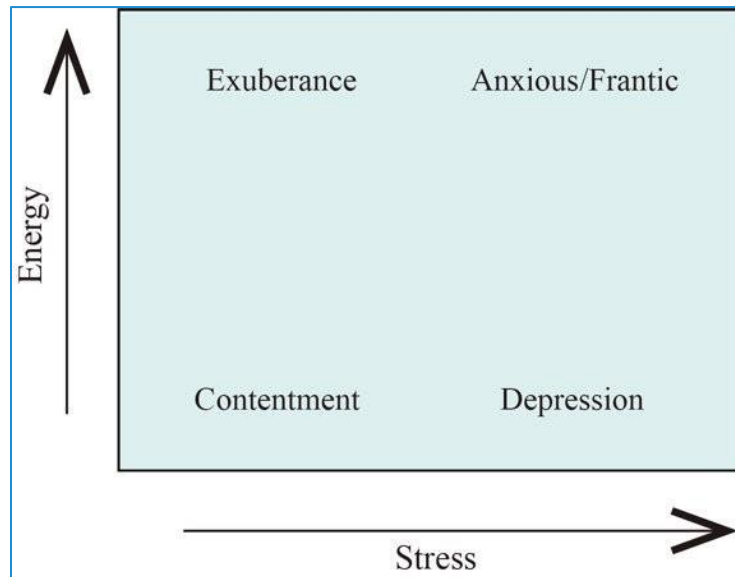
Επίσης, αυτές οι τέσσερις συστάδες είναι διακριτές και σαφείς, και η δισδιάστατη δομή παρέχει χρήσιμες ενδείξεις για τη μοντελοποίηση στον υπολογιστή. Έτσι, στη παρούσα εργασία αντίχενωσης διάθεσης εφαρμόζεται αυτό το μοντέλο συναισθημάτων.







Σχήμα 6.1 & 6.2



Σχήμα 7

## 2.5. Λογιστική Παλινδρόμηση

Η λογιστική παλινδρόμηση (ή logit regression, ή logit model) εισήχθη για πρώτη φορά στη στατιστική βιβλιογραφία από τον D. Cox (1958). Πρόκειται για ένα μοντέλο παλινδρόμησης που αφορά κατηγορικές εξαρτημένες μεταβλητές και ποσοτικές ή κατηγορικές ανεξάρτητες μεταβλητές. Ο προσδιορισμός της δομής των ανεξάρτητων μεταβλητών σε ένα δυαδικό μοντέλο λογιστικής παλινδρόμησης περιλαμβάνει τον προσδιορισμό του αριθμού των επιπέδων της κατηγορικής μεταβλητής, δηλαδή τον προσδιορισμό του αριθμού των δεικτών (δείκτες ή ψευδοδείκτες). Δείκτης είναι κάθε δυαδική μεταβλητή με τιμές 1 και 0.

Η λογιστική παλινδρόμηση (LR) είναι ο πιο διάσημος αλγόριθμος μηχανικής μάθησης μετά τη γραμμική παλινδρόμηση. Η γραμμική παλινδρόμηση και η λογιστική παλινδρόμηση είναι σχεδόν παρόμοιες. Όμως, η μεγαλύτερη διαφορά έγκειται στο για ποιο σκοπό χρησιμοποιούνται.

Στόχος της Λογιστικής Παλινδρόμησης είναι η ανάπτυξη ενός μοντέλου που χρησιμοποιεί τόσο ποσοτικούς όσο και ποιοτικούς ανεξάρτητους παράγοντες για την πρόβλεψη των τιμών της κατηγορικής εξαρτημένης μεταβλητής που μελετάται. Η Λογιστική Παλινδρόμηση έχει ευρύ φάσμα εφαρμογών στην επιστήμη λόγω της φύσης της. Η λογιστική παλινδρόμηση χρησιμοποιείται στην οικονομία και στην επιχειρηματικότητα για να προβλέψει τις μελλοντικές τιμές και τα κέρδη μιας επιχείρησης. Παρακάτω παραθέτουμε μερικά παραδείγματα χρήσης της λογιστικής παλινδρόμησης:

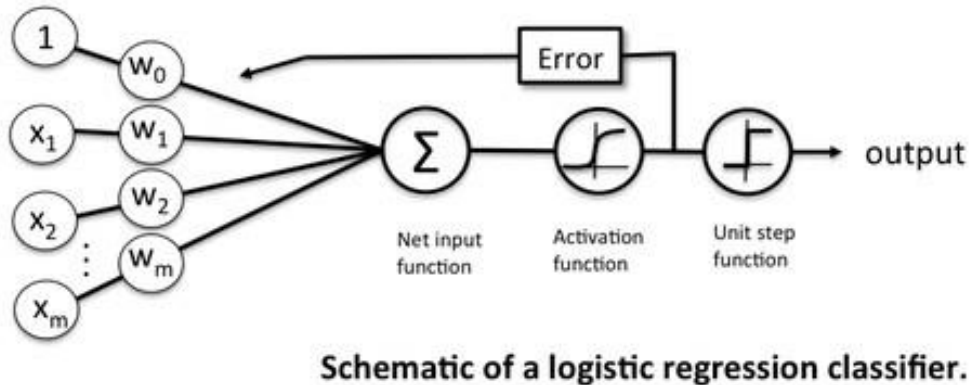
1. Πρόβλεψη της τιμής μιας μετοχής: Η λογιστική παλινδρόμηση μπορεί να χρησιμοποιηθεί για να προβλέψει τη μελλοντική τιμή μιας μετοχής βάσει ιστορικών δεδομένων.
2. Πρόβλεψη των πωλήσεων: Η λογιστική παλινδρόμηση μπορεί να χρησιμοποιηθεί για να προβλέψει το μέγεθος των μελλοντικών πωλήσεων μιας επιχείρησης βάσει παρελθοντικών δεδομένων.

Υπάρχουν τρεις τύποι, ανάλογα με την ιδιαίτερη φύση της εξαρτημένης κατηγορικής μεταβλητής:

- Συγκεκριμένα, στη δυαδική λογιστική παλινδρόμηση, η εξαρτημένη κατηγορική μεταβλητή αποτελείται από δύο κατηγορίες, όπως επιτυχία/αποτυχία, ναι/όχι και απύσα/παρούσα.
- Στην πολυωνυμική λογιστική παλινδρόμηση η εξαρτημένη μεταβλητή κωδικοποιείται σε τρεις ή περισσότερες κατηγορίες. Για παράδειγμα, όταν η εξαρτημένη μεταβλητή είναι η ταξινόμηση του χρώματος των αντικειμένων ως κόκκινο, πράσινο ή κίτρινο κ.λπ. Αυτό σημαίνει ότι οι κατηγορίες, ωστόσο, δεν έχουν φυσική διαβάθμιση, επειδή κάθε μία ξεκινά από ένα ισοδύναμο σημείο του φάσματος.
- Στην Ordinal Regression, η εξαρτημένη μεταβλητή είναι η απόσταση ανάμεσα σε δύο ή περισσότερες κατηγορίες που χαρακτηρίζονται από μια έννοια ανισότητας, για παράδειγμα μια ερώτηση στην οποία οι επιλογές είναι "συμφωνώ/διαφωνώ" με βαθμολογία "καθόλου", "λίγο", "μέτρια", "αρκετά" και "πολύ".

### 2.5.1. Logistic regression as NN

Μπορούμε να θεωρήσουμε τη λογιστική παλινδρόμηση ως ένα νευρωνικό δίκτυο ενός επιπέδου.



Σχήμα 8

### 2.5.2. Μαθηματικά πίσω από τη Λογιστική Παλινδρόμηση

Η παλινδρόμηση χρησιμοποιεί μια πιο σύνθετη συνάρτηση κόστους, αυτή η συνάρτηση κόστους μπορεί ονομάζεται "σιγμοειδής συνάρτηση" ή επίσης γνωστή ως "λογιστική συνάρτηση" αντί για μια γραμμική συνάρτηση.

$$F(x) = \frac{1}{1 + e^{-x}} = \frac{e^x}{e^x + 1}$$

Αυτή η συνάρτηση βοηθά το μοντέλο Λογιστικής Παλινδρόμησης να συμπτύξει τις τιμές από  $(-\infty, \infty)$  σε  $(0,1)$ . Η λογιστική παλινδρόμηση χρησιμοποιείται κυρίως για προβλήματα δυαδικής ταξινόμησης, ωστόσο, μπορεί να χρησιμοποιηθεί για ταξινόμηση πολλαπλών κατηγοριών.

### 2.6. Αλγόριθμος K-πλησιέστερων γειτόνων

Ο αλγόριθμος k-πλησιέστερων γειτόνων (ή k-NN) είναι ένας απλός αλγόριθμος μηχανικής μάθησης για την ταξινόμηση και την παλινδρόμηση των δεδομένων. Ο αλγόριθμος χρησιμοποιείται συχνά στο πλαίσιο της ταξινόμησης εικόνων, αναγνώρισης προτύπων και αναγνώρισης φωνής.

Στον αλγόριθμο k-NN, το k σημαίνει τον αριθμό των πλησιέστερων γειτόνων που θα χρησιμοποιηθούν για να κατηγοριοποιήσουν ένα νέο δείγμα. Οι "γείτονες" αναφέρονται στα k δείγματα στο σετ εκπαίδευσης που είναι πιο κοντά (με βάση την απόσταση) στο δοθέν δείγμα.

Κατά τη διάρκεια της ταξινόμησης, ο αλγόριθμος k-NN υπολογίζει την απόσταση μεταξύ του νέου δείγματος και κάθε δείγματος στο σετ εκπαίδευσης και επιλέγει τα k δείγματα με τη μικρότερη απόσταση. Στη συνέχεια, χρησιμοποιεί τις ετικέτες των k πλησιέστερων γειτόνων για να προβλέψει την ετικέτα του νέου δείγματος.

Μια κλασική εφαρμογή της μεθόδου του πλησιέστερου γείτονα είναι η πρόβλεψη της συμπεριφοράς των χρηστών σε εφαρμογές Τεχνητής Νοημοσύνης, όπως τα συστήματα συστάσεων.

## 2.7. Μετρικές

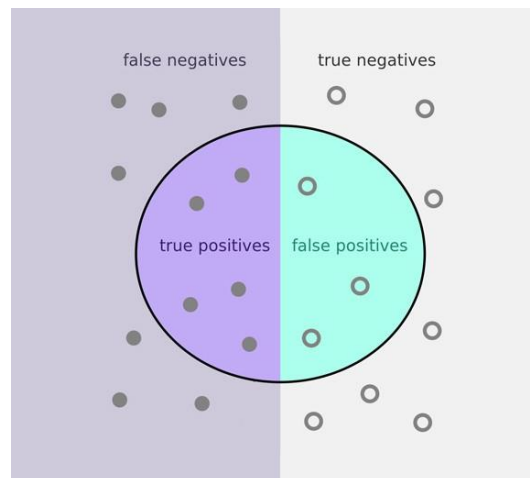
### 2.7.1. Recall

Στην τεχνητή νοημοσύνη, η ανάκληση (recall) είναι μία μετρική αξιολόγησης που χρησιμοποιείται για να μετρήσει πόσο αποτελεσματικά ένα σύστημα αναγνώρισης προτύπων μπορεί να εντοπίσει όλα τα σωστά αποτελέσματα θετικών κλάσεων.

Συγκεκριμένα, η ανάκληση υπολογίζει το ποσοστό των θετικών δειγμάτων που εντοπίστηκαν σωστά σε σχέση με το συνολικό αριθμό των θετικών δειγμάτων. Μαθηματικά, η ανάκληση υπολογίζεται ως εξής:

$$\text{Recall} = \frac{\text{True Positive}}{\text{True Positive} + \text{False Negative}}$$

όπου TP είναι τα πραγματικά θετικά δείγματα που εντοπίστηκαν σωστά, και FN είναι τα πραγματικά θετικά δείγματα που δεν εντοπίστηκαν.



Σχήμα 9

### 2.7.2. Confusion Matrix

Η Confusion Matrix (Πίνακας Σύγχυσης) είναι ένα εργαλείο που χρησιμοποιείται στη μηχανική μάθηση και την τεχνητή νοημοσύνη για να αξιολογήσει την απόδοση ενός μοντέλου πρόβλεψης. Συγκεκριμένα, ο πίνακας σύγχυσης επιτρέπει στους προγραμματιστές και τους ερευνητές να αναλύσουν πόσο αποτελεσματικά λειτουργεί ένα μοντέλο στον εντοπισμό των διαφορετικών κλάσεων.

Ο πίνακας σύγχυσης είναι ένας τετραγωνικός πίνακας που περιέχει τα αποτελέσματα των προβλέψεων ενός μοντέλου και τα πραγματικά αποτελέσματα του συνόλου δεδομένων. Οι διαγώνιες κατηγορίες του πίνακα αντιπροσωπεύουν τις προβλέψεις που είναι σωστές, ενώ οι μη διαγώνιες κατηγορίες αντιπροσωπεύουν τις προβλέψεις που είναι λανθασμένες.

