SYNOPSIS

1. Name / Title of the Project

Song Popularity Prediction Using Data Science and Machine Learning.
This project focuses on leveraging advanced data science techniques and machine learning algorithms to predict the popularity of songs in the dynamic music industry.

2. Statement about the Problem

Despite the exponential growth of music streaming platforms, predicting a song's popularity based on various features remains a complex challenge. With the vast array of music available online, understanding what makes a song resonate with listeners can greatly impact artists and record labels' strategies.

3. Describe the Problem Statement

Popularity prediction is essential for artists and record labels as it helps them understand market trends, enhance marketing strategies, and allocate resources effectively when investing in upcoming artists. Accurately analysing the factors that contribute to a song's success is challenging due to the diverse range of both qualitative and quantitative elements involved. Factors such as the song's tempo, popularity, artist reputation, marketing efforts, lyrical content, and even timing of the release all play significant roles in determining its reach and popularity. In addition, the subjective nature of music consumption requires a nuanced approach to categorise and quantify these influences effectively for predictive modelling.

4. Objective and Scope of the Project

The primary objective of this project is to develop a robust predictive model capable of assessing a song's popularity based on diverse features including tempo, genre, lyrical content, historical performance metrics, and social media engagement. The scope includes comprehensive data collection from reputable platforms such Kaggle, followed by meticulous feature engineering, model training, and evaluation. This project aims to provide actionable insights into the key factors influencing song popularity, thereby equipping artists and labels with the knowledge needed to strategize effectively in a competitive market.

5. Methodology

The methodology for this project encompasses the following phases:

Data Collection and Preprocessing

- **Data Acquisition:** Gathering a comprehensive dataset from reputable sources like Kaggle, encompassing various song attributes, user engagement metrics, and historical popularity data.

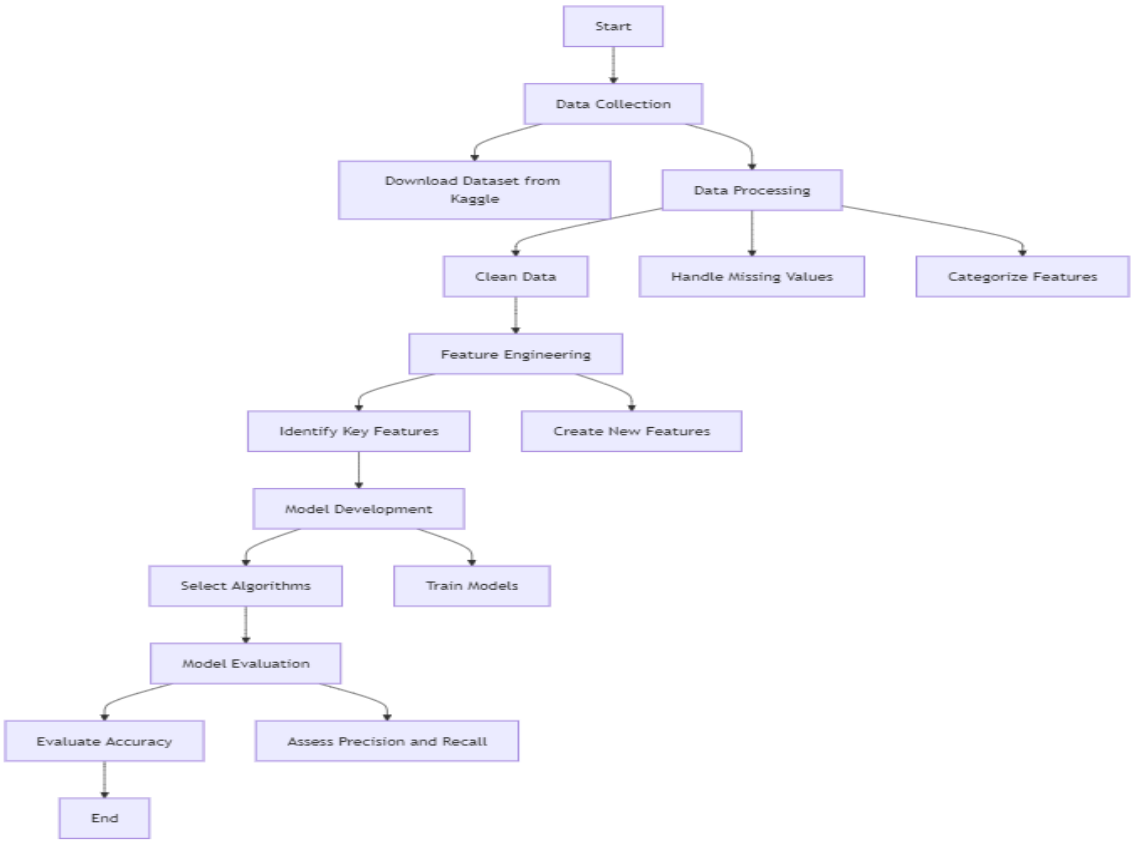
- **Data Cleaning:** Handling missing values, outliers, and inconsistencies within the dataset to ensure data quality and reliability.
- **Feature Engineering:** Creating new or transforming existing features to improve model performance. This might involve feature scaling, normalisation, or generating derived features.

Model Development and Training

- **Traditional Machine Learning:** Employing established algorithms like Linear Regression, Decision Trees, and Random Forests as baseline models to understand the problem and benchmark performance.
- **Deep Learning:** Exploring the potential of deep learning algorithms, particularly to capture complex patterns and dependencies in the data.

