



JOURNAL OF EMERGING TECHNOLOGIES AND INNOVATIVE RESEARCH (JETIR)

An International Scholarly Open Access, Peer-reviewed, Refereed Journal

EMPIRICAL STUDY OF PREDICTION MODEL IN SPORT DOMAIN

1 Dr. V.R. Dhawale, 2 K. P. Raghuvanshi, 3 T. A. Banait, 4 S. Khaire

¹Professor, K. K. Wagh Institute of Engineering Education and Research, Nashik, India

²Professor, Department of MCA, Vidyabharati Mahavidyalaya, Amravati, India

^{3,4}Student, Department of MCA, Vidyabharati Mahavidyalaya, Amravati, India

Abstract: Sports prediction is one of the recent growth areas of interest entailing good prediction accuracy. In this study, we have studied several predictive models for predicting the selection of a player. Coaches require models to assess their players, analyze opponent teams, and formulate winning strategies. We applied an approach based on data mining is proposed for result prediction in sports. It was observed that models were constructed based on the performance of teams in past matches, player performance indices, opposition team information, and external factors; therefore, relevant features are engineered to indicate the same. In many real-life sports games, spectators are interested in predicting the outcomes and watching the games to verify their predictions. In this paper, we present a sports data mining approach, which helps discover interesting knowledge and predict outcomes of sports games such as Football, Cricket, Basketball, and Hockey.

Keywords: Data Mining, Prediction Model, Data Analysis, Big Data.

I. INTRODUCTION

In this study, several predictive models for predicting the selection of a player was compared. Many existing sports data mining research projects focus on scheduling of games, visualization of games or players, sports advising, as well as identification and extraction of interesting moments or players from sports game video. While these research projects are interesting and practical, it is also interesting to predict the outcomes of games (e.g., predict the outcomes of cricket games or football games). For example, pre-game analysis of televised sporting events often includes expert predictions. Usually, the longer the historical data, the more accurate are the results. For example, pre-game analysis of televised sporting events often includes expert predictions. Data mining aims to discover implicit, previously unknown, and potentially useful information or knowledge from data.

II. DATA MINING TECHNIQUES

1. association

Association analysis is the finding of association rules showing attribute-value conditions that occur frequently together in a given set of data.

2. Classification

Classification is the process of finding a set of models that describe and distinguish data classes or concepts, to be able to use the model to predict the class of objects whose class label is unknown.

3.

Prediction

The prediction uses a combination of other data mining techniques such as trends, clustering, classification, etc. It analyzes past events or instances in the right sequence to predict a future event.

4. Clustering

Clustering is the task of dividing the data points into several groups such that data points in the same groups are more similar to other data points in the same group and dissimilar to the data points in other groups.

