RESEARCH PAPER

Machine Learning-Based Cricket Match Winner Prediction

M.Manikanta, manikantamaguluri19@gmail.com

[5]Department of computer science and engineering, Kalasalingam university, Tamil Nadu, India

Abstract: : With the advent of statistical modelling in sports, predicting the outcome of a game has been established as a fundamental problem. Game consists of 11 player team sport played on ground. Cricket has huge fan base in India. With are great spectator support and many people try to predict the outcome of matches based on their individual cricket sense. The games has some rules and scoring system. Factors viz, match location and individual player performance have great impact on outcome of the match. Such various parameters are highly interdependent on each other which makes it heard to make precise prediction of the match. In this project, we are going to build prediction system that takes in data of matches played in past and makes a prediction of future match events such as final score and results in a gain or loss. Our system will predict match outcome by analysing pre-stored match data using various machine learning algorithms . We intend to use more features such as pitch condition, weather condition, outcome of toss, individual player performance with respect to match venue. Our system finally present quantitative results displayed by best suited algorithm having highest accuracy. Also, demonstrating the performance of our algorithms in predicting the number of runs scored which is one of the most important parameter of match outcome This work suggests that the relative team strength between the competing teams forms distinctive feature for predicting the winner. Modeling the team strength boils down to modeling individual player's batting and bowling performances, forming the basis of proposed approach. The career statistics as well as the recent performances of a player to model this. Player independent factors have also been considered in order to predict the outcome of match . The algorithm used are Decision Tree, Logistic regression and Support Vector Classifier (SVC) yields better results in the experimental evaluations.

Keywords: : Prediction ; Cricket Match outcome prediction; Machine learning techniques

INTRODUCTION

Statistical modeling has been used in sports since decades and has contributed significantly to the success on field. Cricket is one of the most popular sports in the hole world, second only to soccer. Various factors affecting the game, enormous media coverage, and a huge betting market have given strong incentives to model the game from various perspectives. However, the complex rules governing the game, ability of players and their performances on a given day, and various other natural parameters play an integral role in affecting the final outcome of a cricket match. This presents significant challenges in predicting the accurate results of game.

The game of cricket played in three formats - Test Matches, ODIs and T20s. We focus our research on ODI's, the most popular format of the game. To predict outcome of ODI cricket matches, we propose an approach where we first estimate the batting and bowling potentials of the 22 players playing the match using their career statistics and active participation is recent games. Since its inception the game has been very successful resulting in its spread around the cricket world and spawned many premier cricket league competitions such as the Indian Premier League. The following are some of the terminologies used in cricket: This template was designed for two affiliations..

PROBLEM DEFINITION:

Statistical modeling has been used in sports since decades and has contributed significantly to the success on the field. Various natural factors affecting the game, enormous media coverage, and a heavy betting market have

given strong incentives to model the game from various perspectives. However, the complex rules governing game, the ability of players and their performances on a given day, and various other natural parameters play an integral role in affecting the final outcome of a cricket match. This presents significant challenges in predicting the correct results of a game. To predict the outcome of ODI cricket matches, an efficient approach to be proposed using the statistics available from cricket websites. Taking some base features into account, namely, run rate and the venue of the match, along with the relative team strength, we adopt machine learning algorithms to predict the winner of the match.

Objective:
relative team strength, we adopt machine learning algorithms to predict the winner of the match.

1. Using different machine learning algorithms to get the good accuracy of the prediction outcome.
2. Prediction of the outcome of the result between the two teams.
3. It can also give the data of the players performance.

RELATED WORK:

A. Predicting the Outcome of ODI Cricket Matches

While team composition is predominantly based on the application of mathematical modeling in sports, predicting the outcome of a game is a fundamental flaw. Cricket is one of the most comprehensive team games on the planet's break. With this lesson, we have a tendency to start predicting one-day international (ODI) match results using a supervised learning approach from a team composition perspective. We have a tendency to use career statistics because player performance has modeled him recently. Players predict the outcome of the match along with the freelance factors.

B. Score and Winning Prediction in Cricket through Data Mining

Currently, the initial innings score supports a predetermination of an international (ODI) match for some purpose. The current run rate can be calculated, resulting in the number of runs bowled. It does not take into account the amount of wickets dropped and the venue for the match. There is no method of predicting match results between the second innings.

