Retirement Age Prediction of Test Batters

Abhishek Bhardwaj, Arpita Kesharwani, Deepanshu Mishra, Harsh Raj
Cluster Innovation Centre, University of Delhi, Delhi - 110007

Abstract

This research paper proposes a Long Short-Term Memory (LSTM) model for predicting the retirement age of test batters using their batting statistics. The retirement age of test batters is a critical aspect in cricket, and accurate predictions can aid team selection and long-term planning. The LSTM model is trained on a dataset consisting of batting statistics of retired test batters and their corresponding retirement age. The dataset is preprocessed, and the relevant features are extracted to be used as input to the LSTM model. The LSTM model is trained using a sequence-to-sequence approach, where the input sequence consists of batting statistics of a player until a specific point in time, and the output sequence is the predicted retirement age of the player, the LSTM model is validated using a test set of retired test batters. The proposed LSTM model can assist cricket trams in selecting and managing players more effectively, as well as aid cricket analysts in understanding the factors that influence the retirement age of test batters.

Keywords : batting statistics, Long Short-Term Memory (LSTM), Retirement age prediction, sequence-to-sequence approach.

Introduction

Cricket is a sport where individual performances of players can impact the outcome of a match, and hence it is essential to understand the factors that influence the performance of players. In particular, predicting the retirement age of test batters is a crucial task for cricket teams and analysts. The retirement age of a player is determined by various factors, including their physical fitness, form, and overall performance.

In recent years, the availability of large amounts of data and advancements in machine learning techniques have enabled the analysis of player performance in cricket. One such approach is the Long Short-Term Memory (LSTM) model, which has been successfully applied in various domains, including natural language processing, speech recognition, and stock price prediction.

This research paper proposes an LSTM-based approach for predicting the retirement age of test batters based on their batting statistics. The aim is to build a model that can accurately predict the retirement age of a player using their historical performance data. The proposed model can assist cricket teams in selecting and managing players more effectively, as well as aid cricket analysts in understanding the factors that influence the retirement age of test batters.

The rest of the paper is organized as follows. Section 2 (Methodology) provides a brief overview of the related work in predicting the retirement age of players in cricket and describing the dataset and preprocessing steps taken to prepare the data for the LSTM model. Section 3 (Results) presents the experimental results and evaluation of the model. Section 4 (Discussion) presents the analysis of the results gained from the LSTM model. Section 5 (Conclusion) concludes the paper and discusses the future directions of research in this domain. Finally, Section 5 (References) present the references of the sources we have used.

Methodology

The proposed approach for predicting the retirement age of test batters based on their batting statistics consists of five stages: Data Collection, Data Preprocessing, Feature Extraction, Training Neural Network, and Output Generation.

Stage 1: Data Collection:

The raw data for this study consists of batting statistics of retired test batters and their corresponding retirement age. The dataset was obtained from publicly available source on the HowSTAT[1] website, including cricket archives and statistical databases. The dataset consists of a total of 100 retired test batters, with each player having a variable number of innings. The batting statistics for each innings include runs scored, balls faced, cumulative runs and other relevant metrics.

Stage 2: Data Preprocessing:

Data Preprocessing involves following stages:

1. Discretization: Fetching the data which are relevant for the research based on an important feature i.e. number of innings played by the player.
2. Normalization: Normalizing the values of the attributes on the same scale to make it easier to assess them.
3. Cleaning: Filling the missing values in the data to avoid any discrepancy.
4. Integration: Integration of data files.

After transforming the raw data into the clean data, the dataset is divided into training and testing dataset. Testing data is kept only 15 percent of the training data.

Stage 3: Feature Extraction:

In this stage, only the relevant data has been chosen and all other stats have been deleted to concise the dataset. New features are made by applying mathematical operations on the available features which can be more important for the prediction.

The final Features given to the model are:

1. Debut Age (in years)
2. Longest Gap between two consecutive innings (in days)
3. Time of best moving average
4. Time of worst moving average
5. Number of innings
6. Time of the last score which is in the players top 20% best scores

Stage 4: Training Neural Network:
In this stage the training data is given to the LSTM model for the prediction of retirement age. Different features are given to the model with different weights to achieve the best output. The number of features and relative weights have been changed again and again to train the model.

