Comparative study on Music Genre Classification using Machine Learning Models

Gitesh Deshmukh
220900436
Big Data Science
Queen Mary University of London

***Abstract*—Music genre classification, despite exhaustive research using machine learning techniques, remains a complex and intriguing domain within audio analysis. This study dives deep into this challenge, primarily leveraging the Support Vector Machine (SVM) model along with the k-NN machine learning model. The primary dataset utilized in this study is the renowned GTZAN collection, which boasts a diverse range of music genres.**

The methodology involves extracting distinctive audio features like Mel-frequency cepstral coefficients (MFCCs), Spectral Contrast, tempogram, chroma features, spectral bandwidth and rms energy.

This comprehensive study aims to explicate the dataset's nuanced characteristics and delivers an insightful assessment of the machine learning models applied within the field of music genre classification.

***Index Terms*—Music Genre Classification, Tempogram, Support Vector Machine (SVM), Mel-frequency Cepstral Coefficients (MFCCs)**

I. INTRODUCTION

Music genre classification arises as a platform for the fascinating intersection of arts and data science. We now face issues that are as much about analyzing musical qualities as they are about data analysis, thanks to the progress of technology and its interaction with music. This fusion not only highlights the interdisciplinary nature of modern problems but also emphasizes that genre classification complexities span both in musical and data realm.

Music's intersection with data analysis exemplifies the blending of arts and sciences, highlighting the interdisciplinary challenges we face today. This amalgamation reveals that the complexities inherent to music genre classification are not solely musical challenges; they are data challenges. Researchers are prompted to understand nuances that traverse both musical and data domains, necessitating a comprehensive viewpoint.This multidisciplinary approach highlights the importance of domain-specific expertise in data activities and advocates for collaborations that connect artists with data specialists.

Each musical sample is an audio signal at its core, composed of distinctive features such as frequency, spectral roll-off, RMS level, bandwidth, zero-crossing rate, among others .These attributes form a structured representation, facilitating the task for machine learning models.[1] Computers process audio signals in formats like wav or mp3, converting them into structured data amenable for machine learning algorithms. Through rigorous analysis, this data can assist in the precise classification of music into distinct genres, thereby enhancing the level of details and the accuracy of genre catalogs on platforms such as Spotify and Apple Music.

Moreover, appropriate genre classification has extensive consequences. The implications are substantial, ranging from enhancing algorithmic music recommendations to aiding in music production and even understanding cultural music evolution. Maintaining accuracy and precision in the attempt has the potential to transform how we engage with music in the digital age. For the detailed classification required in this domain, the dataset choice is fundamental. For this research, I have selected the well-regarded GTZAN dataset, comprising a diverse collection of 1,000 audio tracks distributed evenly across ten genres.[3] My preference for the GTZAN dataset arises from its apt genre representation and its notable mention in previous scholarly endeavors. Furthermore, the fine-tuning of audio features it encompasses makes it ideal for detailed machine learning evaluations.[1]

Among numerous machine learning algorithms, Support Vector Machines (SVM) and k-nearest neighbors (k-NN) stand out for some reasons.SVM efficiently processes high-dimensional data, which is prevalent in audio signals with complex auditory features. On the other hand, k-NN's non-parametric approach adeptly discerns the subtleties of music genres using proximity-based classifications. Combined, these methods integrate the best of boundary and proximity-driven techniques.

Building upon this foundation, the study extends beyond mere categorization. It aims to discern the intricate distinctions delinating one genre from another genre and to comprehend the elements that resonate with listeners on a profound level. Music, at its core, transcends the mere intersection of rhythms and harmonies, embodying a fusion of varied cultures, historical backgrounds, and societal intricacies. Rising trends in digital music consumption underscore the value of precise genre classification. Tailored user experiences, guided by accurate genre recommendations, enhance user engagement and satisfaction on various platforms.

Identify applicable funding agency here. If none, delete this.

In our modern age, music has become an omnipresent force, weaving its melodies and rhythms into the fabric of daily life. The ubiquity of portable devices and streaming platforms has democratized music consumption, granting listeners access to a vast array of genres and artists at their fingertips. This democratization, while a boon for music enthusiasts, poses challenges for service providers. The volume and variety of available music necessitate efficient and precise genre classification systems, ensuring that listeners can easily navigate expansive libraries and discover new tracks that align with their tastes.

When employing machine learning models for this task, they inevitably encounter challenges. These include managing imbalanced datasets, where some genres might be underrepresented, and addressing the inherent subjectivity of genre demarcations. [8] Addressing a track that blends multiple genres presents complexities, emphasizing the need for resilient and adaptive models capable of identifying nuanced patterns.

As with any technological advancement, the automated classification of music genres invites a consideration of its ethical dimensions. While machine learning models strive for objectivity, they are trained on data that may inherently carry biases—reflecting historical imbalances or prevalent societal norms. Thus, there's a critical need to ensure that these models, in their quest for accuracy, do not perpetuate or exacerbate existing biases. Moreover, as genres evolve and intermingle, strict classifications risk pigeonholing artists and their creations. It is vital to approach genre classification with a measure of flexibility, acknowledging the fluidity of musical expression and its resistance to rigid categorization.

