IJMTST is UGC Approved Journal with Journal Id: 43137

International Journal for Modern Trends in Science and Technology, 8(S03):16-21, 2022

ISSN: 2455-3778 online
DOI: https://doi.org/10.46501/IJMTST08S0305
Available online at: https://www.ijmtst.com/vol8si03.html

Efficient Data Analysis using Python with Machine Learning

Vandana Peter | Srivasu | Sandeep | Divya| Eswar

Department of Computer Ccience and Engineering, Godavari Institute of Engineering and Technology(A), JNTUK, Kakinada.

To Cite this Article
Vandana Peter, Srivasu, Sandeep, Divya and Eswar. Efficient Data Analysis using Python with Machine Learning.International Journal for Modern Trends in Science and Technology 2022, 8(S03), pp. 16-21.https://doi.org/10.46501/IJMTST08S0305

Article Info
Received: 26 April 2022; Accepted: 24 May 2022; Published: 30 May 2022.

ABSTRACT

Selection of players is one of the most important aspects of any sport, and cricket is no exception. The competition group, the venue, his current physical condition, and other elements all influence the gamers' overall performance. From a roster of 15 to twenty players, the crew management, coach, and captain select 11 individuals for each in shape. They look at characteristics anddetails about the players to select the ideal betting 11 for each match Each batter contributes by scoring as many runs as feasible, while each bowler contributes by taking as many wickets as possible while allowing the fewest runs possible. This programme seeks to forecast player performance for each team, such as how many runs each batsman will score and how many wickets each bowler will take. Both problems are classified as type problems, with a wide range of runs and a wide variety of wickets being classified at exceptional levels. For both challenges to build prediction styles, we used multiclass SVM and decision tree classifiers. SVM was discovered to be the most accurate for all the issues.

KEYWORDS: Configurable accuracy, Approximate multiplier, area efficiency,low power,scalable, clipping

1.INTRODUCTION

Data mining is an artificial intelligence field that tries to solve real-world engineering problems. It allows you to examine without being expressly coded, and it's completely built on the concept of learning from information. It's so common that we don't even notice it a dozen times a day. System learning (ML) approaches have the advantage of using mathematical models and heuristic learning, acquisition comprehension, and selection timber for selection. As a result, it provides control, observability, and stability. Cricket is played in a variety of formats, including future global, T20, and test matches. The Indian Premier League (IPL) is a Twenty-20 cricket tournament league in India that was founded with the purpose of popularising cricket and nurturing younger and more skilled players.

The league is an annual event in which teams representing various Indian regions compete against one another. An auction is used to choose the IPL teams. In the realm of international sports, player auctions are nothing new. For the first time in India, however, a team was selected from a pool of available players through an auction in the Indian most effective League (IPL). The outcome of matches is vitally crucial to everyone because money, team spirit, local loyalty, and a vast fan base are all involved.

This, in turn, is determined by the sport's complicated rules, the group's luck (Toss), the players' capacity, and their individual performances on any given day. Various natural characteristics, including ancient Information about the players is crucial in determining the outcome of a cricket match. Within the crew selection approach,

there is a method of forecasting the outcomes of fits between distinct groups. However, the multiple characteristics involved provide significant challenges in accurately predicting game outcomes. Furthermore, the precision of a prediction is determined by the number of records used. This research instrument can be used to evaluate player performance. This device allows you to visualize performance.

Due to the following difficult settings, this artwork is unique. At some point during the IPL matches, the evolved fashions might assist decision makers in evaluating the collective power of a group as compared to another. The following are the contributions of the provided paintings:

• To give a statistical examination of gamers according to various characteristics.

• To forecast a crew's performance based on character participation statistics.

• To accurately forecast the ultimate results of IPL matches 2.