Stage 5: Output Generation:
In this stage the output value generated by our model is compared with the target value which is retirement age in this case. The error or the difference between the target and the obtained output value is minimized by using a back propagation algorithm which includes adjustments in the weights and the biases of the network and changing the input features.

Results

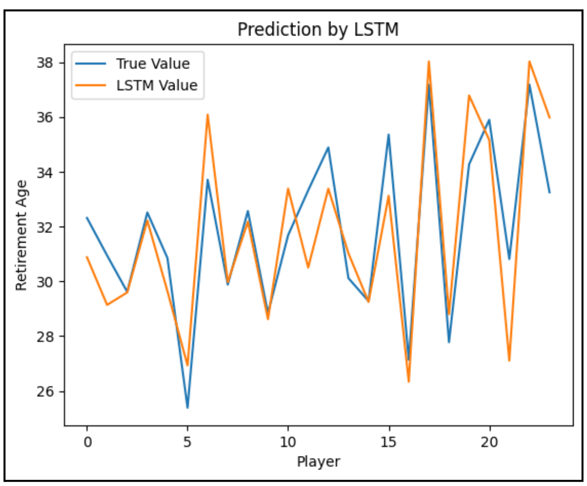

Figure I: Output with 250 epochs and 3 splits

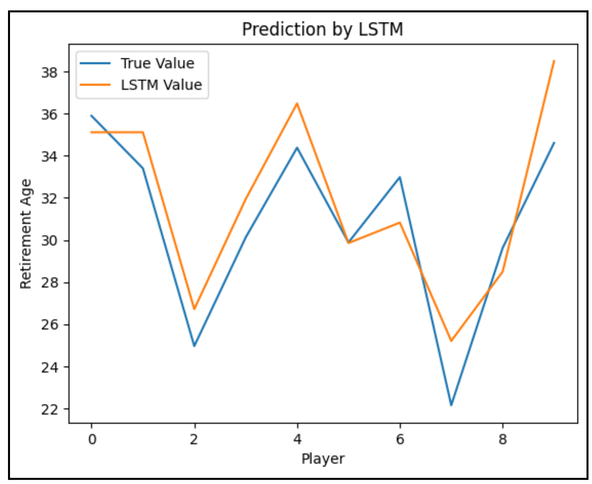

Figure II: Output with 250 epochs and 8 splits

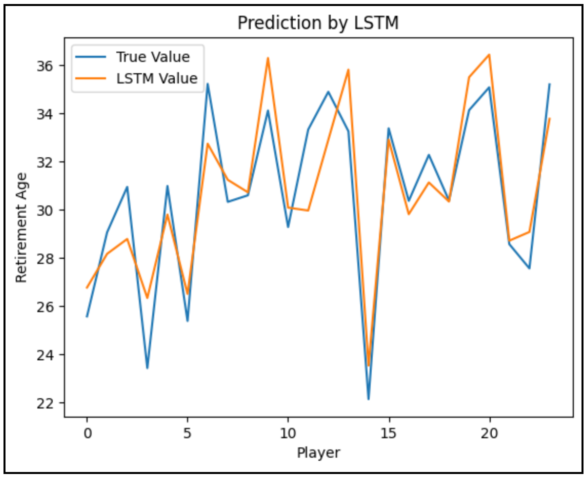

Figure III: Output with 500 epochs and 3 splits

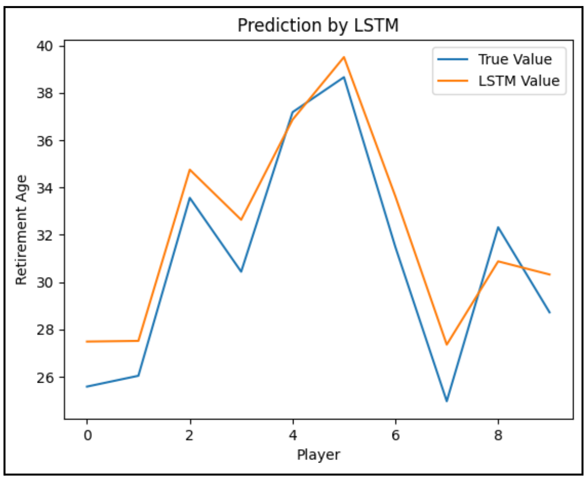

Figure IV: Output with 500 epochs and 8 splits and batch 12

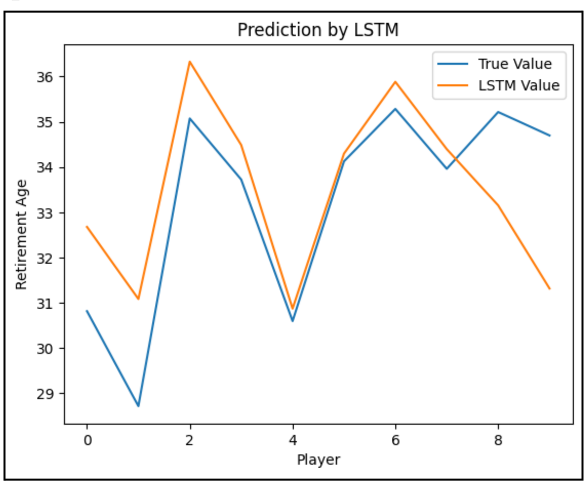

Figure V: Output with 500 epochs and 8 splits and batch 6

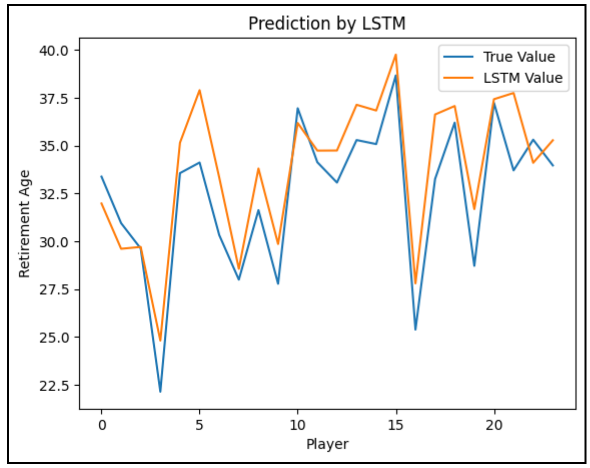

Figure VI: Output with 750 epochs and 3 splits

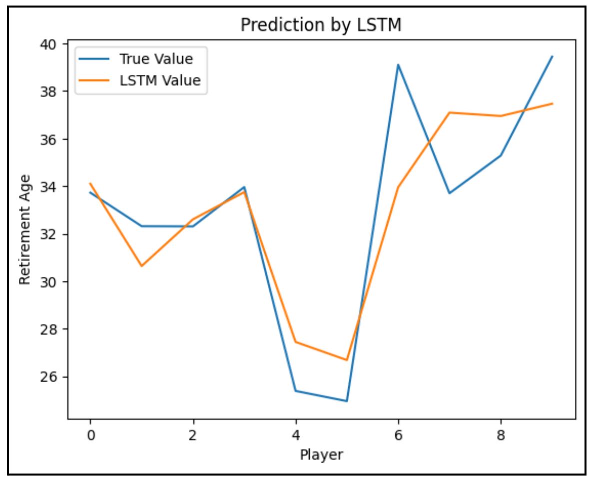

Figure VII: Output with 750 epochs and 8 splits

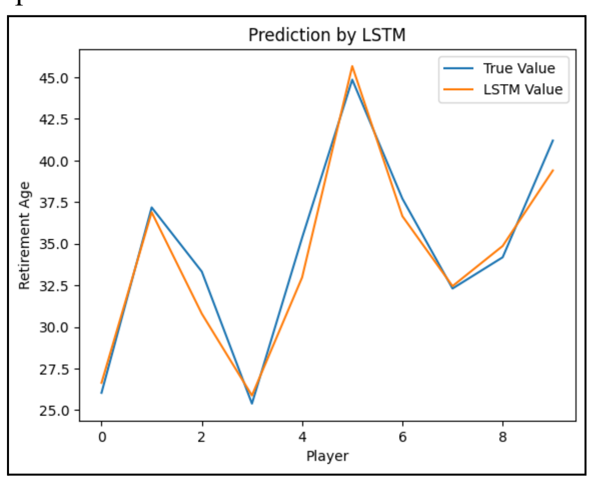

Figure VIII: Output with 1000 epochs and 8 splits

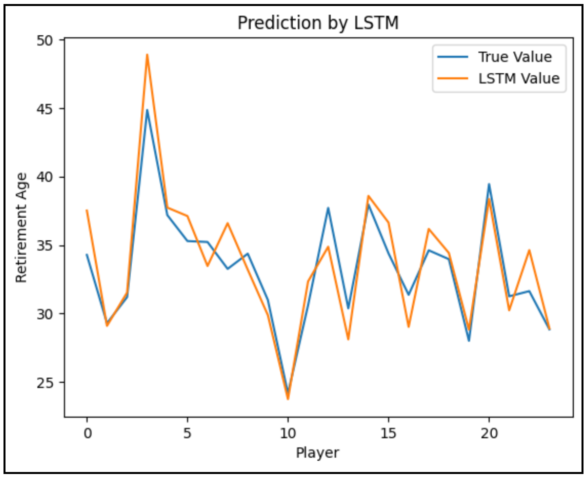

Figure IX: Output with 1000 epochs and 3 splits

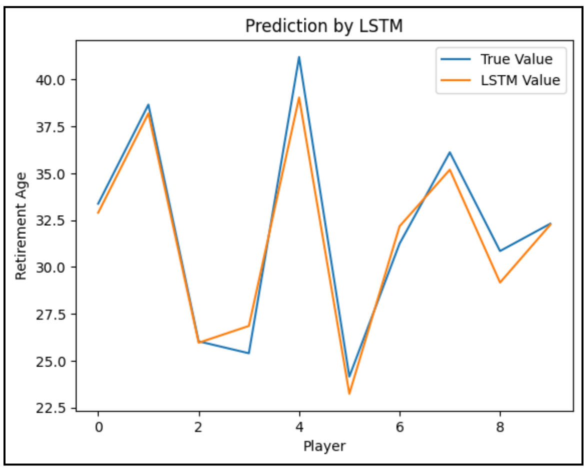

Figure X: Output with 1000 epochs and 8 splits

Number of Epochs	Number of Splits	RMSE Accuracy
