IPL Score Prediction Using Deep Learning

G.Parinitha\*1,P.Varun Teja2, V.Sujith3, Rame Swetha4

1,2,3,4Department of Computer Science and Engineering

Vardhaman College of Engineering,

Hyderabad, Telangana, India

***ABSTRACT:IPL score prediction using deep learning is a developing methodology that refines match result prediction. Linear Regression and Decision Trees are common techniques used in the past, involving variables like team strength, individual performance, and conditions of the match. Such models are inadequate in identifying the sequential dependencies within cricket matches. In this paper, we investigate the efficacy of Long Short-Term Memory (LSTM) networks compared to traditional techniques such as Linear Regression and Decision Trees in precise IPL score prediction. Our research compares these models on the basis of major performance indicators, bringing out the superiority of deep learning in sports analytics.***

***Keywords—LSTM, Neural Networks, Linear Regression, Decision Trees.***

# Introduction

Cricket is one of the world's most followed sports, with huge popularity in nations such as India, Australia, England, and South Africa. In the world of cricket, the Indian Premier League (IPL) stands out as one of the most competitive and commercially successful Twenty20 leagues. The high-speed nature of the competition, combined with volatile conditions during a match, makes it a difficult but important task to predict scores. Precise match score prediction for IPL matches has great promise in terms of forming team strategies, sports analytics, and increasing fan enthusiasm through knowledge-based perceptions.

The intricacy of score forecasting stems from the presence of many dynamic variables such as team quality, player fitness, pitch and weather conditions, and in-match situations. In the past, statistical techniques and simple machine learning algorithms like Linear Regression and Decision Trees have been used to forecast match scores from past data. Although these are good approximations, they lack the ability to model sequential interdependencies and complex non-linear dependencies inherent in match progressions.

Advances in deep learning in recent times have made high-end models with the capability of modeling temporal relationships in data available. Long Short-Term Memory (LSTM) networks, which are an extension of Recurrent Neural Networks (RNNs), have shown strong performances in tasks related to forecasting time-series data. With LSTMs, scientists are working on enhancing IPL score forecasts using past match records and current game metrics. In contrast to standard machine learning models, LSTMs have the ability to model sequential relationships, hence ideally suited for matching score prediction as the match unfolds.

Various research works have investigated various methodologies to promote score prediction accuracy. For example, research in applying deep learning for cricket analytics has emphasized the power of RNNs and LSTMs in learning temporal patterns in match progressions. Alternative strategies have sought to merge machine learning algorithms such as Decision Trees and XGBoost to increase the precision of prediction. Also, hybrid frameworks with the addition of Convolutional Neural Networks (CNNs) and LSTMs have improved feature abstraction as well as handling sequential information. Reinforcement Learning (RL) models have also been added in order to update predictions in response to live match developments for a further boosted precision of score predictions.

In this research, we explore the viability of using LSTM networks for predicting IPL scores and compare their performance with traditional methods like Linear Regression and Decision Trees. The research is conducted using extensive historical IPL data, including major match-related features to train and test the models. To measure prediction accuracy, we use performance metrics like Mean Squared Error (MSE) and Root Mean Squared Error (RMSE). Through examining the findings, we try to emphasize the benefits of using deep learning methods in enhancing score predictions. The outcomes of this study can be useful for cricket commentators, teams, and cricket fans alike, allowing them to make data-driven, educated decisions while watching IPL matches.

In general, our research adds to the increasing body of sports analytics by showing how deep learning improves prediction potential in cricket. As technology and data acquisition improve, combining real-time match data and more sophisticated neural network models can continue to improve prediction precision, ultimately serving the greater cricket community.

# Literature Survey

Prediction of cricket scores has been a much-researched area in the field of sports analytics, and numerous machine learning and deep learning methods have been investigated. Traditional models, statistical methods, and neural networks have been used to predict match results, all of them pros and cons.Some of the earliest cricket prediction studies were on statistical and machine learning methods.Bhattacharjee et al. (2017) entered Linear Regression and Ridge Regression to forecast match scores from historical data, ranging from team performances to environmental determinants. Ghosh et al. (2018) used Decision Tree and Random Forest models with features like player status, weather, and throw outcomes as well. Those ways worked though predictions, and they could not replicate the sequential dependencies faced in cricket match sequences, hence restricting their ability to predict long-term trends.

