



Match Outcome Prediction System Using Machine Learning Algorithm

Syed Kaif Ulla¹, Ms. Jeshma Nishita Dsouza²

^{1,2}St Joseph's College (Autonomous) Bangalore University

ABSTRACT:

The development of statistical modelling in sports has made it clear that accurately predicting a game's outcome is a basic challenge. India has a sizable fan base for cricket. With a lot of backing from the crowd, many fans attempt to anticipate the results of games using their own cricket intuition. There are certain rules and a scoring system in the game. The location of the game and a player's performance, for example, can greatly affect how the match turns out. Such a wide range of characteristics have a strong interdependence, which makes it difficult to anticipate matches with accuracy. In this project, we'll create a prediction system that uses information from previously played matches to forecast future match outcomes, such as the final score and whether a win or loss will occur. By analysing pre-stored match data using a variety of machine learning methods, our system will be able to predict the outcome of a match. In terms of the match venue, we plan to employ more features like pitch quality, weather condition, toss result, and individual player performance. Finally, our system displays quantitative findings using the most accurate and best-suited methodology. Moreover, illustrating the effectiveness of our algorithms in forecasting the number of runs scored, one of the most crucial match-related parameters. According to this study, a distinguishing characteristic for picking the winner is the relative team strength between the contending teams. The core of the suggested approach, which is to model individual players' bowling and batting performances, is to represent team strength. A player's recent performances and career statistics are used to model this. To anticipate a game's outcome, player-independent elements have also been taken into account.

Introduction

Cricket is the second-most popular sport in the world after football, but it is cricket that is most adored in India. Numerous research papers have been published in recent years, and a great deal of work has been done to predict the outcome of a cricket match using the variables that affect the outcome of the match. They use supervised machine learning algorithms to predict the outcome of the match, such as Linear regression, Support vector machines, Logistic regression, Decision Tree, Bayes Network, and Random Forest. Cricket is a fast-paced version of the game that draws spectators to the stadium and watchers at home, the Twenty20 format in particular is immensely popular. The Board of Control for Cricket in India oversees the professional Twenty20 cricket league known as the Indian Premier League (IPL) (BCCI). Each year, the Indian Premier League is held, with each team representing an Indian city. A massive industry like fantasy 11 and betting websites has given the model a lot of importance in addition to the excitement created by the media and several natural variables that have an impact on the game. The ability to accurately anticipate the outcome of a cricket match is greatly influenced by the game's regulations, the players' ability, their physical condition, and numerous other natural elements. People will use the predictions made by the machine learning algorithm as technology advances and apps like Fantasy 11 and betting sites become more popular. In many ways, using machine learning makes life easier. We won't rely on just one machine learning algorithm to forecast the result of a cricket match; rather, we'll use them all. Unsupervised learning and supervised learning are the two types of learning used in machine learning. When using unsupervised learning, the computer must sort the data based on patterns and combinations without any prior training. This is because the data is not properly labelled. In contrast, the data in supervised learning is labelled with the appropriate categorization so that the computer can quickly analyse it and come up with the desired outcome. Due to the well labelled cricket match data, unsupervised learning models are useless for our purpose. So, we'll employ supervised learning models. Regression and classification are the two categories used in supervised learning. Regression is used when the result is a real quantity, such as rupees or height, and classification is used to classify between categories, such as red or blue. It is of the number type. Regression will be used in our model because the winning percentage will be the outcome and numbers represent the outcome's type. Finding the critical elements that influence match outcome is our main goal, and we'll do this by choosing the machine learning model that best matches the data and produces the greatest outcomes. In the field of forecasting the outcome of a cricket match, certain works have already been published. In several papers, the accuracy suffers because only a few significant criteria are used to make predictions. The machine learning model is inappropriate in some papers, though. In order to select the appropriate model for training and testing the data, it is crucial to consider all the important variables that may have a bearing on the match's outcome. The prediction accuracy will dramatically increase as a result.

