



Big Data Project

Team #1

- Ahmed Amin Sayed
- Yousef Mahmoud Ibrahim Gilany
- Mahmoud Khalid Mahmoud
- Ahmed Magdy Abdelaziz

Overview

The project aims to leverage **historical football data** to predict which team will win the championship in a given season. This analysis not only satisfies the curiosity of sports enthusiasts but also provides valuable insights for betting companies, sports analytics firms, and clubs themselves, enhancing decision-making related to investments, betting odds, and team strategies.





Data Filtration



Records length BEFORE filtration

59288

Records

Records length AFTER filtration

28449

Records

Data Filtration

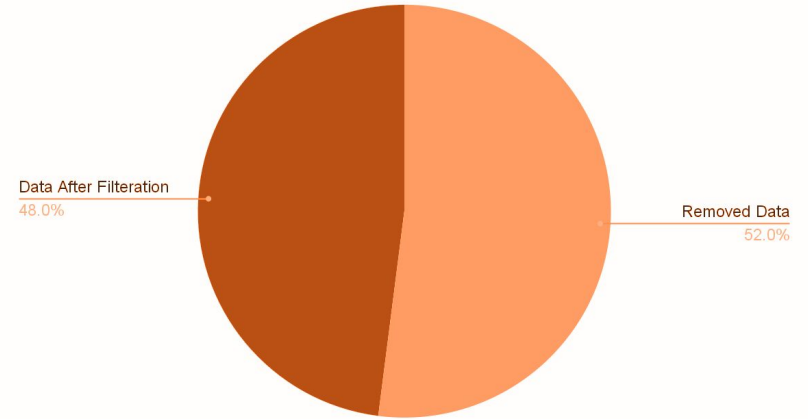


We used

48%

Of the original data

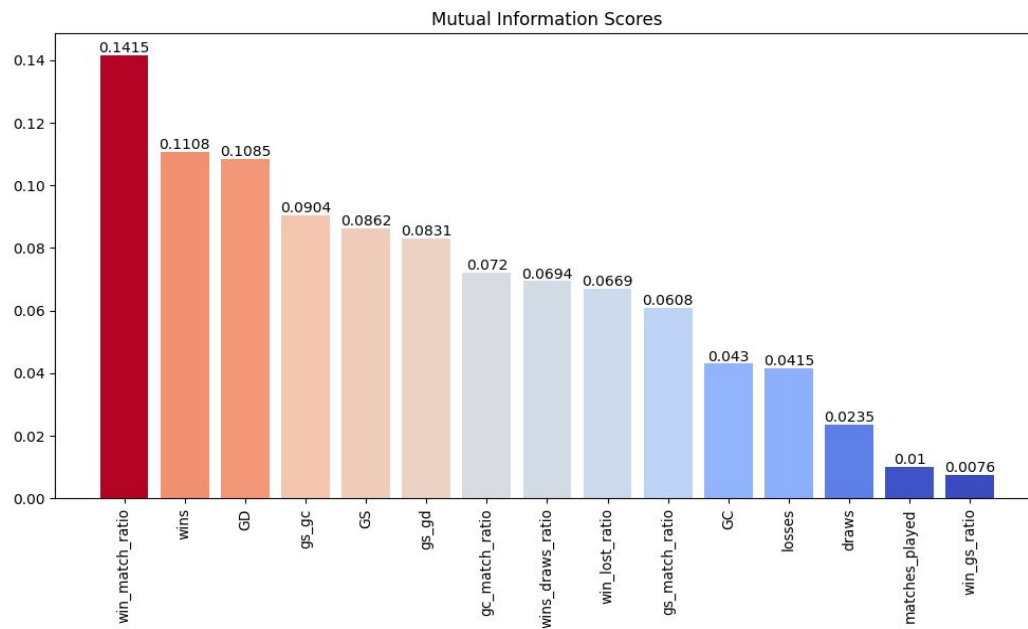
