Statement of Work

Title**:FIFA World Cup Prediction**

Rationale Statement:

Sport prediction is one of the growing areas that requires high predictive accuracy because of the enormous monetary stakes involved in betting. Furthermore, club managers and owners also look for classification models to help them understand and design winning strategies. These models are based on a variety of game-related factors, including past match results, player performance indicators, and opponent information.

Football is one of the famous sports and making a prediction analyser for the game is an interesting topic. As football is a famous sport worldwide and the FIFA world cup is also approaching, everyone is curious about the winner of the world cup .The goal of this project is to predict who is going to win the next FIFA World Cup 2023 using the past datasets.

Problem Statement :

A lot of people watch soccer and they are always keen about these matches and tournaments. Today, people even invest in these matches by betting through different mobile applications. A proper analysis is required for each player's matches.Our project will predict the upcoming FIFA world cup result using the available datasets of each match and player history. This will help the investors to build proper winning strategies.

Objective:

1. To create a machine learning model capable of predicting the outcomes of football games in the 2023 FIFA World Cup.
2. Predict which team is going to win the finals.

Data Sources:

In this project, we are using multiple datasets which are interlinked to each other, two of them having more than 40 thousand records.This dataset contains all available FIFA men's international football rankings from August 1993 to April 2018 and international football results. It is a huge collection of data on matches spanning more than a century of international football games. Currently, the following datasets are being used from Kaggle. Additionally, more data can be collected from official football association websites.

1. FIFA football Rankings from 1993 to 2018

2. International football results from 1870 to 2021

3. World cup 2018

Data Assumption:

We are making the winning team column in categorical data by changing column values to “2” if the home team won, “0” if the away team won, “1” if it’s a tie match.

Data Limitation:

FIFA ranking was created in the 90’s, thus a huge portion of the dataset is lacking. So,we will use the results of historical matches since the beginning of the championship for all participating teams.

A few record values are not available and null, which will further be cleaned.

Data Constraint:

1) The name of the country at the time of the match is used for country names ,even if the names of the home team and the country don't match. For example, Ghana was playing at home in Accra, Gold Coast in the 1950s. The neutral column for those matches reads FALSE, indicating that they did not take place in a neutral site.

2) The current name of the team has been used for both home and away teams. For example, when a team known as Ireland played versus England in 1882, it is referred to as Northern Ireland in this dataset because the current Northern Ireland squad is the heir to the 1882 Ireland side. This is done to make it easier to keep track of a team's history and statistics.

Work Deliverables:

1. We will do the exploratory data analysis on FIFA football Rankings and International football Results datasets.
2. Later, we will do some feature engineering to select the most relevant feature for prediction.
3. In the next stage, we will complete data manipulation to build a machine learning model using logistic regression, random forest algorithm.
4. Train and test our dataset, eventually predicting the result.
5. Add evaluation metrics like accuracy,precision,recall,etc to evaluate the models.
6. Finally, we will deploy the model on the Dataset.

References:

<https://www.sciencedirect.com/science/article/pii/S2210832717301485>

<https://content.iospress.com/articles/journal-of-sports-analytics/jsa200463>

Team Members :

Smriti Raina – 100830255

Devangraj Vaidya – 100780615

Prasann Patel – 100833325

Samarth Patel – 100827936