Literature review

[1] In this paper, they have taken the data of matches from the official website of the Indian Premier League. Because there were several aspects in the data, they examined it and chose a few important ones. They have used several selection models and pre-processed the data using the scikit-learn machine

learning framework. They eliminated the recursive, univariate, and low variance properties. These feature selection models allowed them to choose 5 critical features out of a total of 15. The characteristics are the host team, visiting team, location, toss winner, decision, and winner. They'd taught the Random.10 seasons of data from forests and multiple linear regression IPL data and 11th season training data. 41 out of 60 accurately identify according to their model. Thus, their accuracy not very good, at 68.33%. This model's drawbacks include its reliance on just 5 features and 2 machine learning models are used.[2] As the game goes on, this study uses dynamic data processing to provide a prediction. Prior to the game, their prediction is based on the batsman-performance bowler's during the match as well as the batting order and bowling order of a specific player. Following the game, their prediction is based on these same factors as well as the captains of both teams and the batting-bowling statistics on that surface against that opponent. By utilising 12 logistic regression and K-means clustering, they are only able to predict one-day international (ODI) matches. In this research, they only tested this model on the India vs. Australia match that took place on March 26, 2015, at Sydney Cricket Ground, and their forecast was accurate. Because they fluctuate with each ball, occasionally their forecasts are inaccurate.[3] Using data mining techniques, forecasted IPL player ratings are based on the 2008–2019 datasets. The performance of a player is evaluated using player statistics, and his basic price is established using data mining techniques. Using algorithms like decision tree, naive bayes, and multilayer perceptron, they made predictions about how to choose a player in the IPL based on each player's performance history (MLP). MLP performs more effectively than other algorithms. [5] This paper merely briefly discusses the crucial variables on which a cricket match depends. No machine learning model has been trained to forecast the outcome of a match. This essay discusses the pitch, toss, team strength, previous results, home field advantage, present performance, and weather. Because the pitch completely determines how the ball will perform, it is crucial to the contest. As the target is known and dew enters play after dusk, toss is also crucial because teams chasing first win more games. We can quickly determine the team's strength by averaging all players and looking at each player's current form.[6] employed the accuracy of 95.96%, 97.97%, and 98.98% of the Support Vector Machine (SVM), CTree, and Naive Baiyes classifiers to estimate the likelihood of the matches' winners.[7] By utilising the Nave Bayes classifier, the match result was predicted 14 using the home field, the time of the match, the kind of match, winning the toss, and then batting first.[8] projected the importance of the characteristics of the bowlers and batsmen in the current match. By using prior performances of a player against a certain opposing team and Multiple Random Forest Regression, this would aid in choosing the players for the next matches.[9] Data mining techniques were utilised to predict the results of any cricket match, and solutions were supplied for the strategy used by other authors.[10] The main determinants on which a cricket match depends are just briefly discussed in this essay. To anticipate the match outcome, no machine learning model has been trained. The pitch, toss, and team strengths, past results, home field advantage, present performance, and weather are the factors covered in this essay. The pitch is crucial to the outcome of the game since it completely determines how the ball will behave. The teams that win the toss tend to win more games because they know what they are aiming for and can take advantage of the dew that forms after dark. We can quickly determine the team's strength by computing the average of each player and their present state. Predictions are greatly influenced by past results; how a team performed in person versus a specific opponent is crucial. When a team plays at home, the crowd serves as its 11th player, and the players are more comfortable with the playing environment. Predicting the winner also requires knowledge about the team and players' current form. The ball's swing and the outcome of the game are both impacted by the 15 weather. The model can be trained using these features to improve prediction accuracy.[11] anticipated the result of the game based on the toss and the location.[12] based on the number of wickets and runs, estimated how well each player would perform. Both types of issues are handled as classification problems, with the lists of runs and wickets being categorised into various ranges using machine learning methods. Random forest performs the best of all.

Application

Sports prediction primarily aims to boost team performance and the likelihood of winning a match. The worth of a victory can manifest in several ways, such as trickling down to the fans who fill the stadium seats, broadcast deals, items in fan shops, parking, concessions, sponsorships, enrollment, and retention. any online cricket show platform or betting app

MODEL ARCHITECTURE

The architecture of the model consists of many elements, including datasets, split data, training, testing, supervised learning models, and results.

DATASET

Collecting datasets from diverse sources is the initial phase in the model's architecture. The inputted data determines how the model behaves and responds. Our results or forecasts will be correct if the data is accurate and current. Consequently, we have acquired our dataset from Kaggle.com

Splitting the data

The dataset is divided into two groups in this step: a training group and a testing group. Using supervised learning methods, the machine learning algorithms are trained on the training data. After being trained, the model is put to the test using the methods, and the outcome is anticipated.

Training The Model

Training the model is the most important part. Using training data, the model is trained in this step to look for patterns and forecast outcomes. As a result, the model gains knowledge from the dataset to successfully complete the task.

