Predicting Spotify Track Popularity Using Audio Features

Lara El-Aloul
DATASCI 3000
Western University
London, Canada
Student ID: 251214185

Sarah Huang
DATASCI 3000
Western University
London, Canada
Student ID: 251215072

Vivek Jariwala
DATASCI 3000
Western University
London, Canada
Student ID: 251213353

Swathi Thushiyandan
DATASCI 3000
Western University
London, Canada
Student ID: 251202930

***Abstract*—The rise of music streaming has created extensive datasets that capture detailed information about songs, genres, and listener preferences. Despite this, understanding what drives track popularity remains difficult because popularity is influenced by many musical and non-musical factors. This report investigates how audio features and genre characteristics contribute to Spotify track popularity. Using a dataset of tracks with both numerical audio descriptors and categorical metadata, we perform data preprocessing, exploratory data analysis, regression modeling, and classification modeling. Results show that predicting exact popularity scores is challenging, but classification of popular versus non-popular tracks achieves strong performance. Feature importance analysis highlights loudness, danceability, and energy as positive predictors, while instrumentalness has a strong negative association. Findings suggest that audio features alone cannot fully explain popularity, but they offer meaningful insights into listener engagement and music production patterns. The complete source code, including a detailed record of individual group member contributions, is available at the following GitHub repository.**

***Index Terms*—Spotify, Machine Learning, Audio Features, Regression, Classification**

I. Introduction

Streaming platforms contain millions of songs across a wide range of genres, and listeners generate continuous engagement data as they interact with music. However, despite this abundance of data, the question of why certain songs become popular remains largely unresolved. Popularity is shaped by musical traits, production styles, social trends, and promotional factors, which makes it difficult to predict. Audio features available through Spotify offer a structured way to examine this question since they describe measurable aspects of musical sound, including danceability, valence, instrumentalness, and loudness.

This project examines which audio features and genre characteristics are most predictive of a track's popularity. Understanding these relationships is valuable for several stakeholders, particularly in solving the "cold start" problem in recommendation systems—how to recommend new music that has no historical playback data. By relying on intrinsic audio features, platforms can estimate the potential appeal of a new release before it accrues user interaction data.

The goal of this study is to analyze the Spotify Tracks dataset and determine which audio and genre-related features best predict popularity. We develop regression models that estimate the continuous popularity score, classification models that categorize tracks into popular and non-popular groups, and feature importance analyses to identify the strongest predictors. The results contribute to a more informed understanding of the acoustic factors that shape listener engagement.

II. Dataset Description

The dataset contains Spotify track metadata and audio features. It includes a popularity score ranging from 0 to 100, continuous numerical attributes, and categorical metadata fields such as genre. The presence of 125 genres offers wide musical diversity and allows for genre-specific analysis.

A. Features

The numerical features include:

- Danceability
- Energy
- Valence
- Loudness
- Tempo
- Acousticness
- Instrumentalness
- Speechiness
- Duration

These features capture musical and acoustic characteristics that may influence how listeners perceive a track.

Categorical features include:

- Genre
- Key
- Mode
- Time signature

These variables represent structural and contextual elements of tracks. Genre is particularly important due to distinct musical conventions and audience preferences.

B. Target Variables

Two target variables were used for modeling:

- The continuous popularity score (0-100).
- A binary popularity label, where tracks in the top 30 percent are considered popular.

C. Dataset Limitations

The dataset lacks artist-level metadata, promotional information, and temporal release data. Popularity is influenced by many external factors not captured in audio features, and genre imbalance introduces modeling challenges. These limitations constrain the maximum achievable accuracy of predictive models.

III. Methodology

The methodology follows a full supervised learning workflow, including preprocessing, exploratory data analysis, feature engineering, model development, and evaluation. The approach integrates linear and non-linear modeling techniques and uses multiple validation procedures.

