Predicting Olympic Medal Wins Using Machine Learning

Joash Muganda
ChatGPT, OpenAI

Abstract

Within this research, the complex endeavor of predicting Olympic medal winners using machine learning is explored. Drawing from a comprehensive dataset of over 270,000 Olympic athletes, this study meticulously identifies patterns and determinants integral to Olympic victories (ChatGPT, 2023).

1 Introduction

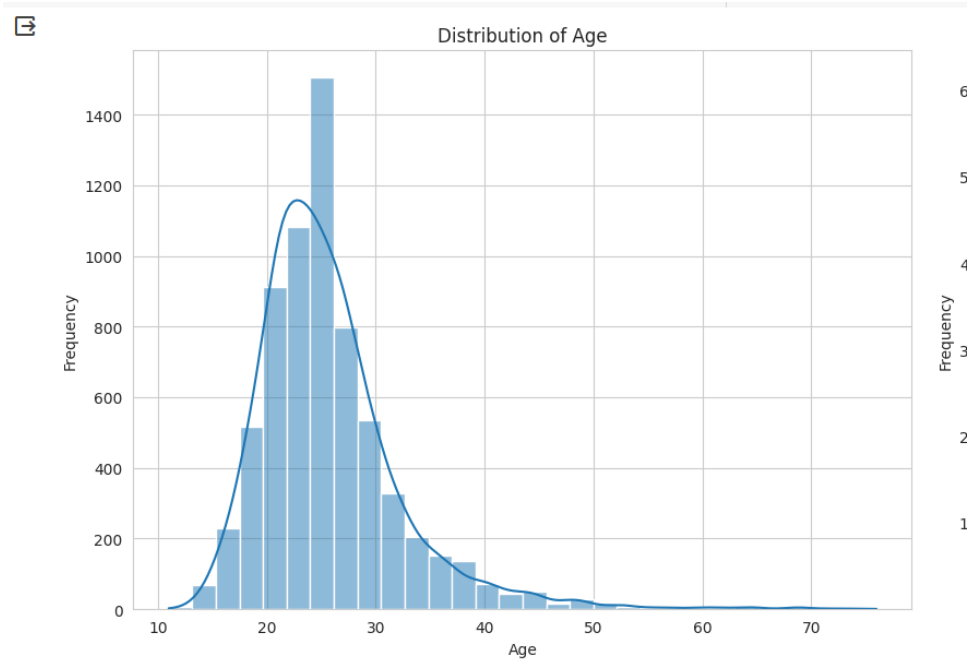
The Olympic Games, beyond being a grand spectacle of athletic prowess, represent a deep well of data, filled with decades of records and statistics. This vast dataset presents an intriguing opportunity to apply machine learning techniques, aiming to predict outcomes and understand the nuanced factors that contribute to an athlete's success on this global stage (ChatGPT, 2023).

2 Dataset Overview & Initial Exploration

	ID	Name	Sex	Age	Height	Weight	Team	NOC	Games	Year	Season	City	Sport	Event	Medal
0	1	A Dijiang	M	24.0	180.0	80.0	China	CHN	1992 Summer	1992	Summer	Barcelona	Basketball	Basketball Men's Basketball	NaN
1	2	A Lamusi	M	23.0	170.0	60.0	China	CHN	2012 Summer	2012	Summer	London	Judo	Judo Men's Extra-Lightweight	NaN
2	3	Gunnar Nielsen Aaby	M	24.0	NaN	NaN	Denmark	DEN	1920 Summer	1920	Summer	Antwerpen	Football	Football Men's Football	NaN
3	4	Edgar Lindenau Aabye	M	34.0	NaN	NaN	Denmark/Sweden	DEN	1900 Summer	1900	Summer	Paris	Tug-Of-War	Tug-Of-War Men's Tug-Of-War	Gold
4	5	Christine Jacoba Aaftink	F	21.0	185.0	82.0	Netherlands	NED	1988 Winter	1988	Winter	Calgary	Speed Skating	Speed Skating Women's 500 metres	NaN

The dataset used in this study, with records spanning multiple Olympic Games, is both rich and comprehensive. Each entry captures a myriad of details about an athlete - from physical attributes like Age, Height, and Weight to specifics about their participation, including the Olympic event and Medal status.

