Winning Cricket Team Prediction Using Machine Learning

*(Under the guidance of **Mr. D.Pavan Kumar**, Assistant.Professor, B.tech, M.tech.,*

Department of Information Technology, Tirumala Engineering College)

R.Jaya Surya

IT Department

suryaramisetty26@gmail.com

Tirumala Engineering College

K.Kalyani

IT Department

karrikalyani77@gmail.com

Tirumala Engineering College

G.Lavanya

ganginenilavanya@gmail.com

Tirumala Engineering College

M.Ramya

ramyamarlapati0@gmail.com

Tirumala Engineering College

Abstract

As kids, we were always fascinated to see our favourite team could get a particular situation inside the match. We discovered that we can in fact achieve it after understanding device learning and with the vast information we have been given today Cricket is the most popular sport in India, and it is played there in all of its regions in various formats such T20, ODI, and Test. The Indian Premier calculates the likelihood of winning by batting first against a certain team at a specific match location. In this research, utilising machine learning algorithms such as SVM, Random Forest Classifier (RFC), Logistic Regression, and we have suggested a model for predicting the results of the IPL matches. The accuracy of the Random Forest algorithm, whichis88.10%, exceed so ther algorithms, according to experimental results. We will organize the data with NumPy and Pandas package, then visualize it with Matplotlib and Seaborn. We tested a number of regression methods in order to find the most appropriate one. Finally, we used linear regression to predict the rating on the surrender of twenty overs.

Keyword–Cricket Prediction, Cricket analysis, Lasso Regression, Naïve Bayes, Logistic Regression, Random Forest Classifier.

I. INTRODUCTION

The act of extricating concealed information from huge measures of crude information is known as information mining. The information should be new and not self - evident, and it should be usable."The nontrivial extraction of beforehand obscure verifiable and potentially gainful data from information "is the means by which information mining is characterized.. It is characterized as "the study of recovering important information from enormous datasets." It is one of the positions engaged with the data set information revelation process.

Information digging is a procedure for extracting information from information and introducing it such that people can comprehend.

The Twenty20 format of cricket is the third and most recent one. This style was developed in 2006, and India won the inaugural world championship in 2007.Theshortgame, whichhas20over, is finished in under three hours. Each team has 20 over to play, and there are only 2 teams involved. Due to the IPL, the t20 format is extremely popular in India. The popularity of the t20 formatinIndiaisaresultofthisevent.

II. LITERATURESURVEY

The consequences of test matchesmight be expected by basic proportions of batting and bowling inputs, as per research utilizing anarranged reaction model to conjecture test cricket results.Rain-delayed chases, effective fourthperiodschases,andsuccessfulfourthsessionrunschasescontestswerethemostfrequentinstancesof the model making inaccurate projections.

Currently, there is a system which can calculate the current run rate and from it calculates the final score of the team. It doesn't consider the fact about the no of wickets and also where the game is being played. The problem with the current system is that it is unable to predict the score of the 2nd team and also unable to predict the win percentage This system which is developed will have 2 model in it the 1st model predict the score a team will get after playing 50 over from the current situation. The second method predicts the win percentage of both teams even before the match has started this done by player selection. We found that error in regression toward the mean classifier could be a smaller quantity than Naïve mathematician in predicting match outcome has been sixty-eight ab initio from 2-15 overs to ninety-one until the top of 42th over

1) CRICKET MATCH OUT COME PREDICTION USING MACHINE LEARNING

In cricket, particularlythetwenty20format is most watched and loved by the people, where noon ecanguess who will win the match until the last ball of the last over. In India, The Indian Premier League (IPL) started in 2008 and now it is the most popular T20 league in the world. So we decided to develop a machine learning model for predicting the outcome of its matches. Winning in a Cricket Match depends on many key factors like a home ground advantage, past performances on that ground, records at the same venue, the overall experience of the players, record with a particular opposition, and the overall current form of the teamandal so the individual player. This paper briefs about the key factors that affect the result of the cricket match and the regression model that best fittest is data and gives the best predictions.

Data Mining and Machine Learning in sports analytics, is a brand-new research field in computer science with a lot of challenge. In this research the goal is to design a result prediction system for a T20 cricket match, in particular for an IPL match while the match is in progress. Different Machine Learning and statistical approach were taken to find out the best possible outcome. A very popular mathematical technique named Multiple Linear Regression is used in order to make comparison of results found. This model is very much popular in predictive modeling. Currently, in Twenty-Twenty (T20) cricket matches first innings score is predicted on the basis of current run rate which can be calculated as the amount of runs scored per the number of overs bowled. It does not

include factors like number of wickets fallen, venue of the match, toss.