Testing The Model

After training, the model's performance is evaluated. The model is tested using novel datasets to achieve this. The testing datasets are the unobserved datasets. As a result, the model is evaluated in this stage by giving testing data that isn't visible.

Supervised Learning

Supervised machine learning techniques come in a variety of forms, including Logistic Regression, Support Vector Machine, K-Nearest Neighbor, Random Forest classifier, and linear Regression, among others.

CONCLUSION

It is particularly difficult and takes very complicated techniques to predict the victor of a sport like cricket. However, this can be greatly simplified and made easier with the introduction of machine learning. The results of the Indian Premier League matches are influenced by a number of factors, which have been identified in this study. The teams participating, the location, the city, the winner of the toss, and the decision to toss are all important factors that affect how an IPL match turns out. We examined IPL data sets and made predictions about game outcomes based on player performance. Logistic regression, Support Vector Machine (SVM), Decision Tree, Random Forest classifier, and K-Nearest Neighbor are some of the techniques employed in the study to arrive at the final test. The other algorithm is outperformed by Random Forest classification (RFC). Regarding the future, the season-long focus can be on each player's performance and its frequent evaluation. It is also possible to anticipate his bowling and batting rankings. The player who will win the game for each team can perhaps be predicted.

References

1. Rabindra Lamsal and Ayesha Choudhary, "Predicting Outcome of Indian Premier League (IPL) Matches Using Machine Learning", arXiv:1809.09813 [stat.AP] (September 2018)
2. Abhishek Naik, Shivance Pawar, Minakshee Naik, Sahil Mulani, "Winning Prediction Analysis in One-Day-International (ODI) Cricket Using Machine Learning Techniques", International Journal of Emerging Technology and Computer Science, Volume: 3 Issue: 2 (April 2018)
3. Arjun Singhvi, Ashish Shenoy, Shruthi Racha and Srinivas Tunuguntla. "Prediction of the outcome of a Twenty-20 Cricket Match." (2015).
4. Swetha, Saravanan.KN, "Analysis on Attributes Deciding Cricket Winning", International Research Journal of Engineering and Technology (IRJET), Volume: 04 Issue: 03 | (March 2017)
5. 5] Geddam Jaishankar Harshit, Rajkumar S, "A Review Paper on Cricket Predictions Using Various Machine Learning Algorithms and Comparisons among Them", International Journal for Research in Applied Science & Engineering Technology (IJRASET), IJRASET17099 (April 2018)
6. Akhil Nimmagadda, Nidamanuri Venkata Kalyan, Manigandla Venkatesh, Nuthi Naga Sai Teja, Chavali Gopi Raju, "Cricket score and winning prediction using data mining", International Journal of Advance Research and Development, Volume: 3 Issue: 3 (2018)
7. Raza Ul Mustafa, M. Saqib Nawaz, M. Ikram Ullah Lali, Tehseen Zia, Waqar Mehmood, "Predicting The Cricket Match Outcome Using Crowd Opinions On Social Networks: A Comparative Study Of Machine Learning Methods", Malaysian Journal of Computer Science, Volume: 30(1) (2017)
8. Muhammad Yasir, LI CHEN, Sabir Ali Shah, Khalid Akbar, M.Umer Sarwar, "Ongoing Match Prediction in T20 International", International Journal of Computer Science and Network Security, Volume: 17 Number: 11 (November 2017)
9. A.N.Wickramasinghe, Roshan D.Yapa, "Cricket Match Outcome Prediction Using Tweets and Prediction of the Man of the Match using Social Network Analysis: Case Study Using IPL Data", International Conference on Advances in ICT for Emerging Regions, ICTer: 442 (2018)
10. Ayush Kalla, Nihar Karle, Sushant Wagle, Sandeep Utala, "AutoPlay - Cricket Score Predictor", International Journal of Engineering Science and Computing, Volume: 8 Issue: 4 (April 2018)
11. Kaluarachchi, Amal, and S. Varde Aparna. "CricAI: A classification based tool to predict the outcome in ODI cricket." 2010 Fifth International Conference on Information and Automation for Sustainability. IEEE, 2010
12. Pedregosa, Fabian, et al. "Scikit-learn: Machine learning in Python." Journal of Machine Learning Research 12.Oct (2011): 2825-2830.
13. Sankaranarayanan, Vignesh Veppur, Junaed Sattar, and Laks VS Lakshmanan. "Auto-play: A Data Mining Approach to ODI Cricket Simulation and Prediction." SDM. 2014
14. Haseeb Ahmad, Ali Daud, Licheng Wang, Haibo Hong, Hussain Dawood and Yixian Yang, Prediction of Rising Stars in the Game of Cricket, IEEE Access, Volume 5, PP. 4104 – 4124, 14 March 2017.