In the fluid world of music, genres are not static entities. They are constantly in a state of flux, influenced by cultural, social, and technological shifts. As music undergoes rigorous analytical processes, the lines demarcating one genre from another begin to blur. Traditional boundaries that once clearly defined genres become increasingly porous.

Sophisticated analytical tools allow for a granular examination of musical elements. These investigations often unveil a tapestry of influences and styles within what might have been previously categorized as a singular genre. Such tools highlight that many tracks contain elements from diverse styles, even if they were initially perceived to belong exclusively to one genre. A prime example of this blending is Reggaeton, a genre that seamlessly fuses the distinct elements of reggae and hip-hop.

This merging of styles is not merely an analytical observation but a reflection of the evolving nature of music. The emergence of hybrid musical forms stands as a testament to the globalized world in which we live. In this interconnected era, cross-cultural exchanges have become the norm, profoundly shaping the realm of artistic expression.

This dissertation ventures beyond the mere confines of categorization, aspiring to weave together rigorous data analysis with profound musical recognition. Following the consequent sections, with all-inclusive literature surveys to methodological explorations, I aim to seamlessly connect data and music within the modern digital context.

II. LITERATURE SURVEY

The music industry, in its evolution, has witnessed rapid advancements driven by technological innovations. This progress has captivated scholars and researchers worldwide. Understanding the intricate nuances embedded within musical compositions, the academic community has delved deep, exploring a myriad of models and techniques. Their primary goal is to refine genre classification and meticulously extract insights from audio signals.

Ghildiyal et al. (2021) made a notable contribution by leveraging the GTZAN dataset, which is considered a benchmark in music classification. By employing various algorithms, including CNN, they achieved an impressive 91% accuracy in their genre classification efforts[1]. Their approach illuminated the ease of distinguishing between genres with distinct acoustic features, such as classical and blues. However, challenges were evident when addressing genres with shared acoustic properties like country and rock.

In another study, Chillara et al. (2019) honed a comprehensive research with a deep dive into extracting pivotal features that spanned both time and frequency domains. This led to the identification of nine salient features. The juxtaposition of spectrogram-driven models with feature-centric models set this work apart for the audio analysis[2]. The results were enhancing, with the achieved accuracies underscoring the versatility and efficacy of adopting diverse modeling strategies in the domain of music genre classification. Further emphasizing the significance of features, Li et al. (2003) delved deep into the intricacies of feature extraction methodologies specific to music genre categorization. Their findings emphasized the pivotal role of FFT and MFCC features, clearly demonstrating their edge over other features, including Beat and Pitch[11]. A central observation was the greater impact of adept feature selection on classification results as opposed to merely focusing on the choice of classification algorithms. Through their systematic assessment, the study sheds light on the vital importance of precision in feature extraction within the broader scope of music genre classification.

In an insightful study, Tzanetakis and Cook (2002) ventured into evaluating the real-world applicability of automated genre classification. The results were promising, with the system's outcomes demonstrating remarkable parity with human judgment levels. By leveraging a concoction of unique feature sets that encapsulated timbral texture, rhythmic content, and pitch variations, the study succeeded in achieving classification accuracies of 61% in non-real-time scenarios and 44% when tested in real-time environments[3]. They emphasized the distinction in performance between controlled settings and real-time scenarios in genre classification models, underscoring the pivotal role of real-world applicability testing.

Delving into machine learning's prowess, Xu et al. (2003) conceptualized and implemented an SVM-based technique tailored for genre classification. Their methodology entailed an initial segregation of musical compositions into broad

categories such as classic, jazz, pop, and rock. This hierarchical strategy proved advantageous, outstripping several conventional techniques in music genre categorization[4].

Meng et al. (2007) implemented cepstral coefficients and temporal features for musical instrument recognition. Their work inspired me to investigate feature extraction approaches and the classifier building. This methodology provided a brighter insights into instrument qualities by merging many features.[7]

Emphasizing deep learning,Bisharad and Laskar (2019) adopted the GTZAN dataset to design a genre classification system anchored around a Residual Neural Network. Interestingly, the model was trained on 3-second music segments but was tested on full-length 30-second clips. The results were noteworthy, with system accuracies clocking in at 82%, 91%, and 94.5% for top-1, top-2, and top-3 genre predictions[5]. This reinforced the observation that certain genres, like classical and jazz, were inherently more distinguishable than others, such as rock.

In a departure from traditional methodologies, Dieleman and Schrauwen (2014) introduced an innovative approach where models were designed to learn directly from raw musical audio. This method significantly reduced the reliance on manual feature extraction, offering a refreshing take on music genre classification[9].

The tools shaping the industry also underwent evolution, with Chawla et al. (2002) introducing "SMOTE," a novel technique to address imbalanced datasets commonly encountered in music genre classification endeavors[8]. McFee et al. (2015) introduced "librosa," a Python package for music and audio analysis[13]. This framework provides tools for various music analysis tasks, making it a cornerstone for researchers and professionals who approach to dissect and understand musical pieces, enhancing genre classification among other tasks.

