ML Mini Project

Name - Vedanta Yadav **SAPID - 60009210185**

Problem Statement

Football Match Winner Prediction using ML

I will be using different prediction models on a dataset which comprises match scores from the Premier League (2001-2016) and determining which is the best model to use for predicting match winner in football

Step 1

Collecting the dataset - The dataset has been taken from github which was converted from sqlite form to csv. Each season from 2001 to 2016 has been merged to form a 'merged_dataset.csv' which contains the scorelines from every match played within those 15 seasons in the Premier League.

Step 2

EDA - We remove all null values (if any) and unnecessary attributes and put them in a cleaned data frame. We then perform some basic analysis using matplotlib on the dataset to see the trend and a better idea of the dataset. We make sure there aren't problems with the dataset by performing EDA so that we can move on with predicting

Step 3

Applying Prediction Models - We will apply different prediction models that have been taught during the course of Machine Learning practicals to the dataset like

1. Regression
2. Decision tree classifier
3. Naive Bayes classifier
4. Random forest

Step 4

Model selection - Based on the outcomes from our prediction models we will compare them with each other and decide what is the best prediction model to use for predicting the winner between two premier league teams..