

 Marwadi University	Marwadi University Faculty of Technology Department of Information and Communication Technology	
Subject: Machine Learning (01CT1509)	Topic: Dynamic cricket match outcome prediction	
Research Paper	Date:	Enrollment no.: - 92110133003, 92000133016

Dynamic cricket match outcome prediction

Rajesh Goel^{a,*}, Jerryl Davis^b, Amit Bhatia^c, Pulkit Malhotra^d, Harsh Bhardwaj^e,
Vikas Hooda^f and Ankit Goel^g

^a*Iris Software Solution pvt Ltd*

^b*Bain Capability Center*

^c*EXL Services India Private Limited*

^d*Nagarro Software pvt. Ltd*

^e*Conduent Technologies*

^f*Hewlett Packard Enterprise Ltd*

^g*Assistant Professor ECE Department, KIET Group of Institutions Delhi-NCR Ghaziabad*

Abstract. In abstract they are trying to explain if we want to predict an IPL match like between two teams which team will won this IPL match so we can predict it. Here they are tracking all the records of batsman and bowlers and its data dynamically change in each ball of each over throughout the match. For predicting the match they tried Logistic Regression, Naïve Bayes, KNN, Support Vector Machine, Decision Tree, Random Forest, Boosting, Bagging and Gradient Boosting with an accuracy of 76.47% (+/-3.77%). With the help of Deep learning, they tried the various flavours of LSTM and GRU, and they train models and get result with an accuracy of 76.13% (+/-2.59). They tested all approaches such as one to one, one to many, many to one and many to many sequencing and they also discussed in their research paper. For to get accurate answers like batsman score, how much bowler take wickets, etc like this it can help to make a good team.

Keywords: Long short-term memory, gated recurrent unit, machine learning, deep learning, IPL, prediction

1. Introduction

In the introduction part they try to explain first rules of the game and list out the formats of cricket played. They also explain on two games first Indian Premier League (IPL) and Twenty20s. (IPL/Twenty20) is embark on predict dynamic outcome of a (IPL/Twenty20) cricket match. (IPL/Twenty20) of match is comparatively shorter than other formats so that's why all are taking data of (IPL/Twenty20) matches. They are also introduce Board of Control for Cricket in India (BCCI) which they control all the venues of (IPL/Twenty20) matches.

*Corresponding author: Rajesh Goel, Iris Software Solution pvt Ltd. E-mail: rajeshgoel.88@gmail.com.

After all of this they explaining how cricket game can play. Toss winning team can wish either they want batting or fielding. After tossing first inning and team will try to make as many runs as possible and second inning and team will try to chase the target to make runs in order to win this game.

2. Related work

By the multivariate regression based solution they evaluating points based on the player past performance. They find seven factors to identified dataset of the model and start to predict match by taking previous matches and live match data and we will get outcome as win or loss. They are trying and applying and doing combination of linear regression and nearest neighbour cluster in the model and by various models using supervised learning method using SVM model with linear and nonlinear poly and RBF kernels algorithms to predict outcome of match. They also do like recommendation of the players for any specific match. Tandon, A. S. Varde and G.de Melo also gave a brief overview on the state of common sense Knowledge in Machine Intelligence and also discussion of open issues. They provide a report of a tutorial and also give brief survey of topics. L. Van den Berg, B. Coetzee, S. Blignaut and M. Mearns et al. study indicated that easily available sources are not effectively in data structure and analysis. They also said that we are lack in technology and their is an absence of a person who is collecting data, shorting the data, collecting live data like that person is responsible for all changes. In G. Melo et al., 2015 their is new scalable techniques for knowledge discovery. They also focus on cloud data mining and machine learning, semi-supervised processing and deep learning. They also given advice on practical base like choosing different methods and discuss and searching problems and concerns. They using data mining tools and techniques in sports, for particular mining the sports match. There are two research base answer give. Firstly, business data mining technique applied in sports games. Secondly analytical and time-consuming exercise suited to the sport world. They are taking match data for the South African cricket team and findings stakeholders in the South African team for data mining exercise. Many data found that data mining tools and techniques can use for unknown patterns in sport match data and it is successfully done. P. Basavaraju and A. S. Varde et al., 2017 gives some review on supervised learning approaches and it is also implemented in mobile apps and they 50% focusing on Androids for the global smartphone market. They also give some description on approaches and portrays and interested for to deploying Android apps, addressing classification and regression problems. In P.U. Maheswari et al, in that they discussed about on an automated framework for to identifying specifics and correlations as per player play

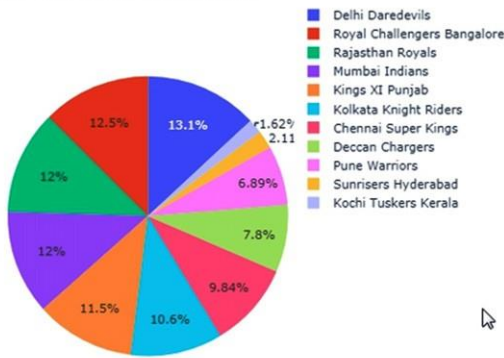
pattern, as per haul knowledge they said an useful information can be modify or improve as per coaching strategies and methodologies increase as at team level or individual. As per information a coach can formulate game strategy for subsequent games. In Kaluarachchi they discuss on winning and losing in ODU as per IPL we are doing. They use classical Bayesian classifiers for prediction in ODU where we use deep neural network to find best prediction.

3. Materials and methods

3.1. Data mining

Here we adopting different approaches so we can get our best predictions outcome. They said prediction is based on second batting team and if team will win or lose is base on the match circumstances. Previous matches data get from the cricket fan site cricinfo.com⁴ for the seasons 2008 to 2009. It is in RAW data and it is cleaned. From these type of dataset features sets are cleared so we can easily constructed our model using automated excel macros. We have two datasets as per formulated input rules in excel we have past data of IPL matches and they use different feature model total 90 attributes they created in the excel macros.

Playing Team wise distribution



Winning Team wise distribution

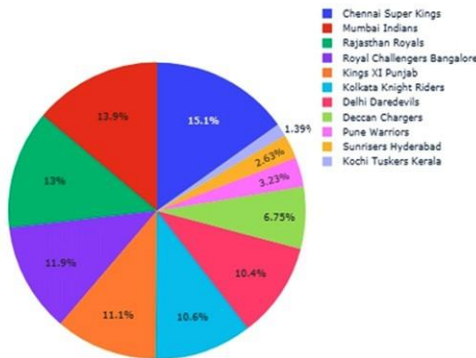


Fig. 1. Exploratory Data Analysis on the constructed features.