While the primary focus of this work is on music genre classification and the associated algorithms, it's worth noting the broader advancements in the realm of machine learning. For instance, in the field of Natural Language Processing (NLP), Kalyan, Rajasekharan, and Sangeetha (2021) undertook a comprehensive exploration of Transformer-based pretrained models titled Ammus: A survey of transformer-based pretrained models in natural language processing"[6]. Their deep dive into the state-of-the-art models sheds light on the potential for leveraging such advanced architectures in related domains. While they mainly looked at text data, the ideas and methods they shared could also be useful for audio studies and music genre sorting. This is because both NLP and audio processing deal with data in sequences.

Meng et al. (2007) examined the temporal feature integration for music genre classification, introducing the novel DAR and MAR features. Leveraging MFCC as a primary short-time feature, the MAR features outperformed traditional methods, especially when paired with classifiers like SVM. Their approach, emphasizing inter-feature relationships and temporal correlations, showcases potential in enhancing genre classification accuracy and aligns with broader MIR applications[10].

III. METHODOLOGY

Music genre classification is a multifaceted endeavor that intricately interweaves data, features, and machine learning models. In this study,I explore this intricate landscape, leveraging the GTZAN dataset and extracting salient audio features to classify genres. This section elucidates the methodology adopted in this research.

A. Dataset Preparation

GTZAN Dataset Acquisition:

The GTZAN dataset, crafted by George Tzanetakis in 2002, has established itself as a keystone in music genre classification research. Over the time, its importance has proliferated as several researchers utilised it in the domain of music and audio analysis. This carefully curated dataset encompasses a total of 1,000 audio snippets, each precisely 30 seconds in duration. These songs cover ten carefully selected musical genres, guaranteeing a wide representation of various musical styles. The balanced constitution of the dataset, with each genre being equally represented, provides a reliable and objective environment for analysis and comparison. This consistency in genre distribution not only reduces biases but also makes it possible for more consistent and repeatable research results. For a easy understanding of its composition, the specific genre-wise distribution is delineated in the subsequent table:

TABLE I
NUMBER OF AUDIO CLIPS IN EACH GENRE

S. No	Class	Clips
1	Blues	100
2	Classical	100
3	Country	100
4	Disco	100
5	Hiphop	100
6	Jazz	100
7	Metal	100
8	Pop	100
9	Reggae	100
10	Rock	100
Total		1000

B. Feature Extraction

Feature extraction stands at the core of the music genre classification. Employing the GTZAN dataset as the bedrock, I embarked on a journey to discern vital characteristics from audio samples. We extracted the following essential features:

- Mel-frequency cepstral coefficients (MFCCs): Integral to the short-term power spectrum of sound, these coefficients play a pivotal role in discriminating musical genres.[11]
- Chroma: Representing the twelve distinct pitch classes, Chroma sheds light on the very basics of harmonics.
- Spectral Contrast: A measure of the 'brightness' of a sound, it locates the center of gravity of the spectrum.

- Tempogram: A diagnostic tool for rhythm, revealing the tempo nuances of music pieces.
- RMS Energy: This metric measures the energy of the sound signal, offering insights into the perceived volume or intensity of a track.
- Spectral Bandwidth: Quantifying the spectrum's width, this provides an understanding of an audio clip's timbral texture.

C. Data Structuring

Loading & Pre-processing: Having curated the extracted features into the 'gtzan.csv' dataset, a meticulous division ensued, segmenting the data into features and targets. To streamline subsequent machine learning tasks, musical genres were transmuted into numerical values. This dichotomy between features and targets established a robust framework, priming the dataset for efficient model training and evaluation.

D. Data Refinement

Data Splitting & Scaling: The quintessence of any machine learning model lies in its ability to generalize. To achieve this, I employed stratified sampling, ensuring an even genre representation during both training and validation phases. Furthermore, to put all features on an equal footing, I standardized them, making the data amenable for in-depth analysis and training.

E. Balancing & Feature Selection

To address the ever-pervasive challenge of class imbalances in datasets, I opted the Synthetic Minority Over-sampling Technique (SMOTE). This technique enriches the dataset by fabricating synthetic samples, ensuring a harmonious genre representation. In the quest for model excellence, I integrated RFECV (Recursive Feature Elimination with Cross-Validation). This approach shifted through the features, spotlighting those critical for the model's predictive acumen.

F. Models

a) Formulating the SVM Model and hyperparameter Optimization: Constructing an SVM model that efficiently handles the high-dimensional data offered by the GTZAN dataset needs a nuanced approach to hyperparameter optimization The initial step follows the separation of features and targets from the dataset. Python libraries such as numpy and pandas are quintessential for this purpose, helping in data manipulation and data analysis. Following data separation, one of the first and most important procedures is to encode genre labels into numbers using LabelEncoder. This is an important step since machine learning models understand numbers better than language.

The subsequent step for the implementation involves generating the polynomial features with the PolynomialFeatures method. This process captures the interactions among distinct features, adding depth to the dataset's dimensionality. While this additional complexity may initially raise concerns about high dimensionality, SVM's kernel technique thrives on exactly this type of data structure.

