May 2021

Mining User Personality from Music Listening Behavior in Online Platforms Using Audio Attributes

Anurata Prabha Hridi
Clemson University, anurata23@gmail.com

MINING USER PERSONALITY FROM MUSIC LISTENING BEHAVIOR IN ONLINE PLATFORMS USING AUDIO ATTRIBUTES

A Thesis
Presented to
the Graduate School of
Clemson University

In Partial Fulfillment
of the Requirements for the Degree
Master of Science
Computer Science

---

by
Anurata Prabha Hridi
May 2021

Accepted by:
Dr. Nina Christine Hubig, Committee Chair
Dr. Bart Knijnenburg
Dr. Taufiquar Khan

Abstract

Music and emotions are inherently intertwined. Humans leave hints of their personality everywhere, and particularly their music listening behavior shows conscious and unconscious diametric tendencies and influences. So, what could be more elegant than finding the underlying character given the attributes of a certain music piece and, as such, identifying the likelihood that music preference is also imprinted or at least resonating with its listener? This thesis focuses on the music audio attributes or the latent song features to determine human personality. Based on unsupervised learning, we cluster several large music datasets using multiple clustering techniques known to us. This analysis led us to classify song genres based on audio attributes, which can be deemed a novel contribution in the intersection of Music Information Retrieval (MIR) and human psychology studies. Existing research found a relationship between Myers-Briggs personality models and music genres. Our goal was to correlate audio attributes with the music genre, which will ultimately help us to determine user personality based on their music listening behavior from online music platforms. This target has been achieved as we showed the users' spectral personality traits from the audio feature values of the songs they listen to online and verified our decision process with the help of a customized Music Recommendation System (MRS). Our model performs genre classification and personality detection with 78% and 74% accuracy, respectively. The results are promising compared to competitor approaches as they are explainable via statistics and visualizations. Furthermore, the RS completes and validates our pursuit through 81.3% accurate song suggestions. We believe the outcome of this thesis will work as an inspiration and assistance for fellow researchers in this arena to come up with more personalized song suggestions. As music preferences will shape specific user personality parameters, it is expected that more such elements will surface that would portray the daily activities of individuals and their underlying mentality.

Dedication

To my parents, who made me the person I am today.

Acknowledgments

First and foremost, I would like to express my immense gratitude to my adviser and committee chair, Dr. Nina Christine Hubig, for putting trust in me. I am grateful to her for her guidance and encouragement to continue the research. Under her supervision, I have become a better learner, and most importantly, a patient researcher. I want to thank Dr. Bart Knijnenburg for his spontaneity in taking so much time out for my thesis preparation and presentation and for always being there for me. I am thankful for all the constructive suggestions I received from him throughout this study. I am grateful to Dr. Taufiquar Khan for joining my committee and assisting me through my master's journey. A vote of thanks goes to my committee for allowing me to defend my thesis in their presence and also for investing time and effort for me.

Second, I am highly indebted to Dr. Eva Zangerle of the University of Innsbruck and her team for their prompt responses regarding music datasets created and shared by them. They went to long lengths to help me with what I needed for the data sources of this study. I also learned relevant important things from them and their research.

Third, I am thankful to my labmates of DZRPT LAB for their continuous support and thoughtful guidelines. I would also like to take this opportunity to thank the faculty and the staff of the School of Computing at Clemson University.

Last but not least, I owe a lot to my parents for believing in me more than myself. I want to express my deepest appreciation to my husband, who helped me in every possible way to complete this journey well. Being a graduate student himself, he was my constant support system, who always stuck with me during all my ups and downs, and inspired me to push through my limits.

Table of Contents

List of Tables

List of Figures

Chapter 1

Introduction

It is possible to learn about an individual's personality by closely observing her/his daily life and regular interaction with different integral mannerisms or/and habits. According to researchers, music preference is one of the most significant facets of everyday life to do so, and modern music recommender systems bank on this subtle relationship [18, 34]. Thanks to the fast advancing technology, our current life has a digital footprint of ourselves everywhere we put our steps in. As a result, the online music streaming and listening platforms and their provision of customization have become part and parcel of the reflection of our personality. There have been studies [69, 39, 44, 75] which identified music to be descriptive of human nature well, and the goal was to understand the relationship between human personality and music preference thoroughly.

There can be multiple faces of an individual's personality, and music has been a trusted element to identify this variation for long [68, 45]. Since a person can possess numerous personality traits simultaneously, the answer is not a straightforward one; instead, it is knitted with a gray texture [59]. Because there are many different types of songs that can be correlated with different kinds of human traits, it is only natural to consider music when it comes to identifying user personality, which goes beyond a mode of entertainment now. In fact, recommender systems that supposedly value users' song preferences drive them towards understanding their personality at a superior level [41]. However, music preference alone cannot determine holistic personality; instead, it can express its relation to relevant personality traits. The personality detection that we do here is based on the types of music that have been registered as the users' playlists, and the subsequent analysis with the recommender system helped us realize the correctness of this process of identifying user personality.

1.1 Motivation

The motivation behind employing music to understand and analyze human personality is mainly because the language of music and its interpretation are universal [20, 36]. Hull explained in his dissertation [39] that there might be positive or negative correlations between different music genres and personality types defined by the Big Five model. However, without any categorical mapping between the contents and the type of music, it cannot be said that music listening behavior can bring out a user's personality. Therefore, our target is to use this underexplored aspect of the research and create a mapping between human personality and track feature values of the songs the individuals listen to digitally. This is only organic that researchers now want to manipulate and use the fact that music defines personality to understand the complexities of human nature entwined with music preferences. This thesis being completely dependent on practical applications, ensures that music reflects the underlying mentality of the individuals. That is why we chose music to look into the future where every music recommendation system can essentially be the emblem of the users themselves as the songs in their playlists would be the direct result of mapping with their personality types.

Existing literature [27, 26] has been exploring methods to relate song genre to human personality; however, not much work has been done to systematically identify the song types from the contents of the tracks themselves. We aim to bridge this gap by showing methods and consequential results for detecting song genres from track feature values. A parallel line of study by Tzanetakis & Cook familiarized us with the detection of song genres from audio signals, where the authors used rhythmic and pitch content features of song excerpts [76]. In nonreal time they classified 61% of the songs of the dataset they used into ten genres, of which 53% were correct. However, they worked not only on music pieces but also on sports announcements and speechy lectures. Moreover, they analyzed the pitch and rhythm of the audio signals, whereas our resources are the acoustic features of songs. Though the inputs are different, their research worked as an inspiration for us to explore this ground. We know the basis of the relationship between personality type and music genre from existing literature [55]. Our target is to establish the connection between music genre and audio attributes. Finally, it will enable us to use audio attributes to do human personality detection. In short, our objective is to identify personality from users' online music listening behavior based on the audio features.

We chose to manipulate the dataset that came with [67] and came up with a full-fledged dataset with real user information and track feature values. While applying unsupervised learning algorithms, we found that the factors causing a song to fall under a particular genre are hugely overlapping. Furthermore, the audio attribute values were fuzzy [76] and abstract enough to convince us to go for subspace clustering [61], where multiple clusters share common spaces. With that thought in mind, we implemented not only K-Means considering this a Gaussian distribution but also several modifications over the baseline via subspace clustering algorithms. While Neuman et al. gave us a vivid idea of the relationship between song genre and user personality [59], we verified our mapping results of song genres from track features using prevalent ground truth (labeled dataset). The dataset also carried real Tweet IDs of the users registered, which we used to get the inherent personality types [31]. We compared our result with it before we went ahead to implement a music recommender system for further verification. It was built based on the personality scores of the users, and the final outputs were the songs suggested for the user having a certain spectrum of personality traits.

1.2 Problem Statement

Our functioning work process can be divided into three steps, which we can find in Figure 1.1, and altogether they create the complete pipeline. The first five sub-steps of the pipeline refer to Pipeline 1, while the next two steps belong to Pipeline 2. The last two sub-steps in Pipeline 3 show the verification process essentially, which validates our effort. To describe:

1. Pipeline 1 shows the steps of identifying genre(s) from track feature values after clustering and classification techniques. This step ends with the validity calculation of cluster assignment to describe particular genres. The idea is to find out how accurately the clusters define the genres of the unlabeled dataset, and the labels refer to their respective clusters. We use this cluster/genre information for the next pipeline.

2. Pipeline 2 is about detecting user personality from the song genre(s) being listened to by mapping the underlying correlation of personality model and song genre. This step points out the users' multifaceted personality traits based on the genre(s) of the songs they listened to. We use Natural Language Processing (NLP) to get the percentage of personality traits from tweets of the associated users and compare our personality results with it.

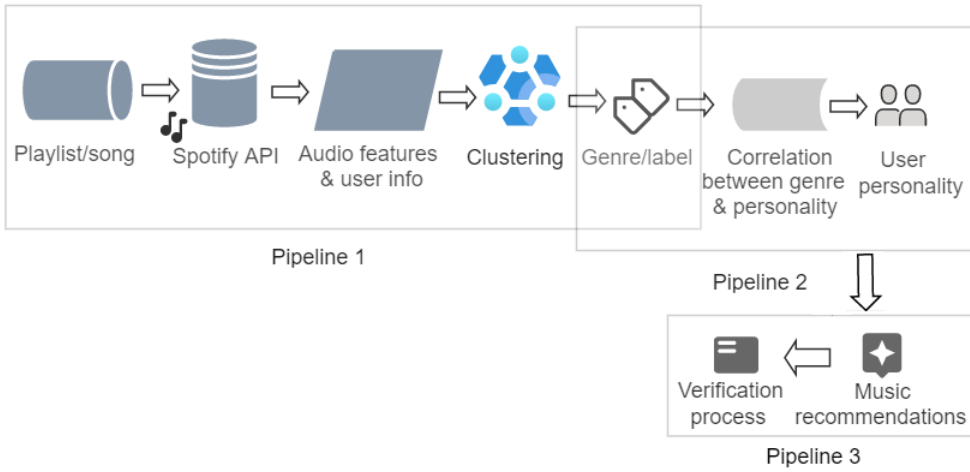


Figure 1.1: The complete pipeline of the workflow. Pipeline 1 works towards identifying the song genre(s) from audio attributes extracted from a user's playlist. Pipeline 2 involves mapping the identified genre(s) with the personality types derived from the 'Big Five factor/Myers-Briggs Personality', leading us to the detection of the personality type(s) of the user under evaluation. The recommender system in Pipeline 3 suggests the songs for the users based on their personality scores. The genre of these suggested songs and the songs typically listened to by the user were compared to verify the personality traits we came up with from the first two steps.

3. Pipeline 3 works towards verifying the experimental results with ground truth via a music recommender system. This step ensures that the results of Pipelines 1 and 2 are reflected in this final pipeline with high accuracy and precision. The personality score of an individual user is given as an input to the system, and the outcome is expected to be the songs that the user prefers listening to. Finally, the system cross-validates the model to show that the recommended songs align with the users' personality types and preferred song genres.

To summarize, the intermediate step of detecting song genres is one of the significant contributions of this research work. We believe this will help the music recommendation platforms make more learned suggestions to their subscribers as this process would allow them to understand the users' different faces of personality, which relate to their preferences in music types.

1.2.1 Research Questions

The research questions of the study sequentially seek to mitigate the research gap between track features and song genres **(RQ1)**, answer the relationship mapping between genre and personality **(RQ2)**, and justify the results via a personality-based music recommender system **(RQ3)**.

They can be articulated as follows:

RQ1 How can we classify songs into genre(s) from their track feature values?

RQ2 What is the best way to map users' personality with the song type(s) they listen to online?

RQ3 How accurately can user personality-specific songs be recommended?

The rest of the chapters in this dissertation are organized as following: Chapter 2 creates the base of the study by briefing about the terms and tools we used, while Chapter 3 extensively studies existing research in this realm. Chapters 4 and 5 answer RQ1 and RQ2 respectively, while Chapter 6 ties our results with ground truth via a verification tool, i.e., the recommender system. Chapter 7 discusses limitations and future prospects before Chapter 8 draws conclusion.

1.3 Who are Stakeholders?

The stakeholders and beneficiaries of this effort are the developers of the music recommender systems and their users. This project will directly influence researchers of automatic recommender systems to suggest more user-mentality befitting songs for virtual music platforms. This will ensure more realistic suggestions of music pieces to the users who are also one of the beneficiaries of our outcomes. Here, it is assumed that the songs registered under the identity of a user are the set of songs that the user listens to on his/her will. That is how we realize that those song genres are the preferred song types of the users to use for personality identification. Besides, that is why we set out to claim that the songs recommended will be user-mentality befitting after calculating performance metrics.

1.4 Contribution

The contributions of this thesis are as follows:

Track features to song genre Identifying song genres from audio attributes of the songs a user listens to online. Automatic identification of song genres given the acoustic feature values of any new song is pledged here with high accuracy.

Song genre to user personality Detecting the users' personality based on the fuzzily labeled genres. This will eventually help the researchers build a system that will automatically understand the individual's nature after extracting the information about her/his playlist(s) using Spotify API.

Personalized recommender system Building an automatic recommender that focuses on the users' individualized personality types gives us the hope of reproducing their preferred song genre(s). Our results indicate that such a platform will enable the music recommender systems to be more accurate and sensitive as well.

Chapter 2

Background

Our approach involves the usage of techniques in data science and applied machine learning before we bring in the concept of recommender systems. By following the procedure we mentioned in Chapter 1 step by step, we want to add new knowledge about using audio attributes to more effectively shed light on personality traits. Figure 2.1 shows a typical flow of knowledge discovery. We exhausted the general procedures in these disciplines, including data classification, regression, clustering, and cross-validation. To justify that our methods align with the processes of competitor approaches and the reproducibility efficacy of the RS reflect our results, we also studied psychology and statistical models. It helped us correlate the methods from multiple domains and optimize our outcomes for knowledge discovery.

