Preprints.org

Review

Prediction for Pakistan Super League Using Machine Learning Techniques: A Survey

Noor ul Ain *, Humaira Ashraf *, NZ Jhanjhi

Review

Prediction for Pakistan Super League Using Machine Learning Techniques: A Survey

Noor ul Ain *, Humaira Ashraf and NZ Jhanjhi

School of Computer Science, SCS, Taylors University, Subang Jaya, Malaysia; humaira.ashraf@ iiu.edu.pk, noorzaman.jhanjhi@taylors.edu.my

*   Correspondence: noorulain.mscs1144@iiu.edu.pk

Abstract: Cricket, the most popular sports in the world to chase by the fans. Many cricket fans are interested in analyzing the elements and activities that affect the outcomes of matches. Machine learning in sports analytics is a relatively new topic in computer science. The purpose of this survey is to anticipate the outcome of a PSL champion team and build a winning strategy. Machine learning (ML) has exhibited favorable outcomes in a variety of sectors for diverse prediction using classification, regression, and so on, and is proven to be accurate. The innovative frameworks based on ML have the ability to learn from previous experiences. The cricket pitch is the most important aspect, alongside home game an edge, coin toss, innings of play, day/night match, physical conditioning, and dynamic plans, among other things. The dataset from previously ESPN Cricifno was used for this purpose. Recent classification approaches including, Random Forest, Gradient Boost, Deep Neural Networks used to conduct a comparative analysis based on their results and performances, as well as team strategies.

Keywords: prediction; machine learning; cricinfo

Introduction

Machine Learning is growing in popularity in statistical data analysis, particularly as an analysis method in cricket, it may be used to forecast both game results and player performance and other related factors. Decision-makers in cricket might greatly benefit from using machine learning as a prediction tool given the growing popularity of the sport and the growing income it generates. Numerous studies have used machine learning methods and other relevant statistical techniques for sports prediction and classification. Machine learning is utilized in predicting the consequence of a match.

There can be a lot of features such as country where the match is being played, cricket pitch, batting averages of each batsman, bowling averages of each bowler, win ratio of both teams against each other and in the specific ground and country.

It is true that the primary objective of machine learning is to provide computers the ability to learn autonomously and adjust their tasks based on the situation. Additionally, recent developments in computing have made it simpler to process and analyses large amounts of data, which has greatly facilitated the application of machine learning to sports analytics.
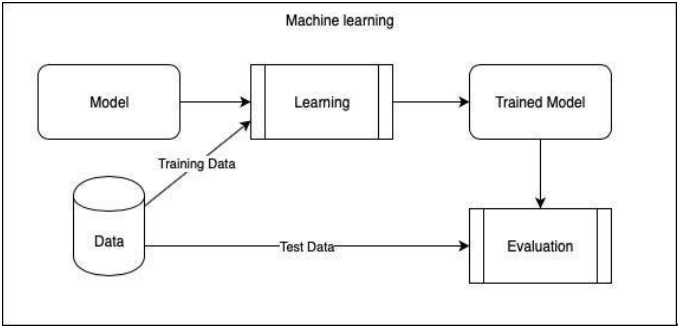
Figure 1. Block model of ML.

Sports analytics is the process of obtaining and examining past game data in order to draw out crucial information that can help with decision-making. Both on and off the pitch, machine learning can be used to predict a team's performance and its outcome against a rival team. Additionally, it can be used to concentrate on players' healthy development and to raise their level of competition, which is advantageous for team owners and other investors. Machine learning is being used by a football team from Portugal, to make decisions based on information that is currently available. In order to construct customized training schedules and improve game planning, the club tracks and estimates practically everything about its players, both on and off the field, including their relaxing, drinking, and training habits.

It is true that systematic reviews aid in the collection of empirical data regarding the ongoing development of research using ML in the cricket area. The author also points out a gap in the literature caused by the lack of a systematic review of the results of two decades of research integrating cricket and ML.

There have been several surveys conducted on the prediction of cricket matches using machine learning techniques. The Table 1 represents the brief overview of already conducted surveys.

Table 1. Previous survey analysis.

Year	The main focus of the Survey	Major contribution	Enhancement in current survey
2022	A Systematic review of Application of Machine Learning in cricket.	The paper most likely investigates and summarizes the many ways whereby algorithms for machine learning have been utilized in the field of cricket.	Current study provides a detailed critical assessment that reveals the flaws in present schemes.
2022	Best strategy to win a cricket match using a hybrid machine learning module	The fundamental contribution of the study is the development and launch of a hybrid network learning-clustering-associative rules model for cricket matches, with a focus on both types of crickets, one-day internationals (ODIs) and Twenty 20s (T20s).	Current study provides a detailed critical assessment focusing on modern day cricket and overcome the flaws present in current scheme.
2022	T20 cricket match prediction using machine learning	The fundamental contribution of this work is to the creation of a precise	Current study provides a detailed

		forecasting method to predict cricket scores in live matches, taking into consideration prior datasets and numerous aspects that influence score prediction.	critical assessment and the enhance the feature to predict better result.
2022	Tournament-level performance prediction and team selection for cricketers using machine learning.	The key objective of this research is to provide a computationally efficient approach for cricket team development.	Current study provides a detailed critical assessment and the optimized solution.
2021	An efficient team prediction using hybrid approach using machine learning algorithms	The goal is to build a large and diversified squad capable of playing effectively in a variety of settings, regardless of location or opponent.	The study propose a realistic and effective approach for squad creation by emphasizing computing efficiency.

