

Research Proposal

Predictive Machine Learning Model for Optimal Player Selection in Fantasy Cricket: Maximizing Points and Winning Fantasy Premier League.

Table of Contents

1. Abstract.....	1
2. Introduction.....	1
3. Research Objectives.....	2
4. Challenges.....	2-3
5. Literature Review.....	3-4
6. Dataset.....	4
7. Methodology.....	5-6
8. Expected Results.....	6
9. Time Management.....	7
10. Conclusion.....	7
11. References.....	8-9

Abstract

As a forum for cricket fans to choose virtual teams and compete for points based on the achievements of real-life cricket players, the Fantasy Premier League (FPL) has become incredibly popular. The main goal for participants is to earn maximum fantasy league points and win the league. In order to increase the likelihood of winning the FPL, this research proposal focuses on creating a predictive machine learning model that analyses previous performance data to choose the top players for fantasy cricket teams. The suggested approach seeks to anticipate player performances and improve participant decision-making via the use of sophisticated machine learning algorithms and statistical methodologies. Participants will have access to a potent tool through this research that enables them to choose players more likely to put up extraordinary performances and make data-driven selections. The findings of this study might change fantasy cricket tactics and advance the area of sports analytics.

Introduction

With the ability to build virtual teams and play against other participants in fantasy leagues, fantasy cricket has completely changed how cricket devotees interact with the sport. The finest player combinations that are likely to perform well in actual contests are crucial to fantasy cricket success. This work is challenging, though, because it necessitates a thorough understanding of player skills, game circumstances, and several other elements that affect player performances. The development of predictive models that help participants choose players wisely has been more feasible in recent years thanks to the rise of machine learning and data analytics. These models can find patterns and trends that lead to good fantasy cricket performances by examining past performance data. To create models that reliably predict player performances and optimise team selection to maximise fantasy league points, machine learning algorithms and statistical approaches can be used. The objective of this research project is to create a predictive machine learning model that has been specially designed for fantasy cricket. A wide range of player traits and match circumstances will be taken into account by the model, which will use past player performance data. The model will be trained to recognise the critical elements that have a major influence on player performances by utilising a variety of machine learning algorithms and statistical methodologies. Participants will be able to choose the best players for their fantasy teams as a result, increasing their chances of maximising fantasy league points and ultimately winning the Fantasy Premier League. According to the research's findings participants will have access to a crucial tool that will help them make better decisions and get a competitive edge in the Fantasy Premier League. Participants may optimise player selection, make data-driven decisions, and raise the possibility that their chosen players will put on extraordinary performances by utilising the power of predictive modelling. This study contributes to the developing field of sports analytics as well as the possibility of changing fantasy cricket strategies. In the following sections of this research proposal, we will outline the objectives, literature review, methodology, expected results, timeline and conclusion for this study.

Research Objectives

The primary objectives of the research proposal are as follows:

- Comprehend the main elements that affect player performance in fantasy cricket and how they affect maximising the number of points in a fantasy league.
- Examine a range of statistical methods and machine learning algorithms to create a predictive model that can predict player performance with accuracy.
- On the basis of predicted player performances, optimise team selection tactics.
- To increase the likelihood of winning the Fantasy Premier League (FPL) and to maximise fantasy league points by utilizing the developed predictive models and selection strategies.
- To Compare forecasts with real fantasy league outcomes to assess how well the predictive model performs in comparison to other traditional methods of selecting players for fantasy cricket teams.
- To provide players in fantasy cricket with practical recommendations and guidelines on how to use the predictive model in order to improve their team selection process and raise their chances of winning the Fantasy Premier League.
- To contribute to the existing knowledge in the field of fantasy cricket by bringing a fresh strategy that blends machine learning methods, statistical analysis, and player performance projection.

