



OLYMPICS

Team 6

Start Slide





INDEX

- 1. INTRODUCTION**
- 2. PROBLEM STATEMENT**
- 3. OBJECTIVES / POTENTIAL USES**
- 4. STEP BY STEP**
- 5. RESULTS**
- 6. ETHICAL CONSIDERATIONS**
- 7. CONCLUSION**





INTRODUCTION

The Olympic Games are the world's only truly global, multi-sport, celebratory athletics competition. With more than 200 countries participating in over 400 events across the Summer and Winter Games, the Olympics are where the world comes to compete, feel inspired, and be together. Sports have the power to change lives, which is why we seek to create a model that predicts the medal count of the world's foremost sports competition.



Learn More >

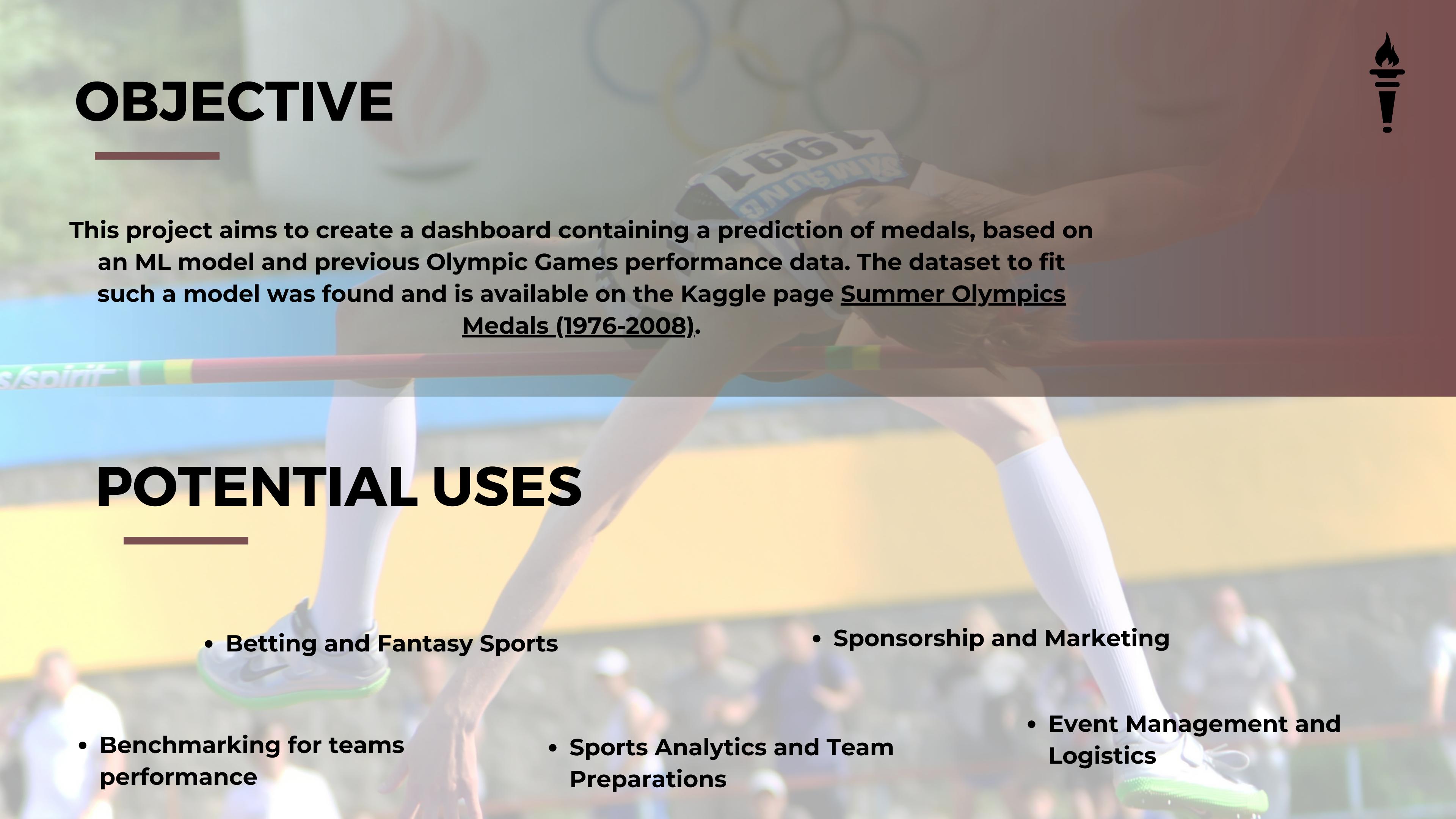


PROBLEM STATEMENT

The Olympic Games represent a critical pinnacle of global sports, where nations compete for prestige, athletes strive for excellence, and fans engage deeply with the event. Accurate predictions of medal counts are essential for nations and sports federations, media, logistics, etc.

