PREDICTING FOOTBALL MATCH OUTCOMES

Presented by Ayan Duisenov,
Galamat Sadr

PROJECT DESCRIPTION

* Match Outcome Prediction project is relevant due to football's global popularity and the growing interest in it.
* It addresses the demand for predictive insights among fans, bookmakers, and analysts.
* It can enhance the accuracy of predictions and unlock new opportunities in sports analytics and betting.

Aim: predict the football match result based on the historical data

ROLES

- Galamat ♚

1) Data validation

2) Buidling & testing architecture (RNN)

- Ayan ♚

1) Data exploration

2) Buidling & testing architecture (MLP)

DATA

- Open source data from Kaggle:
- FIFA World Cup Dataset
- La Liga Dataset

Examples of features:

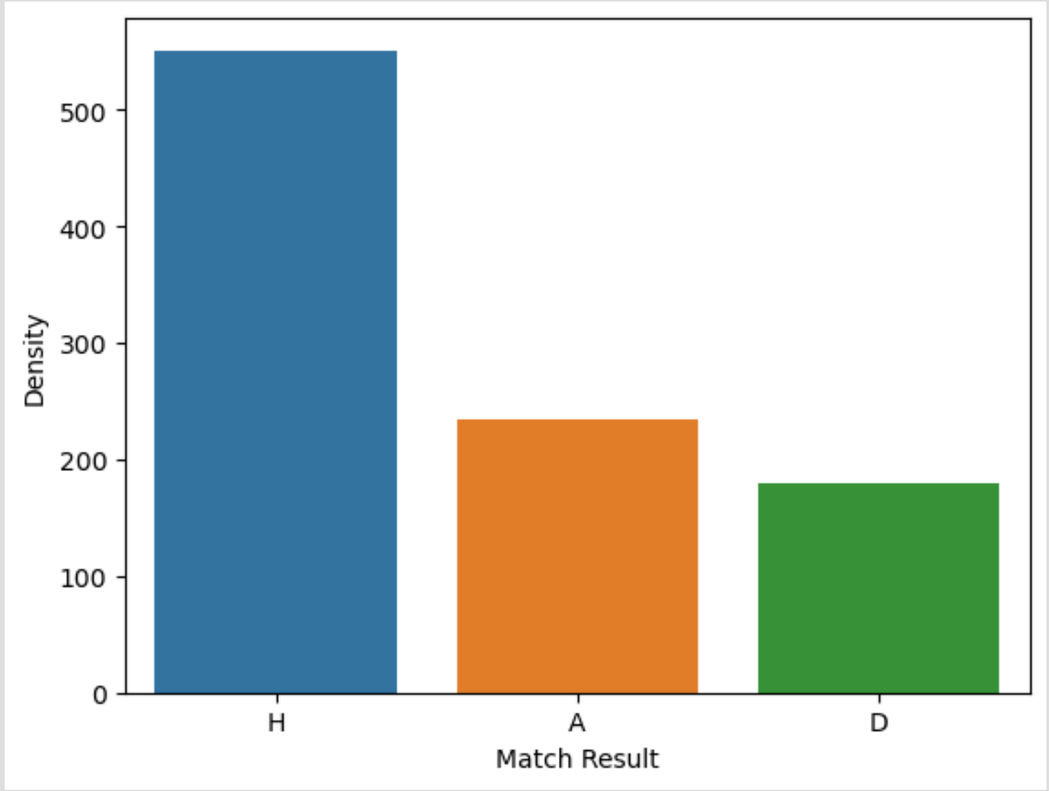
- XG (Expected goals)
- Penalty stats
- Free kick stats
- Captain, Manager

