Final Project Proposal

Team 23

Dataset

https://www.kaggle.com/heesoo37/120-years-of-olympic-history-athletes-and-results

Topic

In this project, we are going to use the dataset of historical Olympic athletes and results to explore how physical characteristics of elite athletes have been changing over time. Gold medal winners do not look the same in the history of 120 years of modern Olympic Games. In the past, short and slim runners were advantageous in sprints, but Usain Bolt, who is 6ft 5 in, dominated sprints in the past decade. Pete Sampras retired at the age of 31 in 2002, while 37-year-old Roger Federer, 32-year-old Rafael Nadal, and 31-year-old Novak Djokovic swept all 8 grand slam titles in past 2 years and these “veterans” are still ruling men’s professional tennis. Consequently, it is highly possible that developing athletic training and sports medicine have increased athletes’ career span and reshaped players’ figure. By statistics we can obtain a model that consists of several features, such as age, gender, height, weight, nationality, and sports, for more successful players in each year’s Olympic Game. The trend of age and career span could indicate that aged athletes are increasingly more capable of maintaining their performance. It would emphasize the significance of sports medical technology nowadays. The trend of body shape could help guide athlete recruitment and athletic training.

What is currently known about this topic

There are lots of factors may determine the probability of winning a gold medal in an Olympic game. In the dataset, we explore the influence of athletes’ features, such as sex, age, height, weight and their nationality; and the influence of where and when the games were held and which event they attended as well. We might have prior knowledge such as there is consistency between the athletes’ performance in recent games; athletes with certain height and weight will be more suitable for some sports and hence can be easier to win.

But some studies also show surprising results. More aging athletes participate and also win in several sports area in recent years [1]. According to the conclusion of the authors, in certain sports areas, athletes can maintain high competence after receiving scientific training. And since aging athletes have more experience, age can also be an advantage. So it is clear that we cannot determine which factor is less important and which factor has definite positive or negative effects before careful analyze.

Another related work we need to notice is that some researchers use athletes’ nationality to predict their performance [2]. The author use the data of total gross domestic product, total population, latitude and overall economic freedom. The reason why this study has good result lays in the fact that athletes may get more scientific training in a more wealthy country. It reminds us to not only pay attention to the features of an individual athlete but also the environment where he get trained.

[1] ’Physiological Characteristics of an Aging Olympic Athlete’ doi: 10.1249 / MSS. 000 000 000 000 0331

[2] ‘Predicting the London Olympics Medal Count and the Why’s Behind It’
http://www.discoverycorpsinc.com/predicting-the-olympic-medal-c/

What is the anticipated impact of this work

Through this work, we will be able to find characteristics that are correlated to winning gold medals in the Olympics. We will examine the distributions and correlations of the features to demonstrate an overview of athletic characteristics. We will analyze the different trends of particular features such as gender, country, and age that lead to success and find out which is more influential. On the other hand, we will find factors that are negatively correlated to understand the characteristics of underperforming athletes. More importantly, these characteristics can potentially help coaches train their athletes to maximize their chance of winning first place. Moreover, by learning the trends, we will be able to predict what elite athletes are going to be like in the future in order to develop a scientific methodology for drafting young players.

Timeline

- Week 6
    - Eric: Clean data using pandas, remove outliers
    - Zifan: Find more useful datasets
    - David: Accumulate datasets into one
- Week 7
    - Eric: Plot data using matplotlib, p-testing
    - Zifan: Analyze results
    - David: Find positive correlation between different features
- Week 8
    - Eric: Plot data using different graphs (bar, line, points)
    - Zifan: Use PCA to reduce dimensionality and decorrelate variables
    - David: Find negative correlations between different features
- Week 9
    - Eric: Analyze graphs and draw conclusions
    - Zifan: Clean up code (add missing docstrings, etc)
    - David: Give possible causes/inferences
- Week 10
    - Eric: Create ⅓ slides of presentation
    - Zifan: Create ⅓ slides of presentation
    - David: Create ⅓ slides of presentation

Github

https://github.com/xericho/ECE143_Team23

Presentation Slides

https://docs.google.com/presentation/d/1Ve-YbZByuTAe7vDpvm3m_CrrO3GLWL7otZkxXzTIg20/edit?usp=sharing