Learn More >

OBJECTIVE



This project aims to create a dashboard containing a prediction of medals, based on an ML model and previous Olympic Games performance data. The dataset to fit such a model was found and is available on the Kaggle page [Summer Olympics Medals \(1976-2008\)](#).

POTENTIAL USES

- Betting and Fantasy Sports
- Sponsorship and Marketing
- Benchmarking for teams performance
- Sports Analytics and Team Preparations
- Event Management and Logistics



STEP BY STEP

DATA LOADING

- ENCODING USED: ISO-8859-1

CLEANING / PREPROCESSING

- DROPPED ~110 EMPTY LINES TO CLEAN THE DATASET.
- CONVERTED YEARS FROM FLOAT TO INTEGER FOR CONSISTENCY.



STEP BY STEP

CLEANING / PREPROCESSING

- DEFINED A SINGLE COUNTRY CODE FOR A COUNTRY LISTED WITH 2 DIFFERENT CODES, ENSURING DATA INTEGRITY.
- AGGREGATED MEDAL COUNTS ENABLE A SINGLE, COMPREHENSIVE OUTCOME FROM THE MODEL.

ML MODEL OPTIMIZATION

- ACHIEVED MARGINAL IMPROVEMENT IN ACCURACY / MEAN SQUARED ERROR, ENHANCING MODEL PERFORMANCE.