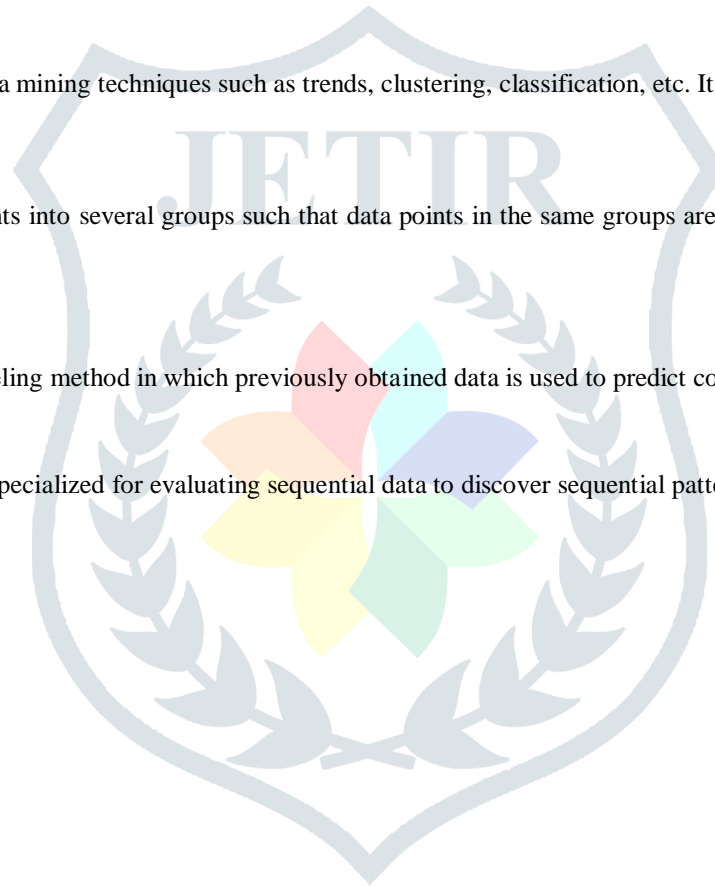
5. Regression

Regression can be defined as a statistical modeling method in which previously obtained data is used to predict continuous quantity for new observations.

6. Sequential Patterns

Sequential pattern is a data mining technique specialized for evaluating sequential data to discover sequential patterns.

III. LITERATURE REVIEW



Sr. No.	Prediction model	Authors	Techniques	Conference/Journal And year	Conclusion
1.	Hybrid Basketball Game Outcome Prediction Model by Integrating Data Mining Methods for the National Basketball Association	Wei-Jen Chen, Mao-Jhen Jhou, Tian-Shyug Lee and Chi-Jie Lu	Prediction model by integrated data mining techniques	MDPI entropy 17 April 2021	This research proposed a hybrid data-mining-based scheme for predicting the final score of an NBA game. The features was designed from original basketball statistics based on game-lag information. The proposed prediction scheme used five data mining algorithms: ELM, MARS, XGBoost, SGB, and KNN.
2.	Predicting Major League Baseball Championship Winners through Data Mining	Brandon Tolbert and Theodore Travis	Prediction model	Athens Journal of Sports December 2016	In this work, SVMs was applied to develop a model to predict championship winners and losers for Major League baseball.
3.	A review of Data Mining Techniques for Result Prediction in Sports	Hamid Rastegari , Maral Haghghat and Nasim Nourafza	Result prediction model/ classification techniques	ASIJ advance in computer science: an international journal November 2013	Considering the popularity of sports in the current world, many organizations disburse large funds to gain better results in sports matches. Therefore, predicting game results has turned into a subject of interest for different sports organizations. Data mining, a widely accepted method to predict and explain events, is an appropriate tool for this purpose. Various data mining techniques such as ANN, decision trees, Bayesian method, logistic regression, SVM, and fuzzy methods have been employed to predict game results in recent years.
4.	Apparent strain measurement of ankle joint in football sports based on data mining	Jinju zhu and meng zhang	Data mining algorithms Techniques	Hindawi, wireless communication mobile 24 March 2022	Firstly, the basic knowledge of imaging equipment technology and image processing technology is roughly sorted out, and then, aiming at the significance of noncontact measurement of flexion angle in ankle flexion movement, a rapid noncontact measurement system of flexion angle based on image technology is designed. Furthermore, using the method of artificial marking, a feature point motion trajectory extraction system based on image technology is designed, and the corresponding

					calculation method of body surface deformation and strain is proposed.
5.	One Day International (ODI) Cricket Match Prediction in Logistic Analysis: India VS. Pakistan	Shanjida Chowdhury, K. M. Anwarul Islam, MD. Mahfujur Rahman, Tahsin Sharmila Raisa, Nurul Mohammad Zayed	Regression Model	International Journal of Human Movement and Sports Sciences 29 December 2020	Conclusions Cricket is now a populous game in the South Asian Sub-continent. India and Pakistan are the most competitive two rivals of this 50-overs format. This study enables to find hindsight in winning matches against team India. Toss winning, venue, day or day-night format, and batting at first are considered independent variables towards the Indian winning strike. For measuring association, Cross tabulation along with Chi-square statistic is observed. Also, for multivariate analysis, binary logistic regression is performed. Day-night format and venue give a statistically significant result in winning the chance of Team India. For better checking out, The robustness was checked under Skewed logistic regression
6.	ICC T20 Cricket World Cup Prediction-Based Data Analytics and Data Mining Technique	Anubha Roy, Sakshi Pandey, Priyanka Bhatia and PROF. M. A Rane	supervised machine learning(SVM) classification, regression	Journal of Emerging Technologies and Innovative Research (JETIR) www.jetir.org May 2020	The outcome of this paper shows the problem of predicting the outcome of an ODI cricket match using the statistics of 366 matches with the help of the KNN classifier and Naive Bayes algorithm. It devise a method to find the cricket match outcome prediction, team structure analysis, and player recommendation system using the statistics of the players extracted from a particular tournament. It was observed that simple features can yield very promising results.
7.	Novel Method for Cricket Match Outcome Prediction using Data Mining Techniques	S.A.D.P Subasingha, S. C. Premaratne, K. L. Jayaratne, and P. Sellappan	Prediction model classification	Blue Eyes Intelligence Engineering & Sciences Publication	The main goal of this research was to learn a model for predicting game progression and outcome in ODI cricket matches. The prediction was separated into two segments. The first one is for predicting the winning possibility by considering the toss effect, ground condition, day-night effect, and opponent. Next,-the batting performances was considered by taking into players' combinations, wicket no, partnership runs total runs

