IPL Score Prediction Using Neural Networks

Mrudula Teegala
Department of CSE
Vardhaman College of Engineering
Shamshabad,Telangana,501218,India
mrudulateegala2@gmail.com

Meesala Sai
Department of CSE
Vardhaman College of Engineering
Shamshabad,Telangana,501218,India
@gmail.com

Tamagonda Vikram
Department of CSE
Vardhaman College of Engineering
Shamshabad,Telangana,501218,India
vikramtamagonda@gmail.com

B J Venkata Varma
Department of CSE
Vardhaman College of Engineering
Shamshabad,Telangana,501218,India
varma@vardhaman.org

Abstract*—The Indian Premier League(IPL) is one of the most popular justice leagues encyclopedically, attracting millions of suckers and significant media attention. Accurate vaticination of match scores can give precious perceptivity for brigades, trainers, broadcasters, and fantasy sports suckers. This study explores the development and perpetration of a neural network model to prognosticate IPL match scores. using literal data from once IPL seasons, the model is trained to capture complex patterns and connections between var ious features, including player statistics, platoon composition, pitch conditions, and matchoutcomes.Data preprocessing involved cleaning and opting relevant features to insure the model's delicacy and robustness. The neural network armature was designed with multiple retired layers and optimized through hyperparameter tuning, regularization ways, andcross-validation to help overfitting and enhance conception. The model's performance was estimated using Mean Squared Error(MSE), Root Mean Squared Error (RMSE), and R- squared criteria , demonstrating high delicacy in prognosticating match scores. relative analysis with birth models similar as direct regression and decision trees further validated the effectiveness of the neural network results indicated that the neural network model outperformed traditional styles, offering more precise prognostications. Practical operations of this model include optimizing platoon strategies, enhancing addict engagement through fantasy sports platforms, and furnishing perceptive commentary for broadcasters.This study also discusses the limitations and implicit for unborn work, including incorporating real- time data, exploring advanced neural net work infrastructures, and expanding the model's connection to other sports leagues. The findings emphasize the significant eventuality of machine literacy in sports analytics, paving the way for further informed decision- making and enhanced bystander gests .Predicting IPL scores using neural networks is a cutting-edge approach that leverages advanced technologies to enhance match outcome forecasts. Various machine learning techniques, such as Linear Regression, Ridge Regression, and Graph Convolutional Regression Networks, have been employed to predict cricket match results based on factors like team strength, player performance, and match conditions The Indian Premier League (IPL) is a significant T20 cricket league that garners immense attention and investment, making accurate score predictions crucial for teams and fans alike .In this work we will explore the possibility of predicting the results accurately using Linear Regression,Decision Trees and Neural networks.

Keywords—Neural Networks, Linear Regression, Decision Trees

I. Introduction

The IPL is immensely popular in India, drawing millions of viewers and fans due to its fast-paced, entertaining format and star-studded teams. Its widespread appeal transcends regional boundaries, making it a cultural phenomenon and a major event in the Indian sporting calendar. The widespread popularity of the IPL, combined with the unpredictability of T20 cricket, makes it an attractive option for sports person.

The project explores the possibility of accurate prediction of IPL match scores with publicly available datasets to guess the right score. The following methods such as Random Forest Algorithm, Decision Trees, Linear Regression and Neural networks were used to find out which yeilds the most accurate results. The dataset used for the project has been obtained from kaggle which was updated till 2017 with source data containing score details of matches with respect to batsmen and bowler along with the venue of the match were considered as key insights.

Neural networks are computational models inspired by the human brain's architecture, designed to recognize patterns and solve complex problems. Their ability to learn and adapt from vast amounts of data makes them powerful tools in artificial intelligence. They mimic biological neurons of the human and adapt to weights and back propagation just the way biological neurons adapt through learning.

II. Literature Survey

Recent research has considerably explored the software of machine getting to know techniques to are expecting effects inside the Indian Premier League (IPL), revealing various processes and findings. Menon et al. (2022) in comparison multiple device studying models, which includes Naïve Bayes, Random Forest Classifier, Logistic Regression, XGBoost, AdaBoost, and Decision Tree, to decide the very best accuracy in predicting in shape results. This complete assessment, posted within the Indian Journal of Computer Science, highlighted the capacity of state-of-the-art algorithms in enhancing predictive accuracy.