A. Preprocessing Pipeline

Preprocessing included:

- Loading and inspecting the dataset to identify schema inconsistencies.
- Removing duplicates to prevent data leakage between train and test sets.
- Handling missing or invalid values.
- Normalizing numerical features to ensure that variables with large ranges (like Duration) do not dominate distance-based algorithms.
- One-hot encoding of categorical variables.
- Creating the binary popularity label.
- Splitting data into training, validation, and test sets.

B. Exploratory Data Analysis (EDA)

EDA provided insights into feature distributions and predictive patterns. As shown in Fig. 1, popularity is strongly skewed, with most tracks occupying lower ranges.

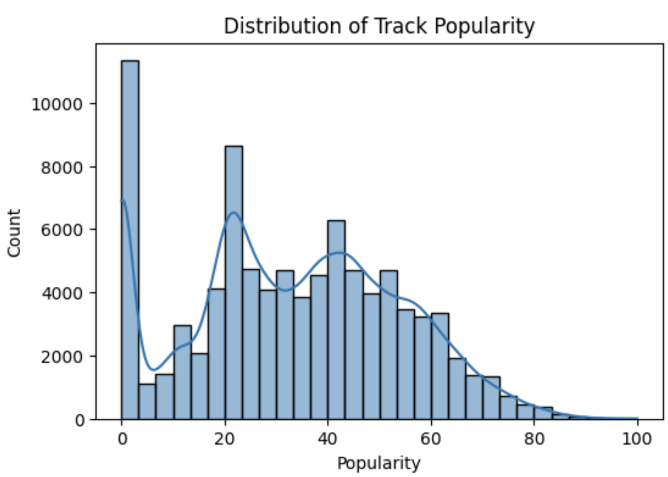

Fig. 1. Distribution of Track Popularity. The data is heavily skewed towards lower popularity scores.

Key observations from the correlation analysis (Fig. 2) include:

- Popular tracks tend to be louder, more danceable, and less instrumental.
- Instrumentalness has the strongest negative correlation with popularity.
- Loudness and danceability show moderate positive correlations.
- Genres such as K-pop, pop-film, metal, and latino tend to have higher average popularity.

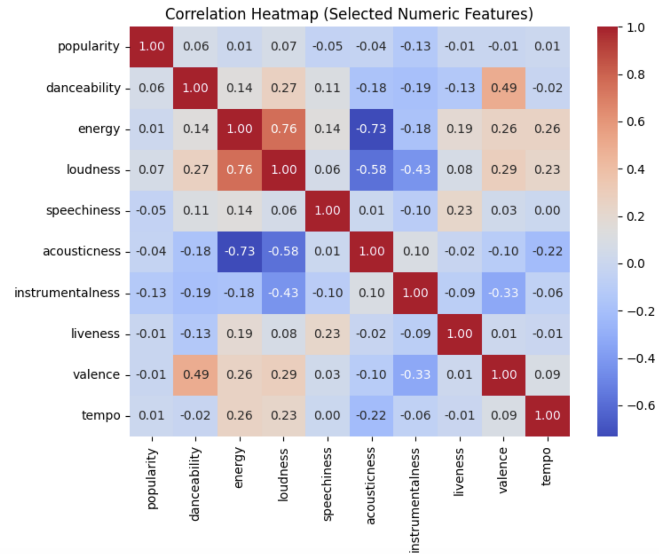

Fig. 2. Correlation Heatmap of Selected Numeric Features.

C. Feature Engineering

To capture complex relationships that simple linear models might miss, we introduced polynomial transformations (e.g., tempo squared, valence squared) and interaction features. For instance, the interaction between energy and loudness was explored, as high-energy tracks are almost invariably loud in modern production. By explicitly engineering these terms, we provided the models with non-linear signals that better represent the complexity of musical perception.

D. Model Architectures

We employed a diverse set of algorithms to tackle both the regression and classification tasks.

1) Regression Models: The following regression models were implemented:

- Linear Regression
- Ridge Regression
- Lasso Regression
- Decision Tree Regression
- Random Forest Regression

Models were evaluated using RMSE, MAE, and R^2.

2) Classification Models: Classification models included:

- Logistic Regression
- Decision Tree Classifier
- Random Forest Classifier
- K-Nearest Neighbour (KNN) Classifier

Models were evaluated using accuracy, precision, recall, F1 score, and ROC-AUC.