C. Player's Performance Prediction in ODI Cricket Using Machine Learning Algorithms

This paper presents a simplified methodology for assessing the performance of a cricket player's upcoming match by applying machine learning algorithms. The job model includes applied mathematical knowledge of Indian national cricket team players sourced from support vector machines with reliable sports websites, algorithm feature elimination and linear vector regression, linear and polynomial kernels. To implement the planned model, the applied mathematical knowledge collected is processed as a numerical value to be implemented between the algorithms. Furthermore, the aforementioned Feature Choice Algorithm applies the field unit to extract properties related to the Area Unit Output feature. In addition, the machine learning algorithm uses the field unit to predict the runs scored by a hitter and the bowler's thought between the coming match. The experimental setup model provides ninety-one-half% accuracy for the hitter Virat and seventy-five% for the bowler Jadeja.

MODULES DESCRIPTION
A.   Decision Trees

Decision Trees (DTs) are a non-parametric supervised learning method used for classification and regression. The goal is to create a model that predicts the value of a target variable by learning simple decision rules inferred from the data features. In decision tree, leaves represent class labels, internal nodes represent features and the branches represent the value of the feature in the internal nodes. Scikit-learn package provides a DecisionTreeClassifier which is the implementation for a decision tree.
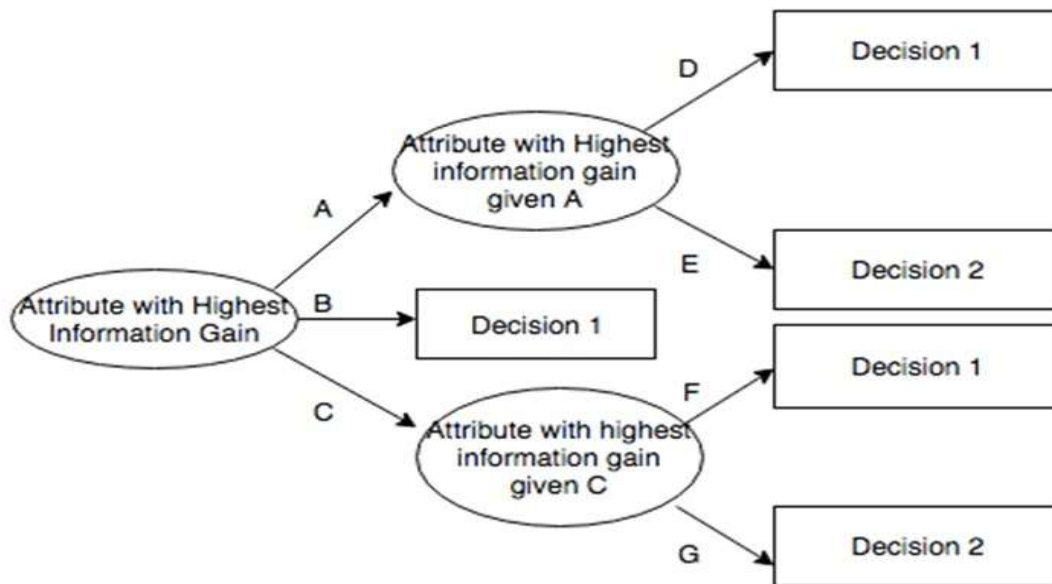
B.   Logistic regression

In simple, linear regression, predict scores on one variable from the scores on a second variable. The variable that predicted is called the criterion variable and is referred to as Y. The variable base for predictions on is called the predictor variable and is referred to as X. When there is only one predictor variable, the prediction method is called simple regression. In simple linear regression, the topic of this section, the predictions of Y when plotted as a function of X form a straight line.

The logistic function, also called the sigmoid function was developed by statisticians to describe properties of population growth in ecology, rising quickly and maxing out at the carrying capacity of the environment.

