Data-Driven Cricket: A Machine Learning Approach to IPL Score Prognostication

S. Lalitha, Anurag Das, Prasanth Ayitapu, Rohit S Nair, Satvik Raghav
Department of Electronics and Communication Engineering
Amrita School of Engineering, Bengaluru
Amrita Vishwa Vidyapeetham, India
s_lalitha@blr.amrita.edu, anuragdas070@gmail.com, payitapu@gmail.com, rohitsnair2004@gmail.com, satvikraghav007@gmail.com

***Abstract*—The Indian Premier League (IPL) is a very popular and globally recognized cricket tournament, attracting many spectators worldwide. The proposed work aims to predict cumulative scores of teams in IPL using attributes such as batting and bowling team data, total runs, wickets, overs, runs scored in the last 5 overs, batting taken in the last 5 overs, and Total Score. This work employs Machine Learning (ML) techniques such as Linear Regression (LR), Random Forest (RF), and Support Vector Machine (SVM) models to predict IPL scores. The database considered in this work was carefully obtained from well-known websites such as ESPN and Cricbuzz, ensuring the reliability and credibility of the research. In addition to the prediction models, a user-friendly graphical user interface (GUI) is implemented for easy interaction with the prediction system. This GUI allows the user to enter the necessary match data and displays the predicted score. Uniquely, this system offers a range of points that can be extended up to 5 runs from the estimated score. This range recognizes the unpredictable nature of cricket matches and provides more realistic and practical predictions. The combination of various ML models and an intuitive GUI makes this system a powerful tool for IPL score prediction.**

***Index Terms*—Linear Regression, Random Forest, Support Vector Machine, One-hot Encoding**

I. Introduction

Cricket, as a sport, has achieved unprecedented global popularity, with leagues such as the Indian Premier League (IPL) emerging as a cultural phenomenon that transcends geographical boundaries. IPL, known for its exciting matches and star-studded lineups, has become a cornerstone of the world of cricket [1], enthralling audiences across the globe. The study is rooted in the dynamic field of the IPL, focusing on predictive modeling of cumulative team scores – an important aspect of the league that contributes to its intense and unpredictable nature.

To achieve this, we used a combination of LR, RF and SVM models. Drawing upon inspiration from diverse applications like the detection of Parkinson Speech [2], Analysis of Heart Disease Prediction [3], these models have established themselves as robust frameworks in the field. The integration of these three models provides a comprehensive and multifaceted approach to predictive modeling, allowing for more robust and accurate predictions of team scores. LR, a basic and widely used statistical method, adds a level of simplicity and interpretability to the model, making it easier to understand the relationships underlying the data. RF, known for its ensemble learning capabilities, enhances the model's ability to handle complex datasets and reduce overfitting. SVM, with its strong classification skills, ensures high accuracy in prediction accuracy. Using all three models gives us a broader perspective and a deeper understanding of the factors influencing IPL team scores. This diversified approach underscores our commitment to accurate analysis and predictive precision, leveraging the strengths of each model to increase overall efficiency and reliability.

The dataset employed in this study is carefully curated from reputable platforms such as ESPN and Cricbuzz, which ensures the reliability and validity of the data used in the predictive model. By integrating LR, RF and SVM with high-quality data, this work aims to provide valuable insights into IPL team score prediction, which contributes to the ongoing development and excitement surrounding this global sporting event.

II. Literature Survey

In the proposed work of IPL score prediction, this literature review evaluates seminal studies using different ML techniques in cricket analytics. Through this brief survey, the goal is to explore concepts, identify effective methodologies, and highlight areas that require further research while contributing to the growing discourse in sports analytics.

Ishan Jain et al. utilized ML algorithms like RF Classifier and Multivariate Polynomial Regression (MPR) to predict the final score and winner of the match. The accuracy of the prediction model was reported to be 67.3% for MPR and 55% for RF classification. A distinctive feature of this work is the integration of a data mining module with a corresponding visualization that improves the predictive and analytical capabilities of the system [4]. Similarly, the study on Emotion Variation Detection [5]has used the RF classifier and obtained remarkable results.

Manoj Ishi et al. used ML techniques such as LR, NB Bayes, KNN, SVM, decision tree, RF, GBM, XGBoost, and CatBoost. The accuracy achieved varies, with SVM reaching 93.54% and the highest when combined with the LR feature optimization reaching 94.28%. A distinctive feature of this work is the use of nature-inspired algorithms for feature selection leading to improved accuracy compared to other works [6].