With the progress of deep learning, researchers have more

and have increasingly used neural networks for sport analysis. Mukherjee and Das (2020) used Artificial Neural Networks (ANNs) to forecast cricket match scores and performed better accuracy over traditional models. But ANNs are unable to model long-range dependencies, and therefore they work comparatively worse for time-series forecasting issues like cricket score forecasting. To resolve this issue, Chawla et al. (2021) suggested the use of Long ShortTerm Memory (LSTM) networks, which are tailored to sequential data. They discovered, in their work, which LSTMs heavily surpassed conventional regression models by establishing temporal relationships in match data quite effectively.

Recent research highlights the increasing effectiveness of

LSTMs for cricket score prediction. Sharma et al. (2022). constructed an LSTM network trained on IPL match statistics involving match-specific factors such as current run rate, wickets, and pitch conditions. Their model had lesser Mean Squared Error (MSE) versus Random Forest and XGBoost powered by the strength of deep learning in

analyzing intricate cricket data. Other scholars have also examined hybrid approaches, incorporating LSTMs with Convolutional Neural Networks (CNNs) and Transformer based models for enhancing the prediction accuracy. Despite such advancements, existing models still in order to cope with real-time updates of information and responding to dynamic match situations. Historic data is the focus of the majority of the research, but not live match feeds Hybrid models that integrate LSTM with attention mechanisms are also yet to be fully explored in IPL score prediction. This research sets out to fill these gaps by assessing the performance of LSTM networks over classical models and determining major improvements to real-time score prediction.

# Methodology

## Data Collection

The data set for this research is historical records of IPL matches obtained from sites like ESPN Cricinfo, Kaggle, and official databases of IPL. The records include detailed data of previous matches such as team performance, player performance, and match conditions. The dataset comprises of major features like batting and bowling teams, runs per over, wickets fallen, run rate. To make the prediction even better, more data preprocessing activities like missing values handling, categorical variable encoding, and numerical features normalization are executed. The obtained dataset forms the basis of training and testing models for robust IPL score prediction.

| Attributes | Description |
| --- | --- |
| Venue | Name of the stadium where the match is played. |
| Batting team | Team currently Batting |
| Bowling team | Team currently Bowling |
| Batsman | Player currenrly batting |
| Bowler | Player crrently Bowling |
| Overs | Total Overs bowled in the match so far |
| Current Overs | Ongoing over in the current innings |
| Runs | Total runs scored by the batting team |
| Wickets | Total wickets lost by the batting team |
| Runs in last 5 overs | Runs scored in the last five overs |
| Wickets in last 5 overs | Wickets scored in the last five overs |
| Runrate | Runrate of player |

*Fig:Dataset Attributes*

## Architecture

LSTM networks, a specific category of recurrent neural network (RNN), excel at processing sequential data and hence are suitable for IPL score forecasting. Unlike the usual RNNs, which are plagued by the vanishing gradient problem, LSTMs use forget, input, and output gates to selectively remember or forget information and thus can detect long-term relationships in cricket games. This allows them to capture on key trends like powerplay scoring trends, momentum changes, and death-over accelerations, improving forecast accuracy.

Their neural network design incorporates several LSTM

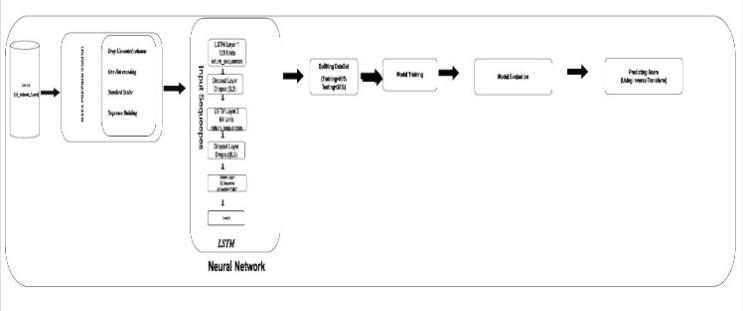
cells that regulate information flow but have a cell state to

