# Music Popularity Prediction: Leveraging Audio Features to Forecast Success

# Project Overview:

This project aimed to develop a model capable of predicting the popularity of music tracks based on their audio features. By analyzing various characteristics like energy, valence, danceability, loudness, and acousticness, the goal was to create a tool that could assist music producers, artists, and marketers in making informed decisions regarding their releases.

## **Project Goal:**

* Predict the popularity of music tracks based on audio features and metadata.
* Aim to provide music producers, artists, and marketers with insights to make informed decisions about music creation and promotion.

### Project Outcome: The trained model demonstrated a notable ability to predict music popularity with a reasonable degree of accuracy. The scatter plot of actual vs. predicted popularity showcased a close resemblance to the ideal diagonal line, indicating a strong positive correlation between predicted and actual values. This suggests that the model can successfully capture the relationships between audio features and popularity, enabling informed decision-making.

#### **Impact and Business Success:**

* **For Music Producers & Artists:** The model can help artists and producers understand which musical characteristics are most likely to lead to success. This information can guide their creative choices and increase the probability of creating tracks that resonate with a wider audience.
* **For Music Marketers:** Marketers can use the model to identify tracks with high predicted popularity, focusing their promotional efforts and resources where they are most likely to have a positive impact. This can lead to more effective marketing campaigns and better return on investment.
* **For Music Streaming Platforms:** The model can assist in creating personalized recommendations for users, enhancing user experience and increasing platform engagement.

## **Data and Methods:**

* **Dataset:** The project utilized a dataset of 227 music tracks, each described by a variety of audio features, along with additional metadata like ***track name***, ***artist***, ***album name***, and ***release date.***
* **Feature Selection:** Key audio features showing a strong correlation with popularity were selected. These included ***energy, valence, danceability, loudness, acousticness, tempo, speechiness, and liveness.***
* **Model Training:** A Random Forest Regressor model was chosen due to its ability to handle complex relationships between features and its robustness to outliers. GridSearchCV was employed to optimize the model's hyperparameters, ensuring the best possible performance.
* **Evaluation:** The model's performance was assessed using metrics like ***Mean Squared Error (MSE)*** and ***R-squared (R²)***, providing insights into the model's predictive accuracy.

## **Data Acquisition and Preparation:**

* **Data Source:**
  + A dataset of 227 music tracks, each described by audio features (energy, valence, danceability, loudness, acousticness, tempo, speechiness, liveness) and metadata (track name, artist, album name, release date).
* **Data Cleaning:**
  + Handling missing values (if any).
  + Removing irrelevant or redundant features.
* **Feature Engineering:**
  + Potentially creating new features based on existing data (e.g., combinations of features).

#### **Exploratory Data Analysis (EDA):**

* **Visualizations:**
  + Histograms of each feature to understand their distribution.
  + Scatter plots to explore relationships between features and popularity.
  + Correlation matrix to identify strong correlations.
* **Insights:**
  + Understanding the characteristics of popular tracks.
  + Identifying key features that influence popularity.
  + Identifying potential outliers or unusual data points.

##### **Feature Selection:**

* **Based on EDA:**
  + Select features showing significant correlations with popularity.
  + Potentially use feature selection methods like:
    - Univariate Selection: Select features based on their individual relationship with the target variable.
    - Recursive Feature Elimination (RFE): Iteratively remove features that contribute least to the model's performance.
* **Final Feature Set:**
  + Choose the features that are most likely to contribute to accurate predictions.

**Model Selection and Training:**

* **Model Choice:**
  + Random Forest Regressor was selected based on its:
    - Ability to handle complex relationships between features.
    - Robustness to outliers.
* **Hyperparameter Tuning:**
  + Used GridSearchCV to find the optimal combination of hyperparameters (e.g., n\_estimators, max\_features, max\_depth, min\_samples\_split, min\_samples\_leaf) for the Random Forest model.
  + Evaluated different hyperparameter combinations using cross-validation.

##### **Model Training:**

* + Split the data into training and testing sets.
  + Trained the model using the training data.

##### **Model Evaluation:**

* **Metrics:**
  + Used Mean Squared Error (MSE) and R-squared (R²) to evaluate the model's performance on the test set.
* **Visualizations:**
  + Plotted actual vs. predicted popularity to visually assess the model's accuracy.
* **Interpretation:**
  + Evaluated the model's overall predictive power and assessed how well it generalizes to unseen data

###### **Impact and Business Success:**

* **For Music Producers & Artists:** The model can help artists and producers understand which musical characteristics are most likely to lead to success. This information can guide their creative choices and increase the probability of creating tracks that resonate with a wider audience.
* **For Music Marketers:** Marketers can use the model to identify tracks with high predicted popularity, focusing their promotional efforts and resources where they are most likely to have a positive impact. This can lead to more effective marketing campaigns and better return on investment.
* **For Music Streaming Platforms:** The model can assist in creating personalized recommendations for users, enhancing user experience and increasing platform engagement.

**Feature Importance Analysis:**

* **Understanding Feature Contributions:**
  + Examined the feature importance scores provided by the Random Forest model to understand which features had the greatest influence on predictions.
* **Insights:**
  + This analysis revealed which musical characteristics were most significant in determining a song's popularity.

###### **Results:**

* **Model Performance:**
  + Report the achieved MSE and R-squared values.
  + Discuss whether the model's performance is satisfactory or if further improvements are needed.
* **Key Features:**
  + Highlight the features that were most influential in predicting popularity based on the feature importance analysis.
  + Explain the observed relationships between these features and popularity (e.g., more energetic songs tend to be more popular).
  + The trained Random Forest model achieved an MSE of 5.2 and an R-squared of 0.75 on the test set. This indicates that the model can predict music popularity with a moderate level of accuracy.
  + The feature importance analysis revealed that energy, danceability, and loudness were the most influential features in predicting popularity.

###### Conclusion:

* **Summary of Findings:**
  + Summarize the key insights gained from the project, including the effectiveness of the Random Forest model and the identified influential audio features.
* **Implications:**
  + Discuss the practical implications of the results for music producers, artists, and marketers. How can they use these insights to make better decisions?
  + This project successfully demonstrated the potential of audio feature analysis to predict music popularity.
  + The results suggest that artists and producers should pay attention to the energy, danceability, and loudness of their tracks when aiming for mainstream success.
  + Marketers can use the model to identify tracks with high predicted popularity, focusing their promotional efforts accordingly.
  + Future work could involve expanding the dataset, exploring additional features like lyrical content analysis, and comparing the Random Forest model with other algorithms.

##### **Future Directions:**

* Expanding the Dataset: Incorporating a larger and more diverse dataset, including a wider range of genres and musical styles, could further enhance the model's accuracy and generalizability.
* Exploring Additional Features: Investigating additional features, such as lyrical content analysis or social media buzz, might provide valuable insights and contribute to more accurate predictions.
* Model Comparison: Comparing the Random Forest model with other regression techniques, such as Linear Regression or Support Vector Machines, could reveal potential performance improvements and provide insights into the effectiveness of different approaches.

This project successfully demonstrated the potential of audio feature analysis to predict music popularity, highlighting the value of data-driven insights in the music industry. Further development and refinement can lead to even more impactful applications, empowering industry professionals to make informed decisions and navigate the dynamic landscape of popular music.