**BaseBall Case Study**

Problem Statement:- This dataset utilizes data from 2014 Major League Baseball seasons in order to develop an algorithm that predicts the number of wins for a given team in the 2015 season based on several different indicators of success. There are 16 different features that will be used as the inputs to the machine learning and the output will be a value that represents the number of wins.

Input features: Runs, At Bats, Hits, Doubles, Triples, Homeruns, Walks, Strikeouts, Stolen Bases, Runs Allowed, Earned Runs, Earned Run Average (ERA), Shutouts, Saves, Complete Games and Errors

Output: Number of predicted wins (W)

**Analysis of the project and Output :-**  We have to predict the wins of Major League Baseball season in 2015 on the basis of 2014 dataset. There are 16 different features as input and output will represents the number of wins.

**Model Building Steps :-**

1. First to load dataset into jupyter using Pandas method.

Graphical user interface, table

Description automatically generated

1. Checking the Null Values in the given Dataset. Null values are the missing data which can be handle using Mean, Median and Mode.

Graphical user interface, text, application, email

Description automatically generated

1. Checking the data types.

Graphical user interface, application

Description automatically generated

1. Graphical user interface

   Description automatically generatedIf there is no null values and no empty columns to drop, Now perform Visualization of Data to get insights of the data.

Graphical user interface

Description automatically generatedGraphical user interface

Description automatically generatedChart, histogram

Description automatically generatedChart, histogram

Description automatically generatedChart, histogram

Description automatically generatedChart, histogram

Description automatically generatedChart, histogram

Description automatically generatedChart, histogram

Description automatically generatedChart, histogram

Description automatically generatedChart, histogram

Description automatically generatedChart, histogram

Description automatically generated

1. After performing all the insights with each column, we came to know that there are lot of outliers present.
2. Before removing the outliers, first Describe the data and then perform correlation. To get better insights of it use Heatmap.

Graphical user interface, application, table

Description automatically generated

Graphical user interface

Description automatically generated with low confidence

1. Darker shades in heatmap represents higher values as compared to lighter shades. Checking the Correlation and performing heatmap.

Graphical user interface, application, table, Excel

Description automatically generatedChart

Description automatically generated

1. Checking the skewness with a threshold value of -/+ 0.5.

Graphical user interface, text, application

Description automatically generated

1. Checking the outliers using Box plot.

Graphical user interface, text, application, email

Description automatically generatedGraphical user interface, chart, box and whisker chart

Description automatically generatedGraphical user interface, box and whisker chart

Description automatically generatedChart, box and whisker chart

Description automatically generated

1. Removing the outliers using zscore or IQR.

Graphical user interface, application, table

Description automatically generated

1. Separating the data into Features and Label.

Graphical user interface, text, application, email

Description automatically generated

1. Transforming the data using power transform
2. Scaling the data using Standard Scaler if result of transformation is between -3 to +3 or else MinMaxScaler.
3. Training and Testing the data and Model Selection

Graphical user interface, text, application, email

Description automatically generatedChart, scatter chart

Description automatically generated

1. Regularization of data

Graphical user interface, text, application, email

Description automatically generated

1. Ensemble Technique

Graphical user interface, text, application, email

Description automatically generated

1. Saving the Model and Concluding the Model.

Graphical user interface, text, application, email

Description automatically generated

Model is giving almost 76% result which is quite good.