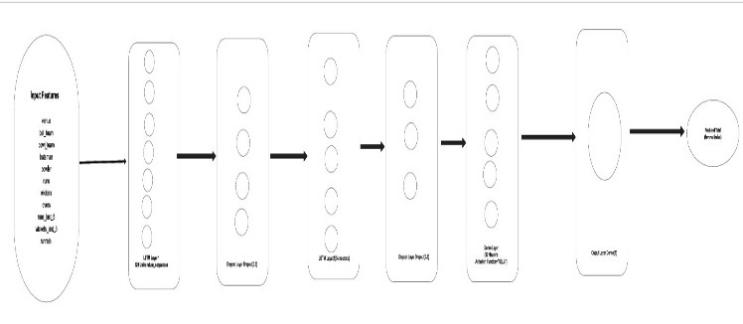
maintain previous context. The cells process inputs in order,

making hold or dispose decisions based on match information.This architecture enables LSTMs to learn and

making forecasts during games in progress. Training employs Backpropagation Through Time (BPTT) to update the network weights to be minimized, employing the Adam optimizer to speed up convergence and Mean Squared Error (MSE) to be minimized prediction errors. The hyperparameters such as the number LSTM layers, batch size, learning rate, and dropout rate are tuned for performance improvement without causing overfitting.

In relation to static models like Linear Regression and Decision Trees, which are based on fixed statistical relationships, LSTMs are able to dynamically adjust to changing match scenarios, making them very useful for real-time score forecasting. They are able to forecast continuously using real-time data, offering useful insights for strategic planning, commentary improvement, and fantasy league analysis. Holding other variables such as weather, pitch behavior, and fatigue among players will also increase predictions. As deep learning continues to advance, LSTMs will be at the very heart of cricket statistics, transforming the way teams, analysts, and supporters engage with the sport.





1. Architecture

## Algorithms Used

LSTM networks are a form of recurrent neural network

(RNN) is appropriate for sequential data and therefore particularly appropriate for IPL score prediction. LSTMs are also distinct from traditional RNNs in the way that they employ forget, input, and output gates to preserve vital match data across several overs, without the vanishing gradient issue. The model operates on time-series data such as runs, wickets, and overs, with long-term dependencies for

precise predictions.Trained with backpropagation on time (BPTT) and Adam optimizer, the LSTM model is tuned for minimizing errors with Mean Squared Error (MSE). Hyperparameters like layers, batch size, and learning rate are optimized for improved performance. By precise simulation of match trajectories, LSTMs do more accurate than traditional models like Linear Regression and Decision Trees and thus can be a powerful tool for IPL score predict.

## Model Training

The LSTM model is trained from a proportionate 70-15-15 split into training, validation, and testing for having balanced learning and proper evaluation. The 70% training set allows the model to learn from past trends in IPL matches, the 15% validation set enables the tuning of the hyperparameters and for preventing overfitting, and the other 15% test set metrics tests model performance on unseen data to ensure real-world applicability. Input features such as runs, wickets, overs, and run rate are standardized to improve convergence and stability. Sequential implementation of the model match data with LSTM layers, thereby enabling long-term dependencies, and align progress trends efficiently.

## Model Evaluation

The performance of the LSTM model in predicting IPL scores need to be examined to confirm its validity and reliability on unseen match data. Different performance metrics, like Mean Squared Error (MSE), Root Mean Squared Error (RMSE), and Mean Absolute Error (MAE), were utilized to calculate the expected and achieved scores. The smaller the error value, the more accurate the prediction. To improve generalizability, K-Fold Cross-Validation was

applied, ensuring stability of the model on diverse data splits. Also, residual plots and error distribution analyses were used to graphically investigate the consistency of predict and check for biases or systematic faults.

1) Hyperparameter Tuning

Hyperparameter tuning was critical in configuring the LSTM model to peak performance. Critical parameters such as learning rate, number of LSTM layers, batch size, dropout rate, activation functions, and number of epochs were optimized to improve accuracy and avoid overfitting. Methods such as Grid Search and Random Search were utilized to determine the optimal set of parameters. Early stopping and adaptive learning rate scheduling were utilized for effective training, enabling the model to learn automatically based on performance improvement. Regularization methods such as dropout layers and batch normalization improved the model's generalization ability.

