**Reviewer’s Report**

|  |  |
| --- | --- |
| **Manuscript No.** | 25AUG-48 |
| **Manuscript Title** | Machine Learning-Based xG Forecasting in the Top 5 soccer Leagues: A Comparative Analysis |

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **Manuscript Title** | |  | | | |
| **1st REVIEW** | | | | | |
| **No** | **REVIEWER’S COMMENTS** | | **Author’s Response** | **Author’s Modifications**  **reflected in the paper** | **Reviewer's confirmation** |
| **1** | **Title: The term "Top 5 Soccer Leagues" is vague. Specify that these refer to European leagues for precision.** | | **We thank the reviewer and have clarified in the title and text that these leagues are European.** | **Revised the title and text to specify European leagues (changes highlighted in red).** |  |
| **2** | **The Abstract lacks quantitative depth in some claims (e.g., vague statements like “notably improved accuracy”). Moreover, resolve some redundancy and verbosity (e.g., restates results multiple times). Revise to be more concise and focused on novelty, methods, and implications.** | | **Thank you for this valuable feedback. We have revised the Abstract to include specific quantitative results, removed redundant statements, and focused the text on the novelty, methodology, and key implications of our study.** | **Revised the Abstract by removing redundant phrases (strikethrough) and adding quantitative details (red font) to highlight novelty, methods, and key results.** |  |
| **3** | **The introduction lacks of clear research gap: The intro references general past limitations but fails to explicitly define what previous models lacked and why the proposed features (coach, tenure, tactics) matter. Define the problem more explicitly. What exactly are you solving? What specific shortcoming of past xG models is this study overcoming? Why are coaching features underexplored, and how are they expected to improve prediction?** | | **Thank you for your feedback. We have revised introduction section according to your comment** | **Revised the Introduction by adding details in the red color** |  |
| **4** | **Missing explicit hypothesis/objective: It does not clearly define testable hypotheses or research questions.** | | **Thank you for this valuable feedback. We have revised the Introduction to include specific explanations removed redundant statements** | **Revised the Introduction by removing redundant phrases (strikethrough) and adding quantitative details (red font) to highlight explanations** |  |
| **5** | **Vague references: Some citations are broad, and a few are non-peer-reviewed sources.** | | **We thank the reviewer for this comment. While our manuscript includes both peer-reviewed and non-peer-reviewed sources—since some statistical and soccer-specific data were obtained from reputable websites and blogs—we have now strengthened the reference list by adding six peer-reviewed studies, listed as references 31–36** | **Added six new peer-reviewed references (31–36) in the References section and updated the Related Work citations to include these studies, highlighted in red; non-peer-reviewed sources are retained only for statistical and soccer-specific data.** |  |
| **6** | **Dataset size: Only 2,917 observations across five leagues is limited, especially for ML models. How was sample representativeness ensured?** | | **We thank the reviewer for this insightful comment. We acknowledge that the dataset, consisting of 2,917 observations across five leagues, is relatively limited for machine learning models. Nonetheless, we ensured representativeness by including all matches with complete player and event data during the study period, covering diverse teams, player roles, and match situations. We have also noted this as a limitation in the manuscript** | **Acknowledged the dataset size as a limitation in the response; no changes were made to the manuscript.** |  |
| **7** | **Feature encoding details are vague: It is unclear how many coach/tactic categories were present and how multicollinearity was handled post-encoding.** | | **We thank the reviewer for this comment. Categorical features such as coaches and tactical approaches are encoded using Label Encoding, with each category converted into a unique numerical value. Although we did not explicitly calculate multicollinearity metrics, our regression models, such as Lasso and Ridge, inherently mitigate multicollinearity through regularization, reducing the impact of highly correlated features on model performance. We also verify the number of unique categories to ensure proper representation and avoid sparse or imbalanced features in the model** | **Clarified that categorical features are label-encoded, number of unique categories is verified and explained in the red color in the section 3.2 and multicollinearity is mitigated through Lasso and Ridge regularization** |  |
| **8** | **No train-test split or cross-validation: The methodology does not clearly explain model training and evaluation protocols (e.g., k-fold CV, train-test ratio).** | | **We thank the reviewer for this valuable comment. We acknowledge that the manuscript did not provide detailed information on the model training and evaluation protocols. In our study, regression models are trained and evaluated using standard practices, including appropriate train-test splits, to ensure reliable performance assessment and prevent overfitting.** | **Clarified that regression models are trained using train-test splits to ensure robust evaluation.** |  |
