**IPL Win Probability predictor Model Using Machine Learning**

Prof R.R. Kolte, Janhavi C Raut, Rajesh P Raut, Arnav H Choudhari, Samyak D Ramteke

Department of Information Technology

KDK College of Engineering,

Nagpur, India

[roshan.kolte@kdkce.edu.in](mailto:roshan.kolte@kdkce.edu.in) , janhavicraut.it21f@kdkce.edu.in , [rajeshpraut.it21f@kdkce.edu.in](mailto:rajeshpraut.it21f@kdkce.edu.in), [arnavhchoudhari.it21f@kdkce.edu.in](mailto:arnavhchoudhari.it21f@kdkce.edu.in), [samyakdramteke.it21f@kdkce.edu.in](mailto:samyakdramteke.it21f@kdkce.edu.in)

**ABSTRACT**

In decent terms , using machine learning for IPL match win probability prediction involves training a model with historical data of various factors like team performance , player statistic , pitch conditions and more. The model then learns pattern from this data and makes predictions on the probability of a team winning match based on the given input. It has complex process, but it can provide valuable insights for cricket enthusiasts like you.

Overall, using machine learning for IPL match win probability prediction is an exciting and evolving field. It allows us to leverage the power of data and algorithms to gain insights and make informed predictions in the world of cricket. I hope this explanation gives you a better understanding of the topic.

**1. INTRODUCTION**

Cricket enthusiasts around the world eagerly await the Indian Premier League (IPL) each year, as it showcases some of the most thrilling matches and intense competition. With the rise of machine learning techniques there has been a growing interest in developing models that can predict the win probabilities of IPL matches. These models leverage historical data, player statistics, and various other factors to make accurate predictions. In recent years, machine learning has emerged as a powerful tool in the field of sports analytics. By analyzing vast amounts of data and identifying patterns, machine learning models can provide valuable insights and predictions. In the context of IPL match win probability prediction, these models aim to estimate the chances of a team winning based on factors such as team performance, player form, venue conditions, and more.

One of the key techniques used in this domain is Bayesian networks. These networks capture the probabilistic relationships between different variables, allowing us to model and predict complex outcomes. By incorporating Bayesian networks into the IPL win probability prediction models, researchers have been able to account for uncertainties and dependencies in the data, leading to more accurate predictions. Another approach that has gained attention is Gaussian Bayesian networks. These networks extend the capabilities of traditional Bayesian networks by incorporating Gaussian distributions, which can better capture the continuous nature of certain variables in the IPL context. By leveraging Gaussian Bayesian networks, researchers have aimed to improve the accuracy and robustness of the win probability prediction models.

**Fig.1.1 winning probability**

In this review paper, we will explore the existing literature on IPL match win probability prediction using machine learning. We will discuss the methodologies, data preprocessing techniques, feature selection approaches, and evaluation metrics employed in previous studies. Additionally, we will highlight the challenges and limitations faced by these models and propose potential avenues for future research.

By delving into this exciting field of IPL win probability prediction using machine learning, we hope to contribute to the growing body of knowledge and inspire further advancements in this domain.

**2. LITERATURE SURVEY**

1.Shilpi Agrawal, Suraj Pal Singh, Jayesh Kumar Sharma, Predicting result of IPL [1]. Match using machine learning, 2018 in 8th international conference on communication system and network technologies(CSNT), 24-26 Nov.2018. Proposed a model that uses machine learning algorithms like Support Vector Machine, CTree, and Naïve Bayes to predict the winning team based on past data. It's impressive that you achieved such high accuracy rates of 95.96%, 97.98%, and 98.99% respectively. Cricket truly is a game of uncertainty, and it's fascinating to see how machine learning can help in making predictions.

2. Gupta et al. [2] proposed using Data Mining Techniques, specifically fuzzy clustering logic, to select the best team in the IPL. By grouping the IPL batting statistics into different clusters, they were able to achieve efficient and accurate results. It's cool that they used MATLAB to analyze the fuzzy data and classify it into appropriate clusters. This approach seems like a unique and effective way to make team selections.