Challenges

It can be difficult to accomplish the research goals mentioned in the proposal. Included in these possible difficulties are:

- **Data Quality and Availability:** It might be difficult to gather accurate information about player performance and other important variables. To create a powerful prediction model, it is essential to ensure the data's correctness and completeness.
- **Feature Selection and Dimensionality:** Selecting the most pertinent characteristics from the given data might be difficult. Careful feature selection and dimensionality reduction approaches could be necessary when dealing with high-dimensional data and probable collinearity problems among variables.
- **Interpretability and Model Complexity:** It's crucial to strike a balance between these two factors. While delivering insights into the factors affecting player performances, the model has to deliver precise forecasts.
- **Generalization and Transferability:** The built prediction model must be able to generalise well beyond the training data in order to be considered transferable. It may be difficult to consider the model's adaptability to various cricket leagues or competitions.
- **Overfitting and Model Performance:** To make sure a model is reliable, it is crucial to avoid overfitting and evaluate the model's performance on unseen data. Appropriate

approaches, such as cross-validation and regularisation procedures should be employed.

- **Ethical Guidelines:** It's essential to follow ethical principles and protect players' anonymity and privacy. To prevent any moral consequences or biases, it is crucial to maintain transparency in the model creation and use .
- **Scalability and Computational Resources:** To manage the computational demands of developing and training machine learning models, particularly with huge datasets or sophisticated algorithms, it is necessary to have sufficient computational resources and efficient algorithms.

The effective completion of the study objectives will depend on addressing these difficulties. These obstacles can be addressed and the validity and reliability of the study findings ensured by careful planning, reliable methodologies, and a methodical approach to data collection, feature selection, model development, and assessment.

Literature Review

- Analytics Vidhya 2023, 'The Science of T20 Cricket: Decoding Player Performance with Predictive Modeling', Analytics Vidhya, viewed [1/06/2023], <
<https://www.analyticsvidhya.com/blog/2023/06/the-science-of-t20-cricket-decoding-player-performance-with-predictive-modeling/#h-model-buildingL>>.

In order predict player performance in T20 cricket matches, the article covers the use of data analytics and machine learning. The project entails gathering and analysing data on player and team performance, creating a predictive model using ridge regression, and then using the model to forecast player performance in a future IPL match. IPL team managers, coaches, and fantasy league players will all have access to a platform that will help them make data-driven player selection and game strategy decisions. The processes for implementing the code and gathering the data are covered in the paper, along with the prediction model's constraints and application scenarios. For clubs, coaches, and other stakeholders, the model can offer insightful information on player performance and support strategic choice-making. The model's effectiveness, nevertheless, is highly dependent on the calibre, volume, and applicability of previous data and it's possible that it doesn't properly account for contextual elements that might affect match outcomes. The model should only be used as a supplementary tool, thus it's critical to recognise its limits.

- Sachin Kumar, S., Prithvi, H. V., & Nandini, C. (2021). Data Science Approach to Predict the Winning Fantasy Cricket Team—Dream 11 Fantasy Sports. International Journal of Advanced Science and Technology, 30(3), 1069-1074.
https://www.researchgate.net/publication/363584809_Data_Science_Approach_to_predict_the_winning_Fantasy_Cricket_Team_Dream_11_Fantasy_Sports

This research paper explores numerous techniques used by authors to forecast results and choose teams, giving readers an overview of sports analytics in baseball and cricket. It draws attention to Michael Lewis' book "Moneyball: The Art of Winning an Unfair Game," which centres on Billy Beane, the Oakland Athletics' general manager, and his record-breaking 20

straight victories done on a shoestring budget. The article also covers Sabermetrics, a method of analysis that gives players with lower costs and greater on-base percentages and slugging percentages priority. To find the best regressor model for their particular issue, the research paper used the PyCaret module. They created a pipeline that combines user input, database queries, data processing, model application, and the delivery of average predictions to the user. The research also makes use of a number of sites that cover various methods and tools for analysing cricket data, including as network analysis, machine learning strategies, and the Yorkpy Python module. To categorise the performance of batsmen and bowlers in One Day International (ODI) matches into five performance categories (1 to 5), the researchers applied a number of machine learning methods, including Naive Bayes, Random Forest, Multiclass SVM, and Decision Trees. With accuracy rates of 90% for batsmen's performance and 92% for bowlers' performance, the Random Forests Classifier Model was shown to produce superior predictions compared to the other classification models. These results demonstrate the promise of machine learning techniques for improving cricket player performance evaluation and forecast accuracy.