For example, survey reviewed various machine learning algorithms used for predicting cricket match outcomes. The study found that ensemble methods such as Random Forest and Gradient Boosting had better performance than other algorithms such as Naïve Bayes, Logistic Regression, and Decision Trees. In another survey it is examined the use of machine learning for predicting cricket player performance. The study found that Random Forest, Decision Trees, and Naïve Bayes were the most commonly used algorithms for player performance prediction. In a survey the authors examined the use of machine learning for predicting cricket match outcomes. The study found that ensemble methods such as Random Forest and Gradient Boosting had better accuracy compared to other algorithms such as Naïve Bayes and Logistic Regression. Overall, these surveys suggest that ensemble methods such as Random Forest and Gradient Boosting have shown better performance in predicting cricket match outcomes and player performance.

However, it is important to note that the accuracy of predictions can vary depending on the quality and quantity of the data used for training the machine learning models.
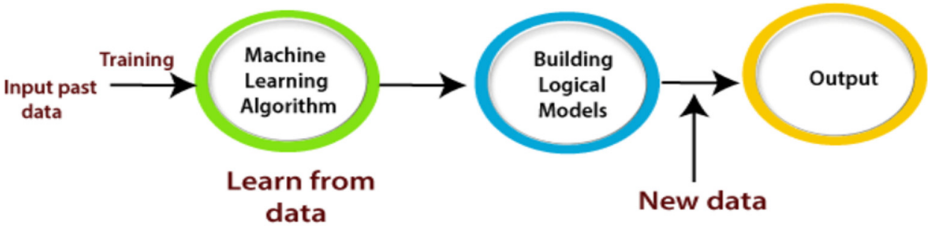
Figure 2. ML model framework.

The cricket domain has seen a lot of research on predicting match winners using machine learning techniques such as KNN, Logistic Regression, SVM, Naïve Bayes, and Random Forest. However, due to inconsistent and complicated datasets, achieving accurate predictions without limitations has been challenging. This study aims to fill the gap in the literature by providing a systematic review of two decades of research findings in combining cricket and ML. Previous studies have achieved varying levels of accuracy. Techniques used include statistical approaches, data of ICC

match ratings, ranking points for batsmen and bowlers, home factor, ICC rating differences, ground effects, player performance, and team strength. Machine learning algorithms used include Naïve Bayes, Random Forest, Decision Tree, and Logistic Regression. The findings of this study will be beneficial to players, coaches, and sports administrators.

The intention of this work is to fill this gap and provide insights that will be beneficial for players, coaches, and sports administrators. Systematic reviews are an essential tool for gathering and synthesizing research findings, particularly in fields where there is a lot of research. By systematically reviewing research on the use of ML in cricket, the study can identify trends, highlight gaps in knowledge, and provide recommendations for future research. The findings of this study will be helpful for players, coaches, and sports administrators who are interested in using ML to enhance cricket performance. For example, the study may identify specific ML techniques that are particularly effective in cricket or highlight areas where more research is needed. Overall, this study has the potential to contribute to the ongoing development of cricket and ML by providing a comprehensive review of existing research. In the current project, the dataset used contained information about the matches of PSL held from 2016 through 2023, and with the help of machine learning, the primary objective of the project was achieved. Machine learning can be very beneficial in sports analytics, providing insights and facilitating good decision-making for teams, managers, and investors.

The motivation behind this paper includes the answers to these questions:

"Which team of PSL will be the winner of season 9 by using the features like particular venue, based on decision to field/bat first, team combinations, winning the toss and pitch conditions? Which suggestions and strategies will help out a team to overcome its weakness and leads it towards the victory stand?"

In this paper we are trying to find out the winner of PSL 2024 and make few strategies to improve the team performance which will leads them towards the victory stands using machine learning techniques like SVM, Random Forest, Logistic Regression etc.
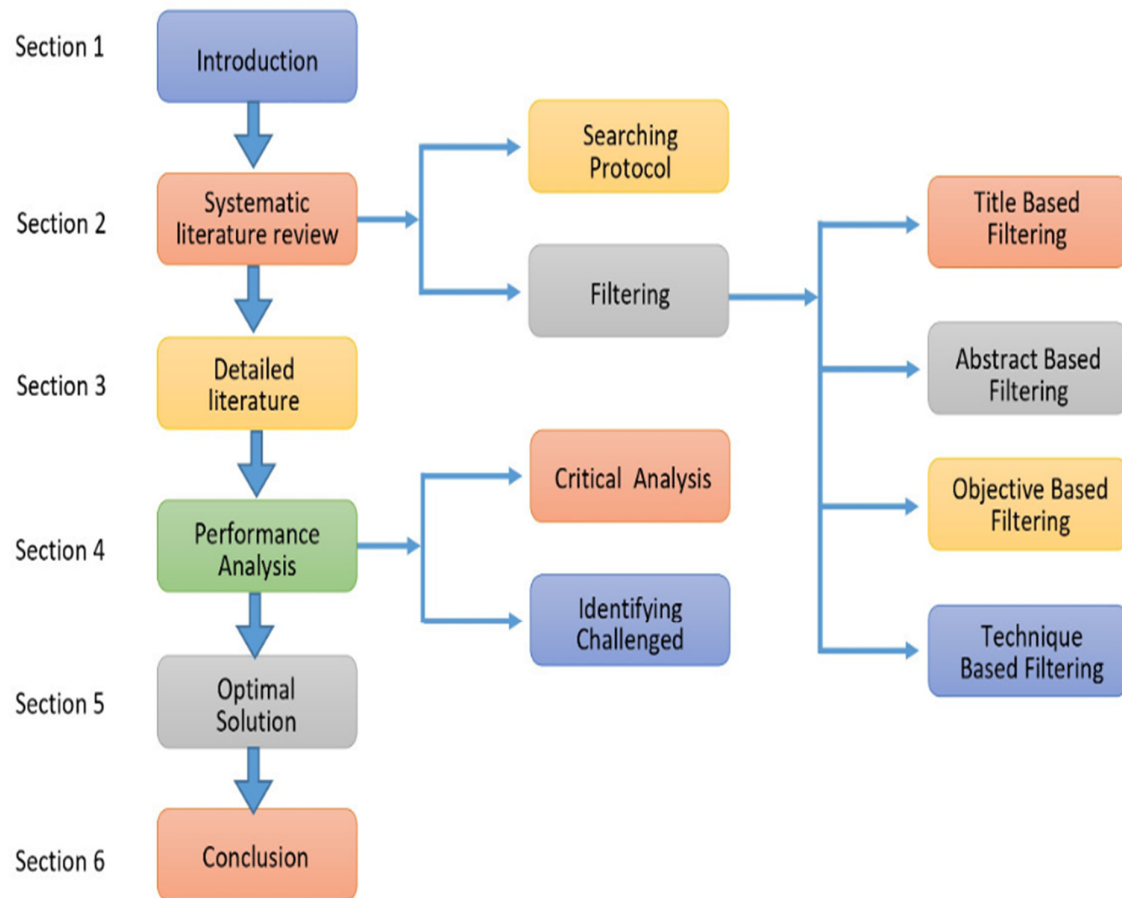
Figure 3. Survey Organization.