Αναλυτικότερα, ο πίνακας σύγχυσης αναφέρεται σε τέσσερις κατηγορίες:

- True Positives (TP): τα θετικά δείγματα που εντοπίστηκαν σωστά.
- False Positives (FP): τα αρνητικά δείγματα που εντοπίστηκαν λανθασμένα.
- False Negatives (FN): τα θετικά

		Actual Values	
		Positive (1)	Negative (0)
Predicted Values	Positive (1)	TP	FP
	Negative (0)	FN	TN

Σχήμα 10

### 2.7.3. Accuracy score

Ο πιο διαισθητικός τρόπος αξιολόγησης της απόδοσης οποιουδήποτε αλγορίθμου ταξινόμησης είναι ο υπολογισμός του ποσοστού των σωστών προβλέψεών του. Και αυτή ακριβώς είναι η λογική πίσω από το σκορ ακρίβειας.

Για να ορίσουμε τον όρο, στη Μηχανική Μάθηση, το “σκορ” ακρίβειας (ή απλώς ακρίβεια - accuracy) είναι μια μετρική ταξινόμησης που χαρακτηρίζει ένα κλάσμα των προβλέψεων που ένα μοντέλο έκανε σωστά. Η μετρική είναι πολύ διαδεδομένη καθώς είναι εύκολο να υπολογιστεί και να ερμηνευτεί. Επίσης, μετράει την απόδοση του μοντέλου με μία μόνο τιμή.

#### 2.7.4. F1 score

Το F1-score συνδυάζει την ακρίβεια και την ανάκληση ενός ταξινομητή σε μια ενιαία μετρική, λαμβάνοντας τον αρμονικό μέσο όρο τους. Χρησιμοποιείται κυρίως για τη σύγκριση της απόδοσης δύο ταξινομητών. Ας υποθέσουμε ότι ο ταξινομητής A έχει υψηλότερη ανάκληση και ο ταξινομητής B έχει υψηλότερη ακρίβεια. Σε αυτή την περίπτωση, τα αποτελέσματα F1 και για τους δύο ταξινομητές μπορούν να χρησιμοποιηθούν για να προσδιοριστεί ποιος από τους δύο παράγει καλύτερα αποτελέσματα.

Το F1-score ενός μοντέλου ταξινόμησης υπολογίζεται ως εξής:

$$\frac{2(P * R)}{P + R}$$

όπου P= η ακρίβεια (precision) & R = η ανάκληση του μοντέλου ταξινόμησης (recall)

## 2.8. State-of-the-art και μειονεκτήματα

Τις τελευταίες δεκαετίες έχουν σημειωθεί σημαντικές πρόοδοι στον τομέα της αναγνώρισης μουσικών συναισθημάτων (MER).

Πιο κάτω, θα εξετάσουμε το τρέχον επίπεδο της τεχνολογίας της αναγνώρισης μουσικών συναισθημάτων και θα διερευνήσουμε μερικές από τις πιο υποσχόμενες τεχνικές που χρησιμοποιούνται για την ανάλυση των συναισθηματικών αντιδράσεων στη μουσική.

### 2.8.1. State-of-the-art Τεχνικές

Μία από τις πιο συχνά χρησιμοποιούμενες μεθόδους για την αναγνώριση μουσικών συναισθημάτων είναι η εξαγωγή χαρακτηριστικών (feature extraction). Η εξαγωγή χαρακτηριστικών περιλαμβάνει την επιλογή σχετικών χαρακτηριστικών ενός μουσικού κομματιού, όπως το τέμπο, το ύψος, το ηχόχρωμα και ο ρυθμός. Αυτά τα χαρακτηριστικά στη συνέχεια χρησιμοποιούνται για τη δημιουργία ενός συνόλου αριθμητικών τιμών, γνωστών ως

διανύσματα χαρακτηριστικών (feature vectors), τα οποία μπορούν να επεξεργαστούν με αλγορίθμους μηχανικής μάθησης.[15]

Μια άλλη πολλά υποσχόμενη τεχνική που χρησιμοποιείται στην αναγνώριση μουσικών συναισθημάτων είναι η βαθιά μάθηση. Η βαθιά μάθηση περιλαμβάνει τη χρήση νευρωνικών δικτύων για την ανάλυση σύνθετων μοτίβων και σχέσεων μέσα στα δεδομένα. Αυτή η τεχνική έχει αποδειχθεί ότι επιτυγχάνει κορυφαίες επιδόσεις σε διάφορες περιπτώσεις έρευνας, συμπεριλαμβανομένης της αναγνώρισης μουσικών συναισθημάτων.[16]

Επιπλέον, πρόσφατες μελέτες έχουν επίσης επικεντρωθεί σε πολυτροπικές (multi-modal) προσεγγίσεις για την αναγνώριση μουσικών συναισθημάτων, οι οποίες συνδυάζουν δεδομένα από πολλαπλές πηγές, όπως φυσικά σήματα και εκφράσεις του προσώπου, για τη βελτίωση της ακρίβειας.[17]

Εκτός από αυτές τις τεχνικές, υπάρχει επίσης αυξανόμενο ενδιαφέρον για την ανάπτυξη πιο εξατομικευμένων προσεγγίσεων στην αναγνώριση μουσικών συναισθημάτων. Μια τέτοια προσέγγιση είναι η συναισθηματική υπολογιστική, η οποία περιλαμβάνει τη χρήση αλγορίθμων μηχανικής μάθησης για τη μοντελοποίηση των ατομικών διαφορών στις συναισθηματικές αντιδράσεις στη μουσική.[18]

## 2.8.2. Δεδομένα

Υπάρχουν διάφορα δημόσια διαθέσιμα σύνολα δεδομένων που χρησιμοποιούνται συνήθως για την ανάλυση μουσικού συναισθήματος, όπως:

- **Million Song Dataset:** Είναι ένα μεγάλο σύνολο δεδομένων ήχου και μετα-δεδομένων για ένα εκατομμύριο τραγούδια, που συλλέχθηκαν από την ηλεκτρονική μουσική υπηρεσία Echo Nest. Το σύνολο δεδομένων περιλαμβάνει πληροφορίες σχετικά με τη μουσική διάθεση, τους στίχους και το είδος της μουσικής, καθιστώντας το χρήσιμο εργαλείο για την ανάλυση μουσικού συναισθήματος.
- **ISMIR:** Η International Society for Music Information Retrieval (ISMIR) παρέχει μια πληθώρα συνόλων δεδομένων για διάφορες εργασίες ανάκτησης μουσικών πληροφοριών, συμπεριλαμβανομένης της ανάλυσης συναισθήματος. Αυτά τα σύνολα δεδομένων περιλαμβάνουν δεδομένα ήχου και κειμένου, όπως στίχους και πληροφορίες καλλιτέχνη, και χρησιμοποιούνται ευρέως στην ερευνητική κοινότητα.
- **EmoMusic:** Πρόκειται για ένα σύνολο δεδομένων από μουσικά κλιπ και αντίστοιχους σχολιασμούς, που συλλέγονται από μια πλατφόρμα πλήθους. Το σύνολο δεδομένων περιλαμβάνει πληροφορίες σχετικά με το είδος της μουσικής, τον ρυθμό και τους στίχους, καθώς και σχολιασμούς συναισθήματος για κάθε κλιπ.
- **Sentiment140:** Πρόκειται για ένα σύνολο δεδομένων με 1,6 εκατομμύρια tweets, τα οποία συλλέγονται από το Twitter και σχολιάζονται με ετικέτες συναισθήματος (θετικά, αρνητικά ή ουδέτερα). Τα σύνολα δεδομένων μπορούν να χρησιμοποιηθούν για την εκπαίδευση μοντέλων μηχανικής μάθησης για την ανάλυση μουσικού συναισθήματος, καθώς τα tweets συχνά περιέχουν μουσικές εκφράσεις και απόψεις για τη μουσική.[19]

Αυτά τα σύνολα δεδομένων αποτελούν πολύτιμη πηγή για την εκπαίδευση και την αξιολόγηση μοντέλων μηχανικής μάθησης και επιτρέπουν στους ερευνητές να συγκρίνουν τις επιδόσεις διαφορετικών μοντέλων και τεχνικών. Ωστόσο μπορεί να έχουν ορισμένα μειονεκτήματα που μπορούν να επηρεάσουν την αξιοπιστία και την απόδοση των μοντέλων ανάλυσης συναισθήματος. [20]

### 2.8.3. Μειονεκτήματα

Παρά τις σημαντικές προόδους στην αναγνώριση μουσικών συναισθημάτων, υπάρχουν ακόμη κάποιες προκλήσεις που πρέπει να αντιμετωπιστούν.

1. Περιορισμένο μέγεθος και παλαιότητα: Πολλά σύνολα δεδομένων για την ανάλυση συναισθήματος μουσικής είναι περιορισμένου μεγέθους και σπάνια λαμβάνουν κάποιου είδους ανανέωση στο πλήθος ή την ποιότητα, γεγονός που μπορεί να δυσχεράνει την εκπαίδευση μοντέλων με μεγάλες ποσότητες δεδομένων και να οδηγήσει σε υπερπροσαρμογή (overfitting), όπου τα μοντέλα αποδίδουν καλά στα δεδομένα εκπαίδευσης, αλλά όχι σε αθέατα δεδομένα.
2. Μεροληψία: Τα δημόσια σύνολα δεδομένων για την ανάλυση μουσικού συναισθήματος μπορεί να είναι μεροληπτικά όσον αφορά τους τύπους μουσικής και συναισθημάτων που περιλαμβάνονται. Για παράδειγμα, τα σύνολα δεδομένων μπορεί να περιλαμβάνουν μόνο δημοφιλή μουσική, γεγονός που μπορεί να περιορίσει την ικανότητα των μοντέλων να αναγνωρίζουν συναισθήματα σε άλλους τύπους μουσικής.
3. Ποιότητα σχολιασμού: Η ποιότητα των σχολίων στα δημόσια σύνολα δεδομένων για την ανάλυση συναισθήματος μουσικής μπορεί να διαφέρει σημαντικά. Για παράδειγμα, οι σχολιασμοί μπορεί να είναι υποκειμενικοί ή ασυνεπείς, και μπορεί να έχει αρνητικό αντίκτυπο στην απόδοση των μοντέλων ανάλυσης συναισθήματος.

Για τους παραπάνω λόγους, στη παρούσα εργασία και όπως παρουσιάζεται παρακάτω, όσον αφορά τα δεδομένα γίνεται χρήση των εργαλείων που παρέχονται από την πλατφόρμα αναπαραγωγής μουσικής Spotify. Η τεχνική που υιοθετείται αποτελεί μία ανάμιξη των state-of-the-art μεθόδων που αναφέρονται πιο πάνω.

## Κεφάλαιο 3: Πειραματική διαδικασία

### 3.1. Συλλογή των δεδομένων



Το Spotify χρησιμοποιεί μια σειρά από διαφορετικά χαρακτηριστικά για την ταξινόμηση των κομματιών.

**Acousticness:** Το Acousticness είναι ένα από τα χαρακτηριστικά που παρέχει το Spotify API για κάθε τραγούδι. Αναφέρεται στο πόσο "ακουστικό" είναι ένα τραγούδι, δηλαδή πόσο κοντά είναι στον ήχο ενός ακουστικού μουσικού οργάνου, όπως η κιθάρα ή το πιάνο. Το acousticness είναι μια μετρική που κυμαίνεται από 0 έως 1, με υψηλότερες τιμές που υποδηλώνουν ότι ένα τραγούδι έχει περισσότερα ακουστικά στοιχεία και χαμηλότερες τιμές που υποδηλώνουν ότι ένα τραγούδι έχει λιγότερα ακουστικά στοιχεία και είναι πιο ηλεκτρονικό.

**Danceability:** Είναι μια μετρική που χρησιμοποιείται για να περιγράψει το πόσο κατάλληλο είναι ένα τραγούδι για χορό. Αυτή η μετρική κυμαίνεται από 0.0 έως 1.0 και υπολογίζεται βάσει διάφορων χαρακτηριστικών του τραγουδιού, όπως ο ρυθμός, η συνοχή, η σταθερότητα και η συνολική ενέργεια του τραγουδιού. Αν ένα τραγούδι έχει μια υψηλή τιμή Danceability, τότε είναι πιο κατάλληλο για χορό και προκαλεί πιο ενθουσιώδη και κινητικές αντιδράσεις από τους ακροατές.

**Energy:** Αναφέρεται στη δύναμη και τον ρυθμό ενός τραγουδιού, καθώς και στην ενέργεια που μεταφέρεται στον ακροατή. Η τιμή του Energy κυμαίνεται από 0 έως 1, με υψηλότερες τιμές να υποδηλώνουν ένα πιο ενεργητικό και ρυθμικό τραγούδι, ενώ χαμηλότερες τιμές υποδηλώνουν ένα πιο ήρεμο και σταθερό τραγούδι.

