|  |  |  |
| --- | --- | --- |
| **Problem Chosen** C | **2025 MCM/ICM Summary Sheet** | **Team Control Number** 2522352 |

**Summary**

This study presents a comprehensive framework for predicting the 2028 Olympic medal tally and analyzing key factors influencing medal distributions. The methodology integrates data preprocessing, predictive modeling, and multi-dimensional evaluation to address the complexity of Olympic performance dynamics. First, historical data spanning past Olympic Games is preprocessed by unifying country codes (ISO 3-letter standards), handling missing values, calculating cumulative historical medals, and creating host country indicators. Athlete-level data is aggregated to national-team granularity for each edition of the Games.

For medal prediction, a hybrid approach is proposed: countries are classified into distinct clusters (e.g., stable performers, declining powers, emerging nations) using K-means and DBSCAN, enabling tailored modeling strategies. Time-series models (ARIMA/LSTM) are applied to stable teams, while emerging nations are analyzed through athlete potential evaluation using VIKOR, a multi-criteria decision-making method. A negative binomial regression model, incorporating features such as host nation advantage, sport-specific dominance, and historical trends, is developed to predict medal counts, addressing overdispersion in medal data. Key factors—host nation effect (validated via difference-in-differences analysis), sport specialization (quantified by national medal concentration in specific disciplines), and event expansion impacts—are rigorously evaluated through statistical significance tests and interaction models.

The study further investigates the "great coach effect" through case studies of nations experiencing performance shifts linked to coach transfers, supplemented by external datasets. Strategic recommendations for coaching investments prioritize nations with untapped potential and historical sensitivity to coaching expertise. Finally, the model reveals novel insights, including the nonlinear impact of event additions on medal redistribution and the diminishing marginal returns of host nation advantages. This work contributes a data-driven paradigm for Olympic strategy formulation, balancing predictive accuracy with interpretability of socio-sporting dynamics.

**Keywords**: Olympic medal prediction; negative binomial regression; host nation effect; sport specialization; coach transfer impact.

**Contents**

[1 Problem Restatement 3](#_Toc188790891)

[2 Data preprocessing 4](#_Toc188790892)

[3 Prediction for the 2028 Los Angeles Olympics 4](#_Toc188790893)

[4 Determinants of Medal Counts Across Nations 4](#_Toc188790894)

[5 Coaching Excellence in National Medal Architectures 4](#_Toc188790895)

# 1 Problem Restatement

The Olympic medal table not only reflects the competitive levels of various countries but also significantly influences the global perception of their athletic capabilities. During the 2024 Paris Summer Olympics, the United States ranked first in total medals with 126, while China and the United States tied for the top position in gold medals. Meanwhile, other countries such as Albania and Dominica earned their first-ever Olympic medals in history. However, over 60 nations have yet to win a single medal. In this context, our study aims to develop a mathematical model to predict the distribution of medals among countries in the 2028 Los Angeles Summer Olympics and analyze potential trends.

Specifically, this study seeks to accomplish the following tasks:

1. Develop a model capable of predicting the medal distribution for the 2028 Los Angeles Olympics, including an assessment of the uncertainty and accuracy of the predictions.
2. Identify countries that are likely to win their first medal in 2028 and estimate the probability of this outcome.
3. Investigate the "great coach effect" on medal counts and provide relevant recommendations.
4. Analyze the significance of various sports to different countries and assess the potential impact of the host country's choice of sports on the medal table.
5. Extract underlying patterns in Olympic medal distributions through historical data analysis to provide actionable insights for national Olympic committees in their strategic planning.

The research process will rely on the provided historical data, including Olympic medal tables, detailed sports data, and athlete performance data, while incorporating reasonable modeling assumptions and external contextual information. The model will account for various factors affecting medal counts, such as athlete performance, host country effects, and other socioeconomic conditions, ultimately providing a scientific basis for medal predictions and strategic decision-making.

# 2 Data preprocessing

# 3 Prediction for the 2028 Los Angeles Olympics

# 4 Determinants of Medal Counts Across Nations

# 5 Coaching Excellence in National Medal Architectures