Systematic Literature Review.

This report presents a systematic review of the research literature. Systematic literature reviews are a research methodology used to summarize and critically evaluate existing literature on a specific topic. The methodology involves a systematic and transparent search and analysis of published studies to answer a research question. First, a descriptive search methodology was created, then systematic searches were carried out in accordance with it. These searches were guided by strings created in accordance with the identified study question. After then, a search method was developed to group all of the searches into categories based on the search journals. Also, according to study articles that were included. Using their title, abstract, and objectives as filters, they were sorted according on their inclusion criteria.

String Development

First the keyword identified and then these key words as it and with three synonyms were used to develop the string to identify the problem and mainly the research question.

Research question: "Win Prediction of cricket match using machine learning Techniques" For this the Table 1 is use to generate the various queries/strings to search the relevant literature.

Table 2. Synonyms for query generation.

WORDS	SYNONYM 1	SYNONYM 2	SYNONYM 3
Win	Victory	Success	Prevail
Prediction	Forecast	Guess	Divination
Techniques	Method	Approach	Strategy

Table 3. Searching queries.

Query NO 1:.	Victory prediction of PSL using ML techniques.
Query NO 2:	Success prediction of PSL using ML techniques
Query NO 3:	Prevail prediction of PSL using ML techniques.
Query NO 4:	Win forecast of PSL using ML techniques.
Query NO 5:	Win guess of PSL using ML techniques.
Query NO 6:	Win divination of PSL using ML techniques.
Query NO 7:	Win forecast of PSL using ML method.
Query NO 8:	Win forecast of PSL using ML approach.
Query NO 9:	Win forecast of PSL using ML strategy.
Query No 10:	Victory forecast of PSL using ML method.
Query NO 11:	Victory guess of PSL using ML strategy.
Query NO 12:	Victory divination of PSL using ML method.
Query NO 13:	Success divination of PSL using ML strategy.
Query NO 14:	Success guess of PSL using ML method.
Query NO 15:	Prevail divination of PSL using ML method.
Query NO 16:	Prevail prediction of PSL using ML strategy.
Query NO 17:	Prevail divination of PSL using ML techniques.
Query NO 18:	Prevail forecast of PSL using ML method.
Query NO 19:	Prevail divination of PSL using ML approaches.
Query NO 20:	Victory divination of PSL using ML approaches.

Searching Protocol:

A searching protocol was designed according to which papers published over four years (2020, 2021, 2022, and 2023) were selected for searching. A searching protocol is designed based on the previous four-year papers published between the duration of 2020-2023 namely IEEE, springer, and google scholar. The papers are selected published in these 4 years and 3 synonyms of each string word is made for searching. In addition, three synonyms of each keyword and three databases (IEEE, Google Scholar, and Science Direct). Table1 and Figure 1 presents the search strategies.
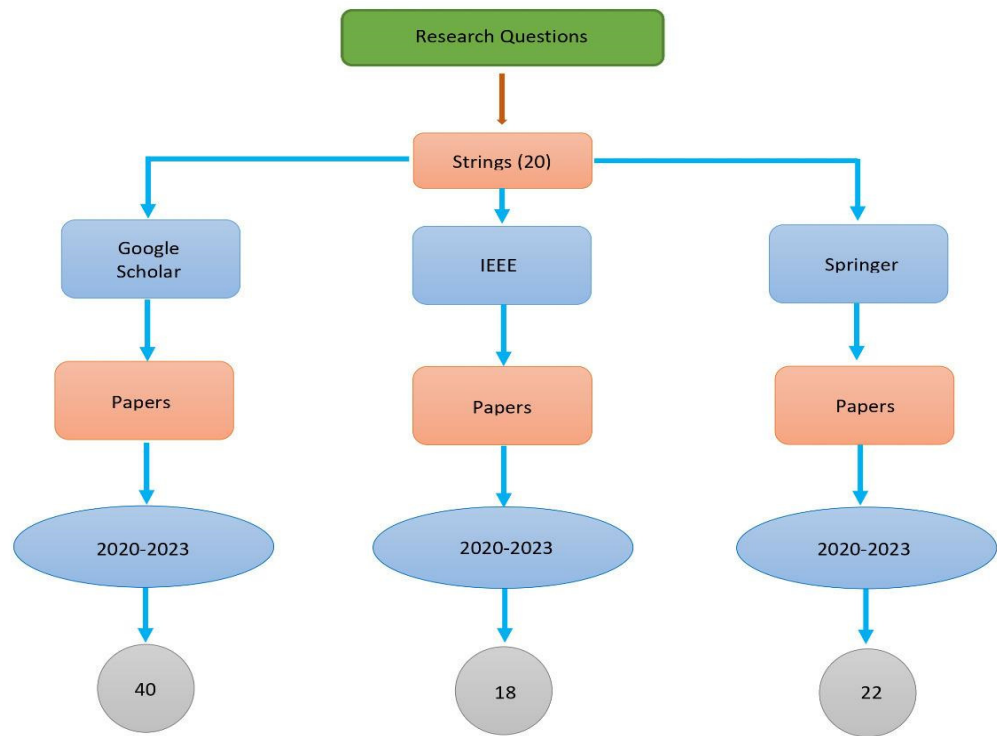
Figure 5. Search Strategy using various Database.