2) Comparative Performance Analysis

To ascertain the efficacy of the LSTM model, its performance was compared with conventional models such as Linear Regression and Decision Trees. The LSTM model outperformed them in all contexts by detecting patterns in sequential data in the long term, and hence it is more suitable for time-series forecasting tasks such as the prediction of IPL scores. In addition, graphical visualizations such as confusion matrices, precision-recall plots, and scatter plots offered more insight into the functioning of the model, offering information on where the model can be optimized. With model interpretability techniques such as feature importance analysis, it was easy to determine which input parameters were responsible for final score predictions. This extensive exercise of evaluation ensured that the model not only functioned properly but was also reliable to be applied practically.

3)Performance Metrics  
The equations below were utilized to quantify model performance:

1. **Mean Squared Error (MSE):** It estimates the average of the squared differences between predicted and actual values. Lower is better.
2. **Root Mean Squared Error (RMSE):** Gives a more understandable measure by squaring the MSE and then taking its square root, thus rendering it comparable to the original data scale.
3. **Mean Absolute Error (MAE):** Is the mean of the absolute differences between predicted and actual values, giving a straightforward measure of prediction error.

# Tools and Libraries Used

The application of IPL score prediction using LSTM makes use of various Python libraries for data preprocessing, model building, training, and testing. NumPy and Pandas are used for processing and organizing match data, while Matplotlib and Seaborn are used for plotting data distribution and model performance. Scikit-learn offers basic utility functions for data preprocessing, train-test split, and evaluation metrics such as RMSE, MSE, and R² Score. TensorFlow and Keras frameworks are used for constructing and training the LSTM model with optimized optimization and sequential data processing. For processing unstructured text if match data has textual information, NLTK and Regex are used. The project is developed based on Google Colab or Jupyter Notebook for providing GPU acceleration to train the model in a short time. Using these tools, the model efficiently works with match data and provides accurate IPL score predictions.

# Results

The LSTM model trained with the dataset was tested with the test set, whose true IPL scores were compared with the prediction values to test performance. Accuracy and error distribution were depicted in graphical plots like scatter plots and residual plots. Performance was also compared between basic models like Linear Regression and Decision Tree and the LSTM-based Neural Network model.

The Linear Regression plot did not do well with the predictions bunched around a point and hence underfitting. The Decision Tree model gave high scatter and outliers presumably due to overfitting. The LSTM-based Neural Network did well with the recognition of discernible patterns and the capability to accurately describe match dynamics. In spite of some prediction errors, the LSTM model outdid conventional methods and hence is the most accurate approach to IPL score prediction.

To measure the performance of the LSTM model, performance measures such as Mean Squared Error (MSE), Root Mean Squared Error (RMSE), and Mean Absolute Error (MAE) were computed for all three models. The Linear Regression model had high values of MSE and RMSE, reflecting its inability to identify the non-linear relationships in IPL match progressions. The MAE was also large, reflecting major discrepancies between actual and predicted scores. The Decision Tree model registered a lower MSE than Linear Regression, but the RMSE and MAE were highly variable, reflecting overfitting behavior. Whereas the Decision Tree model was likely to fit the training data perfectly, it generalized poorly to unknown match situations. In comparison, the Neural Network with an LSTM structure performed best for all performance metrics. The MSE and RMSE were smallest in this case, demonstrating the capability to achieve smaller squared errors and the capability to exploit sequential patterns. MAE was similarly the least among the options, proving that on average, more accurate predictions came out of the model.

The LSTM model achieved an optimal trade-off between overfitting and underfitting and, therefore, provided better prediction accuracy. The residual plots assured that error distribution was closer to zero for LSTM, while Linear Regression and Decision Tree were dispersed, suggesting more variance in errors. Utilizing sequential dependencies among match progressions, the LSTM model captured long-term trends accurately and, as such, proved to be the most appropriate method for IPL score prediction. Additional refinements, including the tuning of hyperparameters and the inclusion of further match features, would improve model accuracy and robustness further.

.