- Mr. MUJAMIL DAKHANI , UMME HABIBA MAGINMANI(2020). PREDICTING ACCURACY OF PLAYERS IN THE CRICKET USING MACHINE LEARNING. International Research Journal of Engineering and Technology (IRJET).

<https://www.irjet.net/archives/V7/i5/IRJET-V7I51265.pdf>

In order to predict the performance of bowlers and batsmen in One Day International matches, the research article emphasises the importance of player selection in cricket and suggests a prediction model that makes use of machine learning techniques. Data from websites like ESPNcricinfo and Cricbuzz are gathered by the system using web scraping technologies, which are then used by PHP to process and store the information in MySQL databases. To forecast the accuracy of batsmen and bowlers, the model uses a variety of machine learning methods, including as Naive Bayes, Decision Trees, Random Forest, and Support Vector Machine. Notably, the Random Forest algorithm is shown to be the best at producing precise forecasts. The article also lists particular uses for these algorithms, such predicting basketball game results and looking at 2003 ICC World Cup data.

Dataset

I have decided to start working on my project with these datasets

<https://www.kaggle.com/datasets/vora1011/ipl-2022-player-statistics>

<https://www.kaggle.com/datasets/akarshsinghh/cricket-player-performance-prediction?select=ball.csv>

But if I face any difficulties I can look for other datasets from my primary and secondary research papers or can collect data and build my dataset from cricsheet

<https://cricsheet.org/> and <https://www.espnecricinfo.com/>

I am also exploring datasets available here

<https://www.kaggle.com/search?q=fantasy+cricket>

Methodology

In order to construct and assess the predictive machine learning model for the best player selection in fantasy cricket, the approach for this study will entail a number of crucial elements. Following elements will be included in the methodology:

1. **Data Preprocessing:** After obtaining the data, I will put it through preprocessing to clean it up and convert it into a format that is appropriate for analysis. Managing missing values, spotting outliers, choosing features, and data normalisation may all be part of this process. In order to prepare the data for further analysis, it must be verified as correct, comprehensive, and ready.
2. **Feature Engineering:** In this stage, I will use statistical methods and my domain expertise to develop new features or change ones that already exist to gather key player performance data. To improve the prediction capability of the model, this may entail combining statistics, developing derived variables, or adding outside data sources.
3. **Model Development:** Based on the methods employed by others in my primary and secondary research articles, which I described in the literature review, I will investigate several machine learning algorithms and statistical methodologies suited for predictive modelling in fantasy cricket. This can include deep learning techniques like neural networks, ensemble methods like random forests, Naïve Bayes, SVM as well as regression models like ridge regression. To determine the most efficient method for forecasting player performances, many models will be developed and assessed.
4. **Tuning of the hyperparameters:** I will adjust the hyperparameters to enhance the performance of the prediction models. To identify the ideal set of hyperparameters for each model, I'm looking forward to using methods like grid search, random search, or Bayesian optimisation.
5. **Model evaluation:** Using relevant assessment measures including accuracy, precision, recall, and F1 score, I will assess the models that have been built. To evaluate how well the models generalise, I shall use cross-validation techniques. Additionally, the model's predictions will be compared to actual player performances from previous games or seasons as part of the grading process.
6. **Ethics:** Ethical principles will be upheld at all times during the study process. All data processing procedures will adhere to legal and ethical requirements, and the privacy and confidentiality of personal data will be protected. To guarantee fairness and openness, any possible biases in the data or models will be thoroughly evaluated and corrected.
7. **Model Deployment:** I will make the model ready for deployment in actual fantasy cricket settings once it has been trained and assessed. To help players choose the best players based on statistics, the model's predictions may be included into current fantasy cricket systems or utilised independently.
8. **Testing and Validation:** In order to evaluate the performance of the deployed model in real-world settings, I will conduct testing and validation on it. To further hone and

enhance the model's predictions and usefulness, feedback from users and participants will be gathered.

This technique will help us create a strong and trustworthy predictive machine learning model that will help fantasy cricket players choose the best players, maximise their fantasy league points, and eventually increase their chances of winning the Fantasy Premier League.

