IJITCE

ISSN 2347- 3657

International Journal of Information Technology & Computer Engineering

Email : ijitce.editor@gmail.com or editor@ijitce.com

International Journal of Information Technology & Computer Engineering

ISSN 2347–3657

Volume 13, Issue 2, 2025

ANALYZING MUSIC STREAMING TRENDS WITH SPOTIFY DATA BY MACHINE LEARNING

Gandham Teja Chandra Mouli, **21PD1A0525**, Department of CSE, West Godavari Institute Of Science and Engineering, Andhra Pradesh, India.

Pachigolla Sowbhagya, 21PD1A0561, Department of CSE, West Godavari Institute Of Science and Engineering, Andhra Pradesh, India.

Gotham Anand Satya Konda, 21PD1A0527, Department of CSE, West Godavari Institute Of Science and Engineering, Andhra Pradesh, India.

J.Eswar Chandu, 22PD5A0529, Department of CSE, West Godavari Institute Of Science and Engineering, Andhra Pradesh, India.

Yepuganti Veera Vignan Dinesh Rahul, 22PD5A0530, Department of CSE, West Godavari Institute Of Science and Engineering, Andhra Pradesh, India

Dr. M.Aravind Kumar, Professor, Department of ECE, West Godavari Institute Of Science and Engineering, Andhra Pradesh, India.

Dr. M.Aravind Kumar, Professor, Department of ECE, West Godavari Institute Of Science and Engineering, Andhra Pradesh, India.

ABSTRACT

Understanding music streaming trends is essential for identifying user preferences, refining recommendation systems, and guiding strategic decisions in the music industry. This study utilizes Spotify data and applies machine learning models—including Regression, Decision Trees, and Random Forest—to forecast streaming trends and uncover the factors driving song popularity. Key variables such as genre, artist, release year, and user engagement metrics are analyzed to determine their impact on a song's success. Regression models reveal linear relationships, while Decision Tree and Random Forest models capture complex interactions for improved predictive accuracy. The dataset undergoes thorough preprocessing to eliminate inconsistencies and enhance model performance. Visualizations are used to interpret listener behavior, streaming patterns, and music attributes associated with popular tracks. By incorporating machine learning, this study boosts the effectiveness of trend prediction and personalized recommendations, offering actionable insights for music platforms, artists, and marketers to drive user engagement and satisfaction.

Problem Statement

With the rapid growth of digital music streaming platforms, understanding user preferences and accurately predicting music trends have become essential for boosting user engagement and refining recommendation systems. However, analyzing vast and complex_streaming datasets presents a significant challenge, as traditional analytical methods often fall short in capturing the intricate relationships between user behavior, song attributes, and listening patterns.

This project addresses these challenges by leveraging machine learning techniques—such as Regression, Decision Trees, and Random Forest—to analyze Spotify streaming data. The

ISSN 2347–3657
Volume 13, Issue 2, 2025

objective is to uncover the key factors driving song popularity, including listener behavior, genre trends, and engagement metrics. The insights derived will help music platforms, artists, and marketers make informed, data-driven decisions. Ultimately, the project aims to enhance the accuracy of recommendation systems, improve music marketing strategies, and deliver a better overall experience for listeners in the digital music landscape.

INTRODUCTION

The growth of music streaming platforms like Spotify has revolutionized how audiences discover and engage with music, making it crucial to derive meaningful insights from user data. Understanding what drives song popularity—such as genre, artist influence, and engagement metrics—requires more than traditional analysis. This project applies machine learning techniques, including Regression, Decision Trees, and Random Forest, to predict and analyze streaming trends. Regression models identify relationships between song features and stream counts, while Decision Tree-based models handle complex interactions for more accurate predictions. Through careful data preprocessing and model training, the system generates reliable insights into listener behavior and trending music attributes. These insights help platforms optimize recommendation engines, personalize user experiences, and guide artists and marketers in making strategic, data-driven decisions.

EXISTING SYSTEM

Current methods for predicting song popularity rely on analyzing a variety of audio and user-related features through machine learning techniques. Researchers typically source data from platforms like Kaggle and use web scraping tools to gather detailed Spotify metrics. Most studies emphasize core audio features—such as Danceability, Tempo, Energy, Loudness, Speechiness, Acousticness, Instrumentalness, Liveness, and Valence—to determine what influences a song's success. Both supervised and unsupervised learning algorithms have been applied, with regression models used to explore linear relationships and Decision Tree or Random Forest models employed to capture more complex feature interactions. Clustering methods have also been used to group songs with similar attributes. While these approaches offer valuable insights, they often face limitations in prediction accuracy and adaptability. Future improvements may include the use of deep learning and neural networks for enhanced performance, along with refined feature selection and loss function optimization. Additionally, this study seeks to explore changes in user song preferences before and after the COVID-19 pandemic, providing a unique perspective on evolving listening behavior.