3.2. Feature selection

They taking selected test dataset of 59 matches of season 2019. There is 7000 data points for prediction. They have total 90 features but still model gives poor results. They use wrapper method for feature selection. Wrapper feature selection methods creates multiple models with different subsets of input features and select best model as per performance metric. They got 30 attributes for their best fitted model. From the Table 1, they got dependent and independent variables, independent available datasets of matches and deliveries, here they constructed 11 batsmen and 10 bowlers metrics and they tracking each ball over through the second inning match. "Target" score is independent attribute because it is dynamically change in every next inning. Now they are splitting data into train dataset and test dataset bases on season. They perform K-Fold Time Series Split to train dataset not use in model. Bias-variance trade-off is property of a set of predictive models with lower and higher bias variance of the parameter estimates across samples and vice versa. Final predictions got from

test data and model performance is evaluated.

4. Results and discussion

Below table shows notation which we followed for our model nomenclature –

4.1. MODEL 1 (1–20 over)

59 matches of season 2019 are tested and 692 matches of all season 2008–2018 are train model. Train size is 79,129 and test size is 6,960. By SVC and Random Forest Classifier show really good results where accuracy is 76.47% with interval of 3.77% and get test accuracies 72.74%. Both the models have very high precision of 83.25% and 81.90%. XGBClassifier also get very high precision of 83.92%. They said one to one sequence vanilla LSTM show good accuracy. Recall and F1 score are comparatively higher than traditional models. Gaussian and Bernoulli Naive Bias show very less train and test scores. Assumption attributes are independent and attributes are continuous in nature. Below comparative bar graph (Fig. 4) we can visualize our result.

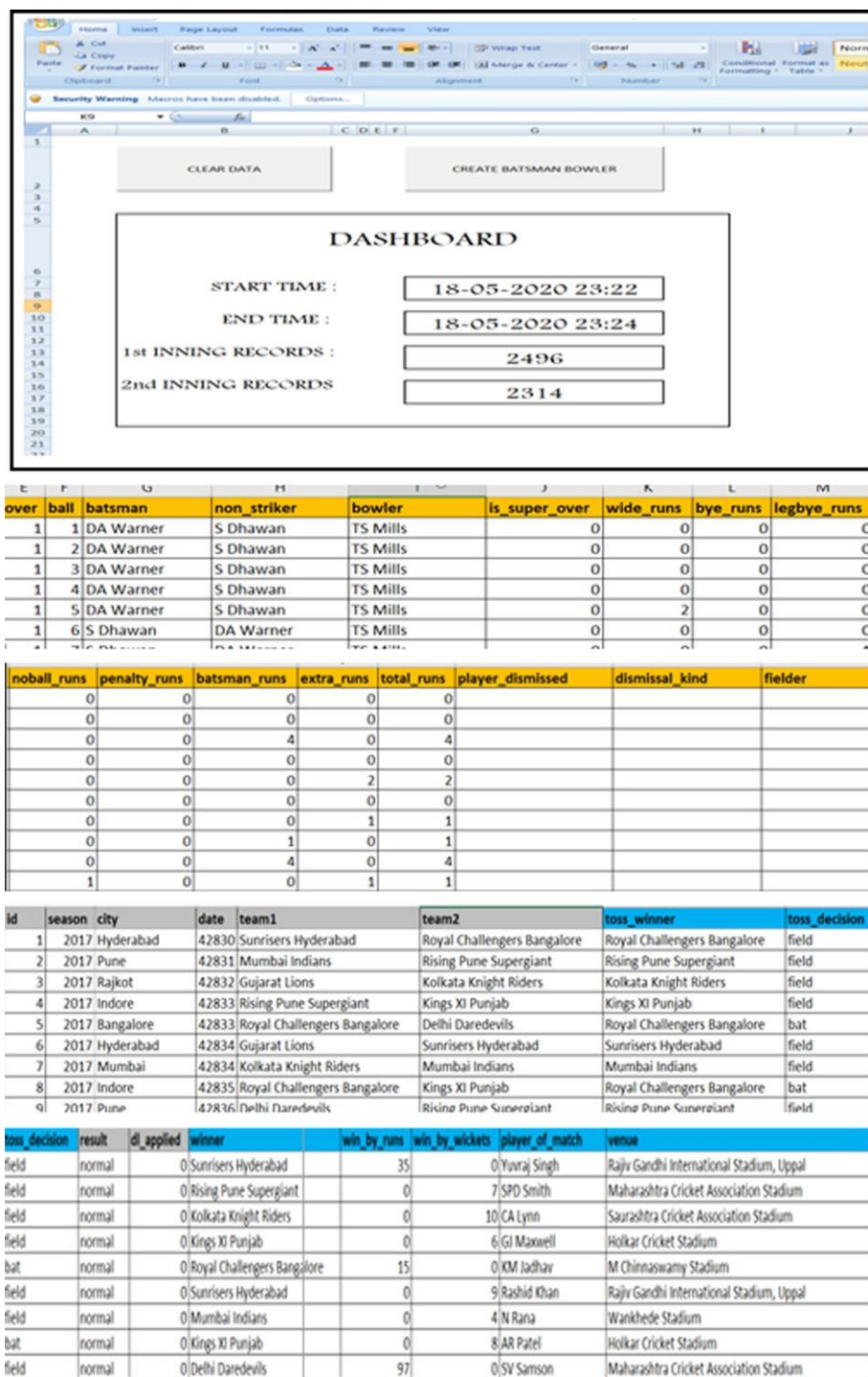


Fig. 2. Excel Macros for features constructions.

