Conversational AI Data Science UCS663

Kaggle-based Lab Evaluation - I

Name: Inderpal Singh
Roll No: 101903343
Group: COE13

Problem Name	Gender Recognition by Voice
Problem Link	https://www.kaggle.com/primaryobjects/voicegender
Problem Type	**Regression**
GitHub Link	**https://github.com/inderpal12/Gender_Voice_Recognition**
Kaggle Link	https://www.kaggle.com/inderpal0013/gender-voice-recognition
Libraries used	**cuPY, cuDF, matplotlib.pyplot, sklearn.metrics, sklearn.model_selection, sklearn.neighbors, sklearn.tree, sklearn.linear_model, sklearn, seaborn**
Models Implemented	**xgboost, Decision Tree Classifier, KNN, SVC, Logistic Regression,**
Evaluation Metrics Used	**Accuracy, R2 score**
Kaggle Rank Achieved with total number of teams (if applicable)	
Proof of Rank (if applicable)	**<Put Screenshot here>**
Tasks done in code	- **Data understanding (preparation and EDA)** - **Data Preprocessing** - **Normalization** - **Data Exploration** - **Various Models Implementation** - **Comparison** - **Detailed Coding documentation**

Various Steps Performed:

1) Basic Importing of Libraries: Various useful libraries are imported to perform our tasks.

2) Data Pre-processing: Data Pre-processing is done to convert the raw data into the useful and efficient format. Firstly, the dataset is loaded. The data is then thoroughly checked with various techniques.

3) Normalization: It is done in order to scale the data values in a specified range (-1.0 to 1.0 or 0.0 to 1.0). Normalization is a scaling technique in which values are shifted and rescaled so that they end up ranging between 0 and 1. It is also known as Min-Max scaling.

4) Data Exploration: Data exploration, also known as exploratory data analysis (EDA), is a process where users look at and understand their data with statistical and visualization methods. This step helps identifying patterns and problems in the dataset, as well as deciding which model or algorithm to use in subsequent steps.

5) Splitting of dataset into train and test sets: Data set is splitted into train and test set on which various models will be applied.

6) Various Models Implementation and their Accuracy: Various models are implemented. The best accuracy is given by XGBoost and KNN Classifier.

7) Comparison of model on the basis of accuracy and R2 score: XGbosot and KNN models are compared and their corresponding graphs have been created.