Hansika Gunasekara
2nd December 2025

Gemini Powered Football Commentators

The New Era of Football Storytelling

Inspiration

- Football is played by 250 million players in over 200 countries
- Football is the most popular sport globally
- The English Premier League(EPL) is the most popular domestic team in the world

Let’s Predict the Winning Football Team

01 Data Manipulation

02 Feature Engineering

03 ML Prediction

04 Performance Evaluation

05 Next Steps

All About the Data

What is the less error-prone SQL query to merge data tables? How the BigQuery Data is accessed from Colab?

01

Data Upload to BigQuery

The data tables include a variety of columns on match results, match statistics & betting data.

Few column names include symbols like `<`. When uploading the dataset set auto detect schema. Then changed the `Column name character map` to V1 to support special characters in columns

02

SQL to Merge Tables in BigQuery

To merge 21 data tables with the same schema there are two options for merge queries.

1. Wildcard table query (Less error prone, simple)
2. `UNION ALL` query (for irregular table names)

03

Load data into Pandas

Merge 21 data tables in BigQuery before pulling it into Colab. It offloads the heavy processing to cloud infrastructure

Feature Engineering

01

Rolling goal features

For each match, a n rolling window is used to count last n goals & last n goals conceded

It is important to prevent data leakage

02

Weighted form score feature

A weighted form score over the last 5 matches:

Win - 3, Draw - 1, Loss - 0

Important to avoid data leakage

03

Home & Away Advantage feature

Features to isolate home specific performance

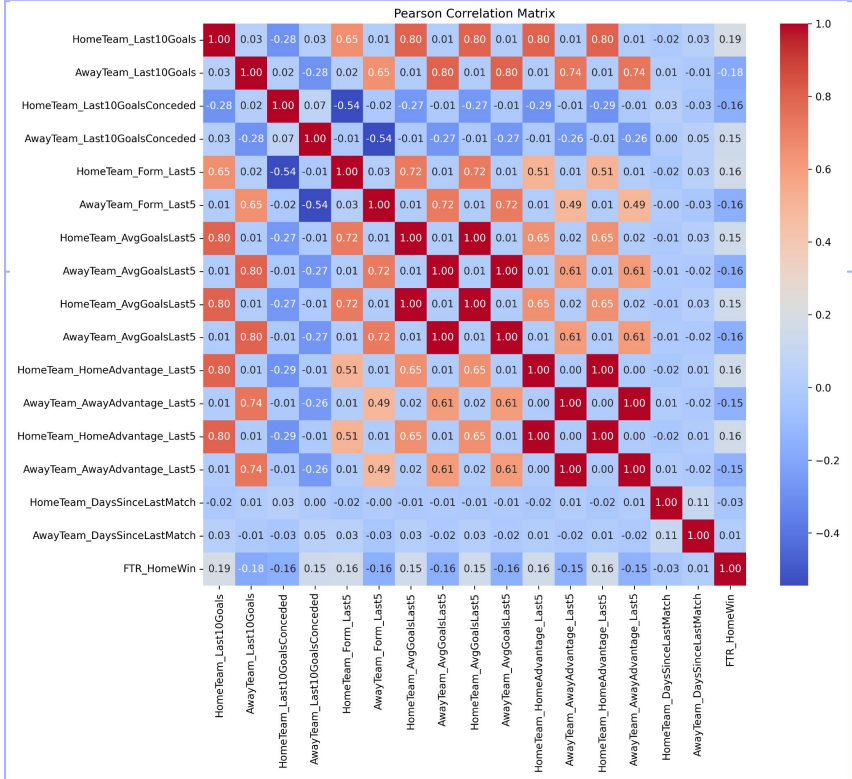
Useful because some teams perform differently away

04

Days since last match feature

For each team compute the difference in days between current match and the previous match