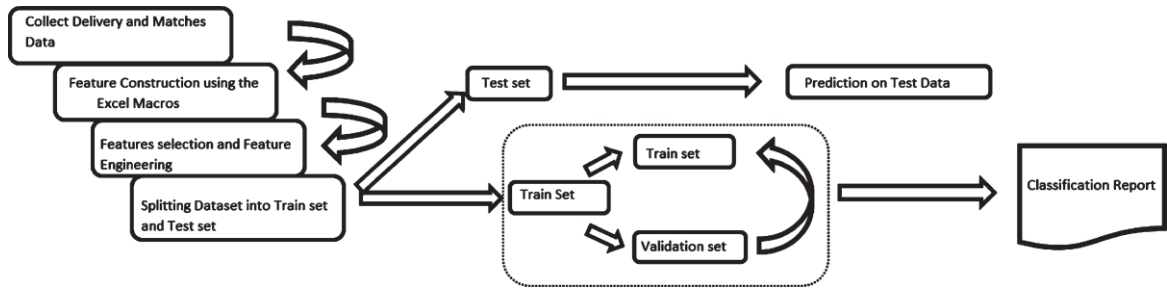


Fig. 3. Architecture of the Model (Flow Diagram).

Table 1

Table showing best fit features fed to the model

Attributes	Description
Match id	The match Id for each match in the IPL
Ball	The current order of the ball bowled
Over	The current over of the match
Balls	Played Balls left to be bowled
Bman1 – Bman11	Runs scored by the batsman on each ball
Bowl1 - Bowl10	Runs scored on every ball bowled by bowler
Extra runs	Extras given by bowlers calculated for each ball
Batsman runs	Runs scored by batsman on each ball
Remaining Balls	The total no of balls left
Wickets Left	The total no of wickets left
Present Score	Current score
Target	Total runs scored by the team batting first

Table 2

Model Nomenclature

Abbreviation	Description
o2o	One to one sequence
m2o	Many to one sequence
o2m	One to Many sequence
o2o	Many to Many sequence
MF	Multivariate Feature
bidirec	Bidirectional
LSTM	Long Short Term Memory
GRU	Gated Recurrent Unit

4.2. MODEL 2 (1–5 over)

Here they are taking train model from 1 to 5 over of all seasons upto 2008-2018 and they are taking test model from 1 to 5 season of 2019. Train and test size is 21,662 and 1,820. Here SVC and one to one sequence vanilla LSTM deep learning models and get good results. Below result in Table 4 taking input as first 5 overs of second inning. Train accuracy get 68.71% and 68.89%, confidence interval get 1.90% and 2.20% and test accuracy 68.68% and 68.65%. By Logistic Regression we get 79.27% high precision. From below (Fig. 5) we can easily visualize bar graph.

4.3. MODEL 3 (6–10 over)

Here they are taking train model from 1 to 10 over of all seasons upto 2008-2018 and they are taking test model from 6 to 9 over of season 2019. Train and test size is 42,793 and 1,820. Here SVC and one to one sequence vanilla LSTM deep learning models and get good results of 5 over of second inning. Below result in Table 5 taking input as next 6 to 10 overs of second inning. Train accuracy get 71.87% and 72.06%, confidence interval get 2.07% and 2.73% and test accuracy 72.36% and 72.09%. By GRU we get 82.75% high recall. From below (Fig. 6) we can easily visualize bar graph of between 6-10 overs.

Table 3
Classification report Results (Over 1–20)

Name	Mean_Train_score	Test_score	Precision_score	Recall_score	F1_score
SVC	76.47% (+/- 3.77%)	72.74%	83.25%	69.85%	75.97%
RandomForestClassifier	75.28% (+/- 3.09%)	72.27%	81.90%	70.64%	75.86%
XGBClassifier	76.34% (+/- 3.93%)	71.67%	83.92%	66.87%	74.43%
GradientBoostingClassifier	76.23% (+/- 3.96%)	71.61%	84.26%	66.36%	74.24%
AdaBoostClassifier	76.40% (+/- 3.87%)	71.49%	83.88%	66.57%	74.23%
Logistic Regression	76.72% (+/- 4.13%)	71.22%	82.80%	67.31%	74.26%
KNeighborsClassifier	70.92% (+/- 2.11%)	70.22%	79.67%	69.41%	74.19%
Bagging Classifier	72.06% (+/- 2.88%)	69.20%	82.21%	63.86%	71.89%
DecisionTreeClassifier	70.41% (+/- 2.14%)	68.71%	80.55%	64.93%	71.90%
GaussianNB	58.15% (+/- 2.63%)	62.66%	62.42%	99.11%	76.60%
BernoulliNB	58.06% (+/- 0.60%)	57.61%	69.10%	56.57%	62.21%
o2o_MF_bidrec_GRU	76.13% (+/- 2.59%)	72.30%	77.66%	77.33%	77.49%
o2o_MF_Stacked_GRU	75.76% (+/- 3.01%)	72.23%	77.22%	77.96%	77.59%
o2o_MF_vanilla_GRU	75.03% (+/- 3.01%)	72.11%	77.05%	78.01%	77.53%
o2o_MF_bidrec_lstm	75.69% (+/- 3.02%)	72.28%	77.55%	77.49%	77.52%
o2o_MF_Stacked_lstm	76.06% (+/- 3.04%)	72.20%	77.11%	78.10%	77.60%
o2o_MF_vanilla_lstm	75.05% (+/- 3.12%)	72.37%	77.50%	77.77%	77.64%
m2o_MF_Stacked_GRU	75.14% (+/- 3.63%)	72.25%	77.51%	77.46%	77.49%
m2o_MF_vanilla_GRU	75.85% (+/- 3.16%)	72.28%	77.05%	78.00%	77.62%
m2o_MF_Stacked_LSTM	75.09% (+/- 2.90%)	71.50%	77.76%	75.31%	76.52%
m2o_MF_vanilla_LSTM	75.94% (+/- 3.84%)	71.80%	76.08%	79.16%	77.59%

