Discovering the Home-field Advantage for France in Forthcoming 2024 Olympic Game.

Shruti Sindhi
Department of Big Data Management and Analytics
Griffith College Dublin
Dublin, Ireland
shrutivijay.sindhi@student.griffith.ie

Amritha Puravelil Manoharan
Department of Big Data Management and Analytics
Griffith College Dublin
Dublin, Ireland
amritha.puravelilmanoharan@student.griffith.ie

***Abstract*—The athletes' success in any sport is considered a source of national prestige, and it hinges on various social and economic factors. It is noticed that all the countries do not have the same capability to win a medal. Also, the most ardent xenophobes comprehend that a single country cannot win all the laurels at a given Olympic Games. Furthermore, there is a limit to authorized entries in an event. Despite all these, there are some edges while participating in front of the home crowd. This study focuses on delivering results showing that participating in front of a home crowd will spur athletic performance. Moreover, it will assist in uncovering the factors influencing and impacting the overall Olympic success of the country. This research paper will report whether home track aids France in obtaining a top rank in the upcoming Paris 2024 Olympic Games. Documented data shows that France's performance has been consistent in past events but has only grabbed the first position on the medal scoreboard once in the 1900 Paris Summer Olympic Games. France has been in the top 3 spots only four times, thrice it was the third position. There are already numerous models implemented in the past to guess the winning country of the Olympic tournament in general. The FP- growth (Frequent pattern growth) algorithm has been utilized to find an association between features of the data in the past. Linear Regression, GBR, Polynomial Regression, ANN, SVM (Support Machine) model, and Random Forest Decision Tree are used on sports data to foresee the winning countries. To significantly find and improve the Olympic medal signifying accuracy, the MLP model utilizing pruning is incorporated for predicting the desired results. The adaption of the MLP model is to enhance the forecasting performance and inspect if the results match the prediction.**

***Index Terms*—Olympic Games, Home field advantage, Multi-layer perceptron model, Data Mining, forecasting winners, Artificial Neural Network, Summer Olympics, and Winter Olympics.**

I. Introduction

There are numerous factors affecting the Olympic medal panel or, to be precise, the overall rank of the countries on the scoreboard. The ability of the countries to perform and excel in different sports in the prestigious event hinges on various social and economic factors. There are considerable ways to measure Olympic success. One prevalent way to measure success is to use the ratio of medals won to the total number of trophies awarded during the Olympics as their dependent variable [6].The hosting gain is estimated to be 1.2 percentage point of the total medal share.There are a [8] few advantages for the players to having sports events hosted on the home ground. And the existence of this home advantage in major team sports has been well established [30].However, antisocial crowd behaviour (swearing, chanting obscenities) in a home game can make home teams commit more infractions [29]. Moreover, the visiting players or teams can generate a dysfunctional aggressive response to the hostile atmosphere of an away game [18]. The crowd size and noise also slightly influence the players' performance. The boost in athlete/team participation and fan presence has not been consistent. It was because of quagmire and dangerous situations like displaced citizens in the 1968 Mexico City Summer Games and terrorist attacks during the 1972 Munich Summer Games, respectively [24].Over the last few decades, the exigency of accurately forecasting Olympic performances in sports and economic literature has procured substantial research interest [13]. There are publications that make in efforts to predict the total national medal count ahead of the start of the Olympic Games. The result of the indagate will be beneficial for the participating nations of the world to enhance their performance. Furthermore, it will aid the bidding agencies in determining which nation they want to sponsor. It can also use to track and judge sports achievement's development law and characteristics [40].