Yogesh Kumar et al. utilized ML techniques such as Multinomial NB Bayes, LR, Ridge Classifier, RF, SGD Classifier, and Decision Tree to predict sentiment in IPL using Twitter data. The accuracy achieved by this method, Ridge Classifier reached the highest accuracy of 90.27%. A distinctive feature of this work is the use of a large dataset of 7 million tweets and the use of an optimization methodology to improve classification accuracy [7].

Parmeet Kaur et al. utilized ML techniques such as KNN. The accuracy of the KNN algorithm with k=4 was observed to be around 71%. The distinguishing feature of this work lies in its dynamic approach to the prediction of match results and the use of the non-relational HBase database for scalability [8].

Amitabha Chakrabarty et al. centered the use of SVM with a linear kernel. For the linear kernel SVM, the model's accuracy is reported to be 92%. Important characteristics are chosen using feature selection algorithms like similarity selection and duplicate feature reduction [9].

E. L. Lekamge et al. proposed a prediction model of cricket player performance using LR, SVM with linear and polynomial kernel. The accuracy of this model was 91.5% for Tamim, the batter, and 75.3% for Mahmudullah, the bowler. Relevant attributes are extracted using feature selection methods like similarity selection and duplicate feature deletion [10].

Saranya G et al. utilized ML techniques such as KNN and RF to predict the performance of players and select the top 15 players for IPL. The work achieved an average accuracy rate of 94% for analyzing IPL match results. The distinctive features of this work include the use of data cleaning techniques, the study of batting and bowling performances, and special data extraction for analysis [11]. The reseach conducted on Mixed-lingual Affective State Recognition System [12] also employed the RF model as well.

From the literature review performed, it is observed that in cricket analytics, a spectrum of ML algorithms has been used to predict scores, player performances, and match outcomes. Some models include LR, Naive Bayes, KNN, SVM with linear and polynomial kernels, Decision Tree, RF, and Gradient Boosting Machine, along with other advanced techniques like Ridge Classifier, SGD Classifier, Multinomial NB Bayes, MPR, GNLM, K-Means Clustering, CNN, and MLP Neural Network. These algorithms are suitable for a wide range of analytical problems, using optimization methodologies, extensive databases, and feature selection techniques to improve accuracy in cricket analytics.

The current literature on cricket analysis reveals several research gaps. First, there are inconsistencies in the reported accuracy of ML models used for tasks such as score prediction and match outcome prediction, suggesting the need for more reliable assessment methods. The prediction models lack real-time data integration, ignoring the potential benefits of timely information during live matches. There is not enough emphasis is placed on the interpretability and explanatory power of the model, which hinders the understanding and acceptance of the prediction model by cricket stakeholders. The current research lacks the integration of a user-friendly GUI website to improve the accessibility and usability of prediction models. Closing these gap can lead to deeper and more reliable insights into cricket analysis, thereby increasing the effectiveness and application of predictive models in real-world scenarios.

III. METHODOLOGY

Fig. 1 shows the detailed workflow of this proposed architecture which is explained below.