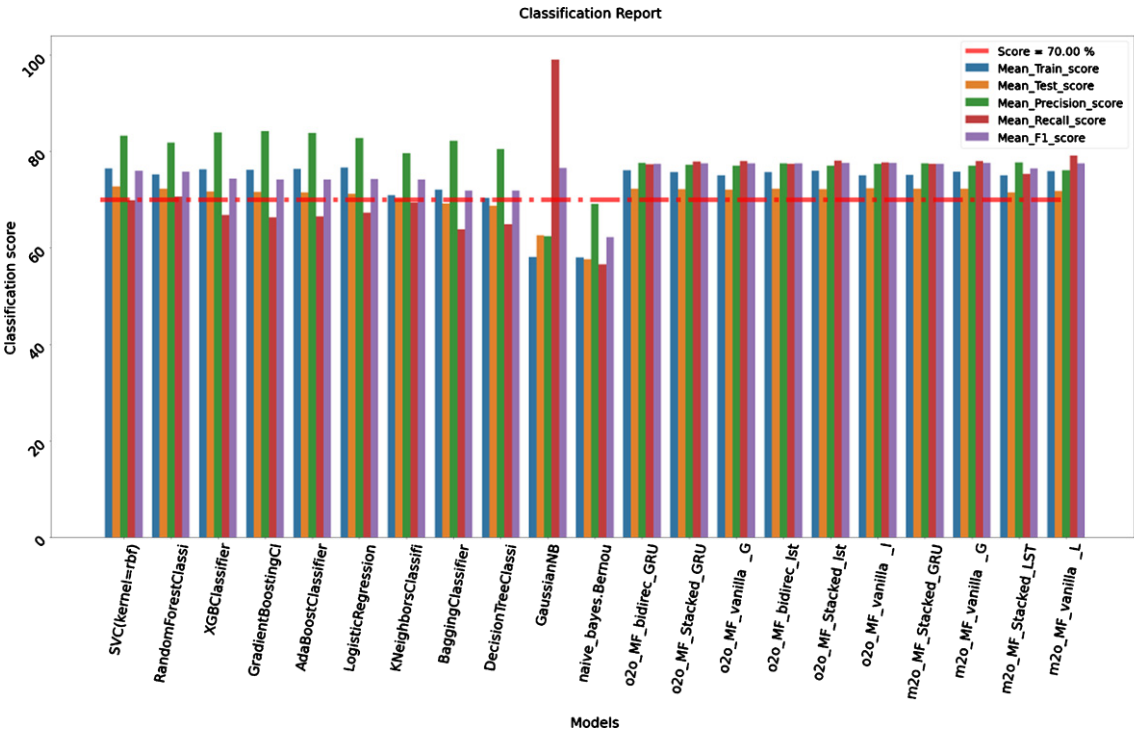


Fig. 4. Graph Classification Report Results (Over 1–20).

4.4. MODEL 4 (11–15 over)

Here they are taking train model from 1 to 15 over of all seasons upto 2008–2018 and they are taking test model from 11 to 15 over of season 2019. Train and test size is 63,007 and 1,829. Here deep learning models are getting best scores. Below result in Table 6 taking input as next 11 to 16 overs of second inning. GRU and Bidirectional LSTM Train accuracy get 76.02% and 75.58%, confidence interval get 3.50% and 3.41%. They getting good precision of recall and F1 scores gets 82.00%. Random Forest Classifier gives good precision score 90.26% on test data. From below (Fig. 7) we can easily visualize bar graph of between 11–15 overs.

Table 4
Classification Report Results (Over 1–5)

Name	Mean_Train_score	Test_score	Precision_score	Recall_score	F1_score
Logistic Regression	67.81% (+/- 2.90%)	66.70%	79.27%	63.24%	70.35%
RandomForestClassifier	68.17% (+/- 2.15%)	66.54%	75.73%	68.34%	71.84%
SVC	68.71% (+/- 1.90%)	68.68%	78.04%	69.39%	73.46%
XGBClassifier	65.14% (+/- 1.92%)	67.53%	77.35%	67.90%	72.32%
m2o_MF_Stacked_GRU	68.85% (+/- 2.00%)	68.59%	78.20%	68.90%	73.26%
m2o_MF_Stacked_LSTM	68.49% (+/- 1.81%)	68.65%	77.61%	69.96%	73.59%
m2o_MF_vanilla_GRU	67.94% (+/- 2.41%)	69.25%	76.67%	72.95%	74.76%
m2o_MF_vanilla_LSTM	67.39% (+/- 3.67%)	67.66%	77.60%	67.75%	72.34%
o2o_MF_Stacked_GRU	68.24% (+/- 2.40%)	68.52%	78.78%	67.90%	72.93%
o2o_MF_Stacked_lstm	68.49% (+/- 2.37%)	68.57%	79.21%	67.37%	72.81%
o2o_MF_bidirec_GRU	68.86% (+/- 2.31%)	68.19%	78.53%	67.55%	72.62%
o2o_MF_bidirec_lstm	68.83% (+/- 2.15%)	68.13%	78.86%	66.93%	72.41%
o2o_MF_vanilla_GRU	68.46% (+/- 2.18%)	68.68%	78.72%	68.34%	73.16%
o2o_MF_vanilla_lstm	68.89% (+/- 2.20%)	68.57%	78.68%	68.16%	73.04%