DATA

Example of Data

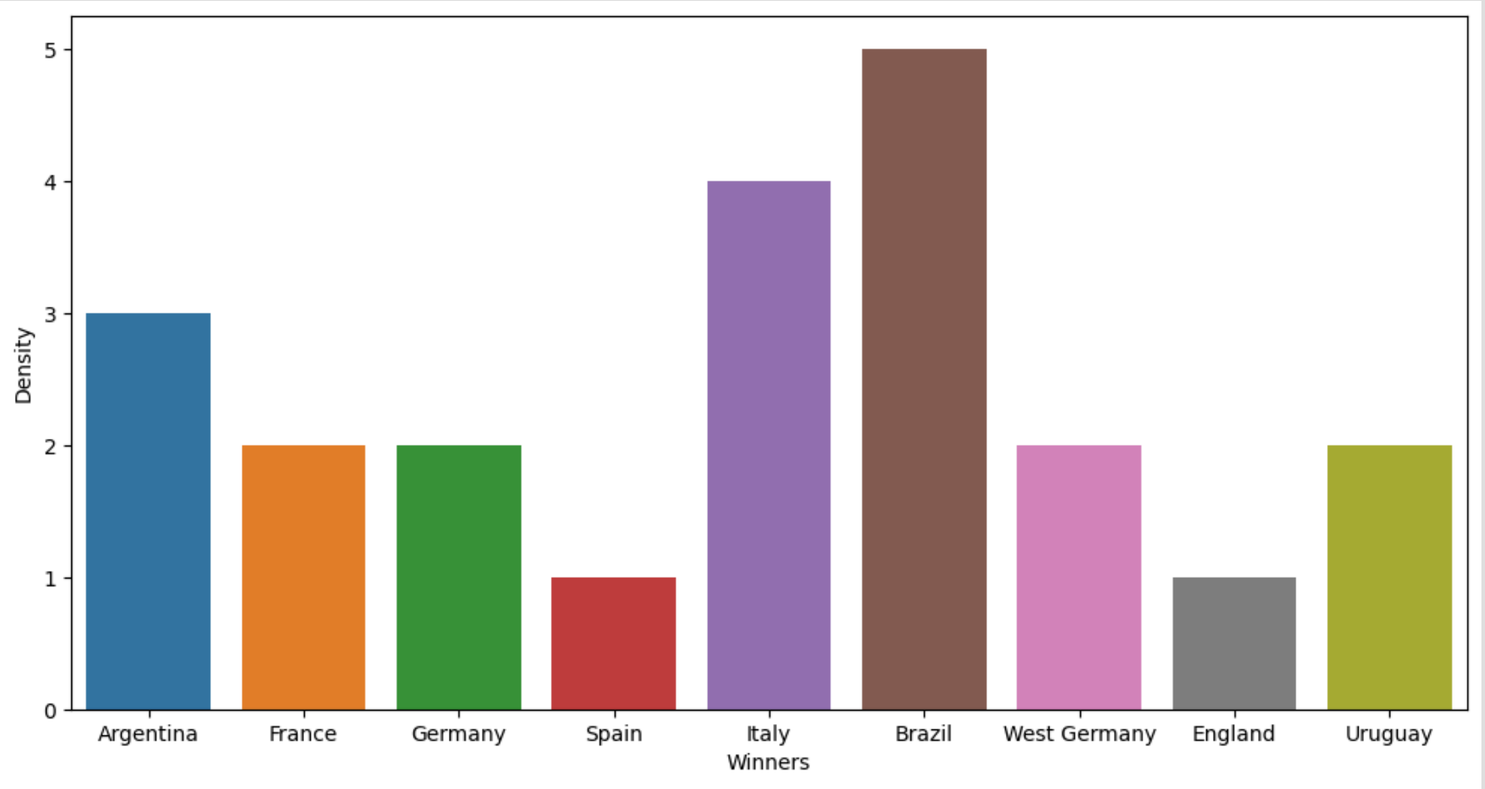
home_team	away_team	home_score	home_xg	home_penalty	away_score	away_xg	away_penalty	
Argentina	France	3	3.3	4.0	3	2.2	2.0	
Croatia	Morocco	2	0.7	0.0	1	1.2	0.0	
France	Morocco	2	2.0	0.0	0	0.9	0.0	D
Argentina	Croatia	3	2.3	0.0	0	0.5	0.0	
Morocco	Portugal	1	1.4	0.0	0	0.9	0.0	

