Berk Demirkan 21802124
Tuğcan Ünalan 21802512

EEE-485 Project Proposal

The "Spotify HUGE database" dataset contains 170,000 instances with features related to music such as energy, danceability, tempo, and sentimentalities. This project aims to build a machine learning model using the dataset to predict either the popularity or genre of a song.

Expected Challenges:

Before discussing the methodology, it is important to consider some of the expected challenges for this project. One challenge is the size of the dataset, which contains 170,000 instances. Handling such a large dataset may require additional computing power and storage resources. Additionally, it may be challenging to select the most relevant features for predicting popularity or genre, as some features may be correlated or redundant. Finally, it is crucial to select the appropriate machine learning algorithms and evaluation metrics to achieve the best possible performance for the given task. These challenges require careful consideration and planning to ensure the success of the project.

Methodology:

The project will involve several challenges that require careful consideration, such as the size of the dataset and selecting the appropriate machine learning algorithms for the task. Additionally, it may be difficult to interpret the results of the sentiment analysis due to the subjective nature of sentiment analysis.

The methodology will consist of the following phases:

1. Data preparation: In this phase, we will analyze the dataset to identify relevant features for predicting popularity or genre. We will pre-process the data to handle missing values and outliers.
2. Feature selection: In this phase, we will select the most relevant features for predicting popularity or genre. We will use techniques such as correlation analysis and feature importance analysis to determine which features are most important.
3. Model development: In this phase, we will train and test various machine learning algorithms for regression or classification tasks. We will evaluate the performance of the model using appropriate evaluation metrics such as accuracy, precision, recall, and F1 scores.
4. Result visualization: In this phase, we will visualize the results and interpret the findings. We will use various visualization techniques such as scatter plots, line graphs, and histograms.

Examples of the machine learning algorithms that we could use for this project include:

- Regression: Linear regression, decision tree regression, random forest regression, and neural network regression.
- Classification: Logistic regression, decision tree classification, random forest classification, and neural network classification.