FLOW CHART OF LOGISTIC REGRESSION ALGORITHM
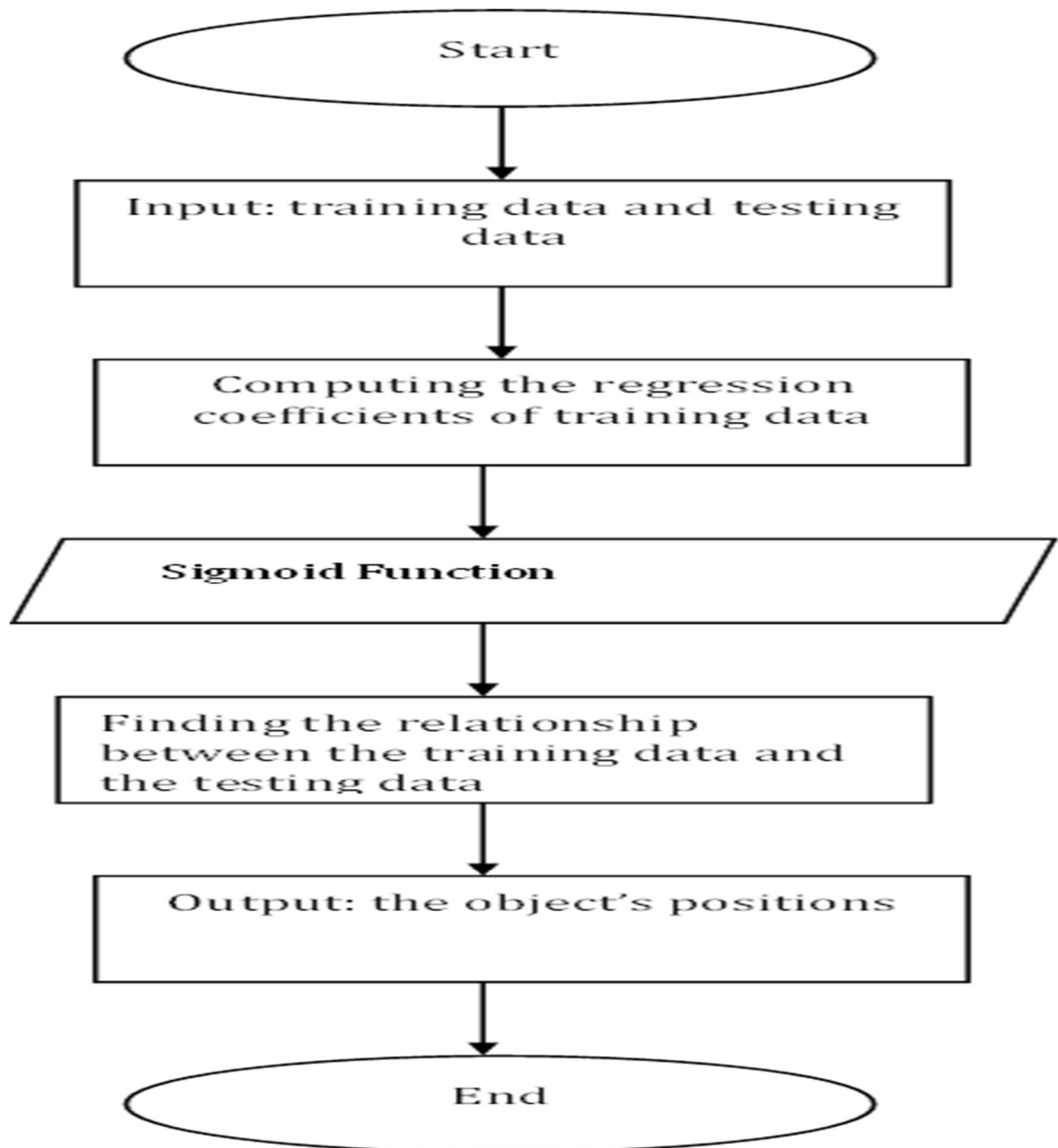
C. SVM ALGORITHM

Support Vector Machine SVM is a set of related supervised learning method used in medical diagnosis for classification and regression. SVM simultaneously minimize the empirical classification error and maximize the geometric margin. SVM is called Maximum Margin Classifiers and it can be efficiently perform nonlinear classification using kernel trick. An SVM model is a representation of the examples as points in space, mapped so that the examples of the separate categories are divided by a large margin gap that is as wide as possible. Given labeled training data as data points of the form.

