Problem Chosen	2025 MCM Summary Sheet	Team Control Number
C		2510086

An MCM Paper Made by Team 2510086

Summary

Here is the abstract of your paper.

Firstly, that is ...

Secondly, that is ...

Finally, that is ...

Keywords: MATLAB, mathematics, LaTeX.

Contents

1 Introduction

1.1 Problem Background

During the most recent 2024 Paris Summer Olympics, aside from watching the competitions, spectators were also very interested in the medal tally of various countries. The nations at the top of the rankings always attract a lot of attention, and everyone hopes their own country can be among the leaders.

Over the past several decades of Olympic Games, many "sporting powerhouses" have emerged, but there are also numerous countries still striving to win their first Olympic medal in history. How will these medal-winning countries perform in the next Olympics? And do those countries without any medals have a chance to win one? These are all questions that people are concerned about.

Predictions of the final medal count are common, but they are typically not based on historical medal totals. Instead, they are often made just before the opening of the upcoming Olympics, when the athletes scheduled to compete are known[1]. However, when the information about the planned participants has not been disclosed, as is the case now with the 2028 Los Angeles Olympics where the competing athletes have yet to be announced, can we still predict the medal outcomes for various countries? This is the question that concerns us.

1.2 Literature Review

Past research has focused on exploring the application of machine learning techniques to predict medal outcomes and analyze medal distribution patterns for the 2024 Summer Olympics[2]. Leveraging a wide range of variables at both athlete and country levels, as well as event-specific metrics, various statistical models are employed to forecast medal counts and identify factors linked to Olympic success. However, such research methods cannot proceed without information on the participants of the next Olympic Games. This necessitates that we seek solutions within the limited dataset available to us.

1.3 Restatement of the Problem

Considering the background, in this paper we are required to solve the following problems:

- **Task 1:** Project the 2028 LA Olympics medal table using your model, including prediction intervals. Identify countries likely to improve or decline compared to 2024.
- **Task 2:** Predict how many countries will win their first medal in 2028 and estimate the odds of this outcome.
- **Task 3:** Analyze how event types and numbers affect medal counts. Identify key sports for different countries and the impact of host-selected events.

- **Task 4:** Investigate the "great coach" effect on medal counts. Suggest three countries and sports where hiring such a coach could have significant impact.
- **Task 5:** Share original insights from your model to guide national Olympic committees' decisions.

1.4 Our work

For ease of description and visualization, we have drawn a flow chart (Figure 1) to represent our work.

2 Assumptions

3 Notations

The primary notations used in this paper are listed in Table 1.

Table 1: Notations

Symbol	Definition
A	the first one
b	the second one
α	the last one

4 Task 1

4.1 Data cleaning

The detail can be described by equation (1):

$$\frac{\partial u}{\partial t} - a^2 \left(\frac{\partial^2 u}{\partial x^2} + \frac{\partial^2 u}{\partial y^2} + \frac{\partial^2 u}{\partial z^2} \right) = f(x, y, z, t) \tag{1}$$

5 Task2: Predicting the Probability of Countries Winning Their First Medal

In this section, we identify patterns in countries transitioning from no medals to their first Olympic medal, predicting how many countries will win their first medal in the upcoming Games and their probabilities. We preprocessed the data, selected three key fea-

tures, and trained an XGBoost model optimized with Bayesian tuning. Finally, we apply the model to the 2028 Olympic data to generate predictions.

5.1 Data Pre-processing

In this task, we focus on analyzing the transition of countries from having no Olympic medals to winning their first medal. Our goal is to uncover patterns in this process. We begin by processing the data from the *summerOly_athletes.csv* file. The steps involved in data pre-processing are as follows:

- **Data Aggregation:** We aggregate the data for each country across different years. For each year, we compute the total number of medals won by the country's team, retaining records where the country has either not won any medals or won their first medal.
- **Handling Team Sports:** For team events, we consolidate the medals into one entry, treating them as if the country won a single medal in the respective event.
- **Handling Anomalies:** We identify and remove anomalous data, such as cases where countries like Russia performed exceptionally well in their first Olympic Games due to historical factors. We also exclude data from the refugee team, which started competing in 2016.
- **Distinguishing First Medal Achievements:** Investigate the "great coach" effect on medal counts. Suggest three countries and sports where hiring such a coach could have significant impact.

  Since countries' first medal achievements vary significantly — for instance, San Marino (SMR) won 2 silvers and 2 bronzes in their first appearance — while many countries win only a bronze, we aim to differentiate these cases. To address this, we apply the Fibonacci Weighted Point System proposed by Sergeyev to quantify performance[3].