Upon dividing the dataset into training and testing sets, the standardization of the data took place using StandardScaler. This ensures that all features are on a consistent scale, essential for distance-based algorithms like SVM.

A major difficulty, though, is the potential imbalance in genre classes. To combat this, the SMOTE method is used, which makes samples in minority classes synthetically, balancing the representation.[8]

The pivotal features is critical to the development of certain models and choosing the important ones and leveraging it becomes indespensable. The relevant method here to implement is called Recursive Feature Elimination with Cross-Validation (RFECV). This method subtly combines feature selection with cross-validation and switches through a plethora of features, assessing their significance and zeroing in on those that have a significant impact on the conclusion.

The deliberate use of RandomForestClassifier as the principal estimator accentuates the method's precision. Random Forest has the ability to sort the features according to the collective reduction in impurity that stems from every tree within the ensemble. A technique of this level becomes necessary while navigating through various elements and with their complicated associations. The way computing over several trees enables a thorough and reliable assessment of feature significance.

Support Vector Machines (SVM) are notable in machine learning for their power and versatility. The thorough construction of hyperparameters is critical to their efficacy. Establishing a complete hyperparameter grid is critical for honing an SVM model to top performance. This grid was not created at random; it is a calculated framework that spans a large number of combinations, encompassing variations of several critical parameters.

The regularization parameter, typically abbreviated as 'C,' is one of these factors. In the world of SVM, the letter 'C' plays an important role. It achieves a balance between gaining a large margin and ensuring correct training example categorization. Depending on the value of 'C,' the model can either choose a larger margin, potentially accepting some misclassifications for the sake of generalization, or it can choose a narrower margin.

Kernel types introduce another dimension of complexity. Kernels transform the input space, potentially projecting it to higher dimensions, thereby enabling the SVM to tackle non-linear problems. The choice between linear, polynomial, radial basis function (RBF), and sigmoid kernels, among others, can significantly alter the model's perception of data.

Individual training instances' influence is determined by gamma values, which are especially important when using the RBF kernel. A decision influenced by a high gamma value will prioritize close locations, whereas a lower value will prioritize distant points as well. This hyperparameter acts as a gatekeeper, determining the amount of influence each data piece has.

Finally, the polynomial degree is taken into account, especially when using the polynomial kernel. This component

determines the intricacy of the decision boundaries and can improve the SVM's ability to recognize intricate patterns in data.

Setting the right balance of these hyperparameters is sensitive. Here, GridSearchCV proves indespensable. It systematically explores the vast configuration landscape defined by the grid, gauging each combination's virtue. Bias from random data division is a potential issue. GridSearchCV works in conjunction with stratified K-Fold cross-validation to mitigate this issue. This approach not only partitions the data into 'K' subsets but also ensures each subset accurately mirrors the overall dataset, especially concerning genre distribution

To summarize, understanding the complexity of hyperparameter tuning is just as crucial as understanding the mechanics of the SVM model-perfecting process. It's a delicate and calculated dance.

b) Harnessing the SVM's Kernel Trick:

The provided script prominently highlights SVM's unparalleled ability to manage intricate non-linear patterns. A standout aspect, the kernel trick, becomes the centerpiece in the hyperparameter optimization phase. Within the vast hyperparameter grid, the kernel option stands out by offering linear, RBF, polynomial, and sigmoid functions as possibilities. This diversity not only alludes to the range of transformations the data might undergo but also the aspiration to ensure data becomes discernible in a transformed space, often of higher dimensionality.

c) Augmenting Predictions: The Ensemble Method with a Voting Classifier:

While SVM holds its prominence, the script doesn't shy away from employing ensemble techniques. The VotingClassifier is introduced as a fusion of the fine-tuned SVM and a RandomForest classifier. This methodology, rooted in a majority voting mechanism, amalgamates predictions from both models, leaning towards the majority's decision. Such a collaborative approach is strategic, pooling the distinct strengths of the SVM and RandomForest. The overarching goal is clear: enhancing the predictive efficacy by mitigating individual model limitations.

d) Scrutinizing Model Performance: Comprehensive Evaluation Metrics:

As the culmination of this machine learning journey, the evaluation phase is crucial. Here, tools like the confusion matrix and classification report are not mere analytical devices; they become the yardstick measuring the model's predictive acumen against new, unseen data. With precision, recall, F1-score, and a broader accuracy metric in play, a holistic picture emerges. This in-depth analysis provides not just a snapshot of the model's capabilities but a panoramic view of its strength and areas ripe for further refinement.

1) k-NN Model Configuration and Hyperparameter Tuning: In the process of developing the k-NN model, it's pivotal to recognize that every dataset has its own nuances and characteristics. Therefore, a generic approach to hyperparameter optimization might not suffice. Instead, the approach utilized in this study is founded upon a dedication to formulating a customized and all-encompassing methodology.

The extensive and deliberate development of the code as part of this initiative demonstrates a blend of technical expertise and contextual awareness. Rather than using pre-developed solutions, the decision to methodically study and fine-tune the hyperparameters emphasizes the scope of the research. Each stage, and subsequently each decision is supported by observational findings as well as methodological considerations, ensuring that the final k-NN model is both robust and finely tuned to the specific problems and subtleties of the dataset under consideration.

