

**ΟΝΟΜΑΤΕΠΩΝΥΜΟ** Βασιλική Στάμου

**ΑΜ** 1059543

**ΕΤΟΣ ΦΟΙΤΗΣΗΣ** 1<sup>ο</sup> – ΣΜΗΝ

# Ψηφιακή Τεχνολογία Ήχου

## Ανάκτηση Μουσικής Πληροφορίας

### *Περίληψη*

Η Ανάκτηση Μουσικής Πληροφορίας (MIR) αναδύεται ως ένα δυναμικό πεδίο, που ενσωματώνει στοιχεία της επιστήμης των υπολογιστών, της μηχανικής μάθησης και της ανάλυσης μουσικής. Σκοπός της είναι η ανάπτυξη και βελτίωση τεχνικών που διευκολύνουν τη συστηματική κατηγοριοποίηση και αναγνώριση μουσικών δεδομένων. Ένας βασικός τομέας αυτού του πεδίου είναι η ταξινόμηση μουσικών ειδών, η οποία επιδιώκει να κατατάξει μουσικά κομμάτια σε διαφορετικές κατηγορίες, με βάση χαρακτηριστικά.

Για την επίτευξη αυτής της ταξινόμησης, απαιτούνται προηγμένες τεχνικές επιβλεπόμενης μηχανικής μάθησης οι οποίες περιλαμβάνουν τόσο κλασικούς αλγορίθμους όσο και σύγχρονα νευρωνικά δίκτυα. Στην παρούσα μελέτη, η διαδικασία της ταξινόμησης μουσικών ειδών εφαρμόζεται χρησιμοποιώντας δεδομένα από το dataset FMA. Ο στόχος είναι η κατηγοριοποίηση 580 μουσικών κομματιών σε τέσσερα διακριτά είδη, αξιοποιώντας τη διαδικασία εξαγωγής και επεξεργασίας συγκεκριμένων χαρακτηριστικών των κομματιών.

Για τη διαδικασία κατηγοριοποίησης, χρησιμοποιούνται τρεις βασικές μέθοδοι επιβλεπόμενης μάθησης: SVM, k-NN και Decision Tree καθώς το MLP νευρωνικό δίκτυο. Κάθε μέθοδος εκπαιδεύεται και αξιολογείται βάσει μετρικών απόδοσης, η ανάλυση των αποτελεσμάτων αποδεικνύει ότι οι μέθοδοι SVM και k-NN προσφέρουν την καλύτερη απόδοση, με ποσοστό σωστής ταξινόμησης που φτάνει το 57%. Αυτά τα ευρήματα επιβεβαιώνουν τη σημασία της επιλογής κατάλληλων τεχνικών κατηγοριοποίησης και την αποτελεσματικότητα των μοντέλων στη συγκεκριμένη εφαρμογή της μουσικής ταξινόμησης.

## I. ΕΙΣΑΓΩΓΗ

Η Ανάκτηση Μουσικής Πληροφορίας (MIR) είναι ένα πολύπλευρο πεδίο που ενσωματώνει την επιστήμη των υπολογιστών, τη μηχανική μάθηση και την ανάλυση μουσικής για τη δημιουργία εργαλείων και τεχνικών που διευκολύνουν την επεξεργασία και κατηγοριοποίηση μουσικών δεδομένων. Ένα βασικό κομμάτι του τομέα αυτού είναι η ταξινόμηση μουσικών ειδών, δηλαδή η κατηγοριοποίηση μουσικών κομματιών ανάλογα με τα είδη τους, με εφαρμογές τόσο στη μουσική βιομηχανία όσο και στην έρευνα. Η διαδικασία αυτή είναι απαραίτητη για τη δημιουργία προτάσεων σε πλατφόρμες streaming, τη διαχείριση μουσικών συλλογών και την ανάλυση από ερευνητές. Περιλαμβάνει την εξαγωγή χαρακτηριστικών από τα μουσικά κομμάτια, όπως ρυθμικών, μελωδικών ή φασματικών στοιχείων, και τη χρήση αλγορίθμων μηχανικής μάθησης για την ταξινόμηση.

Η ανάλυση συχνά απαιτεί κάποια συνοπτική παρουσίαση, και για τη μουσική (όπως και για πολλές άλλες μορφές δεδομένων) αυτό επιτυγχάνεται μέσω της εξαγωγής χαρακτηριστικών, ειδικά όταν αναλύεται το ίδιο το ηχητικό περιεχόμενο και πρόκειται να εφαρμοστεί μηχανική μάθηση. Ο σκοπός είναι η μείωση του τεράστιου όγκου των δεδομένων σε ένα διαχειρίσιμο σύνολο τιμών, ώστε η εκπαίδευση να μπορεί να πραγματοποιηθεί εντός ενός εύλογου χρονικού πλαισίου [1].

Η χρήση σύγχρονων τεχνικών όπως τα νευρωνικά δίκτυα (ANN, CNN, RNN) και οι κλασικοί αλγόριθμοι (SVM, k-NN, Decision Trees) έχουν βελτιώσει σημαντικά την ακρίβεια αυτών των συστημάτων. Οι τεχνικές επεξεργασίας σήματος και η εξαγωγή χαρακτηριστικών είναι κρίσιμες για την επιτυχία, ενώ προκλήσεις όπως η ποιότητα των δεδομένων εκπαίδευσης και ο χειρισμός του θορύβου αποτελούν κεντρικά ζητήματα. Τέλος, η σωστή σήμανση των δεδομένων και η αποφυγή overfitting είναι καίρια για την ευρωστία των μοντέλων σε διαφορετικά μουσικά είδη.

## II. ΑΝΑΠΤΥΞΗ ΘΕΜΑΤΟΣ

### A. *Σύνολο Δεδομένων*

Στην παρούσα εργασία, για την κατηγοριοποίηση μουσικών κομματιών ανάλογα με το είδος τους, χρησιμοποιήθηκαν δεδομένα από το σύνολο δεδομένων Free Music Archive. Το FMA είναι ένα από τα πιο αναγνωρισμένα και συχνά χρησιμοποιούμενα σύνολα δεδομένων στον τομέα της ανάκτησης μουσικής πληροφορίας, και έχει σχεδιαστεί για την υποστήριξη ερευνητικών εργασιών και την ανάπτυξη αλγορίθμων που επικεντρώνονται στην ανάλυση και ταξινόμηση μουσικής. Το πλήρες FMA dataset περιέχει 106.574 μουσικά κομμάτια, διασπασμένα σε 16 διαφορετικά μουσικά είδη και κάθε κομμάτι συνοδεύεται από metadata τα οποία μπορούν να αξιοποιηθούν για την εκπαίδευση αλγορίθμων ταξινόμησης [2].

Στη συγκεκριμένη εργασία, χρησιμοποιήθηκε το dataset FMA small, το οποίο περιλαμβάνει 8.000 κομμάτια με διάρκεια περίπου 30 δευτερολέπτων το καθένα, ταξινομημένα σε 8 διαφορετικά μουσικά είδη. Επικεντρωθήκαμε σε τέσσερα από αυτά τα είδη (Pop, Electronic, Rock, Hip-Hop) και επιλέχθηκαν 154 κομμάτια από κάθε είδος.