II. Related Work

While reviewing the related work in sports and Olympic games specifically, it was learned that there was not enough literature until the 1990s, that discusses about the models executed to recognize Olympic performance or enhance it. The performance analysis could have been impacted as Cold War played havoc with the 1970s and 1980s Olympic Games. A performance analysis restarted in 1993 by Slughart et al, that focuses on the Olympic performance, athletes efforts in transactional economies [36]. Zhang Yuhua executed linear regression to forecast and analyze the result of the 31st Rio Olympic Games [45]. The relationship between attributes in sports science is not linear, as a unit change in an independent variable will not always bring about an equivalent change in the dependent variable [46]. It was discovered that three regression models, GBR (Gradient

Boosting Machine), Polynomial Regression, and Random Forest, adaptable to different regressions, are used to predict the winners of Olympic events [11]. The random forest model is preferred as the best option among the three regression models because it predicts the test set with RMSE (Root Mean Square Error) value of 5.604 [22]. Tobit Regression is considered as the reference model for predicting the Olympic medal table [39]. The regression equation works for large number of countries with no medals in the games [15]. A paper by Lins applied the classic DEA model as a starting point to discover the Olympic ranking of the countries in the game [25]. The decision-making units incorporated in this model are those nations that won at the minimum one medal in Sidney 2000 Olympic Games. Another researcher Lozano also examined the performance in the Olympic games with the aid of Data Envelopment Analysis (DEA) [26]. The study investigated the implementation by monitoring and managing gross national product and population size.

The research did not formally consider home track benefits, but it was conceded as a potential influence [28], unaffected by the authors' main conclusion. Considerable research was conducted to comprehend the performance of the host countries through literature review and mathematical and logical reasoning. It was discovered that the home nations were shown to win around three times more medals than others [4].

And interestingly, about two times more in the Olympic Games on either side of their home game [4]. More detailed research accentuated that 'superior' teams exhibit additional significant home-field advantage than others in team game format [33]. At the same time, it was seen that there was a 61% advantage of winning at the home track, which was more significant than the number of individual sports 54% won away [16]. A further analysis portrayed partial consistency in foretelling the home track benefits in women's sports [17]. However, some other authors in the sports domain have made qualms concerning quality classification [27]. Once the quality of the athlete has been accounted for, home track advantage was not a significant influence on the performance of the individual game [31].Several authors have also discussed how the data is elucidated in the context of the "hot hand" belief, which states that the team or individual who wins the game has greater chances in further attempts [38]. It concludes that the Olympic tournament data are not independent, but instead, they are positively autocorrelated.

Study by researcher Edward Condon, Edward Wasil and Bruce Golden demonstrates that it instigated [11]. Neural Network to make prediction. One issue tackled by performing pruning on ANN back prorogation algorithm is pruning [11]. When an artificial neural network memorizes input training patterns and cannot work with new inputs, it ensues. Pruning is performed on the algorithm because there is a definite number of neurons that are zero. Such zero activation neurons can be withdrawn without affecting the overall accuracy of the Artificial Neural Network. It was also caught that several researchers proceeded in a similar path, trying to add new variables to the model but not obtaining significant results [39]. Finally, it was also learned that with the assistance of Apriori-Feed Forward (AFF), both pattern mining and prediction on large sports datasets can be made by scanning the dataset only once [7].

III. Description of the Dataset

A descriptive analysis of variables used in the model is mentioned in table 1. The data was presented by the IOC Research and Reference Service and available on The Guardian's Datablog. The model is fed and trained by using an overall 206 countries dataset. The number of nations involved in each Olympic game throughout the years is not the same. Officially, the Olympic Committee has never brought out an authorized ranking of the participating nations [25]. The statistical data analysis included 14 countries in the first game and 206 in the last. For analysis, the data is leveraged from 1896 to 2014 instead of one year prior to the event because there is no prior knowledge of how relative weight should be assigned to old and new data. As a result, there is a gambler's fallacy in the outcomes; even though it is independent of previous events, it still depends on the formerly occurred events [38]. Both Summer and Winter Olympics datasets are used in foretelling the results. In addition, both female and male data is analysed to initiate an unbiased home track advantage in the game, even though the females generally did not compete until 1928. The raw dataset comprises 26690 men's and 10245 women's records. 48 different sports categories (inclusive of both individual and team games) are incorporated to examine the dataset. The 48 sports categories can be further sub-categorized into 80 different disciplines. Each sports category does not retain a discipline. The dataset holds game records for each Olympic event in four-year increments, excluding 1940 and 1944 due to World War II. The Olympic game winner dataset has an exception of few records representing more than one hosting nation for the whole event. The records are from 1956 Olympic Games. It is also known as Games of the XVI Olympiad or Melbourne 1956. The two hosting countries were Stockholm and Melbourne. The reason for this anomaly is due to the Australian quarantine regulations by the Australian Quarantine and Inspection Service. The disciplines conducted in Sweden were dressage, eventing, and show jumping.

