AN INSIGHTS ON CRICKET DATA ANALYTICS

Ms.J.S.Sowmiya B.Tech.,M.E[1], Thameem Ansari S[2], Spencer J[3], Vignesh C K[4]

Department Of Computer Science & Engineering, Vel Tech High Tech Dr.Ranagarajan Dr.Sakunthala Engineering

College, Avadi, India[1-4]

Abstract: From this project, we can review performance of the team using previous data of the cricket players. It also helps to track the performance of an individual player, which helps the team management and selection committee to select the best player for a particular tournament by analyzing the data. Data played a key role in analyzing the performance of a team. One team can analyze the teams with whom they are about to compete. They can analyze the opposition team's strengths and weaknesses through rigorous analysis of their scoring patterns, how they scored their runs, when they were vulnerable during innings.

Keywords: performance, analyze

I. INTRODUCTION

The main motto of reducing the time period of the game is to attract the audience and television viewers. Since the introduction of T20s the game has taken a turn worldwide which resulted in many premier leagues that turned out very successful. One of the best examples is the Indian Premier League. In this paper, our aim is to predict each player's performance, overall match's run-rate and suggest suitable best playing eleven players for an upcoming T20 match. Each player's performance includes Strike rate of the bats-man, Economy of the bowler and number of wickets the bowler can take in the upcoming match. For every prediction, there are certain parameters considered accordingly.

For example, prediction of economy of the bowler will be based on the parameters such as bowler's name, team they are playing against, venue of the match and past economies of that particular bowler. Since overall performance of the match is crucial, we predict best playing eleven players based on the past matches and the team they will be playing against. From this project, we can review performance of the team using previous data of the cricket players. It also helps to track the performance of an individual player, which helps the team management and selection committee to select the best player for a particular tournament by analyzing the data. Data played a key role in analyzing the performance of a team. One team can analyze the teams with whom they are about to compete. They can analyze the opposition team's strengths and weaknesses.

II. EXISTING SYSTEM

One of the first studies of semantic segmentation for cricket highlights generation involved the creation of a hidden Markov model (HMM) classifier to detect and classify exciting footage into the categories of 'ball movement,' 'fielding,' and 'wickets' (Kolekar & Sengupta, 2004). Using an HMM is advantageous compared to traditional classifiers because it utilizes previous frames, making it robust to misclassification of outlier footage. By using the estimated motion in each cricket frame as a likelihood function for the classifier, Kolekar and Sengupta achieved an 87.5% average out-of-sample classification accuracy.

In a follow-up study, Kolekar et al. proposed a hierarchical framework to detect and classify exciting footage, which they call "events" (Kolekar et al., 2008). They created a five-level tree classifier that makes use of audio, video, and color features to label frames. By employing a hierarchical tree framework, the authors reduced the computational complexity of their classifier by only considering a small number of features, such as the density of green pixels for detecting 'field views,' at each level of the tree. This classifier achieved an 80+% out-of-sample recall (number of correctly classified events divided by the total number of events) across all five levels. While the accuracy of this classifier was lower than their previous model for certain categories, it greatly improved upon their previous work by delivering much more detailed classifications, ranging from the region of the field displayed to the type of player in frame.

III. PROPOSED SYSTEM

Measuring player performance is a necessary mode of feedback for players and vital for team selection for organizations. For batsmen, the batting average has been an important historical device for quantifying performance, with early work focused on including not-out scores (Damodaran, 2006; Wood, 1945). Since then, several researchers have attempted to diversify performance metrics by introducing new statistics like batting performance or methods like neural networks, HMMs, and Bayesian approaches (Amin & Sharma, 2014; Beaudoin & Swartz, 2003; Iyer & Sharda, 2009; Koulis et al., 2014; Lemmer, 2004; Stevenson & Brewer, 2017, 2021). For example, Stevenson and Brewer (2017) have proposed a Bayesian hierarchical model that measures the quality of a batsman by estimating a batsman's initial batting ability, peak batting ability, and the time required to achieve peak batting ability during an inning. The pair later extended the model by introducing Gaussian processes to capture fluctuations over a player's career and parameters to account for match-specific conditions (Stevenson & Brewer, 2021). Historically, bowler performance has received less attention. The combined bowling rate, built on several bowling statistics such as average runs conceded and average strike rate, remains an important metric in use (Adhikari et al., 2017; Lemmer, 2002). Recent work focusing on Test cricket has attempted to put batsmen and bowlers on a common scale for performance comparison (Akhtar et al., 2015).

IV. OBJECTIVE

To predict the outcome of the cricket match result using data based on the historical and current data. Collecting insights from the dashboards to select a team of 11 players that have a winning combination. Getting visual interpretation of player past performances. It also helps to track the performance of an individual player, which helps the team management and selection committee to select the best player for a particular tournament by analyzing the data. Data played a key role in analyzing the performance of a team. One team can analyze the teams with whom they are about to compete. They can analyze the opposition team's strengths and weaknesses through rigorous analysis of their scoring patterns, how they scored their runs, when they were vulnerable during innings.