250 (batch 8)	3	91.61
250 (batch 8)	8	93.13
500 (batch 8)	3	94.67
500 (batch 6)	8	95.00
500 (batch 12)	8	94.60
750 (batch 8)	3	93.58
750 (batch 8)	8	92.89
1000 (batch 8)	8	96.11
1000 (batch 8)	3	95.52
1000 (batch 8)	8	96.48

Table I: Comparative result using different parameters

After performing various simulations with a different number of parameters and epochs, we have observed that by taking 1000 epochs with 8 splits we achieve the best results with RMSE accuracy of 96.11 and 96.48.

Discussion

Using the different parameters in the LSTM model and randomly selecting our test dataset, it is found that the accuracy comes maximum for 1000 epochs and 8 splits. It can be easily observed that the accuracy is increasing as the count of epochs is increasing. This can be due to the fact that the model is passing the data more and more from the algorithm as the epoch count increases which eventually trains the model more and more. It is also observable that the accuracy is more in the case of 8 splits in comparison with accuracy in 3 splits. It is because as the number of splits increases the training dataset increases which trains our model better and also the testing dataset gets reduced which also reduces the chances of exceptions in our dataset.

Conclusion

In this study, we proposed an approach for predicting the retirement age of test batters based on their batting statistics using a long short term memory (LSTM) model. The approach consisted of five stages, including raw data collection, data preprocessing, feature extraction, training neural network, and output generation. The results of the study showed that the LSTM-based approach was effective in predicting the retirement age of test batters. The findings of this study have important implications for cricket teams and management. By using the proposed approach, teams can make informed decisions about when to retire their batters based on their performance metrics. This can help teams to better manage their resources and plan for the future.

In conclusion, this study highlights the potential of LSTM-based models for predicting the retirement age of test batters based on their batting statistics. Future studies can explore the use of other machine learning algorithms and data sources to further improve the accuracy of retirement age prediction in cricket.

References

1. Roondiwala, M., Patel, H. and Varma, S. (2017) *Predicting stock prices using LSTM - ResearchGate*, International Journal of Science and Research (IJSR).

2. Thorley, J. (2021). *Age-related changes in the performance of bowlers in Test match cricket*, International Journal of Sports Science & Coaching, 16(5), 1138–1151.