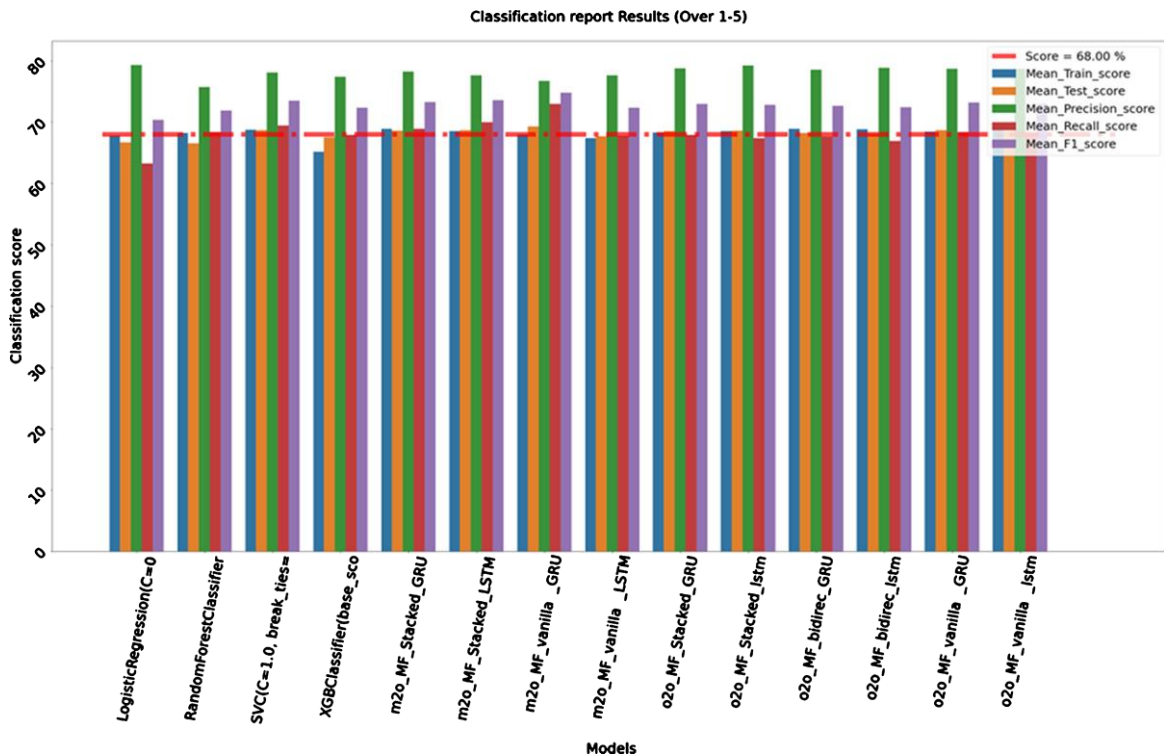


Fig. 5. Graph Classification Report Results (Over 1–5).

Table 5
Classification Report Results (Over 6–10)

Name	Mean_Train_score	Test_score	Precision_score	Recall_score	F1_score
Logistic Regression	71.44% (+/- 3.15%)	69.84%	75.58%	76.71%	76.14%
RandomForestClassifier	71.69% (+/- 3.16%)	70.49%	75.79%	77.85%	76.80%
SVC	71.87% (+/- 2.07%)	72.36%	75.95%	81.87%	78.80%
XGBClassifier	69.17% (+/- 3.65%)	70.49%	74.03%	81.61%	77.63%
m2o_MF_Stacked_GRU	72.28% (+/- 2.76%)	71.51%	75.70%	80.35%	77.96%
m2o_MF_Stacked_LSTM	71.87% (+/- 3.45%)	70.96%	75.80%	78.86%	77.30%
m2o_MF_vanilla_GRU	72.24% (+/- 2.48%)	71.34%	75.94%	79.47%	77.67%
m2o_MF_vanilla_LSTM	71.86% (+/- 2.69%)	71.45%	75.77%	80.09%	77.87%
o2o_MF_Stacked_GRU	71.84% (+/- 2.82%)	72.03%	75.88%	81.26%	78.48%
o2o_MF_Stacked_lstm	72.06% (+/- 2.73%)	72.09%	75.94%	81.26%	78.51%
o2o_MF_bidirec_GRU	71.82% (+/- 2.04%)	71.92%	75.71%	81.35%	78.43%
o2o_MF_bidirec_lstm	71.98% (+/- 2.66%)	71.98%	75.69%	81.52%	78.50%
o2o_MF_vanilla_GRU	72.09% (+/- 2.38%)	71.87%	75.00%	82.75%	78.68%
o2o_MF_vanilla_lstm	71.88% (+/- 2.46%)	71.65%	75.20%	81.79%	78.36%

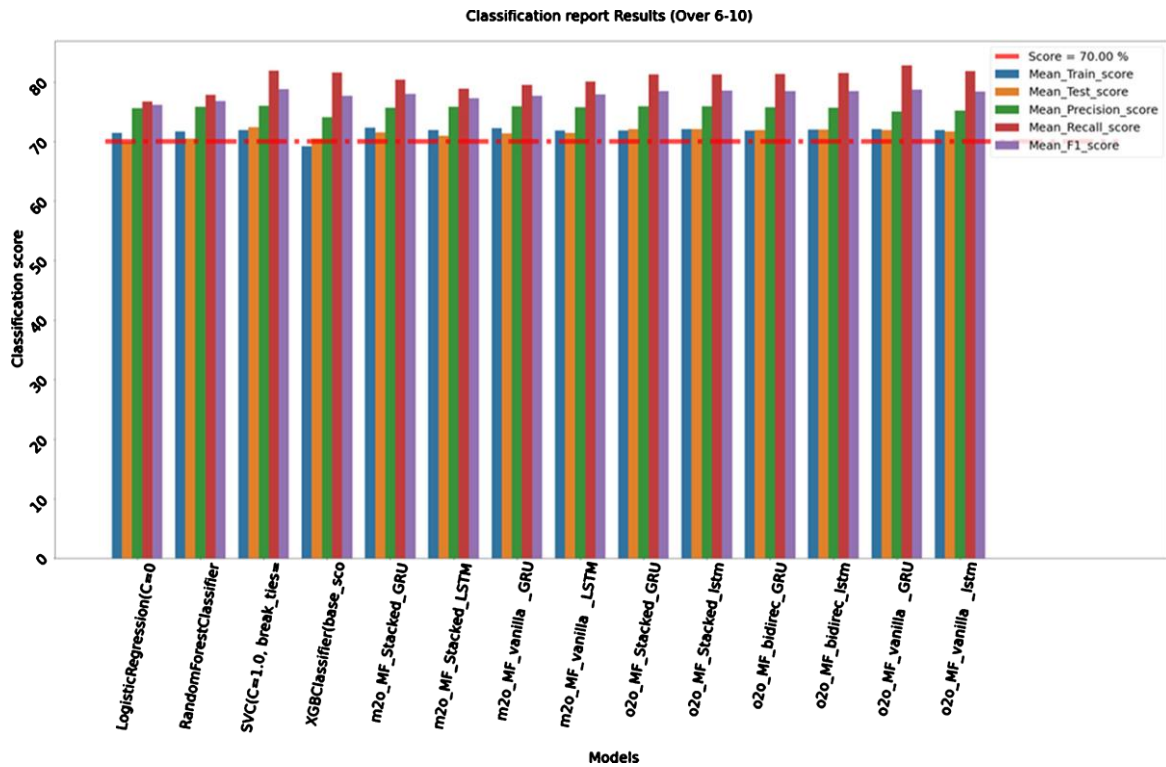


Fig. 6. Graph Classification Report Results (Over 6–10).

4.5. MODEL 5 (16–20 over)

Here they are taking train model from 1 to 20 over of all seasons upto 2008–2018 and they are taking test model from 16 to 20 over of season 2019. Train and test size is 79,129 and 1,491. Logistic regression accuracy gives 76.48% (+/-2.94). SVC and Bidirectional GRU gives accuracy 76.00% and test scores show 77.26% and 77.13% of second inning and we get good Recall score nearly 100.00%. Recall expresses can find relevant instances in a dataset, precision expresses proportion of the data points in

model with relevant. From below (Fig. 8) we can easily visualize bar graph of between 16–20 overs. Below (Fig. 9) shows the graph of train and test score of match and also get various intervals in overs. Blue line show train score in graph and Green line show test score in graph. From the 15th over model gets good prediction. Test scores are higher than train scores. SVC, XGBClassifier and GRU gives good test scores as compared to their train score from the second inning.

