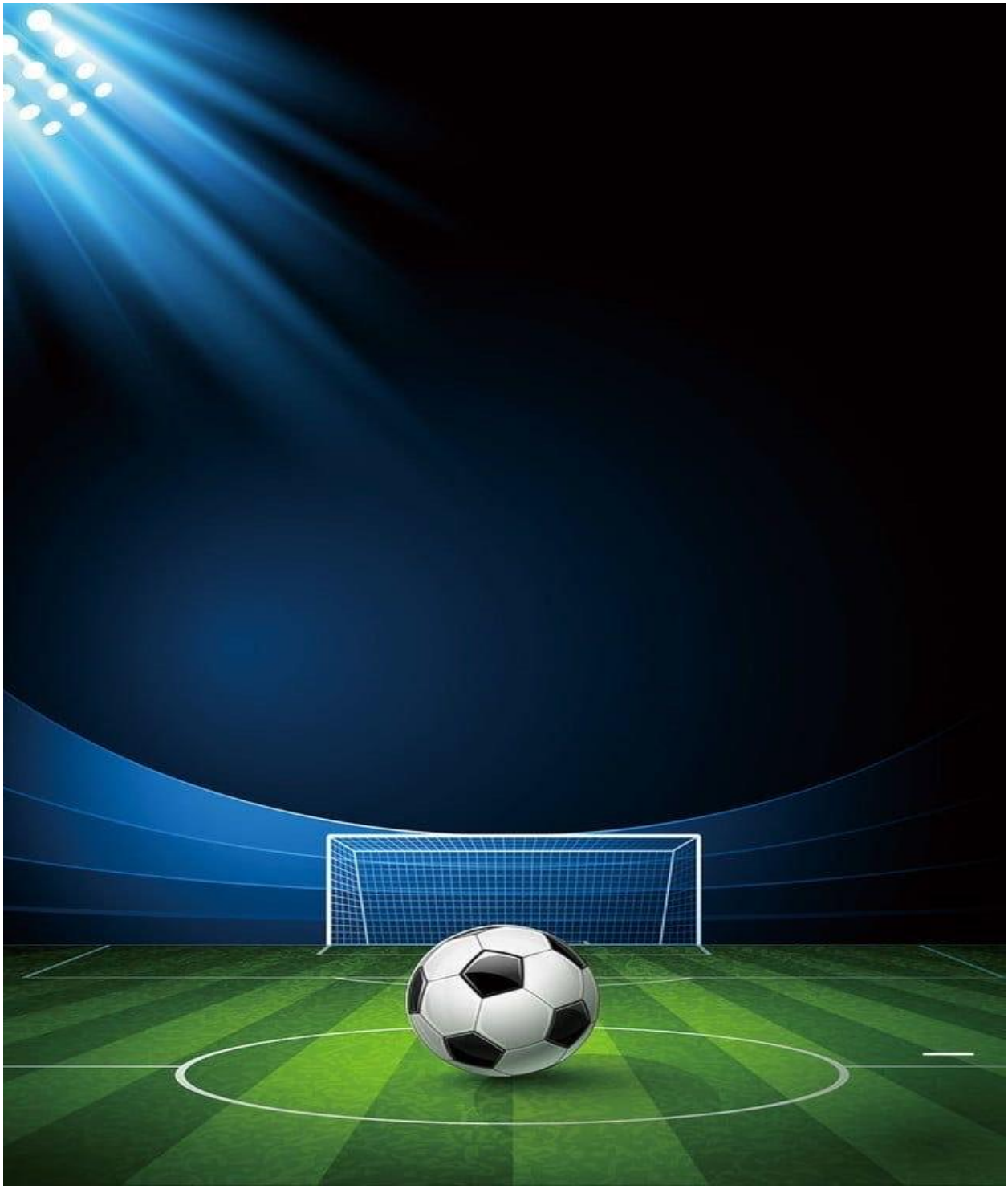
MChezoPesa Limited



Business Understanding.

Introduction

Mchezo Pesa is a sports news technology and sports betting brand company located in Nairobi Kenya. MchezoPesa has a number of products including Sportsbook Betting, Casino, Virtual Games and E-Sports Betting Betika has set the minimum bet at only Kshs 10.

Overview

The FIFA Men's World Ranking is a ranking system for men's national teams in association football. The teams of the men's member nations of FIFA, football's world governing body, are ranked based on their game results with the most successful teams being ranked highest. The rankings were introduced in December 1992.

A points system is used, with points being awarded based on the results of all FIFA-recognised full international matches. It is called the Elo rating system.

Elo ratings are a numerical way to rank the skill level of various competitors based on their performance as well as the relative skill level of their opponents. The idea is that beating a tougher opponent should give you more points than beating an easier opponent.

The study is meant to predict the number of scores made by both the away team and home teams.

The Question

A prediction result of a game between team 1 and team 2, based on who's home and who's away, and on whether or not the game is friendly (include rank in your training).

The Metric of Success.

The research will be viewed and deemed successful if the model could make a prediction of at least 80%.

Data Understanding.

We have been provided with two data sets one for FIFA rankings and another one for results and for Fifa ranking it has data from year 1993 to 2019, for results it has data from 1872 to 2018 , This might pose a challenge when it comes to merging the data. The other thing is the fifa ranking dataset had a lot of null values between 1993 and 2010.The data in Fifa Ranking was being posted monthly and for results it was being posted every date the teams played.

Data Cleaning and Preparation.

Step 1: Remove duplicate or irrelevant observations

We will not use data from 1872 to 2010 in results data, also data in 1993 to 2010 in ranking because it has a lot of null values.

Step 2: Fix structural errors.

We renamed some columns since some were very long and also bring uniformity

Step 3: Filter unwanted outliers and anomalies

We dropped outliers because it is a very special occasion where a home team and away team would score above 5 goals. In this data we had an anomaly of 30 goals.

Step 4: Handle missing data

We dropped the data in our fifa ranking data from 1993 to 2010 because of the missing values.

Step 5: Validate and QA

The data is valid for the project and it makes sense , it is also appropriate in such a way that it can be applied in this field of football. It approves and provides insights to our project.

Merging the data Sets

We merged both data sets.

Exploratory Data Analysis

We performed univariate, bivariate analysis to better understand the data and to see the central tendency , to compare and display relationships in the columns

Modeling.

We did cross validation of the datasets
We performed polynomial regression and logistic Regression

Conclusion,

Our Model in Polynomial approach did best when our degree is 1 for home score and when our degree is 7 for away score .
Cross Validation model number 4 is the best for predicting our home score and away score.
The Logistic approach regression the accuracy test shows that our algorithm is approximately 31.15% accurate in predicting the match status of the home team.

Recommendation.

The first solution brought almost a perfect model which means that there could be changes that could be made. We could have maybe added the weighted points and previous points. Also, some factors such as injury on players should have been included in the dataset and players being given a red card and benched can affect the outcome of their next 2 games.