A. *Original Dataset*

The primary dataset includes information on participants from 1896 to the 2014 Summer and Winter Olympic Games. The winter dataset is recorded from 1924. The original dataset has 9 features, and one new feature called hosting country is added. It was mapped using the city name. The dataset constitutes 7 nominal attributes, one ordinal attribute, and one interval attribute. The year attribute is considered an interval attribute rather than ordinal, nominal data because an absolute 0 value is not expected in this scenario. There are 4 records with no evidence of which country the winner represents for the sport.

B. External Dataset

The primary dataset encloses meaningful features related to athletes and events, but no socio-economic features are present. Hence an external dataset with the name dictionary comprising a list of participating countries, their code, population, and GDP. The population and GDP per capita attributes possess numeric values. The table consists of 2 nominal and 2 numeric attributes. Even on viewing the data and its metadata information, it was not easy to trace back the year or decade the GDP and population of the nation were recorded. The dataset retains 201 unique observations.

IV. Data Preprocessing

Data pre-procesing is a essential step to ensure accurate forecast [41].Data accumulated from the real world might comprise of missing/misplaced values and/or noises. Generally, data cleaning and scrubbing is done to achieve the required end result. Here, the data preprocessing task is split into missing value processing, deduplication of the records, and data integration.

A. Missing Value Processing

Missing data points in a specific year was noticed by inter- / extrapolation, which is a common approach when pre-processing data [10] [9]. The research aims to discover if the home-field advantage will aid France in obtaining a top rank in the upcoming 2024 Olympics. Therefore, the participant's medal information is essential and cannot be bypassed at the analysis time. On preprocessing, it was found that there was no record with NA or displaying a 'Do Not Win' tag in the medal column. There are 4 rows with the winner's country detail missing from the 2012 London Olympic Games. The sport the individual in the missing record belongs to are Athletics, Weightlifting, and Wrestling. Out of 4 records with missing values, 2 records are of men, and the remaining 2 are of women. The percentage of missing data is 0.00010%; thus, the records can be deleted, and it will not affect the overall analysis. Also, additional, more detailed research revealed that individuals from the missing rows were from Olympic non-hosting countries. The newly created dataset consist of 26688 men and 10243 women.

The dictionary dataset does not involve 5 nations' records out of the 206 nations. The missing countries are the Republic of Moldova, Tvala, Federated States of Micronesia, Marshall Islands, and Independent Olympic Athletes. It was scanned employing Python. Mixed Team entry is not considered for the research. A new entry called Refugee Olympic Team (ROT) is overlooked from the dataset because both Mixed Team and ROT do not have a defined GDP and Population. Refugee Olympic Team came into existence in 2016.The code to represent Mixed Team is ZZX. It is assigned by International Olympic Committee (IOC).The country Czech Republic have also represented the Olympic Games in the past with the name Bohemia and Czechoslovakia. Therefore some records have country code TCH or BOH instead of CZE.It was not misassignment of the values in the records. Change in the code was of dissolution issues of Czechoslovakia. Hence such records are not removed from the dataset. Rather changes are made to the table before training it for prediction. Rows with country code as RU1 (Roman Empire) , URS (Soviet Union), and EUN (Unified Teams) were discarded from the dataset. The Russian Empire stretched across northern portions of Asia and Europe. It occupies present Russia, Ukraine, Finland and many more countries. So replacing Country Code, RU1 (Russian Empire) to RUS (Russia) would have affected the overall outcome. Similarly, the URS (Soviet Union Socialist Republics) the communist state that stretched across Eurasia from 1922 to 1991 was made up of 15 territories that are at present different countries engaging in the Olympic Games. Players from Barbados, Jamaica, and Trinidad and Tobago vied as the British West Indies (BWI) at the 1960. And so, records with code BWI are eradicated. List of other rows white out from the dataset encompass are (TPE) Taiwan, (VEN) Venezuela and a few more as their population and GDP per capita value could not be found.