RESULTS

CONFUSION MATRIX

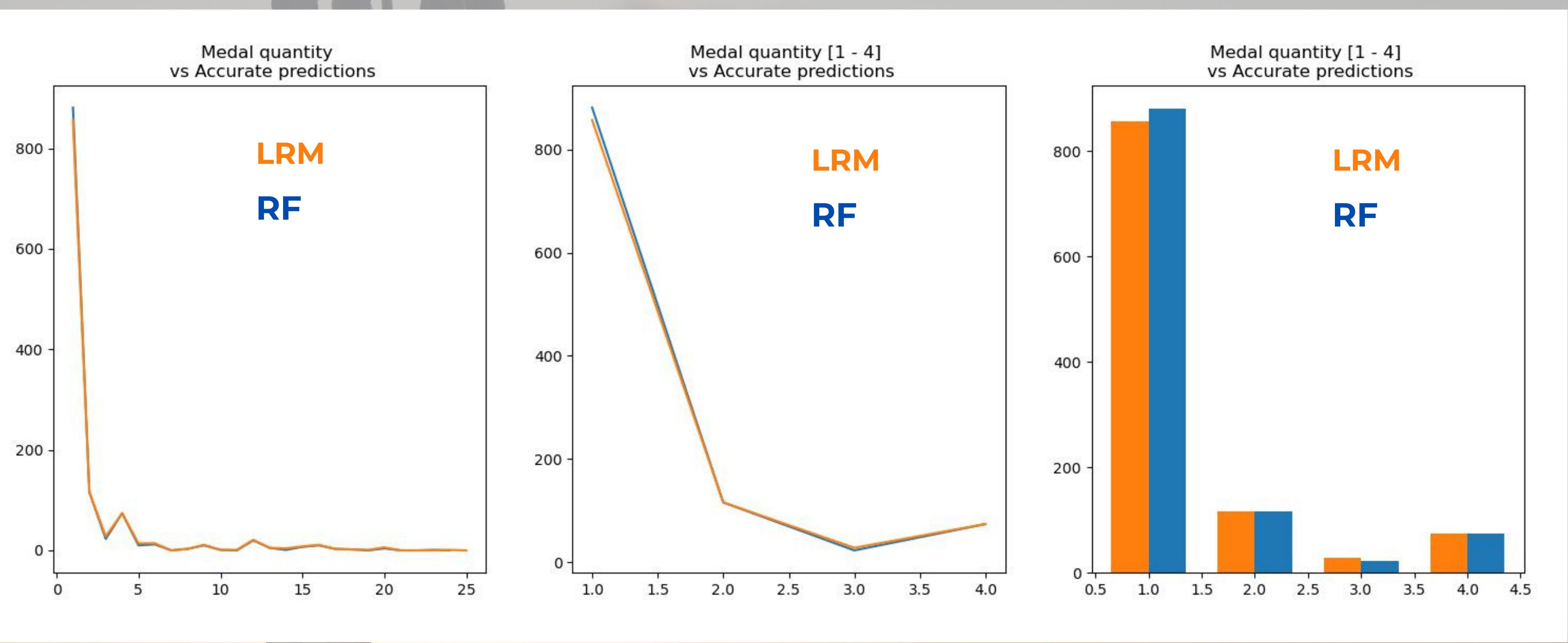


	Predicted 1	Predicted 2	Predicted 3	Predicted 4	Predicted 5	Predicted 6	Predicted 7	Predicted 8	Predicted 9	Predicted 10	Predicted 11	Predicted 12	Predicted 13	Predicted 14	Predicted 15	Predicted 16	Predicted 17	Predicted 18	Predicted 19	Predicted 20	Predicted 21	Predicted 22	Predicted 23	Predicted 24	Predicted 25	Predicted 26	Predicted 32	Predicted 36
Actual 1	903	11	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
Actual 2	117	104	6	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
Actual 3	19	15	22	2	4	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
Actual 4	3	13	14	59	8	5	5	1	0	1	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
Actual 5	0	0	0	0	7	13	2	0	0	2	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
Actual 6	0	0	0	0	3	5	6	1	1	0	2	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
Actual 7	0	0	0	0	0	0	0	1	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
Actual 8	0	0	0	1	5	1	0	3	1	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
Actual 9	0	0	0	0	0	1	0	0	0	6	2	2	1	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0
Actual 10	0	0	0	0	0	1	3	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
Actual 11	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
Actual 12	0	0	0	0	0	2	0	4	1	0	1	5	12	6	0	0	0	0	1	1	0	0	0	0	0	0	0	0
Actual 13	0	0	0	0	0	0	0	0	0	0	0	0	0	1	4	0	0	0	0	0	0	0	0	0	0	0	0	0
Actual 14	0	0	0	0	0	0	0	0	0	0	0	0	0	0	2	0	3	1	2	1	0	0	0	1	0	0	0	0
Actual 15	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	8	0	0	0	0	0	0	0	0	1	0	0
Actual 16	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	2	5	4	0	0	1	0	0	0	0	0
Actual 17	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	5	0	1	0	0	0	0	0	1	0
Actual 18	0	0	0	0	0	0	0	0	1	0	0	1	1	0	0	1	1	1	0	0	0	0	0	0	0	0	0	0
Actual 19	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
Actual 20	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0
Actual 21	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
Actual 22	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
Actual 23	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
Actual 24	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	5	0	0	0
Actual 25	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
Actual 26	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0
Actual 32	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0
Actual 36	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0