Distribution of classes



DATA

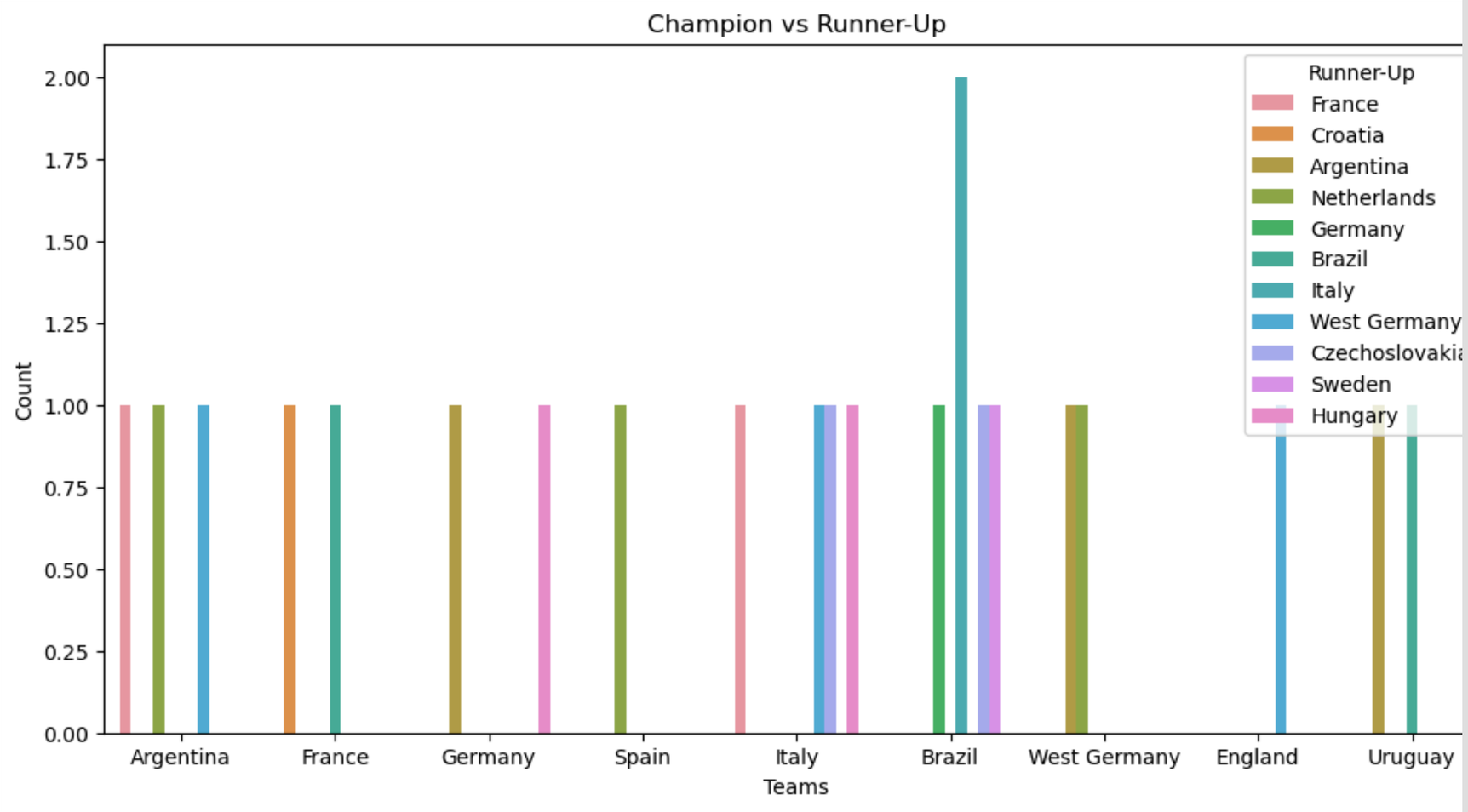
Winners of World Cups 1930 - 2022



The team with the highest number of championship titles is Brazil, followed by Italy and Argentina.