**Instrumentalness:** Είναι μια μετρική που χρησιμοποιείται στο Spotify API για να περιγράψει το επίπεδο του ήχου ενός κομματιού που δεν περιλαμβάνει φωνητικά. Αναφέρεται στο ποσοστό της μουσικής που δεν περιλαμβάνει φωνητικά και καθορίζει το επίπεδο της μουσικής που είναι αποκλειστικά ορχηστρική. Η τιμή του instrumentalness είναι ένας αριθμός από το 0 έως το 1, με το 0 να σημαίνει πλήρως φωνητικό κομμάτι και το 1 να σημαίνει πλήρως ορχηστρικό.

**Liveness:** Είναι ένα χαρακτηριστικό του Spotify API που παρέχει πληροφορίες σχετικά με τον βαθμό ζωντανίας μιας ηχογράφησης μουσικής. Αυτό σημαίνει ότι το Spotify αναλύει την ηχογράφηση και καθορίζει πόσο πιθανό είναι να πρόκειται για μια εκτέλεση κατά τη διάρκεια μιας ζωντανής συναυλίας, αντί για μια στούντιο εκτέλεση. Το Liveness είναι ένα από τα πολλά χαρακτηριστικά που το Spotify παρέχει για να βοηθήσει τους ακροατές να ανακαλύψουν μουσική που ταιριάζει στα γούστα τους..

**Loudness:** Στο Spotify API, η τιμή του Loudness δίνεται σε μονάδες dB (decibels), και χρησιμοποιείται για να βοηθήσει τους ακροατές να διαχωρίσουν τα τραγούδια που είναι πιο δυνατά από τα ήσυχα. Επιπλέον, η τιμή του Loudness μπορεί να χρησιμοποιηθεί από τους developers για να δημιουργήσουν διάφορα εργαλεία επεξεργασίας ήχου. Οι τιμές κυμαίνονται συνήθως μεταξύ -60 και 0 db.

**Speechiness:** Το Speechiness ανιχνεύει την παρουσία προφορικών λέξεων σε ένα κομμάτι. Η μετρική αυτή κυμαίνεται από 0 έως 1, με υψηλότερες τιμές που υποδηλώνουν ότι το τραγούδι έχει περισσότερη φωνητική απόδοση και ελάχιστη επαναληψιμότητα στους στίχους του. Τιμές μεταξύ 0,33 και 0,66 περιγράφουν κομμάτια που μπορεί να περιέχουν τόσο μουσική όσο και ομιλία, είτε τμηματικά είτε σε στρώσεις, συμπεριλαμβανομένων περιπτώσεων όπως η μουσική

ραπ. Τιμές κάτω από 0,33 αντιπροσωπεύουν πιθανότατα μουσική και άλλα κομμάτια που δεν μοιάζουν με ομιλία.

Valence: Χρησιμοποιείται για να αξιολογήσει τη θετική ή αρνητική διάθεση μιας μουσικής κομματιού. Το Valence κυμαίνεται από το -1 (πολύ αρνητική διάθεση) έως το 1 (πολύ θετική διάθεση), ενώ τιμές κοντά στο 0 υποδηλώνουν αδιάφορη διάθεση. Αυτή η μετρική χρησιμοποιείται από το Spotify για να προτείνει μουσική με βάση την προτίμηση των ακροατών και την αίσθησή τους.

Tempo: Χρησιμοποιείται στο Spotify API για να περιγράψει το ρυθμό των τραγουδιών. Το Tempo μετρείται σε χτυπήματα ανά λεπτό (BPM) και χρησιμοποιείται στο Spotify για να προτείνει τραγούδια στους χρήστες με βάση τον ρυθμό της μουσικής που τους αρέσει. Η μετρική αυτή είναι επίσης χρήσιμη για τη δημιουργία playlist με τραγούδια που έχουν παρόμοιο ρυθμό.

[Πηγή: web site Spotify API]

Για να εξάγουμε τα δεδομένα έπρεπε να χρησιμοποιήσουμε μια σειρά συναρτήσεων χρησιμοποιώντας τη βιβλιοθήκη Spotify.

### 3.2. Ετικετοποίηση των δεδομένων

Τα προβλήματα ταξινόμησης χρησιμοποιούν δεδομένα με ετικέτες, οπότε δημιουργήσαμε αυτές τις ετικέτες. Αποφασίστηκε να δημιουργηθούν 4 κατηγορίες για την δεικτοδότηση των κομματιών, οι κατηγορίες αυτές είναι Ενεργητικό (Energetic), Ήρεμό (Calm), Χαρούμενό (Happy) και Θλιμμένό (Sad). Επιλέχτηκαν αυτές οι κατηγορίες με βάση το μοντέλο Thayer που αναλύεται παραπάνω στο κεφάλαιο 2.

Στη συνέχεια έγινε αναζήτηση στο Spotify για κάποιες λίστες αναπαραγωγής (playlists) με μουσικά κομμάτια που έχουν χαρακτηριστεί ως ένα από τα 4 βασικά συναισθήματα - ετικέτες (labels) και, τέλος, όλα αυτά τα κομμάτια ενσωματώθηκαν στο ίδιο data frame με ετικέτες ανά διάθεση. Οι λίστες αυτές έχουν δημιουργηθεί από ανθρώπους - ακροατές και για αυτό το λόγο θεωρούμε πως αποτελούν μία ικανοποιητική βάση αληθείας (ground truth).

Τα δεδομένα έχουν 18 στήλες, αλλά για λόγους μείωσης της πληροφορίας θα χρησιμοποιηθούν τα 10 χαρακτηριστικά (στήλες) που έχουν μεγαλύτερη επιρροή στην ταξινόμηση των κομματιών.

### 3.3. Train/Test Split Method

Η μέθοδος διαχωρισμού εκπαίδευσης/δοκιμής (Train/Test Split Method) χωρίζει το σύνολο δεδομένων σε δύο τμήματα: σύνολο εκπαίδευσης και σύνολο ελέγχου. Το σύνολο εκπαίδευσης (training set) χρησιμοποιείται για την εκπαίδευση του μοντέλου. Μπορεί επίσης να μετρήσει την ακρίβεια του μοντέλου στο σύνολο εκπαίδευσης. Το σύνολο ελέγχου ή δοκιμής (testing set)

χρησιμοποιείται για τη δοκιμή του μοντέλου και την αξιολόγηση της ακρίβειας μετά την εκπαίδευση. Τα δείγματα δεδομένων στο σύνολο δοκιμής δεν εμφανίζονται ποτέ στο μοντέλο κατά τη διάρκεια της εκπαίδευσης. Η ακρίβεια στο σύνολο δοκιμής παρέχει μια καλύτερη ένδειξη για το πώς τα μοντέλα θα αποδώσουν σε νέα δεδομένα.

## 3.4. Τεχνολογίες που χρησιμοποιήθηκαν

### 3.4.1. Python 3

Η Python είναι μια γλώσσα προγραμματισμού υψηλού επιπέδου και γενικού σκοπού, με έμφαση στην ανάπτυξη εύκολα αναγνώσιμου κώδικα. Είναι γνωστή για το ότι διευκολύνει τη διαδικασία της ανάπτυξης έργων λογισμικού (π.χ. Collaboration, Scaling etc.) και της συντήρησής τους (όπως Debugging). Υποστηρίζει διάφορες τεχνικές ανάπτυξης λογισμικού, Διαδικαστικό προγραμματισμό (Procedural), Αντικειμενοστραφή προγραμματισμό (Object-Oriented) και Συναρτησιακό προγραμματισμό (Functional) . Διαθέτει ευρεία Standard Library η οποία παρέχει Δομές Δεδομένων, Διεπαφές Λειτουργικού Συστήματος και άλλα. Παρακάτω περιγράφονται συνοπτικά μερικά από τα modules της Standard Library, που χρησιμοποιήθηκαν για τις ανάγκες της παρούσας εργασίας. Η ανάπτυξη της εργασίας έγινε με χρήση της Python 3.

### 3.4.2. Pandas

Το pandas είναι μια βιβλιοθήκη της Python για τη διαχείριση και ανάλυση δεδομένων. Προσφέρει δομές δεδομένων όπως τα DataFrames και τα Series για την αποθήκευση, την επεξεργασία και την ανάλυση δεδομένων σε Python. Το pandas είναι επίσης πολύ αποδοτικό στην επεξεργασία μεγάλων συνόλων δεδομένων και προσφέρει δυνατότητες για την αναλυτική εξερεύνηση και οπτικοποίηση δεδομένων.

### 3.4.3. Plotly

Το Plotly είναι μια βιβλιοθήκη για την ανάπτυξη διαδραστικών γραφημάτων σε Python. Προσφέρει εύκολα τη δυνατότητα δημιουργίας όλων των ειδών των γραφημάτων, από απλά διαγράμματα έως πιο πολύπλοκα με πολλαπλά στρώματα και αλληλεπιδράσεις.

### 3.4.4. Seaborn

Το Seaborn είναι μια βιβλιοθήκη για την ανάλυση και οπτικοποίηση δεδομένων στην Python, βασισμένη στη βιβλιοθήκη matplotlib με πρόσθετη λειτουργικότητα για την οπτικοποίηση δεδομένων.

### 3.4.5. Scikit-Learn

Η Scikit-Learn είναι μία από τις πιο γνωστές ML βιβλιοθήκες σε Python. Παρέχει διάφορους αλγορίθμους κατηγοριοποίησης, παλινδρόμησης κ.α. όπως για παράδειγμα SVM, Gradient Boosting, k-Means και άλλους. Επίσης, πολύ χρήσιμοι είναι οι αλγόριθμοι προετοιμασίας των δεδομένων όπως ο MinMaxScaler. Επιπλέον, η Scikit-Learn είναι σχεδιασμένη έτσι ώστε να μπορεί να επεξεργαστεί Δομές Δεδομένων των βιβλιοθηκών Numpy - Pandas, πράγμα που διευκολύνει τη διαχείριση των δεδομένων.

### 3.4.6. Visual Studio Code

Ένας επεξεργαστής κώδικα που προσφέρει δυνατότητες ανάπτυξης, συμπεριλαμβανομένων της αποσφαλμάτωσης (debugging), της εκτέλεσης εργασιών (task running) και του ελέγχου έκδοσης (version control), ονομάζεται Visual Studio Code. Στόχος του είναι να παρέχει όλους τους πόρους που χρειάζεται ένας προγραμματιστής.

### 3.4.7. Jupyter Notebook

Το Jupyter Notebook επιτρέπει στους χρήστες να συγκεντρώσουν όλες τις πτυχές μιας εργασίας Python σε ένα μέρος, διευκολύνοντας έτσι την παρουσίαση ολόκληρης της διαδικασίας στο κοινό στο οποίο απευθύνεται.

### 3.4.9. Spotify Web API

Βασισμένα σε απλές αρχές REST, το Spotify Web API επιστρέφουν μετα-δεδομένα JSON για μουσικούς καλλιτέχνες, άλμπουμ και κομμάτια, απευθείας από τη βάση δεδομένων του Spotify.

### 3.4.10. Spotipy

Το Spotipy είναι μια βιβλιοθήκη Python για το Spotify Web API που δίνει πλήρη πρόσβαση σε όλα τα μουσικά δεδομένα που παρέχονται από την πλατφόρμα Spotify.

## Κεφάλαιο 6: Παράθεση κώδικα

### 6.1. Δημιουργία του Dataset

Κάνοντας χρήση του Spotify API και τη βιβλιοθήκη Spotipy για την ανάκτηση χαρακτηριστικών τραγουδιών από διάφορες λίστες αναπαραγωγής στο Spotify. Επιλέγουμε λίστες αναπαραγωγής με βάση τη διάθεση που τις χαρακτηρίζει (ενεργητική, ευτυχισμένη, ήρεμη, λυπητερή).

```
import spotipy
from spotipy.oauth2 import SpotifyClientCredentials
from spotipy import util
import pandas as pd
import math
from IPython.display import clear_output
from typing import List
```

- Αρχικά, γίνεται εισαγωγή των απαραίτητων βιβλιοθηκών και η καταχώρηση μιας λίστας με τα χαρακτηριστικά που θα αναλυθούν για κάθε τραγούδι ("energy", "liveness",

"tempo", "speechiness", "acousticness", "instrumentalness", "danceability", "duration\_ms", "loudness", "valence").