A close examination of the external dataset exhibited 25 countries and 5 countries whose GDP value and Population were missing in dictionary_data_f. In total, there are 25 entries with missing Population or GDP per capita or both. These 25 records are drawn from the dataset before the projection. Discarding those records may not affect the overall analysis as it did not comprise hosting country medal winning details.

B. Deduplication of the records

A new dataset is created and named deduplicated_olympic, which comprises data focusing on sport and discipline rather than the individual post deduplication of the data. As multiple individuals can participate in a group game, each of them will be awarded medals, resulting in inaccuracy in medal statistics. Therefore, for analysis purposes, the team sports records were grouped. In addition, a temporary new feature parameter called "total_athletes" was added to deal with duplicate (from a prediction goal point of view and when the athlete's name feature is dropped) records.

On further detailed screening of the dataset four times, it was uncovered that all sports and discipline group medal count greater than one does not represent one sport. Certain records could be sub-categorized and assessed under a different category. For instance, sport event Athletics can be further classified into 100M, 400M, 800M, and 1500M events. Hence event attribute is also employed for grouping team game records into one. The revised and new updated dataset contains 15553 pieces of data. The number of group events are 733 and sport categories are 48. And percentage of male and female record is 61.25% and 38.75% respectively. The deduplicated data was acquired by using and making modifications to the ‘preprocessed_olympic’ data. A new attribute ‘home_advantage’ is generated where winning country value of each record is compared with the hosting country attribute. If the value matches, then 1 is inserted under the ‘home_advantage’ variable otherwise 0 is included. There are 1696 instances where the player won an Olympic medal in his homeland. There is inclusion number of non-hosting nation countries winning records for the analysis as there is no information available to claim that it will develop any unbiased home advantage. And it is a fair assessment to use all 15553 records for prediction. If the were no data points available for any winning country in the record, there was no leveraging the average of the respective region as benchmark. The reason for this approach is because benchmarking would have worked for some features of smaller countries, which are responsible for 1% of the total Olympic medals [32].

C. Data Transformation and Integration

Data integration was done before missing value processing for the research as the primary key attribute employed by the datasets for the analysis were not consistently embodied applying the same standard for the attribute. The Winter and Summer Game data were accessible separately. They were unified vertically before the preprocessing step. Before merging the dataset, ‘Medal’ attribute of the table was transformed (variable transformation) for good quality analysis. There was no transformation of data done from one format to another while collecting it from different data sources. Variable transformation (normalization and standardization) of all the attributes of the final dataset were done prior breaking the dataset for training, validation, and testing data. It was done to make records(information of Olympic Game) usable with Neural Network [1].

V. Data Analysis

For good analysis outcome and prediction accuracy of the neural network model, the data must be standardized and normalized. The final dataset is analyzed, visualized. Verification of the features in the dataset is done before forecasting the end results.

A. Correlation Coefficient

Fig. 1 displays correlation between the variables without normalization. It could be noticed that the feature ‘Medal’ doesn’t have high linear positive correlation with any feature in the dataset. A moderate correlation is observed between ‘City’ attribute and features related to the sport (‘Sport’, ‘Discipline’, ‘Event’). Even though there are some interesting correlations between variables is seen the focus of our analysis is relationship between ‘Medal’ and other variables.
After normalization of major features there was no much changes in relationship between them in Fig. 2. The correlation between ’medal_numeric’ and ‘GDP per Capita’ is 0.036, indicating that they are weakly related. Same is the case with ‘Home Advantage’ and ‘Population’.