## B. Επιλεγμένα Χαρακτηριστικά

Ένα από τα σημαντικότερα ζητήματα που αντιμετωπίζονται σε αυτή την έρευνα είναι η επιλογή και η εξαγωγή κατάλληλων χαρακτηριστικών από τα μουσικά αρχεία, που θα χρησιμοποιηθούν στην εκπαίδευση των μοντέλων. Στη συνέχεια παρατίθεται μια περιγραφή των κύριων χαρακτηριστικών που επιλέχθηκαν, βασισμένα σε βιβλιογραφικές αναφορές [3], [4], [5].

- **MFCCs (Mel-frequency cepstral coefficients):** Τα MFCCs είναι χαρακτηριστικά που αναπαριστούν τη μορφή του φάσματος ενός ήχου, λαμβάνοντας υπόψη την ανθρώπινη αντίληψη της συχνότητας. Χρησιμοποιούνται ευρέως για την αναγνώριση μουσικών ειδών και την επεξεργασία ομιλίας.
- **Chroma:** Το χαρακτηριστικό Chroma αναλύει τη διανομή ενέργειας σε 12 ημιτόνια της κλίμακας, αντιπροσωπεύοντας έτσι τις αρμονικές ιδιότητες της μουσικής. Είναι χρήσιμο για την ανίχνευση συγχορδιών και την ανάλυση μελωδίας.
- **Spectral Contrast:** Αυτό το χαρακτηριστικό μετρά τη διαφορά ενέργειας μεταξύ των κορυφών και των κοιλάδων του φάσματος συχνοτήτων, προσφέροντας πληροφορία για την αντίθεση μεταξύ διαφορετικών συχνοτικών περιοχών. Συμβάλλει στην ανάλυση της υφής του ήχου.
- **Tonnetz:** Το Tonnetz αναπαριστά τις τονικές σχέσεις μεταξύ των συχνοτήτων, βοηθώντας στην ανίχνευση της αρμονικής δομής και των διαστημάτων σε ένα μουσικό κομμάτι.
- **Tempo:** Το Tempo αναφέρεται στην ταχύτητα ή στον ρυθμό της μουσικής, μετρώντας τα beats per minute. Είναι χρήσιμο για την κατηγοριοποίηση μουσικής βάσει ρυθμικών προτύπων.
- **Spectral Centroid:** Αυτό το χαρακτηριστικό υποδεικνύει το "κέντρο βάρους" του φάσματος, παρέχοντας μια αίσθηση της φωτεινότητας ή της οξύτητας του ήχου. Χρησιμοποιείται συχνά για τη διάκριση μεταξύ διαφορετικών ειδών μουσικής.

### Τεχνικά Χαρακτηριστικά Υλοποίησης

#### Εξαγωγή χαρακτηριστικών από μουσικά κομμάτια:

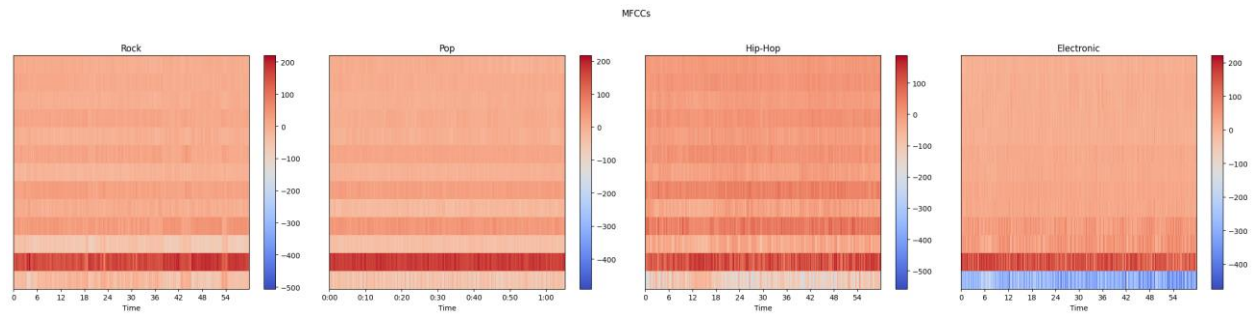
- Με την συνάρτηση `extract_features()` εξάγουμε διάφορα χαρακτηριστικά από ένα μουσικό κομμάτι.
- Χρησιμοποιείται η βιβλιοθήκη **Librosa** [6] για να φορτώσει το αρχείο ήχου και να εξάγει βασικά χαρακτηριστικά από διάρκεια 30 δευτερολέπτων.

Κάθε χαρακτηριστικό υπολογίζεται για όλη τη διάρκεια των 30 δευτερολέπτων, και υπολογίζεται η μέση τιμή για απλοποίηση των δεδομένων. Το tempo είναι βαθμωτός, και όλα τα χαρακτηριστικά συνδυάζονται σε ένα μακρύ διάνυσμα χαρακτηριστικών με τη χρήση της `np.hstack()`.

#### Επεξεργασία πολλαπλών αρχείων ήχου:

- Στη συνέχεια του κώδικα, γίνεται επανάληψη σε έναν κατάλογο (`fma_small_filtered`) που περιέχει αρχεία MP3. Για κάθε αρχείο ήχου:
- Το αρχείο περνάει στη συνάρτηση `extract_features()`, που υπολογίζει τα χαρακτηριστικά του.
- Εάν η εξαγωγή είναι επιτυχής, τα χαρακτηριστικά προστίθενται στη λίστα `features_list`, η οποία αποθηκεύει τα χαρακτηριστικά όλων των κομματιών.
- Το αναγνωριστικό (ID) του κομματιού εξάγεται από το όνομα του αρχείου (το μέρος πριν από την επέκταση `.mp3`), και χρησιμοποιείται για να βρεθεί το είδος από το DataFrame `genre_info`.
- Η ετικέτα είδους του κομματιού προστίθεται στη λίστα `labels_list`.

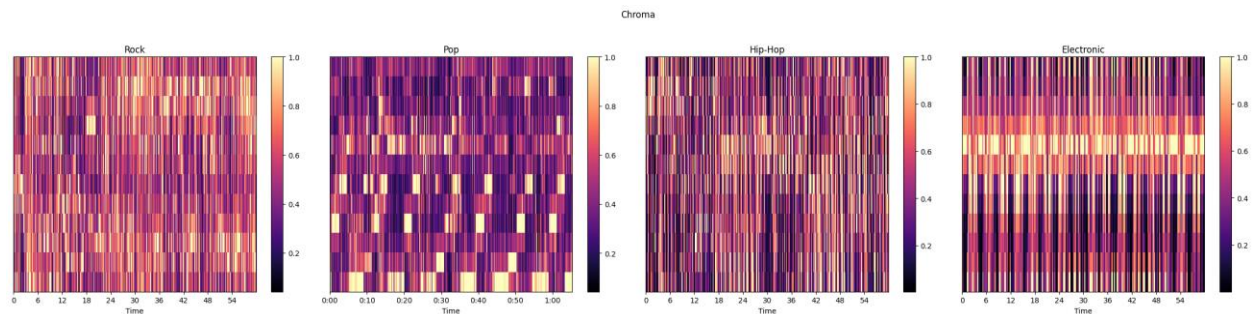
Αυτό δημιουργεί τη βάση για εκπαίδευση μοντέλων μηχανικής μάθησης όπου τα **εξαγόμενα χαρακτηριστικά** θα χρησιμοποιηθούν ως **είσοδος** και οι **ετικέτες ειδών** ως **στόχος** για την ταξινόμηση.



