Rebuttal Table

|  |  |
| --- | --- |
| 1. The Study fails to address how the findings relate/differ to previous research in this area | We have researched and read almost every paper in the domain of IPL Result prediction. However, we tried to replicate the result and failed to do so. Please see the below comments on every paper:  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>;   The paper is well formatted, and the visualization is excellent.  However, there were some flaws in the paper due to which we  haven’t compared the results:   1. The paper lacks information about data preprocessing and feature creation. 2. The author states that they have used Euler Formula for creating the strength, but they haven’t mentioned which factors and components they have considered in their formula. 3. Though the accuracy of the paper is around 58% using Naïve Bayes, which seems legit, the paper is impossible to reproduce. Some bascic information like train-test split, features created and hyperparameters used are missing. 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. Though the authors have mentioned that they have used feature elimination methods on the basis of feature importance to select the features for training the model, there is nowhere mentioned on which features the feature elimination is applied. It is impossible to replicate the model without knowing the features used. 6. The authors have not mentioned K-Fold cross validation for evaluating the model. So, we believe that the model is evaluated on a single split. Hence, there can be a possibility that such a high accuracy is coming due to a single Random State (The issue in our paper). 7. The author states that the accuracy with RandomForest is 89%. However, there is no mentioning of the train-test split ratio, what is the training accuracy(if we are assuming the accuracy given is test accuracy). 8. Moreover, achieving such a high accuracy in sports predictions seems a bit unreliable. 9. 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> 10. Through the images it is possible to know the features used. However, there are a lot of problems that are evident. 11. Firstly, looking at the winner(output) it is clear that the author has label encoded each team with a single number (eg. Mumbai Indians =1). Since, the count of matches win by a team throughout the seven seasons of IPL is different (ref fig 2), there will be a class imbalance problem. 12. The model doesn’t give enough information as to how such high accuracy was achieved. 13. Moreover, k-fold cross validation issue will be here also. There is no mention of train-test split. 14. 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 15. The paper is well structured, and the results seems legit. Unfortunately, the paper comes in the category of ODI cricket (50 overs). Comparing the results of ODI cricket prediction with IPL predictions seems not practical. Reasons: 16. There are huge data available for training a model in ODI cricket. On the other hand for IPL there are around 700 records only. 17. ODI matches are more predictable as compared to IPL (T20 format), as the player of a team remains the same in ODI, while for IPL the player changes each year for a team. 18. There is a significant ratio seen in the win loss ratio for an ODI records (eg. Australia is stronger than Bangladesh and the win percentage is significantly high for Australia), while on the other hand the win-loss ratio for IPL is pretty close when seen between 2 teams.   Conclusion: It’s not reliable to compare a result of a paper which claims to give such a high accuracy without any justification and support. Moreover, comparing different formats of cricket results is meaningless.  Thus, we want to setup a benchmark for other researchers to start predicting the winner of an IPL matches. As the data will grow in the near future the model will become more robust and the accuracy will increase. |
| 2. The dataset can be well understood if there is a visualization of the features distribution in the dataset. | The required changes have been incorporated. Thank you |
| 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 required changes have been incorporated. Thank you |
| 4. The citation for Sklearn is missing despite using their work. | The required citation is now provided at the reference section at number 19. Thank you |
| 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 | The ambiguity parameter was added in the tables of each model and the final table. However, we have added the ambiguity in every result also. Thank you |
| 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. Thus, we took care that model does not underfit. |
| 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 trained SVM, k-nearest neighbor, and Logistic Regression in geometry-based classifiers. All the three model accuracies came in the range of 56-58, on the other hand tree-based algorithms (Random Forest, Ad boost, Extra Tree Classifier) accuracies were above 60 after exhaustive hyper-parameter tuning. We have added the results of SVM and k-nearest neighbor also. |
| 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? | We have analyzed several research papers in sports prediction. Most of the papers uses the same algorithms. Since our problem lies under the supervised learning category, we have used the top 6 algorithms for training. We had trained the SVM and k-nearest neighbor also, but the results were not so promising, and the accuracy was almost same as Naïve Bayes, so we discarded the algorithm to make the paper not very lengthy. However, for now we have included the section of SVM and in our paper. Thank you |
| 10. The Study does not provide the % of train - test split used for the models on the data. | Under “Addressing the Symmetry in Data” section: the training and testing data is split in 90:10 ratio(incorporated the changes).  Under “Data Split” section: We have 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. | Thank you for noticing. The changes have been done. |
| 12. The Study does not provide enough information on how to reproduce this experiment. | We have explained exhaustively all the steps that are used for this project.  Step 1: Collection of data from various sources.  Step 2: Clean the data  Step 3: Using different methods to create features  1: Dream 11 algorithm  2: AHP method  3: Finding ranks using AHP  4: Win Rate  5: Team Points  6: We have defined the formulas to create the intersection features and the need to do it.  Step 4: Addressing the issue of ambiguity and how to deal with it.  Step 5: We have mentioned how we split the data, and all the hyperparameters that we have used for each model tuning. Moreover, the best result is shown with the best parameters. Thus the authenticity of the model is easily verifiable.  It would be very helpful if you can guide me exactly where the confusion is while trying to reproduce. |
| 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. |

Comparing with other models, The model[26] reported accuracy of 89.15 % for Random Forest and 19.02% for Naïve Bayes for the same model, this inconsistency tells us that there might be data leakage in the model. The model[27] is also similar, they reported an accuracy of 75% but the paper doesn’t tell us how they achieved such high accuracy. Our paper has explained the all the features clearly, how they were created and how it deals with the different problems in a small dataset based prediction problem.