

Evaluation of Athletic Events of Olympic History for 100 Years Using Ranking Algorithm

C.Lalitha^{1*}, S. Arulselvarani²

¹Dept. of Computer Science, C.T.T.E College for Women, Chennai, India

²Dept. of Computer Science, Stet Women's College, Mannargudi, India

Corresponding Author: lalithabarathi78@gmail.com

DOI: <https://doi.org/10.26438/ijcse/v7si5.156160> | Available online at: www.ijcseonline.org

Abstract- This paper investigates the result of athletic games of Olympic history for the past 100 years. As a case study, we evaluate ZeroR classification machine algorithms on game datasets. Here we compare the dataset in ranking algorithms in order to determine the results like leading city and the winner of the game. In this paper, we used machine learning data mining tool WEKA for different analysis. We have provided an evaluation based on applying these classification methods on our datasets and measuring the accuracy of test results.

Keywords: *Classification, data mining tool, machine learning, Ranking, WEKA, ZeroR.*

I. INTRODUCTION

Usually to gather tremendous measure of information for every player, team, game and season for some sports. This gives us the plan to classify and ranking few algorithms for data mining on datasets that contain records of sports games. The target of our exploration is to test different information digging calculations for the prediction of the ultimate result. It is hard to locate a correct reason however to utilize a set of outcomes to obscure one. The classifiers and rankers that are utilized in the prediction are actualized in WEKA (Waikato Environment for Knowledge Analysis). Despite the fact that there are many games in the Olympic, we considered just the athletic events (Men and Women) for as long as 100 years (1916-2016). In this paper, prediction methods were used instead of traditional methods. The evaluation results show the accuracy of our sports data mining approach in predicting the outcomes of games in recent seasons.

II. REVIEW LITERATURE

E.W.T Ng et al [1] reviewed & customer relationship management by applying data mining technique CRM. In this paper, the review and classification process were independently verified. Findings indicate that the research area of customer retention received most research attention.

Xuiyiran et al [2] proposed three link analysis algorithms based on the preprocess of support vector machine to estimate an applicant's credit. It is shown that the genetic

link analysis ranking methods have higher performance in terms of classification accuracy.

Joel Brooks [3] developed a data driven player ranking in soccer using predictive model weights. In this paper, a novel method is used to understand the relationship between pars location & shot opportunities.

Carson K. Leung*, Kyle W. Joseph[7] described about a sports data mining approach which helps discover interesting knowledge and predict outcomes of sports games such as college football.

Eftim Zdravevski et. al.[8] analyzed the system for prediction of the winner in the sports game. Here WEKA tool is used to predict the results by using different set of datasets for testing and classifying.

III. METHODOLOGY

In each datum mining process, the underlying data must be experienced couple of phases of processing of system so as to separate valuable data. The below graph demonstrate the phases of the handling of system. The phases of information preparing would be less demanding if the framework is structured separately. By accomplishing this, every one of the modules can be successive and actualized conditionally to one other. Despite the fact that numerous arrangement calculations have been utilized, ZeroR calculation gave best outcome.

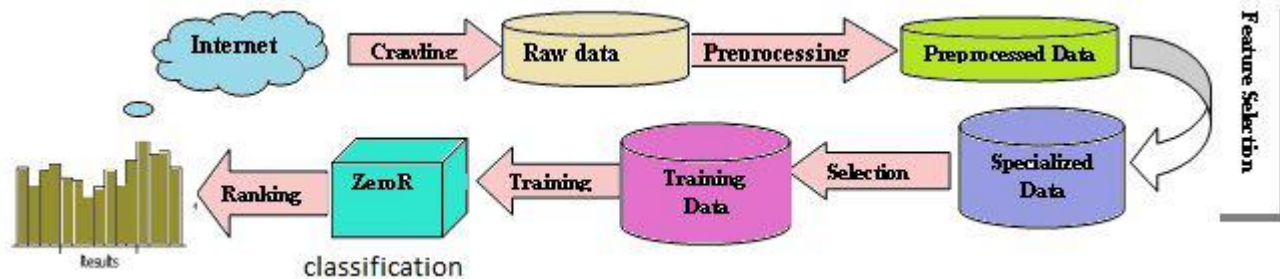


Figure 1: Methodology

IV. DATA COLLECTION

Here the data is about only the athletic events, which has been played in the Olympic for as long as 100 years. Be that as it may, we considered just the athletic events for both Men and Women. The data, which is downloaded from the web, can't be utilized straightforwardly .It needs to change over into csv document and after that it is utilized for preprocessing. Despite the fact that it is simple to discover