MODULES:
- ➢ Data Collection
- ➢ Data Cleaning
- ➢ Data Transformation
- ➢ Data Modelling
- ➢ Data Visualization
- ➢ Data Interpetration
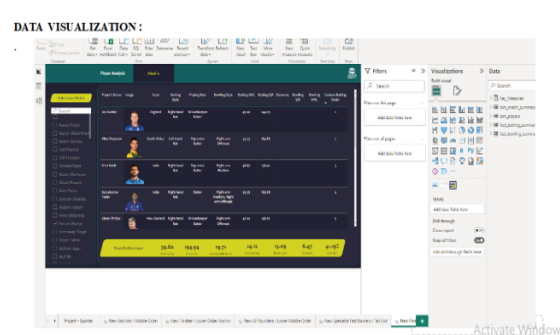
RESULTS:

V. LITERATURE SURVEY

[1] Tripathi et al. (2020) implemented IPL match prediction using machine learning while tackling ambiguity in results. They collected historical data from various sites and created their player and team databases. They implemented feature engineering in which they included city, toss winner and toss decision. Also, for player's data, they added features such as the batting score of a player, bowling score of a player, a total score 7 of a player, and team strength. An analytic hierarchy process was used to obtain the weights used to calculate batting and bowling features. They implemented models using Na¨ıves Bayes, AdaBoost, logistic regression, support vector machine, KNN, XGBoost, extra tree classifier and random forest classifier. These models were compared based on accuracy, cohen kappa and

ambiguity to identify the issue of data symmetry. Models performance was tuned using hyperparameter tuning, and the random forest model was the best model with a standard deviation of 6.3%. **[2]Kapadia et al. (2019)** discussed whether machine learning could help predict accurate match results of IPL matches. They designed models to predict the outcome of the match based on the home ground factor and toss decision. Feature selection was applied to eliminate irrelevant features from the dataset. Home ground and toss winner features were created after the data processing. Their results indicated that toss results features performed slightly better to predict the match's outcome using the KNN algorithm based on accuracy, precision and recall. Na¨ıves Bayes results were low for toss related features, but for home ground related features Na¨ıves Bayes performance was reasonable compared to other algorithms.**[3] Banasode et al. (2020)** implemented a system to predict and analyze the results of IPL matches. Their study used the IPL dataset to analyze the runs scored by batsmen of each team over the years, batsmen performance across all venues to find their favorite venue. They calculated the bowler's economy rate and analyzed bowlers for the economy rate over the years. Further, they analyzed the performance of batters against all bowlers and the distribution of wickets over the years. In their analysis, they found the toss decision influences the outcome of the match.

[4]Sinha et al. (2020) implemented a system to predict wins to improve the team performance using a support vector machine. They scraped the data from the official site of the Indian Premier League. They quantified player performance using multiple statistical indicators, and a rating index was established. The model was then run on the players chosen for both sides, and a prediction of win or loss probability was provided. Similar players were identified using KNN and K-mean using the data of their performance. This model was able to identify the replacement of injured players. They obtained 96.3% accuracy for the SVM model.**[5] Nimmagadda et al. (2018)** used statistical techniques to predict the outcome of a T20 match while it was still in progress. A multiple regression model is evaluated in order to create a prediction model. The key outcome was determined by the impact of the toss winner and the resulting match-winner. The predictive model used the innings score at regular intervals and the final scores to anticipate the match result. The model predicted score and run rate projected score was relatively close to the final score, with the model's score being more accurate when compared to the actual score.

[6]Jayalath (2018) proposed an approach for analyzing one-day international cricket match predictors. External factors sometimes influence the outcome of a match or players performance. Factors such as home advantage, toss result, batting first or second, and day vs day-night game format are discussed in this research. They formulated a logistic regression that predicts a team winning against a particular opponent based on previous results. They included predictors such as a home game, day game or day-night game, toss, batting first, opponent belongs to which continent in their model. This model suggested that home advantage is a significant factor for the majority of the teams. They categorized data based on the day and day-night games for further analysis and found that toss is a significant factor. Based on this result, they decided to use the result of toss to build a classification tree model to predict the outcome of a match. Further, they changed the predictor, which was an outcome of the match, to the margin of victory and performed a regression tree approach as this method helped interpret the results more clearly.

CONCLUSION

Thus, the role of this project is to predict or choose the essential team of players for the future matches Selection of the best team for a cricket match plays a significant role for the team's victory. The main goal of this paper is to analyze the cricket data and predict the players' performance. This knowledge will be used in future to predict the winning teams for the next series matches. Hence using this prediction, the best team can be formed. This project opens scope for future work in the field of cricket and predicts other important things like best team of players, best venue, best city, best fielding decision to win a match.