**Σχήμα 1.** MFCC για τέσσερα κομμάτια, ένα από κάθε είδος.

#### MFCCs

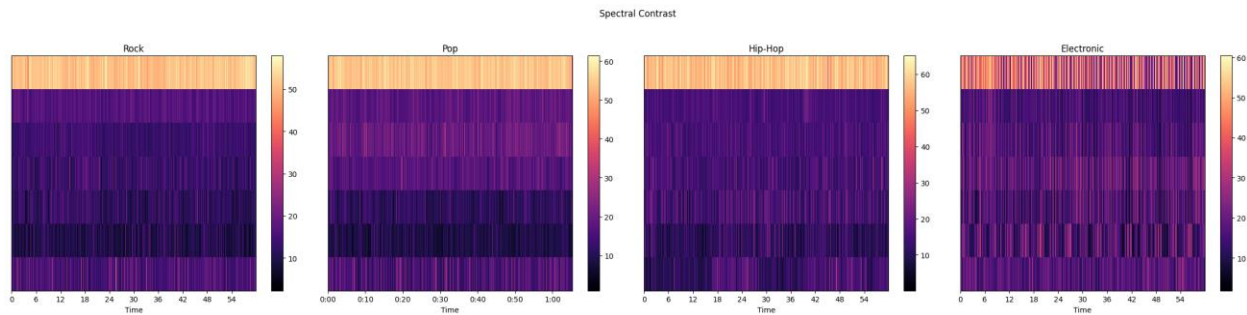
- **Rock:** Εμφανίζει σχετικά ομοιόμορφα MFCCs με μικρή διαφοροποίηση στις χαμηλές και μεσαίες συχνότητες.
- **Pop:** Παρόμοια κατανομή με το Rock, με λιγότερες διαφοροποιήσεις στις υψηλές συχνότητες.
- **Hip-Hop:** Δείχνει έντονες μεταβολές στις χαμηλές και μεσαίες συχνότητες, με λιγότερη πυκνότητα πληροφορίας.
- **Electronic:** Πολύ σταθερή κατανομή, με ελάχιστες αλλαγές στις συχνότητες σε σύγκριση με τα άλλα είδη.



**Σχήμα 2.** Chroma για τέσσερα κομμάτια, ένα από κάθε είδος.

#### Chroma

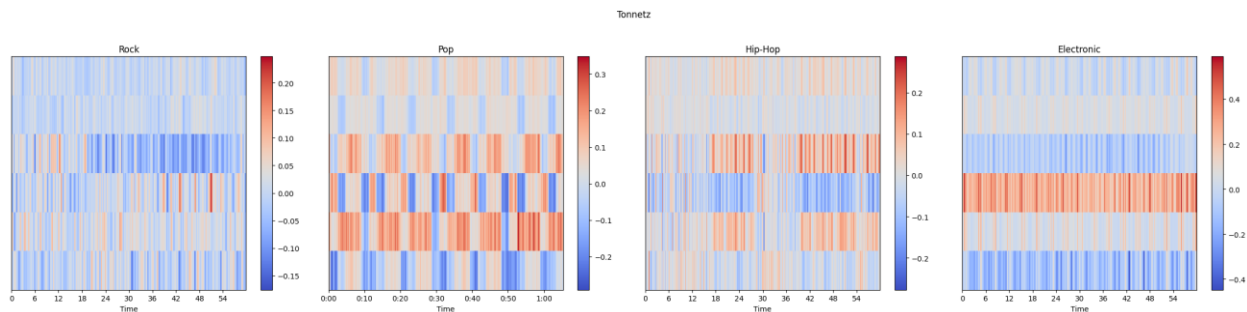
- **Rock:** Εμφανίζει μεγάλη διακύμανση και πολύχρωμες περιοχές, που σημαίνει διαφοροποίηση στις αρμονικές δομές.
- **Pop:** Σημαντικές διαφοροποιήσεις στον chroma δείκτη, αλλά πιο τακτική κατανομή σε σύγκριση με το rock.
- **Hip-Hop:** Πολύ ανομοιογενές μοτίβο, με έντονες μεταβολές, πράγμα που υποδηλώνει διαφοροποιημένη αρμονική δομή.
- **Electronic:** Πιο ομοιόμορφη κατανομή, με περιοχές σταθερότητας, που δείχνει αρμονική συνέπεια στο κομμάτι.



Σχήμα 3. Spectral Contrast για τέσσερα κομμάτια, ένα από κάθε είδος.

#### Spectral Contrast

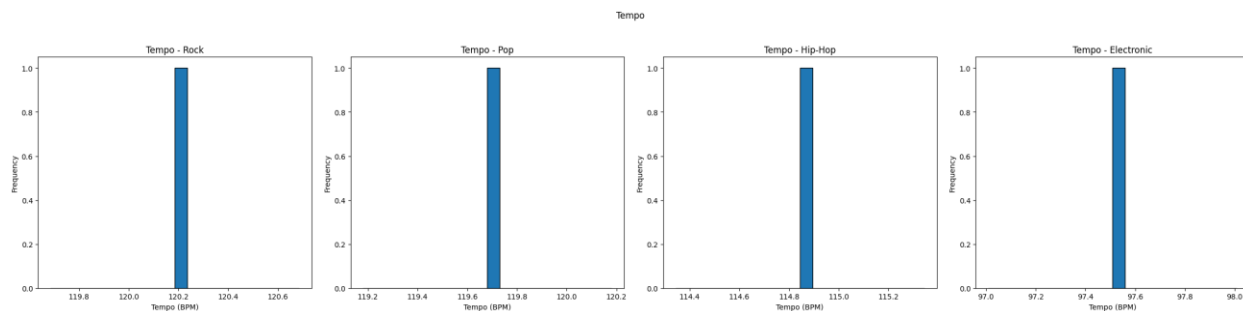
- **Rock:** Μεταβολές στον spectral contrast σε όλες τις συχνότητες, με χαμηλότερη ένταση στις μεσαίες και υψηλές συχνότητες.
- **Pop:** Εμφανίζει λιγότερες μεταβολές, αλλά οι διαφορές είναι πιο εμφανείς στις μεσαίες συχνότητες.
- **Hip-Hop:** Έντονες διακυμάνσεις στις χαμηλές και μεσαίες συχνότητες, λιγότερη ένταση στις υψηλές.
- **Electronic:** Πολύ πιο ομοιογενής κατανομή, με λιγότερες διαφορές στις συχνότητες, γεγονός που δείχνει μια πιο σταθερή δομή.