Data Filtration





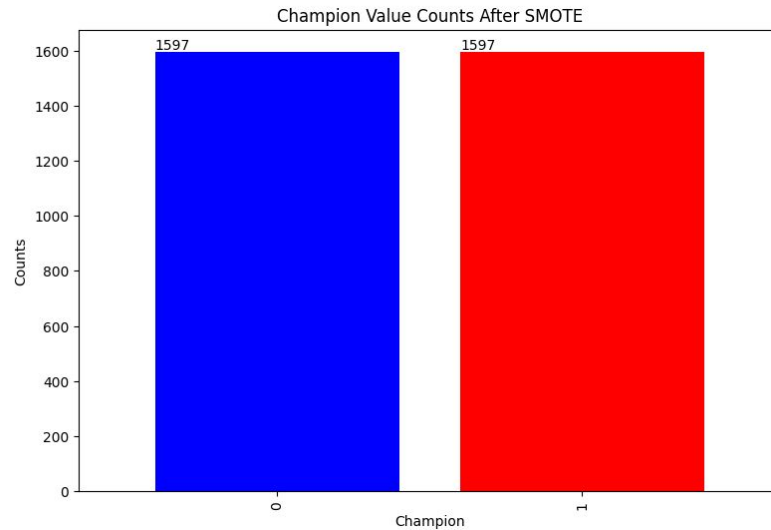
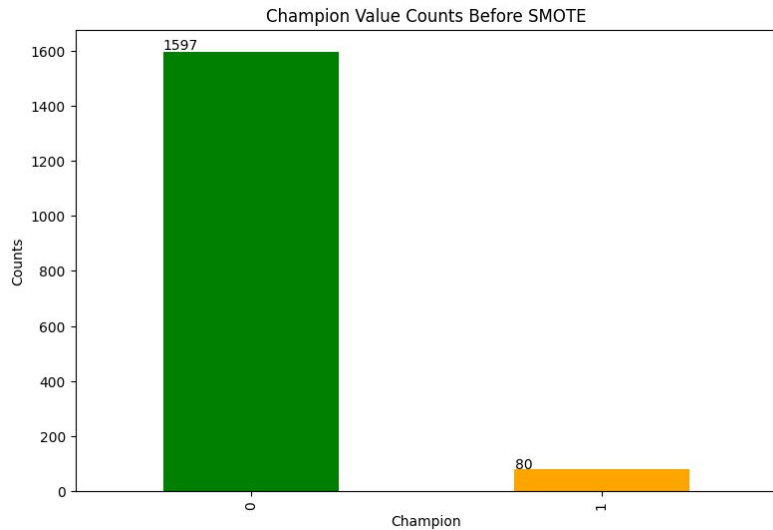
MI Scores

This setup is used to understand how each feature correlates or shares information with the status of being a champion.



SMOTE

SMOTE for balancing data



Models Score Analysis

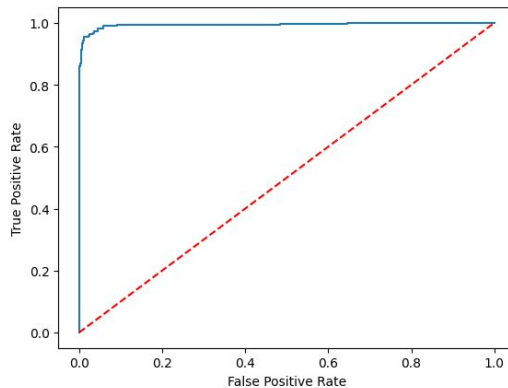
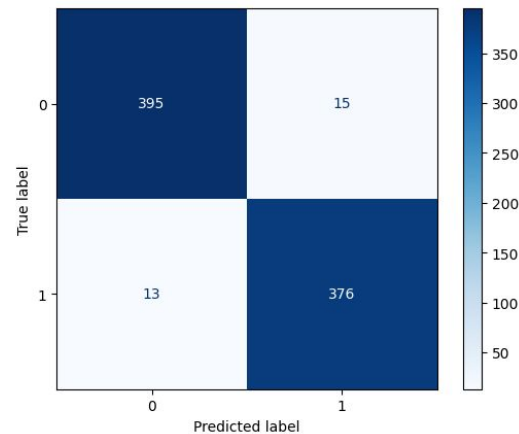
Model	Score
Logistic Regression	0.978712
KNN	0.962845
Random Forest	0.977035
Gradient Boost	0.974543
LightBGM	0.978712

—————



Logistic Regression

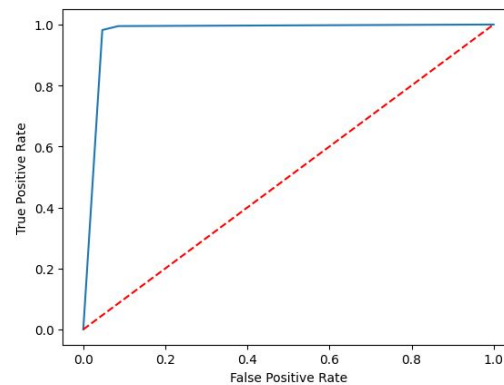
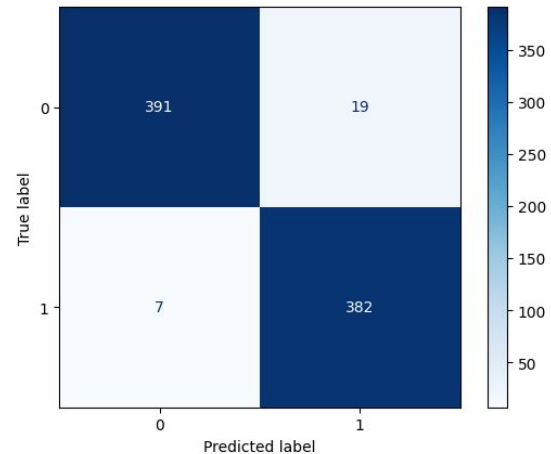
	precision	recall	f1-score	support
0	0.97	0.96	0.97	410
1	0.96	0.97	0.96	389
accuracy			0.96	799
macro avg	0.96	0.96	0.96	799
weighted avg	0.96	0.96	0.96	799