2. REALATED WORK

- Naive Bayes classification is a subset of straightforward "probabilistic classifiers" that are using Bayes' theorem and assumptions of high (naval) independence for many of the characteristics. They are the simplest Bayesian network. designs, however, when combinedthey can attain higher levels of accuracy using kernel density estimation. Especially scalable are Nave Bayes classifiers, as they require some linear parameters inside in a learning process, there are several variables (features/predictors). Most-chance training is possible with the assistance of assessinga linear-time expression with a closed form, rather than with the aid of employing high-priced iterative approximation, which is employed for many other types of classifiers.[16]Naive Bayes models are described in the facts and computer generation literaturealongside simple Bayes and independence Bayes, under various names. Those titles allude to the classifier's choice rule using Bayes' theorem; however, nave Bayes is not always (always) a Bayesian approach.Models that allocate elegance labels to difficulty instances as vectors of function values, with the elegance labels chosen from a limited collection, are built using Naive Bayes. [17]There isn't an unmarried set of rules for education such classifiers, however a circle of relatives of algorithms primarily based totally mostly on a not unusualplace principle: Given the beauty variable, all naive Bayes classifiers assume that the charge of one feature is independent of the charge of another. An apple, for example, is a purple fruit that is round and around 10 centimetres in diameter. A naive Bayes classifier considers each of these abilities to contribute independently to the possibility that this fruit is an apple, regardless of any possible correlations between many of the colouring, roundness, and diameter variables. For a few types of possibility models, naive Bayes classifiers can be trained relatively accurately in a supervised analysis scenario. Estimation of parameters the strategy is used for naïve Bayes models. of most chance in many practical applications, in unusual cases, you can work with the naïve Bayes model while avoiding Bayesian risk. or using any Bayesian procedures. Despite its naïve design and oversimplified assumptions,in a variety of difficult real-world settings, naive Bayes classifiers have performed admirably.In 2004, a study of the Bayesian kind problem revealed that the supposedly improbable the efficacy of naive Bayes classifiers is supported by sound theoretical foundations. Nonetheless, a comprehensive evaluation with several type algorithms in 2006 found that Bayes elegance is outperformed by applying various techniques, such as boosted bushes or random forests. Risks of contemporary systems:
  - Much less accuracy in prediction.
  - Time ingesting procedure.
  - Now not character friendly.
  - No proper modern-day updation.
  - Can't be applied in all datasets.

3. PROPOSED WORK

Support Vector Machine (SVM):

The idea of a guide vector system was added by Vladimir Vapnik, Bernhard Boser, and Isabell Guyon. SVMs are particularly accurate and resistant to overfitting. Each numeric prediction and classification can benefit from SVMs. SVM employs a nonlinear mapping to turn real statistics into a more suitable size. It then looks for a linear advanced hyperplane in this new size, keeping the tuples of one class apart from the tuples of another.

Tuples from classes can be continuously separated using a hyperplane with the necessary mapping to a sufficiently large size. The well-known set of rules demonstrates the employment of manual vectors and margins described by resource vectors in this hyperplane. The resource vectors discovered using the set of policies provide a concise description of the discovered prediction model. Blessings of the proposed device:

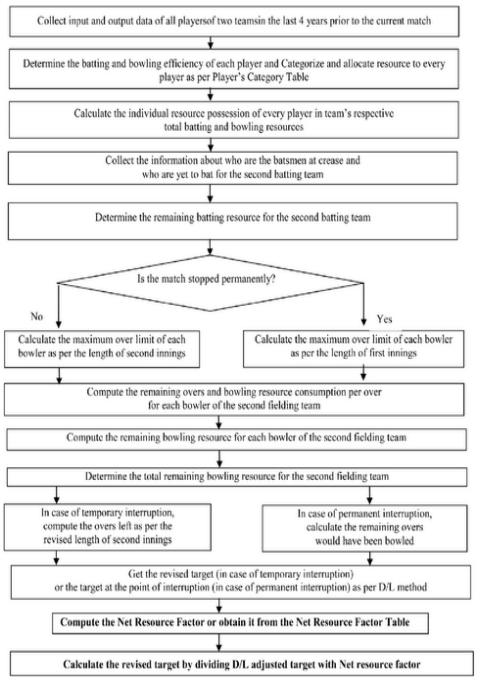
- More accuracy in prediction.
- Much less time consuming manner.
- Character pleasant.
- Proper contemporary-day updation.
- Can be carried out in all datasets.

ALGORITHM OF PROPOSED WORK

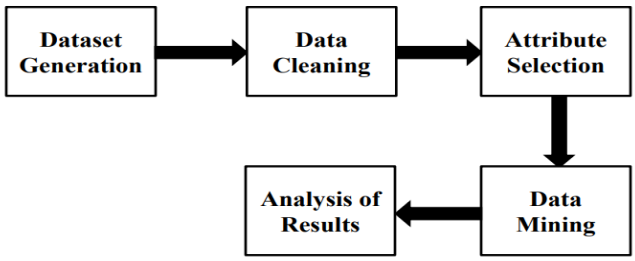
To learn how to apply the SVM technique, follow the steps below:

- To train the model, we start with all n variables in our dataset.
- The model's performance is then calculated.
- We now calculate the model's performance after deleting each variable (n times), i.e., one variable at a time and training the model on the remaining n-1 variables.
- We drop the variable that caused the smallest (or no) change in the model's performance when it was removed.
- Continue until no more variables can be dropped.