Flow chart of SVM algorithm:
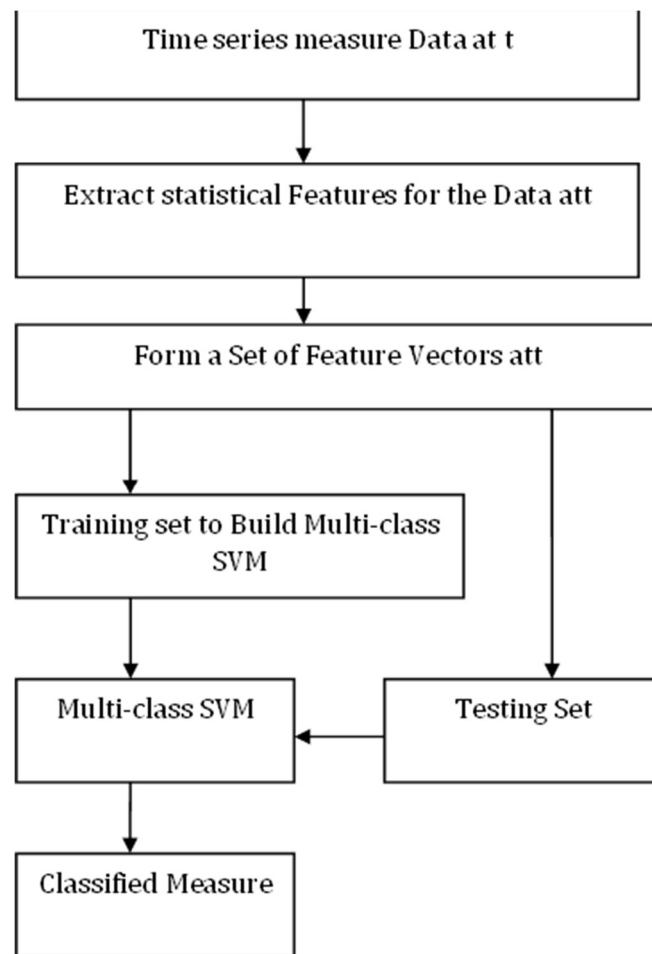
Existing System:

An extensive online search produced very few articles related to players' performance prediction in the game of cricket. The performance of Indian bowlers predicted against seven international teams against which the Indian cricket team plays most frequently. They used back propagation network and radial basis network function to predict how many runs a bowler is likely to concede and how many wickets a bowler is likely to take in a given ODI match. It is easy to apply, requiring nothing more than a single table of numbers and a pocket calculator, and is capable of dealing with any number of interruptions at any stage of either or both innings. The method is based on a simple model involving a twofactor relationship giving the number of runs which can be scored on average in the remainder of an innings as a function of the number of overs remaining and the number of wickets fallen. It is shown how the relationship enables the target score in an interrupted match to be recalculated to reflect the relative run scoring resources available to the two teams, that is overs and wickets in combination.

Proposed System:

Proposed system study on winner prediction using machine learning algorithm. The main aim is to predict the match outcome, performance of each player based on the historical data. In order to achieve a reliable accuracy, we need to analyse a large amount of data. Therefore, the initial step of the implementation was to collect data for all possible matches. Dataset is collected from various websites such as ESPN,kaggle,etc

Cricket Score Prediction using Machine Learning Algorithms.We collected data which provides ball by ball details for all the matches. Various analytical rules are used in order to filter collected data with respect to the selected feature.Features can be match venue, playing 11, weather condition, performance of individual player.Further, cleaned data is split into training (80%) and testing data (20%). Training data is fed to all machine learning model and accuracy of each model is noted. Model having highest accuracy is selected for further prediction.This model is used to make perdition over the input data provided, input data may include home team, opposition team, current weather condition and analysis done on historical data.

RESULTS ANALYSIS:

The proposed work is implemented in Python 3.6.4 with libraries scikit-learn, pandas, matplotlib and other mandatory libraries. The training dataset of espncricinfo data is take for anlaysis. Machine learning algorithm is applied such as Decision tree, and Random forest. We used these machine learning algorithm and indentified intrusion. The result shows that cricket match result prediction is efficient using Decision Tree algorithm.