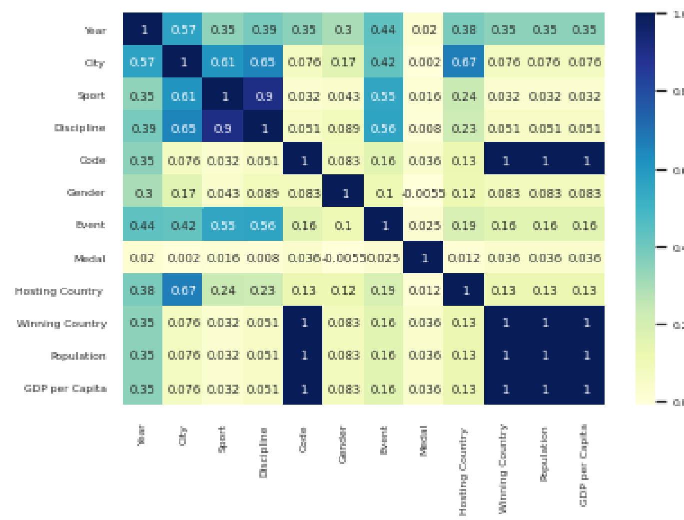

Fig. 1. Correlation coefficient between variables without normalization

Fig. 2. Correlation between certain features after normalization

B. Feature Selection and Verification

The correlograms displays the degree of linear correlation between features. The poor linear relation among the features indicate that they are irrelevant. Linear relationship between them is not obvious. High productivity (determined by GDP per capita) implies that country can pay for sportsmen to partake in the Olympic Games, it may also be associated to better training and equipment [23].The performance of the model can be escalated through selection of a combination of important features that represent maximal separation between the classes [5]. The feature selection and verification process reduce the dimension of the dataset by eliminating irrelevant and redundant features from the data. Through the acquisition of a minimum set of the original features, this technique enables data mining algorithms to operate faster and more effectively, while improving results and comprehensibility of the model [19].

C. Exploratory Data Analysis

Exploratory Data Analysis was done using Tableau software to understand the relationship between the features and discover if there is any other attribute that influences the medal count.Fig. 3 and Fig. 4 display details about the medal.From Fig. 3, it is evident that there is an association between medal count and home track advantage. However, the home-field advantage is not the same for all the countries. But maximum in favor of the United States and the United Kingdom.On comparing the graph results with the general published facts, it was unveiled that the data preprocessing(cleaning, replacing missing values) carried out for the research turned out to be unerring. Therefore, the graph includes medal count only up till 2014. The count of Gold, Silver, and Bronze was segregated with the help of their color in the diagram.

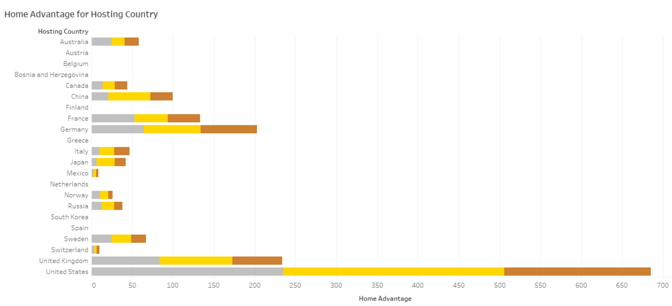

Fig. 3. Home Advantage for Hosting Countries

The fourth diagram portrays the medal count of both men and women for the top 20 winning countries. Again, the medal count of Russia is less as many records related to Russia were dropped in the course of the preprocessing step.

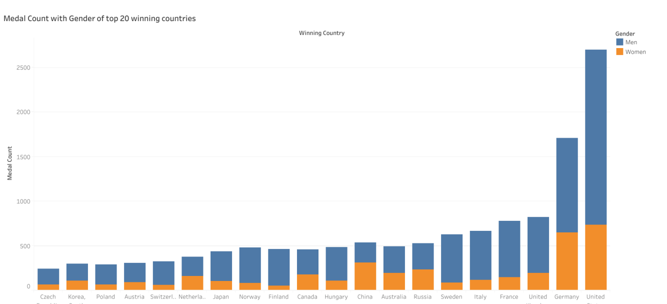

Fig. 4. Medal Count of Men and Women from the Winning Counties

D. Training, Testing, and Validation Set

For proper classification and working, 80% training dataset is select randomly for training, and the other 20% for testing. The training dataset is further allotted in the ratio of 80:20 for training and validation. The remaining 20% of the data will be utilized for validation to avoid overtraining and overfitting of the data [20].

