IPL Match Score Prediction Using Dense Neural Networks: A Sports Analytics Approach

Dr. David Raj Micheal
Division of Mathematics
School of Advanced Sciences
Vellore Institute of Technology Chennai
Tamil Nadu – 600127
davidraj.micheal@vit.ac.in

Vedant Vidhate
Division of Mathematics
School of Advanced Sciences
Vellore Institute of Technology Chennai
Tamil Nadu – 600127
vedantvikas.vidhate2023@vitstudent.ac.in

***Abstract*—This project focuses on developing a deep learning model to predict final scores in Indian Premier League (IPL) cricket matches using historical match data. By leveraging match-specific features like team information, venue, and player statistics, the model forecasts the total score, providing insights that could aid strategic decision-making. Data preprocessing steps include feature selection, label encoding, and target variable extraction. Using a neural network architecture implemented with Keras and TensorFlow, the model is trained and evaluated for accuracy in predicting IPL scores. The results show promising predictive capability, with potential for further optimization and enhancement to improve real-world applicability.**

***Index Terms*—IPL Score Prediction, Dense Neural Networks (DNNs), Keras, TensorFlow, Sports Analytics**

I. Introduction

Predicting cricket scores has been a longstanding challenge in sports analytics, particularly in leagues like the Indian Premier League (IPL), where the dynamic nature of the T20 format introduces additional complexity. In this fast-paced format, scoring patterns can vary significantly based on factors such as team composition, venue conditions, and individual player performance. Accurate score predictions can offer valuable insights to coaches, teams, broadcasters, and fans, enhancing strategic decisions, commentary analysis, and viewer engagement.

The objective of this project is to build a predictive model using deep learning techniques such as Dense Neural Network (DNN) to estimate the final score of an IPL match before it reaches completion. By analyzing historical IPL data and training a model to identify patterns and trends, we aim to forecast the total score based on contextual information available during the match.

The model development involves several key steps: data preprocessing, including feature selection and encoding, neural network design, training, and evaluation. Deep learning frameworks like Keras and TensorFlow are used to create and optimize the model for predictive accuracy. The resulting model demonstrates the potential of machine learning and deep learning in capturing the nuances of cricket data, providing predictions that can complement expert analysis.

This project underscores the capabilities of artificial intelligence in sports and opens doors for future work, including more refined models with additional features and improved real-time applicability.

II. Objectives

The primary objective of this project is to develop a deep learning-based model capable of accurately predicting the final score in an Indian Premier League (IPL) match based on available match data up to a given point in the game. The goal is to create a predictive model that can provide valuable insights to aid strategic decision-making, enhance viewer engagement, and improve analytical commentary. To achieve this, the project will begin with data processing and feature engineering, where historical IPL match data will be preprocessed and relevant features influencing scoring patterns will be selected. Next, a neural network model will be built and trained using Keras and TensorFlow frameworks, with optimization focused on achieving high predictive accuracy. The model's performance will be evaluated using metrics such as Mean Absolute Error (MAE) to assess its effectiveness in predicting final scores. Lastly, the project aims to lay the groundwork for real-time score predictions that can be further adapted and optimized for live IPL matches.

III. Methodology

This section outlines the essential steps of building and evaluating a deep learning-based model for IPL score prediction, ensuring accuracy and practical usability for real-time applications.

A. Data Preprocessing

Data Loading: The IPL dataset was loaded from a CSV file. Columns that were not relevant to score prediction, such as *mid*, *date*, *batsman*, and *bowler*, were removed to focus on essential match-level statistics.

Categorical Encoding: Categorical variables like *venue*, *bat_team*, and *bowl_team* were converted into numeric values using Label Encoding. This transformation allowed the deep learning model to process these features effectively.

Feature Scaling: Since deep learning models benefit from normalized data, we standardized the numerical features using

`StandardScaler`. This transformation ensured that features had a mean of 0 and a standard deviation of 1, improving the model's convergence during training.

Data Reshaping: The data was reshaped into a 2D format suitable for input to a dense neural network. Additionally, the dataset was split into training and testing sets with an 80-20 ratio to evaluate the model's performance on unseen data.

B. Model Selection and Architecture

Model Choice: A Dense Neural Network (DNN) was selected instead of a Convolutional Neural Network (CNN). Given the structured tabular nature of the data, DNNs were more appropriate and likely to yield better accuracy.

Model Architecture:

- The neural network consisted of multiple dense (fully connected) layers, with units of 128, 256, 128, and 64 neurons.
- Dropout layers were introduced after each dense layer to reduce overfitting by randomly deactivating 30% of neurons during training.
- The output layer used a single neuron with no activation function to predict a continuous score (total runs).

Compilation: The model was compiled using the Adam optimizer, with a learning rate of 0.001, which is well-suited for training deep learning models on numerical data. The loss function chosen was Mean Squared Error (MSE), a standard metric for regression, while Mean Absolute Error (MAE) and Mean Absolute Percentage Error (MAPE) were used as additional evaluation metrics.

C. Model Training and Evaluation

Training Process: The model was trained for 200 epochs with a batch size of 32. A validation split was also set up to monitor the model's performance on unseen validation data during training, ensuring it generalizes well.

Evaluation Metrics:

- **Mean Absolute Error (MAE):** Measures the average magnitude of errors in predictions.
- **Mean Absolute Percentage Error (MAPE):** Helps in understanding the average percentage error, making it useful for reporting accuracy as:

$$\text{Accuracy} = 100\% - \text{MAPE} \tag{1}$$

IV. Results

In this section, we present the performance metrics of the IPL Score Prediction Model. These metrics evaluate the model's predictive accuracy on the test data.

- **Mean Absolute Error (MAE):** The Mean Absolute Error on the test data is given by:

  $$\text{Test MAE} = 6.57 \tag{2}$$

  This indicates that, on average, the model's predictions are within approximately 6.57 runs of the actual scores.
- **Mean Absolute Percentage Error (MAPE):** The Mean Absolute Percentage Error, which measures the average percentage error between predicted and actual values, is:

  $$\text{Test MAPE} = 4.26\% \tag{3}$$

  This suggests that the model's predictions deviate by an average of 4.26% from the actual scores.
- **Model Accuracy:** The model's accuracy, calculated as $100 - \text{MAPE}$, is:

  $$\text{Model Accuracy} = 95.74\% \tag{4}$$

  This high accuracy value indicates that the model is able to predict IPL scores with a good degree of reliability.

These results demonstrate that the model has achieved a high accuracy level on the test set, making it a useful tool for score prediction in IPL matches.

V. Key Insights

- **Accuracy:** The model achieved a high approximate accuracy of 95.74% based on the MAPE score, which indicates that it is capable of predicting IPL scores reasonably well.
- **Model Improvements:** Further improvements could involve experimenting with additional feature engineering, such as incorporating historical averages for specific players or venues.
- **Application:** The model provides a foundation for real-time IPL score prediction, which could be extended to applications in sports analytics, betting, or fan engagement platforms.

This project demonstrated the effectiveness of deep learning in predicting cricket scores and highlighted the importance of preprocessing and model selection for structured data.

VI. Conclusion

The dense neural network model built for IPL score prediction achieved satisfactory accuracy, as observed from the MAPE and MAE metrics. The model's architecture was appropriate for the structured dataset, and the inclusion of Dropout layers helped reduce overfitting, ensuring it generalized well to unseen data.