Discovering if home-field advantage for France will aid the nation in obtaining a top rank in the forthcoming 2024 Olympic Game

Shruti Sindhi

Abstract:

The athletes' success in any sport is considered a source of national prestige. 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. The type of literature used in this review is a systematic literature review (meta-synthesis). High-income countries specialize less in a particular sport, as they win medals in a more diversified range of sports. But many other characteristics affect the country's success in any universal game. 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 research paper will assist in foretelling whether home track aids France in obtaining a top rank in the upcoming 2024 Olympic Games. Furthermore, the winter Olympic dataset is incorporated during the analysis to get a better result.

Introduction:

The Olympic game spirit focuses on participation rather than winning. But the ground reality is different; people consider by-country medal count. The ability of the country to perform and excel in numerous sports in this 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.(Bernard & Busse, 2004) Sometimes a nation's Olympic success can be because of comparative advantages.(Tcha & Pershin, 2003) At times the unmeasurable cultural factors can be the reason for comparative advantage. For instance, India's intense fondness for cricket justifies poor attainment in Olympic sports. It explains the unequal distribution of medal numbers, but it can also depend on the relative strength of countries in different sports. (Bian, 2005) There are always some advantages for the players to have a sports event hosted on the home ground. More fans turn up for the event as they do not have to travel far to attend and cheer for the home teams and athletes. The existence of home advantage in major team sports has been well established. (Nevill & Holder, 1999) However, the prevalence of home advantage in both individual sport and unbalanced competition is comprehensively evident. (Balmer et al., 2003)

Related Work:

Even after official claims by the IOC that the Olympic Games are competitions between athletes in individual or team events and not between nations, the rank of countries according to the total medals won are maintained on the scoreboard. (Condon et al., 1999) 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. (Jia et al.) 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. (Jia et al.) Also, Bernard and Busse estimated a Tobit regression equation (which allowed a large number of countries with no medals in a game) employing data from 1996. (Forrest et al., 2010) Many researchers consider Tobit Regression as the reference model for predicting the Olympic medal table.(Tettamanti & Tamagni, 2020) But 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. (Zehr, 2005) Therefore, sports analysts or experts should use non-linear models when uncovering relationships between a different range of features. Some other studies revealed they did predictions using Neural Network. (Condon et al., 1999) One issue tackled by performing pruning on ANN back propagation algorithm is pruning. (Condon et al., 1999) 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. (Tettamanti & Tamagni, 2020)

Methodology:

The methodology incorporated will find the medals France will win by considering the factors influencing national Olympic success, including home track advantage. (Donald, 1972) Data were clustered into three clusters to categorize China's gold medal-winning success while forecasting its performance in Beijing Olympic Games to get a detailed knowledge of the success rate of getting gold medals under all 301 events.(Shibli & Bingham, 2008) Similarly, the utilized data can be clustered to uncover France's success rate of earning Gold under each group in the 2024 Olympic games. It will assist in acquiring better data insight when conducted along with exploratory data analysis. Another sub-objective that will be achieved in the process is locating the percentage of French males and females grabbing the medal.

Multilayer Perceptron (MLP), the multilayer feed-forward network, will be employed to see if France is among the top-ranking countries or not in the 2024 Olympic Games. It is implemented as the single-layer network cannot learn non-linear functions. There is a possibility of the presence of non-linear functions in forecasting the Olympic winners. Deep Learning or deep neural network is better than Multilayer Network as they are good at processing data. But they can cause vanishing gradient problems, so they are not applied to Olympic Games data. Based on estimated results, the total square of error after reaching balance was 4.12 units.(Fazlollahi et al., 2020), 1.48 units less than the Random Forest model used to get the same result. (Jia et al.) The limitation of executing the MLP model is that it does not clarify 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. (Silvestre & Ling, 2014) 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. (Silvestre & Ling, 2014) 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.