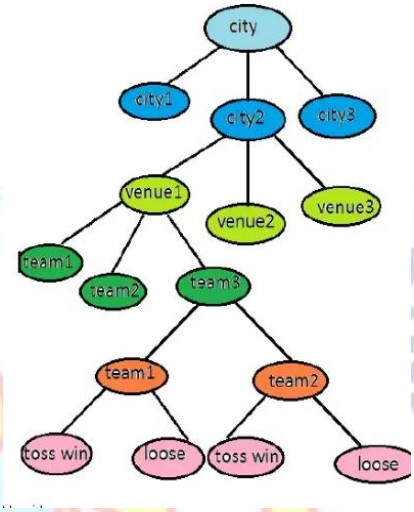
4.2 FLOWCHART



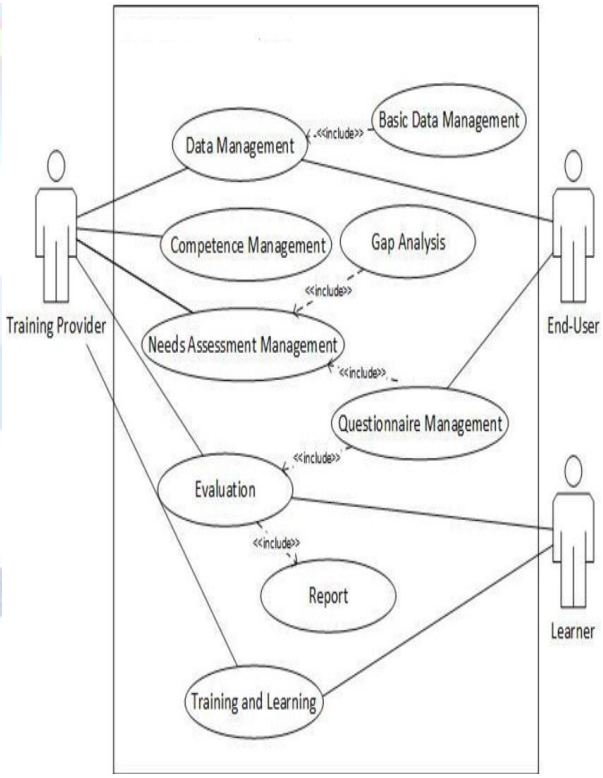
3.2 SYSTEM ARCHITECTURE



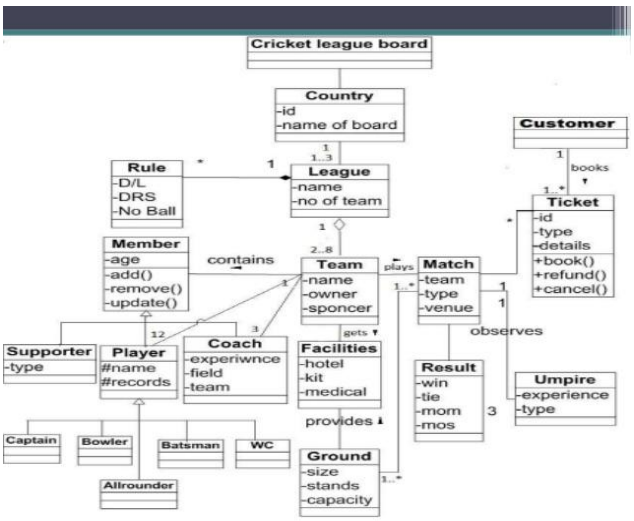
3.3 UML DIAGRAMS



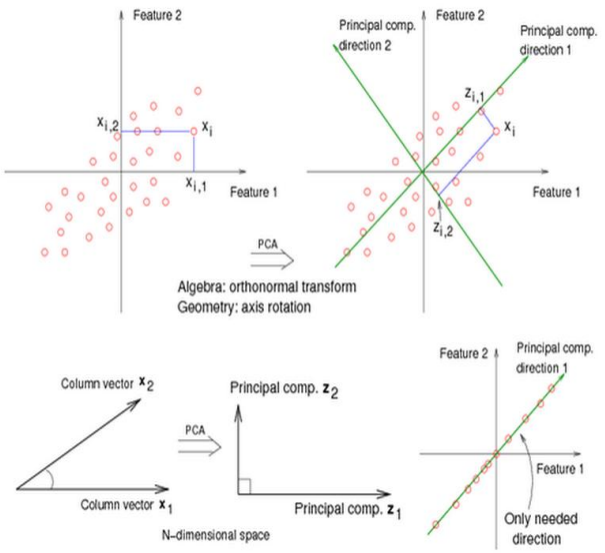
USE CASE DIAGRAM



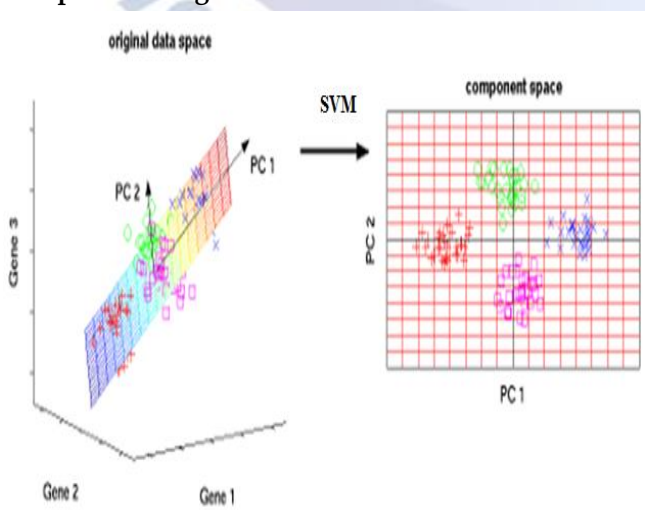
CLASS DIAGRAM



Deployment Diagram



Component Diagram

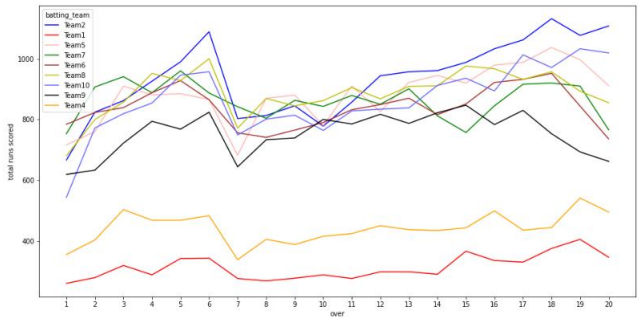


4. RESULTS

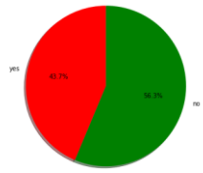
Out[16]:

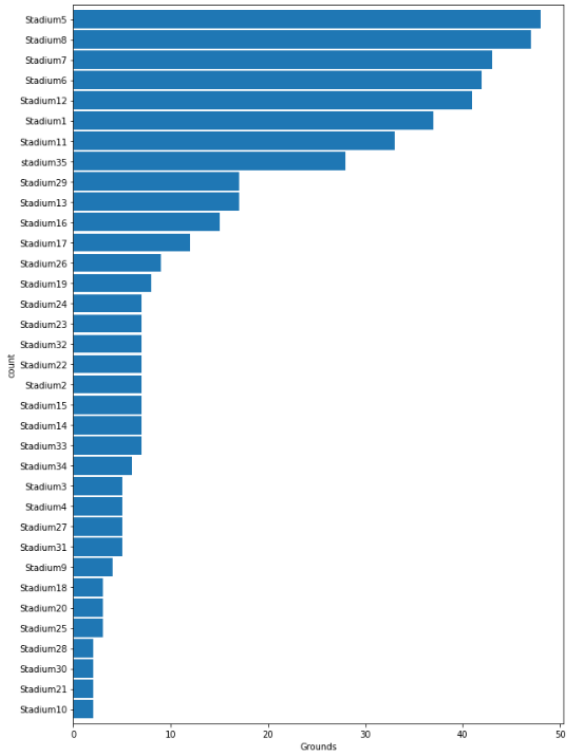
match_id	match_id	Team_A_Batting_Average	Team_A_Bowling_Average	Team_A_Total_Runs	Team_B_Batting_Average	Team_B_Bowling_Average	Team_B_Tota
1	1	3.57	-0.250000	207	2.54	-1.835714	
2	2	2.82	-1.471429	184	3.25	-0.692857	
3	3	3.09	-2.064286	183	3.13	-1.121429	
4	4	2.59	-0.764286	163	2.48	-0.835714	
5	5	2.24	0.228571	157	1.96	0.042857	
...	...	...	...	...	...	...	
496	496	3.18	-1.657143	205	3.70	-1.314286	
497	497	1.98	-0.378571	141	2.18	0.071429	
498	498	3.21	0.085714	201	1.84	-1.371429	
499	499	2.06	0.171429	138	1.89	-0.214286	
500	500	1.82	-0.414286	142	2.40	0.271429	