					or target, and overs to predict the progression of the game. This helps to simulate the game properly. The naive Bayes algorithm is used to predict the match outcome in any given situation. the quality and accuracy was demonstrated by the predictions with an extensive set of experiments on real ODI cricket data. In addition to predicting winning chances for future segments, the winner prediction accuracy is indicated higher value because of calculating winning possibility before starting the match and while playing the match
8.	Applications of modern classification techniques to predict the outcome of ODI Cricket	Neeraj Pathak, Hardik Wadhwa	Classification, prediction model, and ML	Elsevier, Procedia Computer Science 87 (2016)	The motivation for using the Naive Bayesian classifier was based on the fact that it is the most suitable approach when the predictors are independent and is also known to perform decently even in the case of severe class imbalance in the dataset, which happened to be the case in our study.
9.	An Improved Prediction System for Football a Match Result	Igiri, Chinwe Peace; Nwachukwu, Enoch Okechukwu	Prediction model	IOSR Journal of Engineering (IOSRJEN) December 2014	Sports outcomes predictive techniques arise, which motivates the need to find more valuable datasets to improve the prediction accuracy and make precise decisions at key. Past comprehensive statistical data has been kept to assist English premier league games and other sporting events. Both players and teams present varying forms of these statistical facts kept as data season in and off-season. As the dataset set grows with the EPL games, it has become the preferred test platform. This pool of information will keep motivating different groups, ranging from the public, statisticians, and sports enthusiasts to discover embedded knowledge in it. This project also provides an ideal data mining environment, a data mart containing comprehensive EPL game information, which can be reused by future research.

10.	Analysis and Prediction of Football Statistics using Data Mining Techniques	Anurag Gangal ,Abhishek Talnikar ,Aneesh Dalvi Vidya Zope and Aadesh Kulkarni	Prediction model	International Journal of Computer Applications , December 2015	The project aims to attract more users to this game which is Fantasy Football and also aims at improving the general attraction to the Premier League. This happens because, in a predictive model, a user makes a prediction on every game, and ends up watching that game to check if his prediction is going right. Thus the project will not only improve the existing system of Fantasy Football but will also augment the reach of Football in India.
11.	Cricket score and winning prediction using data mining	Akhil Nimmagadda , Nidamanuri Venkata Kalyan , Manigandla Venkatesh , Nuthi Naga Sai Teja and Chavali Gopi Raju	Prediction model and regression	International Journal of Advance Research and Development (2018)	The main goal in this study is to develop a model to predict the outcome of an ODI cricket match while the game is in progress. The data from previous matches played between the team is to design the model. Multiple Variable Linear Regression is used to design this model. Efficiency and error checking is also done in our work. Using multiple linear regression, each innings score is predicted at regular intervals, and finally the winner of the match. This knowledge will help in the future to design a much more accurate prediction
12.	A Data Mining Approach to ODI Cricket Simulation and Prediction	Vignesh Veppur Sankaranarayanan, Junaed Sattar, and Laks V. S. Lakshmanan	Prediction model and regression	Society for industrial and applied mathematics (SIAM) July 2012	The main goal of this paper is to learn a model for predicting game progression and outcome in one-day cricket. Separate models for home runs and non-home runs are being developed using historical features as well as instantaneous match features from past games that are identified. Ridge Regression and attribute bagging algorithms are used on the features to incrementally predict the runs scored in the innings. The quality and accuracy of our predictions is being demonstrated with an extensive set of experiments on real ODI cricket data. In addition to predicting runs for future segments, winner prediction accuracy is by far the highest reported in ODI cricket mining literature

