

# A Project on Python: Winner Prediction of World Cup 2026

**Course No: CSE-346**

**Course Title:** Artificial Intelligence (Sessional)

**Submitted By:**

**Abu Bakkar Siddique Raihan**

ID: 1804004

**Saikat Chowdhury**

ID: 1804013

**Binoy Das Gupta**

ID: 1804032

**Submitted To:**

**Ashim Dey**

Assistant Professor,  
Dept. of CSE, CUET

**Avishek Das**

Lecturer,  
Dept. of CSE, CUET

**Date of Submission:** 08-02-2023



Chittagong University of Engineering and Technology  
Dept. of Computer Science and Engineering

# Contents

<b>1</b>	<b>Introduction</b>	<b>1</b>
<b>2</b>	<b>Methodology</b>	<b>1</b>
<b>3</b>	<b>Tools and Technologies</b>	<b>3</b>
<b>4</b>	<b>Result</b>	<b>3</b>
<b>5</b>	<b>Conclusion</b>	<b>8</b>

## List of Figures

1	Steps of Proposed Model . . . . .	2
2	Group Stage XGBOOST model . . . . .	3
3	Group Stage Tuned XGBOOST model . . . . .	4
4	Knockout Stage Tuned Random Forest model . . . . .	5
5	Knockout Stage Tuned XGBOOST model . . . . .	6
6	Final Result . . . . .	8

# 1 Introduction

The FIFA World Cup is the biggest and most prestigious tournament in the world of football. Scheduled to take place every four years, it brings together the best national teams from around the globe to compete for the title of world champion. The tournament will feature a record 48 teams, who will compete in stadiums across North America, South America, and Central America. The World Cup is an event that generates a lot of excitement and passion among fans in our country and we are among them.

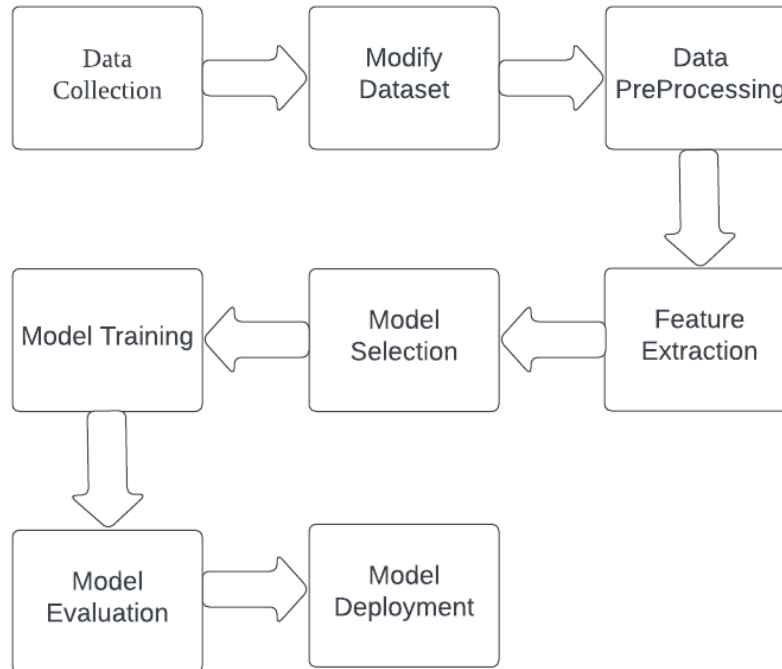
Predicting the outcome of a football game is a task that requires a deep understanding of the teams involved, their playing styles, recent form, and many other factors. In order to make a prediction, it is important to consider all of these factors and to analyze the strengths and weaknesses of both teams. It is a highly complex event with many variables that can influence the outcome of each match, and making accurate predictions can be challenging. However, machine learning algorithms can process large amounts of data, identify patterns, and make predictions based on that analysis.

## 2 Methodology

The following 8 steps can be used to breakdown the process for applying machine learning to make predictions for the 2026 FIFA World Cup:

1. **Data collection:** The first step in making predictions using machine learning is to gather a large and diverse set of data that can be used to train the model. We collect the data from kaggle.
2. **Modify dataset:** Then we modify our collected data by adding almost 130 new row. We add all matches of last FIFA world cup. We use this modified data for predicting the result.
3. **Data pre-processing:** After the data have been modified, they need to be cleaned up and corrected in order to remove any errors or unnecessary informatio.