Table 6
Classification Report Results (Over 11–15)

Name	Mean_Train_score	Test_score	Precision_score	Recall_score	F1_score
Logistic Regression	75.15% (+/- 3.94%)	78.79%	83.27%	82.69%	82.98%
RandomForestClassifier	74.70% (+/- 4.02%)	78.57%	90.26%	73.69%	81.14%
SVC	75.39% (+/- 3.43%)	79.22%	83.33%	83.48%	83.41%
XGBClassifier	72.35% (+/- 4.79%)	76.49%	82.40%	79.37%	80.85%
m2o_MF_Stacked_GRU	76.10% (+/- 3.74%)	78.76%	82.95%	83.10%	83.03%
m2o_MF_Stacked_LSTM	75.78% (+/- 3.55%)	78.65%	82.25%	83.98%	83.10%
m2o_MF_vanilla_GRU	75.48% (+/- 3.53%)	78.82%	82.63%	83.71%	83.17%
m2o_MF_vanilla_LSTM	75.25% (+/- 3.92%)	79.04%	83.97%	82.14%	83.05%
o2o_MF_Stacked_GRU	76.00% (+/- 3.45%)	78.90%	82.56%	84.00%	83.28%
o2o_MF_Stacked_lstm	75.94% (+/- 3.49%)	78.90%	82.96%	83.39%	83.17%
o2o_MF_bidirec_GRU	75.90% (+/- 3.42%)	79.22%	82.87%	84.18%	83.52%
o2o_MF_bidirec_lstm	76.02% (+/- 3.50%)	79.28%	83.17%	83.83%	83.50%
o2o_MF_vanilla_GRU	75.58% (+/- 3.41%)	79.28%	82.50%	84.88%	83.67%
o2o_MF_vanilla_lstm	75.57% (+/- 3.51%)	78.73%	82.80%	83.30%	83.05%

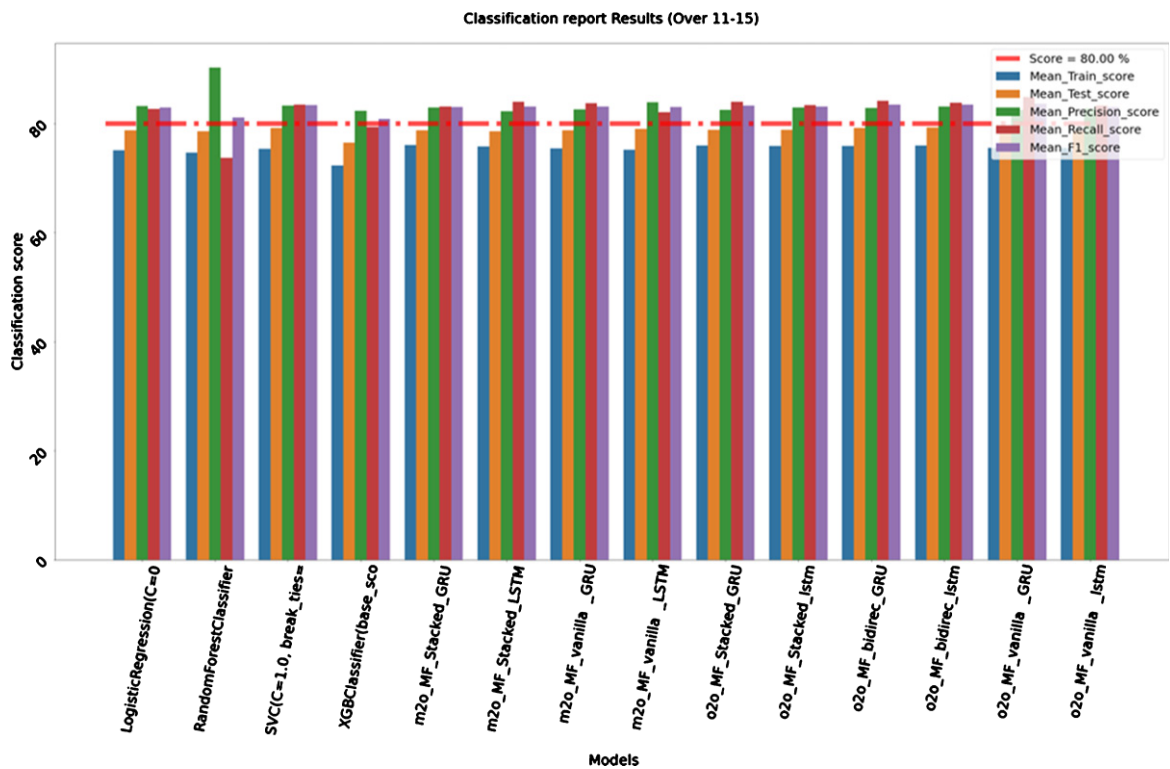


Fig. 7. Graph Classification Report Results (Over 11–15).