Number of Neighbors: The number of neighbors, as the distinguishing pillar of the k-NN technique, is vital. The model investigates a balanced adjustment between sensitivity to noise and generalized predictions by evaluating a range from 1 to 30.

Weight Allocation: I encountered two paradigms – uniform, where each neighbor wields equal influence, and distance, a dynamic paradigm where influence diminishes with increasing distance from the query point.

Metric Selection: An age-old debate in k-NN's realm is the distance metric employed. Here, the tug-of-war is between the geometrically intuitive euclidean distance and the grid-based manhattan distance. Remarkably, the optimal outcome leans towards the manhattan metric, emphasizing the unique intricacies of the dataset in question. The culmination of this phase materializes with the identification of 6 neighbors, utilizing a distance-based weighting scheme under the manhattan metric's watchful gaze.

Fortifying k-NN through Bagging: Beyond the standalone capabilities of k-NN, there's a significant elevation in model robustness when integrated with the Bagging classifier. Bagging—short for Bootstrap Aggregating is an innovation geared towards diminishing model variance. By creating multiple bootstrapped datasets and subsequently training the k-NN on each, Bagging ingeniously averages out individual models' predictions. This strategic pooling inherently reduces overfitting, setting the stage for more stable and consistent predictions.

In sum, while the k-NN stands as a testament to instance-based learning's potential, its true prowess shines when amalgamated with ensemble strategies and critically evaluated through diverse metrics.

IV. RESULTS & COMPARATIVE STUDY

The objective of the research was to delve into the complexities of music genre classification, a challenging task given the intricate and multi-dimensional nature of audio data.The study utilized two machine learning classifiers, namely, the Support Vector Machine (SVM) and k-Nearest Neighbors (k-NN). These classifiers were chosen based on their diverse mechanisms of operation—SVM operates in the realm of maximizing margins between data points of different classes, while k-NN functions on the principle of similarity, where

classification decisions are driven by the proximity of data points in the feature space.

In this section, a comprehensive analysis and discussion on the performances of the SVM and k-NN classifiers is presented. The aim is to provide insights into the effectiveness of each classifier in the domain of music genre classification, elucidate the strengths and weaknesses inherent to each approach, and establish potential avenues for further enhancement in classification accuracy and efficiency.

Performance Metrics: Precision: 75.5% Recall: 74.5% F1 Score: 74.7% Accuracy: 74.5%

The Support Vector Machine(SVM) model achieved an aggregate accuracy of 74.5% on test dataset. While this implies a good level of genre recognition, it's imperative to delve deeper into the genre-specific metrics to better understand the model's performance nuances.

While for the k-NN model, I achieved the accuracy of 67%. The precision, recall, and F1 scores are 0.6612, 0.67, and 0.6567 respectively.

Given the complexity of a 10-class classification, the k-NN model showcased a commendable performance with an accuracy of 67%, signifying its capability in discerning intricate patterns across diverse musical genre.

To further dissect the model's nuances, the confusion matrix serves as an quintessential tool, illuminating genre-specific strengths and weakness.

A. Confusion Matrix Analysis for SVM

The confusion matrix is instrumental in visualizing the performance of an algorithm. Presented below in Table II is the confusion matrix for the SVM model on the GTZAN dataset with hyperparameters $C = 10$, $\gamma = 0.1$, and kernel set to 'rbf'.

TABLE II
CONFUSION MATRIX OF THE SVM MODEL WITH $C = 10$, $\gamma = 0.1$, AND KERNEL 'RBF'

	Predicted									
Actual	Blues	Class.	Country	Disco	Hip-hop	Jazz	Metal	Pop	Reggae	Rock
Blues	15	0	0	0	0	2	0	0	2	1
Class.	0	19	0	0	0	0	0	1	0	0
Country	2	0	16	1	1	0	0	0	0	0
Disco	0	0	0	17	1	0	0	0	1	1
Hip-hop	0	0	1	0	14	0	2	0	1	2
Jazz	0	1	3	0	0	13	0	0	1	2
Metal	1	0	0	0	0	0	14	0	1	4
Pop	0	0	0	2	0	0	0	17	1	0
Reggae	0	0	0	1	1	0	0	1	15	2
Rock	3	0	1	2	0	1	0	1	3	9

The confusion matrix provides an essential perspective for evaluating the efficacy of a machine learning algorithm. As depicted in Table II, the confusion matrix clarifies the performance of the SVM model when tested on the GTZAN dataset. This model, distinguished by hyperparameters $C = 10$ $\gamma = 0.1$ with the kernel set to rbf, showcases patterns that merit thorough analysis.

Detailed Support Vector Machine (SVM) Analysis:

1) High-Dimensional Handling:: The SVM's intrinsic ability to handle high-dimensional data is particularly advantageous for music genre classification. Given that music data can encompass a vast range of features, from tempo and rhythm patterns to spectral characteristics and harmonic content, SVM's capability in such spaces is evident through its achieved accuracy.