- Στη συνέχεια, δημιουργείται μια σύνδεση με το λογαριασμό Spotify και αποκτάται η πρόσβαση στις λίστες αναπαραγωγής με τη χρήση του token, που είναι μια μορφή εξουσιοδότησης που επιτρέπει την πρόσβαση στα δεδομένα του εκάστοτε χρήστη.

```
features_for_mood = ['energy', 'liveness', 'tempo', 'speechiness',
                    'acousticness', 'instrumentalness', 'danceability', 'duration_ms',
                    'loudness', 'valence']

def get_track_features(track_ids, spotify):
    chunk_size = 50
    num_chunks = int(math.ceil(len(track_ids) / float(chunk_size)))
    features_add = []
    for i in range(num_chunks):
        chunk_track_ids = track_ids[i * chunk_size:min((i + 1) * chunk_size, len(track_ids))]
        chunk_features = spotify.audio_features(tracks=chunk_track_ids)
        features_add.extend(chunk_features)

    features_df = pd.DataFrame(features_add).drop(['id', 'analysis_url', 'key', 'mode', 'time_signature',
                                                'track_href', 'type', 'uri'], axis=1)
    features_df = features_df[features_for_mood]
    return features_df

token = util.prompt_for_user_token(username = 'theologia', scope='user-library-read',
                                  client_id='...',
                                  client_secret='...',
                                  redirect_uri='http://localhost:3000')

spotify = spotipy.Spotify(auth=token, requests_timeout=20)
```

- Συγκεκριμένα, παίρνουμε τα τραγούδια από κάθε λίστα αναπαραγωγής και αποθηκεύουμε τα χαρακτηριστικά που χρειαζόμαστε από αυτά (όπως το tempo, τον ήχο, τη διάρκεια κ.λπ.) σε μια Pandas DataFrame. Στη συνέχεια, ξεχωρίζουμε τα χαρακτηριστικά που σχετίζονται με τη διάθεση και τα αποθηκεύουμε σε ξεχωριστά αντικείμενα DataFrame για κάθε διάθεση.

```

playlists = {
  'Energetic' : ["https://open.spotify.com/playlist/4QDWboU5rwpDRXwYprwJf5?si=815ef93592e7475f",
                "https://open.spotify.com/playlist/0V32mTwwBzo6rNIk21owsY?si=1421ae7c97434fbf",
                "https://open.spotify.com/playlist/6112Df05PIoKK4CA61zbQV?si=2da800e88f5e47d3",
                "https://open.spotify.com/playlist/4ito9htFzo0bKJCf4AEeoD?si=5ece19a9f0304b94",
                "https://open.spotify.com/playlist/3XQ85b1Ipl54b6mq3LJDuF?si=eddb8adc010449c7"],
  'Happy' : ["https://open.spotify.com/playlist/7GhawGpb43Ctkq3PRP1f0L?si=5f250f365dc2442a",
             "https://open.spotify.com/playlist/4AnAUkQNrLK1JCInZGSXRO?si=9846e647548d4026",
             "https://open.spotify.com/playlist/0RH319xCjeU8VyTSqCF6M4?si=a1acb06d586242a2",
             "https://open.spotify.com/playlist/4Izx0RxRHM7VaH1XDBgZCa?si=7f83455adc5a455f",
             "https://open.spotify.com/playlist/6M5QoiSaeIq4ZbZY1WJhbF?si=bca50ea0933b4d47"],
  'Calm' : ["https://open.spotify.com/playlist/7fDbhStlkB7nYlGftLYkCg?si=e700cffbc8ac4a54",
            "https://open.spotify.com/playlist/4h2MD8T5fNw2Ss8s05up68?si=25acd37e15ba475a",
            "https://open.spotify.com/playlist/6txDrYS0NwEhdcAt1c3US?si=c06f8ab465664901",
            "https://open.spotify.com/playlist/0OQLmoDiKrLNK0oqAc2SK?si=58cf44cf441d48c9",
            "https://open.spotify.com/playlist/6gPRcy5xb4sHizGeL9EoIo?si=0fefcb050a1e435c"],
  'Sad' : ["https://open.spotify.com/playlist/1hb1Pa5R3QgJGwoazGIbWG?si=d5b9f5796ad148f8",
           "https://open.spotify.com/playlist/4yXfnhz0BRReoVfwwYRtPBm?si=6286c1917ce447fd",
           "https://open.spotify.com/playlist/71dOK1kQ5dxfdi2DFckrq9?si=0ac170d469214371",
           "https://open.spotify.com/playlist/1n2MYypuxk9ueZD2iGYJXB?si=9254983660ab454a",
           "https://open.spotify.com/playlist/371vXUKm0x0azCovhaAuA9?si=49f0ec96cc5a4f64"]
}

```

- Αρχικά, λαμβάνονται τα κομμάτια της λίστας αναπαραγωγής Spotify με τη χρήση του αναγνωριστικού playlist (id). Τα αποτελέσματα αποθηκεύονται στην λίστα pl\_tracks.
- Στη συνέχεια, λαμβάνονται τα ids των κομματιών από τη λίστα pl\_tracks και αποθηκεύονται στη μεταβλητή ids.
- Στο μπλοκ except, εκτελείται μια εναλλακτική ενέργεια για την περίπτωση σφάλματος, δηλαδή στην περίπτωση που ο κώδικας συναντήσει οποιοδήποτε πρόβλημα κατά τη λήψη των κομματιών της λίστας αναπαραγωγής. Συγκεκριμένα, εκτυπώνεται ο σύνδεσμος της playlist που προκάλεσε το σφάλμα και συνεχίζεται η επόμενη επανάληψη του βρόχου.
- Στη συνέχεια, καλείται η συνάρτηση get\_track\_features για να λάβει τα χαρακτηριστικά των κομματιών που αντιστοιχούν στα ids που αποθηκεύσαμε πιο πριν.
- Προσθέτουμε δύο νέες στήλες στο αντικείμενο features που περιλαμβάνουν τα αναγνωριστικά των κομματιών και τη διάθεσή τους (mood).
- Αποθηκεύουμε το αντικείμενο tracks σε ένα αρχείο CSV με όνομα tracks.csv.

```

tracks = pd.DataFrame()
moods = []

for mood, links in playlists.items():
    print (mood)
    for link in links:
        id = link[34:56]
        try:
            pl_tracks = spotify.playlist_tracks(id)['items']
            ids = [foo['track']['id'] for foo in pl_tracks]
        except:
            print (link)
            continue
        features = get_track_features(ids, spotify)
        features['id'] = ids
        features['mood'] = mood
        tracks = tracks.append(features)

tracks.to_csv('tracks.csv')

```

## 6.2. Κανονικοποίηση δεδομένων, εκπαίδευση μοντέλων & δημιουργία γραφημάτων

- Αφού εισάγουμε τις απαραίτητες βιβλιοθήκες για την ανάλυση και την απεικόνιση δεδομένων στην Python: pandas, seaborn, matplotlib και plotly, φορτώνουμε το σύνολο δεδομένων με την χρήση της βιβλιοθήκης pandas, το οποίο περιλαμβάνεται στο αρχείο tracks.csv.

```

# import the libraries

import pandas as pd
import seaborn as sns
import matplotlib.pyplot as plt
import plotly.express as px
from sklearn.preprocessing import OrdinalEncoder

# load the dataset
df=pd.read_csv(r"C:\Users\Theologia\Documents\thesis\tracks.csv")
df.tail()

```

	name	album	artist	id	release_date	popularity	length	danceability	acousticness	energy	instrumentalness	liveness	valence	loudness	speechiness	tempo	key	time_signature	mood
681	windcatcher	windcatcher	Leo Nocta	59VAp8BrS2IADQk4ml5mdo	2020-06-19	36	123066	0.402	0.961	0.2360	0.919000	0.0921	0.1460	-20.615	0.0603	129.736	0	3	Calm
682	yellow is the color of her eyes	yellow is the color of her eyes	Soccer Mommy	4D3nttjPU6L0M2ep7sl6	2019-11-19	5	435080	0.452	0.757	0.5150	0.120000	0.1400	0.1910	-7.351	0.0255	80.537	11	4	Sad
683	you broke me first	you broke me first	Tate McRae	45bE4H000AwGZXZiMp0JR	2020-04-17	87	169265	0.642	0.786	0.3740	0.000000	0.0906	0.0799	-9.386	0.0545	124.099	4	4	Sad
684	you were good to me	brent	Jeremy Zucker	4CxFNSzON70B3VOP8Ybd6P	2019-05-03	76	219146	0.561	0.913	0.0848	0.000026	0.1120	0.2060	-15.099	0.0404	102.128	2	4	Sad
685	aefre	aefre	praam	2irbT1BSYalEF44PlyKaoM	2020-07-17	41	186331	0.377	0.994	0.0156	0.881000	0.0991	0.0804	-28.435	0.0397	140.179	0	4	Calm



```
# finding all the column names of the data
df.columns
```

```
Index(['name', 'album', 'artist', 'id', 'release_date', 'popularity', 'length',
      'danceability', 'acousticness', 'energy', 'instrumentalness',
      'liveness', 'valence', 'loudness', 'speechiness', 'tempo', 'key',
      'time_signature', 'mood'],
      dtype='object')
```

- Το αποτέλεσμα της ομαδοποίησης και του υπολογισμού του αριθμού των εγγραφών αποθηκεύεται στη μεταβλητή `classcnt` ως ένα νέο DataFrame που έχει δύο στήλες: τη στήλη "mood" που περιέχει τις μοναδικές τιμές της στήλης mood και τη στήλη count που περιέχει τον αριθμό των εγγραφών για κάθε μοναδική τιμή της στήλης mood.

```
classcnt = df.groupby(['mood'])['mood'].count()
classcnt
```

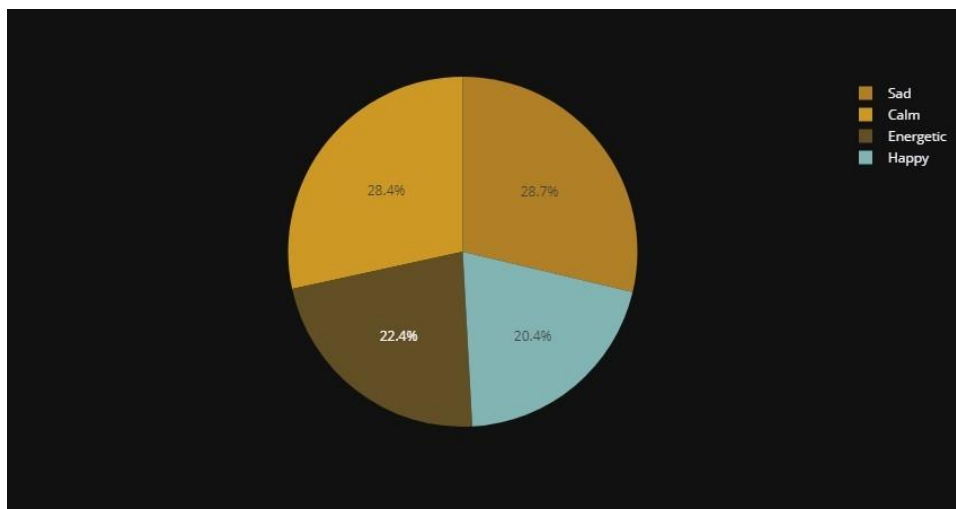
```
# finding all the columns names of data
df.head()
```

	name	album	artist	id	release_date	popularity	length	danceability	acousticness	energy	instrumentalness	liveness	valence	loudness	speechiness	tempo	key	time_signature	mood
0	1999	1999	Prince	2H7PHVdQ3mXqEHKvcvT80	1982-10-27	68	379266	0.866	0.13700	0.730	0.000000	0.0843	0.625	-8.201	0.0767	118.523	5	4	Happy
1	23	23	Blonde Redhead	4HlwL9t6CoXpT0ZtMq0MP	2007-04-16	43	318800	0.381	0.01890	0.832	0.196000	0.1530	0.166	-5.069	0.0492	120.255	8	4	Sad
2	9 Crimes	9	Damien Rice	5GZzeowhvSiefDIR8iQ2im	2006-11-06	60	217946	0.346	0.91300	0.139	0.000077	0.0934	0.116	-15.326	0.0321	136.168	0	4	Sad
3	99 Luftballons	99 Luftballons	Nena	6HA97v4wEGG5TUCRMOXLc	1984-08-21	2	233000	0.466	0.08900	0.438	0.000006	0.1130	0.587	-12.858	0.0608	193.100	4	4	Happy
4	A Boy Brushed Red Living In Black And White	They're Only Chasing Safety	Underoath	47WLBKOKhFmz1FUEUIKE	2004-01-01	60	268000	0.419	0.00171	0.932	0.000000	0.1370	0.445	-3.604	0.1060	169.881	1	4	Energetic

- Στη συνέχεια, ο δημιουργούμε μια λίστα χρωμάτων RGB, με το όνομα `sunflowers_colors` ώστε να χρησιμοποιηθεί στα γραφήματα παρακάτω.
- Έπειτα, χρησιμοποιούμε τη βιβλιοθήκη Plotly για να δημιουργήσει ένα γράφημα πίτας (pie chart), χρησιμοποιώντας τα δεδομένα ως τιμές τα data της στήλης moods και ως ετικέτες τη στήλη names του DataFrame `classcnt` που ορίσαμε πάνω.

```
sunflowers_colors = ['rgb(177, 127, 38)', 'rgb(205, 152, 36)', 'rgb(99, 79, 37)',
                    'rgb(129, 180, 179)', 'rgb(124, 103, 37)']
px.pie(classcnt, values='mood', names= classcnt.index, color_discrete_sequence=sunflowers_colors, template='plotly_dark')
```





- Κρατάμε μόνο τις στήλες του Data Frame που θεωρούμε χρήσιμες ως προς την συναισθηματική ανάλυση.

```
# εκw polla axrista rows opote tha kanw ta filtering
df=df[['acousticness','danceability','energy','liveness','valence','tempo','speechiness','loudness','instrumentalness','mood']]
df.head()
```

	acousticness	danceability	energy	liveness	valence	tempo	speechiness	loudness	instrumentalness
0	0.13700	0.866	0.730	0.0843	0.625	118.523	0.0767	-8.201	0.000000
1	0.01890	0.381	0.832	0.1530	0.166	120.255	0.0492	-5.069	0.196000
2	0.91300	0.346	0.139	0.0934	0.116	136.168	0.0321	-15.326	0.000077
3	0.08900	0.466	0.438	0.1130	0.587	193.100	0.0608	-12.858	0.000006
4	0.00171	0.419	0.932	0.1370	0.445	169.881	0.1060	-3.604	0.000000

- Παρακάτω εφαρμόζεται η κωδικοποίηση των κατηγορικών μεταβλητών με χρήση της μεθόδου OrdinalEncoder στα χαρακτηριστικά που ορίζονται στη λίστα features για το DataFrame με όνομα df.