2.1 Incorporating Personality in Recommender Systems

Personality refers to the individual differences in the characteristic patterns of thoughts, feelings, and behaviors, which make a person unique. It is believed that it remains relatively consistent throughout one's life. According to Ferwerda et al., the mood is temporary, but the personality is not [26, 27]. Therefore, there is a high potential that an individual with certain choices will tend to make particular decisions. That is why suggesting songs according to that will increase his/her experience with the tool or platform. Nunes et al. discussed how incorporating users' characteristics into recommender systems could highly enhance suggestion quality, and user experience [60].
Typically, a **recommender system** refers to an information filtering system that seeks to pre-

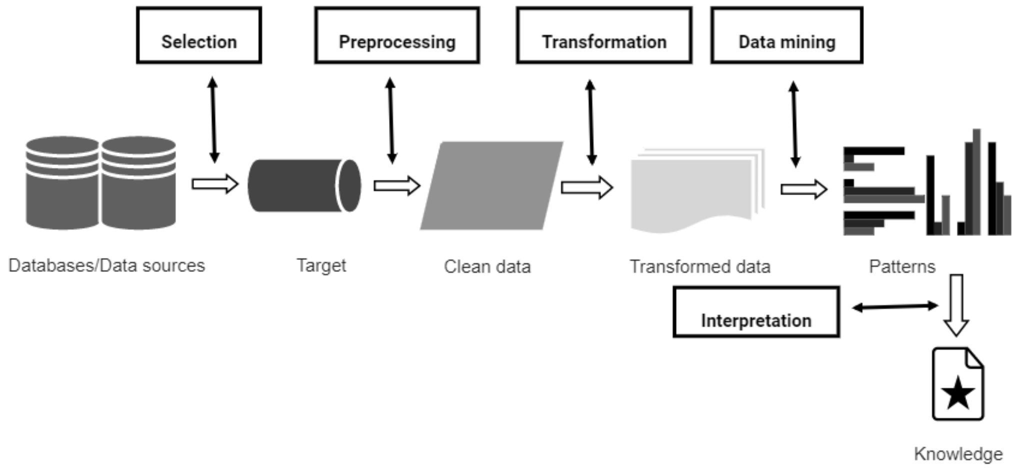


Figure 2.1: Flow of knowledge discovery. The steps mentioned here help us achieve the goal of this study and create new knowledge in the discussed research arena.

dict the 'preference' a user would have for an item. Such an engine usually helps when a person looks around for suggestions in the same bubble; however, recent research [33] argued that an RS could be an excellent medium for an individual to understand himself/herself more accurately and intimately. The process is thoroughly application-based, and this study builds itself upon existing research through experiments, which are verified every step with ground truths.

Chapter 3

Related Work

The investigation in this study involves research works from three tracks. First, we discuss the Myers-Briggs Type Indicator (MBTI) model, where we talk about the Big Five Factors of human personality detection [55] and the famous 16 Personalities model [1]. We argue why and how music is immensely vital in personality detection and consequently create our ground with prior works that music can be a fundamental element to connect the emotional side with technical aspects. We delve deeper into personality detection models using musical attributes to probe music preferences. Second, we discuss how existing literature has contributed to identifying the user personality from the individualistic and customized music data available. To align the literature review with the study's underlying goal, i.e., understanding users' personalities based on their music listening behavior, we choose to explore the existing relevant algorithms extensively.

Finally, we discuss prospective data sources over time and how our study extensively explored nuances in them to find suitable datasets. We sum up with a discussion about the similarities, dissimilarities, experiments, and outcomes of existing recommender systems.

3.1 Music Preferences and Personality Models

Cattell & Anderson made a notion that music preference can help understand human personality [18]. Creating a Music Preference Test, they used it to determine how the users like the songs. Nave et al. used MyPersonality datasets [7] to understand the active and online music listening behavior of participants from different countries [58]. The authors made statistical and graphical deductions based on the Big Five model factors and concluded a reliable correlation between the actual personality traits and the music-based personality predictions. However, no specific mapping was proposed, which allowed us to build our work upon them.

Over the years, the MBTI model has been extensively researched for different aspects and traits of human behavior [62]. Litle & Zuckerman showed a correlation between song genre and user personality [47] while McCown et al. went on to predict preferences for music based on personality elements [52]. However, the 16 personality traits used by [52] needed to be improved after the recent MBTI model started incorporating more personality traits in accordance to song genres. Dollinger confirmed that music genre could be dependent on the psychology of the person who is listening to it, paving the way to perceive the intersection of song genre and personality type as a complex and fuzzy one [23]. Rentfrow & Gosling analyzed the structure of the users' musical preferences based on 14 genres they identified [69]. They pointed out four different categories based on the underlying musical features or dimensions and showed statistical models and consequent significance of their studies. Gosling et al. [32] delved deeper into the Big Five model to understand the participants' stance and opinions about their music preferences via Likert Scale (most widely used approach to scale qualitative responses in survey research) [4] and other computational tools. Skowron et al. [73] used two ensemble regression methods on last.fm dataset [3] to bring out the affecting factors regarding music preferences in a specific socio-cultural condition.

Furthermore, Delsing et al. provided theoretical reasoning behind linking personality patterns with music genres in different age groups of listeners [21]. Dobrota & Ercegovac experimented with a group of college students as they asked them to listen to the same list of 10 instrumental songs [22]. Subsequently, they showed correlations between the song tempo and the Big Five personality traits and concluded that song tempo is a decisive factor in understanding human emotions. Hsu & Hsieh showed how feature selection could be helpful if done through correlation coefficient clustering [37]. We could realize that selecting features beforehand would prove more reasonable and fruitful if we want to cluster based on the inter-relations between certain song types and audio features. However, because of the fuzzy and overlapping nature of the song features in denoting the genre(s), we opted out of feature selection. Mentioning the understudied intersection of song features and human personality, Tully examined the song genre and inherent musical dimensions and came up with interesting and descriptive statistical results [75]. To determine how music preferences were related to the personality factors, the author also provided hypotheses using the chi-square test, Kolmogorov-Smirnov test of normality, Shapiro-Wilk test of normality, etc., and thus examined the correlation existent between personality types and music preferences. Such statistical analyses helped us gain more exposure to the abstract relationship between these two but did not give out any exclusive benchmarking policy. Knowles fragmented the traits of personality dimensions and interpreted the statistical outcomes regarding the correlation of music preference and personality type [44].

Ferwerda et al., among other researchers in recent times, brought the idea of incorporating the music streaming and listening platforms into the discussion of personality traits prediction [27]. By conducting an online user study involving a music application called 'Tune-A-Find' and its users, the authors defined music taxonomy as mood, activity, and genre. They ensured that their deduction after analyzing the user inputs aligned with their hypotheses that personality traits help predicting music preferences. Therefore, it can be mentioned here that from a point where Rentfrow & Gosling studied whether personality is related to preferences for specific music genres [69], Ferwerda et al. made a decisive remark that they are indeed related [27]. Besides, they worked with the mapping between attributes and genres of music and explained how it is shaped from the concept of personality traits of individual users [26]. The authors here extracted information of last.fm users to get to know their online music listening behavior and created graphs and tables to confirm the fact that there exists a relationship between the music genre and personality traits. However, they could not compare their results with prior findings because their work was based on music features or attributes that were not dealt with previously. Langmayer et al. researched whether musical preferences led to proper identification of personality and to what extent, based on four genres [45]. They found out internal correlations among the Big Five factors (O, C, E, A, N); however, they did not provide much information about linking song genres with these five traits.

Goel et al. banked on neural networks and used Discrete Wavelet Transform (DWT) to automatically classify the genres based on the music audio features [30]. Here the researchers did feature extraction and used Parallel Multi-Layered Perceptron with Back Propagation Algorithm. They worked with a supervised learning method to calculate and compare performances via accuracy, which was 85%. In contrast to that, our target is to create a generalized way to link audio attributes to music genres where we are not entitled to any gold standard labels, hence cannot use supervised learning. Nevertheless, this research work [30] comes close to our goal in bridging the territory of music genre and audio attributes. However, the dataset used by the authors was not very large,

and the entire process was practically black-boxed. While the research on bridging the gap between musical attributes and song genre is still on the rise and happening persistently [30], it was only very much organic that as researchers, we would need to investigate the connection between the latent musical features and personality types. Rentfrow et al. mentioned that after doing a follow-up study on the same pool of participants five months later, they found that users' preference for music genre did not change [68]. This suggests that the music preference dimensions are reasonably stable over time and can be reliably used for detecting human personality. In fact, according to Schulte [72], music listening patterns help the online streaming platforms recommend suitable songs to the users, which depend on the song genres the individuals pick over time. Zangerle & Pichl [80] also pointed out how music information retrieval (MIR) supports and empowers music recommendation systems (MRS) with the help of implementations of machine learning algorithms on last.fm dataset. The authors chose to apply K-Means, XGBoosting, and GMM to get content-based models. It is to be mentioned here that content-based structure means incorporating the inclusive features of the songs, and according to them, it helped them learn multiple faces of users' music preferences. However, their goal was more into designing recommender systems based on the different faces of music preferences of a user, and they wanted to model the system for the users.
Nonetheless, this idea resembles ours with respect to our target of understanding different facets of human personality based on music listening behavior. As a result, we looked into the prevalent algorithmic approaches next.

3.2 Techniques of Music Features Analysis

The intersection between audio features and song genre has been under-explored. A python library was developed by a group of researchers to analyze audio signals of a song and to reveal the inherent music type [53]. Wack et al. made a significant impact on the study of song genre detection as they implemented Relevant Component Analysis and Nearest Neighbors to reduce the intra-class variance in the dataset, which in turn allowed them to precisely identify the clusters as genres [77]. However, they did not work with the acoustic feature values we are interested in. Moreover, their task was part of a supervised learning process, unlike ours, where we do not have any gold standard labels. Bogdanov et al. also mentioned similar models in quest of genre recognition from content-based features [14]. However, the authors did not implement any algorithm to achieve this; instead, put the theoretical idea only. Another research work also made an effort towards audio analysis by creating a cross-platform open-source library with the help of Essentia [15]. The authors worked with spectral, temporal, tonal, and high-level music descriptors, unlike the acoustic features we are interested in. The major difference here is that the authors did not focus on a discrete set of values; instead made a time-series analysis of these feature values. Nevertheless, these three studies [77, 15, 14] were very impactful for defining our work process. The study conducted by Murauer et al. [54] again came very close to our concept as they worked with the content-based features of four datasets from Essentia and AcousticBrainz. Linear SVM and Extra Tree algorithms classified the songs into labels, and there were multiple such labels. These were compared to retain the label that got the most percentage of votes for an instance as its genre. Ghosal et al. also came up with a novel algorithm named RANSAC (random sample and consensus), which considered melodic and rhythmic aspects only in a focused way instead of all the song features [29]. Then Nearest Neighbor classifier was used by the authors along with MLP and SVM. Accuracy reported was above 97%, which was better than what Goel et al. reported [30]. Nanni, among many other researchers in this arena of automatic genre classification, also discussed building upon existing works [57, 56]. However, the classification and its high accuracy were not explainable as the authors used neural networks. In contrast, we implement clustering techniques with clear visualization even after the accuracy is not that high. It is to be reiterated that accuracy alone cannot be the component to consider any model the best one.

Zangerle et al. extracted the Million Song Dataset and divided the features into low-level (acoustic), and high-level (abstract) [84]. Combining both the elements, the authors used Neural Networks based on TensorFlow (keras) to predict hit songs; however, our goal is not classifying the hit songs, rather identifying the genres. A similar issue was observed when an extension to an existing RCNN algorithm combining low- and high-level features was discussed [51]. Pichl et al. also made an intriguing analysis of playlist generation behavior of Spotify [8] users by using K-Means clustering techniques to explain the acoustic differences between playlists [65]. They also computed user-cluster correlation, which gave them the advantage to interpret songs for the group of users. This path is inspiring for us as it involves users; however, our idea is to cluster songs explicitly into genres based on their audio attributes. Mai et al. considered a novel algorithm Act-DBSCAN for density-based clustering, which promises to work under constraints of a limited number of pairwise similarities to acquire well comparable clustering result [49]. Englmeier et al. analyzed musical similarity based on explicit semantic analysis followed by feature extraction and K-Means clustering [25]. The difference was again considering signals as features instead of the numerical values of audio attributes. Pichl et al. used spectral clustering to cluster playlists according to the musical features into five clusters [66]. They determined the number of clusters by analyzing the explained variance and an analysis of the eigenvalues. The idea - *Within Cluster Sum of Squares (WCSS) decreases with an increasing number of clusters* was used, and five was an optimal solution. However, the clustering procedure did not come with any performance metrics or codes. Ye et al. proposed a full spectral clustering algorithm FUSE, which generates statistically independent pseudo-eigenvectors eliminating noise and only keeping cluster-separation information [78]. The authors also promised scalability.

Li et al. introduced a sampling-based subspace clustering where through experiments, they put much emphasis on scalability along with a comparison with prevalent clustering methods [46]. Agrawal et al. described CLIQUE, which identifies the subspaces that contain clusters followed by identification of clusters [12]. The authors tested the algorithm's efficiency and accuracy on synthetic datasets with clusters of high density in specific subspaces, and CLIQUE performed better than BIRCH and DBSCAN in terms of retrieving the clusters. However, they discovered meaningful clusters embedded in lower-dimensional subspaces only, whereas our real dataset has high dimensions. Plant described SONAR as an Independent Cluster Analysis; however, the parameters used by the author were restricted to signals, i.e., time-series data, unlike the features we are dealing with [34]. Bezdek et al. introduced scikit cmeans fuzzy algorithms, which were limited to image processing only [13]. Böhm et al. presented RIC, which is a robust clustering technique against noise or outliers [16], and in our opinion, we are looking into achieving something like this that will enable us to cluster the fuzzily distributed song genres, including outliers. Hubig & Plant proposed a novel clustering algorithm, NORD (for NOn-ReDundant), which they said efficiently discovers the truly relevant clusters in complex datasets without requiring any kind of threshold on their redundancy [38]. According to them, the quality of a cluster lies in its contribution to the overall data compression rate. We came across a few more relevant papers on the clustering validation topic that elaborated on metrics [35] or computed and compared errors [17]. Saitta et al. came up with a new index called 'score function', which they said to be performing better than the existing indices in terms of addressing extreme cases, e.g., single and perfect clusters [71]. Liu et al. remarked that the clustering validation index based on nearest neighbors gave the best results concerning clustering validation techniques [48].

Clustering techniques and their validity played critical roles in identifying song genres from track feature values. To do that, we investigated available data sources of different capacities.

3.3 Datasets Linking Music and Personality with Recommender Systems

We chose to work with a public dataset where real user information from online music listening platforms is available. We explored the Twitter [11] -based Spotify dataset [81]; however, it did not have the music attributes values we were looking for. Pichl et al. came up with recommendation systems based on the datasets [63]. Zangerle et al. did a superb job as they put together the acoustic music values and the lyrics, which they used to do both lexical and sentiment analysis [83]. But again, they used both content-and context-based features to make recommendations of playlists based on the songs and their users. Whereas our goal was the other way around, i.e., we wanted to know about the playlist and the values of the attributes and then derive the users' personality traits. Hence, despite this study [83] being very much relevant to our groundwork and processes, the goals were different. Zangerle et al. also took the assistance of a multimodal approach for characterizing a track by analyzing the audio signal and the corresponding lyrics [82]. This could be an excellent resource for our study; however, according to the authors, this research work is yet towards completion and closure. Yahoo! music dataset contained ratings by the users [24] which inclined more towards recommending music to users, rather than choosing song genre. One more dataset based on last.fm was publicly available [10] which dealt with top N song recommendation without user individualization. Pichl et al. [64] also provided a dataset containing user id and de-identified track id along with artist id and playlist id of Spotify; however, it in no way posed us with an opportunity to merge it with any other dataset containing music feature values. Another dataset [6] based on Million Song Dataset was created, which could only give us the mood of the lyrics. We also explored EchoNest based dataset [9]; however, we could not relate the important column values with any other sources, e.g., last.fm or MusicBrainz [5] to augment it for our suitability.