2) Kernel Trick and Decision Boundaries:: The choice of the 'rbf' kernel was crucial for the SVM model's performance. The Radial Basis Function (RBF) kernel can transform the data into a higher-dimensional space, allowing for more complex decision boundaries. This transformation aids in segregating genres that might be harder to distinguish in the original feature space.

3) Robustness to Outliers:: SVM's structure, especially with a well-tuned regularization parameter C, ensures that the model is resilient to outliers. This resilience can be pivotal when dealing with music data, where anomalous tracks or mislabeled samples might exist.

4) Genre-Specific Performances:: A closer inspection of genre-specific metrics reveals SVM's adaptability. For genres like Classical, which possess pronounced musical characteristics, the model's exceptional precision suggests its efficacy in identifying and leveraging these genre-defining features. In contrast, for genres with overlapping characteristics, the SVM still maintains a decent performance, emphasizing its balanced genre recognition capability.

5) Hyperparameter Influence:: The hyperparameters C and γ in an SVM play significant roles in defining the model's flexibility. A well-tuned C strikes a balance between maximizing the margin and minimizing classification errors, while γ determines the influence of a single training sample. Their values, $C = 10$ and $\gamma = 0.1$ in this study, appear to be well-suited for the data, resulting in a robust model performance.

In summation, the SVM model showcases a commendable ability in distinguishing among diverse music genres. While there are nuanced overlaps between some genres, these intricacies only underscore the rich tapestry of musical elements embedded within each song. Given the promising results achieved, it's evident that in the ever-evolving realm of machine learning, there's always potential for refining models and achieving even greater accuracy. Fine-tuning hyperparameters, innovating in feature engineering, or venturing into alternative modeling strategies can further amplify the model's strengths.

B. Confusion Matrix Analysis for k-NN

The confusion matrix provides an insight into the performance of an algorithm. Presented below in Table III is the confusion matrix for the k-NN model on the GTZAN dataset with parameters metric set to 'manhattan', $n\_neighbors = 6$, and weights set to 'distance'.

The confusion matrix for the k-NN (k-Nearest Neighbors) model on the GTZAN dataset offers valuable insights into the model's performance across various music genres. The k-NN model is configured with a 'Manhattan' distance metric, six neighbors, and distance-based weighting. The model exhibits near-perfect accuracy for classical genre, signifying its strong performance for genres with distinct characteristics. However, the confusion matrix also reveals some areas for improvement.

TABLE III
CONFUSION MATRIX OF THE K-NN MODEL WITH METRIC 'MANHATTAN', $n\_neighbors = 6$, AND WEIGHTS 'DISTANCE'

Actual	Predicted									
	Blues	Class.	Country	Disco	Hip-hop	Jazz	Metal	Pop	Reggae	Rock
Blues	10	0	3	1	0	1	0	0	2	3
Class.	0	19	0	0	0	0	0	1	0	0
Country	1	0	17	1	0	0	0	1	0	0
Disco	0	0	1	13	1	0	0	1	3	1
Hip-hop	1	0	1	0	14	0	1	0	2	1
Jazz	1	1	3	0	0	13	0	1	0	1
Metal	0	0	0	0	0	1	17	0	1	1
Pop	0	0	0	1	2	0	0	16	0	1
Reggae	0	0	2	0	2	0	1	1	13	1
Rock	1	0	5	2	1	1	4	2	2	2

Genres like Blues and Reggae show higher rates of misclassification, often confused with Country, Disco, and Rock. overall, the k-NN model's ability to distinguish between a wide range of music genres demonstrates its promise and possibility for further development in future versions.

C. Challenges and Forward Path for my models:

While SVM showcased commendable results, it's worth noting potential challenges. Training times for SVMs can be extensive, especially with larger datasets. Also, SVM's performance heavily relies on the quality of features extracted from the music data. Hence, investing in feature engineering and selection could potentially uplift the model's future performance.

In summation, the SVM's performance in this study underscores its strengths in handling complex musical data. Its combination of high-dimensional data handling, resilience to outliers, and adaptability to genre nuances makes it a formidable model for music genre classification tasks.

While k-NN demonstrates promise in genre differentiation, it faces hurdles with genres exhibiting shared musical elements. Optimizing feature extraction and considering hybrid models could be the next steps for enhancing its performance.To enhance the model's discriminatory capabilities, a deeper dive into feature engineering and a more extensive parameter tuning are recommended. Additionally, exploring ensemble methods that combine k-NN with other classifiers might offer improved accuracy and nuanced genre recognition

In the assessment of the model's proficiency across diverse music genres, certain genres distinctly demonstrated superior recognition accuracy. Classical music, for instance, recorded an outstanding precision and recall, both measuring at 0.95. This suggesting that the model effectively discerns the intricate melodies, harmonies, and specific instrumentations that characterize classical pieces. The Disco genre, characterized by its rhythmic tempo and distinct beats, achieved a commendable recall of 0.85, further evidencing the model's capacity to differentiate unique musical attributes. Moreover, the Metal genre reported a precision of 0.88.