Σχήμα 4. Tonnetz για τέσσερα κομμάτια, ένα από κάθε είδος.

#### Tonnetz

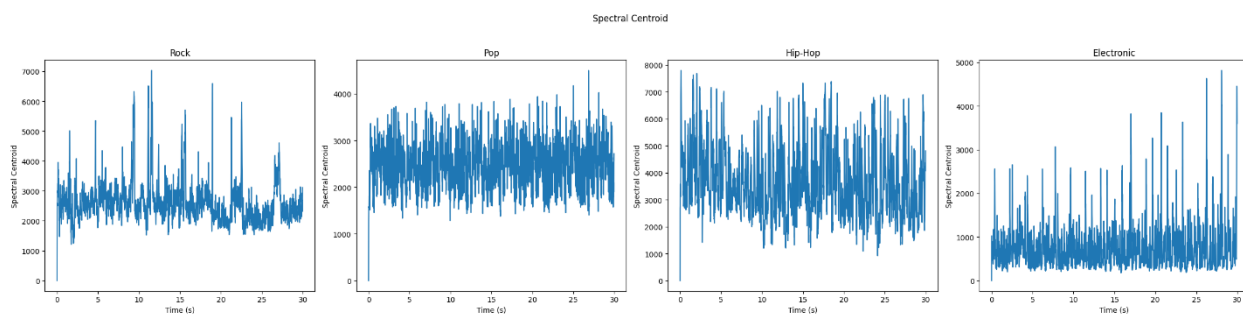
- **Rock:** Το Tonnetz για το rock εμφανίζει περιοδικές διακυμάνσεις, με εναλλαγές μεταξύ θετικών και αρνητικών τιμών. Αυτό δείχνει μεταβολές στην αρμονική δομή, αλλά με έναν σχετικά σταθερό ρυθμό.
- **Pop:** Εμφανίζει πιο τακτική και περιοδική δομή σε σύγκριση με το rock, με πιο ομοιόμορφες και συμμετρικές εναλλαγές μεταξύ θετικών και αρνητικών τιμών. Η αρμονία είναι πιο σταθερή.
- **Hip-Hop:** Το μοτίβο στο tonnetz εδώ είναι πιο διαφοροποιημένο, με μεγαλύτερη ποικιλία και λιγότερη τακτικότητα στις εναλλαγές. Αυτό δείχνει ότι το κομμάτι μπορεί να περιέχει πιο σύνθετες αρμονικές μεταβάσεις.
- **Electronic:** Το tonnetz είναι πιο ομοιογενές και έχει σταθερές εναλλαγές χωρίς ιδιαίτερες διακυμάνσεις. Αυτό υποδηλώνει μια πιο περιορισμένη αρμονική ποικιλία και μεγαλύτερη συνοχή σε σχέση με τα άλλα είδη.



**Σχήμα 5.** Tempo για τέσσερα κομμάτια, ένα από κάθε είδος.

#### Tempo

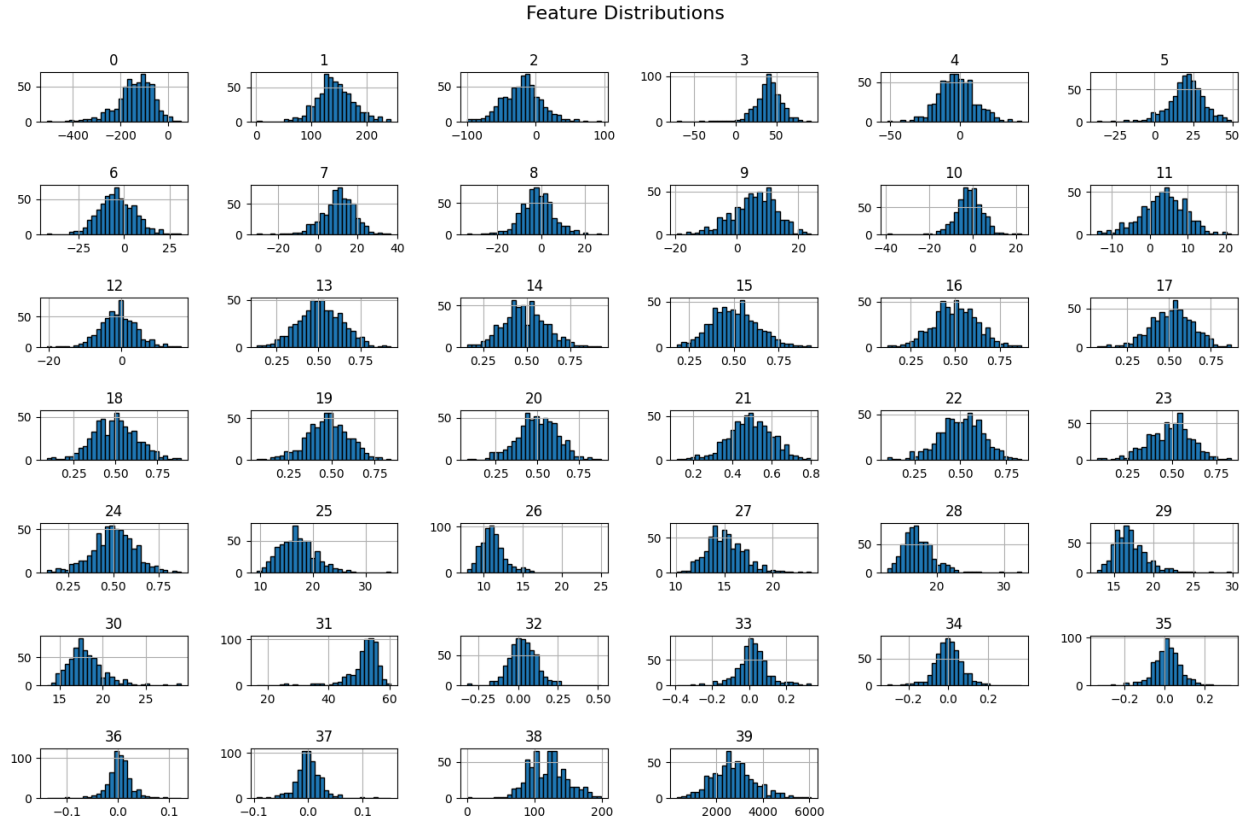
- **Rock:** Σταθερό tempo γύρω στα 102.4 BPM.
- **Pop:** Επίσης σταθερό tempo στα 119.7 BPM.
- **Hip-Hop:** Εμφανίζει σταθερό tempo στα 114.9 BPM.
- **Electronic:** Το tempo είναι σταθερό στα 97.5 BPM, χαμηλότερο από τα άλλα είδη.