13.	Predicting Win-Loss outcomes in MLB regular season games – A comparative study using data mining methods	c.soto Valero	Prediction model, classification, and regression	International journal of computer science in sport 17 Dec 2016	This paper compares the performance of four different data mining methods in the context of predicting outcomes (win or loss) for independent MLB regular season games. First, the inherent difficulty of this particular prediction problem is being proved by showing and characterizing its complexity. To test predictive model, sabermetrics statistics is being used (to measure teams' performance and created a total of 30 datasets (one for each MLB team), corresponding to ten years of free available data (between 2005 and 2014 inclusively). Feature selection methods applied show that the most important predictor variable is the home field advantage. Four popular data mining methods were applied to reduced datasets (classification and regression-based) and were evaluated using the 10-fold cross-validation criterion. Overall, classification schemes outperform regression-based schemes and SVMs results are the best predictor method with accuracy values of nearly 60%. Despite results that were not surprisingly accurate, they became a good starting point for future works in these fields
14.	Expert System for Ice Hockey Game Prediction	Wei Gu Thomas L. Saaty and Rozann Whitaker	Data Mining with Human Judgment	International Journal of Information Technology and Decision Making June 2016	This paper describes an expert system to predict National Hockey League (NHL) game outcomes. A new method based on both data and judgments is used to estimate hockey game performance. Many facts and judgments could influence an outcome. The support vector machine is employed to determine the importance of these factors before incorporating them into the prediction system. System combines data and judgments and use them to predict the win-lose outcome of all 89 post-season games before they took place. The accuracy of prediction with the combined factors was 77.5%. This is to

					date the best accuracy reported of hockey games prediction.
15.	Application of Data Mining Technology in the Subject Tactical Teaching of Badminton	Delin Yang	Data Mining Technology	International Journal of Emerging Technologies in Learning (iJET) (2018)	This paper introduces the application of data mining technology in badminton spot tactics teaching in detail. First of all, the ACARMI algorithm is introduced. It is proved that the efficiency of the ACARMI algorithm is better than that of Reorder algorithm and Separate algorithm. Then, three functional modules, data collection, data preprocessing, and data mining, are divided according to needs. In the data collection module, the application program based on an Android tablet and the collection program based on the windows platform is designed and implemented, and the standardization and efficiency of data collection are improved. In the data preprocessing module, a data preprocessing model suitable for badminton statistics data is designed and implemented
16.	Improvement of table tennis technology based on data mining in the environment of wireless sensor networks	Hongjian Ma	Data mining techniques	International journal of distributed sensor networks (SAGE publication) October 2020	At present, the application of computer technology in sports mainly includes management systems, information release systems, technical and tactical statistical analysis systems, and so on. The application of sports calculation is very few, and the application of table tennis techniques and tactics is very rare. In this article, the application of video technology and data mining technology in table tennis matches under the wireless sensor networks environment is mainly studied based on the current research situation at home and abroad and the knowledge of related disciplines.

In my opinion, I will recommend 16th model proposed out of these models which is about the improvement of table tennis technology based on the data mining in the environment of wireless sensor networks.

At present, the application of computer technology in sports mainly includes management systems, information release systems, technical and tactical statistical analysis systems, and so on. The application of sports calculation is very few, and the application of table tennis techniques and tactics is very rare. In this article, the application of video technology and data mining technology in table tennis matches under the wireless sensor networks environment is mainly studied based on the current research situation at home and abroad and the knowledge of related disciplines.

IV. CONCLUSION

In this paper, we presented a study and review of various sports data mining approach to predict the winners of Football, Cricket, Basketball, and Hockey games. Instead of using the traditional approach of comparing the statistics of the two competing teams and projecting the outcome, the approach studied predicts the outcomes based on the historical results of games. The competing teams are compared to other teams, and game results are pulled from those similar teams and used in the prediction algorithm. The evaluation results show the accuracy of the sports data mining approach in predicting the outcomes of the games in recent seasons.