Conclusion:

It was recently pointed out that a large amount of effort is spent forecasting the outcomes of sporting events. (Boulier & Stekler, 2003) Compared to the previous studies on Olympic game performance, this suggested method focuses on unearthing the socioeconomic variables, including Population, GDP per Capita, and hosting advantage, which significantly impacts the Country's overall performance. (Bian, 2005) The FP growth model is better than the Apriori model as it generates patterns by constructing an FP Tree. The research in the Olympic game domain has resulted in a wide variety of powerful ANNs based on novel formulations of the input space, neuron, type, several synaptic connections direction of information flow in the ANN, cost or error function, learning mechanism, output space, and various combinations of these. (Ghosh-Dastidar & Adeli, 2009). Multilayer perceptron model will be implemented in case study to foreknow about the champions of the sporting event as both neural network and decision tree can model data as the nonlinear relationship between variables. Eventually, with Apriori-Feed Forward (AFF), both pattern mining and prediction on large sports datasets can be made by scanning the dataset only once. (Bhagat et al.) The projection will aid bidding agencies in employing the analyzed data to determine which nation they want to sponsor. The result of the study will also be beneficial for the participating countries of the world to enhance their performance. At the same time, one can track and judge sports achievement's development law and characteristics.(Wang et al.)

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.(Shoval, 2002) (Andranovich et al., 2001)

References:

Andranovich, G., Burbank, M. J., & Heying, C. H. (2001). Olympic cities: lessons learned from mega-event politics. *Journal of urban affairs*, *23*(2), 113-131.

Balmer, N. J., Nevill, A. M., & Williams, A. M. (2003). Modelling home advantage in the Summer Olympic Games. *Journal of sports sciences*, *21*(6), 469-478.

Bernard, A. B., & Busse, M. R. (2004). Who wins the Olympic Games: Economic resources and medal totals. *Review of economics and statistics*, *86*(1), 413-417.

Bhagat, A., Sharma, S., & Pardasani, K. R. November 2010—Feed Forward Neural Network algorithm for Frequent patterns mining‖. *International Journal of Computer Science and Information Security*, *8*(8), 201-205.

Bian, X. (2005). Predicting Olympic medal counts: The effects of economic development on Olympic performance. *The park place economist*, *13*(1), 37-44.

Boulier, B. L., & Stekler, H. O. (2003). Predicting the outcomes of National Football League games. *International Journal of forecasting*, *19*(2), 257-270.

Condon, E. M., Golden, B. L., & Wasil, E. A. (1999). Predicting the success of nations at the Summer Olympics using neural networks. *Computers & Operations Research*, *26*(13), 1243-1265.

Donald, W. B. (1972). Olympic games competition: structural correlates of national success. *International Journal of Comparative Sociology*, *13*, 186.

Fazlollahi, P., Afarineshkhaki, A., & Nikbakhsh, R. (2020). Predicting the Medals of the Countries Participating in the Tokyo 2020 Olympic Games Using the Test of Networks of Multilayer Perceptron (MLP). *Annals of Applied Sport Science*, *8*(4), 0-0.

Forrest, D., Sanz, I., & Tena, J. D. (2010). Forecasting national team medal totals at the Summer Olympic Games. *International Journal of Forecasting*, *26*(3), 576-588.

Ghosh-Dastidar, S., & Adeli, H. (2009). A new supervised learning algorithm for multiple spiking neural networks with application in epilepsy and seizure detection. *Neural networks*, *22*(10), 1419-1431.

Jia, M., Zhao, Y., Chang, F., Zhang, B., & Yoshigoe, K. (2020). A Random Forest Regression Model Predicting the Winners of Summer Olympic Events.

Nevill, A. M., & Holder, R. L. (1999). Home advantage in sport. *Sports Medicine*, *28*(4), 221-236.

Shibli, S., & Bingham, J. (2008). A forecast of the performance of China in the Beijing Olympic Games 2008 and the underlying performance management issues. *Managing Leisure*, *13*(3-4), 272-292.

Shoval, N. (2002). 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.

Silvestre, M. R., & Ling, L. L. (2014). Pruning methods to MLP neural networks considering proportional apparent error rate for classification problems with unbalanced data. *Measurement*, *56*, 88-94.

Tcha, M., & Pershin, V. (2003). Reconsidering performance at the Summer Olympics and revealed comparative advantage. *Journal of Sports economics*, *4*(3), 216-239.

Tettamanti, S., & Tamagni, L. (2020). Predicting Olympic Games: do macro variables still matter? Insight from a prediction model applied to both genders.

Wang, B., Gu, X., Ma, L., & Yan, S. (2016). Temperature error correction based on BP neural network in meteorological wireless sensor network.

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