REFERENCES

1. Williams, J. Cricket and England: A Cultural and Social History of the Inter-War Years; Taylor & Francis: Oxfordshire, UK, 1999.
2. Bailey, M.; Clarke, S.R. Predicting the match outcome in one day international cricket matches, while the game is in progress. J.Sports Sci. Med. 2006,5, 480.
3. Rehma, A.A.; Awan, M.J.; Butt, I. Comparison and Evaluation of Information Retrieval Models. VFAST Trans. Softw. Eng.2018,6,7–14.
4. Alam, T.M.; Awan, M.J. Domain analysis of information extraction techniques. Int. J. Multidiscip. Sci. Eng. 2018,9, 1–9.
5. Kaur, A.; Kaur, R.; Jagdev, G. Analyzing and Exploring the Impact of Big Data Analytics in Sports Sector. SN Comput. Sci.2021,2,1–19. [CrossRef]

6. Ahmed, H.M.; Awan, M.J.; Khan, N.S.; Yasin, A.; Shehzad, H.M.F. Sentiment Analysis of Online Food Reviews using Big DataAnalytics. Elem. Educ. Online 2021,20, 827–836.

7. Zhou, L.; Pan, S.; Wang, J.; Vasilakos, A.V.J.N. Machine learning on big data: Opportunities and challenges. Neurocomputing2017,237, 350–361. [CrossRef]

8. Aftab, M.O.; Awan, M.J.; Khalid, S.; Javed, R.; Shabir, H. Executing Spark BigDL for Leukemia Detection from Microscopic Imagesusing Transfer Learning. In Proceedings of the 2021 1st International Conference on Artificial Intelligence and Data Analytics(CAIDA), Riyadh, Saudi Arabia, 6–7 April 2021; pp. 216–220.

9. Awan, M.J.; Khan, M.A.; Ansari, Z.K.; Yasin, A.; Shehzad, H.M.F. Fake Profile Recognition using Big Data Analytics in SocialMedia Platforms. Int. J. Comput. Appl. Technol. 2021, in press.

10. Anam, M.; Ponnusamy, V.; Hussain, M.; Waqas Nadeem, M.; Javed, M.; Guan Goh, H.; Qadeer, S. Osteoporosis Prediction forTrabecular Bone using Machine Learning: A Review. Comput. Mater. Contin. 2021,67, 89–105. [CrossRef]

11. Ali, Y.; Farooq, A.; Alam, T.M.; Farooq, M.S.; Awan, M.J.; Baig, T.I. Detection of Schistosomiasis Factors Using Association RuleMining. IEEE Access 2019,7, 186108–186114. [CrossRef]

12. Nagi, A.T.; Awan, M.J.; Javed, R.; Ayesha, N. A Comparison of Two-Stage Classifier Algorithm with Ensemble Techniques On Detection of Diabetic Retinopathy. In Proceedings of the 2021 1st International Conference on Artificial Intelligence and Data Analytics (CAIDA), Riyadh, Saudi Arabia, 6–7 April 2021; pp. 212–215.

13. Awan, M.J.; Yasin, A.; Nobanee, H.; Ali, A.A.; Shahzad, Z.; Nabeel, M.; Zain, A.M.; Shahzad, H.M.F. Fake News Data Explorationand Analytics. Electronics 2021,10, 2326. [CrossRef]

14. Gupta, M.; Jain, R.; Arora, S.; Gupta, A.; Javed Awan, M.; Chaudhary, G.; Nobanee, H. AI-enabled COVID-9 Outbreak Analysisand Prediction: Indian States vs. Union Territories. Comput. Mater. Contin. 2021,67, 933–950. [CrossRef]

15. de Souza Junior, A.H.; Corona, F.; Barreto, G.A.; Miche, Y.; Lendasse, A. Minimal learning machine: A novel superviseddistance-based approach for regression and classification. Neurocomputing 2015,164, 34–44. [CrossRef]

16. Javed, R.; Saba, T.; Humdullah, S.; Jamail, N.S.M.; Awan, M.J. An Efficient Pattern Recognition Based Method for Drug-DrugInteraction Diagnosis. In Proceedings of the 2021 1st International Conference on Artificial Intelligence and Data Analytics(CAIDA), Riyadh, Saudi Arabia, 6–7 April 2021; pp. 221–226.

17. Salloum, S.; Dautov, R.; Chen, X.; Peng, P.X.; Huang, J.Z. Big data analytics on Apache Spark. Int. J. Data Sci. Anal.2016,1,145-164

18. Khalil, A.; Awan, M.J.; Yasin, A.; Singh, V.P.; Shehzad, H.M.F. Flight Web Searches Analytics through Big Data. Int. J. Comput.Appl. Technol. 2021, in press.

19. Singh, T.; Singla, V.; Bhatia, P. Score and winning prediction in cricket through data mining. In Proceedings of the 2015International Conference on Soft Computing Techniques and Implementations (ICSCTI), Faridabad, India, 8–10 October 2015; pp.60–66.

20. Kamble, R. Cricket Score Prediction Using Machine Learning. Turk. J. Comput. Math. Educ. 2021,12, 23–28.