IEEE BOMBAY SECTION

9th International Conference for Convergence in Technology (9th I2CT 2024)

Prediction of Shot Selection of a Batsman in the death over of a T20 Cricket Match using Machine Learning

Paper ID: 1298

Deepak Kumar Sahoo, Deheem u Deyar

Amrita Vishwa Vidyapeetham

IEEE

Table of Contents

INTRODUCTION

- ✓ Decision making in sports has been pivotal to the development of the sport, especially seen in cricket
- ✓ Setting up field is important to the context of decision making in Cricket.
- ✓ To set up optimal fielding positions, and studying the pattern of shot selection by the batsman is important to achieve it.
- ✓ T20 Cricket is the shortest version of the game, and it’s easier to analyze the game in this format and the final over is when this decision making is quite prevalent.
- ✓ To do this, we used many machine learning algorithms, and ensemble methods

PROBLEM STATEMENT

Death over of the chasing innings has many uncertainty pertaining towards winning probability of a cricket team, to reduce the uncertainty, prediction shot selection of the batsman is needed such that effective decision making takes place to assist the fielding team to set up optimal fielding positions. To predict shot selection of a batsman, machine learning approach is used.

MOTIVATION

To implement a model which predicts shot selection of the batsman in the death over of a T20 match to enhance the decision making of the fielding team and reduce the uncertainty in the game.

LITERATURE SURVEY

S.NO	Author(s)	Full title of the paper with year	Inference based on the methodology
1	Jadhav, Ranjana, Bhargav Pawar, Nishant Bhat, Shyam Kawale, and Abhijit Gawai	Predicting Optimal Cricket Team using Data Analysis. 2021	Used Data Analysis
2	Tharoor, Vishnusai Viswajith, and N. M. Dhanya	Performance of Indian Cricket Team in Test Cricket: A comprehensive Data Science analysis, 2022	Used Apriori Algorithm and Hypothesis Testing
3	Lekamge, E. L., K. R. Wickramasinghe, S. E. Gamage, T. M. K. L. Thennakoon, Prasanna S. Haddela, and Sandamini Senaratne	CricSquad: A System to Recommend Ideal Players to a Particular Match and Predict the Outcome of the Match. 2023	Used Association rules

S.NO	Author(s)	Full title of the paper with year	Inference based on the methodology
4	[16] Das, Dhruba, Hemanta Saikia, Dibyojyoti Bhattacharjee, and Bhaskar Kushvaha	On estimating shot selection by a batsman in Twenty20 cricket: A probabilistic approach. 2021	Used Probabilistic Approach
5	Tharoor, Vishnusai Viswajith, and N. M. Dhanya	You cannot do that Ben Stokes: Dynamically predicting shot type in cricket using a personalized deep neural network	Used LSTM method by calculating aggressiveness of the shots played by the batsman

ABOUT THE DATASET

TABLE I. DESCRIPTION OF THE DATASET AND FREQUENCY DISTRIBUTION

Attributes	*Description*	*Frequency Distribution*
Total_Runs_on_delivery	Total runs conceded off the ball	0(681), 1(721), 2(252), 3(12), 4(256), 5(2),6(197),7(1)
Innings_balls	Ball number of the innings in numerical form	115 (451), 116 (413), 117(369),118 (338), 119(299), 120(252)
Current_score	Current score of the innings after the ball is played	Current_score <= 150 (790) & Current_score > 150 (1332)
Target	Runs scored by the team which batted in the first innings	Target <= 150 (299) & Target > 150 (1823)
RNPB	Runs required per ball	RNPB < 6 (1398) & RNPB > 6(724)
Previ_ball_type	Ball type on the previous delivery	Full (1171), Good (649), Short (302)
Previ_shot_type	Shot type on the previous delivery	Drive (1216), Loft (702), Pull (204)
Ball_Type	Ball Type bowled by the bowler	Full (1171), Good (649), Short (302)
Shot_Type	Shot selection made by the batsman	Drive (1216), Loft (702), Pull (204)

CLASS BALANCING

- The inferences made from the distribution of these two features are:

  - A batsman is most favored to drive the full ball and pull the short ball.
  - The chances of driving or making a pull shot on a good ball are almost equally likely.
- There are 1171 instances of full ball being bowled, 649 instances of good ball and 302 instances of short balls. Against which, 1215 Drive shots, 703 Lofted shots, and 204 Pull shots were played.