the outcome and the measurements of an athletic event on the web, not openly accessible dataset can be downloaded and imported into some database. This upholds the framework to have a module for getting information for the game of interest from the web and storing it into a database. **Table1** demonstrates the input for the dataset with the attributes of name of the athlete, games, sex, year, season, sport, event and medal.

Table 1 : Input in CSV Format

ID	Name	Sex	Games	Year	Season	City	Sport	Event	Medal
1724	Aristidis Akratopoulos	M	1896 Summer	1896	Summer	Athina	Tennis	Tennis Men's Singles	NA
1724	Aristidis Akratopoulos	M	1896 Summer	1896	Summer	Athina	Tennis	Tennis Men's Doubles	NA
1725	Konstantinos "Kostas" Ak	M	1896 Summer	1896	Summer	Athina	Tennis	Tennis Men's Singles	NA
1725	Konstantinos "Kostas" Ak	M	1896 Summer	1896	Summer	Athina	Tennis	Tennis Men's Doubles	NA
4113	Anastasios Andreou	M	1896 Summer	1896	Summer	Athina	Athletics	Athletics Men's 110 metres Hurdles	NA
4116	Ioannis Andreou	M	1896 Summer	1896	Summer	Athina	Swimming	Swimming Men's 1,200 metres Freestyle	Silver
4189	Nikolaos Andriakopoulos	M	1896 Summer	1896	Summer	Athina	Gymnastics	Gymnastics Men's Rope Climbing	Gold
4431	Georgios Anninos	M	1896 Summer	1896	Summer	Athina	Swimming	Swimming Men's 100 metres Freestyle	NA
4493	Antelothanasis	M	1896 Summer	1896	Summer	Athina	Shooting	Shooting Men's Free Rifle, Three Position	NA
5660	Georgios Aspiotis	M	1896 Summer	1896	Summer	Athina	Cycling	Cycling Men's Road Race, Individual	NA
8620	Edward Battel	M	1896 Summer	1896	Summer	Athina	Cycling	Cycling Men's Road Race, Individual	Bronze
8620	Edward Battel	M	1896 Summer	1896	Summer	Athina	Cycling	Cycling Men's 333 metres Time Trial	NA
8620	Edward Battel	M	1896 Summer	1896	Summer	Athina	Cycling	Cycling Men's 100 kilometres	NA
8747	Albert Baumann	M	1896 Summer	1896	Summer	Athina	Shooting	Shooting Men's Military Rifle, 200 metres	NA
9760	Spyridon Belokas	M	1896 Summer	1896	Summer	Athina	Athletics	Athletics Men's Marathon	NA
12068	Arthur Charles Blake	M	1896 Summer	1896	Summer	Athina	Athletics	Athletics Men's 1,500 metres	Silver
12068	Arthur Charles Blake	M	1896 Summer	1896	Summer	Athina	Athletics	Athletics Men's Marathon	NA
12563	Conrad Helmut Fritz Bcker	M	1896 Summer	1896	Summer	Athina	Gymnastics	Gymnastics Men's Horse Vault	NA
12563	Conrad Helmut Fritz Bcker	M	1896 Summer	1896	Summer	Athina	Gymnastics	Gymnastics Men's Parallel Bars	NA
12563	Conrad Helmut Fritz Bcker	M	1896 Summer	1896	Summer	Athina	Gymnastics	Gymnastics Men's Parallel Bars, Teams	Gold
12563	Conrad Helmut Fritz Bcker	M	1896 Summer	1896	Summer	Athina	Gymnastics	Gymnastics Men's Horizontal Bar	NA
12563	Conrad Helmut Fritz Bcker	M	1896 Summer	1896	Summer	Athina	Gymnastics	Gymnastics Men's Horizontal Bar, Teams	Gold

