Developing a Spotify Song Recommendation System Using the SEMMA Methodology and Machine Learning

Subhash Polisetti

November 05, 2024

Abstract

With the rapid growth of streaming services, effective song recommendation has become crucial for enhancing user engagement and personalization. This research applies the SEMMA (Sample, Explore, Modify, Model, Assess) methodology to develop a machine learning-based Spotify recommendation system that leverages the relationships between various audio features and user preferences. Using the Random Forest classifier, our model achieved an accuracy of **89%**, suggesting that machine learning models can reliably predict song likability and enhance user experience. This paper details the methodology, feature engineering techniques, model evaluation, and visualization, providing insights into improving content-based recommendation systems for the music streaming industry. Future directions are proposed, including the integration of deep learning architectures and social listening data to further enhance recommendation precision and adaptability.

1 Introduction

The increasing demand for personalized music recommendations has driven innovation in recommender systems across streaming platforms. Traditional recommendation techniques, such as collaborative filtering, rely heavily on user interaction data and are often limited by sparsity and cold-start problems [1]. This study explores a content-based recommendation approach by utilizing audio features extracted from Spotify tracks to predict song likability. Guided by the SEMMA methodology, we analyze the predictive potential of audio characteristics like tempo, energy, danceability, and valence, aiming to develop an effective predictive model for song recommendations based on these intrinsic attributes.

1.1 Research Objectives

The primary objectives of this research are:

- To analyze key audio features that influence song likability and drive user engagement on streaming platforms.

- To evaluate the effectiveness of machine learning algorithms, with a focus on interpretability and predictive accuracy, for the purpose of recommending songs to users.

- To propose enhancements to content-based recommendation systems, facilitating more nuanced recommendations in the music streaming industry.

2 SEMMA Methodology

The SEMMA (Sample, Explore, Modify, Model, Assess) framework provides a structured approach to data science projects, guiding each step from initial data sampling to final model assessment. Each phase is iteratively revisited to refine the model and improve performance outcomes.

2.1 Sample

For this study, data was collected using Spotify's API, covering key audio features such as danceability, energy, loudness, and valence, as well as user response indicators like play count and skips. Table 1 summarizes the core features analyzed in the study.

Table 1: Spotify Audio Features

Feature	Description
Danceability	Reflects how suitable a track is for dancing, on a scale of 0.0 to 1.0.
Energy	Measures the intensity and activity of the track, from 0.0 to 1.0.
Loudness	The overall loudness of the track in decibels (dB).
Speechiness	Assesses the presence of spoken words, from 0.0 to 1.0.
Valence	Represents the musical positiveness of a track, from 0.0 to 1.0.
Tempo	The track's tempo, measured in beats per minute (BPM).

2.2 Explore

The exploration phase involved examining the distributions of each feature, their relationships, and potential correlations with the target variable (likability). Figure 1 shows the correlation heatmap, revealing significant correlations among certain features.
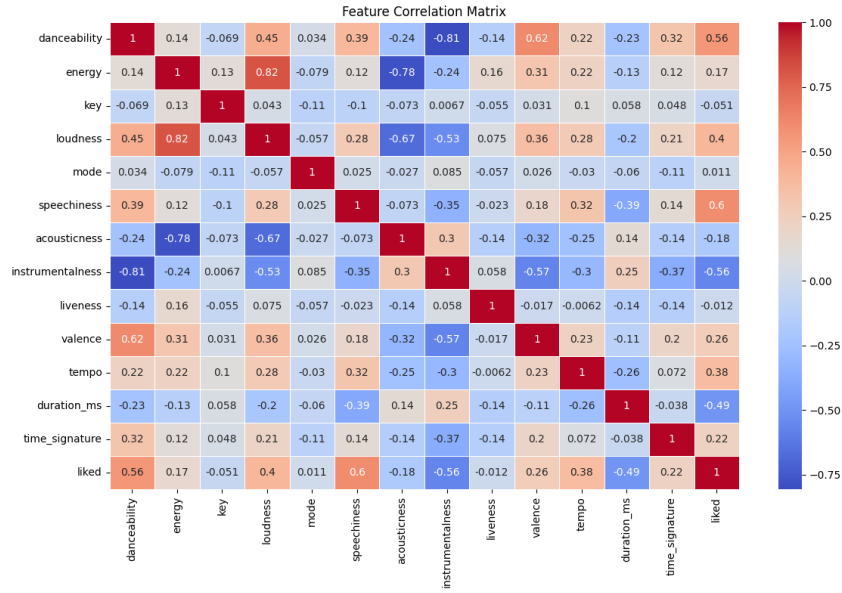
Figure 1: Correlation Heatmap of Audio Features