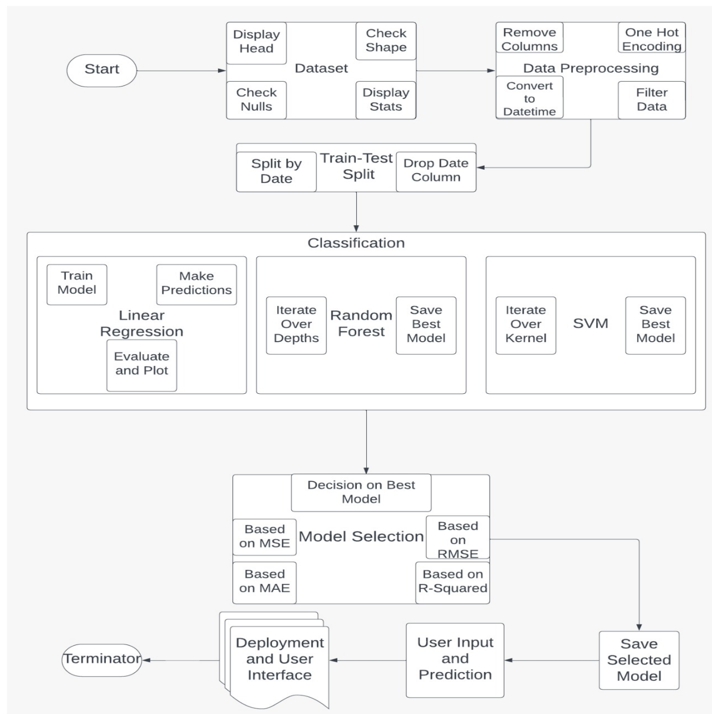

Fig. 1. Architecture of the proposed model

A. Dataset

The dataset used in this work is carefully collected from reputable sources including ESPN [13] and Cricbuzz [14], known for their comprehensive cricket statistics. The dataset consists of a significant size of 76,015 samples and consists of 15 attributes which include Mid, Date, Venue, Bat_Team, Bowl_Team, Batsman, Bowler, Runs, Wickets, Overs, Run_Last_5, Wickets_Last_5, Striker, Non-Striker, and Total. The dataset captures various match-related parameters and player performance metrics. Spanning the years from 2008 to 2017, this comprehensive dataset serves as a rich repository of information for the analysis. Specifically, data up to the year 2016 is designated as the training set, while data after 2017 constitutes the testing set. This strategic division allows for a robust investigation of the temporal evolution of the dataset and enables the development and evaluation of predictive models over different time frames.

B. Data Pre-processing

In this stage, the database is expanded to filter and organize the data for optimal analysis. Unnecessary features have been removed, and focus on important features like Bat_Team, Bowl_Team, Runs, Wickets, Overs, Runs_Last_5, Wickets_Last_5, and Total as it is represented in TABLE I. This selective approach ensures that the database captures the key elements for the predictive model in this research work. To improve the categorical variables, the process of one-hot-encoding [15] is used, changing the categorical characteristics into a suitable format for quantitative analysis. One-hot Encoding is used to convert categorical features ('bat team' and 'bowl team') into binary columns, improving the model's understanding of team data. This process results in a 'cat df' dataframe containing one-hot-encoded columns alongside essential features like 'runs', 'wickets', and 'overs', preparing the data for ML model training. Additionally, to maintain consistency and relevance in the analysis, the database was filtered based on consistent teams across the league. The successive teams selected are Kolkata Knight Riders, Chennai Super Kings, Rajasthan Royals, Mumbai Indians, Kings XI Punjab, Royal Challenger Bangalore, Delhi Daredevils, and Sunrisers Hyderabad. This refined database, now stripped of redundancies and enriched with relevant features, forms the basis for further analyzes and predictive models.

C. Classification

In the classification phase of the analysis, three different ML models are implemented - LR, RF, and SVM.

For LR, this process includes building a model using the training set, making predictions using the built model, and then evaluating the results. The results are carefully evaluated and illustrated using several plots that provide insight into the model's performance and predictive ability.

The RF model undergoes a more complex iterative process, exploring a range of tree depths to determine the optimal configuration. Performance metrics such as RMSE, MSE, MAE, and R^2 are used to determine the effectiveness of the model at each depth. The best model, as determined by these criteria, is selected and retained for further analysis and comparison.

SVM models systematically explore different types of kernels, including linear, polynomial, radial basis function (RBF), and sigmoid kernels. By continuously evaluating the performance of the model with each type of kernel, it is identified, and the most effective kernel is identified based on performance metrics such as RMSE, MSE, MAE, and R^2. This sensible approach ensures that the chosen kernel type improves the prediction accuracy and forms the basis for a detailed evaluation of the SVM's classification capabilities.

The evaluation prominently used for the IPL score prediction are MAE, MSE, RMSE, R^2 [16].

D. Model selection and saving best model

In the critical stage of model selection, it is systematically evaluated and compared the performance of three ML models—LR, RF, and SVM. The decision-making process involves a thorough evaluation based on key performance metrics such as MSE, MAE, RMSE, and R^2. By examining the results of the model against these criteria, the most effective and reliable model is determined. The selected model, which is deemed superior in prediction accuracy and overall performance, is then retained. This systematic approach ensures that the chosen model, whether it is LR, RF, or SVM, represents the optimal solution to the IPL score prediction problem.

IV. IMPLEMENTATION AND ANALYSIS

The evaluation in this work is simulated using Python programming. Initially, the feature vector consists of 15 attributes, which are reduced to 8 after pre-processing. The pre-processed data includes important elements like Bat_Team, Bowl_Team, Runs, Wickets, Overs, Runs_Last_5, Wickets_Last_5, and Total. The work is carried out in two phases:

Phase 1 deals with model implementation and analysis. The three models— LR, RF, and SVM—are applied to predict IPL scores based on selected attributes after pre-processing. The line plots is also being included to have better analysis of the models. The models are compared with respect to year-wise score average and their performance metrics.