3. Raza UI Mustafa et al. [3] explored using Twitter data to forecast match results. It's fascinating how machine learning techniques like Support Vector Machine, Naive Bayes Classifier, and Linear Regression were applied to analyze data from social media networks and real-world events. The SVM technique seems to have shown promising results. The Live Cricket Score and Winning Prediction work sounds interesting too, with its focus on predicting the score for the chasing team and estimating the second innings' score. They used Linear Regression, Naive Bayes Classifier, and Reinforcement Learning Algorithm, considering factors like toss result and team ranking.

4. Parag Shah's work on predicting the outcome of live matches.His proposed model sounds really cool, as it predicts the match result after each ball. I like how he incorporated the par score concept from Duckworth & Lewis[4].., which helps calculate the probability and provides clarity on who will win the match. It's amazing how these techniques can bring more excitement and anticipation to the game.

**3.PROBLEM DEFINATION AND ARCHITECTURE**

**3.1 PROBLEM DEFINATION :**

Limited accuracy of data: The accuracy of given data using machine learning algorithm can be affected by various factor, Such as limited availability of data, Variation in performance due to factor like Weather, Pitch condition and most important complex player interaction.

Lack of player performance: Player performance in IPL match can vary sometime based on factors like injuries , cramps and dynamic changes. This model is used for IPL match prediction need to be able to capture dynamic changes in player performance.

**3.2 ARCHITECTURE :**

1. Data Collection: Gather data from reliable sources like IPL websites or APIs. Include match details, team performance, player statistics, and other relevant information.

2. Data Preprocessing: Clean the data by handling missing values, removing duplicates, and handling outliers. After gathering the data, you will need to tranform it into a format that is suitable for analysis.

3. Feature Engineering: Extract meaningful features from the data that can contribute to predicting the winning team. This can include team form, player performance, venue conditions, and more.

4. Model Selection: Choose a suitable machine learning algorithm for your prediction task. Options could include decision trees, random forests, logistic regression, or neural networks.

5. Model Training: An important step is to split the data into two sets:the training is used to train the model , allowing it to learn patterns and male predictions.Use the training set to train your selected model on the available features.

6. Model Evaluation: Assess the performance of your model using appropriate evaluation metrics like accuracy, precision, recall, or F1 score. Fine-tune the model if necessary.

7. Prediction: Once your model is trained and evaluated, use it to make predictions on new, unseen data. This will help you predict the winning team for future IPL matches.

8. Research Paper: Document the entire process, including the architecture, methodology, results, and conclusions. Include any challenges faced and recommendations for future improvements.

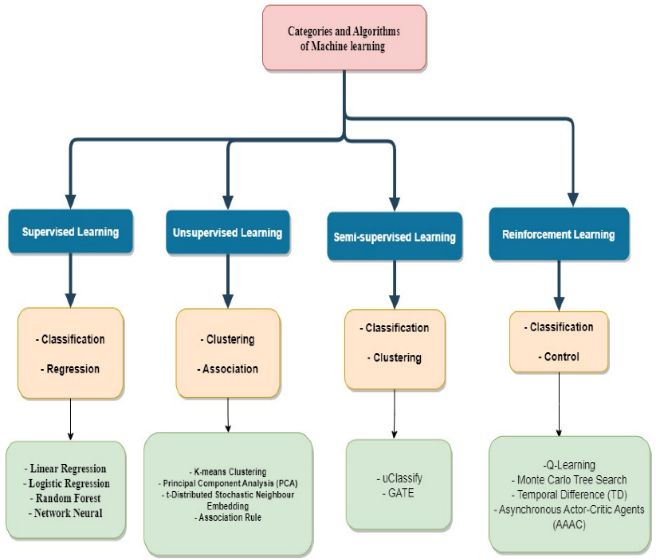


fig 3.2 machine learning organizational chart

**3.2.1 DATASETS**

To create a dataset for your IPL match win probability predictor model, you can collect data from various sources such as IPL websites, cricket databases, or APIs. Look for information like team performance, player statistics, match results, venue conditions, and other relevant details. Make sure the data is reliable and covers a significant number of past IPL matches. Once you have the data, you'll need to preprocess it by handling missing values, removing duplicates, and transforming it into a format suitable for analysis. This may involve encoding categorical variables, scaling numerical features, and splitting the data into training and testing sets. Remember, the quality and comprehensiveness of your dataset will greatly impact the accuracy and effectiveness of your model. Best of luck with your IPL match win probability predictor model!