Accuracy Score: 0.7952861952861953

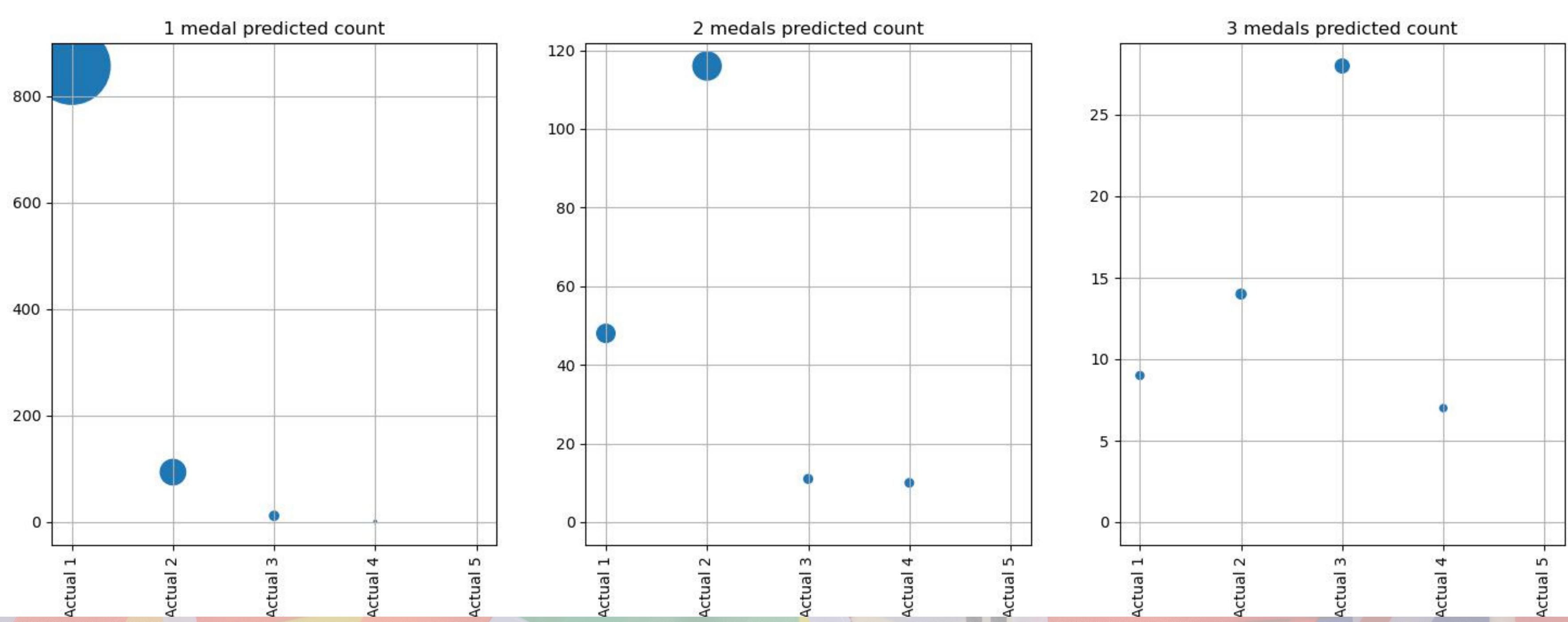
RESULTS

LRM VS RF



RESULTS

RF SCATTER PLOT



ETHICAL CONSIDERATIONS

- In the development of our predictive model, we used a dataset sourced from Kaggle's , ensuring full compliance with their terms and conditions.
- The dataset is free from biases related to individuals, focusing solely on publicly available information.
- No additional measures for the protection of personal data were necessary.





CONCLUSION

In our study to guess how many medals countries would win in the Olympics, we first tried Logistic Regression but then switched to a Random Forest. By changing the method, we saw a big jump in effectiveness in prediction, going from 46% to 79%. This suggests its efficacy in capturing the nuances of predicting Olympic outcomes. Looking ahead, we could make our predictions even better by using more detailed data and trying out more complex models to enhance prediction accuracy.





THANK YOU!