**Σχήμα 6.** Spectral Centroid για τέσσερα κομμάτια, ένα από κάθε είδος.

#### Spectral Centroid

- **Rock:** Το spectral centroid στο Rock εμφανίζει μεγάλες διακυμάνσεις, με αρκετές κορυφώσεις και πτώσεις σε διαφορετικά χρονικά σημεία. Αυτό δείχνει μεγάλη ποικιλία στο ηχόχρωμα, με εναλλαγές ανάμεσα σε πιο "αιχμηρά" και "απαλά" μέρη.
- **Pop:** Το spectral centroid στο Pop είναι πιο συγκεντρωμένο γύρω από συγκεκριμένες τιμές, με μικρότερες διακυμάνσεις σε σύγκριση με το Rock. Παρ' όλα αυτά, παραμένει αρκετά δυναμικό, γεγονός που δείχνει εναλλαγές στις φωτεινότητες των συχνοτήτων, αν και όχι τόσο έντονες όσο στο Rock.
- **Hip-Hop:** Εμφανίζει επίσης πολλές διακυμάνσεις, αλλά οι τιμές είναι πιο σταθερές και χαμηλότερες σε σχέση με το Rock, υποδηλώνοντας έναν πιο σταθερό και βαθύ ήχο χωρίς πολλές υψηλές συχνότητες.
- **Electronic:** Το Spectral Centroid στο Electronic δείχνει πιο σταθερές και χαμηλές τιμές, με μικρές αυξομειώσεις. Αυτό υποδηλώνει έναν πιο σταθερό ήχο που επικεντρώνεται περισσότερο στις χαμηλές συχνότητες, με λιγότερες αιχμηρές μεταβάσεις.

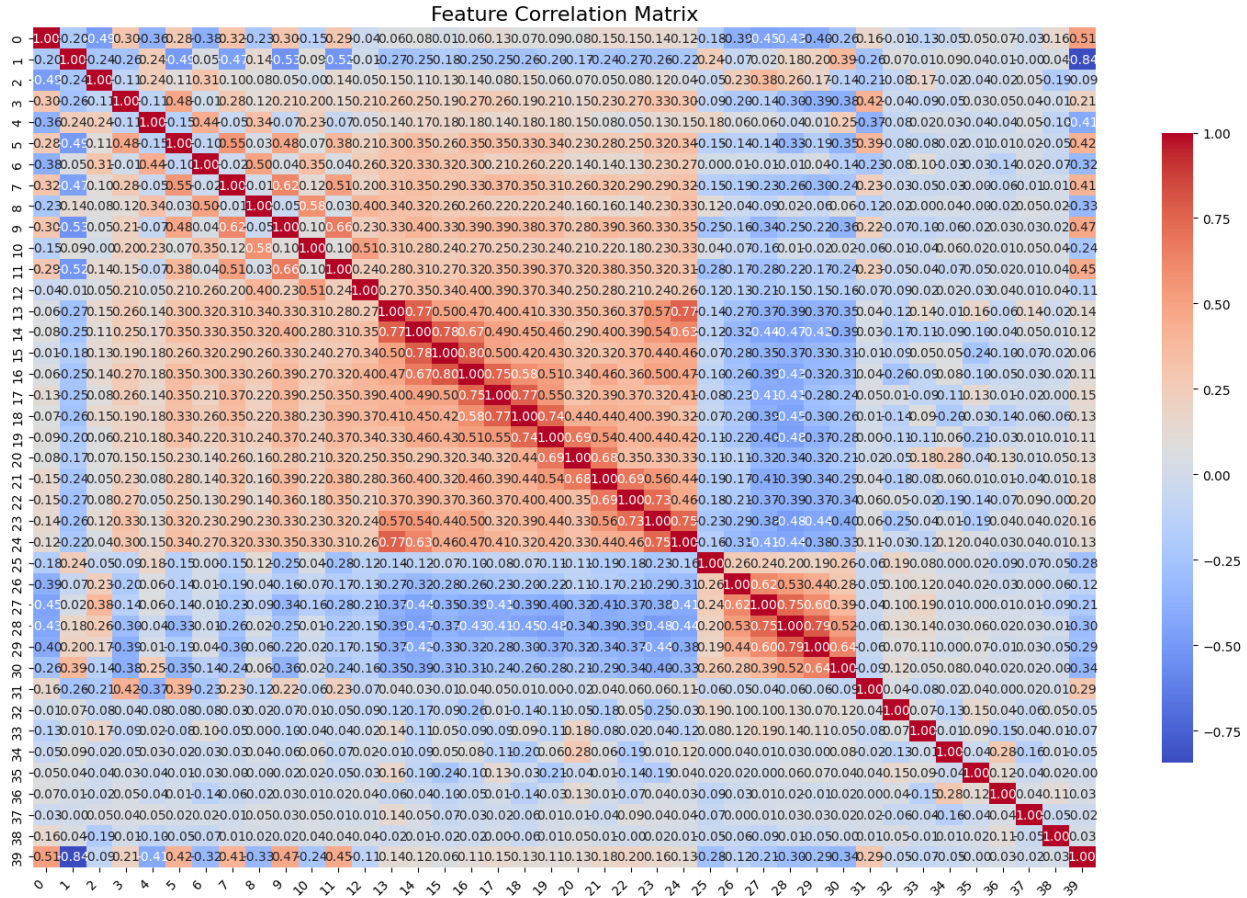


**Σχήμα 7.** Κατανομή Χαρακτηριστικών για κάθε ένα από τα 40 χαρακτηριστικά.

Παρατηρούμε ότι πολλά από τα χαρακτηριστικά ακολουθούν **κανονική κατανομή**, κάτι που υποδηλώνει ότι οι περισσότερες τιμές συγκεντρώνονται γύρω από μια κεντρική τιμή με σταδιακή μείωση στα άκρα.

Υπάρχουν όμως και χαρακτηριστικά που έχουν **ασύμμετρη κατανομή** ή έντονες κορυφώσεις, γεγονός που μπορεί να υποδηλώνει ακραίες τιμές ή εξαρτήσεις από συγκεκριμένα είδη μουσικής.





Σχήμα 8. Μητρώο Συσχέτισης Χαρακτηριστικών για κάθε ένα από τα 40 χαρακτηριστικά.

Το μητρώο συσχέτισης δείχνει τον βαθμό συσχέτισης ανάμεσα στα χαρακτηριστικά, χρησιμοποιώντας την τιμή του συντελεστή συσχέτισης Pearson, η οποία κυμαίνεται από -1 (αρνητική συσχέτιση) έως +1 (θετική συσχέτιση). Οι κόκκινες περιοχές υποδηλώνουν ισχυρή θετική συσχέτιση μεταξύ χαρακτηριστικών, ενώ οι μπλε περιοχές υποδηλώνουν αρνητική συσχέτιση.