Finally, we settled with the resourceful dataset that came with the study by Poddar et al. [67], which was augmented to fit our research questions. We also ensured that the dataset would allow us to verify our results about genres and personality via implementing an RS, which we will elaborate on in the next sections. In addition, we will collate how our data-centric model is giving us optimal solutions without adopting user studies for increasing user experience [43, 42].

Chapter 4

Pipeline One: Audio Features to Genre Classification

4.1 Description of the Dataset

The dataset at hand was the final product after the data source provided by Poddar et al. [67] was augmented by scraping user information from Twitter. A full-fledged dataset was generated, where we could get the real-time user IDs extracted from Twitter API and the corresponding track feature values of the songs they listen to on Spotify. It also contained tweet IDs by the respective users, which referred to the tweets carrying information about those songs registered. These tweets, later on, proved to be very useful as we applied the 'Bag-of-Words' technique to do sentiment analysis. The outcome was considered the ground truth, which we used in Pipeline 2 to compare with our result of personality analysis from song genres. Besides, there were recording IDs collected from MusicBrainz, which were frequently different for the same track. This was because the large database has multiple entries of the same song, which made them redundant. Table 4.1 gives a glimpse of the dataset.

After data augmentation and modification were done, the dataset had a column for user IDs (which could be relayed back to Twitter for easy recognition as Spotify de-identifies user information) and eight columns of features that define the tracks, along with other auxiliary columns. Here, we see no gold-standard label against the song instances that we can call song genres. For classifying

Data column	***Datatype***
tweet_id	int64
user_id	int64
user_name	object
timestamp	object
track_name	object
artist_name	object
recording_id	object
energy	float64
liveness	float64
tempo	float64
speechiness	float64
acousticness	float64
danceability	float64
loudness	float64
valence	float64

Table 4.1: Data columns and their types

song genres, we only used the user ID and the audio features in Pipeline 1, which are:

Energy This is a perceptual measure of intensity and activity. Typically, energetic tracks feel fast, loud, and noisy.

Liveness It detects the presence of an audience in the recording since the higher the value, the greater the probability of the song being performed live.

Tempo It is estimated from the average beat duration of a track.

Speechiness It detects the presence of spoken words in a track. The more instruments-based a piece is, the lesser it will be in terms of speechiness values.

Acousticness It is a confidence measure whether a track is acoustic.

Danceability It describes how suitable a track is for dancing based on a combination of musical elements, including tempo, rhythm stability, and beat strength.

Loudness It is calculated in decibels (dB) after averaging across the entire track.

Valence It measures musical positiveness conveyed by a track.

The associated information about the audio features implies domain knowledge [8]. Fig 4.1 shows the ranges of these values, which lies between 0 (lowest value) and 1 (highest value) after normalization. Normalization helps to reduce data redundancy and improves data integrity by rescaling

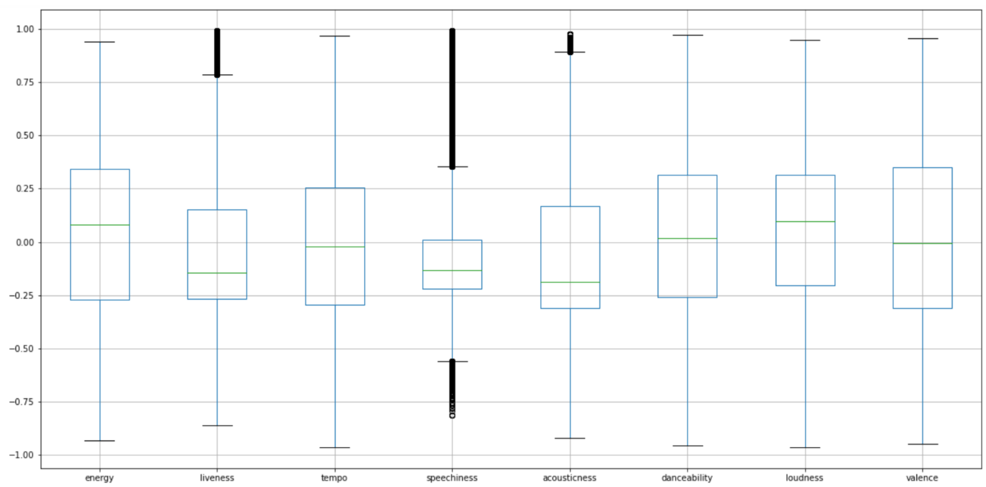


Figure 4.1: Boxplots of the track features with outliers. Median and quartile values within a specific range are shown.

the given values. The figure is essentially a boxplot of the associated track features. A **boxplot** is a standardized way of displaying the distribution of data based on five bars: minimum, first quartile (Q1), median, third quartile (Q3), and maximum. It can tell about the outliers and what their values are. Here, we see outliers, which we initially wanted to get rid of to avoid any skewness in the distribution or incorrect behavior of data points. To that, these apparent outliers of the dataset were put out of consideration with certain thresholds. However, we found a large number of rows getting discarded in this process and knew that these are not necessarily noises as we applied DBSCAN. Being a data clustering algorithm, Density-based Spatial Clustering of Applications with Noise (DBSCAN) groups together closely packed points (points with many nearby neighbors). It also helps in marking outliers that lie alone in low-density regions (whose nearest neighbors are too far away).

4.2 Listing of Genres from Labeled Dataset

We consulted the tiny dataset created by Zangerle et al. [84] where the instances were labeled against a genre, given the track feature values. According to the authors, the dataset did not have any user-identifying information; rather, the listening events were marked by a label, which is the genre. The genres were:

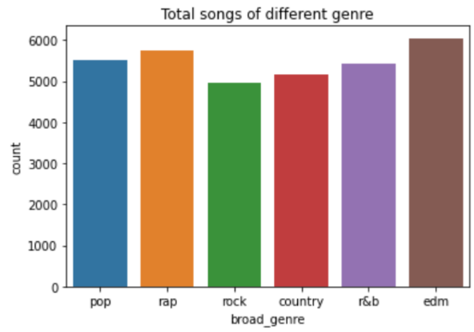
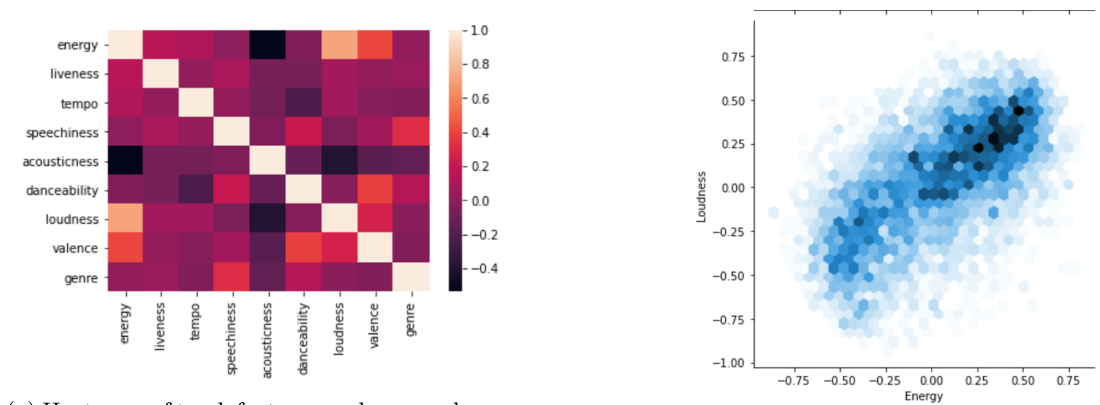


Figure 4.2: Number of songs of different genres in the labeled dataset.

- Country
- Rap
- Pop
- Rock
- EDM
- R&B

This dataset from [84] was created and analyzed by the authors for identifying the hit song numbers in a specific period. Unlike us, they had the genres as inputs, whereas our goal is to identify the genres from the given music attributes. This labeled data source also helped us reach a larger labeled dataset to refer to while labeling the unlabeled dataset. Fig 4.2 shows the 32833 listening events classified under the six genres mentioned above. A **count plot** is used here that shows the counts of observations for each category.

We took help from the dataset to understand correlations between the track features that can be analyzed for genre classification. Fig 4.3a shows the **heat map** of the track features and the corresponding genre. Being a data visualization technique, a heat map shows the magnitude of a phenomenon as color in two dimensions. The color variation gives obvious visual cues to the reader about how the phenomenon is clustered or varies over space. We considered the following one-hot encoding: Country = 1, EDM = 2, Pop = 3, R&B = 4, Rap = 5, and Rock = 6. Here, we find a high positive correlation between Energy and Loudness, which also satisfies the existing domain knowledge [8]. The **joint plot** in Fig 4.3b suggests this correlation in this labeled dataset.



(a) Heat map of track features and genres shows a strong correlation between Energy and Loudness.

(b) Highly correlated Energy and Loudness.

Figure 4.3: Existing positive and negative correlation in the labeled dataset.

By knowing these correlations, we could understand the relationship between song genres and track features. We also found a small p-value (0.0021) for Energy and Danceability for them to be really significant for the model. Liveness and Tempo also had a small p-value (0.0233), and in the latter part of the section, we have used these significances to understand their influence on labeling genres. It is to be mentioned here that the null hypothesis is assumed to be the probability that new variables will not change the model. With the p-value of a variable being lower than 0.05 in 95% confidence interval, the variable is significant for the model, and hence null hypothesis is rejected.

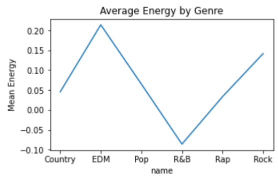
Fig 4.4 contains the subplots of the average attribute values for all genres. Noticeable findings were:

1. Fig 4.4a and Fig 4.4g suggest that EDM genre has the highest energy and loudness values, while R&B has the least of these values. This relays back to the domain knowledge, i.e., EDM is supposed to provoke dancing spirit and energy with loud music, whereas R&B songs are reportedly soft and soothing all the way.

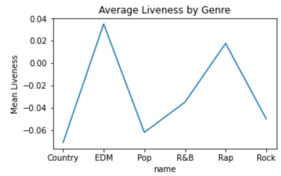
2. Similarly, 4.4c also shows the slowest tempo for R&B (blues) songs.

3. From Fig 4.4d it can be inferred that Rap songs are usually the highest in speechiness, which proves the naming of the genre.

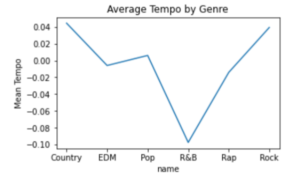
4. Fig 4.4e suggests the highest acoustic values of Country music, which is the essence of the genre as we know of.



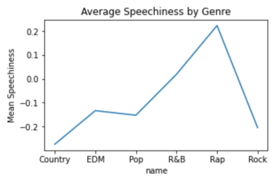
(a) Average Energy



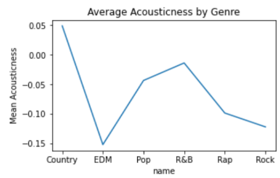
(b) Average Liveness



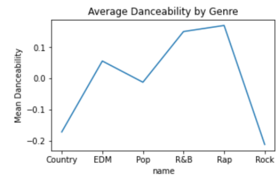
(c) Average Tempo



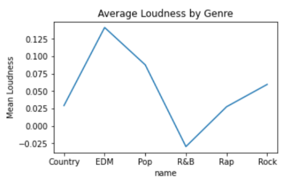
(d) Average Speechiness



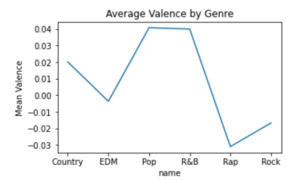
(e) Average Acousticness



(f) Average Danceability



(g) Average Loudness



(h) Average Valence

Figure 4.4: Average track feature values of all song types.

Some of the aspects could also be observed by the data according to domain knowledge [8]; however, these were not prevalent:

1. Rock songs allegedly have higher values for liveness.

2. EDM songs are supposedly higher in danceability values.

Fig 4.5 clarifies that the data points (song instances) distribution grouped by speechiness gives way more prominence to Rap song identification despite the outliers. The other audio features were not distinguishable enough while considering those songs under specific genres. For example, Fig 4.6 shows the boxplots of danceability of the genres where R&B and Rap songs are showed to have the same range of values. This makes it challenging to choose which genre one song would belong to if the value falls under this specified range. Theoretically, EDM should have higher values for danceability, reflecting on the distribution and the min-max-median calculation. However, practically, this is not what we found.Hence, we realized that audio features have large overlaps to rule out a song as a specific genre based on these values and eventually opted out of feature selection. Instead, as an attempt to put thresholds around the feature values to ascertain their genres, we used density plots next as represented in Fig 4.7. A density plot represents the distribution of a numeric

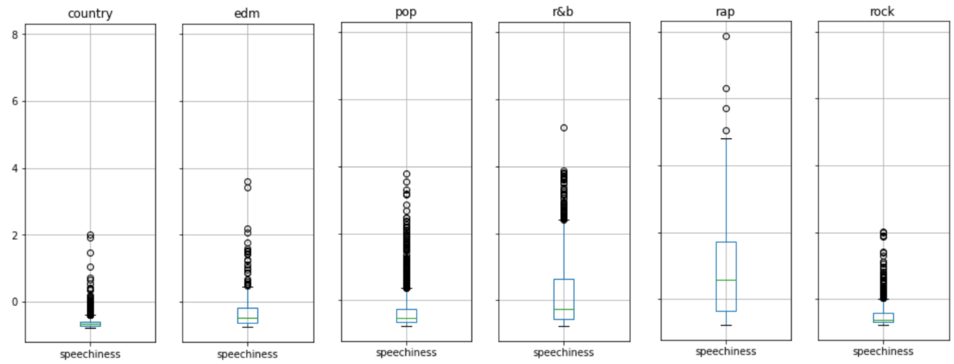


Figure 4.5: Boxplots of Speechiness of different genre of songs

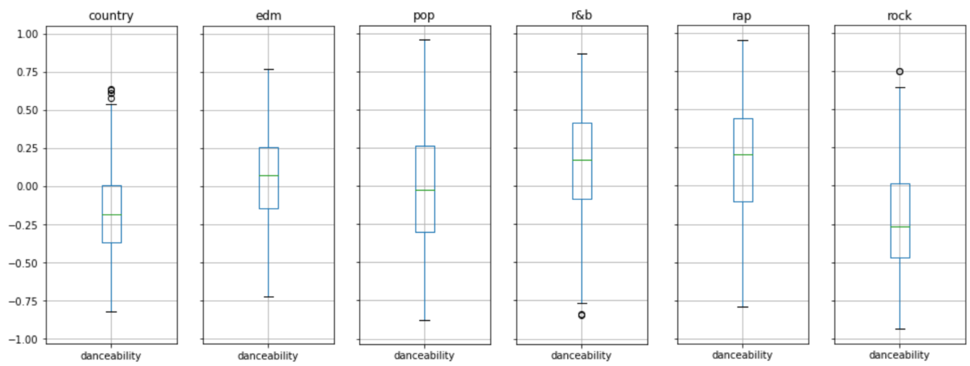
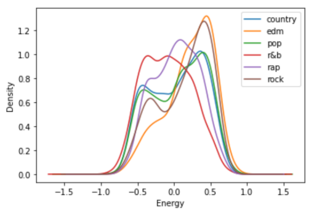


Figure 4.6: Boxplots of Danceability of different genre of songs

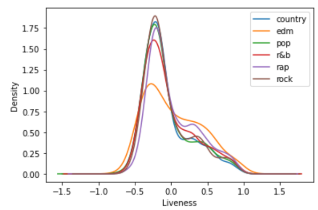
variable. It uses a kernel density estimate (KDE) to show the probability density function of the variable by producing a Gaussian Bell Curve. According to 4.7d where variance is lesser than other features, speechiness for Rap is distinguishably higher. However, looking at 4.7f we find that both R&B and Rap having higher danceability value, which contradicts with the domain knowledge 'rap songs are not danceable enough'. Moreover, in other cases, the ranges of the features coincide with each other making it difficult to rely on the graph-based results of the labeled data. This brings us to the point that feature values cannot be used for benchmarking purposes.