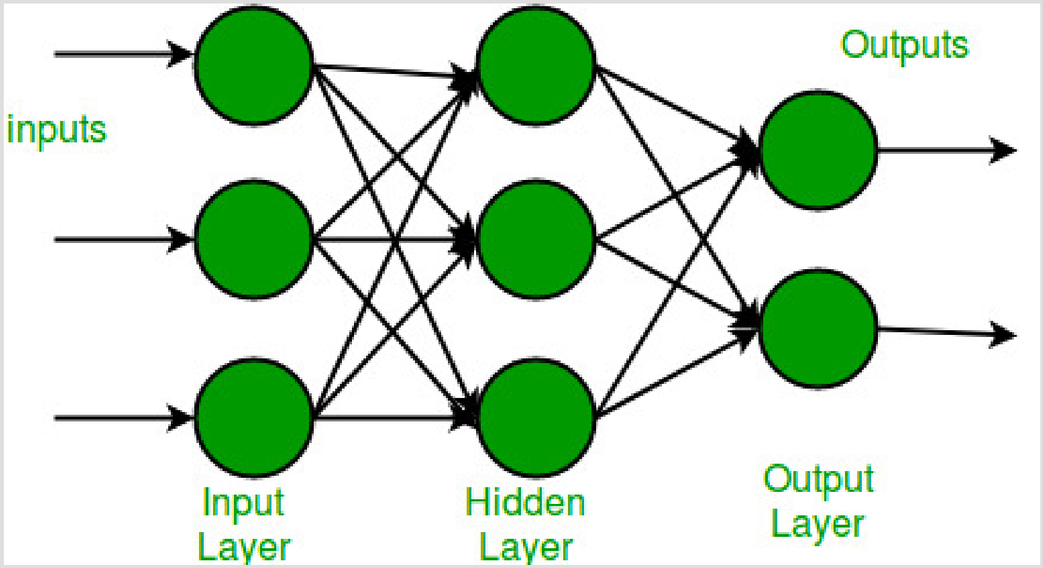
DATA

Teams represented by the number of times they have been Champions and Runners-Up.

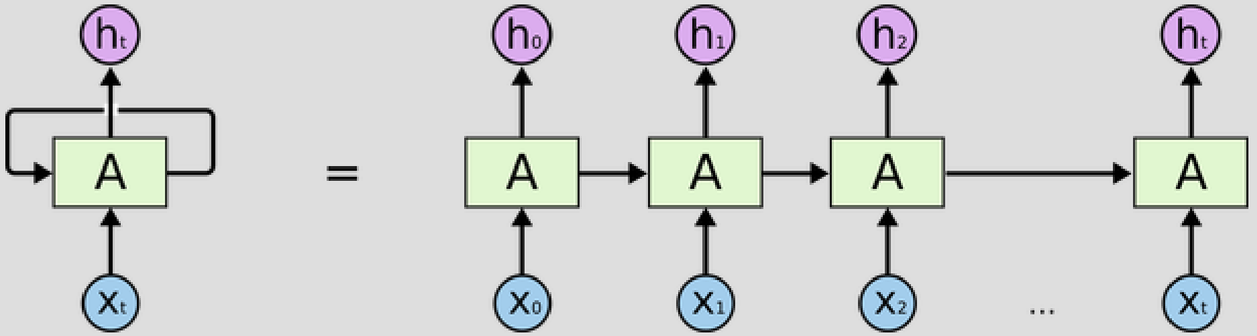


ARCHITECTURES/ALGORITHMS

- MLP (Multilayer perceptron)



- RNN (Recurrent Neural Network)



RELATED WORKS

WORK #1

(Predicting match outcomes of english premier league (EPL))

WORK #2

(Predicting match outcomes of english premier league (EPL) 1500 matches)

If a bet was missed, the profit would decrease by 2 euros. If correct, the profit is calculated according to equation: $Profit = bet_{amount} \times bet_{odd} - bet_{amount}$. For example, in a game with a bet on a tie and the draw odd is 1.5 the profit would be 1 euro. As the value bet is always 2 euros the profit would be equal to 2x1.5-2 = 1. Table 2 presents the results of predictions made with the 8 models developed with the selected algorithms.

Table 4. Analysis of proposed approach

Algorithm	Accuracy (in %)	F1-score (in %)
MLP	73.57	71.45
SVM	58.77	50.07
Gaussian Naive Bayes	65.84	64.26
Random Forrest	72.92	66.07

Table 2. Forecast results with 18 variables.