Πιο συγκεκριμένα, η μέθοδος OrdinalEncoder μετατρέπει κάθε μοναδική τιμή σε κάθε κατηγορική μεταβλητή, σε έναν ακέραιο αριθμό, που αντιστοιχεί στη σειρά της τιμής της κατηγορικής μεταβλητής στο σύνολο των μοναδικών τιμών.

Έτσι, με το κομμάτι κώδικα `df[features] = enc.fit_transform(df[features])`, οι κατηγορικές μεταβλητές που ορίζονται στη λίστα features του DataFrame df αντικαθίστανται από αριθμητικές τιμές, που αντιστοιχούν στη σειρά της τιμής της κατηγορικής μεταβλητής στο σύνολο των μοναδικών τιμών της κάθε μεταβλητής. Το αποτέλεσμα είναι μια αντικατάσταση των αρχικών κατηγορικών μεταβλητών με νέες αριθμητικές μεταβλητές, που μπορούν να χρησιμοποιηθούν ως είσοδος για αλγόριθμους μηχανικής μάθησης.

```
#χρειάζετε ordinal encoder γιατί theloume ta data na einai apo 0 -> 1
features = ['acousticness', 'danceability', 'energy', 'liveness', 'valence', 'tempo', 'speechiness', 'loudness', 'instrumentalness']
enc = OrdinalEncoder()
df[features] = enc.fit_transform(df[features])
df
```

	acousticness	danceability	energy	liveness	valence	tempo	speechiness	loudness	instrumentalness	mood
0	252.0	400.0	361.0	92.0	413.0	313.0	296.0	393.0	0.0	Happy
1	168.0	118.0	403.0	236.0	165.0	341.0	190.0	535.0	284.0	Sad
2	459.0	92.0	116.0	149.0	127.0	526.0	57.0	180.0	95.0	Sad
3	232.0	182.0	248.0	197.0	396.0	663.0	251.0	230.0	35.0	Happy
4	89.0	144.0	467.0	221.0	335.0	622.0	340.0	630.0	0.0	Energetic
...	...	...	...	...	...	...	...	...	...	...
681	487.0	134.0	160.0	138.0	148.0	482.0	248.0	97.0	432.0	Calm
682	396.0	172.0	279.0	224.0	186.0	73.0	3.0	427.0	273.0	Sad
683	407.0	321.0	224.0	129.0	87.0	390.0	223.0	339.0	0.0	Sad
684	459.0	260.0	88.0	196.0	198.0	190.0	131.0	183.0	72.0	Sad
685	519.0	114.0	20.0	179.0	88.0	549.0	125.0	17.0	401.0	Calm

```
# shaping of the datasets
df.shape
```

- Παρακάτω εισάγουμε τη συνάρτηση `train_test_split` από τη βιβλιοθήκη `sklearn.model_selection` και τη χρησιμοποιούμε για να χωρίσουμε το σύνολο δεδομένων σε τρία μέρη: `X_train`, `X_test` και `y_train`, `y_test`. Το **X** είναι ένας πίνακας που περιέχει τα χαρακτηριστικά των δεδομένων (features) και το **y** είναι ένας πίνακας που περιέχει τις ετικέτες των δεδομένων (labels).

Η συνάρτηση `train_test_split` δημιουργεί τυχαίες τεμαχιοποιήσεις του συνόλου δεδομένων και επιστρέφει τα κομμάτια των χαρακτηριστικών και των ετικετών που αντιστοιχούν στο training set και στο test set. Το `test_size = 0.2` δηλώνει ότι το test set θα αποτελεί το 20% του συνόλου δεδομένων και το `random_state = 0` χρησιμοποιείται για να επαναληφθούν οι ίδιες τυχαίες τεμαχιοποιήσεις σε διαφορετικές εκτελέσεις του κώδικα, διασφαλίζοντας έτσι την αναπαραγωγικότητα των αποτελεσμάτων. .

Οι μεταβλητές **X\_train** και **y\_train** περιέχουν τα δεδομένα που θα χρησιμοποιηθούν για την εκπαίδευση του μοντέλου, ενώ οι μεταβλητές **X\_test** και **y\_test** περιέχουν τα δεδομένα που θα χρησιμοποιηθούν για τον έλεγχο της απόδοσης του μοντέλου.

```
# Splitting the datasets into training and test (80/20)
from sklearn.model_selection import train_test_split
X_train,X_test,y_train,y_test=train_test_split(X,y,test_size=0.2,random_state=0)
```

- Εκτυπώνουμε πληροφορίες σχετικά με τις διαστάσεις των πινάκων που περιέχουν τα δεδομένα εισόδου και εξόδου για το σετ εκπαίδευσης και το σετ ελέγχου. Συγκεκριμένα, εμφανίζεται ο αριθμός των δειγμάτων καθώς και ο αριθμός των χαρακτηριστικών ή των ετικετών για κάθε σετ.

```
# Shape of train and test input and outputs
print("Train input data", X_train.shape)
print("Test input data", X_test.shape)
print("Train output data", y_train.shape)
print("Test output data", y_test.shape)
```

```
Train input data (548, 9)
Test input data (138, 9)
Train output data (548,)
Test output data (138,)
```

- Η συνάρτηση `metrics` που υλοποιείται παρακάτω αναλαμβάνει να υπολογίσει και να εμφανίσει τις μετρικές απόδοσης ενός μοντέλου ταξινόμησης, δηλαδή το `accuracy score` και το `f1 measure`, και να εμφανίσει ένα `heatmap` με το `confusion matrix`.

Πιο συγκεκριμένα, η συνάρτηση αυτή παίρνει ως είσοδο ένα μοντέλο (πχ. έναν ταξινομητή) καθώς και ένα σετ δεδομένων εισόδου και ετικετών ( $x$  και  $y$  αντίστοιχα) και εκτελεί τα εξής βήματα:

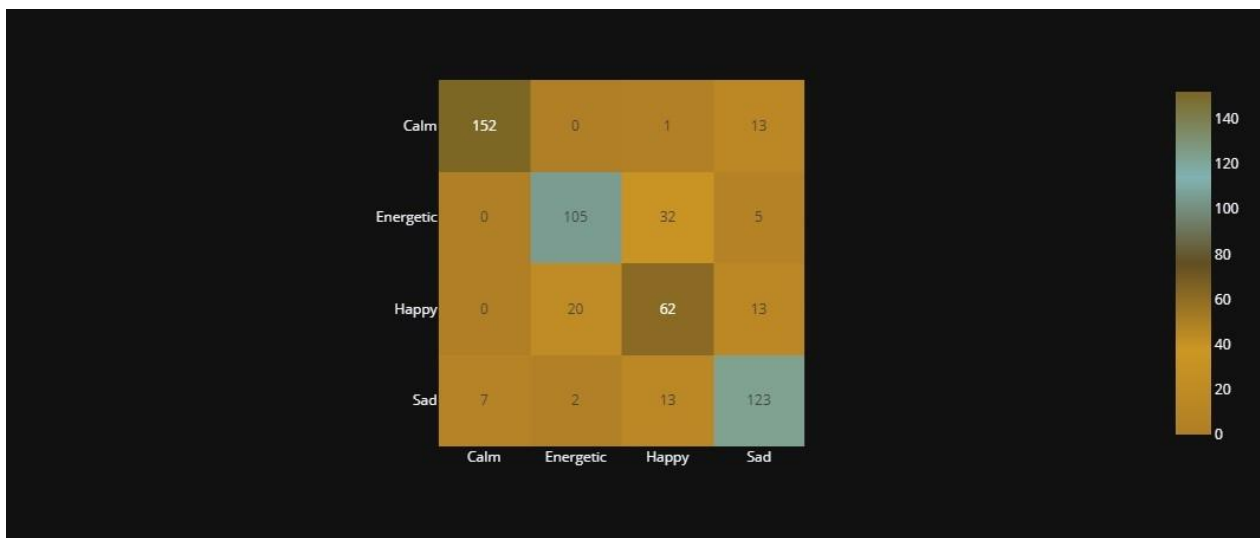
1. Υπολογίζει το `accuracy score` του μοντέλου για το σετ δεδομένων εισόδου και ετικετών που δόθηκε, χρησιμοποιώντας την συνάρτηση `accuracy_score` από τη βιβλιοθήκη `scikit-learn`.
2. Υπολογίζει το `f1 measure` του μοντέλου για το σετ δεδομένων εισόδου και ετικετών που δόθηκε, χρησιμοποιώντας την συνάρτηση `f1_score` από τη βιβλιοθήκη `scikit-learn`.
3. Χρησιμοποιώντας την συνάρτηση `confusion_matrix` από τη βιβλιοθήκη `scikit-learn`, υπολογίζει το `confusion matrix` του μοντέλου για το σετ δεδομένων εισόδου και ετικετών που δόθηκε.

```
from sklearn.metrics import accuracy_score, f1_score, confusion_matrix, recall_score
def metrics(model, x, y):
    print('The accuracy score is: ', round(accuracy_score(model.predict(x),y),3))
    print('The f1 measure is: ', round(f1_score(model.predict(x),y,average='weighted'),3))
    print('The recall is: ', recall_score(model.predict(x),y, average='weighted'))
    px.imshow(confusion_matrix(model.predict(x),y),text_auto=True,template='plotly_dark', color_continuous_scale = sunflowers_colors,x=model.classes_, y=model.classes_).show()
```

- Δημιουργούμε ένα αντικείμενο `log_reg` τύπου `LogisticRegression()` που εκπαιδεύεται στα δεδομένα `X_train` και `y_train` με την χρήση της συνάρτησης `fit()`.
- Στη συνέχεια, υπολογίζουμε με τη συνάρτηση `metrics()` τις διάφορες μετρικές απόδοσης του μοντέλου `log_reg` στα δεδομένα εκπαίδευσης `X_train` και `y_train`.

```
# Creating the model using logistic regression
from sklearn.linear_model import LogisticRegression
# fit the model on training datasets
log_reg=LogisticRegression().fit(X_train,y_train)
metrics(log_reg,X_train,y_train)
```