Phase 2 involves Graphical User Interface (GUI) Integration to improve user interaction. The GUI provides a user-friendly interface for visualizing and interpreting predictions with LR, RF, and SVM models for better exploration and understanding of the analysis results through the GUI, facilitating a seamless and interactive experience. GUI integration increases the practical utility of the model's predictions, making the results more accessible and convenient for stakeholders.

PHASE 1:

A. Proposed IPL score prediction framework using LR

As per the Feature vectors of this work, input is given to the LR model and Output is predicted. MAE is a measure of the average absolute difference between the predicted value and the actual value, was calculated as 12.12. This means, the model predictions on average deviates by approximately 12.12 units from the actual values. MSE, is the measure of average squared difference between the predicted value and the actual value, gave a value of 251.01. RMSE, calculated as the square root of the MSE, was found to be 15.84, indicating the average size of the forecast error. The R^2 value, a statistical measure of model fit, was determined to be 0.75. This R^2 value shows that approximately 75.23% of the variation in the dependent variable (IPL score, in this case) can be explained by a LR model.

TABLE I
PREDICTED EVALUATION METRIC VALUES USING LR MODEL

Evaluation Metrics	Values
MAE	12.118617546193295
MSE	251.00792310417438
RMSE	15.843229566732106
R^2	0.7522633566350527

To have a further in-depth analysis, the line plot is shown in Fig. 2.

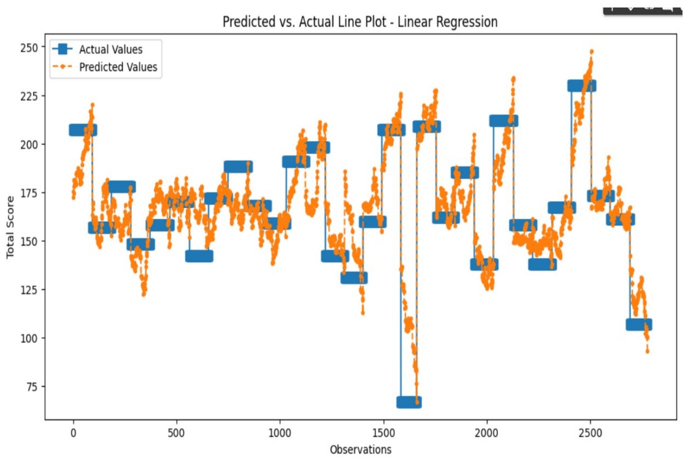

Fig. 2. Graph showing analysis of line plot using LR model

From Fig. 2, it is observed that the line plot shows the range of predicted scores between 60 and 240 and shows the ability of the model to provide diverse predictions across the spectrum of possible scores. In particular, the predicted line consistently matches the actual score, even for outliers, indicating the robust performance of the model in capturing different scenarios. Closer inspection reveals a cluster of data points and prediction lines between the range 140 and 175, indicating a higher frequency of accurate predictions in this range of scores. This clustering confirms the effectiveness of the model in consistently predicting scores, especially for IPL matches in the range of frequent observations.

B. Proposed IPL score prediction framework using RF

Evaluation metrics for the RF model are calculated at different max depth values to determine the optimal depth that gives the best prediction performance as shown in Table II.

For max depth = 9, this model shows an MAE of 13.43, an MSE of 297.56, RMSE of 17.25, and an R^2 value of 0.71. As max depth increased, the performance of the model improved with max depth = 11 showing the best results: MAE 13.06, MSE 291.40, RMSE 17.07, and R^2 value 0.71. This means that about 71.24% of the variation in the IPL score is explained by the RF model with max depth 11. In particular, the R^2 value for the best-performing model closely matches the LR model and shows that it is comparable prediction efficiency. Using the optimized RF model with max depth = 11, the final predicted score for user input is 193.74. This comprehensive measure provides a detailed assessment of the accuracy and robustness of the RF model in various deep configurations.

