PROJECT DELIVERABLE 1

Enrique Aponte
Seif Abdelkefi

1. Project description

We want to build a model that, given an input of two international football teams (Participating in the world cup or even not participating), Team A, and Team B, the model classifies Team A as a winning team or losing team.

2. Methodology

a. Data choice and Processing:

For this model, we will have to analyze a bunch of previous international games, so we can identify the main parameters that decide which team is going to win. In a football game, a lot of parameters can orient the outcome:

- the strength of the team on 5 different levels:
  - ❖ goalkeeping
  - ❖ defence
  - ❖ midfield
  - ❖ offense
  - ❖ average score of the team
- the rank of the team according to fifa
- the historic results of the games between A and B
- is one of teams on a winning or losing streak.
- Home team advantage

Training Set: The kaggle data set contains games that have beel played from the 90’s until june 2022. we can use this data as a training set.
Data: https://www.kaggle.com/datasets/brenda89/fifa-world-cup-2022

Validation Set: For the validation set we could use the games that have been played from june 2022 until the start of the world cup (nation s league and the friendlies)

Testing Set: the test set is the world cup2022 itself.

b. Machine learning model:

Labels or categories :

- Multi-class Classification: Win, lose, tie
- Binary classification: Win or lose and the final outcome can be a tie if p(win) = 0.5

We will predict the chances of winning,drawing and losing from a given match and predict the ranking of teams in the world cup. We will use **Convolutional Neural Networks**

- **Why?**
  - Convolutional Neural Networks are very good at solving classification problems, they are extremely accurate and are mostly used to predict sports games.
- **Pros**
  - Very Accurate
  - Can detect non linear Relationships Between dependent and independent variables
- **Cons**
  - Prone to memorizing the data and not generalizing
  - Computational Heavy
- **Alternatives**
  - Support vector machine
  - Random Forest

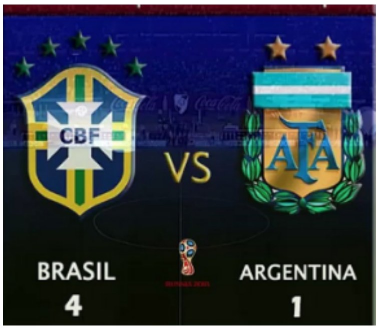
c. Evaluation Metric:

- We will use a Confusion Matrix and accuracy, to measure the number of successful predictions and non-successful ones. The model should be able to to predict the winning it with 55% accuracy

3. Application

- **input:** 2 teams. Dropdown menu with all the teams participating in the WC. ex: (Argentina, Mexico), the program fills the other parameters necessary for the model to work, such as argentina's defence score, midfield score, Mexico's fifa ranking etc...

- **Output:** (team A classification (win,lose or tie), probability of winning?) ex: (Win, 0.56).
  Display:



references

https://www.scirp.org/journal/paperinformation.aspx?paperid=94928
https://towardsdatascience.com/machine-learning-algorithms-for-football-prediction-using-statistics-from-brazilian-championship-51b7d4ea0bc8