Cricket Analysis using machine learning

December 12, 2022

-By Team-A

Abstract

This proposed work is for research purposes and for detecting the overall performance of the team And as the popularity of the sport is increasing day by day it will very helpful for sports players. This proposed work will predict the scores of teams based on team play, batting score, wickets taken, etc. To reduce human intervention and make work feasible

1. Problem Statement

The problem statement is to build the Machine Learning Model that predicts the scores of the teams. To accurately detect the defect performance of the players.

2. Market/Customer/Business Need Assessment

The target market would be all the cricket players around the world and the cricket associations.

The funds for this project can be raised based on the requirements of the model we can make a big model for people like private firms, cricket associations, sports analysts, etc.

Cost for research and development may be provided by the requirements of clients' needs but this proposed work only consists of model building.

3. External Search(Information and Data Analysis)

These are some of the sources I visited for more information links are below:

1. https://www.espncricinfo.com/series/world-cup-league-2-2019-2023-1196667/points-table-standings
2. https://www.cricketworld.com/crickets-popularity-around-the-world/83309.htm#:~:text=Cricket%20is%20the%20most%20widely,total%20population%20of%20the%20planet.

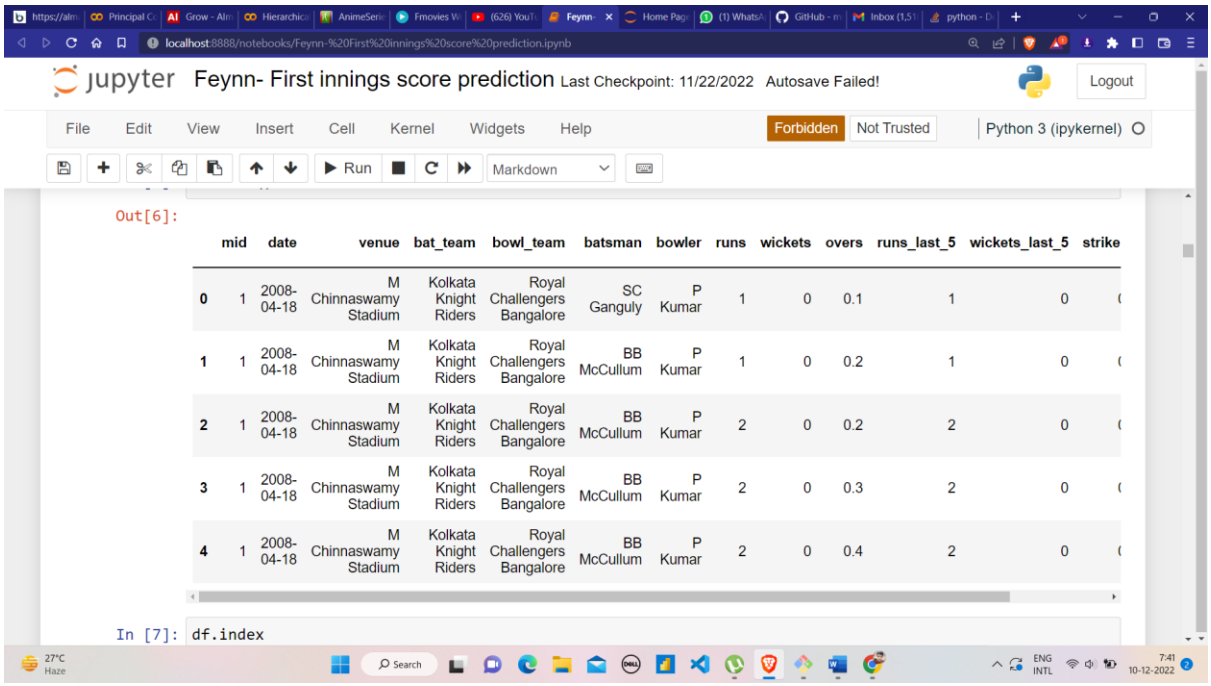
First import the basic libraries for feature extraction from images:

```
import cv2
import numpy as np
import pandas as pd
import matplotlib.pyplot as plt
import seaborn as sns
import plotly.express as px
```

second import the basic libraries for model training from

```
import pandas as pd
from sklearn import metrics
from sklearn.model_selection
import train_test_split
```

