Performance Prediction of Cricket Player Using Blockchain Enabled HMM Model

Sai Radha Krishna G[1,2(✉)], Sandeep Kalakota[1], and Regula Harshavardhan[1]

[1] Department of ECE, Rajiv Gandhi University of Knowledge Technologies, Basar, Telangana, India
sairadhakg@am.amrita.edu, {b171269,b171867}@rgukt.ac.in

[2] Center for Wireless Networks and Applications (WNA), Amrita Vishwa Vidyapeetham, Amritapuri, India

Abstract. Sports play an essential role for human life which keeps physical and mental fitness. Competition in sports is rising day by day. In sports giving opportunity to right talented person is a challenging task. In countries like India, team sport like cricket faces huge competition. This paper introduces the performance prediction of cricket player using blockchain enabled Hidden Markov Model (HMM). In the proposed work, the time series data of matches played by batsmen during past one year is collected, analyzed and performance of a batsman is predicted by using HMM. The data obtained including the predicted results are stored in the proposed blockchain with IPFS based system architecture for data management, decentralization, data security and immutable. It focuses on the players rather than the whole match outcome so best players can team up.

Keywords: Hidden Markov Model (HMM) · Blockchain · InterPlanetary File System (IPFS) · Cricket · Player Performance Prediction

1 Introduction

Sports play a crucial role in a person's growth and development. Unlike ancient days nowadays the competition in sports is significantly rising as many educational institutions are focusing in this area as it not only gives exposure to the students to showcase their talent in bigger platforms but also represent their institutions.

Performance of a player in any sport is a crucial thing and it depends on different factors such as experience, score, physical health, mental well being and training. In a sport like cricket it is more important that the right person gets the opportunity. In conventional method of selecting a sportsperson, a highly experienced authorized team of selection committee takes decision based on his previous performance. But this method of selection of players has some disadvantages like recommendations, bias towards a particular player, political pressure, bribing etc. which is restricting the real talent to get the opportunity.

In order to overcome this, a system is needed to select the best players. This paper introduces a system to predict the performance of a player based on past data so that

B. Raj et al. (Eds.): ICETE 2023, AER 223, pp. 438–451, 2023.
https://doi.org/10.2991/978-94-6463-252-1_47

selection of a player is done on the basis of performance and is not affected by individual decisions. Hidden Markov Model (HMM) is used which gives accurate predictions based on given data and can handle data sparsity, time related data and dynamic nature of data. HMM has hidden states, represents factors such as physical health, confidence, happiness, sad etc. of a player which are not observable in a game.

However as the number of matches played by a player increases, it generates huge data and makes it difficult for traditional centralized data processing system to handle huge data transmission and processing properly. To overcome this drawback a blockchain integrated InterPlanetary File System (IPFS) is introduced for storage and management of data which reduces the data maintenance problems, provides data security and decentralization. Also, the chances of data tampering of the performance prediction results of a player is dealt by the blockchain concept which provides data security and ensures the correctness of data.

A system is designed to predict the batsmen performance based on past data using HMM and storing, managing that data using blockchain technology.

2 Literature Survey

Recommendation of right players is important as they represent the face of a nation in international games like Olympics, commonwealth games, Wimbledon, cricket world cup etc., Present day advanced methods like technology like Machine Learning and Blockchain technology is introduced in place of traditional method of selection process to maintain transparency. P. Cao et al. predicts the performance of athletes using HMM and compares with other prediction models in terms of accuracy. Blockchain technology is introduced to overcome traditional centralized data storage and management problems [1]. Sai Radha Krishna et al. introduces a Food Supply Chain Traceability System (FSCTS) by using Ethereum smart contract and IPFS which can solve several problems like sustainability, not effectively utilizing the resources, wasting the food, data fragmentation, centralization, trust deficit, auditability, verifiability and regulatory issues between different entities [2]. Fei Guan et al. proposes a method of fault alarm recognition based on HMM that takes time series into account and the BIT result is taken as the data source. A case study shows that the method is effective and gives an accurate state of the equipment [3]. Mengxiao Luo et al. predicts the sports performance of primary school students is influenced by factors like body shape, physiology, and physical fitness. The experimental result shows that multiple linear regression model has the best performance on the data set and the result is consistent with previous researches [4]. Gianluca et al. predicts the performance of football player during training using machine learning. Multivariate regression method is used to attempt performance forecasting of football players during training sessions, starting from data about their movements and physiological parameters, extracted through an array of sensors integrated in the vests worn by the athletes [5]. Abdul Basit et al. predicted the ICC T20 world cup 2020 winner using random forest machine learning algorithm with accuracy of 80.86% but no importance to an individual player [13]. Tejinder Singh et al. proposed a model to predict the match outcome in the ODI format of cricket using Linear Regression classifier and Naive bias classifier which doesn't focus on any hidden factors [14].

