B-Tech -/III Year CSE - V Semester

19CSE305-Machine Learning Project Review

Amrita School of Engineering, Coimbatore

TOPIC – Sports Prediction

CB.EN.U4CSE20248	Raghunandan S
CB.EN.U4CSE20238	M Satya Chandana Snehal

Introduction:

FIFA is also a series of video games developed by EA Sports which faithfully reproduces the characteristics of real players. Data from this website for 2019 were scrapped and made available at the Kaggle webpage https://www.kaggle.com/karangadiya/fifa19.

The original dataset contains 89 variables that describe 16,924 players. The variables include information such as age, nationality, club, wage, etc.

We predict two dependent variables:

1.**Market Value**: The market value of a player can determine the amount of money that clubs can demand for a transfer. Factors like personal form, importance to the team, stats, contract length and age can strongly influence the market value of a player.

2. **POSITION:** It is the position in which the player plays, such as attack, middle, defense, or goalkeeper.

Regression:

We predicted the market value of players by regression. The models used are:

    1.Simple Linear Regression

    2.Multiple Linear Regression

    3.KNN

Classification:

We predicted the Position of players using classification. We use several models like Decision tree, Random Forest, etc.

Algorithms used:

Supervised learning algorithms:

SVM

KNN

Decision Tree

Random forest

Naive bayes

Unsupervised learning algorithms:

K-Means Clustering

Birch clustering

Agglomerative clustering

Boosting algorithms:

XGboost

Light gradient boosting Machine (LGBM)

Result analysis:

Accuracy of Decision Tree: 0.8212075954768508

Accuracy of LGBM: 0.8756133987625346

Accuracy of XGBoost: 0.8653723063793471

 Accuracy of SVM: 0.7559206315340303

Accuracy of Random Forest: 0.8724130573927885

Accuracy of Naive Bayes: 0.7277576274802645

Accuracy of KNN: 0.8156603371026243

Accuracy of K-Means: 0.3891615105611265

Accuracy of Agglomerative Clustering: 0.2613612118625987
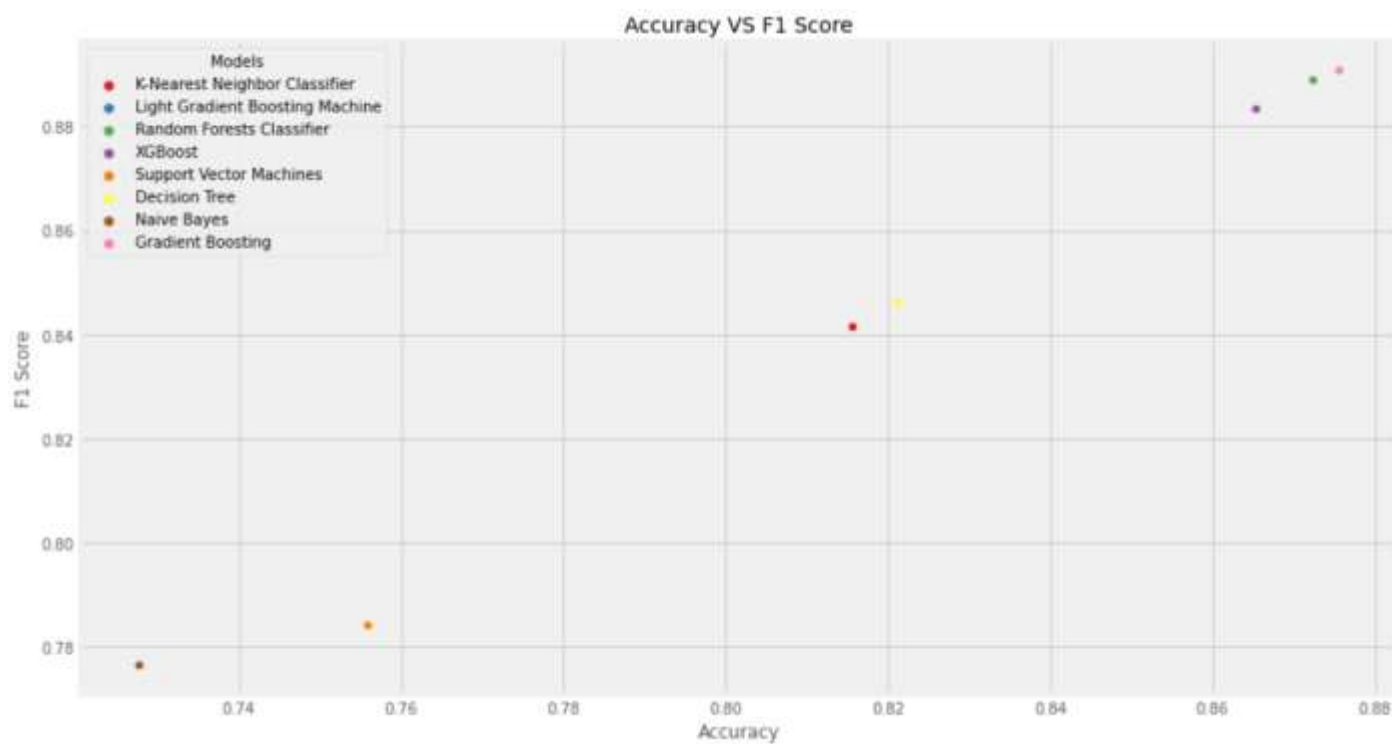
Accuracy of Birch Clustering: 0.2613612118625987

 Bar graph for the Accuracy Scores of all the models:

Bar graph for the F1 Scores of all the models

F1 Score of Different Models

Scatter plot for the Accuracy and F1 scores

Accuracy VS F1 Score

From the above plots and the scores, we can see that LGBM (Light Gradient Boosting Machine) model gave the highest F1 score which means it is the perfect model for predicting.

Two other models which can be used for this data set are Random Forest Classifier and XG Boost as they give a better accuracy after the LGBM model.

Copy of Code:

Please refer this for the colab notebook -

https://colab.research.google.com/drive/1AC96v36bfog2V8A9gAm6BgUSLmDgX-aZ?usp=sharing