E. Model Evaluation

Train-test splitting and cross-validation were used to evaluate generalization. Hyperparameters were tuned, and training versus validation performance was analyzed to detect overfitting.

IV. Results

A. Regression Results

Regression models demonstrated moderate predictive power. The summary metrics are presented in Table I, with full details available in Appendix Table III. Linear and Ridge Regression performed best ($R^2 \approx 0.33$).

The relatively low R^2 values across all regression models highlight a fundamental limitation: acoustic features alone cannot account for external drivers of popularity such as artist reputation, marketing budgets, and viral social media trends. The fact that the Random Forest Regressor ($R^2 \approx 0.24$) performed worse than linear models suggests that the relationship between specific audio features and the exact popularity score is not heavily dependent on complex decision boundaries, but rather on general linear trends (e.g., louder is generally more popular).

TABLE I
Regression Models Summary (Selected Metrics)

Model	RMSE	MAE	R^2
Linear Regression	16.86	12.03	0.327
Ridge (alpha=1.0)	16.86	12.03	0.327
Random Forest Regressor	17.91	13.83	0.240
Lasso (alpha=0.1)	18.56	15.02	0.184

B. Classification Results

Classification results were significantly stronger than regression results. Table II details the performance metrics. The Random Forest Classifier achieved the highest accuracy and ROC-AUC score.

TABLE II
Classification Models Summary (Selected Metrics)

Model	Acc	Prec	F1	ROC-AUC
Random Forest	0.801	0.758	0.630	0.868
Logistic Regression	0.773	0.679	0.591	0.829
KNN (k=15)	0.751	0.641	0.541	0.793
Decision Tree	0.741	0.807	0.358	0.601

It is important to analyze the trade-off between Precision and Recall in the Random Forest model. While the model achieved high Precision (0.758), the Recall was lower (0.539). This indicates that the model is "conservative"—it is very accurate when it explicitly claims a song is popular, but it fails to identify a significant number of actual popular songs (false negatives). This behavior is preferable in scenarios where false positives (recommending a bad song) are more costly to user experience than false negatives. The ROC Curve in Fig. 4 further validates the model's strong separation capability.

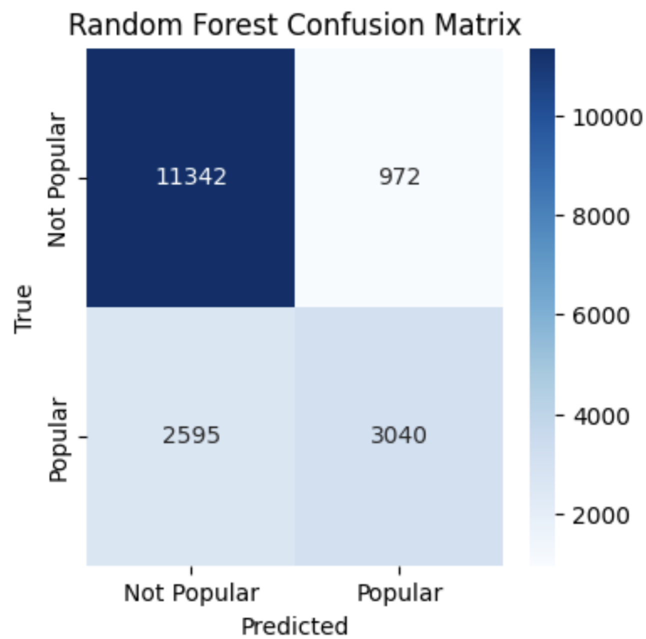

Fig. 3. Random Forest Confusion Matrix. The model shows stronger performance in correctly identifying Non-Popular tracks.