| **9** | **Data leakage risk: Was the same data used to select features and test models?** | | **We thank the reviewer for raising this important point. We confirm that the dataset is split into separate training and testing sets. Feature selection is performed only on the training set, and the testing set is kept entirely independent to evaluate model performance, thereby avoiding any risk of data leakage** | **Clarified that feature selection is performed on the training set only, and the testing set is independent to prevent data leakage** |  |
| **10** | **No statistical significance testing: Differences in MSE or R² are small (e.g., 0.50 → 0.49). Without p-values or confidence intervals, it is unclear if improvements are meaningful.** | | **Thank you for this valuable comment. We did not use k-fold cross-validation in this study; therefore, formal p-values or confidence intervals for the reported metrics cannot be provided. However, the consistent improvements observed across multiple models—particularly the Decision Tree—demonstrate the practical benefit of including the new features. We have added a discussion of these improvements in Section 4.2, highlighted in red in the manuscript, to clarify their impact. Future work may incorporate cross-validation and formal statistical tests to further quantify the significance of these improvements.** | **Added a discussion in Section 4.2 (highlighted in red) clarifying improvements in model metrics** |  |
| **11** | **No error analysis: Which predictions failed? What kinds of players or matches led to large errors?** | | **We thank the reviewer for this comment. Section 5 (highlighted in red) now includes an error analysis, noting that standout players, such as Robert Lewandowski, can cause larger discrepancies between predicted xG and actual goals.** | **Revised Section 4.3 to include a concise error analysis, highlighting that standout players, such as Robert Lewandowski, can cause larger discrepancies between predicted xG and actual goals. Changes are highlighted in red in the manuscript.** |  |
| **12** | **Temporal inconsistency: Training on 2020/21 data and validating against 2022/23 goals without clear explanation of assumptions introduces noise and limits real-time applicability.** | | **We thank the reviewer for this insightful comment. We acknowledge that using training data from the 2020/21 season and validating on the 2022/23 season introduces temporal inconsistencies, which is a limitation of the available dataset. This approach, however, allows us to assess the robustness and generalizability of our xG prediction model across seasons, simulating real-world scenarios where models trained on historical data are applied to future matches. Despite changes in team composition and player performance, the model captures underlying patterns in goal-scoring opportunities, maintaining reasonable predictive capability. We have revised Limitations section in the red color to address data restrictions, temporal gaps** | **We have revised the Limitations section to address data restrictions, temporal gaps, and other potential biases; the changes are highlighted in red in the manuscript.** |  |
| **13** | **Overclaims: Statements like “our model effectively predicts expected goals” are broad and unqualified.** | | **We thank the reviewer for this insightful comment. We have revised the statement in Section 5 to use “reasonable” instead of “effectively,” softening the claim and ensuring a more accurate representation of our model’s predictive performance. The changes are highlighted in red in the manuscript.** | **Revised Section 5 to replace “effectively” with “reasonably” when describing the model’s prediction of expected goals, softening the claim and providing a more accurate and cautious interpretation of the results. Changes are highlighted in red in the manuscript.** |  |
| **14** | **The conclusion implies general applicability but does not acknowledge limitations clearly. There is repetition of earlier claims without added insight. Further, this section does not stress what makes this study *original* in the ML + soccer xG literature.** | | **We thank the reviewer for this insightful comment. We have revised the Conclusion to adopt a more measured tone, emphasizing the study’s contributions and originality in integrating coaching-related features into xG prediction models while avoiding repetition of earlier claims. The Limitations are now clearly addressed in a dedicated section(Section 6), and the Conclusion focuses on the practical implications and key findings of our work. All revisions are highlighted in red in Section 7** | **Revised the Conclusion to emphasize study originality and practical implications, removed repetitive claims, and clearly refer readers to the dedicated Limitations section. Changes are highlighted in red in Section 7.** |  |

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **Manuscript Title** | |  | | | |
| **2nd REVIEW** | | | | | |
| **No** | **REVIEWER’S COMMENTS** | | **Author’s Response** | **Author’s Modifications**  **reflected in the paper** | **Reviewer's confirmation** |
| **1** |  | |  |  |  |
| **2** |  | |  |  |  |
| **3** |  | |  |  |  |
| **4** |  | |  |  |  |
| **5** |  | |  |  |  |
| **6** |  | |  |  |  |
| **7** |  | |  |  |  |
| **8** |  | |  |  |  |
| **9** |  | |  |  |  |
| **10** |  | |  |  |  |
| **11** |  | |  |  |  |
| **12** |  | |  |  |  |
| **13** |  | |  |  |  |
| **14** |  | |  |  |  |
| **15** |  | |  |  |  |