CRICKET ANALYSIS USING MACHINE LEARNING

Vijaya Lakshmi.P

Assistant Professor,CSE

Sreyas institute of engineering and

Technology, Telangana, India.

Vijayalakshmi.p@sreyas.ac.in

Kanike Haripriya

CSE,Sreyas institute of engineering

and Technology, Telangana, India.

Kanikeharipriya1234@gmail.com

Saipavan Gangishetti

CSE,Sreyas institute of engineering

and Technology, Telangana, India.

saipavangangishetti@gmail.com

Thummalapalli VLB Gayatri

CSE ,Sreyas institute of engineering

and Technology, Telangana, India.

Gayatri.t@gmail.com

Snehith Yarlagadda

CSE ,Sreyas institute of engineering

and Technology, Telangana, India.

snehithsai602@gmail.com

ABSTRACT - This project uses machine learning algorithms to set the matter of predicting match results supported historical matches information.

ML is used to predict the match result variable by developing classification models based on certain freelance variables like player's position, their batting and bowling performances, weather, location, etc. However, we have a tendency to 1st analyze the info and train the models then it's straightforward to predict the match results.

We as a team utilize SVM and Random Forest machine learning algorithms to research and predict the result of the sport, and it'll facilitate the coaches of the team acquire knowledge and analyze, wherever really the team goes wrong and also the space of the development. the method involves preparing the model supported previous matches history, then the developed model gets evaluated on associate degree separate future match to calculate its effectiveness.

KEYWORDS: SVM (Support Vector Machine) Random Forest, Prediction.

I. INTRODUCTION

Indian Premier League (IPL) may well be a franchise system based totally, annual cricket tournament. IPL deals with tons of dollars. the number of money spent on the IPL teams imposes high on householders to appear victories, that depends on team performance. Primarily, it's essential to look out the right set of metrics that may cause assemble a team with the simplest likelihood of winning. This study makes a trial to identify the optimum set of attributes, which will facilitate team householders to look for players with these attributes to create a team by that they're going to enhance the win-ability of a cricket team. Most of the

present works targeted on characteristic fully completely different performance metrics supported their domain data of cricket. The planned resolution depends on mathematics analysis and machine learning whereas minimizing the use of domain data. Ball by ball knowledge for all past IPL matches were collected, mass to innings level details for the analysis and conjointly the downside is modelled as a classification drawback. the knowledge set contained a bunch of choices supported the innings level data and win/lose/draw category labels. Machine learning algorithms like Random Forest and SVM algorithms are used, and Random Forest rule achieved the foremost effective accuracy inside the analysis. Then, we tend to tend to examined all possible feature combos victimization machine learning algorithms by victimization k-fold methodology. Finally, the attribute set that yields the simplest accuracy inside the analysis is thought, which might be the optimum set of attributes that impose the high impact on the highest results of a match.

The Author of [3] discusses the analysis role of machine learning within the improvement of performances of players and also the team in several sports and the way wearable technology helps the players to understand their performance levels and any enhancements. The paper describes a system that collects information for every sport, team and player, and it's processed into applied math knowledge. These knowledge sets are clustered and hold on because the knowledge to be hold on is extremely massive

II. LITERATURE SURVEY

A literature survey or a literature review in a project report is that section which shows the various analyses and research made in the field of your interest and the results already published, taking into account the various parameters of the project and the extent of the project.

It is the most important part of your report as it gives you a direction in the area of your research. It helps you set a goal for your analysis - thus giving you your problem statement.
[4]. Kalpdrum Passi and NiravkumarPandey discussed the prediction accuracy in terms of runs scored by batsman and the no. of wickets taken by the bowler in each team. [5]. R.P. Schumaker, discussed different statistical simulations used in predictive situations.
[6]. I. P. Wickramasinghe proposed a methodology to predict the performance of batsman for the previous five years using hierarchical linear model.

III. PROPOSED SYSTEM

The main aim is to predict the match outcome supported the historical information. thus on attain reliable accuracy, we would like to analyse an immense amount of information. Therefore, the initial step of implementation was to collect information for all attainable matches. information set is collected from various websites like ESPN and Kaggle. we tend to collected information that gives ball by ball details for all the matches and in addition the define of the match in CSV format. a lot of the clean information is utilized to educate the model learning k-fold and accuracy of each model is noted. Model having highest accuracy is chosen for additionalprediction.

The best format to use to model the sport is T20, because of the huge range of matches held during a year. Also, T20 remains the toughest format to model because of its extremely unpredictable nature compared to ODIs or take a look at matches. Analysis of IPL information is performed by the author within the paper [1].

A.PROPOSED ARCHITECTURE

A system design diagram would be accustomed show the link between completely distinct components. usually they're created for systems that embody hardware and software package and these are portrayed inside the diagram to point the interaction between them. However, it will even be created for web applications.
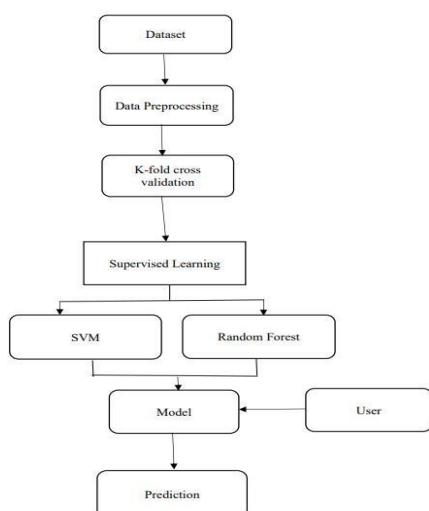

Fig.III.A.1.System Architecture

Dataset is collected from various websites like ESPN, Kaggle, etc. Later the knowledge ar normalized, and it's splitted to learning and testing information by k-fold cross validation technique.