2) IPL Cricket Score And Winning Prediction Using Machine Learning Techniques

Abstract - As cricket is the mostly played game. There are so many series are played in country one of them is Indian Premier League (IPL). Now it is conducted among 8 teams. In these papers the model has been proposed that has two me thuds the first one is prediction of score and the second one is team winning prediction. In these the score prediction in clues line a regression, lassoer egression and rider egress on where as in winning prediction SVC classifier, decision tree classifier and random for estclassi fier are used. The model used the supervised machine learning algorithm to predict the winning. Random Forest Classifier used for good accuracy and the stable accuracy so that desired predicted output is accurate.

3) Sport analytics for cricket game results using machine learning: An experimental study

Abstract-Indian Premier League (IPL) is one of the more popular cricket world tour naments, and its financial is increasing each season, its viewers hi phas increased markedly and the betting market for IPL is growing sign ficantly every year. With cricket being a very dynamic game, bettors and bookies are in centivised to bet on the match results because it is a game that changes ball-by-ball. This paper in vestige esmachine learning technology to deal with the problem of predicting cricket match results based on history calmatch data of the IPL. Influential features of the dataset have been identified using filter-based methods in cluding Correlation based Feature Selection, Information Gain (IG), Relief Find Wrapper. More importantly, machine learning techniques in cluding Na€ıve Bayes, Random Forest, K-Nearest Neighbor (KNN)and Model Trees(classification via regression) have been adopted to generate predictive models from distinctive featuresets derived by the filter-based methods.

4) CRICKET SCORE PREDICTION

Abstract-Now a days the final score of the first inning so fanny cricket match is predicted using CRR (Current Run Rate) method. The number of average runs scored in an over is multiplied by the to tall number of over's to get the final score. These kinds of systems are not useful when the T20 matches are considered because in T20 cricket the match can change its state very quickly irrespective of current run rate. The match can change within 1 or 2 over's. So, to get an accurate score predict on we should have assays temthatcanpredictthefirstinningsscoremoreeffectively.Lotsofpeoplelikewatchingcricket and they also like to predict the final score. This research paper focusesonan accurate prediction of cricket scores for live IPL match reconsidering the previous dataset available and also considers the various factors that play an important role in the score prediction.

The performance of model depends on various features like wickets taken in last 5 over's, runs scored in last 5 over's, over's, overall score and wickets at current ball. The proposed model contains data from IPL matches played between years 2008 and 2019. This paper will give us step by step insights on how one can predict projected score of 1stinning while the match is still in progress. Linear Regression algorithm is used to predict the score.
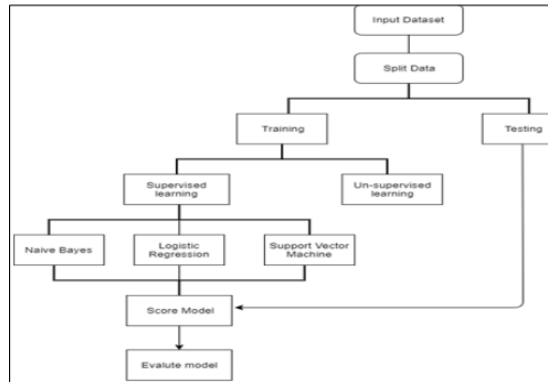
III .METHOLODOGY

A. System Architecture:

Fig.-Flow Diagram for Search System

First the system will be fed with the input data set where it comprises data like player details, players core and place where match is played etc.

Then the data will be processed further and split in to training and testing datasets. Now the training data's etisfur the repletion to supervise dumdum supervised learning.

Here some suitable algorithms will be applied for the super vides learning data's et sand those algoid thesauri Lasso Regression, Naïve bayes, Logistic Regression, Support vector Machine and Random Forest Algorithms. The suitable algorithm will be picked to predict the outcome hence it will be matched with testing datasets and score model will be generated.
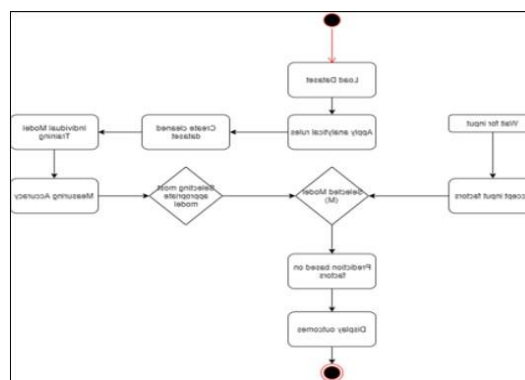
B. Input Pre-processing:

Fig.-Flow Diagram of Input Pre-processing

This tips in input pre-processing include:

First the data set will be loaded and later the analytical rules will be applied. As the data is not pure and cleaned the cleaning process will be perform ed which clears the out liers.Then extstepist ot rain the individual model which is used to me asure the accuracy which in return he lps the score prediction.