Inclusion & Exclusion Criteria

Inclusion and exclusion criteria are important factors to consider when conducting research studies, including those based on abstracts, titles, or objectives. All papers from journals were included according to an inclusion criterion that was created. This inclusion criteria were addressing the relevant tittle, abstract, and objective.   There were no white papers present. The papers which were not published yet are excluded and those papers were also excluded which were not addressing the relevance according to the inclusion criteria. Inclusion criteria may include abstracts that provide a summary of the study's purpose, methods, and findings, and meet the specific topic or population being studied. Exclusion criteria may include abstracts that are not relevant to the topic or do not meet the study's inclusion criteria.

Inclusion criteria may include titles that are relevant to the study's topic. Exclusion criteria may include titles that are not relevant or do not meet the study's inclusion criteria. Inclusion criteria may include studies that have objectives that are relevant to the study.

Filtering:

Filtering, also known as screening, is an important phase in a systematic literature review. It involves assessing the relevance of articles identified through the search process and deciding which ones should be included in the review. One way to filter articles is to use a three-step process that involves screening based on the title, abstract, and full text.

Title-Based Filtering

During the title screening step, papers are reviewed according to when the titles reflect the review's research issue or inclusion criteria. Articles that do not satisfy the requirements are not reviewed further. The initial stage of the filtering phase was title-based filtering, as shown in Figure 2. All papers that were unrelated to the problem's issue were removed from all databases. The preceding literature was then analysed using abstract-based filtering on material acquire from various databases.
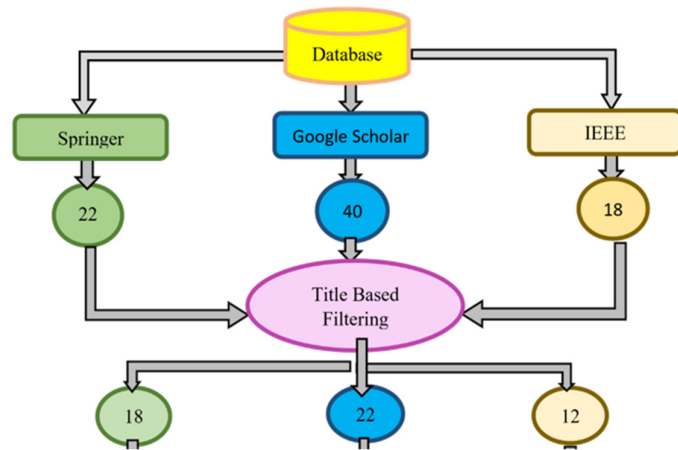

Figure 6. Title base filtering.

Abstract -Based Filtering

The next step is the abstract screening phase, where articles that passed the title screening phase are evaluated based on the relevance of their abstracts. The abstracts are assessed for information on the article's research question, methodology, and findings. Abstract-based filtering was done in the second section. All of the chosen databases did not include any papers whose abstracts did not address the issue.

Objective -Based Filtering

Finally, the objective based screening phase involves a more detailed review of the articles that passed the abstract screening phase. During this phase, the entire article is reviewed to determine if it meets the inclusion criteria of the review. As depicted in Figure 2, objective-based filtering was used in the third stage of the filtering. A table was created showing papers organized by their objectives after all of the papers were filtered according to their objectives. Articles that do not meet the inclusion criteria based on the abstract are excluded. By using a three-step process that involves screening based on the title, abstract, and full text, the review can be conducted in a systematic and transparent manner.

The goals of the research papers were determined and grouped into clusters following the title and abstract-based clustering. This objective-based screening is given in Table 2. The categories of objectives were as follows: Performance (P), Application(A), Efficiency (E), Accuracy(A), Effectiveness (Eff), Cost-Effectiveness (CE).

Table 4. Notations and their definition.