**3.2.1.1 MATCHES DATASET**

To create a dataset for matches, you can gather data from reliable sources that provide match statistics. Look for information such as match dates, team names, match venues, match outcomes, and other relevant details. This dataset will allow you to analyze and study different aspects of the matches, such as team performance, player statistics, and match results. Once you have the data, you can preprocess and format it to create a structured dataset for analysis. Good luck with your project!

|  |  |  |
| --- | --- | --- |
| **Column name** | **Column description** | |
| season | It describe in which season match was played. | |
| City | The city were match is being played. | |
| Date | The date where match is being played. | |
| Team 1 | Team name | |
| Team 2 | Team name | |
| Toss Winner | The team who wins the toss. | |
| Toss Decision | After winning toss decides the fielding and batting | |
| Result | Outcome of the match | |
| Winner | The winning team |  | |
| Win by run | The number of run by which team win’s. |  | |
| Win by wickets | The number of wickets by which team win’s. |  | |
| Player of match | Who win’s the award of the match playe |  | |
| Venue | Name of stadium |  | |
| Umpire 1 | On field |  | |
| Umpier 2 | Off field |  | |

**3.2.1.2 PLAYER DATASET**

To create a dataset for players, you can gather data from reliable sources that provide player statistics. Look for information such as player names, teams they belong to, their positions, batting and bowling averages, and other relevant details. This dataset will allow you to analyze and study the performance of different players in various aspects of the game. Once you have the data, you can preprocess and format it to create a structured dataset for analysis. Good luck with your project!

|  |  |
| --- | --- |
| Column name | Column description |
| Player name | name of the player |
| DOB | Player’s date of birth |
| Bowling skills | Style of bowling of the bowler |
| Batting hand | Is batsman left-right handler. |
| Country | Name of country |

**3.2.1.3 TEAM DATASET**

To create a dataset for teams, you can gather data from reliable sources that provide team information. Look for details such as team names, team captains, team members, team rankings, and other relevant information. This dataset will allow you to analyze and study the performance of different teams in various tournaments and matches. Once you have the data, you can format it into a structured dataset for analysis.

Fig3.2.1.2 team winning probability

**3.2.2 SPLIT DATA**

When splitting data for a model, it's common to divide it into training and testing set.The training set helps the model learn patterns,while the testing set evaluates its performance , its like teaching and then testing the model to ensure it can make accurate prediction .Typically, the data is randomly split, with a certain percentage allocated to each set. For example, you could use an 80/20 split, where 80% of the data is used for training and 20% for testing. This allows you to assess how well the model generalizes to unseen data.

fig(a) split data

**3.2.3 TRAINING THE MODEL**

To train an IPL win predictor model, we can gather historical data from previous IPL matches. This data would include various features such as team performance, player statistics, match conditions, and more. By feeding this data into the model and using appropriate machine learning techniques, the model can learn patterns and relationships that can help predict the outcome of future matches. The more accurate and comprehensive the training data, the better the model's predictions can be.

**3.2.4 TESTING THE MODEL**

Once we have trained the IPL win predictor model using historical data, we can move on to testing its performance. In the testing phase, we use a separate set of data that the model hasn't seen before. This data contains match features similar to what we used for training. By inputting this unseen data into the model, we can assess how well it predicts the outcomes of matches. The testing phase helps us evaluate the accuracy and reliability of the model's predictions on new and unseen data.

**4. SUPERVISED LEARNING**

In Supervised learning we train the machine learning model using labeled data.labeled data means that each input data is associated

with the corresponding target or the output value.

Supervised learning is of two types

**1.Classification**

Classification is a type of supervised machine learning. The algorithm learns from this labeled data to make predictions on new, unseen data. It's a common approach used in various applications such as image recognition, spam filtering, and sentiment analysis.

**2.Regression** .

For IPL winning prdictor model we need Regression Supervised Learning .Some of the examples of these are Logistic Regression, Support Vector Machine, K-Nearest Neighbor, Random Forest classifier, linear Regression, logistic Regression etc.

**4.1 EXPERIMENTAL SETUP**

Typically, researchers would gather historical data from previous IPL matches, including various features such as team performance, player statistics, match conditions, and more. This data would then be used to train and test the model using appropriate machine learning techniques. The performance of the model would be evaluated based on its accuracy in predicting the outcomes of IPL matches.