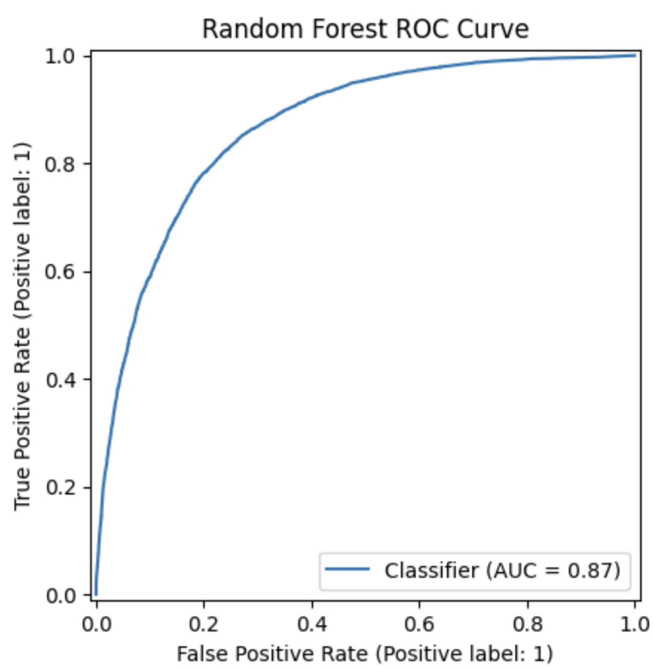

Fig. 4. Random Forest ROC Curve. An AUC of 0.87 indicates strong classification performance.

C. Feature Importance

Across high-performing models, several features emerged as consistently influential. The Random Forest Classifier identified basic audio features as the primary drivers of popularity, with specific genres playing a secondary but notable role.

1) Top Audio Features: As seen in Fig. 5, the most critical features include Duration, Loudness, Valence, and Energy. Instrumentalness also showed a strong negative impact on popularity, reflecting listener preference for vocal-driven music.

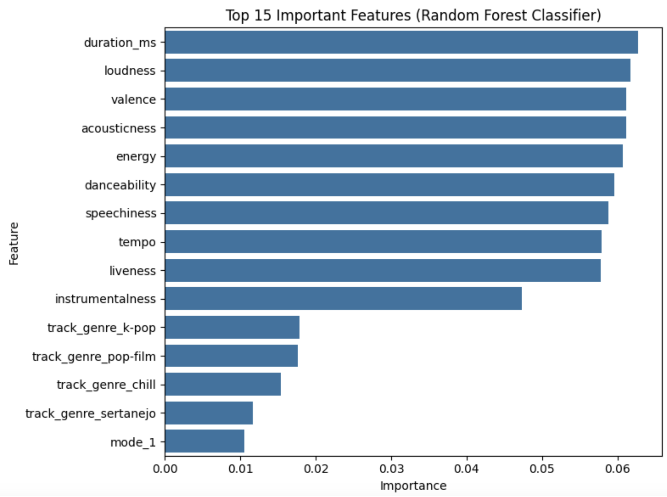

Fig. 5. Top 15 Important Features (Random Forest Classifier).

2) Top Genre Predictors: While audio features dominated the top 10, several genre indicators appeared in the top 20 most important features. K-Pop (0.018), Pop-Film (0.018), and Chill (0.015) were the most predictive genres for popularity. This suggests that while the acoustic quality of a song is paramount, the genre tag serves as a significant signal for listener engagement. A full breakdown of feature importance, including genre-specific metrics, is provided in Appendix Table V.

V. Discussion

A. Interpretation of Findings

The positive associations of loudness, energy, and danceability align with common trends in modern popular music, where energetic and rhythm-focused tracks dominate. The negative impact of instrumentalness suggests that vocal presence remains a strong predictor of widespread appeal.

Furthermore, the high importance of Loudness is consistent with "The Loudness War," a phenomenon where mastering engineers maximize track volume to compete for listener attention, often at the cost of dynamic range [2].

Similarly, the appearance of Duration as a top predictor is likely reflective of the streaming era's economic incentives. As noted by industry analysts, payout models that monetize streams after a 30-second threshold encourage artists to produce shorter tracks [3].

Linear models performed better than tree-based models for regression, which indicates that the relationships among audio features are mostly linear when predicting raw scores. However, the non-linear Random Forest model excelled in classification, suggesting it could better handle the threshold distinctions between hit songs and obscure tracks.