4. **Feature extraction:** From the pre-processed data, significant features that are likely to affect the World Cup's outcome must be found and extracted.
5. **Model selection:** After pre-processing the data and determining the features, we have to choose the best machine learning model for making predictions. Comparing a variety of models and choosing the one that performs the best on validation data.
6. **Model training:** After the model has been chosen, it is trained using the pre-processed data.
7. **Model evaluation:** Once the model has been trained, it must be examined in order to determine its precision and dependability. This could entail comparing the model's predictions to the actual outcomes <https://www.overleaf.com/project/63e2454f7a8c8c28cca87a9b>
8. **Model deployment:** Finally, the machine learning model can be deployed to make predictions for the 2026 World Cup.



**Figure 1:** Steps of Proposed Model

By following this methodology, it should be possible to make a better predictions for the

2026 Football World Cup using machine learning.

### 3 Tools and Technologies

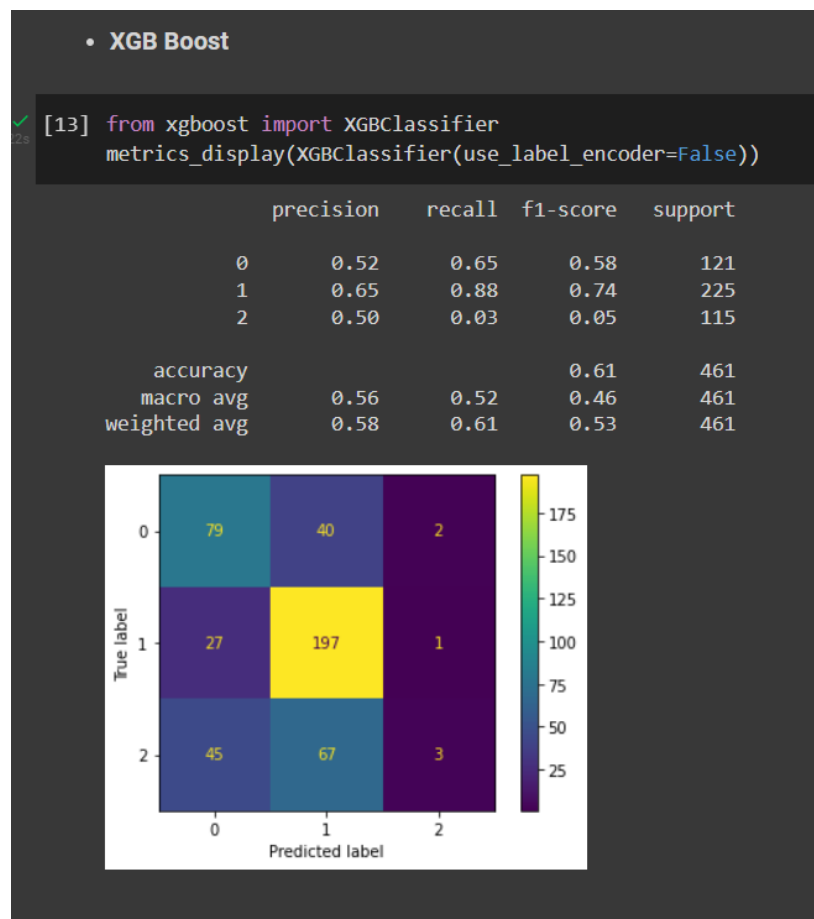
**Languages:** Python

**IDE:** Google Colab

**Python Library Used:** matplotlib, scikit-learn, numpy, pandas, seaborn

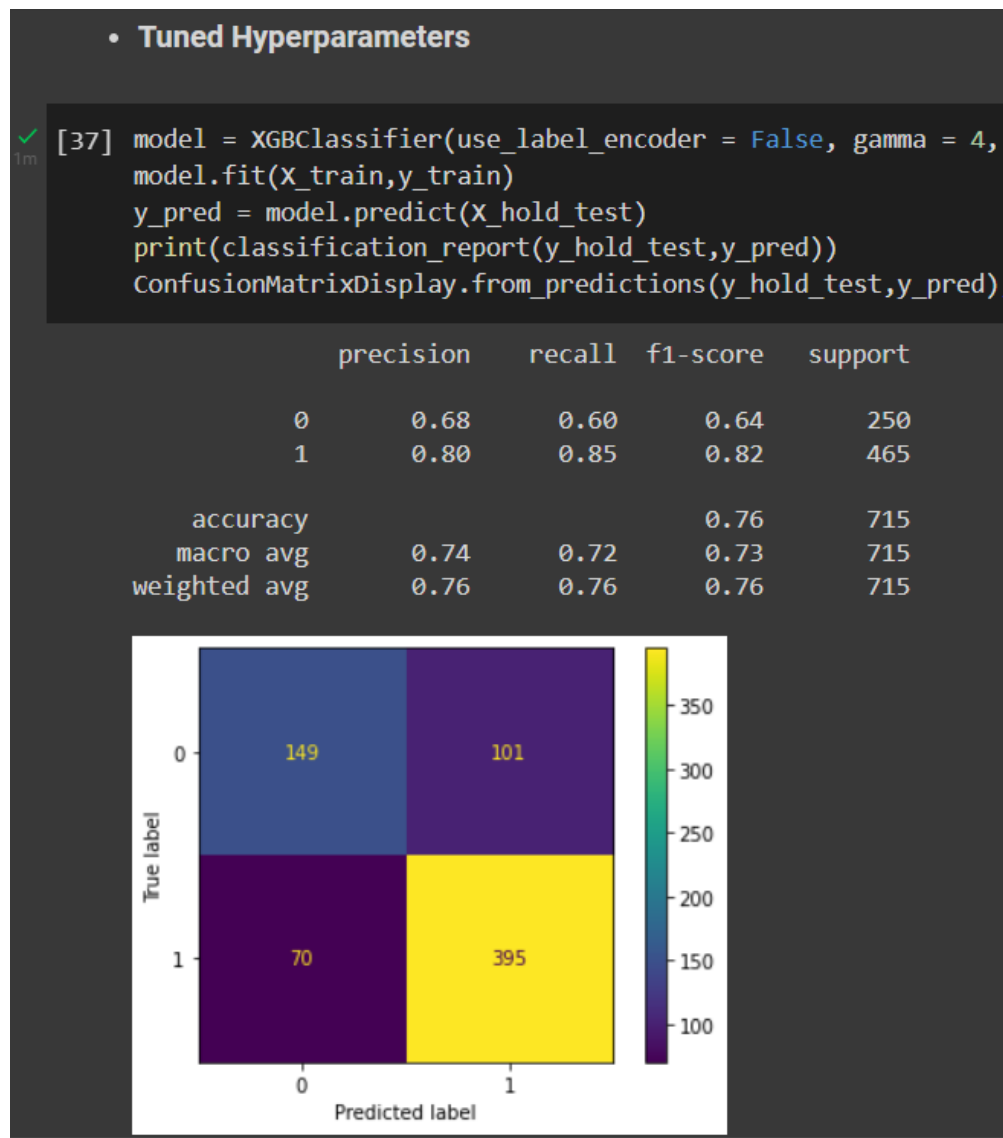
### 4 Result

We checked different models, XGBOOST model performs better than others. We show the confusion matrix and f1-score for better understanding.