Objective-Measures	Acronyms
Performance	P
Application	A
Effectiveness	E

Accuracy	AC
Efficiency	EFC
Cost-Effectiveness	CE

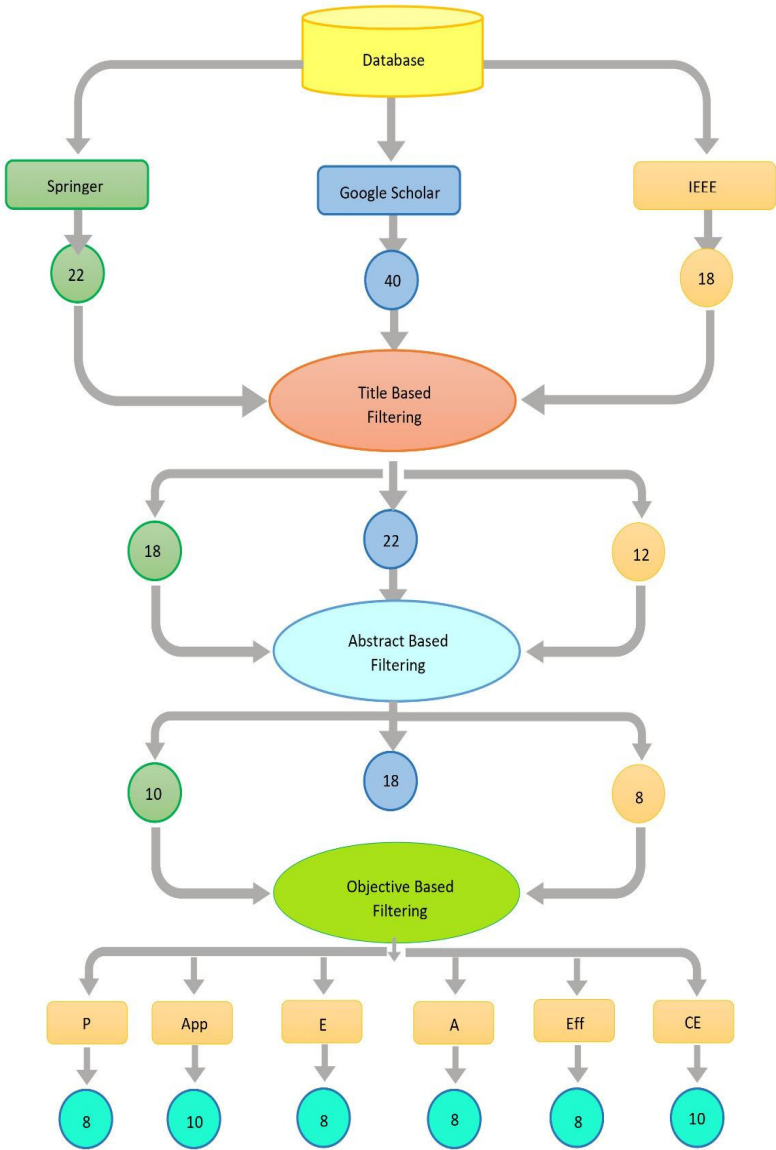
Figure 7. Filtering Process using title, Abstract and objective.

Table 3 presents the literature cited for the review paper. As the machine learning is already being used in

Table 5. Objective based filtering.

Ref.	P	App	E	A	Eff	CE
[1]	✓	✓	–	–	✓	✓
[2]	–	✓	✓	✓	–	✓
[3]	✓	✓	–	–	–	–

[4]	–	–	✓	✓	✓	✓
[5]	✓	✓	–	✓	–	–
[6]	✓	✓	✓	✓	✓	✓
[7]	–	–	–	–	✓	✓
[8]	–	✓	✓	✓	✓	–
[9]	✓	✓	–	–	–	–
[10]	–	–	✓	✓	✓	–

Literature Review for Techniques and Methodology

When conducting a literature review on a specific research topic, it is important to consider the existing work that has been done in the field. This includes examining the techniques and methodologies used in previous studies. The techniques and methodology used in research studies can vary widely, depending on the nature of the research question, the available resources, and the discipline or field of study. Some common techniques and methodologies used in research studies include experimental design, case studies, surveys, and qualitative data analysis.

By examining the techniques and methodology used in previous studies, researchers can gain a better understanding of different approaches, as well as any gaps or areas for improvement in the existing literature.

The evaluation of methods and techniques in SLR involves assessing the quality and validity of the studies that are included in the review. This can help to inform the design and methodology of their own study, and ensure that it is based on the most up-to-date and rigorous approaches in the field. It is important to note that while previous studies can provide valuable insights into techniques and methodology, This can involve examining the quality of the data, the relevance of the study design to the research question, and the potential for bias or confounding factors in the analysis. Overall, a literature review on techniques and methodology can provide a valuable foundation for conducting rigorous and well-designed research in a specific field, and help to advance the state of knowledge in the area of interest. The analysis of predicting the winner of the PSL using machine learning algorithms will provide the contribution of the existing literature and provide a new perspective on the use of these techniques for predicting cricket match outcome

Table 6. Techniques and methodology analysis.

Ref.	Technique	Methodology
[1]	Naïve Bayes Algorithm	To predict the match results between two teams. This is a supervised learning algorithm .by using the testing and training set
[2]	Classification Tree	PSL dataset was used for applying different classification algorithms and Random Forest gave the highest accuracy
[3]	KNN, Random Forest and Naïve Bayes classifier in ML	PSL dataset was used for applying different classification algorithms and Random Forest gave the highest accuracy

[4]	Random Forest Classifier, Naïve Bayes, KNN, Logistic Regression, Decision Tree, SVM, Bagging Classifier, Extra Trees Classifier,	All of these methods are evaluated using historical data is of 15 years (2005 to March 2020). As cricket matches has the Historical cricket statistics without a framework are taken from the ESPN and Cricbuzz websites. Experimental results prove that all approaches are able to predict the best results according to efficiency.
[5]	Neural Network	The use of advanced non-linear modelling techniques such as neural networks to provide an analytical aid in such decision situations.
[6]	Regression Tree The supervised learning approach is used in this to predict the match winner.	
[7]	Decision Trees, Random Forest, and Support Vector Machines	used to predict player performance based on historical data. These models take into account various factors such as batting average, strike rate, and bowling economy rate to predict player performance in future games.
[8]	Decision Trees, Random Forest, Logistic Regression, and Support Vector Machines	used cross-validation to evaluate the performance of the models and selected the best performing model for prediction.