From Fig. 3, it is observed that the plot shows a broader range of predicted values, between 70 and 220, indicating that the model makes diverse predictions in various score scenarios. Unlike the LR plot, the predicted line for RF appears to be less consistent, showing higher variability and inaccuracies in predicting the actual score. The wider the spread of predicted

TABLE II
PREDICTED EVALUATION METRIC VALUES USING RF MODEL

Evaluation Metrics	Max Depth	Values
MAE	9	13.431570993070723
MSE		297.56426740564103
RMSE		17.250051229072945
R^2		0.706313761411314
MAE	10	13.179801689575434
MSE		293.9452995976107
RMSE		17.14483302915519
R^2		0.7098855647477176
MAE	11	13.061816283946026
MSE		291.5606972565823
RMSE		17.075148528097266
R^2		0.7122390895784115
MAE	12	13.052005121403399
MSE		295.2223824819888
RMSE		17.182036622065173
R^2		0.7086251255426039
MAE	13	13.192187916308637
MSE		305.2494186875832
RMSE		17.47138857353883
R^2		0.6987287674446095

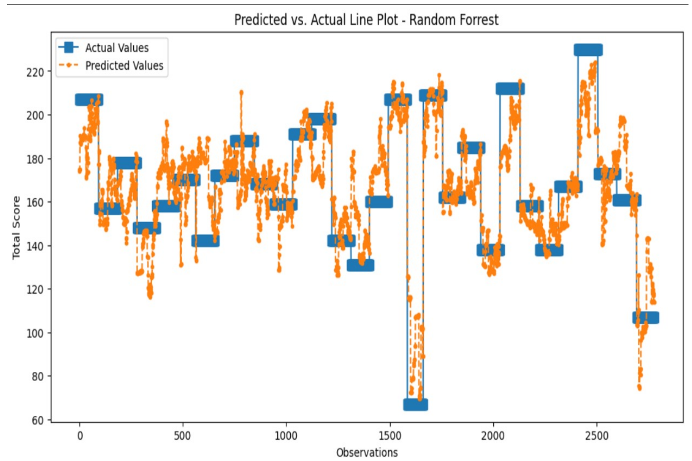

Fig. 3. Graph showing analysis of line plot using RF model

scores for a given actual score, the more inconsistent the model's predictions. This variation indicates that the RF model may struggle to consistently capture the underlying patterns in the data, resulting in a less accurate representation of the actual score.

C. Proposed IPL score prediction framework using SVM

Various kernel functions, namely linear, RBF, sigmoid, and 3rd-degree polynomial are used to evaluate the performance of the SVM model.

Table III illustrates the corresponding RMSE values for each kernel. The linear kernel exhibited competitive performance with an RMSE of 16.44. In contrast, the 3rd-degree polynomial kernel yields an RMSE of 18.16, while the RBF kernel shows an RMSE of 16.96. However, the sigmoid kernel showed a high RMSE of 64.54, indicating a less favorable fit for the given dataset. Further analysis identified the linear kernel as the optimal choice, predicting a score of 187.62 for user input, as detailed in Table IV. Evaluation metrics for the loaded SVM

TABLE III
PREDICTED EVALUATION METRIC VALUES USING SVM MODEL

Evaluation Metrics	Values
SVM Metrics for kernel-linear	
Root Mean Squared Error	16.4385896162757
SVM Metrics for kernel-poly	
Root Mean Squared Error	18.16494478983923
SVM Metrics for kernel-RBF	
Root Mean Squared Error	16.962782579599097
SVM Metrics for kernel-sigmoid	
Root Mean Squared Error	64.5390812299391

model using a linear kernel, MAE, MSE, RMSE, and R^2, highlight the effectiveness of the linear kernel in achieving accurate IPL score predictions.