VI. Model

Post examining the related work and analyzing the correlation between medal and other features, it could be determined that the features don't have linear relationship.Artificial Neural Network model is adapted to study and find the medal list. It has been extensively used for sport prediction [3] [14] [44]. Multilayer Perceptron (MLP), the multilayer, a special type of ANN, is used to see if France is among the top-ranking country or not in the 2024 Olympic Games. The algorithm is briefly described next.

A. Artificial Neural Network (ANN)

Multilayer perceptron neural network (MLP) is a popular ANN model used for prediction. All types of Artificial Neural Network can be applied for classification, clustering, feature mining,pattern recognition, and forecasting. An ANN normally encompasses interconnected components (neurons) that transform a set of inputs into a desired output [42]. MLP training algorithm established on input forward propagation and error backward propagation followed by an apprise of the weights of the network using gradient descent methods. The error is determined at the last layer also called the output layer as the change between the actual and predicted output at each mode [20]. A detailed explanation of neural networks and MLP learning algorithm is accessible in the literature [21].The limitation of executing the MLP model is that it does not clarifies the association between the features making the decision. The main problem with the use of MLP is to define the size of the initial network about its hidden layers and number of neurons [37]. The pruning method can make the classifier faster and determine which neurons in the network can be terminated without the serve impairment of the neural network's performance [37]. Also, the pruning model is targeted to cut neurons displaying only to decrease the misclassification. However, it is not very suitable for issues with unbalanced data. Another major drawback of using pruning in MLP is it does not quantify much beyond the original accuracy. In the final comment, using an MLP pruning network one can reduce inference time, save some power, and reduce the storage requirement.

B. Optimisation

VII. Methodology

Data Mining the experimental science, can have no universal best algorithm [43]. The research study uses the CRISP-DM framework method; it offers a structured way of conducting the data analysis [34]. The methodology has six main steps [12]: apprehending the domain problem and stating the objective of the research; uncovering, accessing, and exploring the data source; preprocessing the applicable data; building the model; evaluating and accessing the validity and utility of the model; implementing the model.

VIII. Evaluation and Results

IX. Conclusion

X. Future Work

The future goal is to predict the winners of the host city elections for the Summer and Winter Olympic games alongside the champions of the Olympic games. The Olympic host city is picked in an exhaustive secret ballot 7 years before the Olympics by the International Olympic Committee (IOC). A type of continental rotation is noticed in hosting locations in the case of the Summer Olympics because no two successive games have been held on the same continent [35] [2].

A. Figures and Tables

a) Positioning Figures and Tables: Place figures and tables at the top and bottom of columns. Avoid placing them in the middle of columns. Large figures and tables may span across both columns. Figure captions should be below the figures; table heads should appear above the tables. Insert figures and tables after they are cited in the text. Use the abbreviation "Fig. 4", even at the beginning of a sentence.

ACKNOWLEDGMENT

XI. REFERENCE

REFERENCES

[1] Paulo JL Adeodato. Variable transformation for granularity change in hierarchical databases in actual data mining solutions. In *International Conference on Intelligent Data Engineering and Automated Learning*, pages 146–155. Springer, 2015.

[2] Greg Andranovich, Matthew J Burbank, and Charles H Heying. Olympic cities: lessons learned from mega-event politics. *Journal of urban affairs*, 23(2):113–131, 2001.

[3] Burak Galip Aslan and Mustafa Murat Inceoglu. A comparative study on neural network based soccer result prediction. In *Seventh International Conference on Intelligent Systems Design and Applications (ISDA 2007)*, pages 545–550. IEEE, 2007.

[4] Nigel J Balmer, Alan M Nevill, and A Mark Williams. Modelling home advantage in the summer olympic games. *Journal of sports sciences*, 21(6):469–478, 2003.

[5] Rezaul Begg and Joarder Kamruzzaman. A machine learning approach for automated recognition of movement patterns using basic, kinetic and kinematic gait data. *Journal of biomechanics*, 38(3):401–408, 2005.