Expected Results

Following is a brief summary of the expected outcomes of this study:

- **Development of a predictive machine learning model:** The main expected outcome is the effective creation of a predictive machine learning model that can predict player performances in fantasy cricket. For the purpose of making the best player selection decisions, the model will make use of historical data, player characteristics, match circumstances, and other pertinent elements.
- **Improved accuracy in player performance predictions:** A considerable improvement in player performance prediction accuracy is expected as a consequence of the use of cutting-edge machine learning algorithms and statistical methodologies. The proposed model ought to perform better than established techniques and baseline strategies, offering more trustworthy perceptions into the expected effectiveness of Fantasy Premier League participants.
- **Optimization of fantasy team selection:** The predictive algorithm recommends the optimal player combination that will increase the number of fantasy league points in order to optimise the choice of a fantasy squad. With consideration for elements including player performance, match circumstances, and team compositions, the intended outcome is to provide participants exact suggestions on which players to choose for their teams.
- **Validation of model performance through experimentation:** Experiments are used to verify the prediction model's efficacy. When participants use the player picks suggested by the algorithm, it is expected that their fantasy league performance would significantly increase. The effectiveness and dependability of the created model will be shown by this validation.
- **Insights and suggestions for participants:** Another expected outcome is to offer helpful participant insights and suggestions for those who play fantasy cricket. These insights may consist of identifying crucial variables that affect player performance, showcasing undervalued players or up-and-coming talent, offering sensible approaches to team composition, and giving timely updates on player injuries or other news that can have an impact on the choice of a fantasy squad.

Overall, this research is expected to produce a highly precise predictive machine learning model for the best player selection in fantasy cricket, improved performance in the Fantasy Premier League for users of the model, and insightful information that will help users make better fantasy cricket decisions.

Time Management

I'll start working on my dissertation assignment on June 13 and have it finished by August 18 as planned. I will have a total of 9 weeks, 3 days to finish the dissertation in this manner. This amount of time, in my opinion, will be enough to successfully complete the essential duties. I have meticulously created a detailed schedule that divides the job into manageable parts and corresponds with the numerous phases and procedures involved in the approaches I want to use in order to ensure efficient use of my time. I have scheduled routine meetings with my supervisor in addition to my self-imposed timetable. Every Wednesday, I will be able to attend these sessions, which will enable me to get timely advice and criticism all during the project. I want to handle any possible issues or obstacles as soon as they arise by asking for my supervisor's guidance early on, avoiding last-minute scrambles that can obstruct progress. I will stay on track and make steady progress towards finishing my dissertation within the allotted term thanks to the planned strategy I've taken and constant communication and feedback from my supervisor. I am confident in my capacity to effectively complete my dissertation within the allocated time period while preserving the highest professional standards thanks to rigorous preparation, a proactive approach, and persistent devotion to academic integrity.

Conclusion

In conclusion, there is a lot of potential for improving the whole fantasy gaming experience with the development of a predictive machine learning model for the best player selection in fantasy cricket with the goal of maximising points and winning the Fantasy Premier League. This research project aims to revolutionise the way fantasy cricket teams are put up by a thorough investigation of machine learning algorithms, statistical methods, and historical player performance data. Our goal is to give fans of fantasy cricket an intelligent tool that can properly estimate player performances and direct them in making decisions when choosing their fantasy teams. Through the use of predictive modelling, we do this. We can identify significant factors, such as batting averages, strike rates, bowling economy rates, and fielding statistics, among others, that have a significant influence on player performance by using machine learning algorithms. The efficiency of the suggested model will be assessed by the teams it can choose and the points it can score in fantasy leagues. We anticipate that our predictive machine learning algorithm will perform better than conventional manual ways after thorough review and testing, which will improve team performance and raise the likelihood that they will win the Fantasy Premier League. Additionally, the goal of this research endeavour is to add to the body of information already known about data science and fantasy cricket. Researchers, sports analysts, and business professionals looking for fresh ways to player selection and performance forecasting may find the facts and insights generated from this study useful in addition to fantasy cricket fans. In conclusion, the creation of an optimum player selection predictive machine learning model for fantasy cricket is a significant progress in the area. It might transform the way that fantasy gamers approach their games and provide them an advantage over one another in the Fantasy Premier League. Our study seeks to improve the entire fantasy cricket experience and increase the pleasure by maximising points and facilitating the selection of the top players.