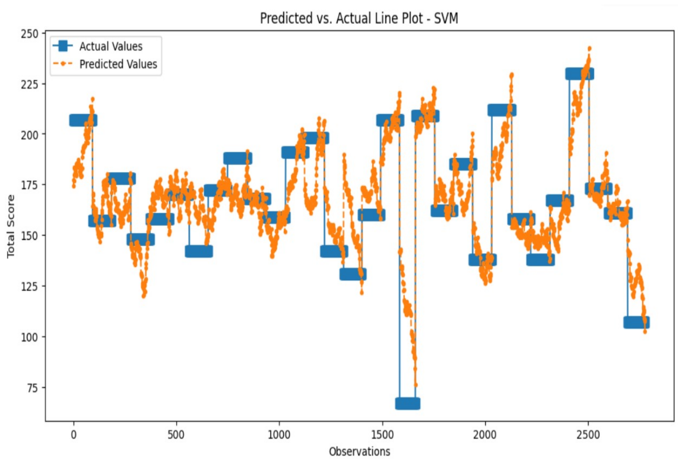

Fig. 4. Graph showing analysis of line plot using SVM model

From Fig. 4, it is observed that this plot displays the characteristics for the LR model with predicted values from 60 to 240. The predicted line consistently matches the actual score, showing the reliability of the model in capturing different scenarios. Similar to LR, even outliers are consistently predicted, confirming the robust performance of the model. In particular, there is a cluster of data points and prediction lines between 140 and 175, indicating a higher frequency of accurate predictions in this scoring interval. Sharing similarities with LR, the unique properties of SVM contribute to its effectiveness of consistently predicting scores across a range of common IPL scores, especially in unique situations.

D. Comparative analysis of the proposed framework across all the models

In the examination of IPL matches in 2023, featuring Mumbai Indians vs Royal Challenger Bangalore and Rajasthan Royals vs Sunrisers Hyderabad, have been reviewed in detail as viewed in Fig. 5. Mumbai Indians were impressive from 99 to 174 for 2 wickets between 10 and 15 overs, scoring 75 runs in the last 5 overs to reach a total of 200 for 4 wickets. On the other hand, Sunrisers Hyderabad scored 217 for 6, with a substantial 64 runs in the last 5 overs. The average

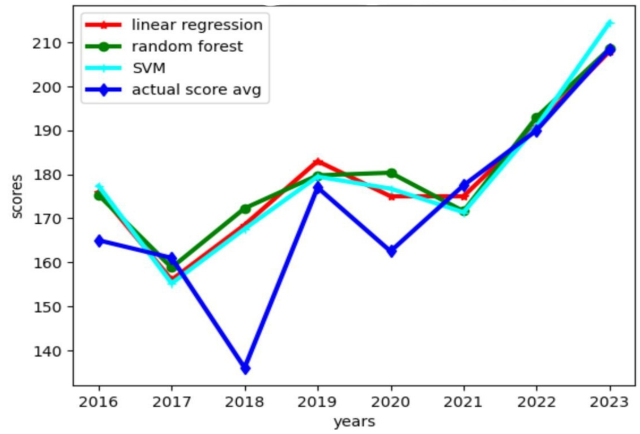

Fig. 5. Comparison of various models with respect to year-wise real score average

score in the two games was 208.5. Predictions were made for each game using LR, RF, SVM models, resulting in scores of 208, 208.58, and 214.5, respectively. This prediction is then compared with 217 points, providing a comprehensive assessment of the model's performance from 2016 to 2023. In the same way, random matches were chosen from each edition of the IPL from 2016 to 2023, and a similar analysis was done to get the points, which were plotted on the graph.

E. Analysis of proposed framework using evaluation metrics with respect to all the models

TABLE IV
COMPARISON OF PERFORMANCE METRICS FOR DIFFERENT ML TECHNIQUES

Metrics	LR	RF	SVM
MAE	12.1186	13.0618	12.1651
MSE	251.0079	291.5607	270.2272
RMSE	15.8432	17.0751	16.4386
R^2	0.7523	0.7122	0.7333

Evaluating three ML models - LR, RF, and SVM - the main performance metrics are MAE, MSE, RMSE, and R^2. LR emerges as a top performer. LR shows the lowest MSE at 251.01, surpassing both SVM (MSE: 270.23) and RF (MSE: 291.56). LR has a higher MAE (12.12) and RMSE (15.84) compared to RF and SVM, which shows slightly higher error. When evaluating the goodness of fit, LR has the highest R^2 value of 0.75, better than RF (0.71) and SVM (0.73). The results as displayed in Table V show LR as a robust model, providing accurate and robust predictions for IPL scores compared to SVM and RF.

Phase 2: Deployment and user interface

In creating an online cricket match final score prediction app, the user interface will allow users to enter various details such as the batting team, bowling team, current runs, wickets, score in the last 5 overs, and wickets fallen in the last 5 overs. This user interface will be designed using HTML, CSS, and possibly JavaScript to provide users with a visually appealing and interactive experience.

The back-end of the application will be powered by Flask, a web framework for Python. In Flask, the stack file model that was previously trained and saved as a .pkl file will be loaded using the pickle module. This model will be responsible for processing the user inputs and predicting the final score based on the details provided.