[6] Andrew B Bernard and Meghan R Busse. Who wins the olympic games: Economic resources and medal totals. *Review of economics and statistics*, 86(1):413–417, 2004.

[7] Amit Bhagat, Sanjay Sharma, and KR Pardasani. November 2010—feed forward neural network algorithm for frequent patterns mining‖. *International Journal of Computer Science and Information Security*, 8(8):201–205.

[8] Xun Bian et al. Predicting olympic medal counts: The effects of economic development on olympic performance. *The park place economist*, 13(1):37–44, 2005.

[9] Shuixia Chen, Jian-qiang Wang, and Hong-yu Zhang. A hybrid pso-svm model based on clustering algorithm for short-term atmospheric pollutant concentration forecasting. *Technological Forecasting and Social Change*, 146:41–54, 2019.

[10] Charisios Christodoulos, Christos Michalakelis, and Dimitris Varoutas. Forecasting with limited data: Combining arima and diffusion models. *Technological forecasting and social change*, 77(4):558–565, 2010.

[11] Edward M Condon, Bruce L Golden, and Edward A Wasil. Predicting the success of nations at the summer olympics using neural networks. *Computers & Operations Research*, 26(13):1243–1265, 1999.

[12] Dursun Delen, Douglas Cogdell, and Nihat Kasap. A comparative analysis of data mining methods in predicting ncaa bowl outcomes. *International Journal of Forecasting*, 28(2):543–552, 2012.

[13] Paul Downward, Bernd Frick, Brad R Humphreys, Tim Pawlowski, Jane E Ruseski, and Brian P Soebbing. *The SAGE handbook of sports economics*. SAGE, 2019.

[14] Jürgen Edelmann-Nusser, Andreas Hohmann, and Bernd Henneberg. Modeling and prediction of competitive performance in swimming upon neural networks. *European Journal of Sport Science*, 2(2):1–10, 2002.

[15] David Forrest, Ismael Sanz, and Juan D Tena. Forecasting national team medal totals at the summer olympic games. *International Journal of Forecasting*, 26(3):576–588, 2010.

[16] William F Gayton and Guy Langevin. Home advantage: Does it exist in individual sports. *Perceptual and Motor Skills*, 74(3):706–706, 1992.

[17] William F Gayton, Sharon A Mutrie, and Joseph F Hearns. Home advantage: Does it exist in women's sports. *Perceptual and motor skills*, 65(2):653–654, 1987.

[18] Francis D Glamser. Contest location, player misconduct, and race: A case from english soccer. *Journal of Sport Behavior*, 13(1):41, 1990.

[19] Jiawei Han, Jian Pei, and Micheline Kamber. *Data mining: concepts and techniques*. Elsevier, 2011.

[20] Md Rafiul Hassan, Sadiq Al-Insaif, M Imtiaz Hossain, and Joarder Kamruzzaman. A machine learning approach for prediction of pregnancy outcome following ivf treatment. *Neural computing and applications*, 32(7):2283–2297, 2020.

[21] Simon Haykin and N Network. A comprehensive foundation. *Neural networks*, 2(2004):41, 2004.

[22] Mengjie Jia, Yue Zhao, Furong Chang, Bofeng Zhang, and Kenji Yoshigoe. A random forest regression model predicting the winners of summer olympic events. In *Proceedings of the 2020 2nd International Conference on Big Data Engineering*, pages 62–69, 2020.

[23] Daniel KN Johnson and Ayfer Ali. A tale of two seasons: participation and medal counts at the summer and winter olympic games. *Social science quarterly*, 85(4):974–993, 2004.

[24] Timothy Koba. *Comparing the Success of Official Sponsors and Ambush Marketers: An Event Study Analysis of Brazil Following the 2014 Fifa World Cup and 2016 Rio De Janeiro Summer Olympic Games*. PhD thesis, University of South Carolina, 2020.