Model Evaluation and Comparison

- **Performance Metrics:** Employing metrics such as RMSE (Root Mean Squared Error), MAE (Mean Absolute Error), R-squared, and F1-score to evaluate model performance.
- **Model Comparison:** Comparing the performance of different models, including traditional machine learning and deep learning approaches, to identify the most effective approach for predicting song popularity.
- **Visualisation:** Creating visualisations to compare the predictions of different models, helping to understand their strengths and weaknesses.



The provided flowcharts workflow for the "Song Popularity Prediction" project. The flowchart details the sequential steps involved in the project, starting from data collection through to model evaluation. It emphasises the importance of cleaning and processing data, identifying key features, and selecting appropriate algorithms for model training. Overall, this help in understanding the project's structure, timeline, and critical tasks necessary for successful completion.

User Interface Development

- **Overview of the UI Interface:** This project includes the development of an intuitive user interface using Gradio, which enables users to interact with the song popularity prediction model seamlessly. The UI is designed to provide an easy-to-use platform for users, including artists, record labels, and music enthusiasts, to input song features and receive real-time predictions of a song's popularity.

- **Features of the Gradio UI User Input Fields:** The interface allows users to input various song attributes, such as tempo, acousticness, loudness, singer, and social media engagement metrics. Prediction Display: Users can view the model's predictions in an easily interpretable format, showing estimated popularity scores.

6. Hardware & Software to be used

- Hardware: The project will be executed on a HONOR laptop equipped with an Intel i5 processor, 16 GB RAM, which provide enhanced processing capabilities, particularly for running complex algorithms efficiently.

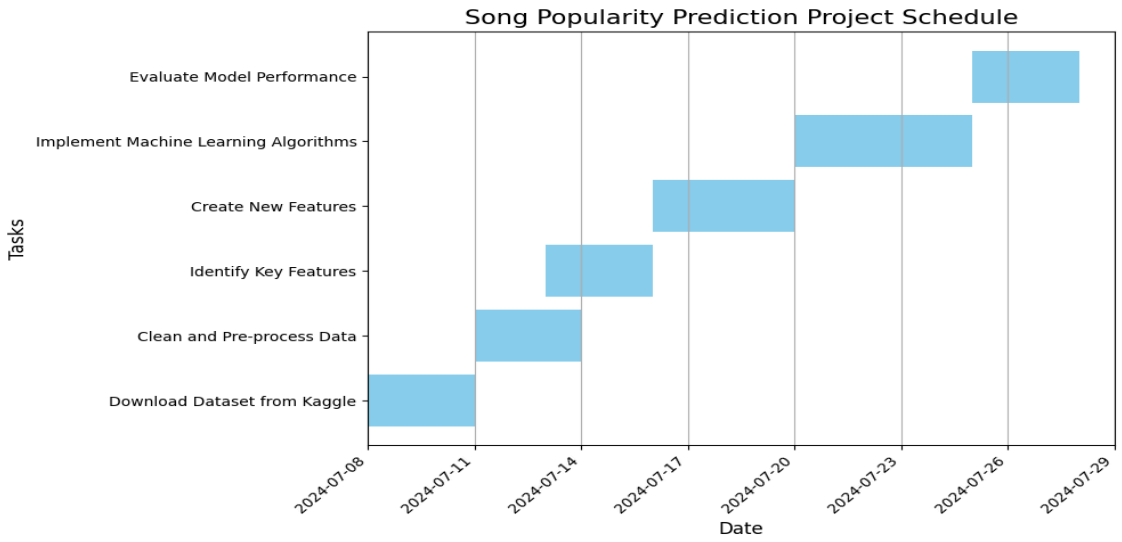
- Software: The implementation will be performed using the Python programming language, with Jupyter Notebook as the main development environment. Key libraries such as Pandas for data manipulation, NumPy for numerical operations, Scikit-learn for machine learning, and Matplotlib and Seaborn for data visualisation will be extensively used.

7. Future Work of this Project

Future work may delve into the incorporation of advanced deep learning techniques for enhanced prediction accuracy, especially for sequential data analysis. Additionally, the project could evolve to include real-time analytics, allowing users to predict song popularity dynamically as new data comes in. There is also potential for developing an interactive dashboard that visualises trends, predicts future hits, and allows users to input various song attributes to gauge potential success.

8. The Schedule of the project (Gantt chart / PERT chart)

A Gantt chart will be crafted to outline the project timeline for each phase. The anticipated duration for each phase is as follows:



The Gantt chart will serve as a valuable tool for project management, enabling the team to track progress, stay on schedule, and ensure that deadlines are met efficiently.

9. References/Bibliography

- Harrison, T. (2019). *Data Science for Music*. Journal of Music Data Science.
This reference provides a comprehensive overview of data science methodologies applied in the music sector.
- Kumar, A., & Reddy, M. (2020). *Machine Learning for Music Genre Classification*. Neural Computing and Applications.
This paper discusses the application of machine learning in classifying music genres, providing insights relevant to understanding popularity predictors.
- Jabbar, B. (2021). *Understanding music streaming algorithms: A study of Spotify and beyond*. Journal of Digital Media.
This article explores how streaming algorithms can affect music visibility and popularity, informing future research directions.