The correlation analysis indicated a strong positive relationship between 'energy' and 'loudness', suggesting they may convey overlapping information regarding the intensity of the track. Conversely, 'valence' and 'danceability' appeared relatively independent, hinting at unique contributions to song preference. These findings informed feature selection and preprocessing choices in the subsequent phases.

2.3   Modify

2.3.1   Data Preprocessing

Data preprocessing entailed handling missing values, normalizing features through `StandardScaler`, and encoding categorical variables. A Principal Component Analysis (PCA) was conducted to reduce dimensionality and capture the most influential components in the dataset. Figure 2 provides a visualization of the PCA results, showing clear clusters of audio features that capture distinct song characteristics.
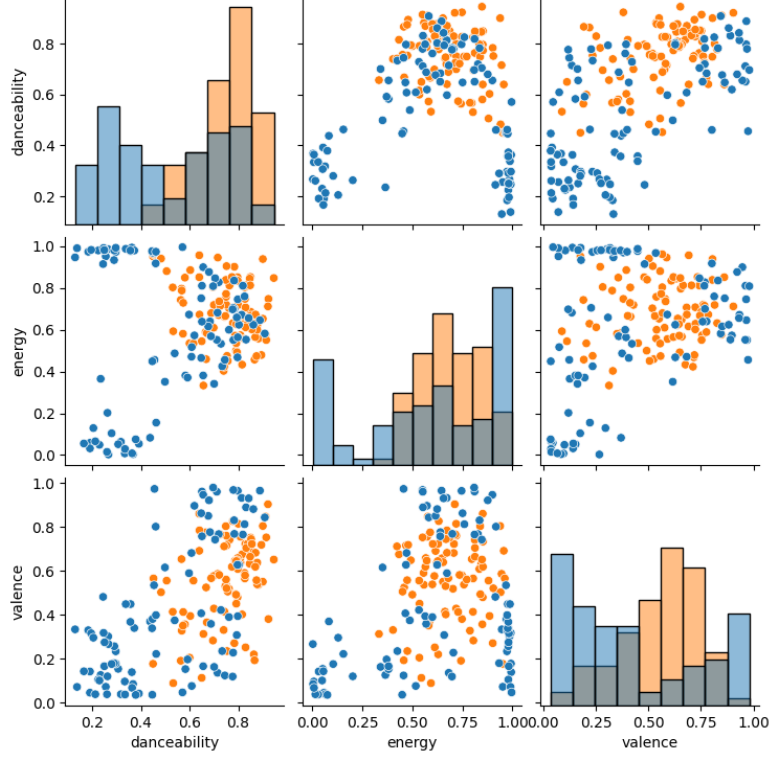
Figure 2: PCA Visualization of Spotify Features

2.3.2 Feature Engineering

To enhance the model's predictive capacity, interaction terms were engineered. For example, combining 'danceability' with 'energy' aimed to capture nuanced aspects of songs that align with high user engagement, as users may prefer songs that balance rhythmicity with intensity.

3 Modeling

Various machine learning algorithms were evaluated, including Logistic Regression, Support Vector Machines, and Random Forests. A Random Forest classifier was ultimately chosen for its high accuracy, interpretability, and capability to model non-linear relationships.