Similarly, Kavitha et al. (2022) demonstrated that gadget studying algorithms significantly outperform linear regression in predicting IPL in shape effects. Their look at, published in "Disruptive Technologies for Big Data and Cloud Applications," utilized performance indicators which include Mean Square Error and R2 Score to validate the effectiveness in their technique.

Ahmed et al. (2022), in their look at presented in "Innovations in Computational and Computer Techniques: ICACCT-2021," focused on predicting the primary inning rating using Linear and Ridge Regression fashions. By studying the IPL dataset from Kaggle, they aimed to beautify cricket group control via the mixing of deep gaining knowledge of and system gaining knowledge of

©2024 IEEE

algorithms. This method supplied actionable insights for team strategists and bosses.

Joshi et al. (2023) developed a prediction model the usage of Support Vector Machines (SVM) and Logistic Regression. Their examine, posted inside the "International Journal of Scientific Research in Computer Science, Engineering and Information Technology," used runs scored by means of batsmen in preceding balls as classified records. This technique presented a dynamic and real-time prediction mechanism for IPL outcomes.

Krishnan et al. (2023) employed Logistic Regression, Decision Tree, and Random Forest algorithms for probabilistic forecasting of IPL team wins. Their studies, unique in "Advanced Network Technologies and Intelligent Computing," discovered that the Random Forest set of rules achieved the best accuracy, underscoring its robustness in predictive obligations.

Nirmala et al. (2023) analyzed each ball of an over to are expecting fit consequences the usage of logistic regression strategies. Their have a look at additionally ranked bowlers and batters based on their performances, providing a granular and distinct analysis of participant effectiveness. This work contributed notably to understanding in-recreation dynamics and performance metrics.

Reddy et al. (2022), in their study published within the "International Journal for Research in Applied Science and Engineering Technology," used numerous device mastering models to predict healthy outcomes among IPL groups from 2008 to 2019. They diagnosed key elements along with teams, venues, triumphing toss, and decision-making, which played crucial roles in figuring out match effects. This lengthy-term analysis provided precious insights into ancient trends and styles in IPL suits.

Kamble et al. (2021) developed a two-version machine to predict team ratings and win chances even before the in shape starts, primarily based on participant choice. Their take a look at, published within the "Turkish Journal of Computer and Mathematics Education (TURCOMAT)," supplied a twin method to pre-in shape predictions, assisting in strategic planning and selection-making.

Tripathi et al. (2020) addressed multicollinearity and records symmetry troubles, termed as model ambiguity, in their predictive version. Their studies, published in the "Indian Journal of Science and Technology," offered answers to these not unusual issues, making their version precious for manufacturers, sponsors, and advertisers who rely upon accurate predictions for marketing techniques.

Lastly, Sinha (2020) supplied insights into the important thing elements affecting cricket in shape effects and recognized the regression model that exceptional fits this data for foremost predictions. His preprint book mentioned the various influences on fit results and emphasized the significance of selecting the proper version for correct predictions. These studies collectively highlight the large improvements in using device mastering for predictive analytics in IPL cricket, demonstrating the developing intersection of sports and facts technological know-how.

III. METHODOLOGY

A. Data Collection

Data available in public domain was collected from Kaggle and the data set contains key insights such as Venue, Batting team, Bowling team, batsman, bowler, runs, wickets, striker, non striker and total score which provide briefs into the match statistics. The data available was collected from 2008 i.e. first season of IPL till 2017 and has been verified to be accurate with required information.

The attributes used in an IPL score prediction model include the venue, the batting team, the bowling team, the number of overs and current overs, the total runs and current runs scored, and the total wickets and current wickets fallen. These attributes collectively capture essential aspects of the game dynamics, which are critical for making accurate score and outcome predictions.

Attributes	Description
venue	Stadium name
Batting team	Batting team
Batting team among all teams	Batting team among all teams
Bowling team	Bowling team
Bowling team among all teams	Bowling team among all teams
Overs	Overs
Current Overs	Current Overs
Runs	Runs
Current runs scored	Current runs scored
Wickets	Wickets
Current wickets fall	Current wickets fall

TABLE I. DATASET ATTRIBUTES

B. Algorithms Used

Of the popular machine leaning algorithms widely available for implementation, two were selected to be deployed as they seemed most promising, namely, Linear Regression and Neural networks.