### C. Προεπεξεργασία Δεδομένων

Η προεπεξεργασία δεδομένων αποτελεί ένα κρίσιμο στάδιο στην ανάπτυξη μοντέλων μηχανικής μάθησης, καθώς διασφαλίζει την ποιότητα και την καταλληλότητα των δεδομένων για την εκπαίδευση των αλγορίθμων. Τα δεδομένα σπάνια είναι έτοιμα να χρησιμοποιηθούν απευθείας, καθώς συχνά περιέχουν ανωμαλίες όπως ελλιπή ή θορυβώδη στοιχεία, διαφορετικές κλίμακες μεταξύ των χαρακτηριστικών, ή ανισορροπία στις κατηγορίες. Μέσω της προεπεξεργασίας, βελτιώνονται η ακρίβεια και η αποδοτικότητα των μοντέλων, επιτρέποντας την ομαλότερη εκπαίδευσή τους και την καλύτερη γενίκευση των αποτελεσμάτων. Βασικές τεχνικές περιλαμβάνουν τον χειρισμό των ελλειπόντων δεδομένων, την κλιμάκωση των χαρακτηριστικών, την κωδικοποίηση των κατηγορικών δεδομένων, καθώς και την αντιμετώπιση της ανισορροπίας των κατηγοριών μέσω τεχνικών υπερδειγματοληψίας. Η σωστή προεπεξεργασία είναι καθοριστική για τη βελτίωση της απόδοσης και της σταθερότητας των μοντέλων ταξινόμησης.



- Εφαρμόστηκε **κανονικοποίηση (feature scaling)** μέσω της τεχνικής StandardScaler ώστε τα χαρακτηριστικά να αποκτήσουν μηδενική μέση τιμή και τυπική απόκλιση 1, διευκολύνοντας τη σωστή λειτουργία των αλγορίθμων ταξινόμησης.
- Οι **ετικέτες κατηγοριών μουσικών ειδών κωδικοποιήθηκαν (label encoding)** σε ακέραιους αριθμούς χρησιμοποιώντας το LabelEncoder για την προετοιμασία των δεδομένων προς εκπαίδευση.
- Για την αντιμετώπιση του προβλήματος της **ανισοκατανομής των κλάσεων**, εφαρμόστηκε η τεχνική **SMOTE** (Synthetic Minority Over-sampling Technique) [7], δημιουργώντας συνθετικά δείγματα για τις υποεκπροσωπούμενες κατηγορίες.
- Τέλος, ελέγχθηκε η νέα κατανομή των κλάσεων, διασφαλίζοντας την ισορροπία μεταξύ των κατηγοριών, και αποφεύγοντας τη μεροληψία των μοντέλων προς τις πολυπληθέστερες κλάσεις.

**Πίνακας I.** Δείγματα ανά κλάση

Κλάσεις	Πλήθος Δειγμάτων	Πλήθος Δειγμάτων μετά από Smote
Hip-Hop	154	154
Pop	154	154
Electronic	143	154
Rock	129	154

#### *D. Μοντέλα Μηχανικής Μάθησης*

Τα μοντέλα που χρησιμοποιήθηκαν για την ταξινόμηση μουσικών ειδών είναι τα ακόλουθα:

- **Support Vector Machine (SVM):** Το SVM είναι ένας αλγόριθμος που στοχεύει να βρει την καλύτερη υπερεπίπεδη διαχωριστική γραμμή (hyperplane) ανάμεσα σε δύο ή περισσότερες κατηγορίες δεδομένων. Είναι κατάλληλο για αυτό το πρόβλημα, καθώς διαχειρίζεται καλά πολυδιάστατα δεδομένα και είναι αποτελεσματικό σε περιπτώσεις με μεγάλες διαστάσεις χαρακτηριστικών, όπως αυτά που προκύπτουν από την ανάλυση ήχου [8].
- **k-Nearest Neighbors (k-NN):** Ο k-NN ταξινομεί τα δείγματα με βάση τα πιο κοντινά τους γειτονικά σημεία στον χώρο χαρακτηριστικών. Εξαρτάται από την απόσταση μεταξύ των χαρακτηριστικών και μπορεί να προσαρμοστεί εύκολα. Κρίθηκε κατάλληλο για το συγκεκριμένο θέμα, επειδή επιτρέπει την ταξινόμηση δεδομένων με σχετικά απλό τρόπο και προσαρμόζεται σε διαφορετικές δομές δεδομένων [9].
- **Decision Tree:** Το Decision Tree δημιουργεί ένα μοντέλο που λαμβάνει αποφάσεις βασισμένες σε μία ιεραρχία χαρακτηριστικών. Είναι εύκολο να κατανοηθεί, ενώ παρέχει ερμηνεύσιμα αποτελέσματα. Είναι κατάλληλο για την κατηγοριοποίηση μουσικών ειδών, καθώς επιτρέπει την ανάλυση σύνθετων σχέσεων μεταξύ χαρακτηριστικών [10].
- **MLP Neural Network (Multilayer Perceptron):** Το MLP είναι ένα είδος τεχνητού νευρωνικού δικτύου που περιλαμβάνει πολλαπλά επίπεδα νευρώνων. Είναι ισχυρό στην ανάλυση μη γραμμικών σχέσεων και στη διαχείριση μεγάλων δεδομένων. Κρίθηκε κατάλληλο λόγω της ικανότητάς του να συλλαμβάνει περίπλοκες σχέσεις στα μουσικά χαρακτηριστικά, κάτι που απαιτείται για την ταξινόμηση σε πολύπλοκα μουσικά δεδομένα [11].

## E. Cross Validation Μοντέλων

Η τεχνική του cross-validation είναι μια σημαντική διαδικασία στην εκπαίδευση μοντέλων μηχανικής μάθησης, που βοηθά στην καλύτερη εκτίμηση της απόδοσης του μοντέλου και την αποφυγή του overfitting. Ο στόχος του cross-validation είναι να αξιολογήσει πώς θα αποδίδει το μοντέλο σε δεδομένα που δεν έχει "δει" κατά την εκπαίδευση, διασφαλίζοντας έτσι την γενίκευση των αποτελεσμάτων [12].

**Διαχωρισμός δεδομένων:** Ο διαχωρισμός του συνόλου δεδομένων γίνεται σε training set και test set. Στην περίπτωση μας, το 70% των δεδομένων χρησιμοποιείται για την training και το 30% για το testing.

**Χρήση διαφορετικών ταξινομητών:** Εξετάζονται διάφορα μοντέλα όπως SVM, k-NN, Decision Tree, και MLP Neural Network. Αυτό δίνει τη δυνατότητα να συγκριθούν οι επιδόσεις διαφορετικών αλγορίθμων.

**Μετρικές αξιολόγησης:** Εκτός από το accuracy, χρησιμοποιούνται και άλλες σημαντικές μετρικές όπως precision, recall και f1-score, οι οποίες λαμβάνουν υπόψη διάφορες πτυχές της απόδοσης του μοντέλου, όπως η ικανότητα του μοντέλου να ανιχνεύει σωστά τις κατηγορίες.

**Cross-validation:** Η διαδικασία του cross-validation γίνεται με τη χρήση του cross\_validate με 5-fold cross-validation (cv=5). Τα δεδομένα του training set χωρίζονται σε 5 μέρη, εκπαιδεύοντας το μοντέλο σε 4 από αυτά και αξιολογώντας το στο 1, και επαναλαμβάνοντας αυτή τη διαδικασία για κάθε υποσύνολο. Αυτό επιτρέπει μια πιο αξιόπιστη εκτίμηση της απόδοσης του μοντέλου.