B. Implications

Music Industry: Artists and producers may use these insights to design tracks with stronger commercial appeal. Music platforms can integrate these models into recommendation engines to improve song discovery. Understanding feature importance can also help identify early-stage trends across genres.

C. Limitations

- Absence of artist-level metadata limits predictive power.
- Popularity is influenced by non-acoustic factors such as marketing and social media engagement.
- Dataset suffers from genre imbalance.
- One-hot encoding inflates dimensionality and increases variance.

D. Future Work

Future improvements could incorporate:

- Artist-level and listener-level metadata.
- Genre clustering to reduce dimensionality.
- Time series analysis to detect popularity trends.
- Natural Language Processing (NLP) on song lyrics to determine if sentiment (positive vs. negative lyrics) correlates with popularity.

VI. Conclusion

This project explored the extent to which audio features predict Spotify track popularity. While exact popularity scores were difficult to model due to noise and external influences, classification models achieved strong performance and offered meaningful insights. Loudness, energy, and danceability contributed positively to popularity, while instrumentalness reduced it. These findings show that audio features alone are not sufficient to fully predict popularity, but they provide valuable information for understanding listener behavior, industry trends, and music production strategies.

VII. Declaration of Generative AI Use

Generative AI tools were utilized in this project exclusively for assistance with coding syntax and troubleshooting LaTeX formatting errors. All conceptual work, model implementation, and analysis remain the original work of the authors.

References

[1] M. Pandya, "Spotify Tracks Dataset," *Kaggle*, 2022. [Online]. Available: https://www.kaggle.com/datasets/maharshipandya/-spotify-tracks-dataset. [Accessed: Dec. 08, 2025].

[2] D. Miraglia, "What Is The Loudness War & Is It Still Relevant In 2025?," *Unison Audio*, Oct. 28, 2024. [Online]. Available: https://unison.audio/loudness-war/. [Accessed: Dec. 09, 2025].

[3] Z. Mack, "How streaming affects the lengths of songs," *The Verge*, May 28, 2019. [Online]. Available: https://www.theverge.com/2019/5/28/18642978/music-streaming-spotify-song-length-distribution-production-switched-on-pop-vergecast-interview. [Accessed: Dec. 09, 2025].

APPENDIX A
DETAILED MODEL RESULTS

This section contains the comprehensive performance metrics for all models tested, as well as the detailed feature importance breakdown including genre-specific predictors.

TABLE III
DETAILED REGRESSION MODELS SUMMARY

Model	**RMSE**	**MAE**	R^2
Linear Regression	16.859739	12.029411	0.326653
Ridge (alpha=1.0)	16.859905	12.034881	0.326640
RF Regressor (fast)	17.909901	13.832112	0.240157
Lasso (alpha=0.1)	18.555319	15.022649	0.184406

TABLE IV
DETAILED CLASSIFICATION MODELS SUMMARY

Model	**Acc**	**Prec**	**Rec**	**ROC-AUC**
Random Forest	0.801	0.758	0.539	0.868
Logistic Reg.	0.773	0.679	0.524	0.829
KNN (k=15)	0.751	0.641	0.467	0.793
Decision Tree	0.741	0.807	0.230	0.601

TABLE V
TOP FEATURE IMPORTANCE (AUDIO & GENRE)

Feature	**Importance Score**
Duration (ms)	0.062749
Loudness	0.061677
Valence	0.061203
Acousticness	0.061188
Energy	0.060701
Danceability	0.059603
Speechiness	0.058757
Tempo	0.057945
Liveness	0.057769
Instrumentalness	0.047317
Mode	0.010538
Genre: K-Pop	**0.017889**
Genre: Pop-Film	**0.017602**
Genre: Chill	**0.015407**
Genre: Sertanejo	**0.011672**
Genre: Anime	0.008390
Genre: Grunge	0.008195
Genre: Deep House	0.008056
Genre: Prog. House	0.007705
Genre: Ambient	0.007141