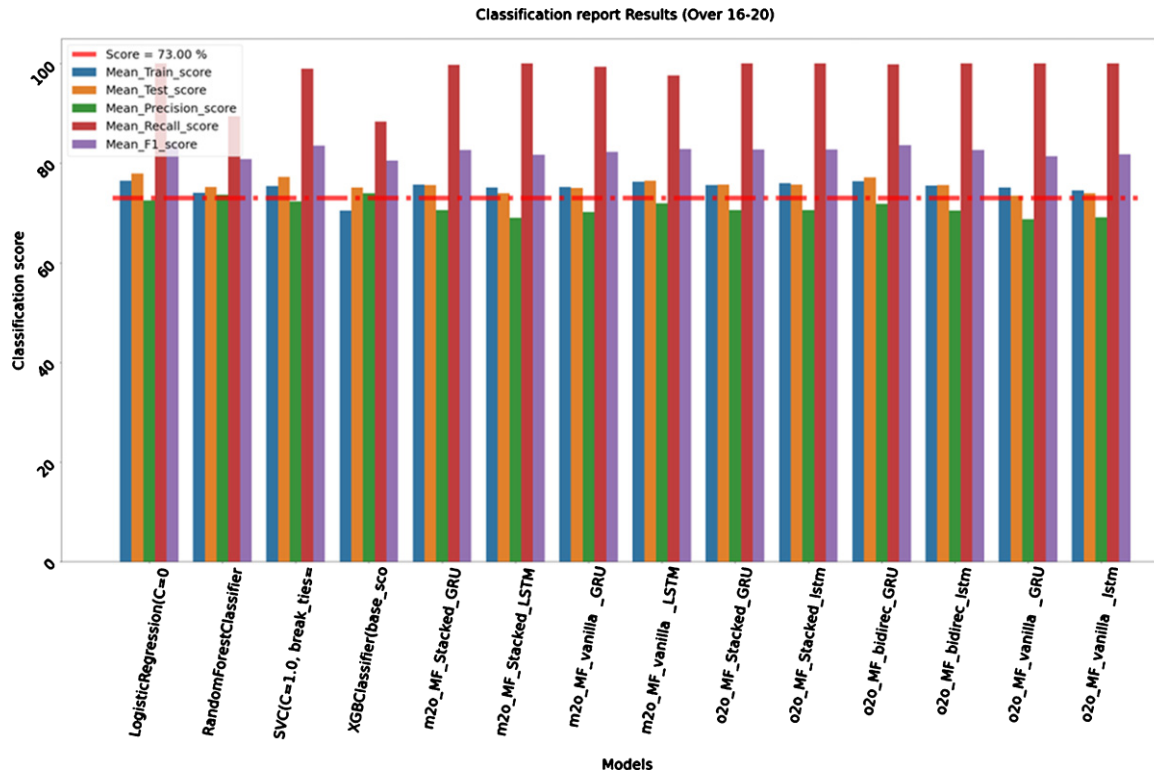


Fig. 8. Graph Classification Report Results (Over 16–20).

Table 7
Classification Report Results (Over 16–20)

Name	Mean-Train-score	Test-score	Precision-score	Recall-score	F1-score
Logistic Regression	76.48% (+/- 2.94%)	77.93%	72.54%	100.00%	84.08%
RandomForestClassifier	74.01% (+/- 3.59%)	75.18%	73.65%	89.41%	80.77%
SVC	75.38% (+/- 2.85%)	77.26%	72.27%	98.96%	83.54%
XGBClassifier	70.46% (+/- 4.21%)	75.12%	73.99%	88.38%	80.55%
m2o_MF_Stacked_GRU	75.65% (+/- 2.71%)	75.62%	70.55%	99.77%	82.66%
m2o_MF_Stacked_LSTM	75.10% (+/- 3.54%)	73.94%	69.08%	100.00%	81.72%
m2o_MF_vanilla_GRU	75.24% (+/- 2.82%)	75.02%	70.17%	99.31%	82.23%
m2o_MF_vanilla_LSTM	76.32% (+/- 3.19%)	76.43%	71.94%	97.58%	82.82%
o2o_MF_Stacked_GRU	75.57% (+/- 3.11%)	75.65%	70.54%	100.00%	82.72%
o2o_MF_Stacked_lstm	76.02% (+/- 2.69%)	75.72%	70.59%	100.00%	82.76%
o2o_MF_bidirec_GRU	76.33% (+/- 2.54%)	77.13%	71.85%	99.88%	83.58%
o2o_MF_bidirec_lstm	75.51% (+/- 3.32%)	75.59%	70.48%	100.00%	82.68%
o2o_MF_vanilla_GRU	75.12% (+/- 3.58%)	73.44%	68.70%	100.00%	81.44%
o2o_MF_vanilla_lstm	74.58% (+/- 3.49%)	73.98%	69.13%	100.00%	81.75%

5. Conclusion and future work

From these research paper they take 692 matches of all seasons 2008-2018 to train model take 59 matches of season 2019 for to test. In future they start beginning of the first inning. In these research they only focus on IPL match but it can apply any other sports games like football, tennis etc. They also take care and handle results of the matches which are interrupted by rainfall or other natural calamities. For to handle such type of scenarios they take help from Duckworth-Lewis method. Winning or losing is depend on the players. For to manage team we have

the information about on batsman and bowler like how many runs a batsman is likel score and how many wickets a bowler is likely take. In the cricket match it carries a lot of randomness, because every over score will change. IPL is at infantry stage it decade old league has less number of matches compared to test score. IPL is the popular leage and it has the wealthiest brand value which was \$6.7 billion in 2015,according to BCCI IPL contributed around \$160 million to the GDP of indian economy there is scale given above is scale of betting industries worldwide this study can benefit cricket club manager, sport data analyst and many more.

