**CMPE 255**

**PROJECT PROPOSAL**

**SONG RELEASE YEAR PREDICTION**

*Team Members:*

Charanveer Singh

Moxank Prakashbhai Patel

**Project Description**

In this project, we will aim to predict the release year of the song using audio features provided within the dataset. This will be accomplished using data mining techniques learned in the course and applying these techniques to 90 attributes/audio features provided within the dataset.

**Proposed Methodology**

*Dataset Used:* YearPredictionMSD Data Set

*Dataset Link:* https://archive.ics.uci.edu/ml/datasets/YearPredictionMSD

To start off, the dataset will go through some pre-processing steps such as Dimensionality Reduction, Feature Subset Selection, Feature Creation, and Attribute Transformation. Additionally, we will eliminate any noise, outliers, and missing or duplicate data to guarantee the highest possible data quality. Finally, the use Data Mining techniques such as Classification and Regression will be used to finish the task. The prediction's accuracy may be improved by using one or more strategies. A few metrics used for Model performance may include:

* Accuracy Score
* Root Mean Squared Error
* Mean Absolute Error
* R Square Value

About Dataset:

The dataset includes total of 90 features which are nothing but timbre of the songs. A timbre is nothing but the tone of the music. The songs varies from year 1992 to 2011. The total number of songs in the dataset is around 500000.