15. Haryong Song, Vladimir Shin and Moongu Jeon, Mobile Node Localization Using Fusion Prediction-Based Interacting Multiple Model in Cricket Sensor Network, *IEEE Transactions on Industrial Electronics*, Volume: 59, Issue: 11, November 2012.
16. Sarbani Roy, Paramita Dey and Debajyoti Kundu, Social Network Analysis of Cricket Community Using a Composite Distributed Framework: From Implementation Viewpoint, *IEEE Transactions on Computational Social Systems*, Volume: 5, Issue: 1, PP. 64-81, March 2018.
17. Priyanka S, Vysali K, K B PriyaIyer, Score Prediction of Indian Premier League- IPL 2020 using Data Mining Algorithms, *International Journal for Research in Applied Science & Engineering Technology (IJRASET)*, Volume 8, Issue II, PP. 790-795.
18. Prince Kansal, Pankaj Kumar, Himanshu Arya, Aditya Methaila, Player valuation in Indian premier league auction using data mining technique, *International Conference on Contemporary Computing and Informatics (IC3I)*, 27-29 Nov 2014
19. Shilpi Agrawal, Suraj Pal Singh, Jayash Kumar Sharma, predicting results of IPL T-20 Match using Machine Learning, 2018 8th International Conference on Communication Systems and Network Technologies (CSNT), 24-26 Nov. 2018.
20. Harshit Barot, Arya Kothari, Pramod Bide, Bhavya Ahir, Romit Kankaria, Analysis and Prediction of Indian Premier League, 2020 International Conference for Emerging Technology (INCET), 5-7 June 2020.
21. Amal Kaluarachchi, S. Varde Aparna, CricAI: A classification-based tool to predict the outcome in ODI cricket, 2010 Fifth International Conference on Information and Automation for Sustainability, 17-19 Dec. 2010.
22. Kalpdrum Passi and Nirav Kumar Pandey, predicting player's performance in one day international cricket matches using machine learning, Volume 8, *Computer Science & Information Technology*, 2018.
23. Nigel Rodrigues, Nelson Sequeira, Stephen Rodrigues, Varsha Shrivastava, Cricket Squad Analysis Using Multiple Random Forest Regression, *IEEE Xplore*, 1st International Conference on Advances in Information Technology, 2019.
24. M. B. Wright, Scheduling fixtures for New Zealand Cricket, *IMA Journal of Management Mathematics* 16, PP. 99-112, 2005.
25. Manuka Maduranga, Hatharasinghe, Guhanathan Poravi, Data Mining and Machine Learning in Cricket Match Outcome Prediction: Missing Links, 5th International Conference for Convergence in Technology (I2CT), Mar 29- 31, 2019.
26. Monali Shetty, Sankalp Rane, Chaitanya Pandita, Suyash Salvi, Machine learning-based Atlantis Highlights in Computer Sciences, volume 4 40 Selection of Optimal sports Team based on the Players Performance, *Proceedings of the Fifth International Conference on Communication and Electronics Systems (ICCES 2020)*, IEEE Conference Record, ISBN: 978-1-7281-5371-1, 2020.
27. Balasundaram A, Ashokkumar S, Jayashree D, Magesh Kumar S, Data Mining based Classification of Players in Game of Cricket, proceedings of the International Conference on Smart Electronics and Communication (ICOSEC 2020), IEEE Xplore Part Number: CFP20V90- ART; ISBN: 978-1-7281-5461-9.
28. Jalaz Kumar, Rajeev Kumar, Pushpender Kumar, Outcome Prediction of ODI Cricket Matches Using Decision Trees and MLP Networks, 2018 First International Conference on Secure Cyber Computing and Communication (ICSCCC).
29. Prabu, S., Balamurugan Velan, F. V. Jayasudha, P. Visu, and K. Janarthanan. "Mobile technologies for contact tracing and prevention of COVID-19 positive cases: a cross-sectional study." *International Journal of Pervasive Computing and Communications* (2020).
30. Subramani, Prabu, K. Srinivas, R. Sujatha, and B. D. Parameshachari. "Prediction of muscular paralysis disease based on hybrid feature extraction with machine learning technique for COVID-19 and post-COVID-19 patients." *Personal and Ubiquitous Computing* (2021): 1-14