4.3 Preliminary Analysis of Dataset

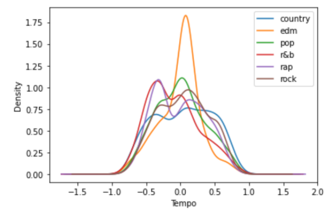
Initially, there were 11609883 rows in our unlabeled dataset. After necessary data cleaning (e.g., getting rid of null values, removing rows with undecipherable categorical values, etc.) and data processing (e.g., regularization, standardization, etc.) a basic Exploratory Data Analysis (EDA) on



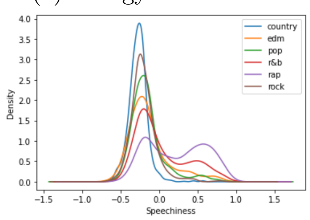
(a) Energy distribution



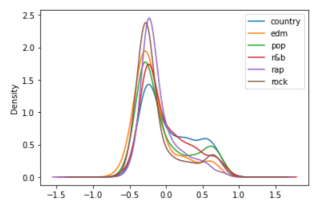
(b) Liveness distribution



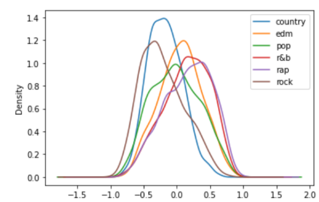
(c) Tempo distribution



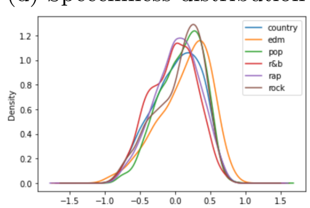
(d) Speechiness distribution



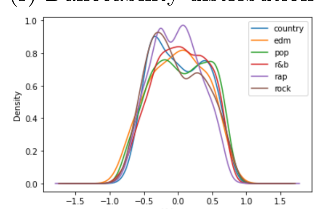
(e) Acousticness distribution



(f) Danceability distribution

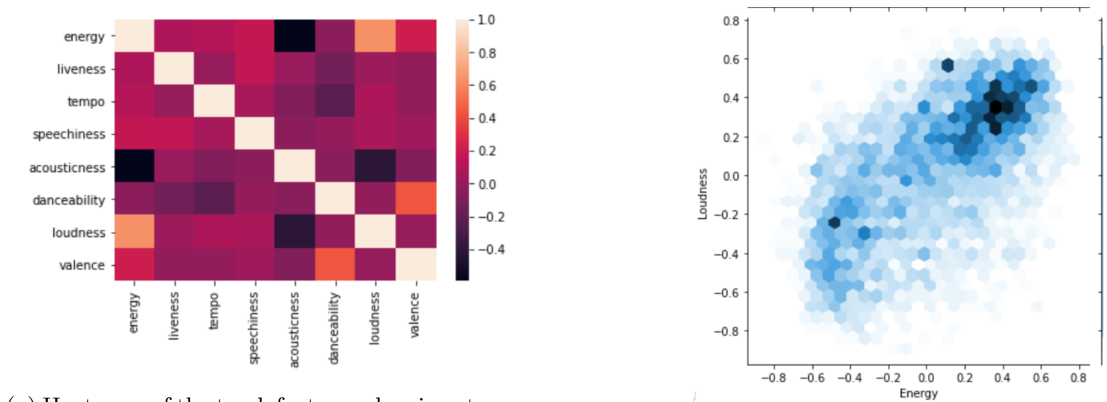


(g) Loudness distribution



(h) Valence distribution

Figure 4.7: Density plots of track features of labeled dataset.



(a) Heat map of the track features showing strong correlation between Energy and Loudness

(b) Highly correlated Energy and Loudness

Figure 4.8: Existing correlation in the unlabeled dataset.

the 77062 rows of the dataset was done. There are 1237 unique user IDs and 4436 unique songs. Getting rid of the repetitive recording IDs, which were 60024 in total, we were left with 5752 instances.

Fig 4.8a shows the correlation matrix for the track features of our dataset. Here, we find a strong correlation between Energy and Loudness again as found in Section 4.2, which is supported by [8]. The joint plot in Fig 4.8b suggests this correlation in this unlabeled dataset. However, we see some irregularities here, which we can owe to the large number of outliers present in the dataset. Furthermore, analyzing significance for the model, we found the p-value to be 0.0099 for Energy and Danceability. We also found a small p-value (0.033) for Liveness and Loudness, making them significant as well for the model. Moreover, we got another small p-value (0.0008) for Acousticness and Danceability, which justified the domain knowledge [8] about Country songs being both acoustic and danceable. We also plotted these more significant features to show how their correlation helped us explore genre classification more deeply. According to the labeled dataset, we already reported that the correlation matrix does not show consistent strong positive or negative correlations between attributes and genres. As a result, feature selection was not made. Even after some feature engineering, it did not give better RMSE values than adopting dimensionality reduction, as shown in the latter section. Root mean squared error (RMSE) is the square root of the mean of the square of all of the errors, and it is considered an excellent general-purpose error metric for numerical predictions. However, it was clear that classification or clustering will help us predict the genres at that point, given this was unsupervised learning. The cross-entropy loss is suited to classification problems since

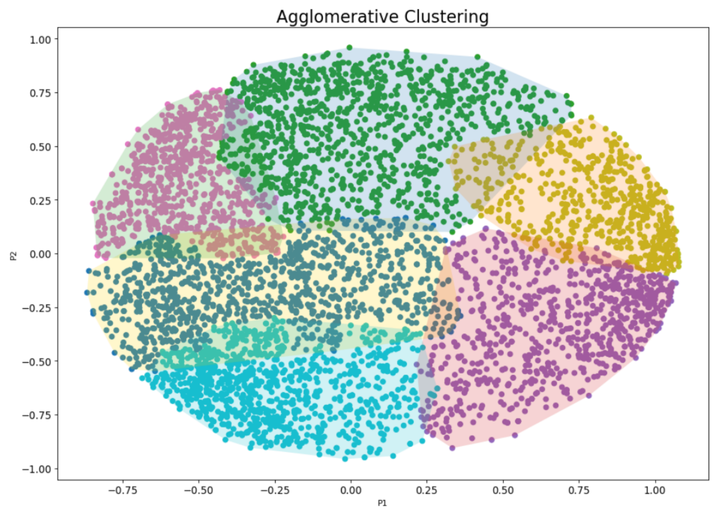


Figure 4.9: Agglomerative Clustering gives us six clusters after correlations indicate their similarity.

we can attain it from the likelihood of a classification model.

4.3.1 K-Nearest Neighbor

At this point, we had a large labeled dataset and a small unlabeled one from the same source and with the same range of values. Among other classifiers (e.g., Random Forest, Extra Tree, Decision Tree, and XGBoosting), we used K Nearest Neighbor (KNN) regression algorithm to predict the labels.

4.4 Clustering of Dataset

4.4.1 Principal Component Analysis

We used Principal Component Analysis (PCA) on the unlabeled dataset to reduce the dimensionality of the dataset from eight to three. This was done for the ease of visualization, keeping the principal components of the dataset even in the reduced dataset.

4.4.2 Agglomerative Clustering

Another meaningful visualization was provided by Agglomerative Clustering, which is shown in Fig 4.9. This is a form of hierarchical clustering, which builds upon hierarchy of clusters.

4.4.3 K-Means Clustering

We then used K-Means clustering, where n number of observations are divided into K clusters, and each of them belongs to the cluster with the nearest mean. Fig 4.10a shows the clustered data points in 3D after K-Means clustering is applied. Here, we also get to see the clusters and their corresponding genres. We find that the clusters are largely overlapping, which means that there are data points that share common spaces because of the hard assignment of clusters.

4.4.4 Independent Subspace Analysis and Clustering

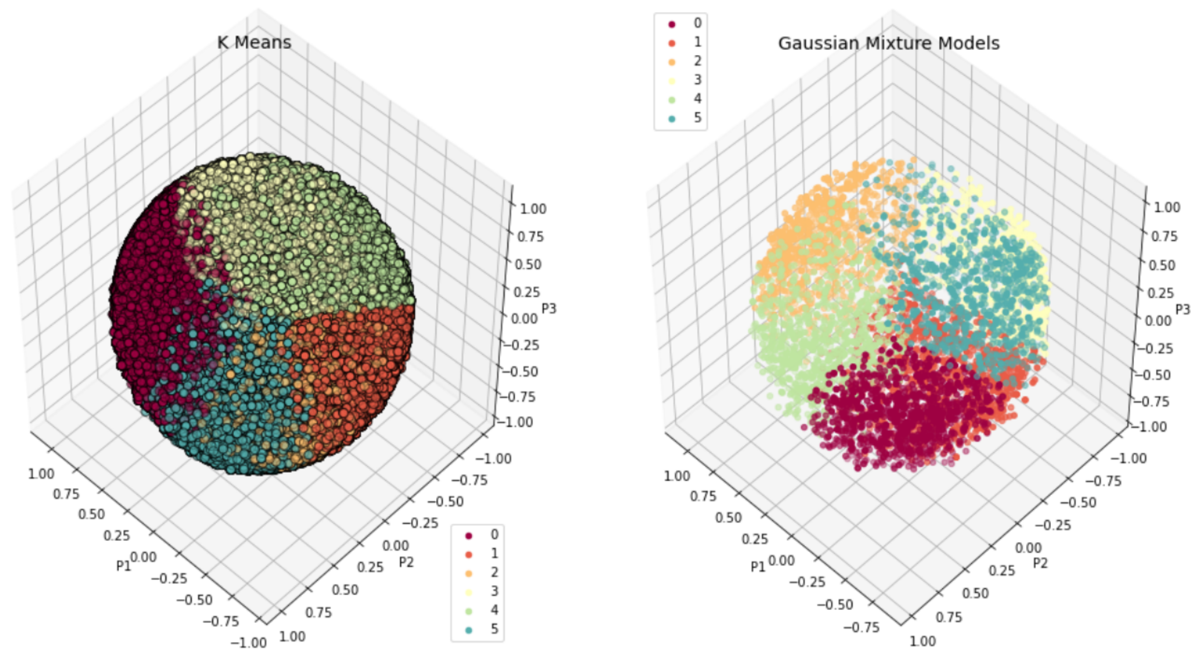
Hence, we implemented subspace clustering, which finds clusters within different subspaces of the same or different dimensions. Using Independent Subspace Analysis and Clustering (ISAAC) by [79], we found clusters with shared subspaces in 2D space.This algorithm needed the set of target values, which we had from K-Means. However, this was not an optimal solution since the clusters did not provide much information about the clustered data points. So, an improvement of K-Means was called for.

4.4.5 SUBKMEANS

Mautz et al. initiated Subspace K-Means or SUBKMEANS, an improvisation of the partition-based clustering techniques, where the dimensionality of the clustered space is determined automatically [50]. In SUBKMEANS, the goal is to simultaneously find a sufficient K-Means style clustering partition and transform the clusters into a common subspace, which is optimal for the cluster structure. Feeding the set of target values we had from K-Means as targets, this algorithm gave us the predicted targets, and **Normalized Mutual Information (NMI)** was found to be 0.5. A normalized measure of MI, i.e., NMI, is computed for clustering results of a training set with known labels [40]. The clustering quality of community finding algorithms is often tested using NMI, and that is why the higher the NMI value, the better for clustering validity. Therefore, NMI here being 0.5 shows the validity of the clustering technique is 50%.

4.4.6 Gaussian Mixture Model

In order to check the validity of NMI calculated from SUBKMEANS, we next chose another clustering algorithm Gaussian Mixture Model (GMM), where each cluster is modelled according to a



(a) Clusters in 3D space and their respective labels after K-Means clustering.

(b) Clusters in 3D space and their respective labels after implementing GMM.

Figure 4.10: P1, P2 and P3 are the three principal components after reducing dimensionality of track features. Labels 0-5 refers to the six clusters formed of different colors.

different Gaussian distribution. This flexible and probabilistic approach to modelling the data means that we have soft assignments of clusters, instead of the hard assignment as seen in Fig 4.10b. We predicted the target value (i.e., genre), and when they were compared with the values after applying GMM, data had 50% correctness. To recall, we got 50% as the NMI after using SUBKMEANS clustering algorithm to predict the genres. These are precisely equal but less than the percentage shown by Tzanetakis & Cook [76], which was 53%. They identified genres from pitch and rhythm and showed a slightly higher correctness measure.

Nevertheless, the clusters we got are distinct, and the complexity of the overlapping nature of the data points was handled well here. In fact, with respect to visualization, GMM shows the six clusters better than K-Means. However, to keep consistency in algorithms used, we stick to K-means for clustering purposes after we got predicted labels from the KNN classifier. In Section 4.6, we will show how clusters found from K-Means algorithm accurately match the separate genres.

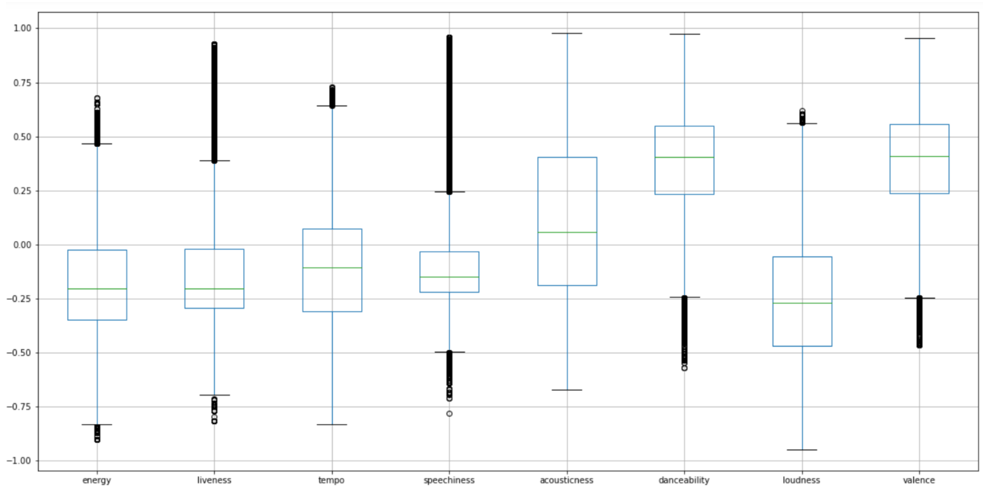


Figure 4.11: Boxplots of the track features of Country songs.