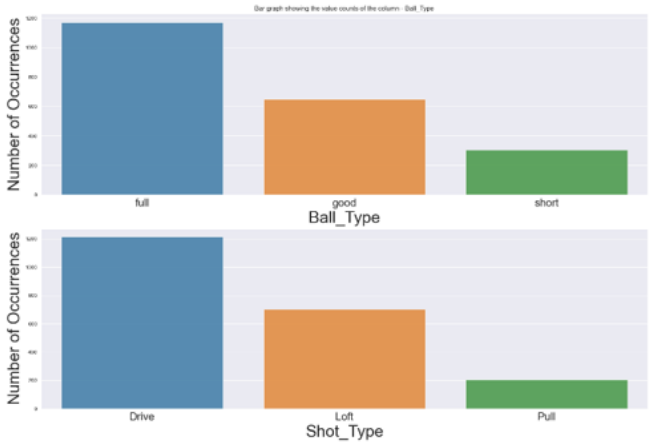
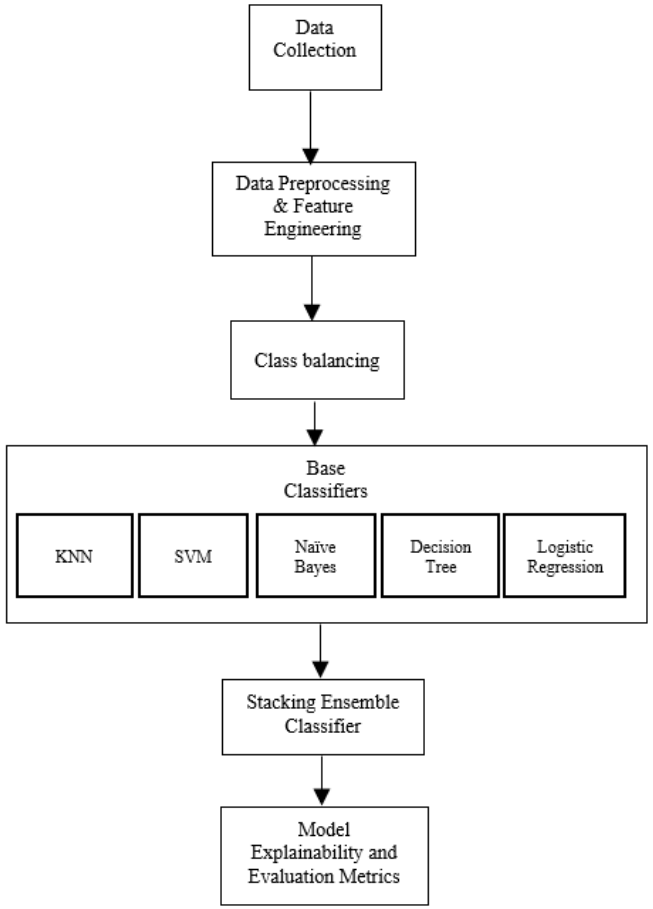