KNN

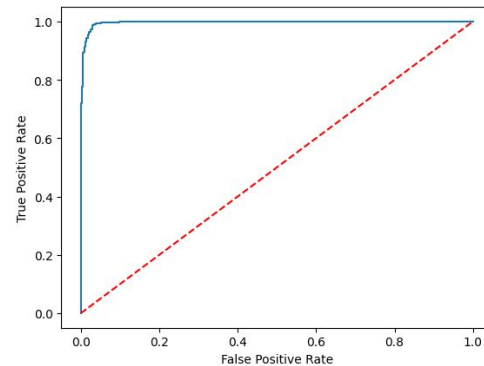
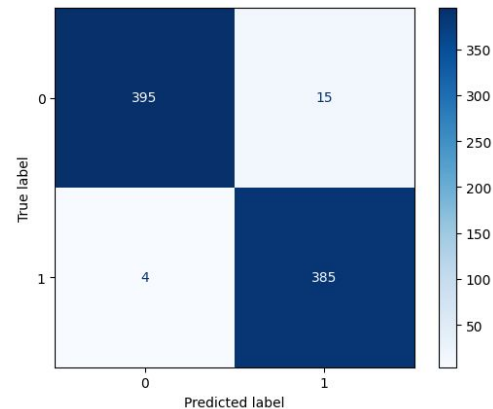
	precision	recall	f1-score	support
0	0.98	0.95	0.97	410
1	0.95	0.98	0.97	389
accuracy			0.97	799
macro avg	0.97	0.97	0.97	799
weighted avg	0.97	0.97	0.97	799





Random Forest

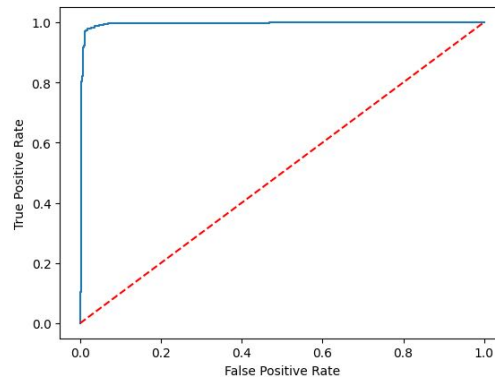
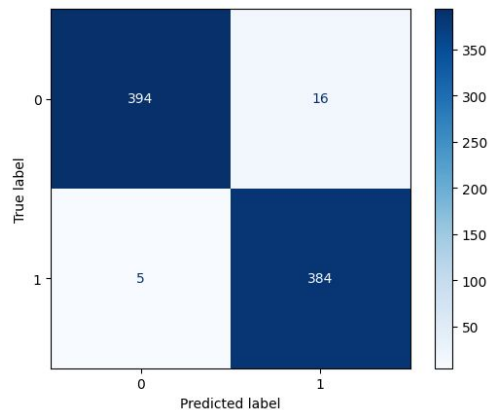
	precision	recall	f1-score	support
0	0.99	0.96	0.98	410
1	0.96	0.99	0.98	389
accuracy			0.98	799
macro avg	0.98	0.98	0.98	799
weighted avg	0.98	0.98	0.98	799





Gradient Boost

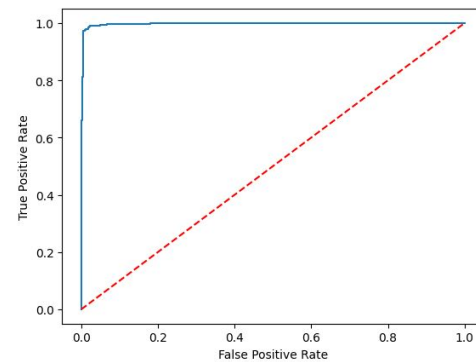
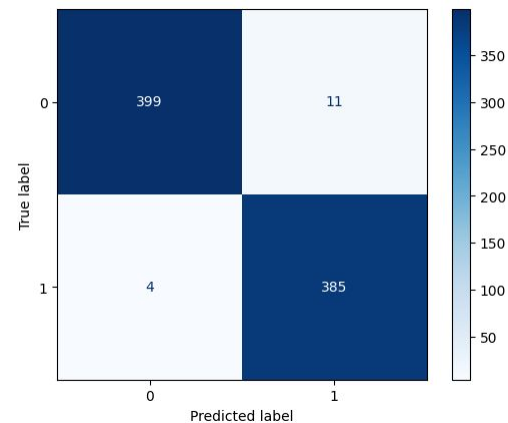
	precision	recall	f1-score	support
0	0.99	0.96	0.97	410
1	0.96	0.99	0.97	389
accuracy			0.97	799
macro avg	0.97	0.97	0.97	799
weighted avg	0.97	0.97	0.97	799





LightBGM

	precision	recall	f1-score	support
0	0.99	0.97	0.98	410
1	0.97	0.99	0.98	389
accuracy			0.98	799
macro avg	0.98	0.98	0.98	799
weighted avg	0.98	0.98	0.98	799





Thank you.