

# Research Report on Data Analytics, Data Science, and AI in Music Popularity Prediction

This report provides the theoretical background for using Data Analytics, Data Science, and Artificial Intelligence (AI) to predict song popularity based on audio features.

## Define Key Terms

**Data Analytics:** The process of examining datasets to extract meaningful insights using statistical and computational techniques.

**Data Science:** An interdisciplinary field that uses algorithms, statistics, and computing to extract knowledge from data.

**Artificial Intelligence (AI):** The simulation of human intelligence in machines, enabling them to learn and make predictions.

**Song Popularity:** A measurable score representing how well a song performs, often derived from streaming counts, chart rankings, or user ratings.

**Audio Features:** Quantitative descriptors of a song's sound, such as danceability, energy, tempo, and acousticness.

**Regression Analysis:** A statistical approach for modeling the relationship between a dependent variable and one or more independent variables.

**Random Forest:** An ensemble learning method using multiple decision trees for improved predictive accuracy.

**Gradient Boosting:** An ensemble technique that builds models sequentially, each correcting errors of its predecessor.

## Role in the Music Industry

Predicting song popularity helps record labels, artists, and streaming platforms make informed decisions about marketing, playlist curation, and promotional efforts.

## Factors Influencing Song Popularity

Popularity can be influenced by both quantifiable and qualitative factors, including audio features, artist reputation, marketing campaigns, timing of release, and cultural trends. In this project, we focus on audio features as key predictors.

## Regression Techniques for Prediction

Suitable regression methods for predicting song popularity include:

- **Linear Regression:** Models linear relationships.
- **Polynomial Regression:** Captures non-linear patterns.
- **Random Forest:** Handles complex, non-linear relationships with high accuracy.
- **Gradient Boosting:** Excels in predictive performance by iteratively improving models.

## Audio Feature Analysis

Audio features like tempo, energy, and danceability are extracted using tools such as Spotify's API. These features quantitatively describe a song's rhythm, mood, and style.

## **Feature Engineering**

In music data, feature engineering may involve combining related audio metrics, creating interaction terms, or normalizing tempo and loudness to enhance model performance.

## **Model Evaluation in Regression**

Common evaluation metrics for regression include Mean Squared Error (MSE), Root Mean Squared Error (RMSE), Mean Absolute Error (MAE), and R-squared, which collectively assess accuracy and explanatory power.