References

- Analytics Vidhya 2023, 'The Science of T20 Cricket: Decoding Player Performance with Predictive Modeling', Analytics Vidhya, viewed [1/06/2023], <
<https://www.analyticsvidhya.com/blog/2023/06/the-science-of-t20-cricket-decoding-player-performance-with-predictive-modeling/#h-model-buildingL>>.
- Sachin Kumar, S., Prithvi, H. V., & Nandini, C. (2021). Data Science Approach to Predict the Winning Fantasy Cricket Team—Dream 11 Fantasy Sports. *International Journal of Advanced Science and Technology*, 30(3), 1069-1074.
https://www.researchgate.net/publication/363584809_Data_Science_Approach_to_predict_the_winning_Fantasy_Cricket_Team_Dream_11_Fantasy_Sports
- Mr. MUJAMIL DAKHANI , UMME HABIBA MAGINMANI(2020). PREDICTING ACCURACY OF PLAYERS IN THE CRICKET USING MACHINE LEARNING. *International Research Journal of Engineering and Technology (IRJET)*.
<https://www.irjet.net/archives/V7/i5/IRJET-V7I51265.pdf>
- Kataara, K., Krishnan, G. S., Shetty, S., & Vanahalli, M. K. (2020). Analysis and Prediction of Fantasy Cricket Contest Winners Using Machine Learning Techniques. In *Evolution in Computational Intelligence, Frontiers in Intelligent Computing: Theory and Applications (FICTA 2020)*, Volume 1.
https://www.researchgate.net/publication/344265243_Analysis_and_Prediction_of_Fantasy_Cricket_Contest_Winners_Using_Machine_Learning_Techniques
- Behera, A. (2021). A Scientific Method to Select Your Fantasy Sports Team: Using machine learning and optimization techniques to select your team on the fantasy sports platform — Dream11.com. Analytics Vidhya.
<https://medium.com/analytics-vidhya/a-scientific-method-to-select-your-fantasy-sports-team-b23726136256>
- Prakash, C. D., & Verma, S. (2022). A new in-form and role-based Deep Player Performance Index for player evaluation in T20 Cricket. *Decision Analytics Journal*, 2, 100025.
<https://www.sciencedirect.com/science/article/pii/S2772662222000029#d1e3710>
- Ahmed, F., Deb, K., Jindal, A.: Multi-objective optimization and decision making approaches to cricket team selection. *Appl. Soft Comput.* 13(1), 402–414 (2013).
- Amin, G.R., Sharma, S.K.: Cricket team selection using data envelopment analysis. *Eur. J.Sport Sci.* 14(sup1), S369–S376 (2014)
- Bhattacharjee, D., Saikia, H.: An objective approach of balanced cricket team selection using binary integer programming method. *Opsearch* 53(2), 225–247 (2016)
- Burney, S.M.A., Mahmood, N., Rizwan, K., Amjad, U.: A generic approach for team selection in multi-player games using genetic algorithm. *Int. J. Comput. Appl.* 40(17), 11–17 (2012). Full text available
- Trawinski Krzysztof . “A fuzzy classification system for prediction of the results of the basketball game.” *Fuzzy system(FUZZ)* 2010. IEEE International conference on IEEE 2010s

- Swetank Pathak.(2021).Machine Learning Approach In Fantasy Sports: Cricket.Analytics India Magazine
<https://analyticsindiamag.com/machine-learning-approach-in-fantasy-sports-cricket/>
- E. Nsolo, 'Prediction models for soccer sports analytics',Dissertation, 2018.
- Sports Performance, Measurement and Analytics: the Science Of Assessing Performance, Predicting Future Outcomes, Interpreting Statistical Models, and Evaluating the Market Value Of Athletes Lorena Martin- Pearson Education, Inc. – 2016
- S. Muthuswamy and S. S. Lam, "Bowler Performance Prediction for One-day International Cricket UsingNeural Networks," in Industrial Engineering Research Conference, 2008.

Excluding References word count = 3173.