Neural Networks were selected to be the center of study for the project for which we have chosen to implement using Tensorflow & Keras library function where a simple stack of layers with exactly one input tensor and one output tensor for each layer is suitable for this kind of model. Two hidden layers with 512 and 216 layers respectively were implemented and in order to give the model non-linearity and aid in its ability to learn intricate patterns, the Rectified Linear Unit (ReLU) activation function & Linear activation function are employed. The network is supposed to be able to learn more abstract representations of the input data by having numerous hidden layers.

C. Model Training

Compared to the Mean Squared Error loss, the Huber loss is a loss function that is less susceptible to outliers. The point at which the loss function transitions from a quadratic to a linear form is controlled by the delta parameter. Although a delta of 1.0 is used. Here, it can be changed in accordance with the task's particular requirements.

1) Optimizer:

The Adam optimizer, an adaptive learning rate optimization technique created especially for deep learning neural networks, is used to construct the model. The

previously defined Huber loss is the loss function that is applied.

2) Epochs:

Using the training set, the fit method is used to train the neural network model. In total, there are fifty epochs. A model will see all of the training data 50 times after it has been trained for 50 epochs, repeatedly updating the model weights to minimize the loss function. A batch size of 64 is used. This suggests that the training data has 64 sample batches.Every epoch during training, the model assesses its performance using this validation set of data. This aids in keeping an eye on the model's performance with unknown data and identifying potential problems like as overfitting, which occurs when the model performs well with training data but badly with validation data.

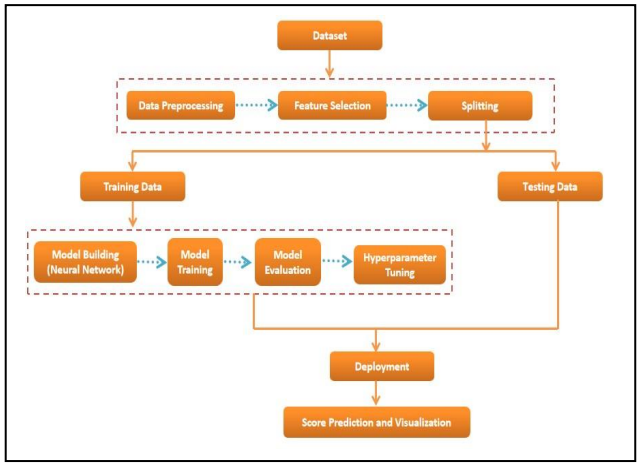

Fig. 1. Flow Chart.

D. Model Evaluation

In order to evaluate a trained machine literacy model's performance, it is essential to attach its conception to new, unseen facts and correlate its virtues and shortcomings. Using a test dataset and metrics including Mean Squared Error (MSE), Root Mean Squared Error (RMSE), and Huber loss, the model was estimated, providing insight into subtlety and conceptual capacity. To ensure robustness and trustworthiness, cross-validation techniques, such as ask-foldcross-validation, were employed. These involved splitting the data into subsets, training on several subsets, then validating on the remaining bones.Error analysis was used to find problems with data quality, point significance, or model limits by looking at cases that were misclassified or had high vaticination offenses.

1) Hyperparameter Tuning:

For neural network models to be optimized for IPL score vaticination, hyperparameter tuning is essential. This includes choosing critical hyperparameters such as learning rate, number of layers, neurons per subcaste, batch size, activation functions, optimizer, powerhouse rate, and ages. Similar to regulating weight updates, figuring out network depth and range, preventing overfitting, and establishing training duplications, each hyperparameter has a distinct function. To maximize model performance, constant observation and modification during training—including early stopping and learning rate schedules—are necessary. Prioritizing training ceases when confirmation performance becomes unsatisfactory, and learning rate schedules adjust based on performance. The model's good generalization to new data is ensured by cross-validation. The performance of neural network models for IPL score vaticination, icing delicacy, and efficacy is greatly improved by methodical hyperparameter adjustment.