The observed metrics emphasize the model's prowess in genre classification and highlight its ability to differentiate its nuanced variances within musical compositions. The pervasive high scores across the majority of genres reinforce the notion that an SVM model, given a well-preprocessed dataset, stands as a potent tool for distinguishing between a wide array of musical genres suggesting the model is adept at identifying patterns unique to these genres.

The SVM model showcases competent proficiency across a myriad of genres post preprocessing method, thereby emphasizing the model's capacity as an efficient tool for musical file classification. Delving into genre-specific performances (refer to Table - performance/ metrics), it's intriguing to observe that Disco registers a recall of 0.85. This suggests that the model accurately identified 85% of the genuine Disco samples.

Comparative Performance Analysis

A tabular representation is provided below to offer a comparative view of the performance metrics achieved by both SVM and k-NN models.

TABLE IV
RESULT ANALYSIS OF MY MODELS

Metric	SVM	k-NN
Accuracy	74.5%	67%
Precision	75.5%	66.12%
Recall	74.5%	67%
F1 Score	74.7%	65.67%

The confusion matrix stands as an invaluable tool in understanding the intricacies of how well a machine learning model fares. Taking a moment to gaze upon Table III, the k-NN model's performance on the GTZAN dataset unfurls before us. It's especially noteworthy that this model is fine-tuned with the metric set to 'manhattan', $n\_neighbors = 6$, and weights calibrated to distance.

Delving deeper into the matrix's results reveals the varying degrees of precision, the k-NN model demonstrates across musical genres. For genres like Classical and Country, the model illustrates an outstanding proficiency. The Classical genre comes close to impeccable classification, with only a minor lapse where it's classified as Pop. This indicates that the classical tracks in the GTZAN dataset have such distinct characteristics that they are rarely confused with other genres.

While some genres are clearly and precisely characterized, the model encounters complex issues with others. Consider the 'Blues' genre: it exhibits a broad range of misclassifications, demonstrating that it has been periodically confounded with Country, Disco, and even Rock. This behavior suggests that blues records in the GTZAN dataset may share key musical characteristics with these genres, putting the k-NN's differentiating capabilities to the test. Similarly, the Rock genre unfolds a complicated mesh of accurate and incorrect identifications, highlighting potential overlaps and shared traits in the dataset's rock songs.

Another genre that beckons attention is Reggae. Here, the model seems to grapple with differentiating it from 'Hip-hop' and even 'Metal'. Such confusions shed light on the potential similarities in rhythm patterns, beats, or even instrumentation that may exist between these genres in the GTZAN collection.

Upon consideration, analyzing the k-NN model in relation with the manhattan metric and its distinctive weighting

approach yields good outcomes and some challenges. The confusion matrix underscores the duality faced in music classification while there's the potential for remarkable accuracy, and on the other, the mutable boundaries of music genres, shaped by their profound depth and varied expressions. Based on this analysis, it can be inferred that music classification through machine learning navigates the fine line between artistry and algorithmic precision.

D. Comparative with Other Models

TABLE V
COMPARATIVE ANALYSIS WITH OTHER MODELS.

Model	**Model Used**	**Accuracy %**
[1]	Support Vector Machine	68.9
[1]	Convolution Neural Network	91
[12]	K Nearest Neighbours	61.50
[12]	Random Forest	61.90
[2]	Logistic Regression	60.89
[12]	Support Vector Machine	72.60

The investigation into the effectiveness of various machine learning algorithms for music genre classification reveals distinct performance metrics. As illustrated in Table V, the Support Vector Machine (SVM), proposed by Ghildiyal et al. [1], delivers decent performance with an accuracy of 68.9%. This achievement signifies SVM's inherent capability to discern intricate patterns within musical data. In contrast, using a Convolutional Neural Network, they achieved a notably higher accuracy of 91%. This result underscores the potential of neural network architectures in processing multi-dimensional audio data.

K Nearest Neighbours and Random Forest, influenced by the findings of Elbir et al. [13], present accuracies of 61.50% and 61.90%, respectively. These metrics suggest potential challenges these algorithms encounter in achieving optimal generalization across varied musical genres.

The Logistic Regression model, adapted from Chillara et al.'s work [2], registers an accuracy of 60.89%. Such a metric implies the model's limitations in capturing the nuances of the dataset, especially in contrast to more advanced models.

Elbir et al. (2018) reported an SVM achieving an accuracy of up to 72.70%[12], which is comparable to my results, underlining SVM's robustness in music genre classification.

Within the intricate domain of music genre classification, achieving exceptional accuracy remains a formidale task. However, as evidenced by Table V, there's evident progression in the field. Although advances bring their own set of challenges, the benchmarks set in my study highlight the significant potential of blending machine learning with the intricacies of musical artistry. My results, which compare favorably with many established methodologies, underscore the transformative power of machine learning in redefining music genre classification. This investigation not only illuminates the capabilities of current techniques but also sets the stage for future innovations. As demonstrated by my findings, combining music with machine learning suggests a promising future for better genre recognition and beyond.

V. CONCLUSION

Navigating the challenges of music genre classification necessitated an examination of multiple machine learning methodologies, taking GTZAN dataset as the foundation. This dive into the subtle domain of audio processing and genre delineation not only revealed interesting pathways, but also stressed the promise of infocentric methodologies in the advancement of music analytics.