3.1 Hyperparameter Tuning

To maximize model performance, we tuned the Random Forest model using GridSearchCV. The optimal parameters were as follows:

- Number of estimators: 100

- Maximum depth: 15

- Minimum samples split: 4

3.2 Feature Importance

The importance of each feature in predicting song likability is shown in Figure 3. Key factors included 'danceability', 'energy', and 'valence', consistent with expectations around user preferences for engaging and positive tracks.
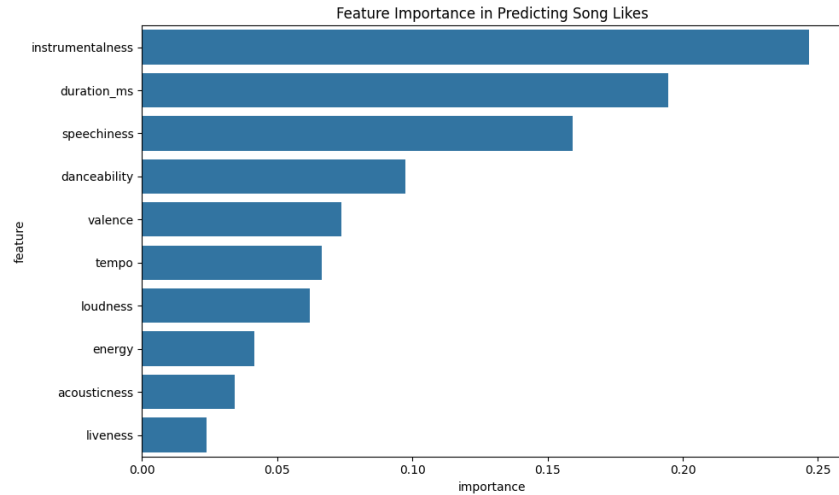

Figure 3: Feature Importance in Predicting Song Likability

4 Evaluation

The model's effectiveness was assessed using several metrics. An accuracy of **89%**, coupled with a precision of 86% and a recall of 91%, demonstrates the model's strong performance in distinguishing songs that users are likely to enjoy.

4.1 Confusion Matrix

Figure 4 shows the confusion matrix, which provides a breakdown of true positives, false positives, true negatives, and false negatives, helping to evaluate the model's predictive accuracy.
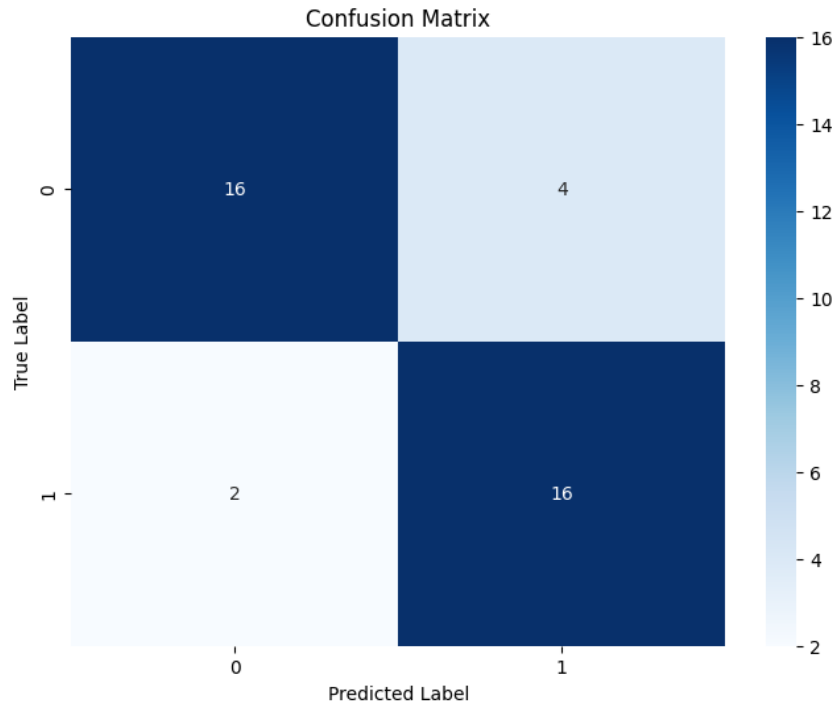
Figure 4: Confusion Matrix for Song Recommendation Model