	Algorithm	Accuracy (%)
1.	Decision Tree	86.56
2.	Logistic regression	76.56
3.	SVM	85.93

CONCLUSION AND FUTURE SCOPE:

FUTURE SCOPE

Traditionally, single algorithm is used in such prediction system and respective performance is measured. Instead, we intend to use multiple machine learning algorithms and relatively measure their performance**.**

CONCLUSION

The career figures as well as a player's recent performance in model this. To predict a match result, player-independent variables were also considered. The algorithm used are Decision Tree, Logistic regression and Support Vector Classifier (SVC) yields better results in the experimental evaluations. The advantage of using decision trees is fast and easy to implement, reusable, modular. The advantage of using logistic regression is that the target is discrete; works well with diagonal decision boundaries and can produce probabilistic outcomes.

REFERENCES:

[1] S. Muthuswamy and S. S. Lam, "Bowler Performance Prediction for One-day International Cricket Using Neural Networks", Proceedings of IIE Annual Conference and Expo 2008, pp.1391-1395/

[2] I. P. Wickramasinghe, "Predicting the performance of batsmen in test cricket", Journal of Human Sport & Exercise, vol. 9, no. 4, pp. 744-751, May 2014.

[3] G. D. I. Barr and B. S. Kantor, "A Criterion for Comparing and Selecting Batsmen in Limited Overs Cricket", Operational Research Society, vol. 55, no. 12, pp. 1266-1274, December 2004.

[4] S. R. Iyer and R. Sharda, "Prediction of athlete's performance using neural networks: An application in cricket team selection", Expert Systems with Applications, vol. 36, pp. 5510-5522, April 2009.

[5] M. G. Jhanwar and V. Pudi, "Predicting the Outcome of ODI Cricket Matches: A Team Composition Based Approach," in European Conference on Machine Learning and Principles and Practice of Knowledge Discovery in Databases (ECMLPKDD 2016 2016), 2016.

[6] Pandey, Niravkumar, "Predicting Players' Performance in the Game of Cricket Using Machine Learning" in Master of Science [7] Kalpdrum Passi and Niravkumar Pandey, "Predicting Players' Performance in One Day International Cricket Matches Using Machine Learning", Department of Mathematics and Computer Science, Laurentian University, Sudbury, Canada

[7] Passi, Kalpdrum & Pandey, Niravkumar. (2018) "Predicting Players' Performance in One Day International Cricket Matches Using Machine Learning" 111-126. 10.5121/csit.2018.80310.

[8] I. P. Wickramasinghe et. al, "Predicting the performance of batsmen in test cricket," Journal of Human Sport & Exercise", vol. 9, no. 4, pp. 744-751, May 2014.

[9] R. P. Schumaker, O. K. Solieman and H. Chen, "Predictive Modeling for Sports and Gaming" in Sports Data Mining, vol. 26, Boston, Massachusetts: Springer, 2010.

[10] J. McCullagh, "Data Mining in Sport: A Neural Network Approach," International Journal of Sports Science and Engineering, vol. 4, no. 3, pp. 131-138, 2012.

[11] Bunker, Rory & Thabtah, Fadi. (2017) "A Machine Learning Framework for Sport Result Prediction. Applied Computing and Informatics", 15. 10.1016/j.aci.2017.09.005.

[12] Ramon Diaz-Uriarte and Sara, "Gene selection and classification of microarray data using random forest, BMC Bioinformatics", doi:10.1186/1471-2105-7-3

[13] Rabindra Lamsal and AyeshaChoudhary, "Predicting Outcome of Indian Premier League (IPL) Matches Using

Machine Learning"

[14] Akhil Nimmagadda et. Al, "Cricket score and winning prediction using data mining", IJARnD Vol.3, Issue3.

[15] Ujwal U J et. At, "Predictive Analysis of Sports Data using Google Prediction API" International Journal of Applied Engineering Research", ISSN 0973-4562 Volume 13, Number 5 (2018) pp. 2814-2816.

[16] Rameshwari Lokhande and P.M.Chawan, "Live Cricket Score and Winning Prediction", International Journal of Trend in Research and Development, Volume 5(1), ISSN: 2394-9333.

[17]