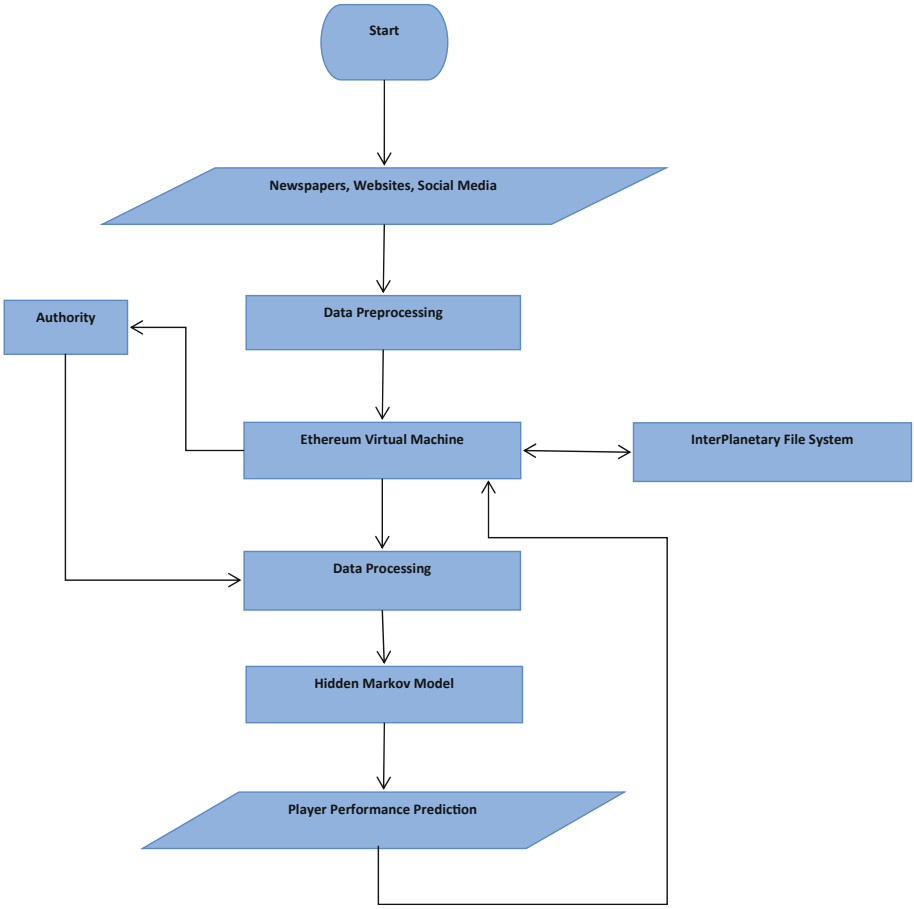

Fig. 1. System Architecture.

3 System Architecture

As can be shown from Fig. 1, the proposed system consists of different blocks each doing specific work. Initially, system collects necessary data about cricket from different resources like websites, social media, newspapers etc. Then, preprocessing of the raw data is done which organizes and process the data which is necessary for further implementation. This data is forwarded to the Ethereum Virtual Machine which stores data securely in IPFS. Data processing module loads the data from Ethereum Virtual Machine and processes it according to the given inputs by the authority and then data is send to the HMM which gives player performance prediction which is stored securely.

Table. 1. Categorization of runs and assigning points

Runs	Points
0–29	0
30–49	1
50–99	2
100–149	3
150–199	4
200 - above	5

Table. 2. Categorization of strike rate and assigning points

Strike rate	Points	Strike rate	Points
0–24.9	0	125–149.9	5
25–49.9	1	150–174.9	6
50–74.9	2	175–199.9	7
75–99.9	3	200-above	8
100–124.9	4		

4 Implementation Framework

4.1 Data Preprocessing

The raw data collected from different sources such as social media, websites, newspapers is organized and the data corresponding to those players required by the authority is extracted and points are assigned to players performance by taking into account, different criteria like runs and Strike rate as shown Fig. 2. For runs scored and Strike rate by a player in a match, the allocation of points are shown in Tables. 1 and 2

4.2 Ethereum Virtual Machine & InterPlanetary File System

The data after preprocessing is stored in the IPFS using Ethereum Virtual Machine which is the run time environment of the blockchain. Blockchain is used for the purpose of decentralization, security, privacy and to prevent data tampering. As the number of matches played by a particular player increases, the data to be stored increases. Due to the high gas fee of Blockchain Technology, the system cannot handle such huge data. To accomplish this, a decentralized system called IPFS is introduced which finds the data in the system using content addressing whenever required and thus improves the availability of the data in the system. In the proposed system data is stored in the IPFS and the corresponding hashes obtained by using cryptography are stored in the blockchain.