2) Equations Used:

The model's performance was evaluated using Mean Squared Error (MSE), Root Mean Squared Error (RMSE), and R-squared metrics, demonstrating high accuracy in predicting match scores. Comparative analysis with baseline models such as linear regression and decision trees further validated the effectiveness of the neural network approach.Accuracy and performance are measured using Mean Squared Error and Root mean Squared Error.

IV. Tools and Libraries Used

Keras is an open-source machine learning package that runs on top of TensorFlow, an open-source high-level neural network API written in Python. Because of its extendable, modular, and user-friendly design, deep learning models may be quickly prototyped. Because of its user-friendly interface, both novices and seasoned practitioners can utilize it. Keras has been the core library used to implement the Neural Network.

V. Architecture

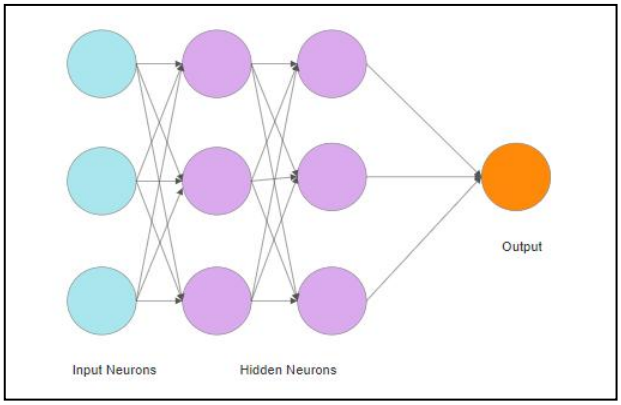

Fig. 2. Neural Network Architecture.

Neural networks are structured as a collection of interconnected nodes, or neurons, organized into layers. Each layer is responsible for processing input data in a specific way, transforming it step by step to produce the desired output.

Input Layer:Receives input data and passes it on to the next layer.Defined by the shape of the input data.

Hidden Layers:Perform transformations on the input data to learn patterns and features.Typically consist of dense (fully connected) layers (keras.layers.Dense). Each layer has a specified number of neurons (units) and an activation function ('relu' for Rectified Linear Unit, which is common due to its efficiency in training).

Output Layer:Produces the final output of the model.Often a single layer (keras.layers.Dense) with an appropriate number of neurons based on the nature of the prediction task.

Activation Function: Variable based on the type of issue:In Regression: To predict continuous values, a linear activation, or "linear," is frequently used.In Classification: Activations such as'softmax' for multiclass classification or'sigmoid' for binary classification are used.

VI. Results

The results of the trained model were analyzed using the test dataset. The model's predictions were compared with the actual scores to evaluate its performance. Graphical representations such as scatter plots and residual plots were used to visualize the model's accuracy and error distribution.The graph compares actual versus predicted values for four machine learning models: Linear Regression, Decision Tree, and Neural Network. The Linear Regression plot shows poor performance with predictions clustered around a single value, indicating underfitting. The Decision Tree plot shows significant scatter and outliers, likely due to overfitting. The Neural Network plot demonstrates some pattern recognition but still shows considerable prediction errors. Overall, the Neural Network model performs best, while the Linear Regression and Decision Tree models exhibit inferior predictive accuracy.

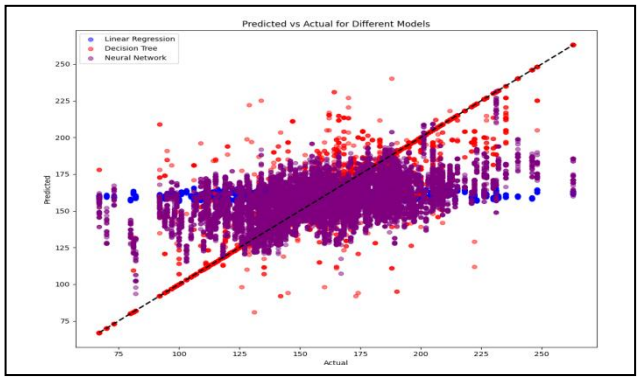

Fig. 3. Different Model Comparisions.