Cricket has evolved, and this has made it a very popular topic for sports experts. Many studies have been conducted on cricket, but they have been unable to reliably forecast match winners because of inconsistent and complex data sets. Although several methods, including KNN, Logistic Regression, SVM, and Naive Bayes, have been used to predict match winners, no method has yet reached accuracy without any restrictions. [2] author claim that they used numerous statistical methodologies for the creation of datasets and experimented with different classification strategies to determine who will win a one-day cricket (50 over) match. He correctly predicted the winner of the time. [3] It used information from the ICC match ratings, ICC ranking points for bowlers and batsmen, home factor, ICC rating disparities, and ground influences on the match to predict One Day International match outcomes. On the basis of this data, they applied Logistic Regression, and their predictions of match outcomes were accurately identified the winning team. In this paper [4] used the machine learning techniques Nave Bayes random forest & Support vector Machines to forecast the result of an ODI Cricket match based on several criteria. These competitions were all held between 2001 and 2015.

The Nave Bayes model is most appropriate among all, as shown by the projected results. In forecasting the victor of the One Day International cricket match, applied binary classification models such decision trees, KNN, Random Forest, and Logistic Regression. There was no cross validation process used. Author has conducted research on methods for determining who would win a game at the end of an over. These methods include using player performance both recent and past as well as

other statistics. Calculating the first team's score at the end of the first inning is the first challenge. To assess and compare the strength of the opposing teams, divide Team A's relative strength by Team B's relative strength, when features are combined to forecast the outcome of the match. The accuracy of the Random Forest classifier (R.F.C.) is 84%. [5] conducted an analysis of the One Day International matches played between 2006 and 2016 and found that, more accurately than earlier searches and models, the top three batsmen are hot candidates for the Man of the Match award time.

The methods used to forecast player performances in a game include Random Forests, Decision Trees, KNN, and Logistic Regression. [5,6] predicted the result of a cricket match. Using characteristics of a dynamic team, such as player history, weather conditions, ground history, and winning %, for winner prediction, he suggested a system for forecasting team results and elaborated on how it functions. On 100 matches, he used this method, and the forecast rate was good. [8] began looking into the extent to which County T-20 match results may be predicted. He forecasted the results of English County Twenty Over Cricket Matches using the machine learning algorithms Nave Bayes & Random Forest.

two thirds of findings using Naive Bayes, by using the Naive Byes Classifier to predict the outcome of the ODI following the conclusion of the first innings. He used datasets from roughly 15 feature rich series that were broadcast between September 30, 2012, and January 1, 2020. As a result, the accuracy rate was a respectable. [9] uses machine learning techniques like Decision Tree & Random Forest to forecast the winner of the T20 World Champion. The dataset was composed of T20 results from ICC tournaments held between 2005 and the present. He took the information from the ESPN Cricinfo website. His work produced satisfactory results, with Australia's team taking first place. [9] predicted that the team with the home-field advantage will perform the best among the eight teams in the ICC rankings. To examine the top performer out of everyone, he employed the machine learning techniques Classifier Tree, Logistic Regression, and Regression Tree. He made use of a dataset that included every series between the top 8 ICC-ranked sides from 2000 to 2014. The best team overall was Team Australia.

The primary goal of this study is to predict the winner of the PSL amongst six teams using machine learning algorithms to analyses previously stored data, which is not being met by the already available literature. We'll use and preprocess the PSL dataset from 2016 to the present. Following preprocessing, the data will be trained on several models to produce the results. In order to determine the likely winner, we will examine the different datasets and employ important criteria like strike rate, bowler economy, etc.

Different attribute to influence the match

Cricket matches are decided by a variety of criteria, including batting, bowling, fielding, team and individual performances, and others. It's never simple to pick the winner of a cricket match. However, there are always some match-specific factors that may benefit one team and occasionally do not, such as home pitch advantage, pitch conditions, and weather. The cricket pitch is crucial to the game of cricket. The captain chooses whether to bat first or go up first according on the pitch report. It is the primary problem with the game. Because these grassy pitches absorb moisture, which is necessary for swing bowling, the ball swings more on such tracks, which is advantageous for fast bowlers. Cricket balls fly off the surface more quickly and with better bounce on hard cricket fields. Yet, grassy fields can aid in the skidding off of new balls.

Home Field Advantage: This is related to the cricket field and surrounding terrain. Home teams have won more than 1.5 times as many games as away teams, winning 40.42% of games to 25.97% for away teams. Individual teams do occasionally lose their home advantage.

Effects of dew: The moisture that forms during the late hours of the night is called dew. Both during daytime and nighttime matches, a coating of moisture can be observed on the grass pitch. The bowling crew is substantially impacted by dew. A damp ball requires more effort to grasp, which puts bowlers at a disadvantage. This is typically exceptional in ball-grabbers who use their fingers to generate the proper amount of rotation.

Day/Night with weather effect: This aspect focuses on the time of day or night on which the game is being played. Batsmen are able to score more runs in day light or in flood lights. The bowling side can also benefit from cloudy and humid weather. The faster bowlers especially enjoy bowling in cloudy conditions because the ball can swing and glide in the air under specific circumstances.

Winning the toss: Each match is decided by the coin toss. The winning captain gets to pick which team gets the opening inning. When a team flips the coin, they win, tie, or lose 68.6% of the time. According to the toss, the team that won has opted to bat 75% of the time, scoring an average of 361 runs as opposed to 326 runs when a team is sent in by the opponent. The likelihood of winning is affected by the runs scored by the side batting first throughout the first inning [14]. In our survey on predicting outcomes in the Pakistan Super League (PSL) using machine learning techniques, our research builds upon foundational insights presented in [15–26].

