Project Proposal

Title Page

- **Title:** Analysis of Factors Influencing Song Popularity on Music Streaming Platforms.
- **Team Members:** Harsh Sahay, hsahay2@wisc.edu; Srijan Gupta, sgupta285@wisc.edu; Sahil Mehta, smehta27@wisc.edu
- **Team Name:** Skyfall21

1. Overview

This project aims to analyze the factors that influence a song's popularity on streaming platforms like Spotify, Apple Music, and Deezer, leveraging a rich dataset of the most streamed songs on Spotify. We plan to explore relationships between song attributes, artist popularity, release timing, and presence across multiple platforms to predict song popularity.

2. Background

With the rise of digital music streaming, understanding what makes a song popular has become crucial for artists and record labels. Previous studies have focused on audio features and social media presence, but few have combined these with multi-platform streaming data and temporal release factors. Our project seeks to fill this gap, providing insights into effective strategies for maximizing song popularity.

3. Statement of Work

3.1 Datasets

The primary dataset for this project includes detailed attributes of top songs on Spotify, such as track name, artist(s), release date, playlist inclusion, streaming statistics, and presence on Apple Music and Deezer. Preliminary examination suggests the data is well-documented and clean, with a diverse range of songs from various years and genres.

3.2 Method

Our methodology will involve:

- Data Preprocessing: Cleaning data, handling missing values, and encoding categorical variables.
- Exploratory Data Analysis (EDA): Visualizing data to identify trends and correlations between song popularity and other features.
- Feature Engineering: Creating new variables (e.g., artist popularity index, genre diversity score) to enrich the analysis.
- Predictive Modeling: Employing machine learning techniques (e.g., regression analysis, random forests, gradient boosting machines) to predict song popularity.
- Validation and Evaluation: Using cross-validation and performance metrics (e.g., RMSE, R^2) to assess model accuracy and interpretability.

3.3 Outcome and Performance Evaluation

- **Metrics:** Model performance will be evaluated using RMSE (Root Mean Square Error) and MAE (Mean Absolute Error).
- **Success Criteria:** High accuracy in predictions and actionable insights for enhancing renewable energy adoption.

3.4 Outcome and Performance Evaluation

We aim to identify key predictors of song popularity and develop a predictive model with high accuracy. Success will be measured by the model's ability to predict popularity rankings and streaming counts, providing actionable insights for artists and music industry professionals.

4. Project Plan

A Gantt chart outlines the project timeline over 7 weeks, including dataset collection, preprocessing, EDA, model development, evaluation, and report preparation. The project will be documented on GitHub, with the repository link shared with instructors and TA.

Github Link: https://github.com/sahilmehtx/ECE539_Project/

Grant Chart:

Weeks	1	2	3	4	5	6	7
Dataset collection and preprocessing.	☒	☒					
Exploratory data analysis.			☒				
Model development.				☒	☒		
Performance evaluation.						☒	
Final report and presentation preparation.							☒

5. References

1. Velmuruga, Dr.A. (2023). Machine Learning Approaches for Predicting Song Popularity: A Case Study in Music Analytics. INTERNATIONAL JOURNAL OF SCIENTIFIC RESEARCH IN ENGINEERING AND MANAGEMENT. 07. 1-11. 10.55041/IJSREM27361. This study explores various computational models to predict the popularity of songs. https://www.researchgate.net/publication/376387750_Machine_Learning_Approaches_for_Predicting_Song_Popularity_A_Case_Study_in_Music_Analytics/

2. Alison Salerno (2020): "Prediction of Spotify Song Popularity". This paper examines different machine learning techniques to predict the popularity of Spotify songs. Explore on IEEE Xplore. https://medium.com/analytics-vidhya/predicting-song-popularity-71bc3b067237/

3. Jacob Devlin, Ming-Wei Chang, Kenton Lee, and Kristina Toutanova (2019): "BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding". While primarily focused on natural language processing, BERT's methodologies can be adapted for analyzing textual data in song lyrics to predict popularity. https://aclanthology.org/N19-1423/

4. Elena Georgieva, Shiva Pentyala, Marco Giunta, Paolo Papotti, and K. Selçuk Candan (2018): "Feature Engineering for Predicting Billboard Hits". This study discusses feature engineering techniques crucial for predictive models in the context of Billboard hits, applicable to Spotify song popularity prediction. Find on ACM Digital Library. https://ccrma.stanford.edu/~egeorgie/documents/HitPredict_Final.pdf

6. Contributions

Harsh: Led the project concept development, identified datasets, and outlined the analytical methods.

Srijan: Conducted preliminary data inspection, contributed to the methodological approach, and will lead the model development phase.

Sahil: Responsible for the literature review, identifying key references, and will oversee the final report preparation and presentation.