Fig 5. Distribution of Shot Type and Ball Type

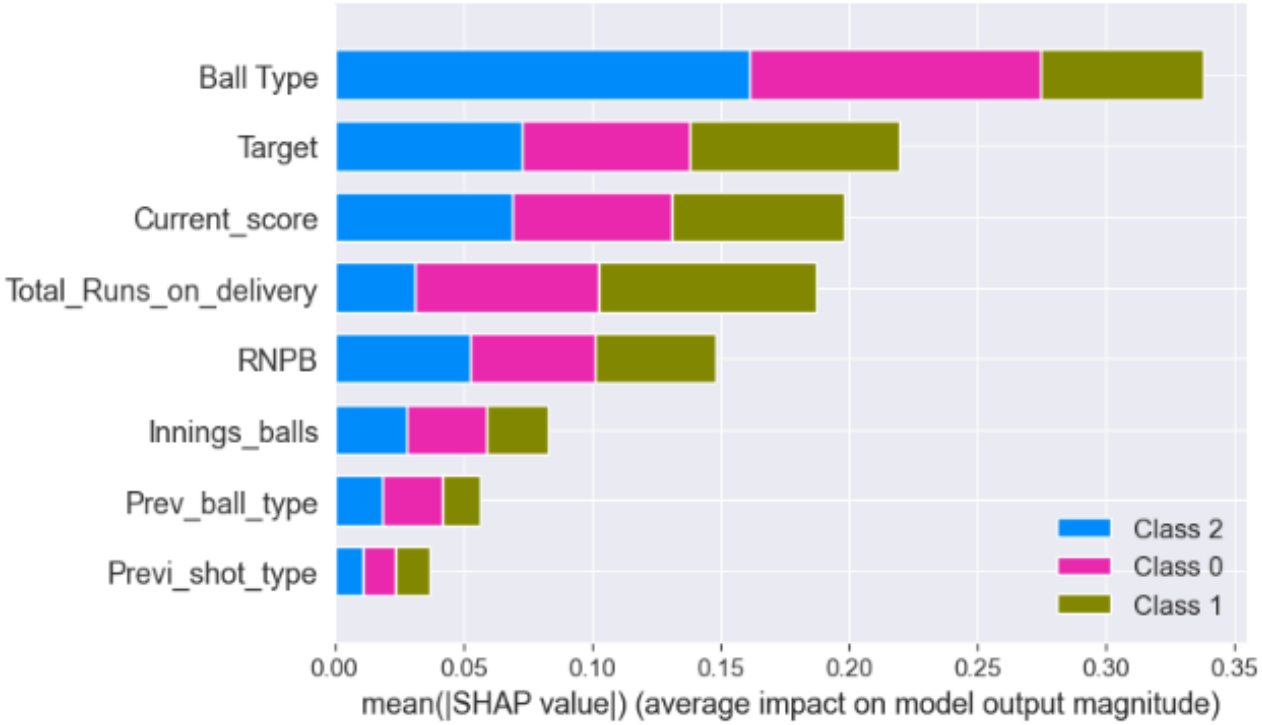
METHODOLOGY



RESULTS AND ANALYSIS

TABLE II. COMPARISON OF PERFORMANCE MEASURES OF THE APPLIED CLASSIFIERS

Classifiers	Performance Metrics				
	Accuracy	Precision	Recall	F1 Score	Specificity
KNN	0.639	0.621	0.639	0.617	0.588
SVM	0.664	0.663	0.664	0.630	0.626
Naive Bayes	0.649	0.640	0.649	0.637	0.821
Decision Tree	0.589	0.594	0.589	0.590	0.529
Logistic Regression	0.647	0.635	0.647	0.619	0.593
Random Forest	0.660	0.650	0.660	0.647	0.607
XGBoost	0.637	0.631	0.637	0.631	0.500
SMOTE + KNN	0.708	0.704	0.708	0.704	0.704
SMOTE + SVM	0.689	0.686	0.689	0.683	0.686
SMOTE + Naive Bayes	0.695	0.701	0.695	0.693	0.807
SMOTE + Decision Tree	0.719	0.718	0.719	0.716	0.718
SMOTE + Logistic Regression	0.653	0.638	0.653	0.624	0.604
SMOTE + Random Forest	0.767	0.764	0.767	0.763	0.764
SMOTE + XGBoost	0.777	0.771	0.777	0.775	0.776
SMOTE + Voting Classifier with RF, KNN,and XGBoost	0.791	0.787	0.790	0.786	0.787
SMOTE + Stacking Classifier with RF, KNN,and XGBoost + SVM (meta learner)	0.796	0.795	0.796	0.795	0.762
SMOTE + Stacking Classifier with RF, KNN,and XGBoost + LR (meta learner)	0.793	0.791	0.793	0.791	0.792