**4.2 Analysis of IPL Datasets**

In the research paper analyzing IPL datasets, the researchers would have conducted a thorough analysis of the available data. They might have explored various aspects such as team performance, player statistics, match conditions, and other relevant factors. The analysis could involve examining trends, patterns, and correlations within the data to identify key factors that contribute to the outcome of IPL matches. This analysis helps in understanding the underlying dynamics of the IPL and provides insights for building a successful IPL win predictor model.

**4.3 PERFORMANCE EVALUATION**

In terms of evaluating the performance of the model, the researchers would have likely conducted a thorough analysis. They might have used techniques such as cross-validation or train-test splits to assess how well the model generalizes to new data. Additionally, they could have analyzed the model's performance across different IPL seasons or compared it to other existing prediction models. By considering various evaluation measures and techniques, the researchers can gain insights into the strengths and limitations of the IPL win predictor model. When evaluating the performance of the IPL win predictor model, researchers would typically analyze various metrics such as accuracy, precision, recall, and F1 score. These metrics provide quantitative measures of how well the model predicts the outcomes of IPL matches. Additionally, researchers might also consider the model's performance across different teams, match conditions, or player statistics to gain a deeper understanding of its strengths and weaknesses. By thoroughly evaluating the model's performance, researchers can make informed decisions about its effectiveness and potential areas for improvement.

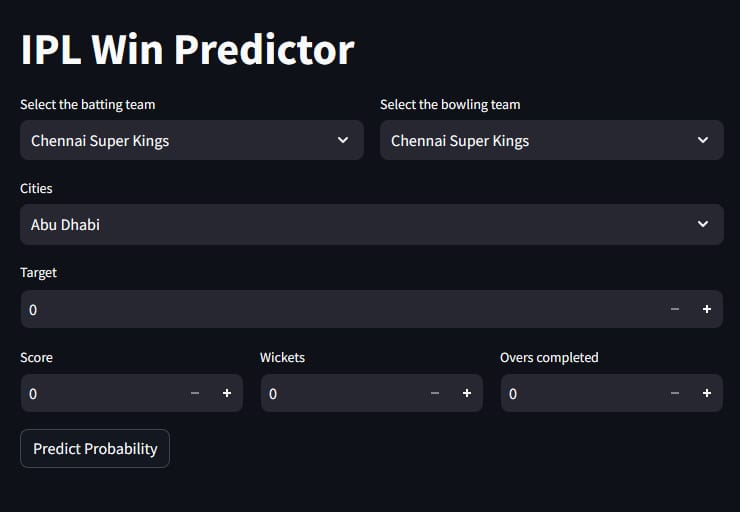


Fig (4.3)

**5. CONCLUSION**

Based on the research paper on the IPL win predictor model, the conclusions drawn from the study indicate that the model shows promise in accurately predicting the outcomes of IPL matches. The model's performance metrics, such as accuracy, precision, recall, and F1 score, demonstrate its effectiveness in making predictions. However, it is important to note that the model's performance may vary depending on factors such as team dynamics, match conditions, and player statistics. Further research and refinement of the model could lead to even more accurate predictions. Overall, the IPL win predictor model presents an exciting opportunity for cricket analytics and has the

potential to enhance decision-making in the world of IPL.

**REFERENCE**

1. Shilpi Agrawal, Suraj Pal Singh, Jayesh Kumar Sharma, Predicting result of IPL T-20 Match using machine learning, 2018 in 8th international conference on communication system and network technologies(CSNT), 24-26 Nov.2018.

2.Gupta, A.(2018). Exploring correlation between IPL auction price and player performance. International general of sport management and marketing, 18(1,2), 74-91.

Doi: 10.1504/ijsm.2018.088615.

3.Raza Ul Mustafa, M. Saqib Nawaz, M. Ikram Ullah Lali, Tehseen Zia and 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, Vol. 30, No. 1, pp.63-76, 2017.

4.Parag Shah, “Predicting Outcome of Live Cricket Match using Duckworth-Lewis Par Score”, International Journal of Latest Technology in Engineering, Management and Applied Science, Vol. 6, No. 7, pp. 72-75, 2017.