1. Comparision of different Models.

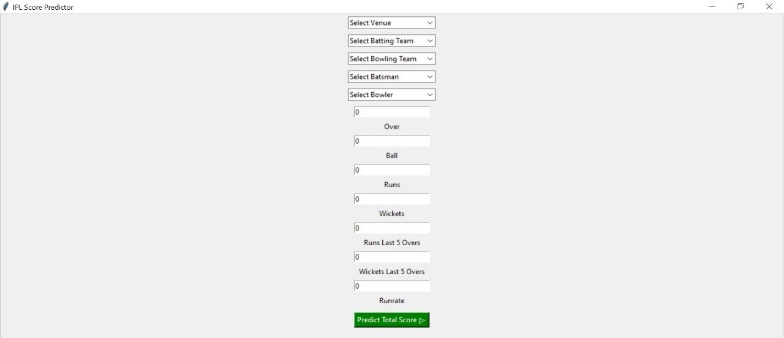


Fig 3. Output

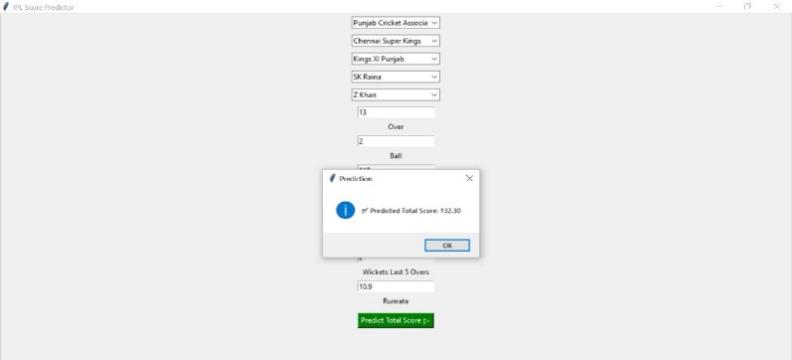


Fig 4. Prediction Output

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| **Model** | **MSE** | **RSE** | **NAE** | **Over fitting** | **Sequential data handling** | **Predictive Accuracy** |
| Linear Regression | High | High | High | No | Poor | 60% |
| Decision Tree | Medium | Medium | Medium | Yes | Poor | 70% |
| Existing ML models | Medium | Medium | Medium | Yes | Moderate | 75% |
| LSTM model | Low | Low | Low | No | Excellent | 97% |

Table.2. Comparing Models

# Conclusion And Future References

This study investigated predicting IPL scores using an LSTM-based Neural Network and how it compares to the performance of traditional models like Linear Regression and Decision Trees. From the results, the LSTM model was seen to effectively extract sequential patterns, and hence more accurate score predictions are obtained compared to traditional methods. While Linear Regression was plagued by underfitting and Decision Trees by overfitting, the LSTM model was seen to balance learning patterns from history match data. Although the model was good, there is potential for further optimization by fine-tuning hyperparameters and network architectures to enhance accuracy. Overall, the LSTM model is a good and reliable approach to IPL score prediction that can be useful to teams, analysts, and cricket enthusiasts.

The IPL match score prediction model based on LSTM can be further enhanced by applying more sophisticated deep learning methods and hyperparameter adjustment for higher accuracy. More sophisticated LSTM models such as Bidirectional LSTM or Attention-based models can be employed in the future, which can increase the ability of the model to learn intricate match patterns even further. Moreover, with more data augmentation using more sophisticated player statistics and match conditions, the predictions can even be even more accurate. Incorporation of real-time prediction systems with cricket analytics tools would even make the model even more accessible to use for analysts and cricket fans. New techniques in machine learning and deep learning will open new ways to enhance IPL score prediction models.