V. DATA PREPROCESSING

After every single required datum is stored in a database, it must be preprocessed. The preprocessing may refer to normalization and discretization of a few parameters in a given range. Preprocessing is done in which the data is being collected, since the technique is progressively efficient. The majority of the preprocessing methods are implemented as stored procedures and functions in the database. Furthermore, works in the database. A few of them represent

potential features, while others are facilitating the computation of the previous. The feature computation methods are invoked before the start of each training stage or test stage. Each time they are summoned their outcome is utilized as an input to the CSV[9] producing module that prepares the input to the WEKA. The framework results from the feature computation strategies are not stored in the database for flexibility and adaptability reasons.

VI. FEATURE SELECTION

The feature selection is manual. Notwithstanding the problem domain what's more, the classification or ranking algorithm, the training and test data must be represented as a lot of data points. Every datum point is N dimensional space and each coordinate of the data point represents a feature. It is executed as a stored function in the database that returns a table as a result. Every section in the subsequent table represents to an estimation of one feature and each row as a data point.

VII. IMPLEMENTATION OF ZEROR CLASSIFICATION

To summon order, classification or clustering or ranking algorithms from WEKA, an interface must be actualized. These algorithms can be invoked in the form of command line with a few explicit parameters [10].The CSV documents

contain the input dataset for the algorithm that is being invoked. They are generated utilizing the results from the feature selection module. The output format can be arranged with a same command. The output has to be caught, parsed and then stored.

ZeroR is the simplest classification method which relies and ignores all predictors. It simply predicts the majority class. Although there is no predictability power in ZeroR, it is useful for determining a baseline performance as a benchmark for other classification methods. Here we have considered some important classifiers for classifying only the athletic games for both men and women Among all the classifiers the ZeroR classifier gave a best precision value and classified the medals appropriately.(Gold =38,Silver=44 and Bronze =.41). **Figure2** shows the output of ZeroR for both Men and Women.