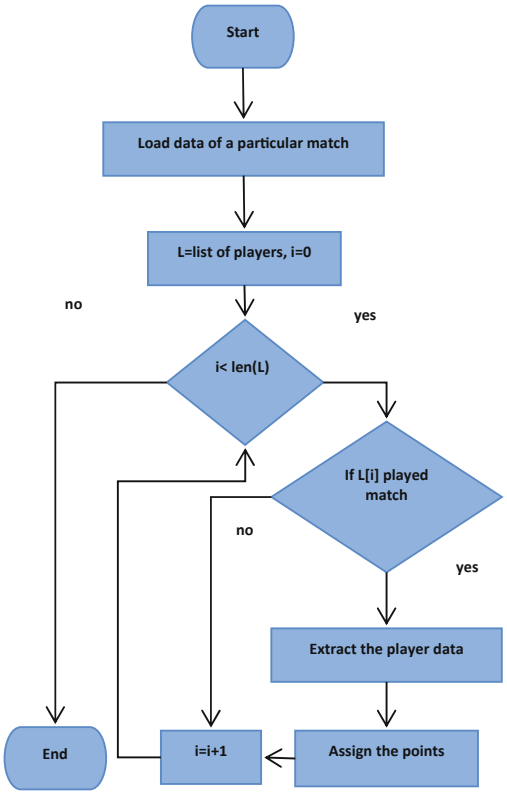

Fig. 2. Flow chart of data preprocessing.

4.3 Data Processing

When authority asks for the performance prediction of a player, it takes data from the Ethereum Virtual Machine and process the data based on the inputs(Player Name, Format) given by the authority which produces the observation sequence and forward the observation sequence to the HMM as shown in Fig. 3.

4.4 Hidden Markov Model

HMM predicts the (T + 1)th, (T + 2)th,….time observations by analyzing and processing the time series data up to Tth time in terms of probabilities. It considers the hidden states such as physical and mind state of a player which are not observable in a game. The probability of going from hidden state Si at time 't' to the hidden state Sj at time 't + 1' is called transition probability and the probability of going from hidden state Si at time t to the observation state Oj at time (t + 1) is called emission probability. Transition matrix is values of probability of transition from one hidden state to another hidden state. Emission matrix is values of probability of emission from one hidden state to observation. At the beginning the transition and emission matrices considered with random values. Initial

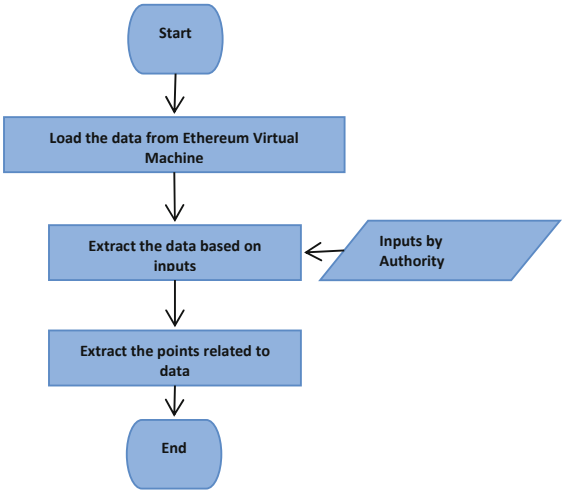

Fig. 3. Flow chart of data processing.

probabilities are the probabilities of going to any hidden state from time zero. In this system all initial probabilities are considered with 0.5 value.

For analysis in the proposed system, Baum-Welch Algorithm [6], Viterbi Algorithm [9] and Prediction Algorithm as shown in Fig. 4 are considered. Baum-Welch algorithm takes observation sequence generated from data processing and transition, emission matrices with random values as an input and gives trained transition, emission matrices as shown in Fig. 5. Viterbi algorithm takes observation sequence and trained transition, emission matrices as an input and gives sequence of hidden states as shown in Fig. 6.

In prediction algorithm, it takes trained transition and emission matrix as an input and gives probability of observations. This system categorizes the runs and strike rate into 6 and 9 observations as points respectively (Tables 1 and 2) and two hidden states (Happy, Stressed) are considered. For each observation, all the possible paths from initial state to observation state are considered as shown in Fig. 7 and for every path all the transition probabilities and emission probabilities in that path are multiplied then values of those paths are added which is the probability of getting that observation as an outcome. Same procedure is done for each and every observation as shown in Fig. 8. Probabilities of all the observations are compared and observation with maximum probability is considered as an outcome. All three algorithms are applied to both runs & strike rate categories (Table 3).

