IDS 702 - Final Project Proposal

Pranav Manjunath

**Title: Prediction of Genre based on Song Characteristics/Features**

Objective: The objective of this project is to create an R Shiny Application based on a statistical model that can accurately predict a song’s genre based on its musical features and characteristics.

Background: Many musicians incorporate multiple elements of various genres into their music. While composing and creating, there would not have a clear idea as to what genre their music belongs to.

Hence this project aims to accurately predict the song’s genre by using it’s musical characteristics.

Scope: This project will be mainly used by musicians who compose music. The scope of the project would be building an appropriate model that can classify and predict song genres. The next step would be building an R Shiny application where musicians can input their song characteristics and the application would predict the respective genre of the song.

Major Research Questions:

1. What are the most important musical features that impact the genre of a song?
2. Which model would produce better results in predicting the genre of a song?

A picture containing text

Description automatically generated

Data:

The dataset that will be used for this project is taken from Kaggle and consists of 131562 rows (songs) and 15 musical features. There are 2800 Genres present in the dataset. Except for the genre, there are no factor variables present in the data. There are no missing values present in the data. Grouping of similar genres must be done.

Questions to answer when working with the data:

1. How to combine and group genres together?
2. What is the minimum number of songs a genre should have?

Project Plan:

As the response variable (Genre) is a multiclass factor variable, I will use Multinomial Logistic Regression to build a classifier. I would also plan to use a neural network and compare the results of both models.

Timeline:

Week1 (Oct 18th - Oct 24th): EDA and Understanding the Data

Week2 (Oct 25th - Oct 31st): Model implementation and Assessment

Week3 (Nov 1st - Nov 7th): Model implementation and Assessment

Week4 (Nov 8th – Nov 14th): Working on Front End and Report Writing

Data Columns