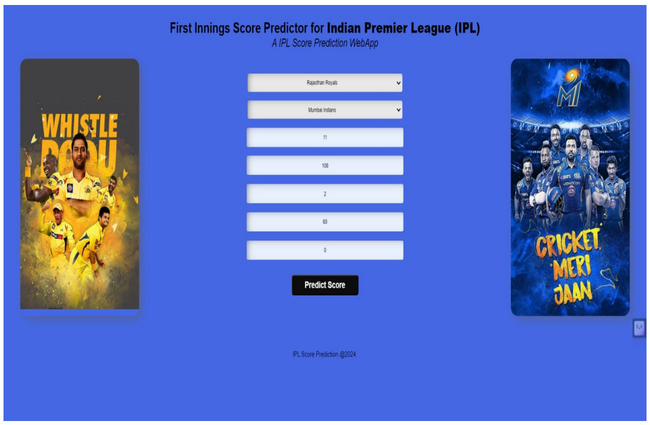

Fig. 6. Website User Interface of Proposed Model

Once the user submits a form with the required details, Flask processes the form submission, processes the input data, and passes it to the loaded stack file model for prediction. The predicted final score is then displayed back to the user, either on a new page or on the same page, giving them the valuable insight they are looking for.

This approach not only provides an intuitive and user-friendly experience but also uses the power of machine learning to provide accurate predictions, increasing overall user engagement and satisfaction with the app.

An innovative approach has been taken to improve the clarity of predicted IPL scores in the website layout. The final score is displayed as a 10-digit range rather than as a single point estimate. This range is achieved by offering a wider range of potential outcomes, including +5 and -5 to the predicted score. By presenting that section, the placement strategy recognizes the uncertainty involved in predicting cricket scores and gives users a more complete understanding of possible score changes. This thoughtful presentation not only provides the model's forecast but also provides transparency and informed decision-making for users interested in the website's score forecast, as well as the level of uncertainty.

In this work, specific input values were provided to the trained ML model, including details such as the batting team being Rajasthan Royals, bowling team as Mumbai Indians, 11 overs played, 108 runs scored, 2 wickets taken, 60 runs in the last 5 overs, and no wickets in the last 5 overs. The model generated a predicted score ranging between 175 and 190. Upon comparison with the actual score of 193, it reveals a reasonably accurate prediction, with a slight underestimation of the final score by the model. This instance showcases

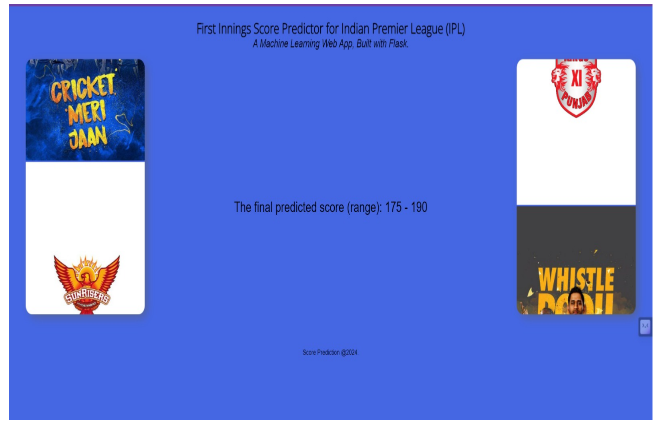

Fig. 7. Output of the Predicted Proposed Model

the model's ability to provide valuable insights into match outcomes based on given match parameters.

V. Conclusion

In conclusion, this work serves as a demonstration of ML application for predictive insights in sports data. While LR has made a significant contribution to sports analytics, the consideration of more complex models such as RF and SVM suggests the preference of LR in this context. This discovery not only enriches the field of sports analytics but also shows the integral role of machine learning in shaping the dynamic landscape of sports analytics, creating a solid foundation for future research.

This work demonstrates the significant potential of ML, especially LR, in sports analytics, with a special focus on IPL score prediction. While the accuracy of the predictions shows promise, there is continual room for improvement. The incorporation of additional features such as player shape, team composition, and game conditions is an important way to increase the efficiency of the model. In addition, the exploration of more sophisticated ML models, including decision trees, RF, and neural networks, promises to capture complex relationships in the data, thus increasing the accuracy of predictions. Overall, this research contributes to the development of sports analytics, laying the foundation for future progress and methodology.

VI. Future Scope

The future scope of this work involves advancements in cricket predictive analytics. Continuous improvement, realistic predictions during live IPL matches, and improving model performance by integrating additional data sources offer opportunities for development. A user-friendly interface improves accessibility and extends the model to provide insight into player choices, adding valuable strategic information. By exploring this area, this work aims to make a significant contribution to cricket predictive analytics, team management, sports analysts, and cricket enthusiasts.