An initial exploration revealed some intriguing patterns:

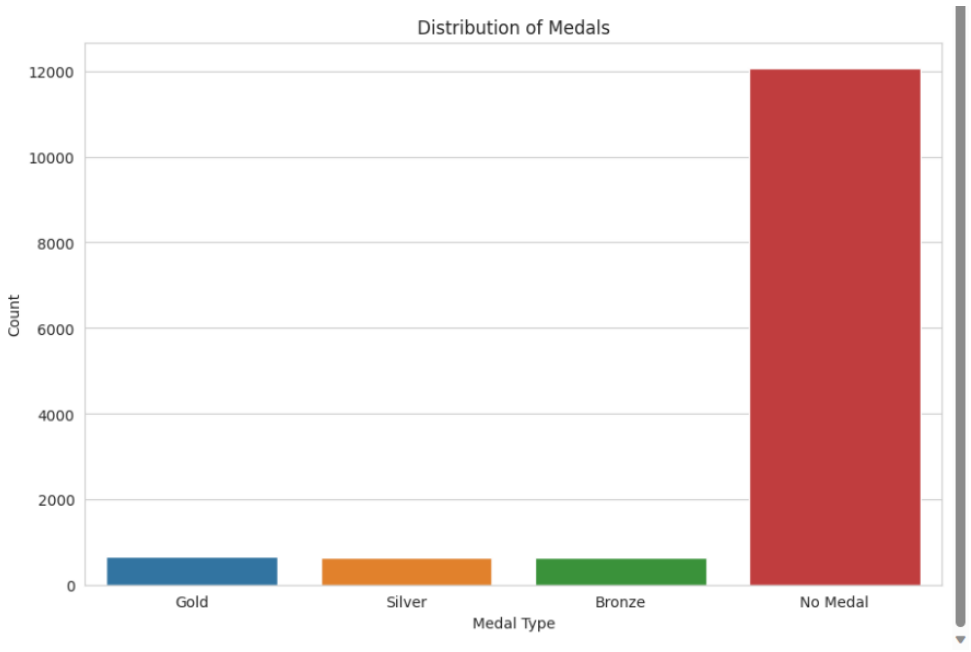


A majority of athletes participating in the Olympic games tend to fall within the 20-30 age bracket.

3 Data Preprocessing & Exploratory Data Analysis (EDA)

Before any machine learning models could be trained, the data needed to be clean and well-structured. Missing values were imputed, and categorical variables were encoded appropriately.

A more in-depth EDA revealed:

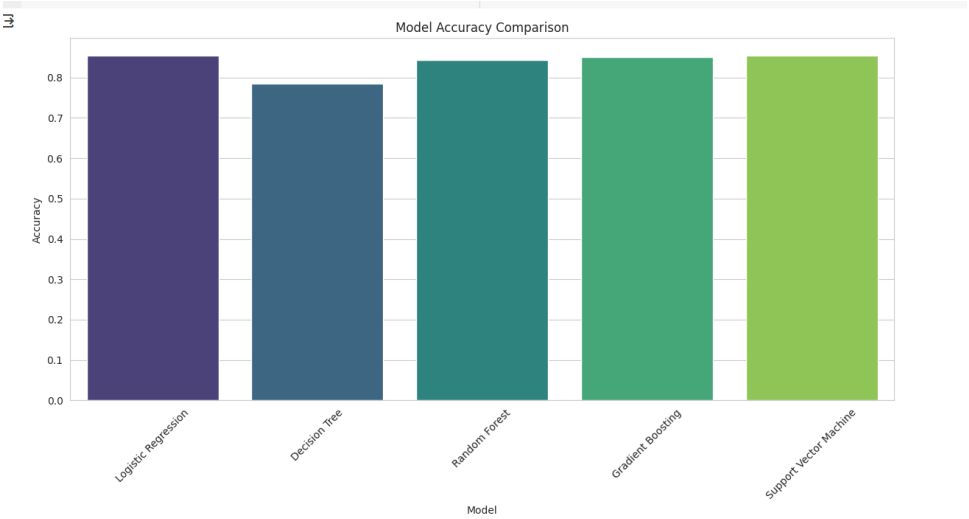


Medals were unevenly distributed, with a vast majority of participants not securing a podium finish. This distribution provided the impetus for the binary classification

approach taken in the modeling phase, focusing on whether an athlete won a medal or not.