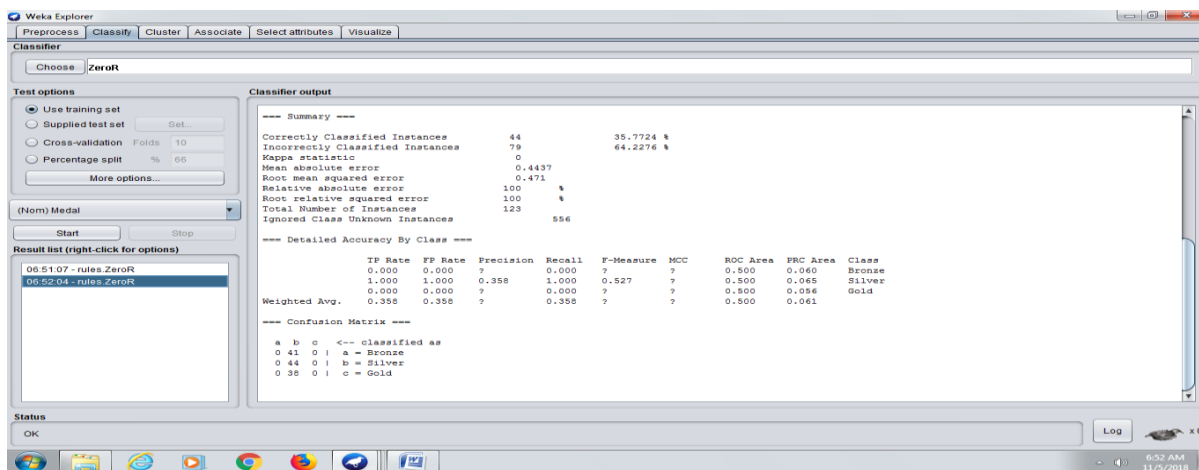


Figure 2:ZeroR classification

Table 2: Result of classifier Method

Game	Classifier	FP Rate	Precision	Recall	F-Measure
Athletic Women	ZeroR	1	0.892	1	0.943
	Rules Decision Table	0	1	1	1
	WA5000M	0.51	0.348	0.727	0.471
	WA1500M	0.017	0.444	0.364	0.444
	Ruleone	0	1	1	1
	WA5000M	0.048	0.3	0.545	0.387
	WA1500M	0.014	0.556	0.445	0.5
	Rules PART	1	0.692	1	0.943
	WA5000M	0	0.623	1	0.847
WA1500M	0	0.634	1	0.744	
Athletic Men	ZeroR	1	0.81	1	0.895
	Rules Decision Table	0.109	0.228	0.75	0.35
	MA5000M	0.002	0	0	0
	MA5000M	0.002	0	0	0
	MA1500M	0	1	0.273	0.429
	Rules PART	1	0.819	1	0.9
	MA5000M	0.111	0.226	0.75	0.347
MA1500M	0.039	0.133	0.25	0.12	

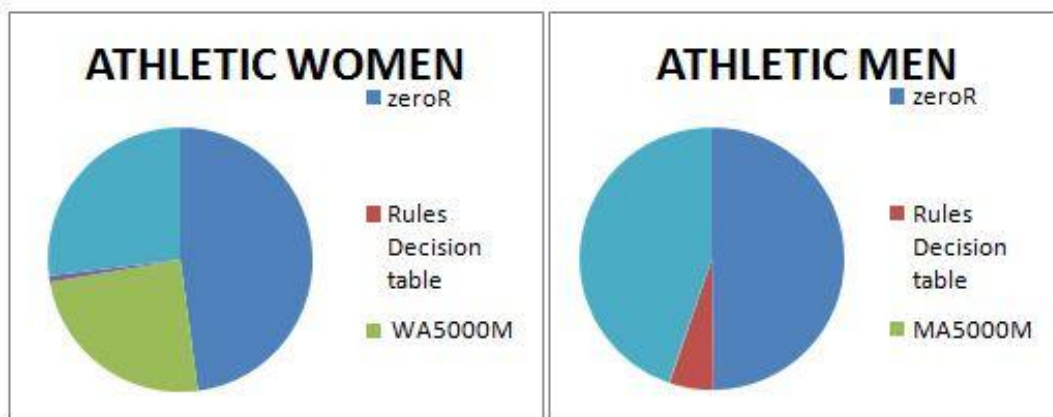


Figure 3 : Output of Classifier Method

VIII. RANKER IN CLASSIFICATION ALGORITHM

The ranker search method sorts attributes according to their evaluation parameters such as default, retain all, ignore, etc. OneR, C4.5, are the metrics for evaluating the attributes.

Here the InfoGainEval evaluator has given best result in case of Ranker Algorithm. The city has been ranked with the games and thus it ranks the dataset according to the value of the weights.

IX. RESULTS AND CONCLUSIONS

In this paper, we displayed the sports data mining approach to deal with prediction of the winners of world Athletic games. More often than not, the traditional methodology has been pursued to look at the competing teams and project the results. However, our methodology predicts the weights of attributes dependent on the historical results of games. Contrasting the four classification algorithms (OneR, ZeroR, Rule PART and Decision Table Tree) in 10 fold cross

validation, we have found the advantages of every measurement. According to the previous performance analysis, we have discovered the outcome that ZeroR performed well in categorize the test mode of Athletic Sports. Athletic ranking system can provide point that is more interesting for the overall people. As future work, we intend to investigate the data for future, world level team, games, seasons and other sports games.