Critical Analysis

Critical analysis involves examining something carefully and thoroughly to understand its strengths, weaknesses, and limitations. It often requires a deep understanding of the subject matter and an ability to look at it objectively, without bias or preconceptions. When conducting a critical analysis, it's important to consider different perspectives and viewpoints, as well as the underlying assumptions and values that inform them. Ultimately, a successful critical analysis provides a balanced understanding of the subject matter, highlighting its strengths while also acknowledging its limitations and potential areas for improvement.

Table 7. Analysis for limitations/shortcomings.

Ref.	Technique	Limitation/Shortcomings
[1]	Naïve Bayes Algorithm	Supervised learning needs past history.
[2]	Classification Tree	Accuracy Compromised
[3]	KNN, Random Forest and Naïve Bayes classifier from ML techniques.	Accuracy level is not upto mark/
[4]	Random Forest Classifier, Naïve Bayes, KNN, Logistic Regression, Decision Tree, SVM, Bagging Classifier, Extra Trees Classifier,	These techniques on unstructured data. Many main features are missing.
[5]	Neural Network	Not efficient with nonlinear data.
[6]	Regression Tree	Only efficiently work with linear data
[7]	Decision Trees, Random Forest, Logistic Regression, and Support Vector Machines	Only Random Forest Model was the best performing model rest of the model were not up to the mark regarding the accuracy.

Decision Trees, Random ML techniques such as Decision Trees, Random Forest Machines	Forest, and Support Vector and Support Vector Machines have been used to predict player performance, while clustering techniques have been used to analyse team strategy.

Critical Analysis of different techniques for predicting Cricket Winners in PSL

Naive Bayes: Naive Bayes uses the Bayes probability theorem and makes the potentially incorrect assumption that each feature is independent of each class label (or predicted variable). Recursive feature removal combined with a naive Bayes model.

Regressor with Decision Tree: Decision By employing a tree node system to learn from the noise in the data, a tree regression has been utilized to check for overfitting. Decision tree regressors use noise in training data to extract details when the maximum depth of the tree is high. Using the tree node principle, decision trees classify data by grouping instances into a tree node system. This hierarchy breaks down complex decision-making systems into smaller, simpler options, resulting in a straightforward solution that is simple to carry out.

Support Vector Machine (SVM): Support Vector Machine has established itself as the most popular component classifier for a variety of prediction approaches, including facial identification, picture recognition, and medical health diagnosis. An ideal hyperplane is produced by the SVM classifier using the training data to categories new samples. A hyperplane is a plane that splits a line into two halves, each with a class on either side. Regularization parameters are used to optimize SVM. The SVM Optimization is covered by the regularization parameter. A subset of supervised machine learning methods known as SVM requires training on classes of output that have already been determined. An ideal hyperplane for categorizing fresh samples is produced by the SVM classifier using training data.

Classifier with Random Forests: A tool for classification and regression is the Random Forest classifier. The Random Forest Classifiers use a number of trees working at random in a forest to categories new instances. Each tree's job is to vote for the class by assigning a class label or target variable. And the Random Forest Classifier will decide that which node gets the most votes. In order to increase projected accuracy and manage over-fitting, Random Forest estimate and average the strategy on samples of dataset fitted with various numbers of decision tree classifiers. The subsamples used here are still of the same size as the original input.

Research Gaps

Table 8. Research Gaps & solutions.

Ref.	Technique	Challenges	Solutions
[1]	Naïve Bayes Algorithm	Extra Feature/Irrelevant features	PCA used for feature extraction
[2]	Classification Tree	Accuracy Compromised	Increase dataset
[3]	KNN, Random Forest and Naïve Bayes classifier from ML techniques.	Accuracy level is not up to mark	Increase dataset

| [4]	Random Forest Classifier, Naïve Bayes, KNN, Logistic Regression, Decision Tree, SVM, Bagging Classifier, Extra Trees Classifier,	These techniques on unstructured data. Many main features are missing.	Include the more feature to enhance the results.
[5]	Neural Network	Not efficient with lot of empty values data.	Reduce the empty values by applying different feature extraction methods.
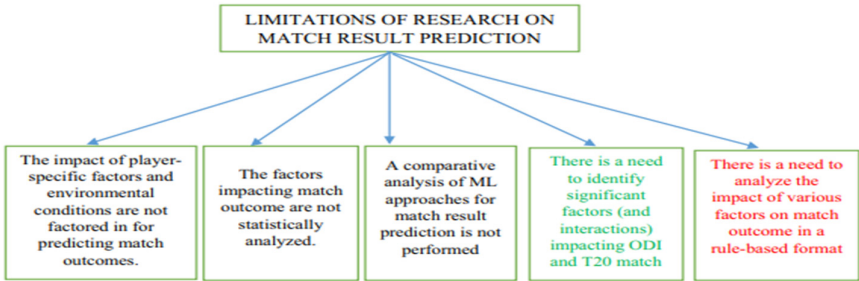
Figure 8. Research limitations.