Algorithm	Accuracy	Profit	% Victories Home Team	Draws	% Victories Away Team
Bayes	53,42%	17,40€	51,87%	30,95%	73,79%
KNN	57,63%	78,02€	78,07%	15,48%	55,05%
RF	59,21%	85,20€	75,40%	21,43%	60,55%
SVM	61,32%	95,06€	88,77%	3,57%	58,72%
C5.0	55,26%	42,52€	72,73%	23,81%	49,54%
Xgboost	59,47%	72,80€	77,54%	10,71%	66,06%
RLM	57,63%	32,56€	78,07%	5,95%	62,34%
RNA	50,00%	18,28€	58,29%	30,95%	50,46%

EVALUATION METRICS

Precision - the quality of a positive prediction made by the model

For each class:

- $\text{Precision}_H = \frac{\text{TP}_H}{\text{TP}_H + \text{FP}_H}$
- $\text{Precision}_A = \frac{\text{TP}_A}{\text{TP}_A + \text{FP}_A}$
- $\text{Precision}_D = \frac{\text{TP}_D}{\text{TP}_D + \text{FP}_D}$

The F1-score is the harmonic mean of precision and recall

$$F1_c = 2 \cdot \frac{\text{Precision}_c \cdot \text{Recall}_c}{\text{Precision}_c + \text{Recall}_c}$$

$$F1_{\text{weighted}} = \frac{\sum_{c \in \{H,A,D\}} (F1_c \cdot \text{Support}_c)}{\sum_{c \in \{H,A,D\}} \text{Support}_c}$$

RESULTS

MLP Results

```
F1 Score: 0.827187603802889
Precision: 0.8269052448565634

Classification Report:
              precision    recall  f1-score   support

           A       0.79      0.77      0.78        35
           D       0.67      0.67      0.67        27
           H       0.89      0.90      0.90        83

    accuracy                           0.83       145
   macro avg       0.78      0.78      0.78       145
weighted avg       0.83      0.83      0.83       145
```

```
Predictions for the Qatar Final match with actual results:
   home_team away_team result predicted_result
0  Argentina    France      H                 H
```

RNN results

```
6/6 [==============================] - 0s 1ms/step
F1 Score: 0.4201085534548268
Precision: 0.445093201754386

Classification Report:
              precision    recall  f1-score   support

           A       0.50      0.02      0.04        47
           D       0.00      0.00      0.00        36
           H       0.57      0.99      0.72       109

    accuracy                           0.57       192
   macro avg       0.36      0.34      0.25       192
weighted avg       0.45      0.57      0.42       192
```

```
Predictions for the separated Qatarfinal match dataset:
   home_team away_team result predicted_result
0  Argentina    France      H                 H
```

FURTHER WORK

Future work:

- Add more data about matches in club league games
- Test on new football matches
- Data augmentation allows us to use LSTM
- Add context encoders on dataset

Work on RNN results

```
6/6 [==============================] - 0s 1ms/step
F1 Score: 0.4201085534548268
Precision: 0.445093201754386

Classification Report:
              precision    recall  f1-score   support

           A       0.50      0.02      0.04        47
           D       0.00      0.00      0.00        36
           H       0.57      0.99      0.72       109

    accuracy                           0.57       192
   macro avg       0.36      0.34      0.25       192
weighted avg       0.45      0.57      0.42       192
```

```
Predictions for the separated Qatarfinal match dataset:
   home_team away_team result predicted_result
0  Argentina    France      H                 H
```

REFERENCES

- Rudrapal, Dwijen & Boro, Sasank & Srivastava, Jatin & Singh, Shyamu. (2020). A Deep Learning Approach to Predict Football Match Result. 10.1007/978-981-13-8676-3_9.

- Naccarato, K., & Zanni-Merk, C. (2022). Smart mobility: Towards new public transportation systems based on big data and artificial intelligence. Procedia Computer Science, 204, 864–873. https://doi.org/10.1016/j.procs.2022.08.057
- Sherstinsky, A. (2020). Fundamentals of recurrent neural network (RNN) and long short-term memory (LSTM) network. Physica D: Nonlinear Phenomena, 404, 132306.
- Zou, J., Han, Y., & So, S. S. (2009). Overview of artificial neural networks. Artificial neural networks: methods and applications, 14-22.