4.5 Visualization of Clusters based on Genres Predicted

Fig 4.11, 4.12, 4.13, 4.14, 4.15 and 4.16 show the boxplots of the predicted genres i.e., Country, EDM, Pop, R&B, Rap and Rock respectively.

The boxplots let us compare the predicted labels with the prevalent gold standard labels, and we found the similarity in the audio feature range and values. We see the overlaps in danceability values dominant in Country and EDM songs in Fig 4.17 which is really intriguing and consistent with the results we calculated. In fact, we measured that Country songs have the highest danceability values, whereas domain knowledge points towards EDM. We also found that the cluster dedicated to Country songs had 13% EDM songs of the total population.

Next, from Fig 4.18, we understand the density plots for all kinds of song genres. We plotted speechiness for Rap in Fig 4.19a where we notice that the inclusion of outliers does not give the expected result, as shown in Fig 4.18d. Hence, the spikes in other genres disrupt the regular nature of Rap songs having the highest speechiness values. In fact, we found that the cluster dedicated for Rap songs had 8% of total songs which were not actually of Rap genre. We also looked at Loudness in Fig 4.18g, and it was found that EDM has higher values for it as seen in Fig 4.19b. Interestingly, the quartile values of danceability of Rock songs in Fig 4.6 refer to the lower danceability values found in Rock songs. We can verify this from Fig 4.18f and Fig 4.20a as well. Moreover, Rock songs tend to be higher in liveness values according to Fig 4.20b. Both findings match with the

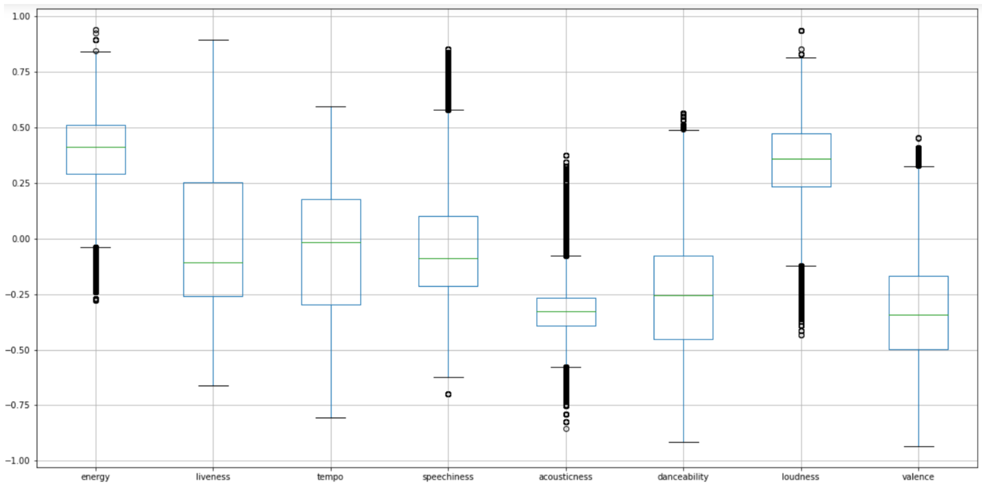


Figure 4.12: Boxplots of the track features of EDM songs.

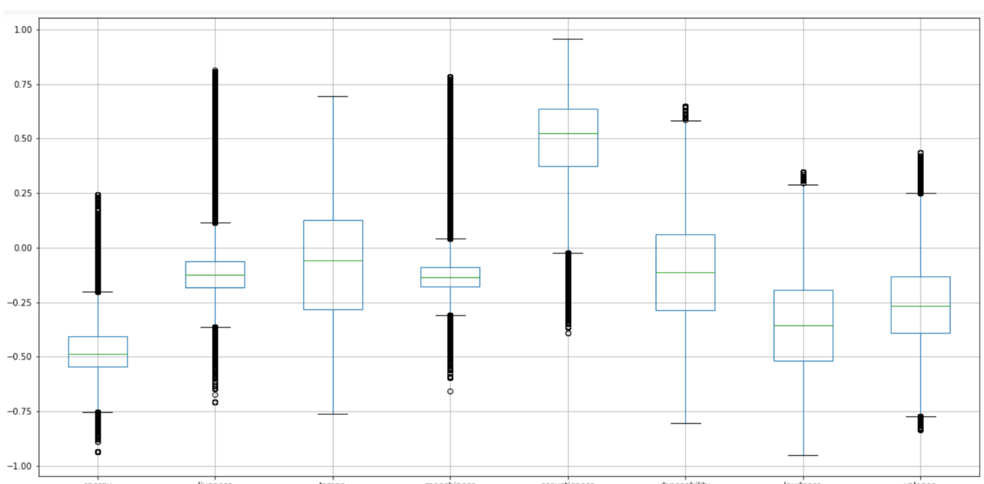


Figure 4.13: Boxplots of the track features of Pop songs.

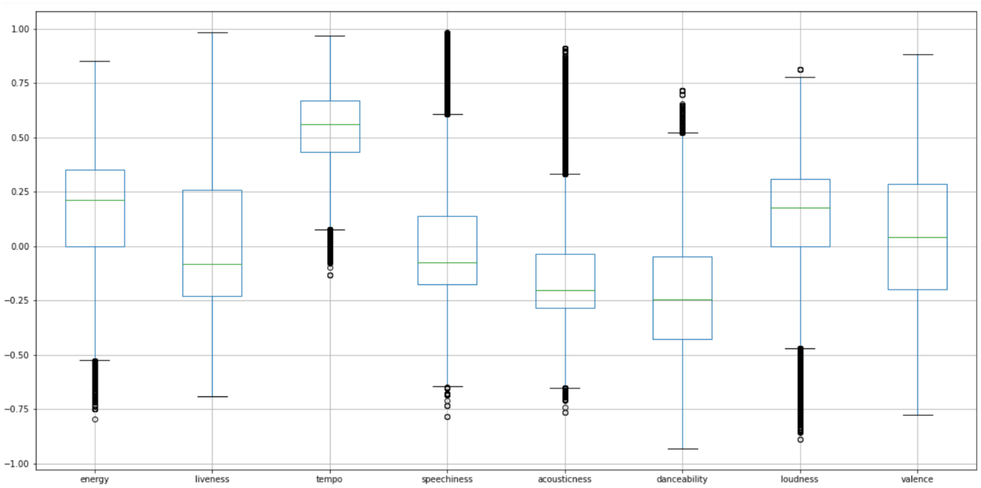


Figure 4.14: Boxplots of the track features of R&B songs.

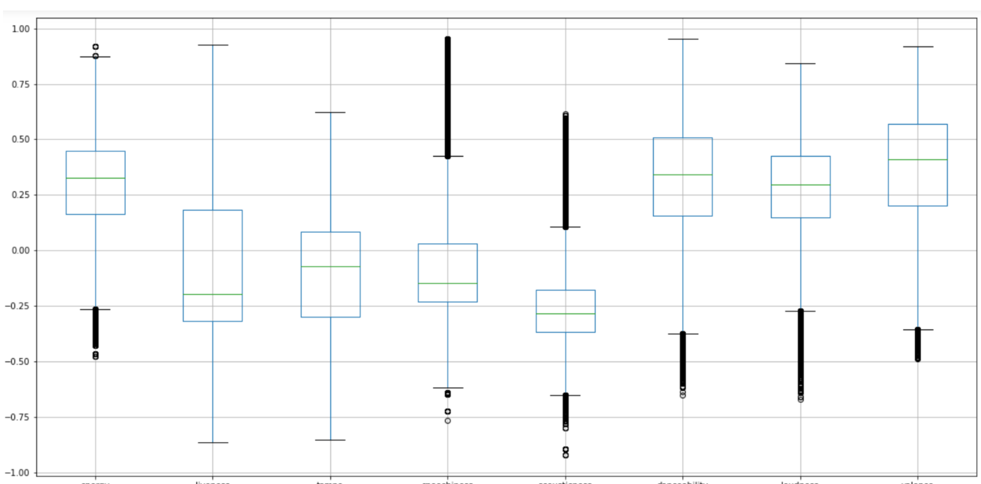


Figure 4.15: Boxplots of the track features of Rap songs.

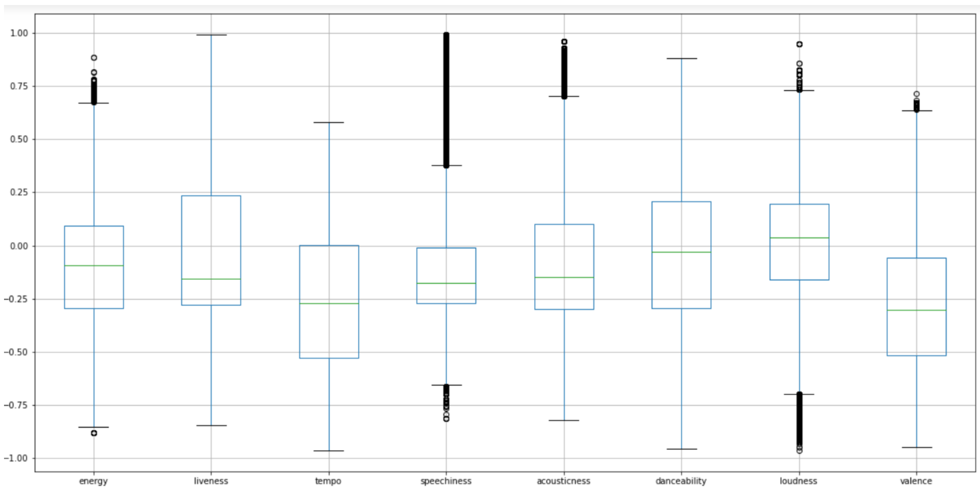


Figure 4.16: Boxplots of the track features of Rock songs.

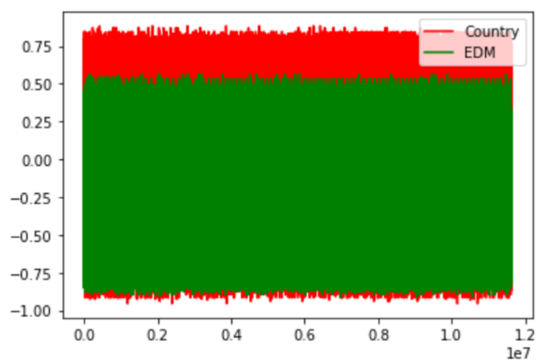
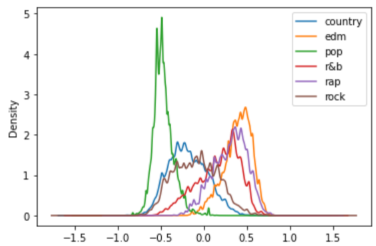
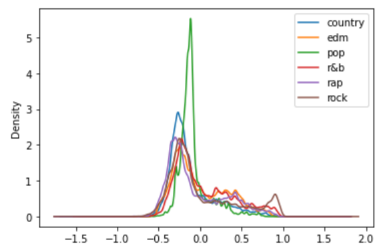


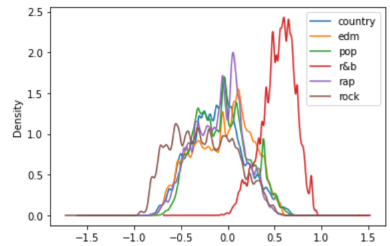
Figure 4.17: Largely overlapping values of Danceability for Country and EDM. This phenomenon made it difficult to distinguish these two song types in the unlabeled dataset.



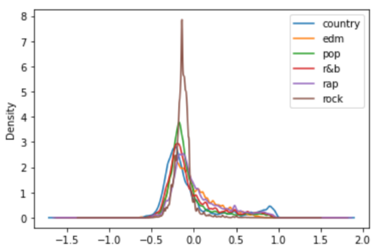
(a) Energy values



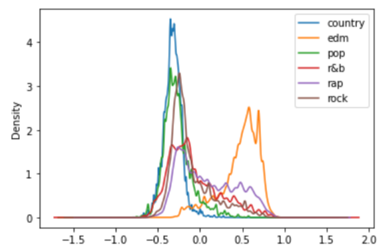
(b) Liveness values



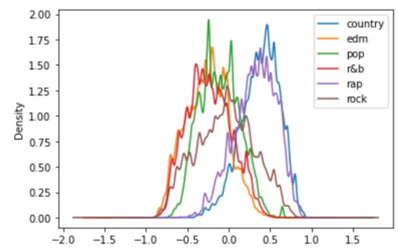
(c) Tempo values



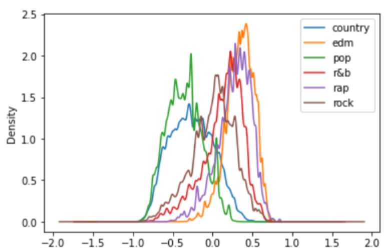
(d) Speechiness values



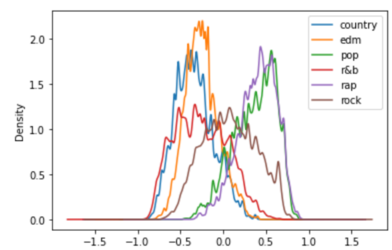
(e) Acousticness values



(f) Danceability values

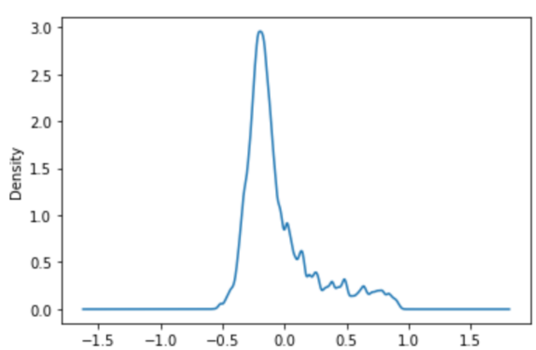


(g) Loudness values

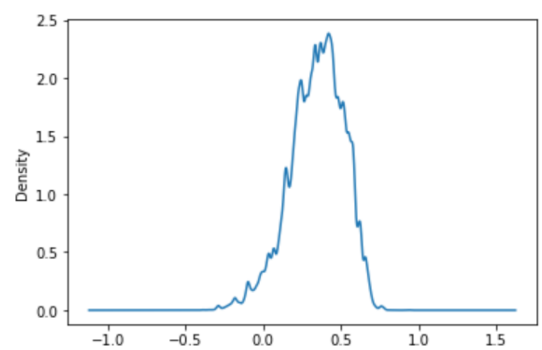


(h) Valence values

Figure 4.18: Density plots for all song types.