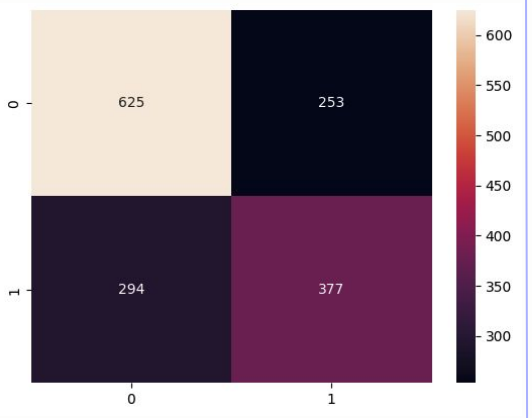
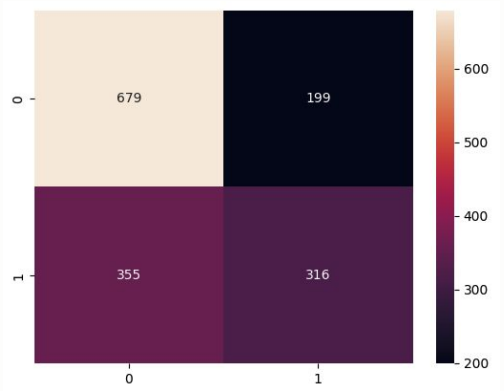
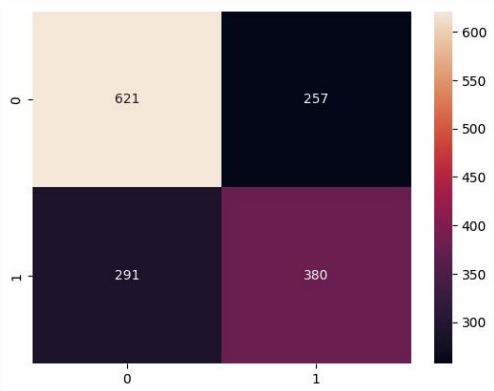
Feature Engineering



Home Team Win Rate

45.59%

ML Prediction

	Logistic Regression	Random Forest	Ensemble Learner
Confusion Matrix			
Accuracy	65%	64%	65%

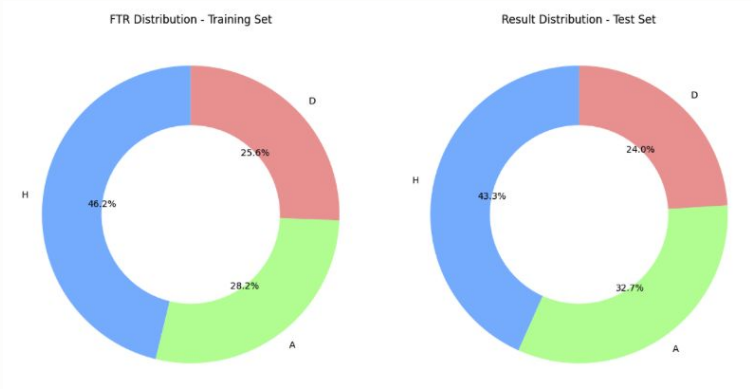
Performance Evaluation

16

Number of Features

5800: 1500

Train Dataset : Test Dataset



65%

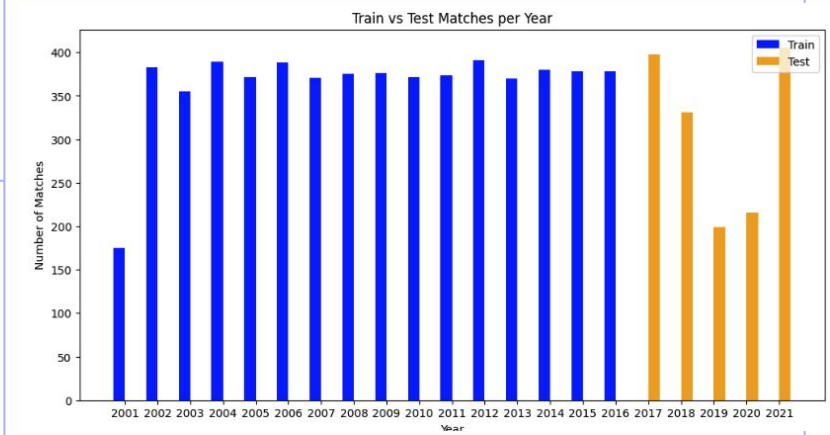
Accuracy

64%

Precision

64%

Recall



NEXT STEPS

STEP 1

Explore more features - Win streak, Fail Streak, etc..
Parameter Tuning
Model Porting
Bootstrap model as a Sliding Window for training & prediction

STEP 2

AI Agent Solutions
Kaggle Writeup / Check the youtube video

STEP 3

Lead Generation - Football coaching, Football commentators

THANK YOU

CONTACT ME
hansikagunasekara@gmail.com