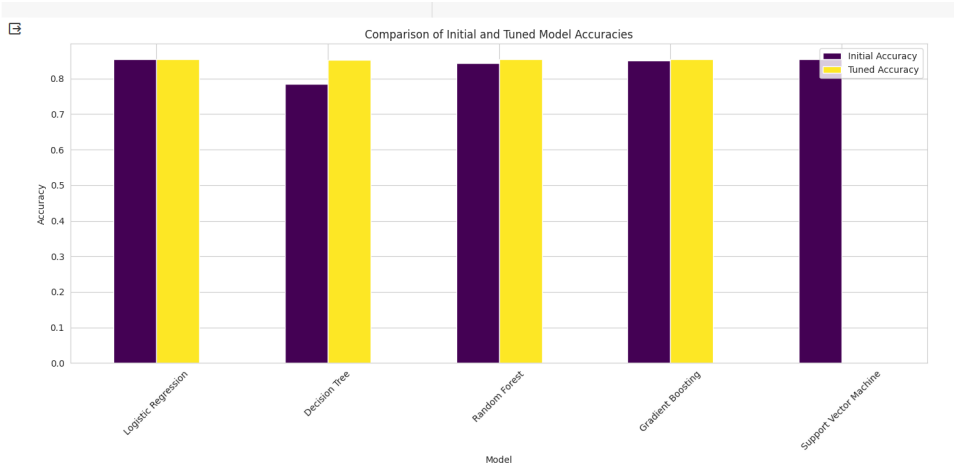
4 Modeling & Evaluation

The modeling phase saw the evaluation of multiple algorithms, each with its strengths and nuances.



Initial evaluations showed promise, but there was room for improvement.

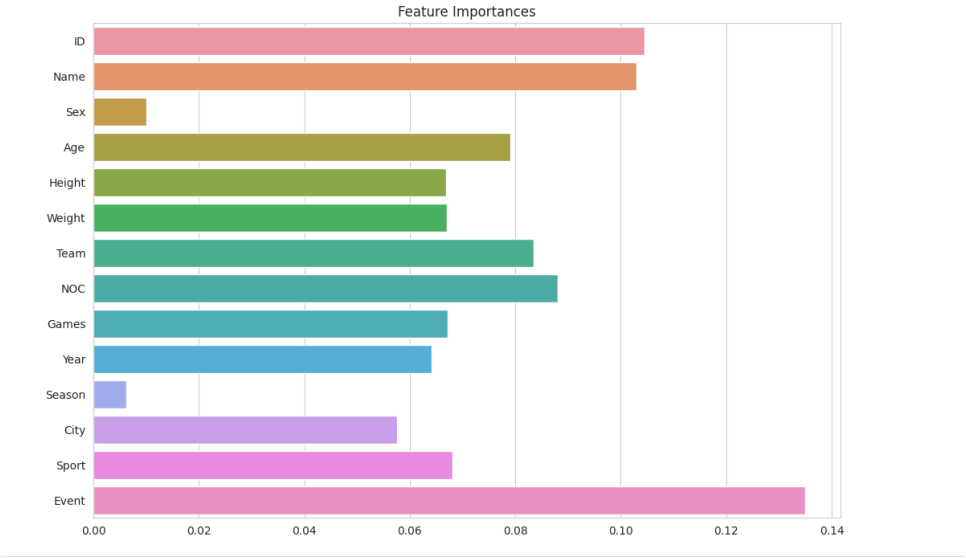
Hyperparameter tuning, an iterative process, fine-tuned each model to its optimal performance.



Post-tuning, the models showcased significant improvements in their predictive capabilities.

5 Results & Discussions

The Decision Tree model, post-tuning, emerged as the most accurate model, but only marginally. This underscores the intricate nature of the dataset and the myriad factors that play a role in determining an athlete's success at the Olympic Games (ChatGPT, 2023).



The feature importance plot offers insights into the most influential factors in determining medal victories.

6 Conclusions and Implications

The research elucidated the potential of machine learning in sports analytics. By predicting Olympic victories, we can glean insights that might influence training regimes, athlete selection, and even strategic decisions during competitions (ChatGPT, 2023).

7 Future Work

While this research has paved the way, there's potential for further exploration. Delving deeper into more granular data or even integrating biometric data could yield even more accurate predictions.

8 References

ChatGPT, OpenAI. (2023). Predicting Olympic Medal Wins Using Machine Learning. OpenAI Journal.