SNo	City	Weight
1	London	5661
2	Athina	3664
3	Rio de Janeiro	3485
4	Syndey	3303
5	Beijing	3285
6	Los Angels	3271
7	Atlanta	3233
8	Barcelona	3123
9	Seoul	3005
10	Munich	2481
11	Montreal	2213
12	Mexico city	2152
13	Roma	2119
14	Helsinki	1993
15	Paris	1960
16	Moskva	1856
17	Tokyo	1797
18	Berlin	1503
19	Amsterdan	1192
20	Antwerpon	1148

Figure 4:Output of Ranking Algorithm

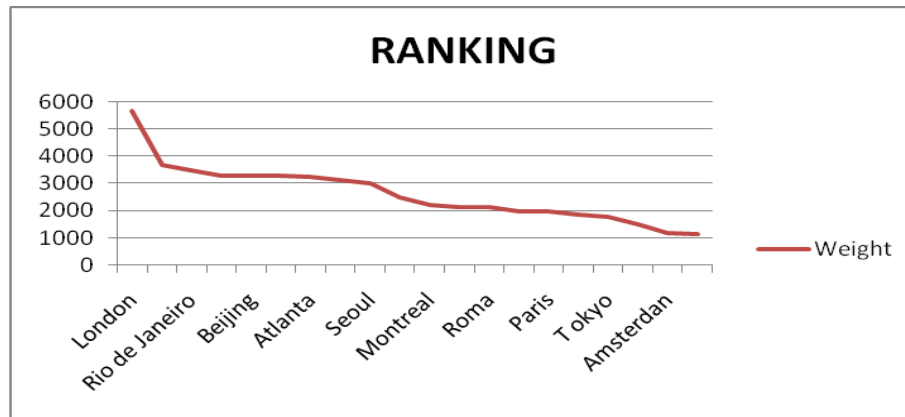


Figure 5: Graph of Ranking Algorithm

REFERENCES

- [1] Osama K. Solieman, "Data Mining in Sports: A Research Overview" MIS Masters Project August 2006.
- [2] E. W. T. Ngaia Li Xiub D. C. K. Chaua, "Application of data mining techniques in customer relationship management: A literature review and classification" IEEE, March 2018.
- [3] Lorena Siguenza-Guzman et. al. "Literature Review of Data Mining Applications in Academic Libraries", The Journal of Academic Librarianship, Volume 41, Issue 4, 2015, pp. 499-510.
- [4] T. Femina Bahari, M. Sudheep Elayidom, "An Efficient CRM-Data Mining Framework for the Prediction of Customer Behaviour", Procedia Computer Science, Volume 46, 2015, pp. 725-731.
- [5] Mark A. Hall, Geoffrey Holmes, "Benchmarking Attribute Selection Techniques for Discrete Class Data Mining", IEEE Transactions on Knowledge and Data Engineering, Vol. 15, No. 3, May/June 2003.
- [6] András London, József Németh and Tamás Németh. "Time-dependent Network Algorithm for Ranking in Sports", Acta Cybernetica.
- [7] Carson K. Leung and Kyle W. Joseph / Procedia Computer Science 35 (2014) 710 – 719.
- [8] Eftim Zdravevski et.al. "System for prediction of the winner in a sports game" <https://www.researchgate.net/publication/226597761>.
- [9] Witten, I.H., Frank, E. "Data Mining: Practical machine learning tools and techniques (2nd Edition)". Morgan Kaufmann. (2005).
- [10] <https://ingsistemastelesup.files.wordpress.com/2017/03/zeror-algorithm.pdf>.

Authors Profile

Dr.C.Lalitha completed her doctorate degree in Vel's University, TamilNadu, India in 2015. Currently she is working as an Assistant Professor in Chevalier T.Thomas Elizabeth College for Women, Chennai. She has published more than 15 research papers in reputed international journals including Scopus and presented papers in both National & International conferences and it's also available online. Her main research work focuses on Data Mining, Spatial Mining, Cloud Computing, Fog Computing. She has 15 years of teaching experience.

Dr.S Arulselvarani has completed her doctorate degree under Mother Teresa Women's University, Kodaikannal, Tamil Nadu, India (2014). Currently she is working as Assistant Professor in STET Women's College, Mannargudi. She has published many National and International Journals and presented papers in various conferences. She has 14 years of teaching experience. She is a member in IEEE. Her area of interest includes E-Learning, Simulation, Data mining and Machine Learning.