FIFA 20 Player Value Prediction

Introduction

Fifa is one of the most popular games in the world. Every year, EA Sports releases its ratings for each FIFA player based on their current form. It's a great data resource to explore for football fans (which happens to be myself). Every FIFA player has been given a score for dozens of attributes such as dribbling, pace, shooting, etc. And more importantly, it has provided the market value and wages for each player. This project is about to predict the value of players and compare the result between two models which are linear regression and decision tree.

Dataset Description

The datasets provided include the player's data from FIFA 20. This dataset contains total 16 columns. They are name, age, height, weight, overall, potential, wages, weak foot, skill move, pace, shooting, passing, dribbling, defending, physic and values.

Content

- 16 attributes
- Players attributes with statistics as skill moves, pace, shooting, passing, dribbling, physic, etc.
- Player personal data like name, wages, Salary, values, etc.

Descriptions of the Models

1. **Linear Regression:**
   It is used to predict the relationship between two variables by applying a linear equation to observed data. There are two types of variables, one variable is called an independent variable, and the other is a dependent variable. Linear regression is commonly used for predictive analysis. The main idea of regression is to examine two things. First, does a set of predictor variables do a good job in predicting an outcome (dependent) variable. The second thing is which variables are significant predictors of the outcome variable.
   For the regression line where the regression parameters b0 and b1 are defined, the following properties are applicable:
   - The regression line reduces the sum of squared differences between observed values and predicted values.
   - The regression line passes through the mean of X and Y variable values.
   - The regression constant b0 is equal to the y-intercept of the linear regression.
   - The regression coefficient b1 is the slope of the regression line. Its value is equal to the average change in the dependent variable (Y) for a unit change in the independent variable (X)

2. **Decision tree:**
   Decision Tree is a supervised learning technique that can be used for both classification and Regression problems, but mostly it is preferred for solving Classification problems. It is a tree-structured classifier, where internal nodes represent the features of a dataset, branches

represent the decision rules and each leaf node represents the outcome. In a Decision tree, there are two nodes, which are the Decision Node and Leaf Node. Decision nodes are used to make any decision and have multiple branches, whereas Leaf nodes are the output of those decisions and do not contain any further branches. The decisions or the test are performed on the basis of features of the given dataset.

Result Table

Linear Regression and Decision tree, these two models are used to measure the result. To compare the results four performance matric is executed. They are mean absolute error, mean squared error, root mean squared error and r2 score.

The result comparison table is given below.

Classifiers	MAE	MSE	RMSE	R2 Score
Linear Regression	6.95128	111.1972	10.5450	0.7240
Decision Tree	11.3043	371.0652	19.2630	0.0790

Conclusion

After the model analysis, here in the linear regression classification the mean absolute error is 6.95128, mean squared error is 111.1972, root mean squared error is 10.5450 and r2 score is 0.7240.

For the decision tree classification the mean absolute error is 11.3043, mean squared error is 371.0652, root mean squared error is 19.2630 and r2 score is 0.0790.

So, Linear Regression works well for the project.