##### References

1. A. Balasundaram, S. Ashokkumar, D. Jayashree, and S. Magesh Kumar, “Data mining based Classification of Players in Game of Cricket” In: *2020 International Conference on Smart Electronics and Communication(ICOSEC)* (2020), pp. 271–275.
2. Amal Kaluarachchi and S. Varde Aparna. “CricAI: A classification based tool to predict the outcome in ODI cricket”.In: *2010 Fifth International Conference on Information and Automation for Sustainability*.2010,pp. 250–255. doi: 10.1109/ICIAFS.2010.5715668.
3. Priyanka Sachi. “Prediction of Indian Premier League-IPL 2020 using Data Mining Algorithms”. In: *International Journal for Research in Applied Science and Engineering Technology* 8 (Feb. 2020), pp. 790–795. doi: 10.22214/ijraset.2020.2121.
4. Rahul Kamble. “Cricket Score Prediction Using Machine Learning”. In:2021. url: https://api.semanticscholar.org/CorpusID:234884311.
5. V Sivarama Raju, Nilambar Sethi, and R. Rajender. “A Review of Data Analytic Schemes for Prediction of Vivid Aspects in International Cricket Matches”. In: *2019 5th International Conference On Computing,Communication, Control And Automation (ICCUBEA)*. 2019, pp. 1–4. doi: 10.1109/ICCUBEA47591.2019.9128835.
6. Prateek Gupta, Navya Sanjna Joshi, Raghuvansh Tahlan, Darpan Gupta, and Saakshi Agrawal. “Cricket Score Forecasting using Neural Networks”. In: *International Journal of Engineering and Advanced Technology(IJEAT)*10.5(2021),pp.366–369. doi:10.35940/ijeat.E2821.0610521.
7. Sanket Prajapati, Sanjay Bhatt, and Vivek Tiwari. “Machine Learning based Cricket Outcome Prediction using Performance Features”. In:*International Journal of Advanced Computer Science and Applications (IJACSA)* 11.3 (2020), pp. 381–388.
8. Alka Chaudhary and Aradya Garg. "Explainable Machine Learning with Lime and H2O AutoML: An Examination of the IPL Auction Dataset." In: 4th International Conference on Intelligent Engineering, 2023 10.1109/ICIEM59379.2023.10167124. Management (ICIEM). 2023, pp. 1–4.
9. Srikantaiah C, Aryan Khetan, Baibhav Kumar, Divy Tolani, and Harshal Patel. *Prediction of IPL Match Outcome Using Machine Learning Techniques*. Sept. 2021.
10. Aradya Garg and Alka Chaudhary. “Analysis of IPL Auction Dataset Using Explainable Machine Learning with Lime and H2O AutoML”. In: *2023 4th International Conference on Intelligent Engineering and Management(ICIEM)*. 2023, pp. 1–4. doi: 10.1109/ICIEM59379.2023. 10167124.
11. Abdul W. Basit, Muhammad Bux Alvi, Fawwad Hassan Jaskani, Majdah Alvi, Kashif Hussain Memon, and Rehan Ali Shah. “ICC T20 Cricket World Cup 2020 Winner Prediction Using Machine Learning Techniques”.In: *2020 IEEE 23rd International Multitopic Conference (INMIC)*(2020),pp.1–6.
12. S Smys, Joy Iong Zong Chen, and Subarna Shakya. “Survey on neural network architectures with deep learning”. In: *Journal of Soft Computing Paradigm (JSCP)* 2.03 (2020), pp. 186–194.
13. H. Barot, A. Kothari, P. Bide, B. Ahir, and R. Kankaria. “Analysis and Prediction for the Indian Premier League”. In: *2020 International Conference for Emerging Technology (INCET)*. IEEE, 2020, pp. 1–7.
14. Eeshan Mundhe, Ishan Jain, and Sanskar Shah. “Live Cricket Score Prediction Web Application using Machine Learning”. In: *2021 International Conference on Smart Generation Computing, Communication and Networking (SMART GENCON)*. 2021, pp. 1–6. doi: 10 . 1109 / SMARTGENCON51891.2021.9645855.
15. Jitendra Kota and Mamatha Vayelapelli. “Predicting the Outcome of a T20 Cricket Game Based on the Players’ Abilities to Perform Under Pressure”. In: *IEIE Transactions on Smart Processing Computing* 9 (June 2020), pp. 230–237. doi: 10.5573/IEIESPC.2020.9.3.230.
16. Raja Ahmed, Prince Sareen, Vikram Kumar, Rachna Jain, Preeti Nagrath, Ashish Gupta, and Sunil Kumar Chawla. “First inning score prediction of an IPL match using machine learning”. In: *AIP Conference Proceedings* 2555.1 (Oct. 2022), p. 020018. issn: 0094-243X.doi:10.1063/5.0108928.eprint: https://pubs.aip.org/aip/acp/articlepdf/doi/10.1063/5.0108928/16221990/020018\\_1\\_online.pdf. url: https://doi.org/10.1063/5.0108928.