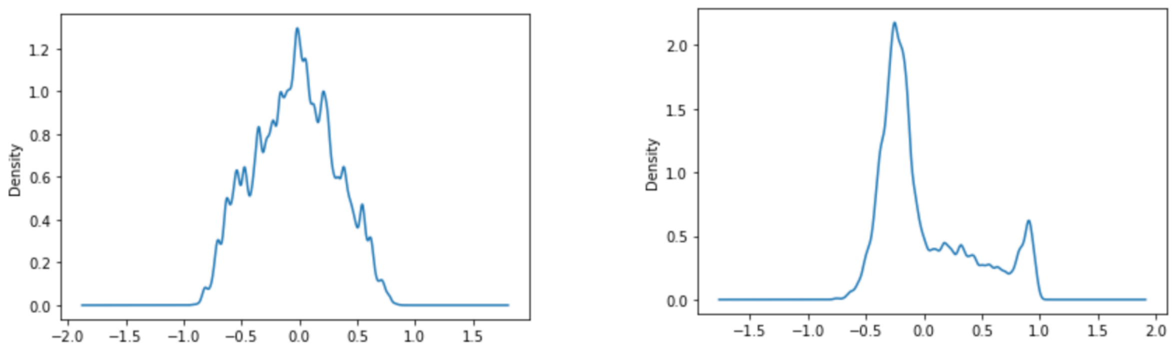


(a) Range of Speechiness values for Rap songs, where the distribution reaches its peak at 3



(b) Range of Loudness values for EDM songs, where the distribution reaches its peak at 2.5

Figure 4.19: Density plots of the two significant feature values.



(a) Distribution shows Danceability values

(b) Distribution shows Liveness values

Figure 4.20: Density plots for Rock songs.

domain knowledge of Rock songs being less danceable and more full of a live audience.

4.6 How do We Know that the Clusters Refer to the Genres?

The predicted labels from the KNN algorithm were used for K-Means clustering approach, where we calculated the silhouette score to be 0.502. **Silhouette score** for clustering is calculated using the mean intra-cluster distance to validate consistency within clusters of data. The coefficient here means that data points within clusters are 50% consistent. But the question remained whether the clusters dedicated to a particular genre actually contained songs of that predicted genre. It was also to be verified whether the songs of the same genres were clustered together.

To justify our cluster assignment, we went ahead to find out the percentage of these two categories and their respective accuracy scores by merging the dataset containing ground truth and the dataset with predicted genre labels. Since unsupervised learning algorithms do not have any straightforward evaluation metrics, we wanted to answer these queries, e.g., what percentage of EDM-labeled songs end up in the EDM cluster? What percentage of the labeled items in the EDM cluster is EDM-labeled songs?

Using the K-Nearest Neighbor Regression algorithm, the labels were predicted for the unlabeled dataset before merging both the datasets. Next, we grouped the listening events by genres. The combined dataset was clustered into six using the K-Means algorithm. The textual value of genres were one-hot encoded, and the six genres (1-6) are Country, EDM, Pop, R&B, Rap, and Rock respectively, while the six clusters (1-6) sequentially refer to EDM, Rap, Country, R&B, Rock, and

Pop. At that point, our target was to find out the validity of cluster assignment after predicting the labels. We found out that out of 5477 Country labeled songs, 5155 fell into its cluster (based on the gold-standard labels). Interestingly, based on the labeled dataset, three EDM-labeled songs did not belong to the EDM cluster. For the merged dataset, 6040 songs were EDM out of 6228 songs in the EDM cluster. Out of 6470 Pop labeled songs, 6125 fell into its cluster (based on the gold-standard labels). Out of 6588 R&B labeled songs, 5643 fell into its cluster (based on the gold-standard labels). Out of 6836 Rap labeled songs, 6104 fell into its cluster (based on the gold-standard labels). Out of 6986 Rock labeled songs, only 5176 fell into its cluster (based on the gold-standard labels). Frequently, EDM songs are labeled as Country songs due to their similar range of values in **Danceability**, which we could visualize and verify from Fig 4.17.

4.6.1 Accuracy Calculation of Cluster Assignment

The percentage of Country-labeled songs ending up in the Country cluster is **94%**, while the percentage of the labeled items in the Country cluster being actually Country songs is **97%**. The percentage of EDM-labeled songs ending up in the EDM cluster is **97%**, while the percentage of the labeled items in the EDM cluster being actually EDM songs is **80%**. The percentage of Pop-labeled songs ending up in the Pop cluster is **95%**, while the percentage of the labeled items in the Pop cluster being actually Pop songs is **98%**. The percentage of R&B-labeled songs ending up in the R&B cluster is **86%**, while the percentage of the labeled items in the R&B cluster being actually R&B songs is **82%**. The percentage of Rap-labeled songs ending up in the Rap cluster is **89%**, while the percentage of the labeled items in the Rap cluster being actually Rap songs is **93%**. The percentage of Rock-labeled songs ending up in the Rock cluster is **74%**, while the percentage of the labeled items in the Rock cluster being actually Rock songs is **89%**. Rap songs are the easiest to identify as it appears because of their distinguishing **Speechiness** values. The low percentage of Rock songs getting identified correctly is due to the overlapping values of **Liveness** with Rap and R&B. Track features have overlapping influences on the song labels or genres. Domain knowledge about song attributes playing important roles in labeling into genres is not always accurate when it comes to real-life data analysis. Finally, a 10-Fold Cross Validation is done to justify cluster assignment, and the accuracy score is **78%**.

Table 4.2 shows the performance metrics of cluster assignment validation processes. The accuracy we got after CV got reduced a little, which refers to the removal of overfitting (if any).

Metric	Value
Accuracy score	78%
Silhouette score	50%
NMI score	50%

Table 4.2: Clustering validity measures. Here, we find from our computation that out of 100 songs, 78 are identified accurately, and with reasoning behind that.

Our result is better than [70] and [76]. However, the performance is not higher than what Srinivas et al. [74], Ghosal et al. [29] and Goel et al. [30] reported (99.41%, 97% and 85% respectively). Nonetheless, unlike their approaches, our model did not involve Neural Network or Deep Learning methods, so cluster assignment is explainable with statistics and visualizations.

For example, from Table 4.3 we see the playlist of a particular user with five listening events (i.e., five rows of data frame). The rows have values of music attributes, which were standardized and normalized. Each song has been labeled with the genre it belongs to. This was made possible because our proposed model clustered songs upon the prediction of the broad genres based on gold-standard labels. Then cross-validation of the model classified the songs into labels or genres. Since the track feature values are abstract and overlapping, we chose to use all the genres identified to go into the next step of our pipeline, which is about getting to know about the user's personality.

Track name	Energy	Liveness	Tempo	Speechi -ness	Acoustic -ness	Dancea -bility	Loudness	Valence	Genre
Revival in the Land	-0.539132	0.232695	0.309419	-0.351254	-0.082312	-0.114512	-0.221367	-0.605953	R&B
Jesus Loves the Little Ones	-0.535516	-0.126395	0.194306	-0.097016	0.573668	0.272720	-0.166049	-0.468013	Pop
Climbing Higher and Higher	-0.416101	-0.090920	-0.358511	-0.124581	0.604605	-0.343620	-0.270854	-0.342867	Rock
Im Not Lisa	0.161551	-0.231705	0.084680	-0.218226	-0.418177	-0.292435	0.320240	0.708857	Rap
Great God	-0.492463	-0.214282	0.187086	-0.181992	0.616203	0.116661	0.012303	-0.499973	Pop

Table 4.3: Summary of track feature values of a real anonymized user '10095402'. The user supposedly listens to Pop, Rock, R&B and Rap songs, which can be identified from the tracks registered with his ID.

Chapter 5

Pipeline Two: Genres to Personality Detection

In his research, Celli showed how unsupervised learning could help us understand user personality from online social sites [19]. We also worked with unlabeled data in pipeline one before we clustered the dataset into genres. We will determine the personality types from the genres predicted in pipeline one with reasonable accuracy for pipeline two. For making that happen, we assume that the users listen to the songs as registered in the given dataset according to their own will, which is why the playlist is important to deduce their personality pattern. Looking back at the basic statistical information about the numbers of occurrences registered against a user in the dataset, there could be two possibilities:

- The user listens to the same genre(s) of songs over and over, leading herself/himself towards a singular personality trait.
- The user listens to a different genre(s) of songs, making her/him a person of multifaceted personality traits.

From Table 4.3 we found that the user listens to song genres of Rap, R&B, Rock, and Pop. The user may listen to these songs more frequently or different kinds of songs; however, since we are relying on the static dataset, not on the runtime information of the users' music listening history, we believe our diagnosis is the most accurate one in this scenario. Our next task is to understand the underlying personality based on the existent and universally acknowledged personality models.

5.1 Linking Personality with Genre(s) Identified

5.1.1 Personality traits

Based on Myers-Briggs Type Indicator (MBTI), there are two of the most popular and unanimously accepted personality models - Big Five and 16Personalities [55, 1].

16 Personality model ESFJ, ISFJ, ESFP, ISFP, ENFJ, INFJ, ENTP, INTP, ESTP, ISTP, ESTJ, ISTJ etc.

Big Five traits are characterized as following [31]:

Openness to Experience (O) curious, intelligent, imaginative. High scorers tend to be artistic and sophisticated in taste and appreciate diverse views, ideas, and experiences.

Conscientiousness (C) responsible, organized, persevering. Conscientious individuals are extremely reliable and tend to be high achievers, hard workers, and planners.

Extroversion (E) outgoing, amicable, assertive. Friendly and energetic extrovert people draw inspiration from social situations.

Agreeableness (A) cooperative, helpful, nurturing. People who score high in agreeableness are peace-keepers who are generally optimistic and trusting of others.

Neuroticism (N) anxious, insecure, sensitive. Neurotics are moody, tense, and easily tipped into experiencing negative emotions.

The notations for the Myers-Briggs model can be elaborated as given in Table 5.1. Furnham proposed a mapping between the MBTI and the Big Five in 1996 [28]. He considered MBTI as a Big Four model without Neuroticism. However, the mapping was improvised later on, as we see from Table 5.2. Nevertheless, the fifth row not being included in the original mapping, we relied more on Big Five for strengthening the relationship between genres and traits.

5.1.2 Mapping of Music Genre and Human Personality (Big Five)

Rentfrow & Gosling came up with music preference dimensions [69] which were:

Acronym	***Description***	***Acronym***	***Description***
E	Extravert	I	Introvert
N	iNtuitive	S	Sensing
F	Feeling	T	Thinking
J	Judging	P	Perceiving
T	Turbulent	A	Assertive

Table 5.1: Notations and their meanings in Myers-Briggs Personality Model

MBTI	***Big Five***	***Correlation***
I**N**tuition/Sensing	**O**penness to Experiences	O correlates with N
Feeling/Thinking	**A**greeableness	A correlates with F
Perception/**J**udging	**C**onscientiousness	C correlates with J
Introversion/**E**xtraversion	**E**xtroversion	E correlates with E
Turbulent/Assertive	**N**euroticism	N correlates with T

Table 5.2: Mapping between Myers-Briggs and Big Five Personality Model

1. Reflective and Complex

2. Intense and Rebellious

3. Upbeat and Conventional

4. Energetic and Rhythmic

Based on the classification, the following list shows the relationship between different song genres and the Big Five traits [75]. Here, the terms in bold refer to the genres we have used in this thesis. The rest of the genres could not be considered due to a lack of ground truth to support their mapping for both the pipelines.

O classical, **rock**, heavy metal

C soul

E rap, pop, EDM, country

A latin

N hip-hop, **blues**

5.1.3 Mapping of Music Genre and Human Personality (16Personalities)

The following list shows the relationship between these definite genres and the associated personality types [1].

Country ESFJ, ESFP, ENFJ

EDM ESFP, ENFP, ENTJ

Pop ESFP, ESFJ, ISFP, ESTJ

R&B ENFP, ENFJ, ESFJ

Rap ESTP, ESFP, ESTJ

Rock INTP, ENTP, INFP, INFJ

From such a mapping, we can say that it is rightly perceived that rock listeners are mostly introverted compared to the rest of the listeners in the pool.

5.2 Mapping between Personality Traits and Song Genres

Table 5.3 shows the mapping between the six genres and the personality traits. Here, ND means not detected - which might attribute to the genres we left out as we took only six of them as the sample size. It is quite clear from this table that if a person likes to listen to more than one music genre (which is pretty common, as we saw in the dataset), the person has multiple personalities existent in herself/himself. It is also observed here that the rock song listeners are not much inclined to listen to any other song genres.

5.3 Final results

Continuing from the outputs shown in Table 4.3 and using the relationship between the song genres and the personality models, we show the final values in Table 5.4. It leads to the fact that the user is thoroughly extroverted (E) with an attitude of openness to experiences (O) and slight neuroticism (N). In fact, due to the spectral nature of human personality, this person can be termed as - ESTP, ESFP, ESTJ, ESFJ, ENFJ, etc., and a little bit of INTP/INFP. This is to be

Personality trait	***Genre***	***Personality trait***	***Genre***
O	classical, rock, heavy metal, jazz, folk, soul, alternate	ENFJ	country
C	soul, country	ENTP	rock
E	rap, pop, EDM, country	ENTJ	EDM
ENFP	EDM, R&B	A	latin
N	hip-hop, blues, R&B	ESTP	rap
ESTJ	rap, pop	ISFP	pop
ESFP	country, EDM, rap	INTP	rock
INFJ	rock	INFP	rock
ESFJ	country, pop, R&B	ISFJ	ND
ISTJ	ND	ISTP	ND
INTJ	ND		

Table 5.3: Exhaustive list of apparent relationship between song genre and personality type

#	***Genre***	***Big Five Model***	***16Personalities Model***
1	Rap	E	ESTP, ESFP, ESTJ
2	R&B	N	ENFP, ENFJ, ESFJ
3	Rock	O	INTP, ENTP, INFP, INFJ
4	Pop	E	ESFJ, ESFP, ENFJ

Table 5.4: Identifying personality trait(s) from song genre(s) listened by the user

mentioned that the minor changes in acronyms here owe to the different faces of a person's traits. For more than one instance of a particular genre, the frequency is more effective in understanding the primary personality pattern. It is, in fact, helpful to consider a specific type of personality for a user if the playlist is long or the songs the user listens to are overlapping. This also helps us understand that a person can have multiple personalities instead of assuming that they belong to a single personality type.

The portrayal of the example above helped us realize the spectrum of personality traits any individual can carry. Observing a user's music listening behavior (e.g., navigating any music listening application, prioritizing a list of songs over others, etc.) physically would have given us an edge over narrowing down a user's personality a bit more. The current final output might sound generalized and static; however, we need to keep in mind that the Big Five and the MBTI models we are using here lead to a spectrum of personalities. Moreover, any two persons can have similar personality traits even if they do not have identical music preferences, and that is what our result is about. It is not claimed that the outputs would be individualized; instead, it is the individualized choice of music that ultimately leads the users to the personality spectrum. Therefore, the contribution here is that the user would get to know her/his personality type(s) based on the songs he listened to,

Five Notations	*From Tweets*	*From Track Features*
O	1	0.1
C	0.18	-
E	0.73	0.75
A	0.5	-
N	0.91	0.6

Table 5.5: Pairwise comparison of Big Five notations. '-' points to the absence of the trait in result.

which would help us get more refined music recommendations online.