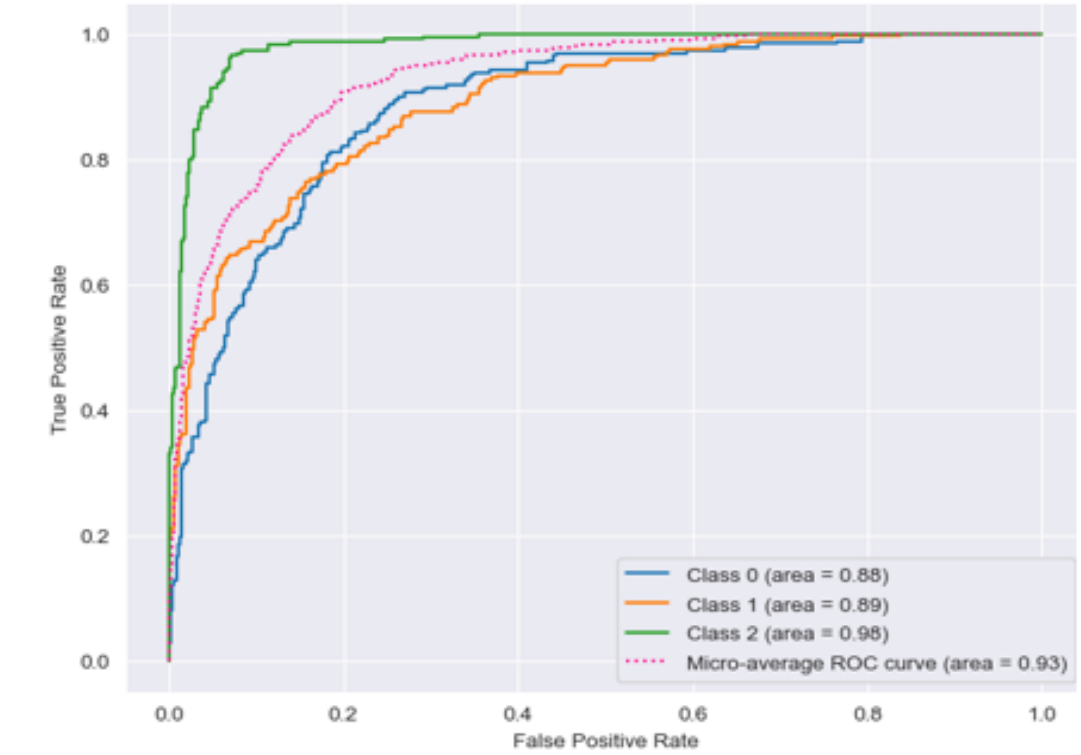


Fig 7. ROC Curve of Stacking Classifier to predict shot selection of the batsman.

CONCLUSION AND FUTURE SCOPE

- Stacking Classifier performed the best among all the classifier model.
- Use of commentary was effective
- Further study can extend to a whole match and different formats of the game
- Opponent team details can be included.
- NLP based architecture could be used to extract important features from commentary

REFERENCES

1. Jadhav, Ranjana, Bhargav Pawar, Nishant Bhat, Shyam Kawale, and Abhijit Gawai. "Predicting Optimal Cricket Team using Data Analysis." In 2021 International Conference on Emerging Smart Computing and Informatics (ESCI), pp. 269-274. IEEE, 2021.
2. Tharoor, Vishnusai Viswajith, and N. M. Dhanya. "Performance of Indian Cricket Team in Test Cricket: A comprehensive Data Science analysis." In 2022 International Conference on Electronic Systems and Intelligent Computing (ICESIC), pp. 128-133. IEEE, 2022
3. Das, Dhruba, Hemanta Saikia, Dibyojyoti Bhattacharjee, and Bhaskar Kushvaha. "On estimating shot selection by a batsman in Twenty20 cricket: A probabilistic approach." Communications in Statistics: Case Studies, Data Analysis and Applications 8, no. 2 (2022): 354-367.
4. 4. Gürpınar-Morgan, Will, Daniel Dinsdale, Joe Gallagher, Aditya Cherukumudi, and Patrick Lucey. "You cannot do that Ben Stokes: Dynamically predicting shot type in cricket using a personalized deep neural network." arXiv preprint arXiv:2102.01952 (2021).

THANK YOU