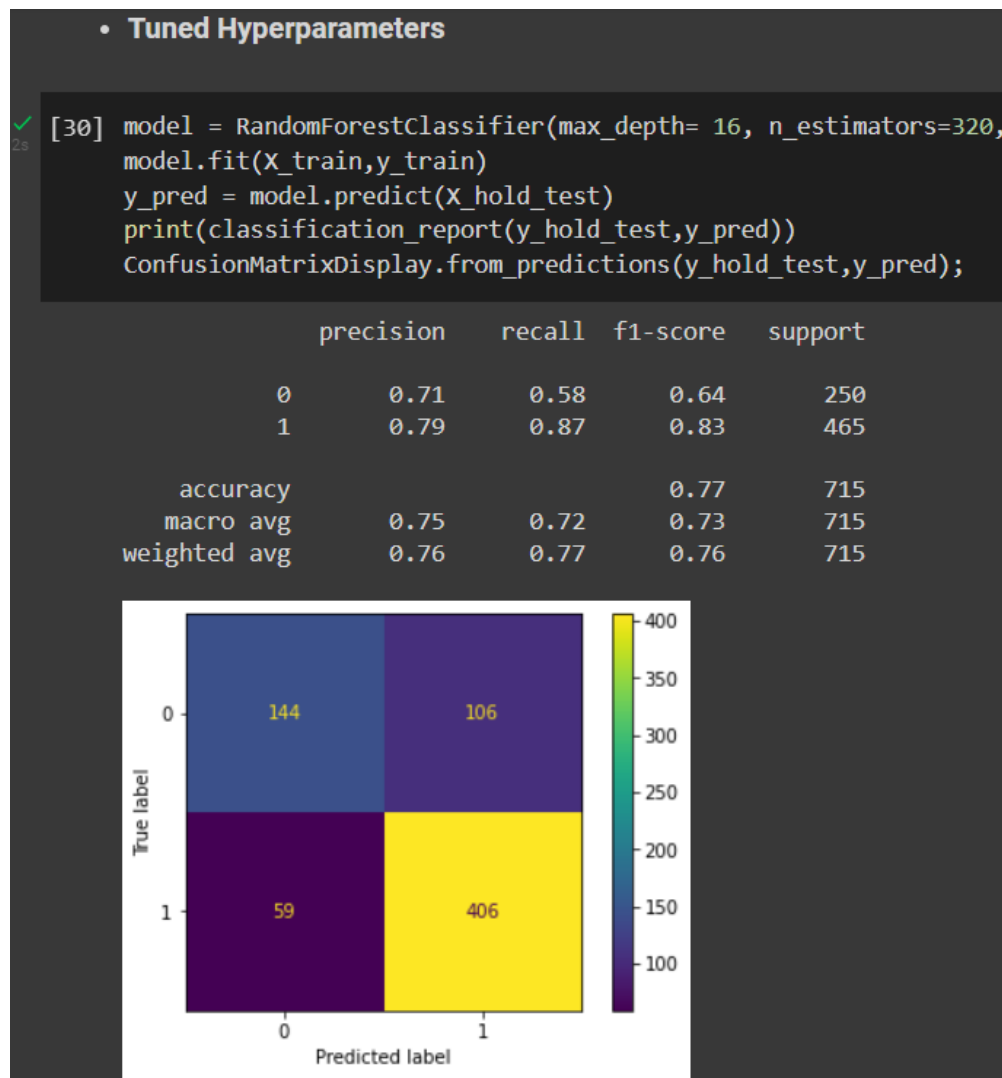
**Figure 2:** Group Stage XGBOOST model

This is the tuned XGBOOST model which performs better than default XGBOOST. We show the confusion matrix and f1-score for better understanding.