C. Algorithm

1) Lasso Regression:

Lasso regression is a regularization technique. it's used overeggs ression strategies for accurate prediction. This mode lushes shrinkage. Shrinkage is where data value saris run ken to a central point as the mean. The lasso procedure encourages simple, straight forward, thin models (i.e. models with fewer parameters). This specific sty leof regressionis well-suited for models showing high levels of multiple regression or when you want to auto matisecert ainelements of model, like varia blese lection/parameter elimination.

2)   Random Forest Classifier:

Random Forest is a classifier that contains variety of decision trees on varied subsets of the givendatasetandtakestheaveragetoimprovethepredictiveaccuracyofthatdataset.Insteadofhopingonone   decision tree, the random forest takes the prediction from each and every tree and based on the majority votes of predictions, and it predicts the final output. The more the number of trees within the forest results in higher accuracy and prevents the matter of over fitting.

3)   Naïve Bayes Algorithm:

Naive Bayes is a simple technique for constructing classifiers: models that as sign class label lst o problem instances,   represented   as   vectors of feature values, where the class labels are drawn from some finite set. There is   not   a single algorithm for training such classifiers, but a family of algori thmsbase donacom mon principle: all naive Bayes classifiers assume that the value ofa particular feature is independent of the value of any other feature, given the class variable. Fore example, a fruit may be considered to be an apple if it is red, round, and about 10 cm in diameter. A naive Bayes classi fier conside rseach of these featuresto contribute independently to the probability that this fruitisan apple, regardless of

Any possible correlations be twin the color, roundness, and diameter features.

4)   Log is tic Regression Algorithm:

Log is tic regressionisa classi fication tech niquebor rowed by machine learning from the field of statis tics. Logistic Regress on is as tat is teal method for analyzing a dataset in which there are one or more in depend invariables that determine an out come. The intention behind using logistic regression is to find the best fitting model to describe the relationship bet wean the dependent and the in dependent variable.

IV.RESULTS

It is advantageousformanystakeholderstousemachinelearningtoanalysecricketgamesbytakingpreviousgamedata, player performance, natural parameters, pre-game conditions, and other feature es into account. Predicting the result of a game in adynamic format like T20, when the scenario in agame changes with every ball, is difficult. We have looked into machine learning technologies to see if it can increase the accuracy with which results of matches are predicted for T20 cricket matches. In order to better un darts and the issue, we have divide did into two scenarios: the Home Team features set and the Toss Winner determination features set.

The model constructed on Toss related feat uresyiel dsmar ginally better out comes than Home Advantage in terms of the assessment measure stylized, according to an analy sis of the result so brained using four different machine learning approacheson 10years' worth of T20matches(Accuracy, Precision, Recall, FPs, FNs, etc). When processing the Toss Winner feat reset, Lasso Regression, Random Forest Classifier, Naive Bayes Algorithm, and Logistic Regression Algorithm per formed better than the other algorithms because they

produced more accurate predictive models than Decision Trees, Probabilistic, and Statistical models. Further more, the number of occurrences that the a for ementionedal girths, both FPs and FNs, wrongly classi fyis low, leading to improved Precision and Recall rates. The a fore mentioned approach successfully identified 134 instances that were in correctly cite gorse dabs longing to the "Lose" class and 105 instances—or around 35%—were in correctly label ledas "Wins." The classing dependence assumption of the procedure, ho wever, makes the out comes of Nave Bayes on the Toss Decisionsubsetun promising. How ever, the Home Team subset yielded higher results using Nave Bayes. Team management and acade micsinte rested in cricket data analytics will help people to anal yes and bet well.

V.CONCLUSION

The goal of this research is to use past data to fore cast the final score and match winner. Data Pre-processing, Data Visualizations, Data Preparation, Data Selection, and Machine Learning Model Implementation are some of the fields of Data Science that will come to get her to conduct the study and fore cast the match's score. To accurately forecast the score of innings and obtain the desired out come, anumbero fmachine learning models will be applied to specified data

REFERENCES

[1]   Rabindra Lamsal and Ayesha Choudhary, "PredictingOutcome of Indian Premier League (IPL) Matches Using Machine Learning",arXiv:1809.09813"

[2]   Abhishek Naik, Shivanee Pawar, Minakshee Naik, Sahil Mulani, "Cricket Score Prediction using Machine Learning Algorithms"

[3]   "Cricket Score and Winning Prediction"

[4]   Dhonge, N., Dhole, S., Wavre, N., Pardakhe,M., &Nagarale, A (2021). "IPL Cricket Score and Winning Prediction Using Machine Learning Techniques"

[5]   Kumash Kapadia, Hussein Abdel-Jabber, Fade Thatch, Wael Hadi (2019)."Sport analytics for cricket game results using machine learning: A experimental study"