References

[1] Indian Premier League (IPL), https://www.iplt20.com/

[2] Jayashree, R.J., Ganesh, S., Karanth, S.C., Lalitha, S. (2021). Automatic Detection of Parkinson Speech Under Noisy Environment. In: Thampi, S.M., Gelenbe, E., Atiquzzaman, M., Chaudhary, V., Li, KC. (eds) Advances in Computing and Network Communications. Lecture Notes in Electrical Engineering, vol 736. Springer, Singapore.

[3] K. Joshi, G. A. Reddy, S. Kumar, H. Anandaram, A. Gupta and H. Gupta, "Analysis of Heart Disease Prediction using Various Machine Learning Techniques: A Review Study," 2023 International Conference on Device Intelligence, Computing and Communication Technologies, (DICCT), Dehradun, India, 2023, pp. 105-109, doi: 10.1109/DICCT56244.2023.10110139.

[4] E. Mundhe, I. Jain and S. Shah, "Live Cricket Score Prediction Web Application using Machine Learning," 2021 International Conference on Smart Generation Computing, Communication and Networking (SMART GENCON), Pune, India, 2021, pp. 1-6, doi: 10.1109/SMART-GENCON51891.2021.9645855.

[5] S. Lalitha, D. Gupta, M. Zakariah, and Y.A. Alotaibi "Mental Illness Disorder Diagnosis Using Emotion Variation Detection from Continuous English Speech," Comput. Mater. Contin., vol. 69, no. 3, pp. 3217-3238. 2021. https://doi.org/10.32604/cmc.2021.018406.

[6] Manoj Ishi, Jayantrao Patil, Vaishali Patil, An efficient team prediction for one day international matches using a hybrid approach of CS-PSO and machine learning algorithms, Array, Volume 14, 2022, 100144, ISSN 2590-0056.

[7] Y. Kumar, H. Sharma and R. Pal, "Popularity Measuring and Prediction Mining of IPL Team Using Machine Learning," 2021 9th International Conference on Reliability, Infocom Technologies and Optimization (Trends and Future Directions) (ICRITO), Noida, India, 2021, pp. 1-5, doi: 10.1109/ICRITO51393.2021.9596405.

[8] Shubhra Singh, Parmeet Kaur,IPL Visualization and Prediction Using HBase, Procedia Computer Science, Volume 122, 2017, Pages 910-915, ISSN 1877-0509.

[9] A. I. Anik, S. Yeaser, A. G. M. I. Hossain and A. Chakrabarty, "Player's Performance Prediction in ODI Cricket Using Machine Learning Algorithms," 2018 4th International Conference on Electrical Engineering and Information Communication Technology (iCEEiCT), Dhaka, Bangladesh, 2018, pp. 500-505, doi: 10.1109/CEEICT.2018.8628118.

[10] E. L. Lekamge, K. R. Wickramasinghe, S. E. Gamage, T. M. K. L. Thennakoon, P. S. Haddela and S. Senaratne, "CricSquad: A System to Recommend Ideal Players to a Particular Match and Predict the Outcome of the Match," 2023 3rd International Conference on Advanced Research in Computing (ICARC), Belihuloya, Sri Lanka, 2023, pp. 42-47, doi: 10.1109/ICARC57651.2023.10145677.

[11] S. G, A. Swaminathan, J. B. J, S. R and L. Nelson, "IPL Data Analysis and Visualization for Team Selection and Profit Strategy," 2023 7th International Conference on Computing Methodologies and Communication (ICCMC), Erode, India, 2023, pp. 592-598, doi: 10.1109/ICCMC56507.2023.10083736.

[12] Lalitha, S. and Gupta, Deepa. 'Investigation of Automatic Mixed-lingual Affective State Recognition System for Diverse Indian Languages'. 1 Jan. 2021 : 5467 – 5476.

[13] ESPN, https://www.espncricinfo.com/.

[14] Cricbuzz, https://www.cricbuzz.com/.

[15] Scikit-Learn One Hot Encoding Documentation, https://scikit-learn.org/stable/modules/generated/sklearn.preprocessing. OneHotEncoder.html.

[16] Bishop, Christopher M. "Pattern recognition and machine learning." Springer google schola 2 (2006): 1122-1128.