References

1. Wickramasinghe, I. (2022). Applications of Machine Learning in cricket: A systematic review. Machine Learning with Applications, 10, 100435.
2. Kapadia, K., Abdel-Jaber, H., Thabtah, F., & Hadi, W. (2020). Sport analytics for cricket game results using machine learning: An experimental study. Applied Computing and Informatics, (ahead-of-print).
3. WEERADDANA, N., & PREMARATNE, S. (2021). Unique approach for cricket match outcome prediction using xgboost algorithms. Journal of Theoretical and Applied Information Technology, 99(9), 2162-2173.
4. Tekade, P., Markad, K., Amage, A., & Natekar, B. (2020). Cricket match outcome prediction using machine learning. International journal, 5(7).
5. Das, K., & Awais, M. (2021). A Predictive Analysis of PSL Match Winner Using Machine Learning Techniques. UMT Artificial Intelligence Review, 1(1), 1-1.
6. Bhatia, V. (2020, November). A review of Machine Learning based Recommendation approaches for cricket. In 2020 Sixth International Conference on Parallel, Distributed and Grid Computing (PDGC) (pp. 421-427). IEEE.
7. Awan, M. J., Gilani, S. A. H., Ramzan, H., Nobanee, H., Yasin, A., Zain, A. M., & Javed, R. (2021). Cricket match analytics using the big data approach. Electronics, 10(19), 2350.
8. Tyagi, S., Kumari, R., Makkena, S. C., Mishra, S. S., & Pendyala, V. S. (2020,December). Enhanced predictive modeling of cricket game duration using multiple machine learning algorithms. In 2020 international conference on data science and engineering (ICDSE) (pp. 1-9). IEEE.
9. Jayanth, S. B., Anthony, A., Abhilasha, G., Shaik, N., & Srinivasa, G. (2018). A team recommendation system and outcome prediction for the game of cricket. Journal of Sports Analytics, 4(4), 263-273.

10. Goel, R., Davis, J., Bhatia, A., Malhotra, P., Bhardwaj, H., Hooda, V., & Goel, A. (2021). Dynamic cricket match outcome prediction. Journal of Sports Analytics, 7(3), 185-196.

11. Carloni, L., De Angelis, A., Sansonetti, G., & Micarelli, A. (2021). A machine learning approach to football match result prediction. In HCI International 2021-Posters: 23rd HCI International Conference, HCII 2021, Virtual Event, July 24–29, 2021, Proceedings, Part II 23 (pp. 473-480). Springer International Publishing.

12. Ashraf, J., Bhatti, S., & Talpur, S. Predicting the Best Team Players of Pakistan Super League using Machine Learning Algorithms. International Journal of Computer Applications, 975, 8887.

13. Puram, P., Roy, S., Srivastav, D., & Gurumurthy, A. (2022). Understanding the effect of contextual factors and decision making on team performance in Twenty20 cricket: an interpretable machine learning approach. Annals of Operations Research, 1-28.

14. Bunker, R., & Susnjak, T. (2022). The application of machine learning techniques for predicting match results in team sport: A review. Journal of Artificial Intelligence Research, 73, 1285-1322.

15. Almuayqil, S. N., Humayun, M., Jhanjhi, N. Z., Almufareh, M. F., & Khan, N. A. (2022). Enhancing sentiment analysis via random majority under-sampling with reduced time complexity for classifying tweet reviews. Electronics, 11(21), 3624.

16. Pal, S., Jhanjhi, N. Z., Abdulbaqi, A. S., Akila, D., Almazroi, A. A., & Alsubaei, F. S. (2023). A hybrid edge-cloud system for networking service components optimization using the internet of things. Electronics, 12(3), 649.

17. Tayyab, M., Marjani, M., Jhanjhi, N. Z., Hashem, I. A. T., Usmani, R. S. A., & Qamar, F. (2023). A Comprehensive Review on Deep Learning Algorithms: Security and Privacy Issues. Computers & Security, 103297.

18. Alwakid, G., Gouda, W., Humayun, M., & Jhanjhi, N. Z. (2023). Diagnosing Melanomas in Dermoscopy Images Using Deep Learning. Diagnostics, 13(10), 1815.

19. Shahid, H., Ashraf, H., Javed, H., Humayun, M., Jhanjhi, N. Z., & AlZain, M. A. (2021). Energy optimised security against wormhole attack in iot-based wireless sensor networks. Comput. Mater. Contin, 68(2), 1967-81.

20. Wassan, S., Chen, X., Shen, T., Waqar, M., & Jhanjhi, N. Z. (2021). Amazon product sentiment analysis using machine learning techniques. Revista Argentina de Clínica Psicológica, 30(1), 695.

21. Almusaylim, Z. A., Zaman, N., & Jung, L. T. (2018, August). Proposing a data privacy aware protocol for roadside accident video reporting service using 5G in Vehicular Cloud Networks Environment. In 2018 4th International conference on computer and information sciences (ICCOINS) (pp. 1-5). IEEE.

22. Kok, S. H., Azween, A., & Jhanjhi, N. Z. (2020). Evaluation metric for crypto-ransomware detection using machine learning. Journal of Information Security and Applications, 55, 102646.

23. Shafiq, M., Ashraf, H., Ullah, A., Masud, M., Azeem, M., Jhanjhi, N. Z., & Humayun, M. (2021). Robust Cluster-Based Routing Protocol for IoT-Assisted Smart Devices in WSN. Computers, Materials & Continua, 67(3).

24. Lim, M., Abdullah, A., Jhanjhi, N. Z., & Supramaniam, M. (2019). Hidden link prediction in criminal networks using the deep reinforcement learning technique. Computers, 8(1), 8.

25. Gouda, W., Sama, N. U., Al-Waakid, G., Humayun, M., & Jhanjhi, N. Z. (2022, June). Detection of skin cancer based on skin lesion images using deep learning. In Healthcare (Vol. 10, No. 7, p. 1183). MDPI.

26. Adeyemo Victor Elijah, Azween Abdullah, NZ JhanJhi, Mahadevan Supramaniam and Balogun Abdullateef O, "Ensemble and Deep-Learning Methods for Two-Class and Multi-Attack Anomaly Intrusion Detection: An Empirical Study" International Journal of Advanced Computer Science and Applications(IJACSA), 10(9), 2019. http://dx.doi.org/10.14569/IJACSA.2019.0100969