shown below is our dataset

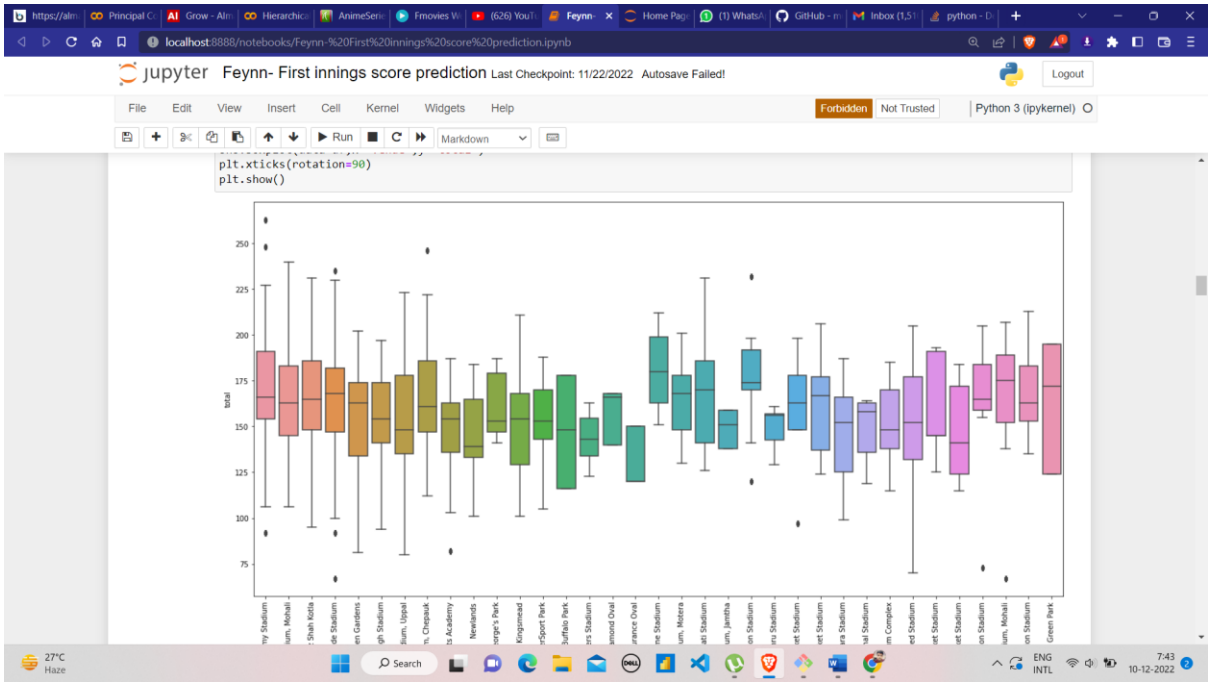


	mid	date	venue	bat_team	bowl_team	batsman	bowler	runs	wickets	overs	runs_last_5	wickets_last_5	strike
0	1	2008-04-18	M Chinnaswamy Stadium	Kolkata Knight Riders	Royal Challengers Bangalore	SC Ganguly	P Kumar	1	0	0.1	1	0	
1	1	2008-04-18	M Chinnaswamy Stadium	Kolkata Knight Riders	Royal Challengers Bangalore	BB McCullum	P Kumar	1	0	0.2	1	0	
2	1	2008-04-18	M Chinnaswamy Stadium	Kolkata Knight Riders	Royal Challengers Bangalore	BB McCullum	P Kumar	2	0	0.2	2	0	
3	1	2008-04-18	M Chinnaswamy Stadium	Kolkata Knight Riders	Royal Challengers Bangalore	BB McCullum	P Kumar	2	0	0.3	2	0	
4	1	2008-04-18	M Chinnaswamy Stadium	Kolkata Knight Riders	Royal Challengers Bangalore	BB McCullum	P Kumar	2	0	0.4	2	0	

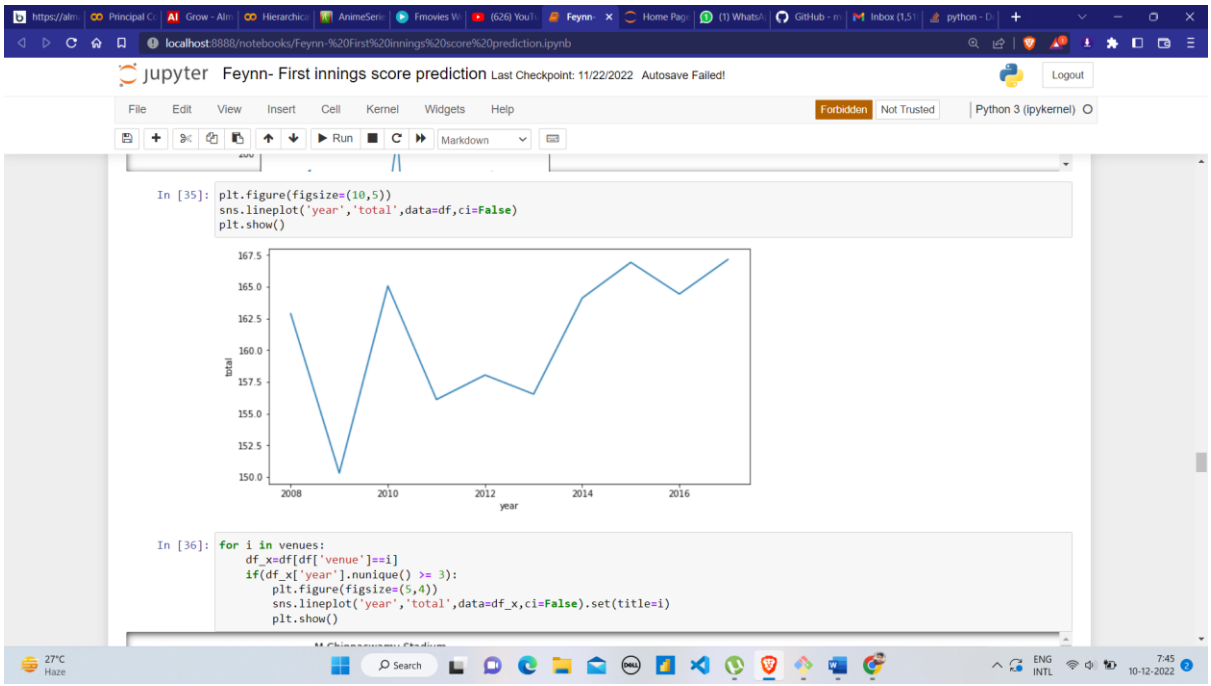
The data set is divided into training 80% and testing 20%,80%will be our training data and 20% will be our test data.