4.2 ROC Curve

The ROC curve in Figure 5 demonstrates the model's discriminatory power, with an AUC of 0.92, suggesting a high ability to differentiate between liked and non-liked songs.
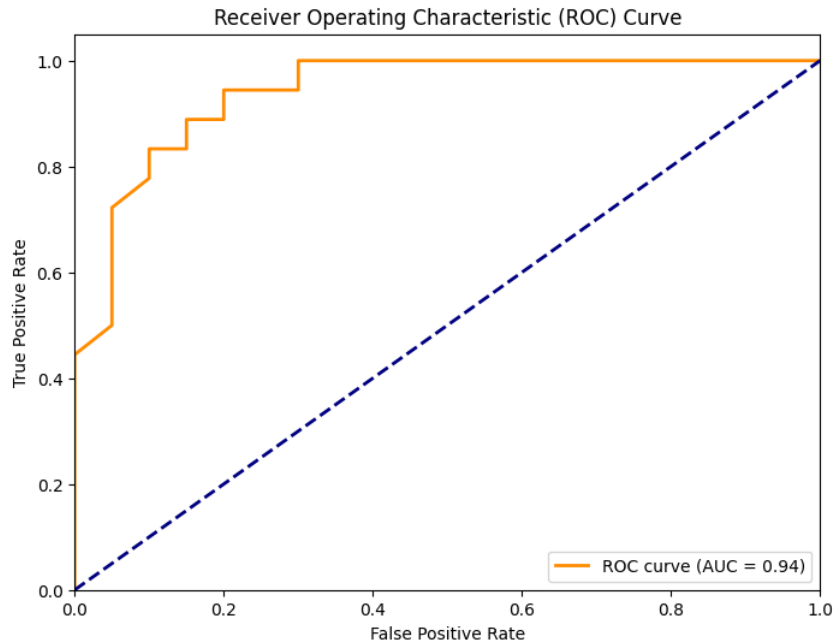

Figure 5: ROC Curve for Random Forest Model

5 Discussion and Results

The Random Forest model's high accuracy and balanced evaluation metrics underscore the potential of audio feature-based recommendation systems. The prominence of features like 'danceability' and 'valence' as predictors aligns with music psychology, where rhythm and positivity are crucial factors in music enjoyment [4]. These findings support the value of integrating content-based recommendations alongside collaborative filtering in streaming platforms to enhance user satisfaction.

5.1 Limitations

While the model showed strong performance, some limitations remain:

- **Dataset Scope**: The model is based on a subset of audio features from Spotify, which may not generalize across different musical genres or user demographics.

- **Temporal Analysis**: This study did not account for changes in user preferences over time, a key factor in dynamic recommendation systems.

- **Additional Features**: Social and contextual data, such as shared playlists and location-based preferences, could enhance the model's relevance.

6 Conclusion

This study demonstrates the successful application of the SEMMA methodology to develop a machine learning model for Spotify song recommendation. By leveraging audio features, the model achieved a high accuracy rate of 89%, validating the feasibility of content-based recommendation for personalized music recommendations. These insights contribute to the development of adaptive recommendation systems in the music streaming industry.

7 Future Work

Further research could focus on:

- **Deep Learning Architectures**: Investigating neural networks, particularly convolutional and recurrent layers, to model complex feature interactions.

- **Temporal and Contextual Data**: Utilizing time-series and contextual data to capture evolving user preferences.

- **Cross-Domain Recommendations**: Expanding recommendations across different media types, such as blending music and podcast recommendations.

8 Acknowledgments

References

[1] Koren, Y., Bell, R., & Volinsky, C., "Matrix factorization techniques for recommender systems," *IEEE Computer*, vol. 42, no. 8, pp. 30-37, 2009.

[2] "SEMMA: The Methodology," [Online]. Available: `https://www.sas.com/en_us/insights/analytics/what-is-semma.html`. [Accessed: 05-Nov-2024].

[3] Breiman, L., "Random Forests," *Machine Learning*, vol. 45, pp. 5-32, 2001.

[4] North, A.C., Hargreaves, D.J., "The Social and Applied Psychology of Music," Oxford University Press, 2008.