**SONG RELEASE YEAR PREDICTION**

San Jose State University

CMPE255 GROUP PROJECT

Group Members: Charanveer Singh, Moxank Prakashbhai Patel, Xinyu He, Hongjin Cheng

1. **INTRODUCTION**

* **Motivation**

A song is a musical composition intended to be performed by the human voice. Each song has its own unique melody or tone, etc. These characteristics make up the song. Everyone has their own favorite type of song, and this is because there are songs in each genre, and most of these songs in the same genre share some similar characteristics. In addition to personal preferences, there are some very popular songs from each era that represent the characteristics of the most popular song genres of that era. Most of the songs from different eras also have the characteristics of the song style belonging to the era. By analyzing the characteristics of these songs and combining them with the time of their release, we can get an era in which we can judge the possible release of different songs.

* **Objective**  
  In this project, we will use the audio features provided in the dataset to predict the year of release of a song. This will be done using the data mining techniques learned in the course and applying these techniques to the 90 attributes/audio features provided in the dataset. We will design a prediction model and train it with the dataset. By refining this model, we hope it can predict the release time of the input songs.
* Literature / Market review

1. **System Design & Implementation details**

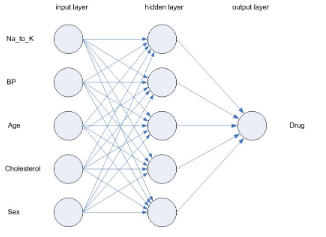
* **Algorithms considered/selected  
  Simple Linear Regression**

Linear regression analysis is used to predict the value of a variable based on the value of another variable. The variable you want to predict is called the dependent variable. The variable you are using to predict the other variable's value is called the independent variable.

Simple linear regression is a linear regression model with a single explanatory variable. That is, it concerns two-dimensional sample points with one independent variable and one dependent variable (conventionally, the x and y coordinates in a Cartesian coordinate system) and finds a linear function (a non-vertical straight line) that, as accurately as possible, predicts the dependent variable values as a function of the independent variable. The adjective simply refers to the fact that the outcome variable is related to a single predictor.

**Neural Network Model**

Neural networks are simple models of the way the nervous system operates. The basic units are neurons, which are typically organized into layers, as shown in the following figure



A neural network is a simplified model of the way the human brain processes information. It works by simulating a large number of interconnected processing units that resemble abstract versions of neurons.

The processing units are arranged in layers. There are typically three parts in a neural network: an input layer, with units representing the input fields; one or more hidden layers; and an output layer, with a unit or units representing the target field(s). The units are connected with varying connection strengths (or weights). Input data are presented to the first layer, and values are propagated from each neuron to every neuron in the next layer. Eventually, a result is delivered from the output layer.

* **Technologies & Tools used**Jupyter Notebook (python3)
* **Architecture-related decisions  
  Data set selection**

The entire dataset of the experiments is

280GB, the subset of 10,000 songs(1%) is 1.8 GB. However, it is still very large, and we only use the data to predict the year of song by some input attribute. Hence, we continue to select our dataset YearPredictionMSD, which is only 449MB.

**Model selection**

We tried the simple linear regression model first. However, it has very low accuracy. Then we try to use the neural network model, and it has much better performance.

**Model Performance**

…

* **System design/architecture/data flow**
* **Component details**
* **Use cases / GUI / screenshots**

1. **Experiments / Proof of concept evaluation**

* **Dataset used**  
  Name: YearPredictionMSD Data Set  
  Source: <https://archive.ics.uci.edu/ml/datasets/YearPredictionMSD>  
  Type of data: Multivariate  
  Size of data: 427MB  
  Number of instances: 515345
* **Methodology followed**
* **Graphs showing different parameters /algorithms evaluated in a comparative manner, along with some supportive text**
* **Analysis of results**

1. **Discussion & Conclusions**

* **Decision made**During the first discussion, we had two proposals. One is about online shopping research on websites and the other is song release year prediction. After voting we chose the second one.  
    
  During the second discussion, we decided which part each person would be responsible for. Each part was discussed in depth to determine the general direction of our project in the future.  
    
  During the third discussion, we have made some progress in each section and have generally completed the project. But we have not yet integrated each part, nor have we discussed and improved them.  
    
  During the fourth discussion, we discussed the content of the PowerPoint, divided up the content of the presentation that each group member needed to speak, and how we should improve the part we were responsible for.
* **Difficulties faced**
* **Things that worked**
* **Things that didn’t work well**

simple linear regression model

* **Conclusions**

1. **Project Plan / Task Distribution**

* **Charanveer Singh**
* **Moxank Prakashbhai Patel**
* **Xinyu He 012491544**
* **Hongjin Cheng 016032406**

**GitHub Link:**

**Reference:**

1.Comparison Experiencement results <https://github.com/raghav1810/Music-analysis-and-Year-prediction/blob/master/release-year-prediction-on-msd.ipynb>

2.PCA Guidance <https://scikit-learn.org/stable/auto_examples/preprocessing/plot_scaling_importance.html#sphx-glr-auto-examples-preprocessing-plot-scaling-importance-py>

3.Wiki Simple Linear Regression <https://en.wikipedia.org/wiki/Simple_linear_regression>

4.Wiki Neural Network <https://en.wikipedia.org/wiki/Artificial_neural_network>