5.4 Analysis of Tweet Data

We needed to verify our result about personality identification with some ground truth, and analyzing tweets seemed to be the most relevant and organic way to do so. There are total of 5751 unique tweets in the data source. First, we applied for and received permission to crawl tweets from Twitter API. Second, we wanted to do queries using Tweet IDs to get the associative tweets. However, the way the dataset was built, the ID associated with tweets contained only hashtags, i.e., *#nowplaying*, among others, which did not help us in terms of sentiment analysis of tweet contents. In this scenario, we took the help of the Python libraries and did a user query to find out the existent user IDs, and the number was 990. Third, we extracted tweets after doing a timeline query on Twitter API based on the unique users. Fourth, we chose to work on 25 recent tweets of a user and used **bag-of-words** method to find out the Big Five personality traits as a percentage for a user. The bag-of-words model is represented as the bag of its words, disregarding grammar and even word order but keeping multiplicity. It is most suitable when modeling text with machine learning algorithms, i.e., Natural Language Processing (NLP). In our study, we had a-priori dictionary and converted the collection of tweets into a matrix of word counts. Getting rid of the common 'stop words' we created a sparse representation of the appearance counts. Table 5.5 gives out the comparison between our calculated output and the ground truth. Here, the user's personality score from his/her tweets can be seen as the average percentage value. The dominance of certain personality traits can be observed, and to refer to the apparent absence of a trait, '-' was put. Our conclusion about the user matches the ground truth from Twitter personality analysis in **74%** cases and we could validate Spotify users' personalities with their activity on Twitter.

Chapter 6

Pipeline Three: Cross-Validation with a Recommender System

Our study with users' playlists containing songs they prefer listening to and have tweeted about has given us the opportunity to:

- identify song genres based on a reference labeled dataset and justify the predictions **(Pipeline 1)**, and
- determine individual's personality from song genres and verify that with sentiment analysis on tweets **(Pipeline 2)**.

In **Pipeline 3** we implemented a personality-based music recommender system to validate our results about cluster formation and genre classification through recommending relevant songs to the users. It is to be reiterated here that the pipelines work and produce results assuming that these real users tend towards listening to these songs registered under their IDs. Since this is a static dataset with no runtime updates to the songs, we could reliably use the existing tracks for developing the model. It helped us with understanding the accuracy of the RS, too.

6.1 Calculating Personality Scores

The user's personality score was calculated with the Big Five percentage based on the songs listened to. From the previous example, let us suppose that a user mostly listens to Pop, Rap, Rock, and R&B songs according to his/her registered playlist. In that case, the user is Extrovert, Neurotic, and Open to Experience besides having a little bit of Conscientiousness and Agreeableness. Upon knowing the dominant features of the user's personality, we go through a binary check and then add the percentage values up together, considering their presence above a threshold of 50%. This resembles the user-based Matrix Factorization approach, where the 'score' of the recommender system is essentially the personality score, not 'listen count' or 'popularity'.

6.2 Traditional Matrix Factorization

We started out with conventional **Matrix Factorization (MF) via Singular Value Decomposition (SVD)** for recommendation purpose. MF refers to a collaborative filtering algorithm, where the user-item interaction matrix is fragmented into a product of multiple matrices. SVD is an algorithm that decomposes a matrix into the best lower rank (i.e., smaller or simpler) approximation of the original matrix. We had 'songs' to pivot 'userID' matrices with. Next, we wanted to do a K-Fold Cross-Validation (here, K=5) to determine the performance measure with a small hold-out set. **Cross-validation** is a resampling procedure used to evaluate machine learning models on a limited data sample.

For example, we had 60024 songs in total, and user '15518784' had 3774 songs he/she listened to according to the original dataset. We created a dummy instance based on the samples to make it 3775. After that, we did an 80-20 split to get the train sets and testing sets, respectively. We used 3020 for training 100 and 755 for testing purposes. Next, we calculated 1000 recommendations using latent factors. This leads us to the question - what percentage of those recommendations is also in the test set? We found the precision score at N (here, N = 1000) to be **3.5%** on an average. As a matter of fact, for some users, we got a higher precision and lowered for some. Fig 6.1 shows the histogram of precision scores according to users' frequency. The baseline precision being 1% with the total number of songs being 60024 and the number of songs in the testing set being 755, it performs better than that. However, when the number of songs sampled (N) is decreased, e.g., 10 or 100, we get a precision at N of 0%.

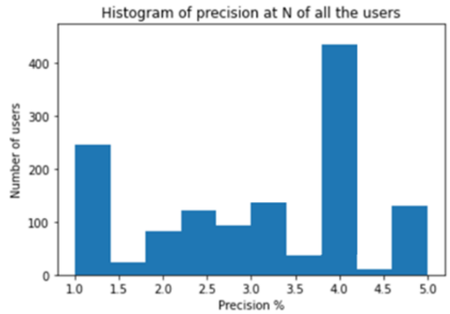


Figure 6.1: Histogram showing precision at 1000 values for all the users.

6.3 Item Similarity Matrix

A customized recommender system was implemented based on the similar songs the person listened to, according to our data source. Getting rid of the redundant tracks, we found only a handful of users listened to the same song more than once. It was very random and very less frequent, so it did not impact the counts of songs per user. The modified dataset also saw shared songs among multiple users. From such characteristics, we could explore the similarity of the songs the users were listening to. After that, we followed these steps for evaluation purposes:

1. For a particular user, we iterated through pipeline 1 to find out the genre(s) of the songs s/he listens to as registered in the dataset. Our clustering model gave us the predicted genres.

2. For the same user, we made a list of top-N (here, N = 1000) with ranking based on their similarity of music attributes. The top song on the list was the most similar one with the songs of the user's registered playlist, and this value was kept between 0 and 1.

We got an accuracy of 51% regarding the appropriate use of the algorithm to find out similar songs to the user. However, we were in need of improvising this algorithm to find out song suggestions based on individual personalities. As a result, we explored personality scores mentioned above to implement a personality-based RS next.

6.4 Personality-based RS

Using personality scores as indices for pivot purposes, we implemented a personality-based MF to get song recommendations for a user based on his/her Big Five personality model. We

	track_name	genre
0	Revival in the Land	r&b
1	Im Not Lisa	rap
2	Great God	pop
3	Climbing Higher and Higher	rock
4	For My Broken Heart	r&b
5	Por el amor de una mujer	pop
6	Obscure Verses for the Multiverse	rap
7	Into the Blue	rock
8	Glorious Unfolding	r&b
9	Beat It	rock
10	Gannet	r&b
11	Love Has Found a Way	r&b
12	La valle dei Re	pop
13	Yellow River	r&b

(a) A part of already rated/listened songs.

	track_name	genre
0	Paradise in Distress	r&b
1	Dont Go Messin With My Heart	rock
2	They Dont Care About Us	edm
3	Luv U More	pop
4	We Are the World	rap
5	When You Come Back to Me	r&b
6	Nothing Like the Rain	rock
7	I Cant Sleep Without You	country
8	Night in Motion	r&b
9	Ray of Light	rap

(b) A part of newly recommended songs.

Figure 6.2: Partial lists of already listened (left) and suggested songs for real anonymized user '10095402'.

calculated the precision at N, which is around **0.9%**. MF does not scale particularly well to massive datasets, and our large shaped dataset could be a reason behind the low precision. It is to be recalled here that precision is the degree to which a process will repeat the same value, whereas precision at N refers to the proportion of recommended items in the top-N list that are relevant. We understand that a better precision score will always empower us with a reasonable model. The value is lower than the traditional personalized MF; however, the solution is more explainable than the competitor approaches [34], where the authors did not report any performance measure. Fig 6.2a shows a glimpse of the already rated songs by our sample user, while Fig 6.2b shows the top 10 recommended songs. Looking at the new genres - Country and EDM in Fig 6.2b, we find that the user gets suggestions about songs besides his/her typical genres of Rap, R&B, Rock and Pop. It is unlike what user or item-based matrices would do, exploiting the similarity in user behavior or song features. Rather, we focus on the overlapping personality traits for multiple song genres. As a result, the user having a spectrum of personalities is being recommended songs based on similar personality score values.

Step #	***Purpose***	***Process/Algorithm***	***Performance Metric***	***Value***
1	Cluster validity	SUBKMEANS	NMI score	50%
1	Cluster validity	Gaussian Mixture Model	Correctness	50%
1	Cluster assignment	K-Means Clustering	Silhouette score	50%
1	Genre classification	10-fold CV with K-Means	Accuracy	78%
2	Personality detection	Bag-of-Words	Accuracy	74%
3	Songs reproduction	Personality-based MF	Precision at N	0.9%

Table 6.1: Different performance metrics at different stages of the pipeline. Personality-based Matrix Factorization process has less mean precision values than the traditional RS (3.5%), but it has explainability unlike competitor approaches [34].

6.5 Interpretation of the Outputs

In Table 6.1 the performance metrics are registered. After one iteration of pipelines 1 and 2, we found that the genres and the consequent personality traits we found for a user could be contextual and not constant. In fact, according to [27, 26], due to having non-singular personalities, users listen to different types of songs, depending on their 'current mood'. As a result, it is always more effective to give insights on the genres they listen to and subsequently the spectrum of personality they belong to than the other way round.

We introduced cross-validation, where we took the average of the predicted genres after splitting the dataset into train and test. According to the Big Five model, the user we sampled here has a personality spectrum of Openness to experience, Extroversion, and Neuroticism. According to Table 5.2, 16 personalities model also expects the person to be highly extrovert, perceiving, and intuitive, along with a tad bit of judgemental behavior. In short, with the help of the personality-based music recommender system, we could justify how our model of song classification into genre(s) from their track feature values (RQ1) and determination of the users' personality from those identified song type(s) based on prevalent correlations (RQ2) is accurate. The whole process of feeding songs into our proposed model to find out associated personality was validated when we used personality-based RS and recovered songs listened to by the user (RQ3).

Chapter 7

Limitations and Future Work

7.1 Constraints of the Dataset

The dataset was the result of merging related data from different data sources. As a result, it had to be augmented in a way so that it served our purpose. After getting rid of a large number of repetitions, unwanted characters, and strings, we kept the common identifiers to make the instances left meaningful and valuable. However, there were different recording IDs for the same song simultaneously for the same user. It generated confusion whether the same user listened to the same song repeatedly for multiple cycles or not. Tracing the recording IDs from MusicBrainz, we found that this vast database can have multiple identifiers for the same track depending on how they were inserted into it. Therefore, MusicBrainz has more than one instance based on whether they are the original version or a cover of the song, whereas Spotify has a single instance of that song. This led to losing some valuable data regarding music attributes as a studio, or acoustic version of the same song has differences in track features.

7.2 Scope of Natural Language Processing

While working on this thesis, we also augmented the dataset with lyrics for later work. Natural Language Processing (NLP) can help understand underlying sentiments in the songs' words; in fact, song lyrics might be directly linked to identifying song genre(s). Nonetheless, we proceeded with the audio features only, which are available via Spotify API, and predicted the genres by analyzing the patterns. Still, we found that there could be more inherent factors to it that could make more robust detection.

7.3 Inability to Capture the Whole Domain

It is to be recalled that song genres are not limited to the discussed six only [2]; rather, there could be many more, and those could be termed as sub-genres of these prime types mentioned in this study. All those song types combined would have strengthened the relationship between the song genres and the personality models. In our next studies, we will bridge the gaps found in this study regarding the shortness of genres discussed and the correctness of the psychology models. We also hope to incorporate user studies in the future, which would help us capture more nuances of human personalities.

7.4 Future Work

Our future work can be two-folded:

- Increasing the accuracy of correctly identified song genres from track feature values by involving more content and context-based song features.
- Improving the personality-based music recommender system will automatically reproduce correct song suggestions to the user based on the individual's music listening behavior.

By understanding an individual's nature after extracting the information about her/his playlist(s) using Spotify API presumably, such a platform will enable the music recommender systems to be more accurate and sensitive. It will also empower researchers to envision this system as the digital versions of the users themselves.

Chapter 8

Conclusions

Music being a significant component of human personality analysis, has driven our study here, enabling us to manipulate the track feature values. This research work's novelty was identifying song genres from audio attributes of the songs a user listens to online with reasonable explainability and performance. Our model has helped us maintain and improvise the current black-boxed accuracy reported by [30, 29, 74]. We have also cross-validated our results to determine the closest ground truth. Such automatic genre classification based on the model mentioned and implemented in the study helped the consequent task of detecting user personality. It is to be reiterated that music features are always very abstract, and hence, music pieces cannot go through the 'hard assignment' of labels. Therefore, we kept the overlapping nature of music features that influenced genre detection and made optimal decisions about the listeners' personality traits. The outputs of our experiment and their nonlinear relationships led to the evidence that humans are prone to carrying more than a singular personality. This means that an individual can have a certain percentage of neuroticism even if s/he is out and out a happy-to-go person. Therefore, having any personality trait up to some probability does not rule out the existence of an opposite trait in that person. As a result, the prediction task is very complex and might not give any straight result. Our study made the highest effort to understand users' personalities based on their music listening history and behavior. Finally, the personality-based music recommender system incorporated the validity of our experiments as it gave an accuracy of 81.3% in terms of suggesting songs based on individual personality. This leaves ample room for improvement in developing better models for analyzing song features, detecting user personality, and recommending relevant songs. Moreover, we also learned that human personality is too broad to contain and sometimes too ambiguous to describe with only music preferences as parameters. We believe that multiple aspects of human lives combined can make a more confident statement of human personality detection and our contribution here will provide a better user experience.

Bibliography

[1] *16Personalities*, Accessed September 29, 2020.

[2] *Kaggle*, Accessed September 29, 2020.

[3] *last.fm*, Accessed September 29, 2020.

[4] *Likert scale*, Accessed September 29, 2020.

[5] *MusicBrainz*, Accessed September 29, 2020.

[6] *Musixmatch*, Accessed September 29, 2020.

[7] *MyPersonality*, Accessed September 29, 2020.

[8] *Spotify Data Visualization*, Accessed September 29, 2020.

[9] *Taste profile songs*, Accessed September 29, 2020.

[10] *Top N songs*, Accessed September 29, 2020.

[11] *Twitter*, Accessed September 29, 2020.

[12] Rakesh Agrawal, Johannes Gehrke, Dimitrios Gunopulos, and Prabhakar Raghavan. Automatic subspace clustering of high dimensional data for data mining applications. In *Proceedings of the 1998 ACM SIGMOD international conference on Management of data*, pages 94–105, 1998.

[13] James C Bezdek, James Keller, Raghu Krisnapuram, and Nikhil Pal. *Fuzzy models and algorithms for pattern recognition and image processing*, volume 4. Springer Science & Business Media, 1999.

[14] Dmitry Bogdanov, Alastair Porter, Julián Urbano, and Hendrik Schreiber. The mediaeval 2017 acousticbrainz genre task: Content-based music genre recognition from multiple sources. CEUR Workshop Proceedings, 2017.