Με αυτόν τον τρόπο μπορούμε να συγκρίνουμε την απόδοση διαφορετικών μοντέλων σε μια πιο ολοκληρωμένη βάση, λαμβάνοντας υπόψη διάφορες μετρικές και χρησιμοποιώντας όλα τα δεδομένα του training set για εκπαίδευση και αξιολόγηση.

Πίνακας II. Μετρικές Απόδοσης Μοντέλων – Training Set

Μοντέλα	Accuracy	Precision	Recall	F1
SVM	0.57	0.57	0.57	0.56
k-NN	0.52	0.53	0.51	0.50
Decision Tree	0.48	0.48	0.48	0.47
MLP Neural Network	0.55	0.56	0.55	0.55

## F. Testing Μοντέλων

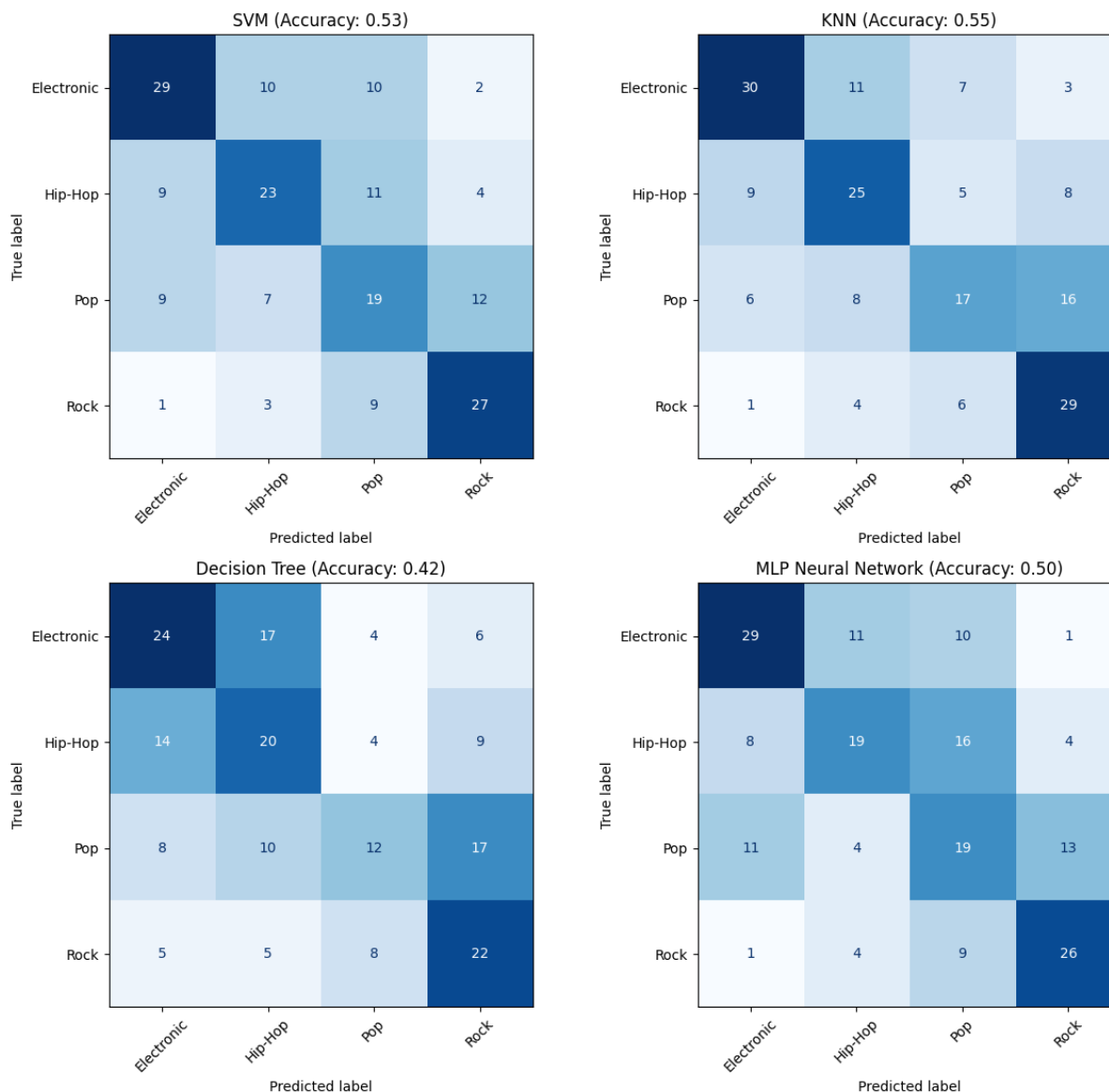
Στο τμήμα του testing, αξιολογούμε την απόδοση των μοντέλων που έχουμε εκπαιδεύσει, χρησιμοποιώντας το test set. Αυτό είναι το σύνολο δεδομένων που δεν έχει χρησιμοποιηθεί κατά την εκπαίδευση, και με τη βοήθειά του μπορούμε να δούμε πόσο καλά μπορεί το μοντέλο να γενικεύσει σε νέα δεδομένα.

**Για κάθε μοντέλο:**

- Το εκπαιδεύουμε με τα δεδομένα του training set.
- Χρησιμοποιούμε το μοντέλο για να κάνουμε προβλέψεις στο test set.
- Υπολογίζουμε το accuracy του μοντέλου και το classification report για κάθε κλάση.
- Υπολογίζεται το confusion matrix για κάθε ταξινομητή, το οποίο δείχνει πόσα παραδείγματα ταξινομήθηκαν σωστά και πόσα λάθος για κάθε κλάση.

Με αυτό τον τρόπο, μπορούμε να συγκρίνουμε την απόδοση κάθε μοντέλου τόσο ποσοτικά όσο και οπτικά.

## Confusion Matrices for Different Classifiers



**Σχήμα 9.** Confusion Matrix για κάθε μοντέλο – Test Set.

Παρατηρούμε ότι σε όλα τα μοντέλα, τα είδη **Rock** και **Electronic** ταξινομούνται με μεγαλύτερη ακρίβεια σε σύγκριση με τις υπόλοιπες κατηγορίες. Αυτό φαίνεται από τις υψηλές τιμές στην αντίστοιχη γραμμή του **confusion matrix**, όπου τα παραδείγματα από αυτές τις κατηγορίες αναγνωρίζονται σωστά με μεγαλύτερη συχνότητα.

Επιπλέον, αξίζει να σημειωθεί ότι το **accuracy** του μοντέλου **k-NN** κατά το testing είναι υψηλότερο από το **SVM**, παρόλο που το **SVM** έδειξε καλύτερα αποτελέσματα κατά το **cross-validation**. Αυτό δείχνει ότι ο k-NN καταφέρνει να αποδώσει καλύτερα σε νέα δεδομένα από ό,τι αναμενόταν βάσει της εκπαίδευσης.