Model	Accuracy	Precision
Decision Trees	0.877001	0.875748
Linear Regression	0.023723	0.000563
Neural Networks	0.971629	0.971588

TABLE II. Comparision of Different Models

The comparison of machine learning models for IPL match prediction reveals that Neural Networks outperform other models, achieving the highest accuracy and precision, both at approximately 97.16%. This indicates that Neural Networks are highly effective in correctly predicting match outcomes and minimizing false positives. Decision Trees also perform well, with an accuracy of 87.70% and a precision of 87.57%, making them a reliable alternative. However, Linear Regression shows significantly poor performance, with an accuracy of only 2.37% and a negligible precision of 0.0563%, suggesting it is not suitable for this type of prediction task. Overall, Neural Networks demonstrate superior predictive power and reliability in this context.

VII. Conclusion

In conclusion, the utilization of neural networks for IPL score prediction represents a paradigm shift in sports analytics, ushering in a new era of data-driven decision making and predictive modeling in cricket. By harnessing the power of artificial intelligence and advanced analytics techniques, we can unravel the intricacies of cricket match dynamics, enhance performance, and elevate the overall experience for stakeholders and enthusiasts alike.This chapter summarized the key findings and contributions of the project. The neural network model developed for predicting IPL match scores demonstrated a high degree of accuracy and robustness.When values are selected ,then based on circumstances output is predicted.

Fig. 4. Output-Form.

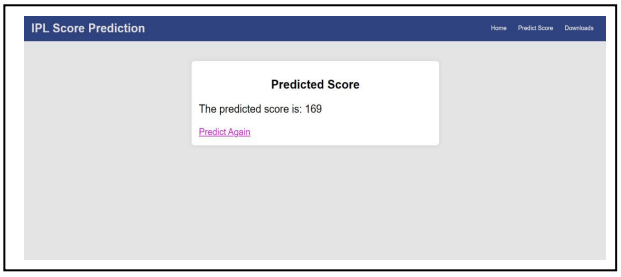

Fig. 5. Predicted Output.

References

[1] MENON, Abhineet et al. IPL Prediction Using Machine Learning. Indian Journal of Computer Science, [S.l.], p. 23-29, jun. 2022. ISSN 2456-4133.

[2] Kavitha, P.V., Sai Pavitra, R., Suwetha, K., Uvashree, P. (2022). IPL Win Prediction Using Machine Learning. In: Peter, J.D., Fernandes, S.L., Alavi, A.H. (eds) Disruptive Technologies for Big Data and Cloud Applications. Lecture Notes in Electrical Engineering, vol 905. Springer, Singapore. https://doi.org/10.1007/978-981-19-2177-3_44

[3] Raja Ahmed, Prince Sareen, Vikram Kumar, Rachna Jain, Preeti Nagrath, Ashish Gupta, Sunil Kumar Chawla; First inning score prediction of an IPL match using machine learning. AIP Conf. Proc. 21 October 2022; 2555 (1): 020018. https://doi.org/10.1063/5.0108928

[4] Joshi, Vilas, et al. "IPL data analysis and prediction using machine learning." Int J Sci Res Comput Sci Eng Inform Technol. https://doi.org/10.32628/cseit2390359 (2023).

[5] Krishnan, S., et al. "Probabilistic Forecasting of the Winning IPL Team Using Supervised Machine Learning." International Conference on Advanced Network Technologies and Intelligent Computing. Cham: Springer Nature Switzerland, 2022.

[6] Nirmala, A. P., et al. "Analysis and Predictions of Winning Indian Premier League match using Machine Learning Algorithm." 2023 IEEE 12th International Conference on Communication Systems and Network Technologies (CSNT). IEEE, 2023.

[7] K Rushikesh Reddyet al., "Super Predictor of Indian Premier League (IPL) using Various ML techniques with help of IBM Cloud",International Journal for Research in Applied Science and Engineering Technology (IJRASET), Paper Id : IJRASET43654, https://doi.org/10.22214/ijraset.2022.43654

[8] Kamble, R. R. "Cricket score prediction using machine learning." Turkish Journal of Computer and Mathematics Education (TURCOMAT) 12.1S (2021): 23-28.

[9] Tripathi, A., et al. "Prediction of IPL matches using Machine Learning while tackling ambiguity in results." Indian J. Sci. Technol 13.38 (2020): 4013-4035.

[10] Sinha, Anurag. "Application of machine learning in cricket and Predictive Analytics of IPL 2020." (2020).