[15] Dmitry Bogdanov, Nicolas Wack, Emilia Gómez Gutiérrez, Sankalp Gulati, Herrera Boyer, Oscar Mayor, Gerard Roma Trepat, Justin Salamon, José Ricardo Zapata González, Xavier Serra, et al. Essentia: An audio analysis library for music information retrieval. In *Britto A, Gouyon F, Dixon S, editors. 14th Conference of the International Society for Music Information Retrieval (ISMIR); 2013 Nov 4-8; Curitiba, Brazil.[place unknown]: ISMIR; 2013. p. 493-8.* International Society for Music Information Retrieval (ISMIR), 2013.

[16] Christian Böhm, Christos Faloutsos, Jia-Yu Pan, and Claudia Plant. Ric: Parameter-free noise-robust clustering. *ACM Transactions on Knowledge Discovery from Data (TKDD)*, 1(3):10–es, 2007.

[17] Marcel Brun, Chao Sima, Jianping Hua, James Lowey, Brent Carroll, Edward Suh, and Edward R Dougherty. Model-based evaluation of clustering validation measures. *Pattern recognition*, 40(3):807–824, 2007.

[18] Raymond Bernard Cattell and Jean C Anderson. *IPAT Music Preference Test of Personality*. Institute for Personality and Ability Testing, 1953.

[19] Fabio Celli. Unsupervised personality recognition for social network sites. In *Proc. of sixth international conference on digital society*, 2012.

[20] Deryck Cooke. The language of music. 1959.

[21] Marc JMH Delsing, Tom FM Ter Bogt, Rutger CME Engels, and Wim HJ Meeus. Adolescents' music preferences and personality characteristics. *European Journal of Personality: Published for the European Association of Personality Psychology*, 22(2):109–130, 2008.

[22] Snježana Dobrota and Ina Reić Ercegovac. The relationship between music preferences of different mode and tempo and personality traits–implications for music pedagogy. *Music Education Research*, 17(2):234–247, 2015.

[23] Stephen J Dollinger. Research note: Personality and music preference: Extraversion and excitement seeking or openness to experience? *Psychology of music*, 21(1):73–77, 1993.

[24] Gideon Dror, Noam Koenigstein, Yehuda Koren, and Markus Weimer. The yahoo! music dataset and kdd-cup'11. In *Proceedings of KDD Cup 2011*, pages 3–18. PMLR, 2012.

[25] David Englmeier, Nina Hubig, Sebastian Goebl, and Christian Böhm. Musical similarity analysis based on chroma features and text retrieval methods. *Datenbanksysteme für Business, Technologie und Web (BTW 2015)-Workshopband*, 2015.

[26] Bruce Ferwerda, Marko Tkalcic, and Markus Schedl. Personality traits and music genres: What do people prefer to listen to? In *Proceedings of the 25th Conference on User Modeling, Adaptation and Personalization*, pages 285–288, 2017.

[27] Bruce Ferwerda, Emily Yang, Markus Schedl, and Marko Tkalcic. Personality traits predict music taxonomy preferences. In *Proceedings of the 33rd Annual ACM Conference Extended Abstracts on Human Factors in Computing Systems*, pages 2241–2246, 2015.

[28] Adrian Furnham. The big five versus the big four: the relationship between the myers-briggs type indicator (mbti) and neo-pi five factor model of personality. *Personality and Individual Differences*, 21(2):303–307, 1996.

[29] Arijit Ghosal, Rudrasis Chakraborty, Bibhas Chandra Dhara, and Sanjoy Kumar Saha. Perceptual feature-based song genre classification using ransac. *International Journal of Computational Intelligence Studies*, 4(1):31–49, 2015.

[30] Anshuman Goel, Mohd Sheezan, Sarfaraz Masood, and Aadam Saleem. Genre classification of songs using neural network. In *2014 International Conference on Computer and Communication Technology (ICCCT)*, pages 285–289. IEEE, 2014.

[31] Jennifer Golbeck, Cristina Robles, Michon Edmondson, and Karen Turner. Predicting personality from twitter. In *2011 IEEE third international conference on privacy, security, risk and trust and 2011 IEEE third international conference on social computing*, pages 149–156. IEEE, 2011.

[32] Samuel D Gosling, Peter J Rentfrow, and William B Swann Jr. A very brief measure of the big-five personality domains. *Journal of Research in personality*, 37(6):504–528, 2003.

[33] Lijie Guo. Beyond the top-n: algorithms that generate recommendations for self-actualization. In *Proceedings of the 12th acm conference on recommender systems*, pages 573–577, 2018.

[34] Sonali Gupta, Payal Gulati, Surbhi Bhatia, and Rosy Madaan. An automatic approach to music recommendations based on individual personality traits. *Available at SSRN 3565276*, 2020.

[35] Maria Halkidi, Yannis Batistakis, and Michalis Vazirgiannis. On clustering validation techniques. *Journal of intelligent information systems*, 17(2-3):107–145, 2001.

[36] Kathleen Marie Higgins. The music of our lives. 2011.

[37] Hui-Huang Hsu, Cheng-Wei Hsieh, et al. Feature selection via correlation coefficient clustering. *JSW*, 5(12):1371–1377, 2010.

[38] Nina Hubig and Claudia Plant. Information-theoretic non-redundant subspace clustering. In Jinho Kim, Kyuseok Shim, Longbing Cao, Jae-Gil Lee, Xuemin Lin, and Yang-Sae Moon, editors, *Advances in Knowledge Discovery and Data Mining - 21st Pacific-Asia Conference, PAKDD 2017, Jeju, South Korea, May 23-26, 2017, Proceedings, Part I*, volume 10234 of *Lecture Notes in Computer Science*, pages 198–209, 2017.

[39] Robert K Hull. The relationship between personality and music preference. 2009.

[40] Nezamoddin N Kachouie and Meshal Shutaywi. Weighted mutual information for aggregated kernel clustering. *Entropy*, 22(3):351, 2020.

[41] Bart P Knijnenburg, Saadhika Sivakumar, and Daricia Wilkinson. Recommender systems for self-actualization. In *Proceedings of the 10th acm conference on recommender systems*, pages 11–14, 2016.

[42] Bart P Knijnenburg and Martijn C Willemsen. Evaluating recommender systems with user experiments. In *Recommender Systems Handbook*, pages 309–352. Springer, 2015.

[43] Bart P Knijnenburg, Martijn C Willemsen, Zeno Gantner, Hakan Soncu, and Chris Newell. Explaining the user experience of recommender systems. *User Modeling and User-Adapted Interaction*, 22(4):441–504, 2012.

[44] Christopher L Knowles. The correlation of music preference and personality. 2013.

[45] Alexandra Langmeyer, Angelika Guglhör-Rudan, and Christian Tarnai. What do music preferences reveal about personality? *Journal of individual differences*, 2012.

[46] Weiwei Li, Jan Hannig, and Sayan Mukherjee. Subspace clustering through sub-clusters. *arXiv preprint arXiv:1811.06580*, 2018.

[47] Patrick Litle and Marvin Zuckerman. Sensation seeking and music preferences. *Personality and individual differences*, 7(4):575–578, 1986.

[48] Yanchi Liu, Zhongmou Li, Hui Xiong, Xuedong Gao, Junjie Wu, and Sen Wu. Understanding and enhancement of internal clustering validation measures. *IEEE transactions on cybernetics*, 43(3):982–994, 2013.

[49] Son T Mai, Xiao He, Nina Hubig, Claudia Plant, and Christian Böhm. Active density-based clustering. In *2013 IEEE 13th International Conference on Data Mining*, pages 508–517. IEEE, 2013.

[50] Dominik Mautz, Wei Ye, Claudia Plant, and Christian Böhm. Towards an optimal subspace for k-means. In *Proceedings of the 23rd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, pages 365–373, 2017.

[51] Maximilian Mayerl, Michael Vötter, Hsiao-Tzu Hung, Bo-Yu Chen, Yi-Hsuan Yang, and Eva Zangerle. Recognizing song mood and theme using convolutional recurrent neural networks. In *Working Notes Proceedings of the MediaEval 2019 Workshop, ceur-ws. org*, volume 12, 2019.

[52] William McCown, Ross Keiser, Shea Mulhearn, and David Williamson. The role of personality and gender in preference for exaggerated bass in music. *Personality and individual differences*, 23(4):543–547, 1997.

[53] Brian McFee, Colin Raffel, Dawen Liang, Daniel PW Ellis, Matt McVicar, Eric Battenberg, and Oriol Nieto. librosa: Audio and music signal analysis in python. In *Proceedings of the 14th python in science conference*, volume 8, pages 18–25. Citeseer, 2015.

[54] Benjamin Murauer, Maximilian Mayerl, Michael Tschuggnall, Eva Zangerle, Martin Pichl, and Günther Specht. Hierarchical multilabel classification and voting for genre classification. In *MediaEval*, 2017.

[55] Isabel Briggs Myers. The myers-briggs type indicator: Manual (1962). 1962.

[56] Loris Nanni, Yandre MG Costa, Rafael L Aguiar, Carlos N Silla Jr, and Sheryl Brahnam. Ensemble of deep learning, visual and acoustic features for music genre classification. *Journal of New Music Research*, 47(4):383–397, 2018.

[57] Loris Nanni, Yandre MG Costa, Alessandra Lumini, Moo Young Kim, and Seung Ryul Baek. Combining visual and acoustic features for music genre classification. *Expert Systems with Applications*, 45:108–117, 2016.

[58] Gideon Nave, Juri Minxha, David M Greenberg, Michal Kosinski, David Stillwell, and Jason Rentfrow. Musical preferences predict personality: evidence from active listening and facebook likes. *Psychological Science*, 29(7):1145–1158, 2018.

[59] Yair Neuman, Leonid Perlovsky, Yochai Cohen, and Danny Livshits. The personality of music genres. *Psychology of Music*, 44(5):1044–1057, 2016.

[60] Maria Augusta SN Nunes and Rong Hu. Personality-based recommender systems: an overview. In *Proceedings of the sixth ACM conference on Recommender systems*, pages 5–6, 2012.

[61] Lance Parsons, Ehtesham Haque, and Huan Liu. Subspace clustering for high dimensional data: a review. *Acm Sigkdd Explorations Newsletter*, 6(1):90–105, 2004.

[62] Sampo V Paunonen and Michael C Ashton. Big five factors and facets and the prediction of behavior. *Journal of personality and social psychology*, 81(3):524, 2001.

[63] Martin Pichl, Eva Zangerle, and Günther Specht. #nowplaying on #spotify: leveraging spotify information on twitter for artist recommendations. In *International Conference on Web Engineering*, pages 163–174. Springer, 2015.

[64] Martin Pichl, Eva Zangerle, and Günther Specht. Towards a context-aware music recommendation approach: What is hidden in the playlist name? In *2015 IEEE International Conference on Data Mining Workshop (ICDMW)*, pages 1360–1365. IEEE, 2015.

[65] Martin Pichl, Eva Zangerle, and Günther Specht. Understanding playlist creation on music streaming platforms. In *2016 IEEE International Symposium on Multimedia (ISM)*, pages 475–480. IEEE, 2016.

[66] Martin Pichl, Eva Zangerle, and Günther Specht. Understanding user-curated playlists on spotify: A machine learning approach. *International Journal of Multimedia Data Engineering and Management (IJMDEM)*, 8(4):44–59, 2017.

[67] Asmita Poddar, Eva Zangerle, and Yi-Hsuan Yang. #nowplaying-rs: a new benchmark dataset for building context-aware music recommender systems. In *15th Sound and Music Computing Conference*, 2018.

[68] Peter J Rentfrow, Lewis R Goldberg, and Daniel J Levitin. The structure of musical preferences: a five-factor model. *Journal of personality and social psychology*, 100(6):1139, 2011.

[69] Peter J Rentfrow and Samuel D Gosling. The do re mi's of everyday life: the structure and personality correlates of music preferences. *Journal of personality and social psychology*, 84(6):1236, 2003.

[70] D. Rönnow and Theodor Twetman. Automatic genre classification from acoustic features. 2012.

[71] Sandro Saitta, Benny Raphael, and Ian FC Smith. A comprehensive validity index for clustering. *Intelligent Data Analysis*, 12(6):529–548, 2008.

[72] Muriel Schulte. *Examining the link between personality and music preferences using clustering, feature extraction and prediction.* PhD thesis, Tilburg University, 2018.

[73] Marcin Skowron, Florian Lemmerich, Bruce Ferwerda, and Markus Schedl. Predicting genre preferences from cultural and socio-economic factors for music retrieval. In *European Conference on Information Retrieval*, pages 561–567. Springer, 2017.

[74] Mettu Srinivas, Debaditya Roy, and C Krishna Mohan. Music genre classification using online dictionary learning. In *2014 International Joint Conference on Neural Networks (IJCNN)*, pages 1937–1941. IEEE, 2014.

[75] David Tully. Examining the relationship between music preference and personality type. 2012.

[76] George Tzanetakis and Perry Cook. Musical genre classification of audio signals. *IEEE Transactions on speech and audio processing*, 10(5):293–302, 2002.

[77] Nicolas Wack, Enric Guaus, Cyril Laurier, O Meyers, R Marxer, Dmitry Bogdanov, Joan Serra, and Perfecto Herrera. Music type groupers (mtg): generic music classification algorithms. *Music Information Retrieval Evaluation eX-change (MIREX) extended abstract*, 2009.

[78] Wei Ye, Sebastian Goebl, Claudia Plant, and Christian Böhm. Fuse: Full spectral clustering. In *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, pages 1985–1994, 2016.

[79] Wei Ye, Samuel Maurus, Nina Hubig, and Claudia Plant. Generalized independent subspace clustering. In Francesco Bonchi, Josep Domingo-Ferrer, Ricardo Baeza-Yates, Zhi-Hua Zhou, and Xindong Wu, editors, *IEEE 16th International Conference on Data Mining, ICDM 2016, December 12-15, 2016, Barcelona, Spain*, pages 569–578. IEEE Computer Society, 2016.

[80] Eva Zangerle and Martin Pichl. Content-based user models: Modeling the many faces of musical preference. In *19th International Society for Music Information Retrieval Conference*, 2018.

[81] Eva Zangerle, Martin Pichl, Wolfgang Gassler, and Günther Specht. #nowplaying music dataset: Extracting listening behavior from twitter. In *Proceedings of the first international workshop on internet-scale multimedia management*, pages 21–26, 2014.

[82] Eva Zangerle, Michael Tschuggnall, Stefan Wurzinger, and Günther Specht. Analyzing coherent characteristics in music playlistsanalyzing coherent characteristics in music playlists.

[83] Eva Zangerle, Michael Tschuggnall, Stefan Wurzinger, and Günther Specht. Alf-200k: Towards extensive multimodal analyses of music tracks and playlists. In *European Conference on Information Retrieval*, pages 584–590. Springer, 2018.

[84] Eva Zangerle, Michael Vötter, Ramona Huber, and Yi-Hsuan Yang. Hit song prediction: Leveraging low-and high-level audio features. In *ISMIR*, pages 319–326, 2019.