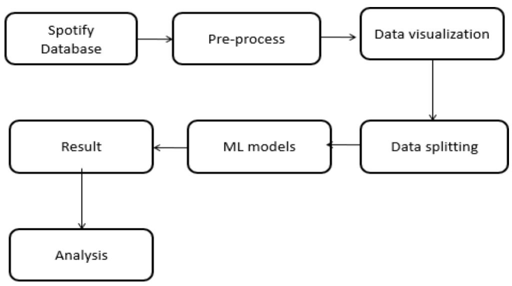
PROPOSED METHOD

The proposed approach employs machine learning models—Regression, Decision Tree, and Random Forest—to analyze and predict music streaming trends using Spotify data. The process begins with data collection and preprocessing, which involves handling missing values, normalizing numerical features, and encoding categorical

International Journal of Information Technology & Computer Engineering
ISSN 2347–3657
Volume 13, Issue 2, 2025

variables to prepare the dataset for modeling. Regression is used to explore linear relationships between song attributes such as genre, tempo, and duration and their corresponding streaming counts. Decision Tree models provide interpretable segmentations, highlighting the key factors that contribute to a song's popularity. Random Forest, as an ensemble method, enhances predictive accuracy by mitigating overfitting and capturing complex feature interactions. Model performance is assessed using evaluation metrics like Mean Absolute Error (MAE) and Root Mean Squared Error (RMSE). Visualization techniques are integrated to interpret listener behavior, identify trends, and highlight influential song characteristics. These insights support the improvement of recommendation systems and the development of targeted marketing strategies. Overall, this machine learning-driven methodology enables more accurate forecasting of music trends and delivers actionable insights for streaming platforms, artists, and industry professionals.

ARCHITECTURE



The methodology begins by collecting music streaming data from sources like the Spotify API, Kaggle datasets, and web scraping, encompassing track metadata, audio features, user engagement metrics, and temporal data. Preprocessing involves cleaning the data by handling missing values, removing duplicates, normalizing numerical values, encoding categorical variables, and eliminating outliers to ensure model readiness. Feature engineering focuses on extracting influential attributes such as stream counts, audio characteristics, release timing, genre, and social media influence, with techniques like correlation analysis, PCA, and RFE used for optimal feature selection. The selected data is then used to train machine learning models—Linear Regression, Decision Tree, and Random Forest—with an 80/20 training-testing split and hyperparameter tuning through Grid Search and K-Fold Cross-Validation to boost accuracy. Models are evaluated using MAE, RMSE, R^2 for regression, and metrics like accuracy, precision, recall, F1-score, and ROC analysis for classification. Visualization tools such as time series plots, heatmaps, bar graphs, and scatter plots are applied to interpret patterns in listener behavior and song popularity. Finally, the best model is deployed as an API or dashboard integrated with Spotify for real-time predictions, with future enhancements planned through deeper learning models and more dynamic data integration.

CONCLUSION

This study highlights the effectiveness of machine learning in uncovering music streaming trends using Spotify data. By applying models like Regression, Decision Tree, and Random Forest, it identifies key factors influencing song popularity, such as genre, artist, release year, and user engagement. Through careful preprocessing, feature selection, and model evaluation, accurate and reliable predictions were achieved. The insights gained support personalized recommendations and strategic decisions for music platforms, artists, and marketers, while visualizations of listener behavior further enhance understanding of audience preferences.

FUTURE SCOPE

While this study offers meaningful insights into music trend prediction, there are several opportunities for enhancement to improve the depth and accuracy of the analysis. Incorporating advanced machine learning models such as Recurrent Neural Networks (RNNs) and Long Short-Term Memory (LSTM) networks can significantly enhance the forecasting of time-based streaming trends. Additionally, hybrid approaches that combine traditional machine learning with deep learning techniques may lead to more accurate and robust predictions. Real-time data integration from the Spotify API can further refine model responsiveness, allowing it to adapt quickly to new releases and shifting listener preferences. Moreover, the inclusion of social media data from platforms like Twitter, YouTube, and Instagram can provide valuable context on how external factors, such as viral trends and artist engagement, influence song popularity.

REFERENCES

1. **Schedl et al. (2018)** – Provides a comprehensive overview of recent advancements and practical applications in Music Information Retrieval (MIR), covering techniques for audio analysis, user modeling, and music recommendation.
2. **Ferwerda et al. (2017)** – Explores the correlation between personality traits and music genre preferences, revealing how music tastes vary with age and gender demographics.
3. **Herremans et al. (2017)** – Proposes a functional taxonomy for categorizing music generation systems, offering a structured framework to evaluate and compare generative music models.
4. **Schedl (2019)** – Introduces the LFM-2b dataset, comprising 2 billion music listening events, aimed at supporting large-scale analysis in MIR research and personalized recommendation systems.
5. **Jiang et al. (2021)** – Demonstrates the use of machine learning algorithms to predict song popularity on Spotify, emphasizing features like audio content, artist metadata, and user engagement metrics.
6. **Van den Oord et al. (2013)** – Develops a deep learning-based content model for music recommendation, improving personalization by learning high-level audio features.
7. **Koops et al. (2020)** – Surveys machine learning approaches for music genre classification, discussing challenges, datasets, and evaluation methods in the field.