While the initial exploration encompassed the SVM and k-NN models, the remarkable competency of the Support Vector Machine (SVM) emerged distinctly. With an impressive accuracy of 74.5%, SVM set itself apart, demonstrating its expertise, especially when processing intricate audio data.The prudent selection of audio features significantly enhanced the model's ability to process and classify data accurately.

By considering genre-specific metrics and embracing an encompassing evaluation method that integrated precision, recall, and the F1-score, a comprehensive view of performance was achieved. This method emphasized SVM's capability to grapple with classification intricacies. The implications of this prowess are substantial for platforms like music streaming services, heralding an era characterized by tailored music suggestions, advanced search mechanisms, and enhanced user engagement.

In a comparative analysis with models such as k-NN, SVM's elevated efficiency is evident. Even with k-NN's notable performance, SVM's superiority, as validated by its good metric performance, strengthens its position as a benchmark model for this dataset.

In conclusion, this study demonstrates the importance of SVM in music genre classification. A promising future exists at the intersection of musical creativity and cutting-edge machine learning, with the potential to change the digital audio universe and boost user experiences.

VI. FUTURE WORK

The trajectory recorded by this research acknowledges existing advancements while spotlighting potential future endeavors. The use of deep learning in music genre classification, though not novel, leaves ample room for the future innovation. Deep learning's inherent capability to determine the complex patterns has already solidified its role in numerous classification assignments. Nevertheless, the potential to calibrate these techniques, capitalizing on evolving neural designs and training strategies, remains vast.

The progression of music genre classification necessitates a persistent reevaluation and expansion of the auditory features analyzed. By harnessing advanced signal processing methods, future endeavors can extract richer auditory descriptors. These refined features could offer a profound understanding of unique attributes across musical genres, playing a pivotal role in enhancing the process of classification.

VII. ACKNOWLEDGMENT

I would like to extend my heartfelt gratitude to my supervisor, Dalia Senvaitye. Her professional guidance and detailed instructions help me to perform with better mindset and gain overall valuable insights about the project.

REFERENCES

[1] Ghildiyal, A., Singh, K. and Sharma, S., 2020, November. Music genre classification using machine learning. In 2020 4th international conference on electronics, communication and aerospace technology (ICECA) (pp. 1368-1372). IEEE

[2] Chillara, S., Kavitha, A.S., Neginhal, S.A., Haldia, S. and Vidyullatha, K.S., 2019. Music genre classification using machine learning algorithms: a comparison. Int Res J Eng Technol, 6(5), pp.851-858.

[3] Tzanetakis, G. and Cook, P., 2002. Musical genre classification of audio signals. IEEE Transactions on speech and audio processing, 10(5), pp.293-302.

[4] Xu, C., Maddage, N.C., Shao, X., Cao, F. and Tian, Q., 2003, April. Musical genre classification using support vector machines. In 2003 IEEE International Conference on Acoustics, Speech, and Signal Processing, 2003. Proceedings.(ICASSP'03). (Vol. 5, pp. V-429). IEEE.

[5] Bisharad, D. and Laskar, R.H., 2019, October. Music genre recognition using residual neural networks. In TENCON 2019-2019 IEEE Region 10 Conference (TENCON) (pp. 2063-2068). IEEE.

[6] Kalyan, K.S., Rajasekharan, A. and Sangeetha, S., 2021. Ammus: A survey of transformer-based pretrained models in natural language processing. arXiv preprint arXiv:2108.05542.

[7] Meng, A., Ahrendt, P., Larsen, J. and Hansen, L.K., 2007. Temporal feature integration for music genre classification. IEEE Transactions on Audio, Speech, and Language Processing, 15(5), pp.1654-1664.

[8] Chawla, N.V., Bowyer, K.W., Hall, L.O. and Kegelmeyer, W.P., 2002. SMOTE: synthetic minority oversampling technique. Journal of artificial intelligence research, 16, pp.321-357.

[9] Dieleman, S. and Schrauwen, B., 2014, May. End-to-end learning for music audio. In 2014 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP) (pp. 6964-6968). IEEE.

[10] Meng, A., Ahrendt, P., Larsen, J. and Hansen, L.K., 2007. Temporal feature integration for music genre classification. IEEE Transactions on Audio, Speech, and Language Processing, 15(5), pp.1654-1664.

[11] Li, T., Ogihara, M. and Li, Q., 2003, July. A comparative study on content-based music genre classification. In Proceedings of the 26th annual international ACM SIGIR conference on Research and development in informaion retrieval (pp. 282-289).

[12] Elbir, A., Çam, H.B., Iyican, M.E., Öztürk, B. and Aydin, N., 2018, October. Music genre classification and recommendation by using machine learning techniques. In 2018 Innovations in intelligent systems and applications conference (ASYU) (pp. 1-5). IEEE.

[13] McFee, B., Raffel, C., Liang, D., Ellis, D.P., McVicar, M., Battenberg, E. and Nieto, O., 2015, July. librosa: Audio and music signal analysis in python. In Proceedings of the 14th python in science conference (Vol. 8, pp. 18-25)