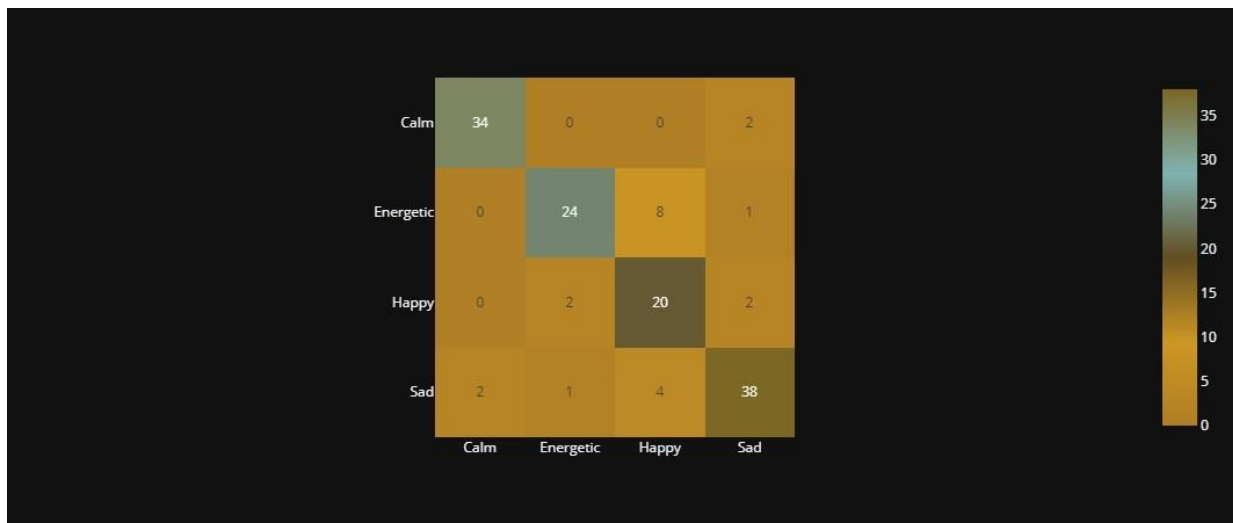
```
The accuracy score is: 0.807
The f1 measure is: 0.809
The recall is: 0.8065693430656934
```



- Χρησιμοποιούμε την `metrics()` και για την αξιολόγηση του ταξινομητή logistic regression στα δεδομένα ελέγχου (test set) `X_test` και `y_test`.

```
metrics(log_reg,X_test,y_test)
```

```
The accuracy score is: 0.841
The f1 measure is: 0.844
The recall is: 0.8405797101449275
```



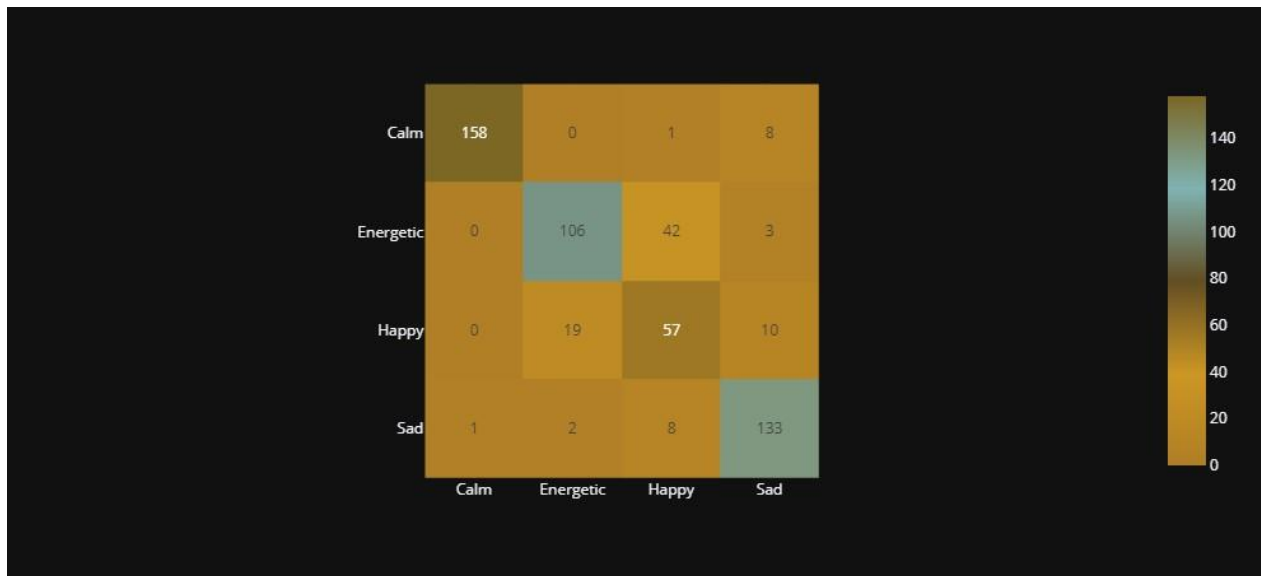
- Παρομοίως με πριν, δημιουργούμε ένα αντικείμενο `KNeighborsClassifier` το οποίο χρησιμοποιείται για την κατηγοριοποίηση παραδειγμάτων μέσω του αλγορίθμου k-NN (k-Nearest Neighbors).

Στη συνέχεια, το αντικείμενο αυτό εκπαιδεύεται με τα δεδομένα εκπαίδευσης `X_train` και `y_train` χρησιμοποιώντας τη μέθοδο `fit()`. Επίσης, εμφανίζουμε τις αντίστοιχες μετρικές.

```
from sklearn.neighbors import KNeighborsClassifier
knn = KNeighborsClassifier().fit(X_train,y_train)
```

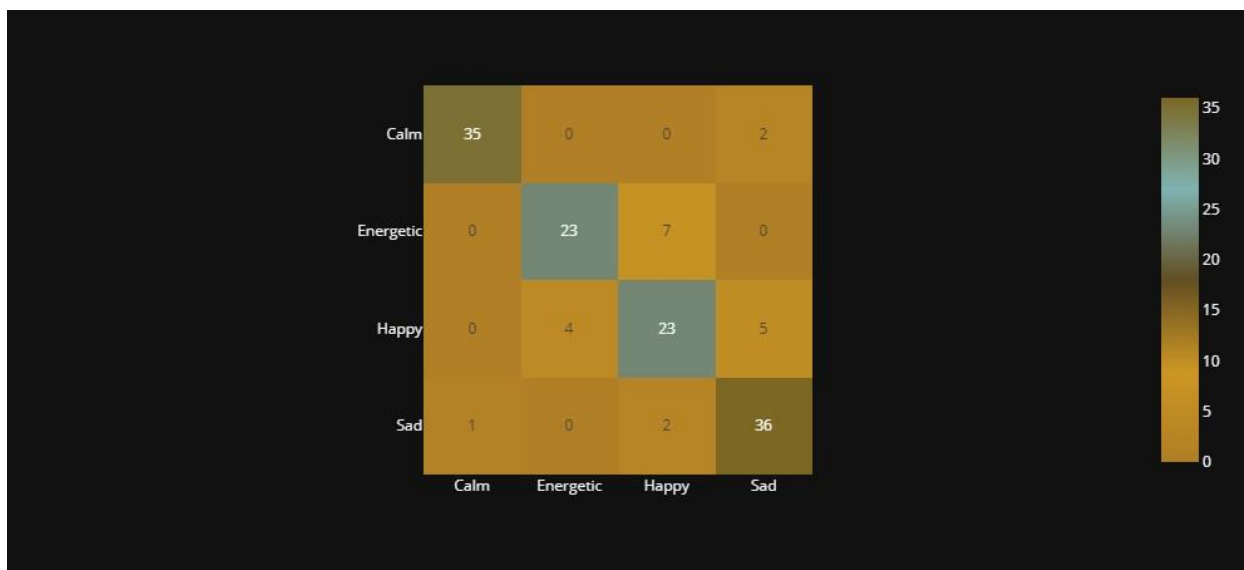
```
metrics(knn,X_train,y_train)
```

```
The accuracy score is: 0.828
The f1 measure is: 0.832
The recall is: 0.8284671532846716
```



```
metrics(knn,X_test,y_test)
```

```
The accuracy score is: 0.848
The f1 measure is: 0.847
The recall is: 0.8478260869565217
```



✓ Παρατηρούμε πως οι μετρικές των δεδομένων ελέγχου και εκπαίδευσης είναι πολύ κοντά μεταξύ τους.

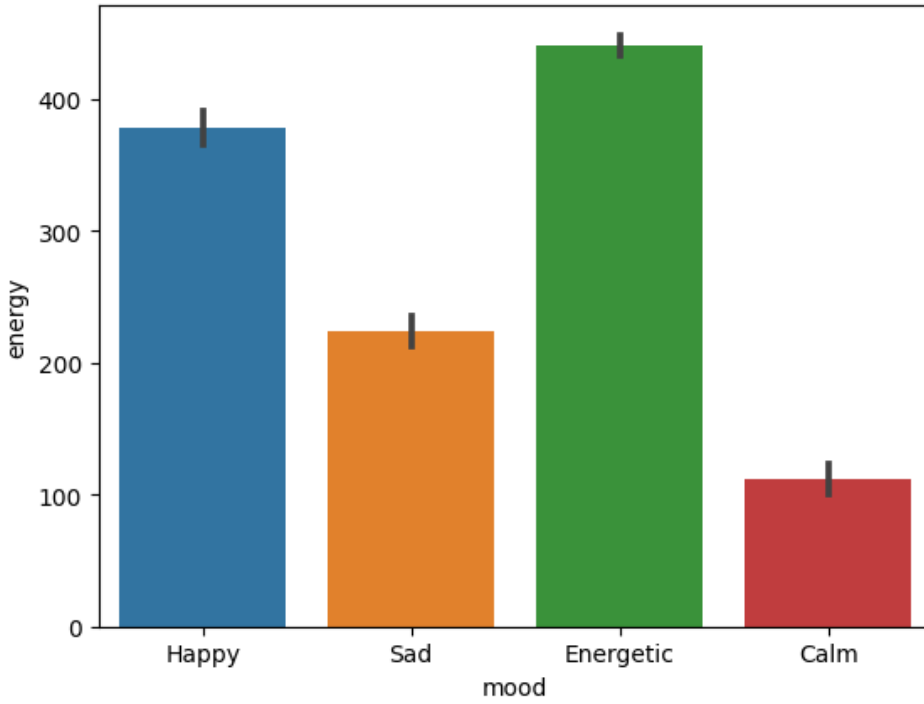
Όταν η ακρίβεια εκπαίδευσης ενός αλγορίθμου είναι πολύ κοντά στην ακρίβεια δοκιμής, αυτό σημαίνει ότι ο αλγόριθμος έχει μάθει να γενικεύει καλά τα δεδομένα εκπαίδευσης σε νέα δεδομένα που δεν έχει “ζαναδεί”. Αυτό είναι ένα σημαντικό κριτήριο για την απόδοση του αλγορίθμου, καθώς ο στόχος είναι να επιτυγχάνει υψηλή ακρίβεια όχι μόνο στα δεδομένα εκπαίδευσης αλλά και στα νέα δεδομένα που θα χρησιμοποιηθούν στην πράξη.

Εάν ο αλγόριθμος έχει υψηλή ακρίβεια εκπαίδευσης αλλά χαμηλή ακρίβεια δοκιμής, αυτό μπορεί να σημαίνει ότι έχει μάθει να "απομνημονεύει" τα δεδομένα εκπαίδευσης και δεν είναι σε θέση να γενικεύσει σε νέα δεδομένα. Αυτό ονομάζεται υπερπροσαρμογή (overfitting) στα δεδομένα εκπαίδευσης και μπορεί να οδηγήσει σε κακή απόδοση στην πράξη.

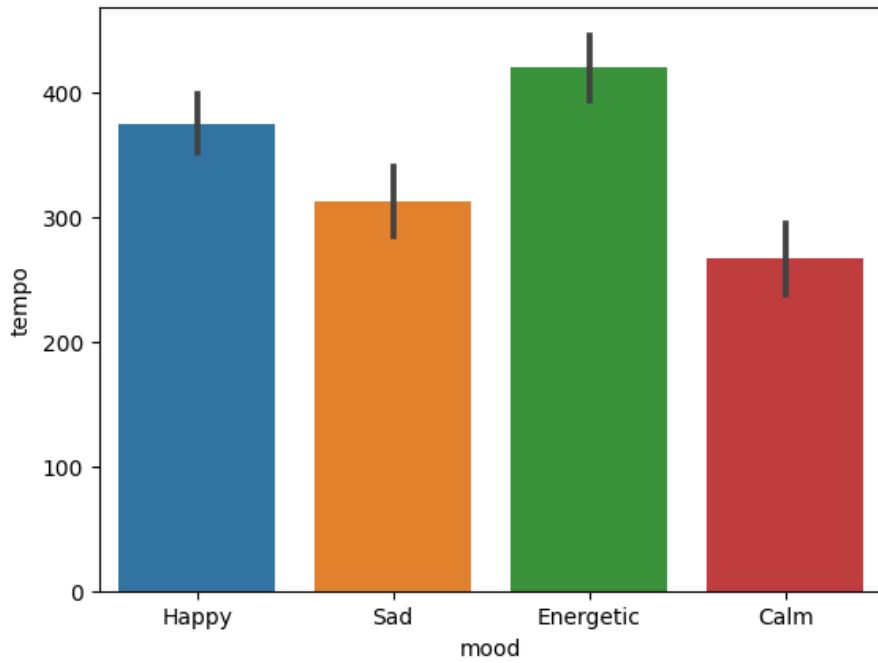
- ✓ Μετά την εκπαίδευση του αλγορίθμου με τα δεδομένα εκπαίδευσης, μπορεί να χρησιμοποιηθεί για την κατηγοριοποίηση νέων παραδειγμάτων χρησιμοποιώντας τη μέθοδο `predict()`.