  The performance score for a country's team in the i^{th} Olympic Games is calculated as follows:

$$PE_i = 3g_i + 2s_i + b_i$$

  where PE_i is the performance score for the i^{th} country's team, calculated based on the number of gold (g_i), silver (s_i), and bronze (b_i) medals won.

- **Data Representation:** Finally, we organize the data into a vector representation for each country's participation in the Olympics. The vector Xi for the i^{th} country is defined as:

$$X_i = (N_i, P_i, E_i, F_i, PE_i)$$

  where N_i is the number of Olympic Games the country has participated in, P_i is the number of athletes sent, E_i is the number of events participated in, F_i is a binary variable indicating whether the country won a medal (1) or not (0), and PE_i is the performance score for the country's team in that specific Olympic Games.

In total, we process and organize 1,244 vectors that represent the various features of each country's participation and medal achievement across different Olympic Games.

5.2 Bayesian-Optimized XGBoost Medal Predictor (BO-XGMP)

The instance of long and wide tables are shown in Table 2.

Table 2: Basic Information about Three Main Continents (scratched from Wikipedia)

Continent	Description	Information
Africa	Africa Continent is surrounded by the Mediterranean Sea to the north, the Isthmus of Suez and the Red Sea to the northeast, the Indian Ocean to the southeast and the Atlantic Ocean to the west.	At about 30.3 million km^2 including adjacent islands, it covers 6% of Earth's total surface area and 20% of its land area. With 1.3 billion people as of 2018, it accounts for about 16% of the world's human population.
Asia	Asia is Earth's largest and most populous continent which located primarily in the Eastern and Northern Hemispheres. It shares the continental landmass of Eurasia with the continent of Europe and the continental landmass of Afro-Eurasia with both Europe and Africa.	Asia covers an area of 44,579,000 square kilometres, about 30% of Earth's total land area and 8.7% of the Earth's total surface area. Its 4.5 billion people (as of June 2019) constitute roughly 60% of the world's population.
Europe	Europe is a continent located entirely in the Northern Hemisphere and mostly in the Eastern Hemisphere. It comprises the westernmost part of Eurasia and is bordered by the Arctic Ocean to the north, the Atlantic Ocean to the west, the Mediterranean Sea to the south, and Asia to the east.	Europe covers about 10,180,000 km^2, or 2% of the Earth's surface (6.8% of land area), making it the second smallest continent. Europe had a total population of about 741 million (about 11% of the world population) as of 2018.

Figure **??** gives an example of subfigures. Figure **??** is on the left, and Figure **??** is on the right.

5.3 Future Prediction

6 Sensitive Analysis

7 Strengths and Weaknesses

7.1 Strengths

- First one...
- Second one ...

7.2 Weaknesses

- Only one ...

Memorandum

To: Heishan Yan
From: Team 1234567
Date: October 1st, 2019
Subject: A better choice than MS Word: LaTeX

In the memo, we want to introduce you an alternate typesetting program to the prevailing MS Word: **LaTeX**. In fact, the history of LaTeX is even longer than that of MS Word. In 1970s, the famous computer scientist Donald Knuth first came out with a typesetting program, which named TeX . . .

Firstly, . . .

Secondly, . . .

Lastly, . . .

According to all those mentioned above, it is really worth to have a try on LaTeX!

References

[1] Nielsen. Nielsen's Gracenote Expects USA, China, Great Britain, France and Australia to Lead 2024 Paris Olympic Games Medal Table. https://www.nielsen.com/news-center/2024/virtual-medal-table-forecast/.

[2] Moolchandani, Jhankar, et al. "Predictive Analytics in Sports: Using Machine Learning to Forecast Outcomes and Medal Tally Trends at the 2024 Summer Olympics." *2024 4th International Conference on Technological Advancements in Computational Sciences (ICTACS)*. 2024. https://ieeexplore.ieee.org/document/10840553.

[3] Sergeyev, Yaroslav D. "The Olympic Medals Ranks, lexicographic ordering and numerical infinities." *arXiv preprint arXiv:1509.04313*, 11 Sep. 2015. Available: https://arxiv.org/abs/1509.04313.

Appendix A: Further on LaTeX

To clarify the importance of using LaTeX in MCM or ICM, several points need to be covered, which are ...

To be more specific, ...

All in all, ...

Anyway, nobody **really** needs such appendix ...

Appendix B: Program Codes

Here are the program codes we used in our research.

test.py

```python
# Python code example
for i in range(10):
    print('Hello, world!')
```

test.m

```
% MATLAB code example
for i = 1:10
    disp("hello, world!");
end
```

test.cpp

```
// C++ code example
#include <iostream>
using namespace std;

int main() {
    for (int i = 0; i < 10; i++)
        cout << "hello, world" << endl;
    return 0;
}
```