By training the models we'll get the accuracy of the dataset, that helps in giving fascinating outcome.

This analysis is employed by the team for framing game-winning plans. In [2] the author shows that a much bigger information set will improve the accuracy of the prediction.

B. IMPLEMENTATION

This model offers a lot of accuracy of prediction and calculation of score by taking numerous fields like toss winners, players, venue, DL methodology etc. Here we tend to area unit using Random forest algorithm that works on creating decision trees to administer correct result. SVM wherever it helps in differentiating algorithm.

1)Random forest rule

2)SVM algorithm

1) Random forest algorithm

Random Forest could be a supervised machine learning methodology that operates by constructing multiple decision trees. it's composed of various decision trees, each with the same nodes, but using completely different knowledge that results in different leaves. It merges the selections of multiple decision trees so as to search out a solution, that represents the average of of these decision trees. The final decision is created based on the majority of the trees and is chosen by the random forest. The below fig. 3 depicts how random forest rule merges the outputs and offers final decision.
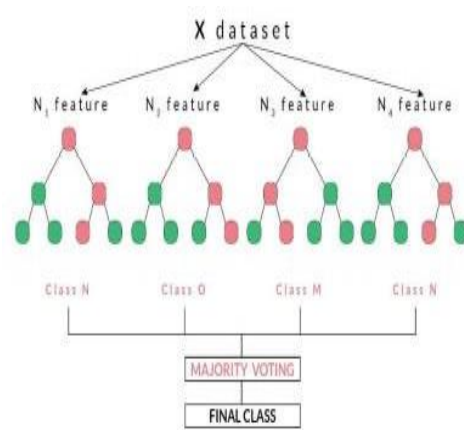

Fig.III.B.1.1.RandomForest Algorithm

2)SVM algorithm

Support Vector Machine (SVM) could be a supervised machine learning rule which may be used for each classification or regression challenges. Howeverit is generally utilized in classification issues. In this algorithm, plot each data item is plotted as a degree in ndimensional space (where n is variety of features) with the value each

feature being the value of a specific coordinate. Then, classification is performed by finding the hyper-plane that differentiate the 2 classes very well. This hyperplane is made with the help of data items that are nearer to it. Based on support vectors, the hyperplane or margin is made.
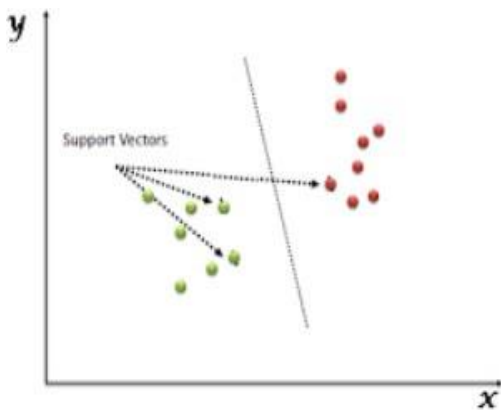

Fig.III.B.2.1.SVM(Support Vector Machine)

Inorder to administer best accuracy of prediction and calculation of score our projected model undergoes four steps includes Data Collection, Data PreProcessing, Feature Extraction, Evaluation Model.

Data Collection- Data utilized in this project is collected from Kaggle.com in google dataset engine.

Data Pre-Processing- The pre-processing refers to the transformations applied to the data before feeding it to the algorithm.

Feature Extraction- Feature extraction is an attribute reduction method.

Evaluation Model- Model evaluation is an integral part of the model development method. It helps to search out the best model that represents data and how well the chosen model will work.

IV.   CONCLUSION

Selection of the best team for a match plays a major role for the team's finish. The main goal of our is to analyze the IPL cricket data and predict the end result of a match. Here, 3 classification algorithms ar used and compared to seek out the best correct algorithm.

The implementation tools used are anaconda navigator and Jupyter. Random Forest is observed to be the best correct classifier with 85.582% to predict the end result of a match. Hence using this prediction, the best team can be formed.

V.   FUTURE SCOPE

The project currently takes into consideration the vital factors from previous matches then predicts the result before the match starts. But we can also take the ongoing match details to predict the result.

So the future work will be to combine both the previous knowledge with current match knowledge to possess far more higher results. Also the prediction can be extended to not only predict the winner of the match but also the expected runs to be scored by both teams.

VI.   REFERENCES

[1] Harshit Barot, Arya Kothari, Pramod stay, Bhavya Ahir, Romit Kankaria," Analysis and Prediction for the Indian Premier League",2020 International Conference for emerging Technology (INCET) Belgaum, India. Jun 5-7, 2020.

[2] Manuka Maduranga Hatharasinghe, Guhanathan Poravi"Data Mining and Machine Learning in match Outcome Prediction: Missing Links"2019 cricket match International Conference for Convergence in Technology (I2CT) Pune, India. Mar 29-31, 2019.

[3] P. Sri Harsha Vardhan Goud, Y. Mohana Roopa, B. Padmaja," Player Performance Analysis in Sports: with Fusion of Machine Learning and wearable Technology" Proceedings of the Third International Conference on Computing Methodologies and Communication (ICCMC 2019) IEEE Xplore part Number: CFP19K25-ART; ISBN: 978-1-5386-7808-4

[4] Passi, Kalpdrum & Pandey, Niravkumar. (2018) "Predicting Players' Performance in in some unspecified time in the future International Cricket Matches learning Machine Learning" 111126.10.5121/csit.2018.8031.

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

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