- Παρακάτω κάνουμε χρήση της βιβλιοθήκης `seaborn` (η οποία βασίζεται στη βιβλιοθήκη `matplotlib`) για να δημιουργήσουμε διαγράμματα τύπου `barplot` που δείχνουν τη σχέση μεταξύ ενός χαρακτηριστικού των τραγουδιών και των συναισθημάτων που εξετάζουμε αντικατοπτρίζονται στα δεδομένα που περιέχονται στο `DataFrame df`.

```
sns.barplot(df,x='mood',y='energy')  
plt.show()
```

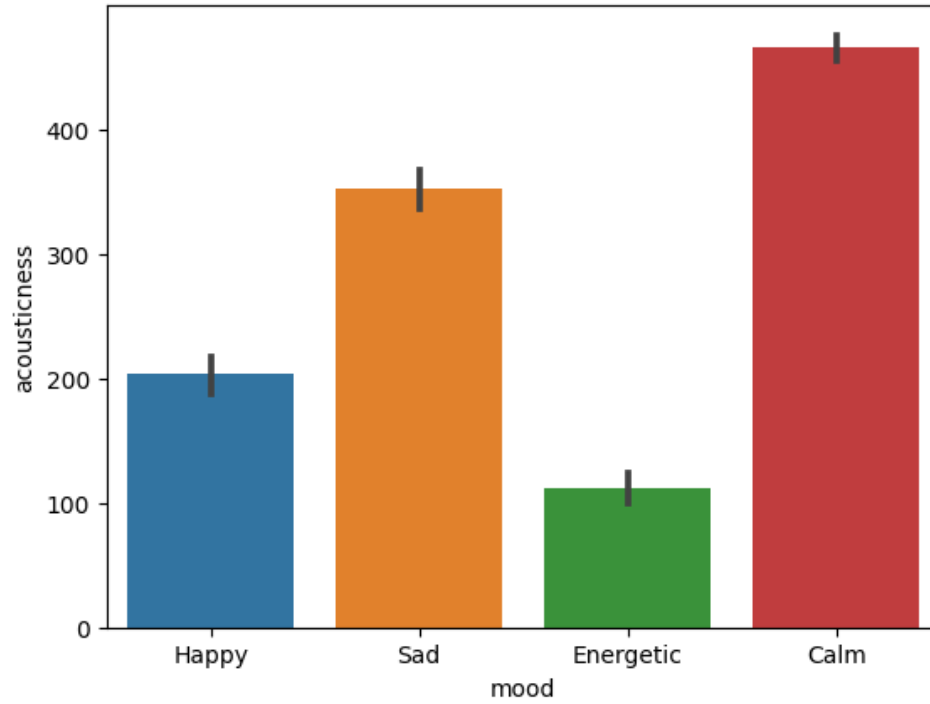


```
sns.barplot(df,x='mood',y='tempo')  
plt.show()
```

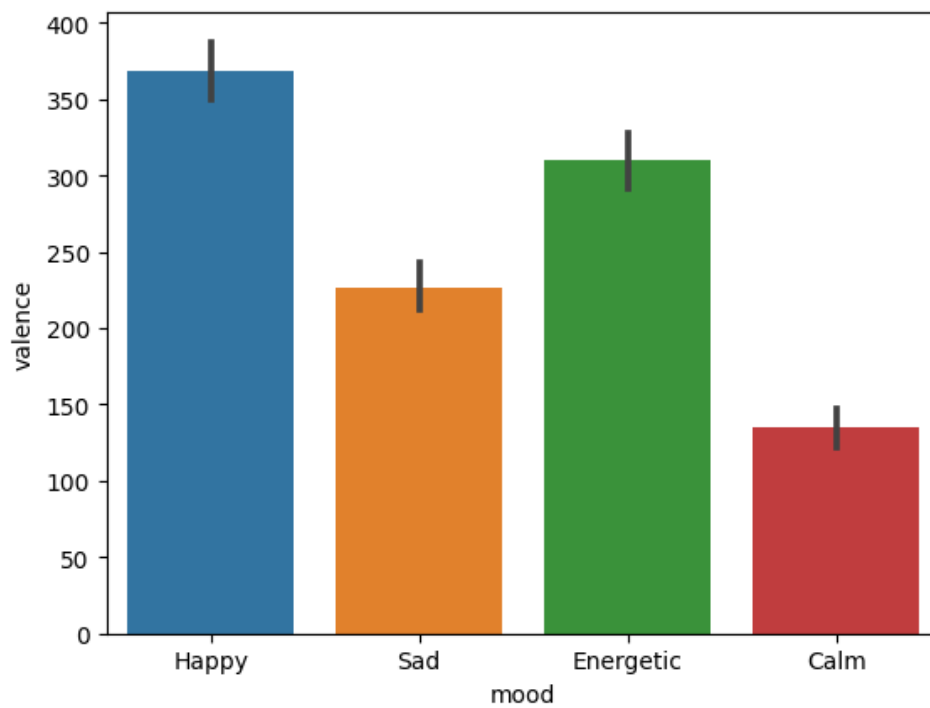




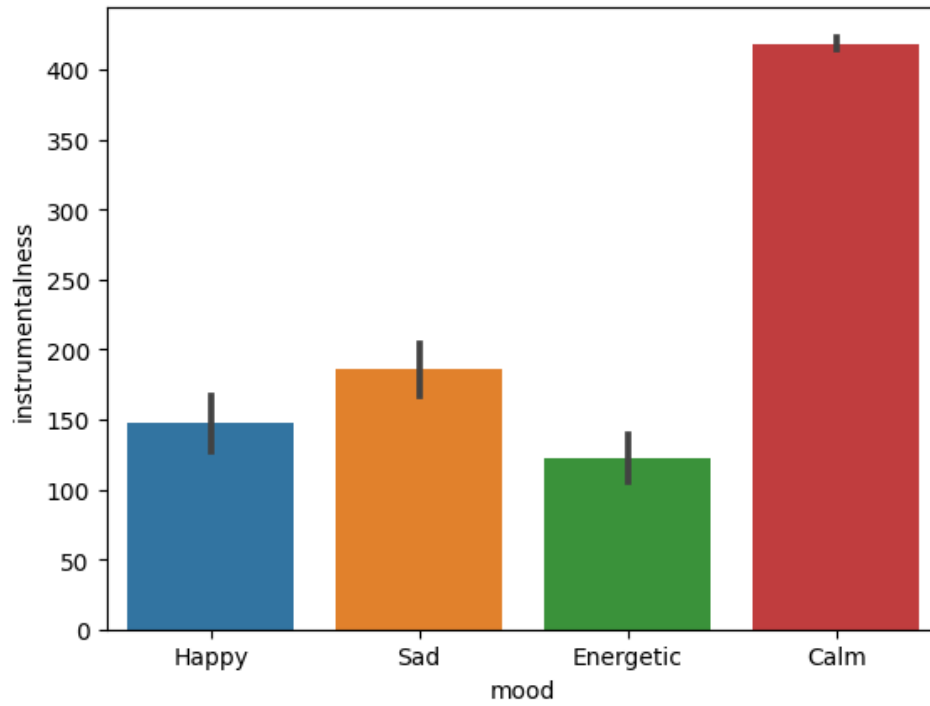
```
sns.barplot(df,x='mood',y='acousticness')
plt.show()
```



```
sns.barplot(df,x='mood',y='valence')
plt.show()
```



```
sns.barplot(df,x='mood',y='instrumentalness')  
plt.show()
```



## Κεφάλαιο 7: Παρατηρήσεις & Συμπεράσματα

### 7.1. Αξιολόγηση

#### 7.1.1. Ακρίβεια

Δεν υπάρχει ένα ακριβές ποσοστό για το πόσο καλά θα αποδώσει ένα μοντέλο μηχανικής μάθησης, καθώς αυτό εξαρτάται από διάφορους παράγοντες, όπως ο τύπος του μοντέλου, οι παράμετροί του, τα δεδομένα που χρησιμοποιούνται για την εκπαίδευσή του, η πολυπλοκότητα του προβλήματος και ποιότητα των δεδομένων ελέγχου. Ωστόσο, ένα καλό ποσοστό ακρίβειας για ένα εκπαιδευμένο μοντέλο μπορεί να κυμαίνεται από το 70% έως το 99%, ανάλογα με το πρόβλημα και τον τύπο του μοντέλου. Για παράδειγμα, σε κάποια προβλήματα αναγνώρισης εικόνων, το καλό ποσοστό ακρίβειας μπορεί να είναι κοντά στο 99%, ενώ σε άλλα προβλήματα, όπως η πρόβλεψη τιμών μετοχών και η περίπτωση μας, ένα καλό ποσοστό ακρίβειας μπορεί να είναι κοντά στο 70%.

Επιπλέον, είναι σημαντικό να σημειωθεί ότι ένα μοντέλο μπορεί να έχει υψηλό ποσοστό ακρίβειας στα δεδομένα εκπαίδευσης αλλά να μην γενικεύει καλά σε νέα δεδομένα ελέγχου. Η αξιολόγηση της ακρίβειας ενός μοντέλου αναγνώρισης βασίζεται στην ακρίβεια του test set, καθώς το test set περιλαμβάνει δεδομένα που το μοντέλο δεν έχει δει κατά τη διάρκεια της εκπαίδευσης. Αν η απόδοση του μοντέλου αξιολογηθεί μόνο με βάση την ακρίβεια του train set, τότε το μοντέλο μπορεί να φανεί υπερβολικά βελτιωμένο ή υπερβολικά απλοϊκό, καθώς έχει προσαρμοστεί στα δεδομένα εκπαίδευσης.

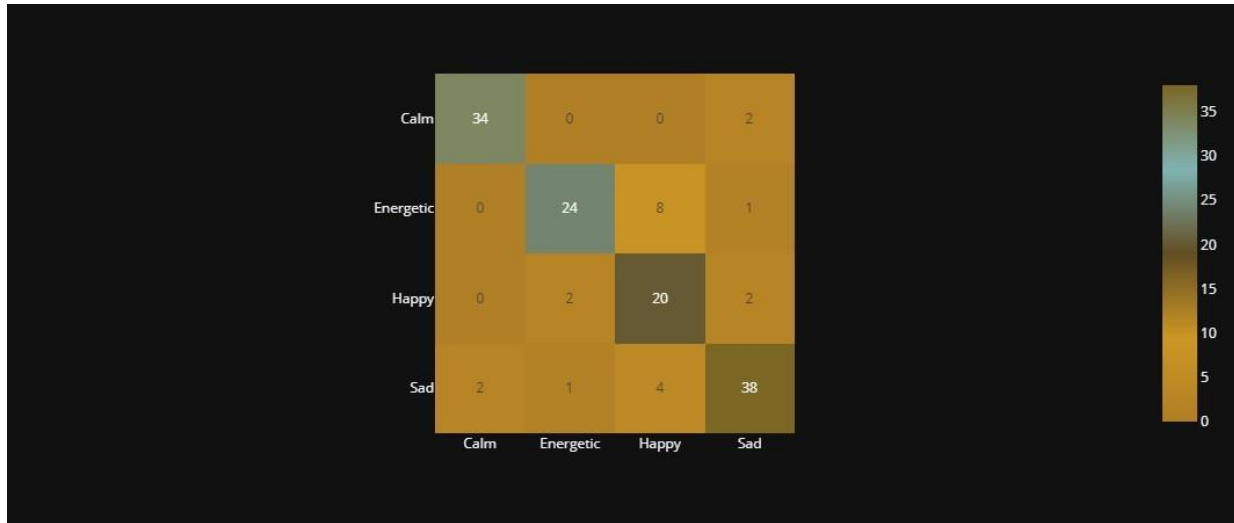
Το test set πρέπει να είναι ανεξάρτητο από το train set και να περιλαμβάνει δεδομένα που είναι αντιπροσωπευτικά της πραγματικής χρήσης του μοντέλου. Η αξιολόγηση της ακρίβειας του μοντέλου πρέπει να γίνεται σε αυτά τα αντιπροσωπευτικά δεδομένα για να είναι έγκυρη και αξιόπιστη. Η χρήση του test set για την αξιολόγηση της ακρίβειας του μοντέλου βοηθά να διασφαλιστεί ότι το μοντέλο είναι ικανό να γενικεύσει την εμπειρία του σε νέα δεδομένα και δεν προσαρμόζεται αποκλειστικά στα δεδομένα εκπαίδευσης.

- Για την περίπτωση που μελετούμε, τα μοντέλα αναγνώρισης – ταξινόμησης που εκπαιδεύσαμε μας δίνουν ποσοστό ακρίβειας 84%, το οποίο είναι αρκετά ικανοποιητικό.