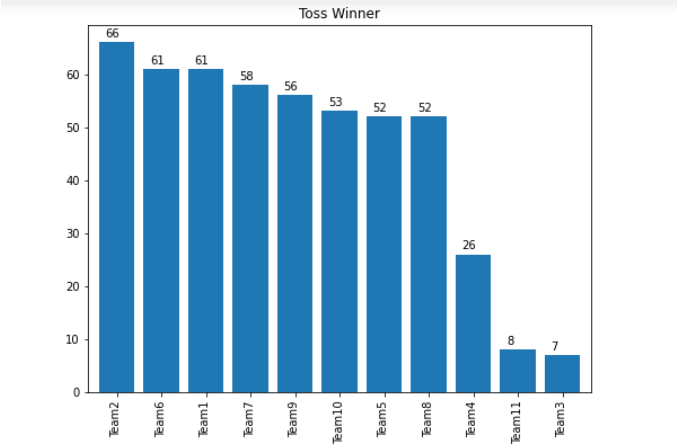
500 rows × 10 columns

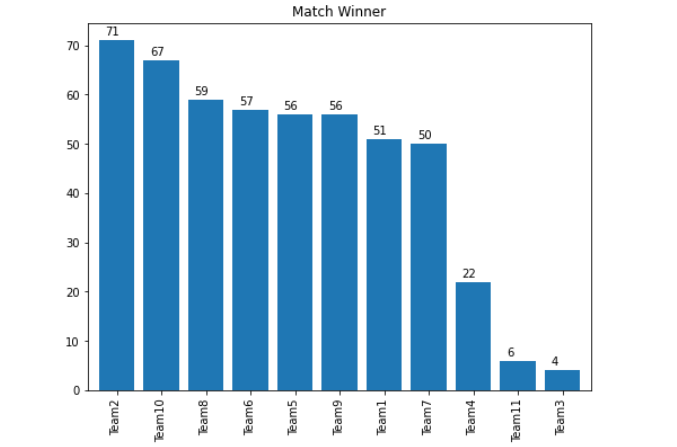



```
In [18]: df=TRM[TRM['toss_winner']==TRM['winner']]
         slices=[len(df),(577-len(df))]
         labels=['yes','no']
         plt.pie(slices,labels=labels,startangle=90,shadow=True,explode=(0,0),autopct='%1.1f%%',colors=['r','g'])
         fig = plt.gcf()
         fig.set_size_inches(6,6)
         plt.show()
```





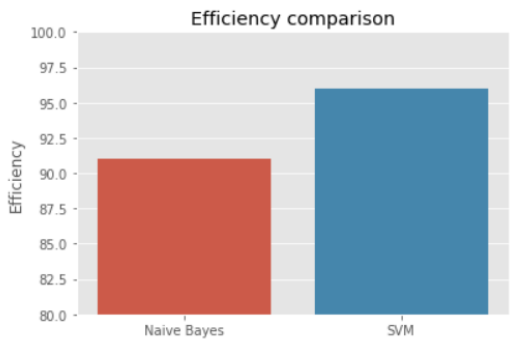




	match_id	Team_A_Batting_Average	Team_A_Bowling_Average	Team_A_Total_Runs	Team_B_Batting_Average	Team_B_Bowling_Average	Team_B_Tota
match_id							
1	1	3.57	-0.250000	207	2.54	-1.835714	
2	2	2.82	-1.471429	184	3.25	-0.692857	
3	3	3.09	-2.064286	183	3.13	-1.121429	
4	4	2.59	-0.764286	163	2.48	-0.835714	
5	5	2.24	0.228571	157	1.96	0.042857	
...	...	...	...	...	...	...	
496	496	3.18	-1.657143	205	3.70	-1.314286	
497	497	1.98	-0.378571	141	2.18	0.071429	
498	498	3.21	0.085714	201	1.84	-1.371429	
499	499	2.06	0.171429	138	1.89	-0.214286	
500	500	1.82	-0.414286	142	2.40	0.271429	