After loading the dataset we our doing some eda cleaning the data, removing outliers, removing na values etc.

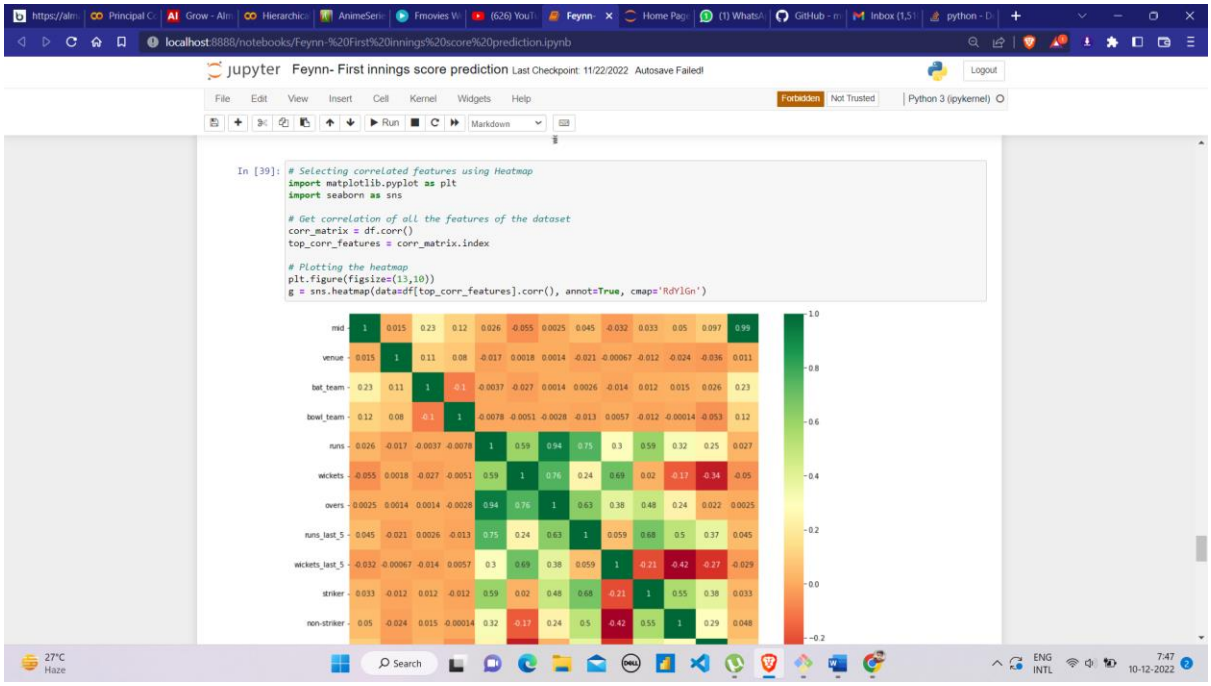
Below the data shown some outliers.



Checking the trend of the data set.



Checking the corelations in the data set



After doing the eda we have perform some feature selection process on the data set:

1. Applying one hot encoder on the dataset
2. Splitting the dataset into train_test_split
3. Applying the different algorithms to increases the accuracy of the model.

A) Feasibility

This project can be developed and deployed within a few months it can be built by python or using only machine learning techniques.

B) Viability

As the cricket industry grows in India and the world, there will always be small cricket associations or team clubs that can use this service to improvise on their techniques. So, it is

viable to survive in the long-term future as well but improvements are necessary as new technologies emerge.

6. Prototype Development

GitHub Link:
https://github.com/ajaysable976/cricket_analysis