Rebuttal Table

|  |  |
| --- | --- |
| 1. The Study fails to address how the findings relate/differ to previous research in this area | We compared our model with the papers given in the reference:   1. Sunil Bhutada, Subhani Shaik, Aishwarya Laxmi Nethi, Pabba Adeeshwar, Sairam Parshi. (2020). IPL Match Prediction using Machine Learning. International Journal of Advanced Science and Technology, 29(05), 3438 - 3448. Retrieved from <http://sersc.org/journals/index.php/IJAST/article/view/12036>; 2. The paper lacks information about data preprocessing and feature creation 3. A significant amount of it is filled with redundant graphs, the paper is not reproduceable and doesn’t give any information how such high accuracy was achieved. 4. Predictive Analysis of IPL Match Winner using Machine Learning Techniques. Ch Sai Abhishek, Ketaki V Patil, P Yuktha, Meghana KS, MV Sudhamani. International Journal of Innovative Technology and Exploring Engineering (IJITEE), Volume-9 Issue-2S, December 2019. <http://www.ijitee.org/wp-content/uploads/papers/v9i2S/B10431292S19.pdf> 5. Same as the previous one, the paper doesn’t give enough information as to how such high accuracy was achieved which doesn’t seem feasible with such small data set. 6. Sarvani Anandarao, B. Manvitha Bramarambika, K.Lakshmi Prahla, Sk. Kushbu Kalam. (2020). ANALYZING AND ESTIMATING THE IPL WINNER USING MACHINE LEARNING. International Journal of Advanced Science and Technology, 29(04), 1940 - 1946. Retrieved from <http://sersc.org/journals/index.php/IJAST/article/view/7921> 7. The accuracy is hugely inconsistent among the different models 8. The model doesn’t give enough information as to how such high accuracy was achieved 9. Naive Bayes approach to predict the winner of an ODI cricket game. I. Wickramasinghe. Journal of Sports Analytics 6 (2020) 75–84DOI 10.3233/JSA-200436 10. This is a good paper but comparing a paper made on ODI cricket with one made on IPL doesn’t seem right because the amount of data available for ODI is towering. 11. There are tons of features which can be created for ODI but not for IPL because the players keep changing to often in IPL. 12. The players play very unpredictably as the stakes are quite different. |
| 2. The dataset can be well understood if there is a visualization of the features distribution in the dataset. | We thought adding figures would not be that useful as we tried a lot of different methods to make the model robust so to not make the paper lengthy we decided not to add graphs. |
| 3. The abbreviations of KTK and GL are not given on their first occurrence (note: this helps if the reader is not familiar with IPL) | The abbreviations are given in the specified places |
| 4. The citation for Sklearn is missing despite using their work. | The citation is now provided at the reference section numbered 20 |
| 5. The ambiguity % is missing in the results despite mentioning that it is one of the parameter on which the evaluation for the model was made | Ambiguity was already sent in the tables file sent along with the paper, now those results are also added in the paper in the results section. |
| 6. The study does not show what measures were taken to make sure model is not underfit as it was mentioned that the dataset was small. | We started with creating some basic fundamental features like strength of the team based on the players’ different performance metrics i.e. number of sixes, number of fours type of features. After this we used these features and created some second level features out of this and then use these features to create difference features by subtracting one from the other. We exhausted all the different features that can be made using the available data and then applied various models. We trained the models with different hyper parameters using grid search and chose the ones which gave the highest and robust results and least amount of ambiguity. This was added in the conclusion section as the whole paper explains this very process. |
| 7. The Study does not provide details on why tree-based classifiers performed better on the data | It is not always possible to decide what classifier to use just by seeing the data so we took some inspiration from the research papers we read and found that decision trees work best on cricket-based data. We trained our model on both geometric classifiers and tree-based classifiers and found the tree-based classifiers outperforming the geometric ones. Moreover, you can see that our data is not continuous, some features have continuous data and some don’t so it makes more sense to use tree-based classifier as tree-based classifier are known to work better on data that is not continuous. Tree-based classifier also work better on small dataset |
| 8. The Study needs more experiments on models to prove its claim that tree-based classifiers perform better on IPL data than geometric based classifiers. | We have done experiments on more geometric based classifier but seeing the trend of tree-based classifiers working better we presented the results of mostly tree-based ones. |
| 9. The study does not mention how the model selection was made? Was it random? Or was there any decision process to not include models like SVM? | It was not random, we tried different models and after tuning the hyperparameters using grid search we found out that tree-based work better. We have added SVM results in the results now. |
| 10. The Study does not provide the % of train - test split used for the models on the data. | We used stratified k-fold with n\_splits = 10 and n\_repeats = 2. It is mentioned in the paper, we used stratified k-fold instead of a single split to make the model more robust. |
| 11. The results are mentioned under section “3. Dataset Split” of the manuscript. There should be a separate section for results. | It has been fixed. |
| 12. The Study does not provide enough information on how to reproduce this experiment. | We have provided all the information needed to recreate this model or use it for more advanced research |
| 13(Reviewer’s 1’s comments): Many such work already reported; what novelty this work provides should be well highlighted in abstract (novelty) in introduction (to justify the study) under discussion (how your work is better) and also state in conclusion | Our model is unique because what we have seen is that most of the models built by different people have absurd accuracies, they are reporting accuracies of 80%, 90% which doesn’t seem possible for this dataset and also for predicting cricket based tournament unless the dataset has been compromised. Our model doesn’t have such sky rocketing accuracy but it gives a balanced accuracy around all the models which shows how robust the model is, we tuned the hyperparameters of all the models for tons of different combinations and chose the best ones. |