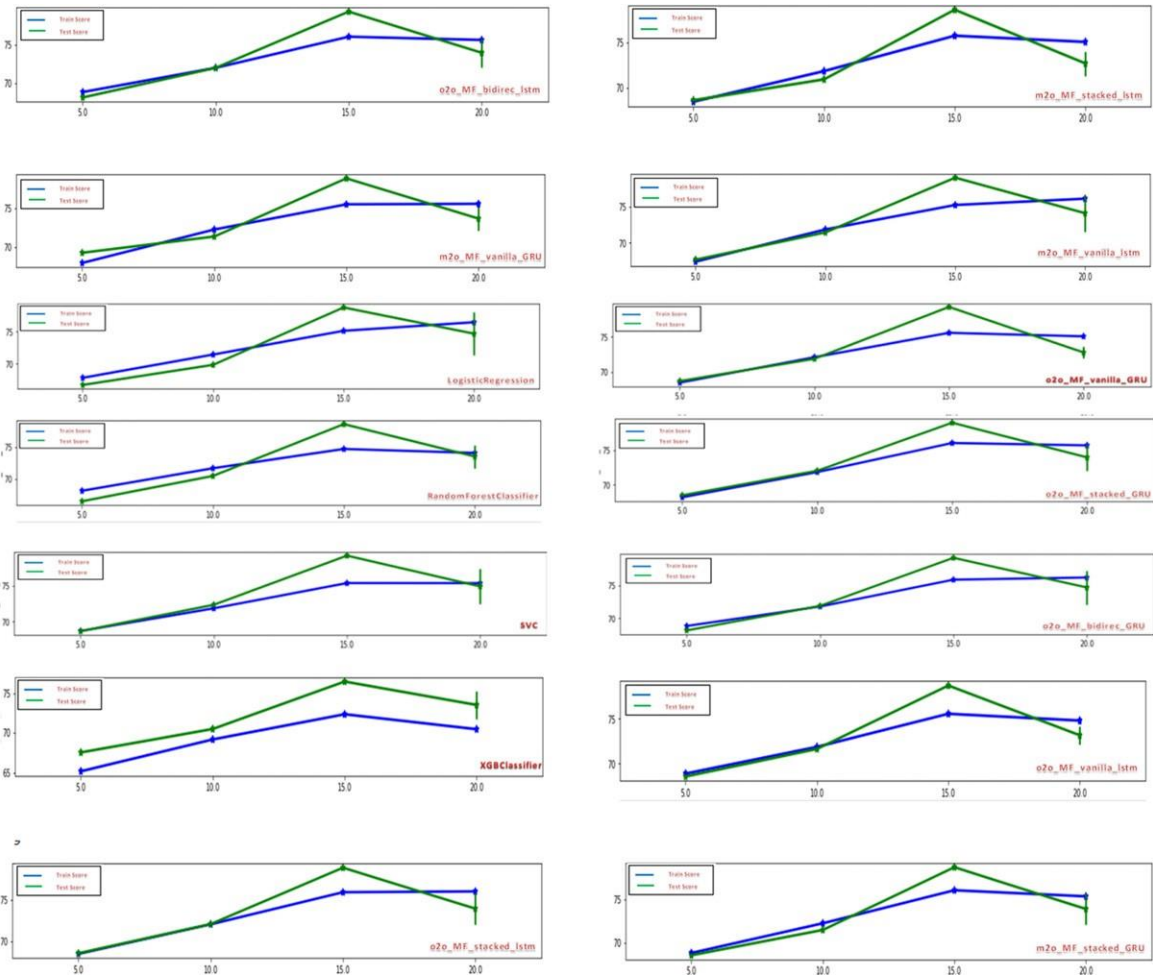


Fig. 9. Trend chart of Train and Test scores during the second inning.

References

- Bailey, M., and Clarke, S.R., 2006, Predicting the match outcome in one day international cricket matches, while the game is in progress, *J Sports Sci Med*, 5(4), 480–48.
- Basavaraju, P., and Varde, A.S., 2017, Supervised learning techniques in mobile device apps for androids, *SIGKDD Explor*, 18(2), 18–29.
- Bhattacharjee, D., and Talukdar, P., 2019, Predicting outcome of matches using pressure index: evidence from Twenty20 cricket, *Commun Stat Simul Comput*, pp. 1–13.
- Deep, C., Patvardhan, C., and Vasantha, C., 2016, Data analytics based deep mayo predictor for IPL-9, *Int J Comput Appl*, 152(6), 6–11.
- De Melo, Gerard and Varde, Apama, 2015, Scalable Learning Technologies for Big Data Mining.
- Jayanth, S.B., Anthony, A., Abhilasha, G., Shaik, N., and Srinivasa, G., 2018, A team recommendation system and outcome prediction for the game of cricket, *J Sports Anal*, 4(4), 263–227.
- Kaluarachchi, A., and Apama, S.V., 2010, CricAI: A classification based tool to predict the outcome in ODI cricket, *2010 Fifth International Conference on Information and Automation for Sustainability*, pp. 250–255.
- Kampakis, S., and Thomas, W., 2015, Using machine learning to predict the outcome of English county twenty over cricket matches, arXiv [stat.ML].
- Lamsal, R., and Choudhary, A., 2018, Predicting outcome of Indian Premier League (IPL) matches using machine learning, arXiv [stat.AP].
- Maheswari, P.U., A Novel Approach for Mining Association Rules on Sports Data using Principal.
- Manivannan, S., and Kausik, M., 2019, Convolutional neural network and feature encoding for predicting the outcome of cricket matches, in *2019 14th Conference on Industrial and Information Systems (ICIIS)*.
- Passi, K., and Pandey, N., 2018, Increased prediction accuracy in the game of cricket using machine learning, arXiv [cs.OH].
- Passi, K., and Pandey, N., 2018, Predicting players' performance in one day international cricket matches using machine learning, in *Computer Science & Information Technology*.
- Pathak, N., and Wadhwa, H., 2016, Applications of modern classification techniques to predict the outcome of ODI cricket, *Procedia Comput Sci*, 87, pp. 55–60.
- Percy, D.F., 2015, Strategy selection and outcome prediction in sport using dynamic learning for stochastic processes, *J Oper Res Soc*, 66(11), 1840–1849.
- Sankaranarayanan, V.V., Sattar, J., and Lakshmanan, L.V.S., 2014, Auto-play: A data mining approach to ODI cricket simulation and prediction, in *Proceedings of the 2014 SIAM International Conference on Data Mining, Philadelphia, PA: Society for Industrial and Applied Mathematics*, pp. 1064–1072.
- Schoeman, J.H., Matthee, M.C., and Van der Merwe, P., 2006, The viability of business data mining in the sports environment: cricket match analysis as application, *S.A. j. res. sport phys. educ. recreat*, 28(1).
- Singh S., and Kaur, P., 2017, IPL visualization and prediction using HBase, *Procedia Comput Sci*, 122, 910–915.
- Singh, T., Singla, V., and Bhatia, P., 2015, Score and winning prediction in cricket through data mining, in *2015 International Conference on Soft Computing Techniques and Implementations (ICSCTI)*.
- Tandon, N., Varde, A.S., and de Melo, G., 2018, Commonsense knowledge in machine intelligence, *SIGMOD Rec*, 46(4), 49–52.
- Ul Mustafa, R., Nawaz, M.S., Ullah Lali, M.I., Zia, T., and Mehmood, W., 2017, Predicting the cricket match outcome using crowd opinions on social networks: A comparative study of machine learning methods, *Malays J Comput Sci*, 30(1), 63–76.
- Van den Berg, L., Coetzee, B., Blignaut, S., and Mearns, M., 2018, The competitive intelligence process in sport: data collection properties of high-level cricket coaches, *Int J Perform Anal Sport*, 18(1), 32–54.
- 2019, Forecasting the outcome of the next ODI cricket matches to be played, *Regular Issue*, 8(4), 10269–10273.