**Figure 3:** Group Stage Tuned XGBOOST model

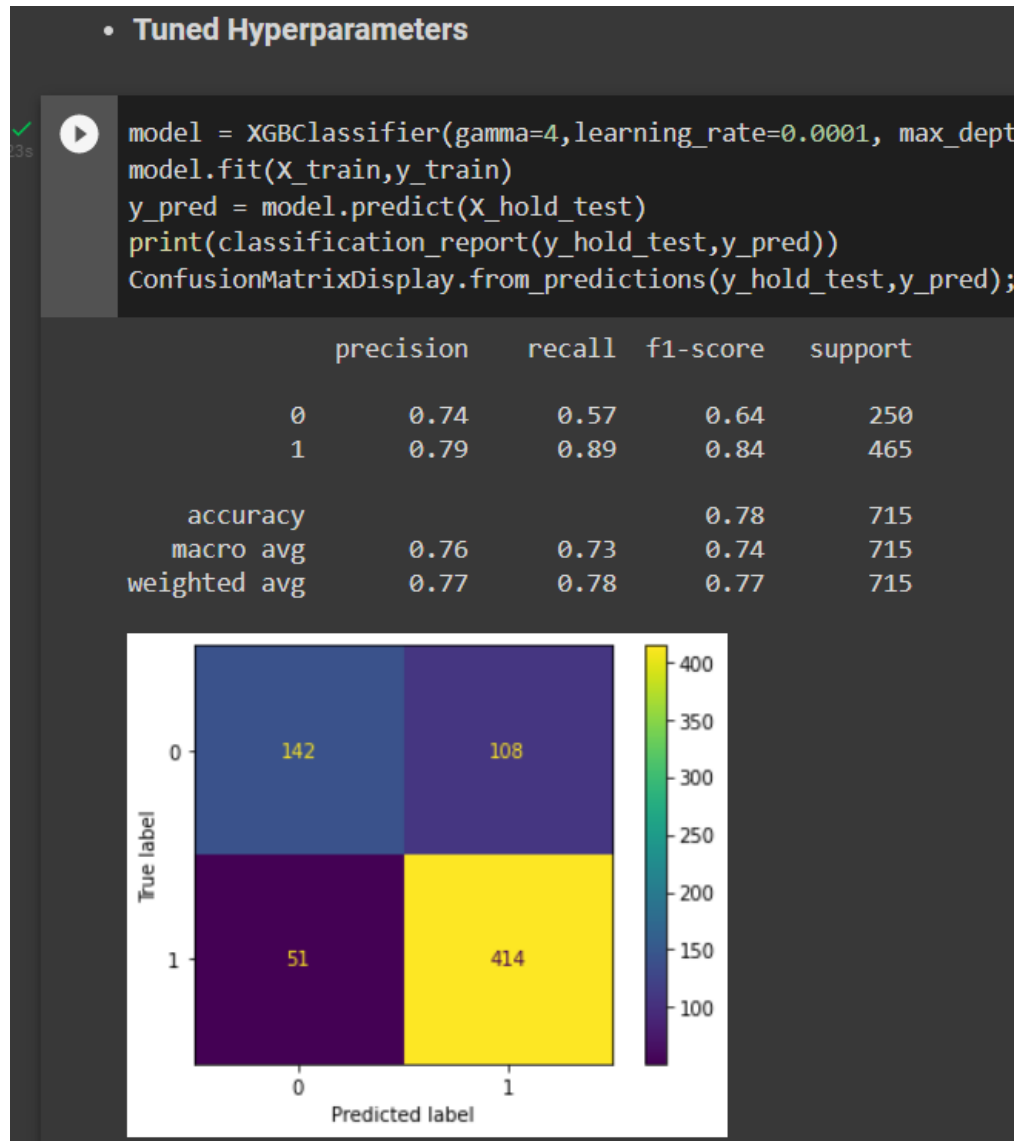
For knockout stage we show here the tuned Random Forest mdoel. We show the confusion matrix and f1-score for better understanding.



**Figure 4:** Knockout Stage Tuned Random Forest model



This is the tuned XGBOOST model for knockout stage which performs better than default XGBOOST or other model. We show the confusion matrix and f1-score for better understanding.



**Figure 5:** Knockout Stage Tuned XGBOOST model

Models	Average F1 Score
Random Forest	59%
Gradient Boosting	60%
Neural Network	53%
Ada Boost	59%
XGBoost	61%
Tuned XGBoost	78%

**Table 1:** Group Stage Models Accuracy

Models	Average F1 Score
Random Forest	73%
Gradient Boosting	76%
Neural Network	75%
Ada Boost	73%
XGBoost	76%
Tuned Random Forest	77%
Tuned XGBoost	78%

**Table 2:** Knockout Stage Models Accuracy

This is the comparison of the models we used in group stage and knockout stage. In both cases, XGBoost gives best accuracy. Tuned XGBOOST gives 78

Final result of our predicted model.



**Figure 6:** Final Result

## 5 Conclusion

A comprehensive dataset and utilizing sophisticated algorithms and techniques, machine learning models can provide a reliable predictions that can inform fans, analysts, and decision-makers. We modified the existing dataset and used this dataset for worldcup 2026. This Predictions are trained on, and that their accuracy can be affected by many factors, such as the quality and quantity of the data, the choice of algorithms and models.