|  |  |
| --- | --- |
| Reviewer 3 comments to authors:  Generic Discussions, Novelty is leveraged by brands which will always changing. This can be applied to a static database, and not dynamic. Whenever there is a change in playground, weather conditions and injured players etc the system may not work.  Dataset is not properly mentioned and non-clarity of prepossessing process.  Random forest and tree classifiers only two algorithms used. No comparative reasons shown of previous years ipl matches with the current ones. Broad recommendations and few conclusions. Literature survey not much impressive. | **Under Novelty Section**- We have incorporated the changes. Under Application Section –“It can be leveraged by brands, sponsors and advertisers to keep up their marketing strategies.”Here we mean that the model which we have designed can be used by the agencies for their benefits. It is true that we have trained this model to a database which is up to the year 2019. Each year for IPL the database needs to be updated and the new values should be appended. The model is designed such that all the matches will be predicted before the match and not in between. The database should incorporate the updated conditions like weather, player status etc. **Dataset**  Thank you for your valuable input. We have now added the features table to easily understand the data that were scraped from the sites.  **We have used 8 algorithms for training:** Naïve Bayes, Logistic Regression, Support Vector Machines, k- Nearest Neighbours, ADABOOST, XGBOOST, ExtraTreesClassifiers, Random Forest Classifier. As per your guidance, we have now added 2 more algorithms in our result: Support Vector Machines, k- Nearest Neighbours. Earlier, we added graphs that were relevant and helpful for the prediction of the results to make our paper crisp. As per your guidance, we have now included features distribution graphs as well. These graphs show the range of values and the counts of each feature.    In our paper, we have **recommended** the use as: ***“***It can be leveraged by brands, sponsors and advertisers to keep up their marketing strategies.*”* The above use case is only possible if the companies are able to predict the result of a match. The **conclusion** states the effectiveness of the model while predicting the result of a match.  **Literature survey**  We have researched and read almost every paper in the domain of IPL Result prediction. However, we tried to replicate the result but couldn’t do so due to the following reasons/comments on every paper that we studied:  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 soon the model will become more robust and the accuracy will increase.  Hence, we have written only 2 papers in our related work section.  We have incorporated the changes in the introduction section and included the part of how this paper is different than the others. Thank you. |
| Reviewer 4 comments to authors:    Comments to authors: The study would have been conducted on statistics of 8 regular playing IPL playing teams. The teams like GL and RPS can be avoided. There are several grammatical mistakes. Sentence formation is not up to mark. Spellings are misspelled, like in the second last line of the conclusion section "can" is written as "cab". Better to check the research paper on Grammarly. Avoid the use of words like "we", "our". Keep the terminology of evaluation parameters the same. For eg. in Table 11 "Real Test" is used which should have been "Real Test Accuracy". | Under Class Imbalance Section we have discussed this issue: “We noticed that predicting the results using the team's name is not feasible as it can cause a massive Class Imbalance between the groups. For example – MI (Mumbai Indians) winning more than 100 matches whereas KTK(Kochi Tuskers Kerala) wining less than 10 matches is a Class Imbalance problem. To rule out the class imbalance, we decided to design our model to predict the winner based on the essential features instead of the Team names, declaring either Team 1 or Team 2 as a winner. Moreover, we also discovered that the number of times Team 1 won is more than Team 2. To resolve this issue and balance the Team 1 winning and Team 2 winning in the label column, we interchanged a few values of column Team 1 with column Team 2.” Since GL team was almost same as the CSK, and RPS was like Pune Warriors we have assumed them to be the same. Moreover, most of the features in our model depends upon the past performance of the player irrespective of the team. However, while calculating the Rank using the AHP process we have considered this issue and removed the teams which haven’t played against each other.  Grammatical mistakes: Thank you for noticing the errors and suggesting the tool. We have corrected all the mistakes using Grammarly now.  We have corrected the terminology of evaluations name, and removed words like “we”, “our”. |
| Reviewer 5 comments to authors:    The article need minor modification:    Q1: The authors have not followed format in a proper manner. It should be corrected and the every new headings should be numbered instead of only keeping in bold format.    Q2: In abstract, the text are not properly spaced. Each new words should be separated by whitespace. | A1: Thank you for the review, we have incorporated the changes.  A2: Thank you for the review, we have incorporated the changes. ExtraTreeClassifier and XGBoost are written together only. |
|  |  |