Το ποσοστό ακρίβειας 84% σημαίνει ότι το μοντέλο κάνει σωστές προβλέψεις στο 84% των περιπτώσεων. Αυτό είναι αρκετά υψηλό και μπορεί να βοηθήσει στη μείωση των λανθασμένων αποφάσεων και των σφαλμάτων. Η ακρίβεια 84% μπορεί να επιτευχθεί μόνο με μία καλά εκπαιδευμένη μηχανή μάθησης και με μεγάλο όγκο δεδομένων για εκπαίδευση. Επομένως, ένα τέτοιο μοντέλο είναι αποδοτικό και αξιόπιστο.

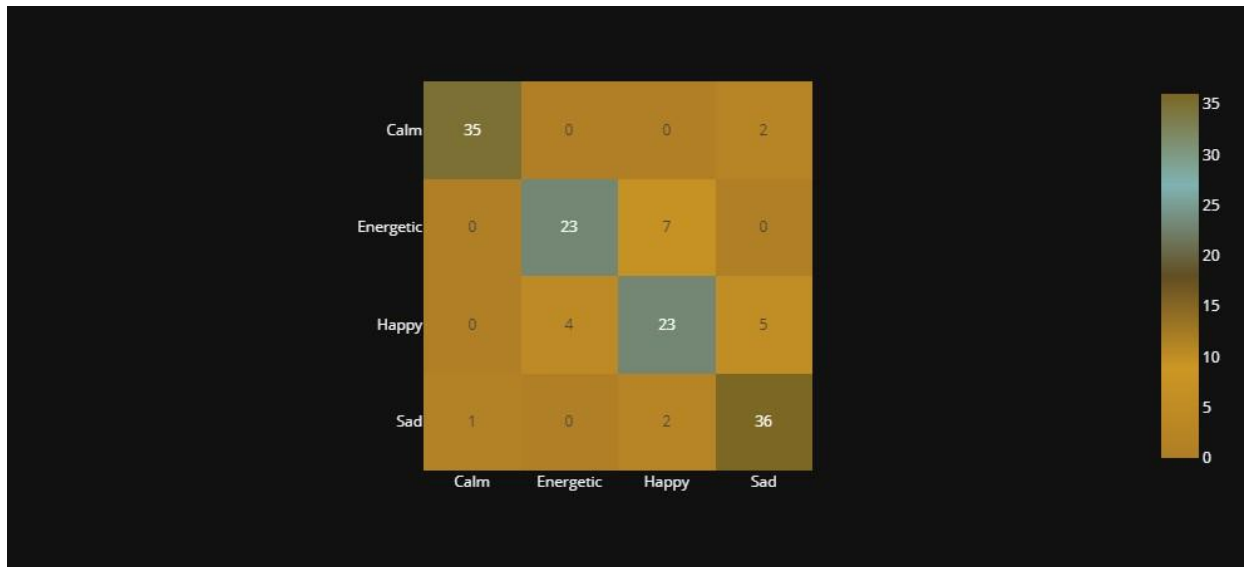
Επιπλέον, το ποσοστό ακρίβειας είναι κάτι που μπορεί να βελτιωθεί στο μέλλον, με τη συνεχή εκπαίδευση και βελτιστοποίηση του μοντέλου, είναι δυνατό να επιτευχθεί ακόμα μεγαλύτερη ακρίβεια.

### 7.1.2. Πίνακας Σύγχυσης (Confusion matrix)



Παρακάτω αναλύουμε τα αποτελέσματα του πίνακα σύγχυσης που προέκυψε από τη δοκιμή του μοντέλου Λογιστικής Παλινδρόμησης:

- Στην πρώτη γραμμή του πίνακα, το μοντέλο προέβλεψε ότι 34 από τα 36 κομμάτια της κατηγορίας Calm είναι στην πραγματικότητα Calm, ενώ 2 από αυτά προβλέφθηκαν λανθασμένα ως Sad.
- Για την κατηγορία Energetic, ο ταξινομητής προέβλεψε σωστά σε 24 περιπτώσεις, ενώ σε 8 περιπτώσεις προέβλεψε λανθασμένα ως Happy και σε μία περίπτωση ως Sad.
- Για την κατηγορία Happy, ο ταξινομητής προέβλεψε σωστά σε 20 περιπτώσεις, ενώ σε 2 περιπτώσεις προέβλεψε λανθασμένα ως Sad και σε μία περίπτωση ως Energetic.
- Για την κατηγορία Sad, ο ταξινομητής προέβλεψε σωστά σε 38 περιπτώσεις, ενώ σε 2 περιπτώσεις προέβλεψε λανθασμένα ως Calm και σε μία περίπτωση ως Energetic.



Παρακάτω αναλύουμε τα αποτελέσματα του πίνακα σύγχυσης που προέκυψε από τη δοκιμή του μοντέλου K-πλησιέστερων γειτόνων:

- Η κατηγορία Calm έχει προβλεφθεί σωστά 35 φορές, ενώ έχουν προβλεφθεί λανθασμένα ως Energetic 2 φορές.
- Η κατηγορία Energetic έχει προβλεφθεί σωστά 23 φορές, ενώ έχουν προβλεφθεί λανθασμένα ως Happy 7 φορές.
- Η κατηγορία Happy έχει προβλεφθεί σωστά 23 φορές, ενώ έχουν προβλεφθεί λανθασμένα ως Sad 5 φορές.
- Η κατηγορία Sad έχει προβλεφθεί σωστά 36 φορές, ενώ έχει προβλεφθεί λανθασμένα ως Calm μία φορά και ως Happy 2 φορές.

## 7.2. Ευρήματα

Βάσει της των γραφημάτων που δημιουργήσαμε με τη χρήση των δεδομένων στο κεφάλαιο 6, μπορούμε να καταλήξουμε στα παρακάτω συμπεράσματα.

- ✓ Τα μουσικά κομμάτια που διαθέτουν υψηλή ενέργεια (energy) φαίνεται να επηρεάζουν θετικά την ψυχική κατάσταση του ακροατηρίου, προκαλώντας αίσθημα ενεργητικότητας (energetic) και χαράς (happy).

- ✓ Σύμφωνα με το γράφημα που βασίζεται στον δείκτη ακουστικότητα (acousticness) των μουσικών κομματιών, τα τραγούδια με υψηλότερο acousticness φαίνεται να διαθέτουν χαρακτηριστικά που τα καθιστούν πιο συγκινητικά (sad) και ήρεμα (calm).
- ✓ Τα κομμάτια με υψηλό βαθμό αρμονικής έντασης (valance) εμφανίζουν τάση προς την αίσθηση της χαράς (happy) και της δυναμικότητας (energetic), ενώ ταυτόχρονα μπορεί να προκαλούν και συναίσθημα μελαγχολίας (sad).

### 7.3. Συμπέρασμα

Η αναγνώριση συναισθημάτων στη μουσική είναι ένας πολύ ενδιαφέρων και επίκαιρος τομέας έρευνας. Χρησιμοποιώντας τεχνητή νοημοσύνη και μηχανική μάθηση, μπορούν να αναπτυχθούν αλγόριθμοι που μπορούν να αναγνωρίζουν τα συναισθήματα που εκφράζονται στη μουσική.

Η έρευνα στον τομέα της αναγνώρισης συναισθημάτων από τη μουσική έχει πολλές εφαρμογές σε διάφορους τομείς, όπως η ψυχολογία, η ψυχοθεραπεία, ο κινηματογράφος, το μάρκετινγκ και τα βιντεοπαιχνίδια.

Για να επιτευχθούν καλά αποτελέσματα στην αναγνώριση συναισθημάτων στη μουσική, είναι σημαντικό να υπάρχει ένα μεγάλο και καλά δεικτοδοτημένο σύνολο δεδομένων και να χρησιμοποιούνται κατάλληλες μέθοδοι επεξεργασίας σήματος και αλγόριθμοι μηχανικής μάθησης. Ωστόσο, η βελτίωση της ακρίβειας της αναγνώρισης συναισθημάτων στη μουσική παραμένει μια πρόκληση. Οι ερευνητές πρέπει να αναπτύξουν μοντέλα μηχανικής μάθησης και να συλλέξουν δεδομένα που καλύπτουν ένα ευρύ φάσμα ανθρώπινων συναισθημάτων, λαμβάνοντας παράλληλα υπόψη πολυπλοκότητες όπως οι ατομικές διαφορές και η πολυσημία στη μουσική.

## Κεφάλαιο 8: Βιβλιογραφία

- [1] A. Barr, E. Feigenbaum, and C. Roads, “The Handbook of Artificial Intelligence, Volume 1,” *Computer Music Journal*, vol. 6, no. 3, 1982, doi: 10.2307/3680201.
- [2] I. H. Witten, E. Frank, M. A. Hall, and C. J. Pal, *Data Mining: Practical Machine Learning Tools and Techniques*. 2016. doi: 10.1016/c2009-0-19715-5.
- [3] T. Mitchell, *Machine Learning*. New York: McGraw Hill. 1997.
- [4] G. Brewka, “Artificial intelligence—a modern approach by Stuart Russell and Peter Norvig, Prentice Hall. Series in Artificial Intelligence, Englewood Cliffs, NJ.,” *Knowl Eng Rev*, vol. 11, no. 1, 1996, doi: 10.1017/s0269888900007724.
- [5] A. B. Goldberg, X. Zhu, A. Singh, Z. Xu, and R. Nowak, “Multi-manifold semi-supervised learning,” in *Journal of Machine Learning Research*, 2009, vol. 5.
- [6] S. B. Kotsiantis, I. D. Zaharakis, and P. E. Pintelas, “Supervised Machine Learning : A Review of Classification Techniques General Issues of Supervised Learning Algorithms,” *Informatica (Ljubljana)*, vol. 31, 2007.
- [7] Z. R. Yang and Z. Yang, “Artificial Neural Networks,” *Comprehensive Biomedical Physics*, vol. 6, pp. 1–17, Jul. 2014, doi: 10.1016/B978-0-444-53632-7.01101-1.
- [8] M. Schedl, E. Gómez, and J. Urbano, “Music information retrieval: Recent developments and applications,” *Foundations and Trends in Information Retrieval*, vol. 8, no. 2–3. 2014. doi: 10.1561/15000000042.
- [9] S. Hizlisoy, S. Yildirim, and Z. Tufekci, “Music emotion recognition using convolutional long short term memory deep neural networks,” *Engineering Science and Technology, an International Journal*, vol. 24, no. 3, 2021, doi: 10.1016/j.jestch.2020.10.009.
- [10] Y.-H. Yang and H. H. Chen, *Music Emotion Recognition*. 2011. doi: 10.1201/b10731.
- [11] P. Gilbert, “The Biopsychology of Mood and Arousal. Robert E. Thayer ,” *Q Rev Biol*, vol. 67, no. 3, 1992, doi: 10.1086/417761.
- [12] G. Tzanetakis, G. Essl, and P. Cook, “Automatic Musical Genre Classification Of Audio Signals.”
- [13] E.-S. Jee, Y. J. Cheong, C. H. Kim, D.-S. Kwon, and H. Kobayashi, “Sound Production for the Emotional Expression of Socially Interactive Robots,” 2009. doi: 10.5772/6838.
- [14] J. A. Russell, “A circumplex model of affect,” *J Pers Soc Psychol*, vol. 39, no. 6, 1980, doi: 10.1037/h0077714.
- [15] D. Han, Y. Kong, J. Han, and G. Wang, “A survey of music emotion recognition,” *Front Comput Sci*, vol. 16, no. 6, p. 166335, 2022, doi: 10.1007/s11704-021-0569-4.
- [16] W. K. Ngai, H. Xie, D. Zou, and K. L. Chou, “Emotion recognition based on convolutional neural networks and heterogeneous bio-signal data sources,” *Information Fusion*, vol. 77, 2022, doi: 10.1016/j.inffus.2021.07.007.

- [17] R. Panda, R. Malheiro, B. Rocha, A. P. Oliveira, and R. P. Paiva, "Multi-Modal Music Emotion Recognition: A New Dataset, Methodology and Comparative Analysis," in *10th International Symposium on Computer Music Multidisciplinary Research – CMMR 2013*, 2013.
- [18] E. Cambria, "Affective Computing and Sentiment Analysis," *IEEE Intell Syst*, vol. 31, no. 2, 2016, doi: 10.1109/MIS.2016.31.
- [19] P. Helmholtz, M. Meyer, and S. Robra-Bissantz, "Feel the moosic: Emotion-based music selection and recommendation," in *32nd Bled eConference Humanizing Technology for a Sustainable Society, BLED 2019 - Conference Proceedings*, 2020. doi: 10.18690/978-961-286-280-0.11.
- [20] C. Lv, S. Li, and L. Huang, "Music Emotions Recognition Based on Feature Analysis," in *Proceedings - 2018 11th International Congress on Image and Signal Processing, BioMedical Engineering and Informatics, CISP-BMEI 2018*, 2019. doi: 10.1109/CISP-BMEI.2018.8633223.