[25] Marcos P Estellita Lins, Eliane G Gomes, João Carlos CB Soares de Mello, and Adelino José R Soares de Mello. Olympic ranking based on a zero sum gains dea model. *European Journal of Operational Research*, 148(2):312–322, 2003.

[26] Sebastián Lozano, Gabriel Villa, Fernando Guerrero, and Pablo Cortés. Measuring the performance of nations at the summer olympics using data envelopment analysis. *Journal of the Operational Research Society*, 53(5):501–511, 2002.

[27] Robert Madrigal and Jeffrey James. Team quality and the home advantage. *Journal of Sport Behavior*, 22(3), 1999.

[28] R Hugh Morton. Who won the sydney 2000 olympics?: an allometric approach. *Journal of the Royal Statistical Society: Series D (The Statistician)*, 51(2):147–155, 2002.

[29] Alan M Nevill, Nigel J Balmer, and A Mark Williams. The influence of crowd noise and experience upon refereeing decisions in football. *Psychology of sport and exercise*, 3(4):261–272, 2002.

[30] Alan M Nevill and Roger L Holder. Home advantage in sport. *Sports Medicine*, 28(4):221–236, 1999.

[31] Alan M Nevill, Roger L Holder, Andrew Bardsley, Helen Calvert, and Stephen Jones. Identifying home advantage in international tennis and golf tournaments. *Journal of Sports Sciences*, 15(4):437–443, 1997.

[32] Christoph Schlembach, Sascha L Schmidt, Dominik Schreyer, and Linus Wunderlich. Forecasting the olympic medal distribution during a pandemic: a socio-economic machine learning model. *arXiv preprint arXiv:2012.04378*, 2020.

[33] Barry Schwartz and Stephen F Barsky. The home advantage. *Social forces*, 55(3):641–661, 1977.

[34] Colin Shearer. The crisp-dm model: the new blueprint for data mining. *Journal of data warehousing*, 5(4):13–22, 2000.

[35] Noam Shoval. A new phase in the competition for the olympic gold: the london and new york bids for the 2012 games. *Journal of urban affairs*, 24(5):583–599, 2002.

[36] William F Shughart and Robert D Tollison. Going for the gold: property rights and athletic effort in transitional economies. *Kyklos*, 46(2):263–272, 1993.

[37] Miriam Rodrigues Silvestre and Lee Luan Ling. Pruning methods to mlp neural networks considering proportional apparent error rate for classification problems with unbalanced data. *Measurement*, 56:88–94, 2014.

[38] Herman O Stekler, David Sendor, and Richard Verlander. Issues in sports forecasting. *International Journal of Forecasting*, 26(3):606–621, 2010.

[39] STEFANO TETTAMANTI and LUCA TAMAGNI. Predicting olympic games: do macro variables still matter? insight from a prediction model applied to both genders. 2020.

[40] Baowei Wang, Xiaodu Gu, Li Ma, and Shuangshuang Yan. Temperature error correction based on bp neural network in meteorological wireless sensor network. *International Journal of Sensor Networks*, 23(4):265–278, 2017.

[41] Yichuan Wang, LeeAnn Kung, and Terry Anthony Byrd. Big data analytics: Understanding its capabilities and potential benefits for healthcare organizations. *Technological Forecasting and Social Change*, 126:3–13, 2018.

[42] Ian H Witten and Eibe Frank. Data mining: practical machine learning tools and techniques with java implementations. *Acm Sigmod Record*, 31(1):76–77, 2002.

[43] David H Wolpert and William G Macready. No free lunch theorems for optimization ieee transactions on evolutionary computation. *E997*, 1997.

[44] William A Young, William S Holland, and Gary R Weckman. Determining hall of fame status for major league baseball using an artificial neural network. *Journal of Quantitative Analysis in Sports*, 4(4), 2008.

[45] ZHANG Yuhua. Prediction of chinese delegation medal number in the thirty-first session of olympic games by linear regression dynamic model. *Journal of Henan Normal University (Natural Science Edition 2013-02*.

[46] E Paul Zehr. Neural control of rhythmic human movement: the common core hypothesis. *Exercise and sport sciences reviews*, 33(1):54–60, 2005.