V. FUTURE

As we are approaching towards the future goals, we will analyze the data for the future of different types of games played during Olympic.

REFERENCE

- [1] Song, K.; Zou, Q.; Shi, J. Modelling the Scores and Performance Statistics of NBA Basketball Games. *Commun. Stat. Simul. Comput.* 2018, 49, 2604–2616.
- [2] Jordan, S.E.; Hovet, S.E.; Fung, I.C.H.; Liang, H.; Fu, K.W.; Tse, Z.T.H. Using Twitter for Public Health Surveillance from Monitoring and Prediction to Public Response. *Data* 2019, 4, 6.
- [3] Ben-Hur A, Weston J (2010) A User's Guide to Support Vector Machines. In O Carugo, F Eisenhaber (eds.), *Data Mining Techniques for the Life Sciences* (pp. 223-239).
- [4] S. B. Hong, "Effects of closed-sling exercise on muscle activity and balance; football club player with chronic ankle instability," *Journal of International Academy of Physical Therapy Research*, vol. 9, no. 4, pp. 1608–1613, 2018.
- [5] Y. S. Jeong and J. H. Kim, "Effects of whole body vibration exercise on lower extremity muscle activity and balance ability in a football player with chronic ankle instability," *The Journal of Korean Physical Therapy*, vol. 29, no. 6, pp. 293–298, 2017.
- [6] J. Peacock, "Maintaining a firm ankle: an effective coaching cue for improving football kicking?" *ISBS Proceedings Archive*, vol. 36, no. 1, pp. 98–98, 2018.
- [7] A classification-based tool to predict the outcome in ODI cricket, *Information and Automation for Sustainability (ICIAFs)*, 2010 5th International Conference on 17-19 Dec. 2010, pp. 250-255
- [8] I. Bhandari, E. Colet, and J. Parker, "Advanced Scout: Data mining and knowledge discovery in NBA data. *Data Mining and Knowledge Discovery*", vol. 1, no. 1, pp. 121-125, 1997.
- [9] G. Gartheeban and J. Guttag, "A data-driven method for in-game decision making in MLB: when to pull a starting pitcher", *In Proceedings of the 19th ACM SIGKDD international conference on Knowledge discovery and data mining*, pp. 973-979, 2013.
- [10] J. Surowiecki and P. Silverman, "The wisdom of crowds", *American Journal of Physics*, vol. 75, no. 2, pp. 190-192, 2005.
- [11] S. Kampakis and W. Thomas, "Using Machine Learning to Predict the Outcome of English County twenty over Cricket Matches", 2018.
- [12] Vignesh Veppur Sankaranarayanan, Junaed Sattar and Laks VS Lakshmanan, "Auto-play: A Data Mining Approach to ODI Cricket Simulation and Prediction", *SDM*, 2014.
- [13] A. Tumasjan, T. O. Sprenger, P. G. Sandner and I. M. Welpe, "Predicting elections with Twitter: What 140 characters reveal about political sentiment", *Proceedings of the Fourth International AAAI Conference on Weblogs and Social Media*, pp. 178-185, 2010.

- [14] M. Bailey and S. R. Clarke, "Predicting the match outcome in one-day international cricket matches while the game is in progress", *Journal of Sports Science and Medicine*, vol. 5, no. 4, pp. 480-487, 2006.
- [15] T. Singh, V. Singla and P. Bhatia, "Score and winning prediction in cricket through data mining", *International Conference on Soft Computing Techniques and Implementations (ICSCTI)*, pp. 60-66, 2015.
- [16] R. D. Brooks, R. W. Faff and D. Sokulsky, "An ordered response model of test cricket performance", *Applied Economics*, vol. 34, no. 18, pp. 2353-2365, 2002.
- [17] I. Bhandari, E. Colet, and J. Parker, "Advanced Scout: Data mining and knowledge discovery in NBA data", *Data Mining and Knowledge Discovery*, vol. 1, no. 1, pp. 121-125, 1997.
- [18] Zuccolotto P., Manisera M., Sandri M. Big Data Analytics for Modeling Scoring Probability in Basketball: The Effect of Shooting under High-Pressure Conditions. *Int. J. Sports Sci. Coach.* 2018