Τέλος, οι τιμές του **accuracy** στο **testing** κινούνται σε παρόμοια επίπεδα με αυτές του **training**, κάτι που υποδηλώνει ότι τα μοντέλα γενικεύουν καλά και δεν παρουσιάζουν φαινόμενα overfitting ή underfitting. Αυτό σημαίνει ότι τα μοντέλα μπορούν να κάνουν ακριβείς προβλέψεις τόσο στα δεδομένα εκπαίδευσης όσο και στα δεδομένα ελέγχου.

**Πίνακας III.** Classification Report SVM – Test Set

<b>Κλάσεις</b>	<b>Precision</b>	<b>Recall</b>	<b>F1</b>	<b>Support</b>
Electronic	0.60	0.57	0.59	51
Hip-Hop	0.53	0.49	0.51	47
Pop	0.39	0.40	0.40	47
Rock	0.60	0.68	0.64	40

**Πίνακας IV** Classification Report k-NN – Test Set

<b>Κλάσεις</b>	<b>Precision</b>	<b>Recall</b>	<b>F1</b>	<b>Support</b>
Electronic	0.65	0.59	0.62	51
Hip-Hop	0.52	0.53	0.53	47
Pop	0.49	0.36	0.41	47
Rock	0.52	0.72	0.60	40

**Πίνακας V** Classification Report Decision Tree – Test Set

<b>Κλάσεις</b>	<b>Precision</b>	<b>Recall</b>	<b>F1</b>	<b>Support</b>
Electronic	0.47	0.47	0.47	51
Hip-Hop	0.38	0.43	0.40	47
Pop	0.43	0.26	0.32	47
Rock	0.41	0.55	0.47	40

**Πίνακας VI** Classification Report MLP – Test Set

<b>Κλάσεις</b>	<b>Precision</b>	<b>Recall</b>	<b>F1</b>	<b>Support</b>
Electronic	0.59	0.57	0.58	51
Hip-Hop	0.50	0.40	0.45	47
Pop	0.35	0.40	0.38	47
Rock	0.59	0.65	0.62	40

### G. *Hyperparameter Tuning*

Στο κομμάτι του αυτό, γίνεται βελτιστοποίηση των παραμέτρων των ταξινομητών, ώστε να επιτευχθεί η καλύτερη δυνατή απόδοση. Με τη μέθοδο του cross-validation, δοκιμάζονται διαφορετικοί συνδυασμοί τιμών για τις υπερπαραμέτρους και αξιολογείται η ακρίβεια του μοντέλου σε κάθε περίπτωση. Παρακάτω αναφέρονται οι βέλτιστες εκτιμήσεις για κάθε ταξινομητή.

**KNN:** neighbors: 15, weights: uniform.

**SVM:** C=10, Gamma=0.00648, kernel= rbf.

**Decision Tree:** MaxDepth=7, MinSamplesSplit=5.

**MLP:** HiddenLayerSizes=(100, 100), Activation=relu, Solver=sgd, LearningRate=constant, Alpha=0.0001.

Ο σκοπός του hyperparameter tuning είναι να αυξηθεί η γενικευσιμότητα των μοντέλων, δηλαδή να επιτυγχάνουν υψηλή ακρίβεια τόσο στο train set όσο και στο test set [13].

## III. ΣΥΝΟΨΗ

Στην παρούσα εργασία, εξετάζεται η κατηγοριοποίηση μουσικών κομματιών από το σύνολο δεδομένων Free Music Archive (FMA), εστιάζοντας σε τέσσερα μουσικά είδη: Pop, Electronic, Rock και Hip-Hop. Από 8.000 κομμάτια, επιλέχθηκαν 154 από κάθε είδος. Η εξαγωγή χαρακτηριστικών περιλαμβάνει MFCCs, Chroma, Spectral Contrast, Tonnetz, Tempo και Spectral Centroid, τα οποία χρησιμοποιούνται για την εκπαίδευση μοντέλων μηχανικής μάθησης. Στη διαδικασία, εφαρμόστηκαν τεχνικές προεπεξεργασίας, όπως κανονικοποίηση και SMOTE για την αντιμετώπιση ανισορροπιών στις κατηγορίες.

Διαφορετικά μοντέλα (SVM, k-NN, Decision Tree, MLP Neural Network) αξιολογήθηκαν μέσω cross-validation, με το SVM να επιτυγχάνει καλύτερα αποτελέσματα σε ακρίβεια (57%) στο training set. Στο testing, ο k-NN πέτυχε την καλύτερη ακρίβεια με 55%, ενώ τα είδη Rock και Electronic ταξινομήθηκαν με μεγαλύτερη ακρίβεια. Τέλος, μέσω hyperparameter tuning, βελτιστοποιήθηκαν οι υπερπαραμέτροι των μοντέλων, ενισχύοντας τη γενικευσιμότητά τους. Τα αποτελέσματα δείχνουν ότι η επιλογή και η επεξεργασία των χαρακτηριστικών είναι κρίσιμες για την επιτυχία της ταξινόμησης μουσικών ειδών.

## IV. ΒΙΒΛΙΟΓΡΑΦΙΑ

- [1] H. EIDENBERGER, "Fundamental media understanding. Norderstedt," 2011.
- [2] M. DEFFERRARD, "FMA: A dataset for music analysis.," no. arXiv preprint arXiv:1612.01840, 2016.
- [3] T. LI and M. OGIHARA, "Toward intelligent music information retrieval," *IEEE Transactions on Multimedia*, vol. 8.3, pp. 564-574, 2006.
- [4] M. ZANONI, "Searching for dominant high-level features for music information retrieval," in *2012 Proceedings of the 20th European Signal Processing Conference (EUSIPCO). IEEE, 2012. p. 2025-2029.*
- [5] J. H. JENSEN, Feature extraction for music information retrieval, 2009.
- [6] B. MCFEE, "librosa: Audio and music signal analysis in python," *SciPy*, pp. 18-24, 2015.
- [7] N. V. CHAWLA, "SMOTE: synthetic minority over-sampling technique," *Journal of artificial intelligence research*, vol. 16, pp. 321-357, 2002.
- [8] M. A. HEARST, "Support vector machines," *IEEE Intelligent Systems and their applications*, vol. 13.4, pp. 18-28, 1998.
- [9] L. E. PETERSON, " K-nearest neighbor," *Scholarpedia*, vol. 4.2: 1883., 2009.
- [10] Y.-Y. SONG and L. U. YING, "Decision tree methods: applications for classification and prediction.," *Shanghai archives of psychiatry*, vol. 27.2: 130, 2015.
- [11] D. W. RUCK, S. K. ROGERS and M. KABRISKY, "Feature selection using a multilayer perceptron.," *ournal of neural network computing*, vol. 2.2, pp. 40-48, 1990.
- [12] D. BERRAR, "Cross-validation," 2019.
- [13] L. YANG and A. SHAMI, "On hyperparameter optimization of machine learning algorithms: Theory and practice," *Neurocomputing*, vol. 415, pp. 295-316, 2020.