500 rows × 10 columns

Out[25]: (80.0, 100.0)



5. CONCLUSION

Choosing the best team for a cricket match is critical to the success of the team. The main purpose of this study is to study IPL cricket statistics and predict player performance.To get the most accurate results, the SVM algorithm is employed and compared. Anaconda navigator and Jupyter were utilised as implementation tools. SVM has been found to be the most accurate classifier to Predict the best player performance. This information will be used to predict the winning team of the IPL series in the future. As a result, the best team can be built using this forecast.

Future Scope

There is some potential future improvements to this project .

- The dataset contains the following external factors based on this data: B. Injury to a player, tiredness to a player, a team's winning streak, the overall winning streak, and the average run the team recorded for a particular team in the previous game. It may contain several factors. Make a prediction and see if the accuracy improves
- In addition to using general data about various games such as: Forecasts such as draw winners, draw decisions, home teams, etc. can also consider the performance of the team's players The player's overall number of runs in the event, his or her form, and the number of Men of the Match awards won, and more.

My project has no web/mobile application or user interface. As a result, a web/mobile application that takes in the entire informational collection as information and results the forecast outcome for each example to a pdf or text document might be made.

Conflict of interest statement

Authors declare that they do not have any conflict of interest.

REFERENCES

[1] S. Muthuswamy and S. S. Lam, "Bowler Performance Prediction for One-day International Cricket Using Neural Networks," in Industrial Engineering Research Conference, 2008.

[2] I. P. Wickramasinghe, "Predicting the performance of batsmen in test cricket," Journal of Human Sport &Excercise, 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 athletes 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] H. H. Lemmer, "The combined bowling rate as a measure of bowling performance in cricket," South African Journal for Research in Sport, Physical Education and Recreation, vol. 24, no. 2, pp. 37-44, January 2002.

[7] D. Bhattacharjee and D. G. Pahinkar, "Analysis of Performance of Bowlers using Combined Bowling Rate," International Journal of Sports Science and Engineering, vol. 6, no. 3, pp. 1750-9823, 2012.

[8] S. Mukherjee, "Quantifying individual performance in Cricket - A network analysis of batsmen and bowlers," Physica A: Statistical Mechanics and its Applications, vol. 393, pp. 624-637, 2014.

[9] P. Shah, "New performance measure in Cricket," ISOR Journal of Sports and Physical Education, vol. 4, no. 3, pp. 28-30, 2017.

[10] D. Parker, P. Burns and H. Natarajan, "Player valuations in the Indian Premier League," Frontier Economics, vol. 116, October 2008.

[11] C. D. Prakash, C. Patvardhan and C. V. Lakshmi, "Data Analytics based Deep Mayo Predictor for IPL- 9," International Journal of Computer Applications, vol. 152, no. 6, pp. 6-10, October 2016.

[12] M. Ovens and B. Bukiet, "A Mathematical Modelling Approach to One-Day Cricket Batting Orders," Journal of Sports Science and Medicine, vol. 5, pp. 495-502, 15 December 2006.

[13] 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.

[14] M. Haghighat, H. Ratsegari and N. Nourafza, "A Review of Data Mining Techniques for Result Prediction in Sports," Advances in Computer Science: an International Journal, vol. 2, no. 5, pp. 7-12, November 2013.

[15] J. Hucaljuk and A. Rakipovik, "Predicting football scores using machine learning techniques," in International Convention MIPRO, Opatija, 2011.

[16] Parvathi, D. S. L., Leelavathi, N., Ravikumar, J. M. S. V., & Sujatha, B. (2020, July). Emotion Analysis Using Deep Learning. In 2020 International Conference on Electronics and Sustainable Communication Systems (ICESC) (pp. 593-598). IEEE.

[17] Kumar, J. R., Sujatha, B., &Leelavathi, N. (2021, February). Automatic Vehicle Number Plate Recognition System Using Machine Learning. In IOP Conference Series: Materials Science and Engineering (Vol. 1074, No. 1, p. 012012). IOP Publishing."