5 Results & Analysis

A dataset is created based on the performance of fourteen current Indian batsmen in which the player match details like format, runs, strike rate, fours, sixes are mentioned as shown in Fig. 9. In this model the performance of a player is considered based on the runs and Strike rate categories in a particular format as shown in Fig. 10. This dataset is given to HMM and the results of every player in all three formats T20, ODI and

Table. 3. List of notations used in the figure 7 and figure 8

Notation	Explanation
x	No of hidden states
y	No of observations
p	Initial State
H	Happy
S	Stressed
a_model	Transition Matrix
b_model	Emission Matrix
O	Observation State
D	List of observation probabilities

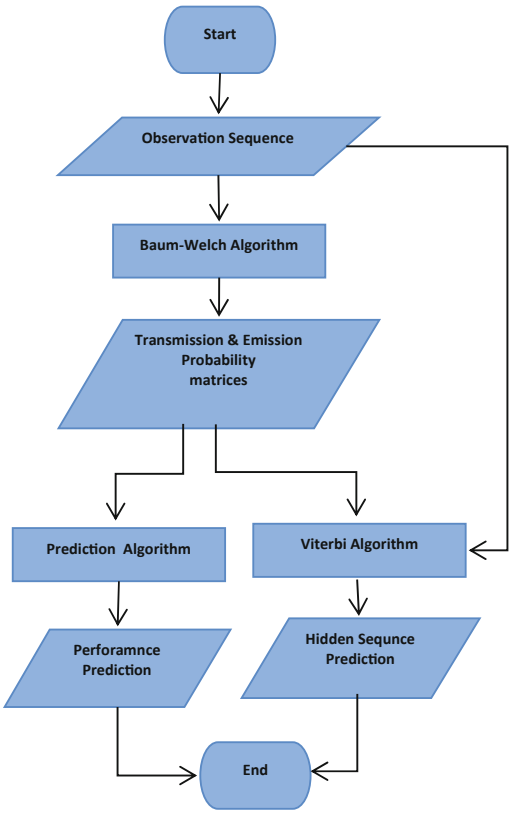

Fig. 4. Flow chart of training the model, predicting outcome and hidden sequence.

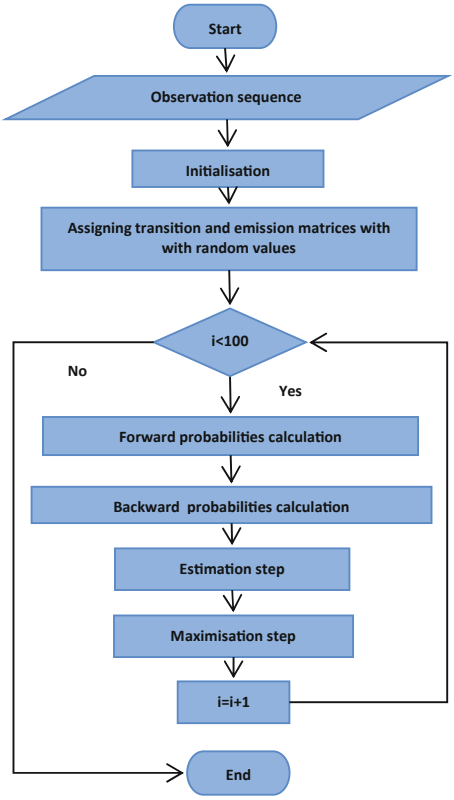

Fig. 5. Flow chart of Baum-Welch algorithm to train the model.

Test is predicted in both runs and strike rate categories. As shown in Fig. 13 these are the runs scored by a player Suryakumar Yadav in his previous matches and in Fig. 14 is corresponding Strike rate of those matches in the T20 format. The data is further processed based to get observation sequence which is given to HMM which gives the probability of runs and strike rate by the player in the next match as shown in Figs. 15 and 16.

Average model is also applied to dataset and results of HMM and average model are compared as shown in Figs. 11 and 12. Average model predicts a particular value whereas HMM predicts the probability of an outcome over a range of values. Probability of all the observations is compared and observation with maximum probability is taken as outcome.

In Figs. 11 and 12, for Virat Kohli in an ODI format average model shows 27.4 (blue) runs with strike rate 75.1(blue) respectively where as HMM shows runs in the range of 100(orange) to 149 (green) and strike rate in the range of 75 (orange) to 99.9(green) respectively. Similarly for Suryakumar Yadav in an ODI format average model shows 21.6(blue) runs with strike rate 99.7(blue) respectively where as HMM shows runs in the range of 0(orange) to 29(green) and strike rate in the range of 75(orange) to 99.9(green) respectively. Results(Figs. 13, 14, 15 and 16) are predicted format wise i.e. there is no

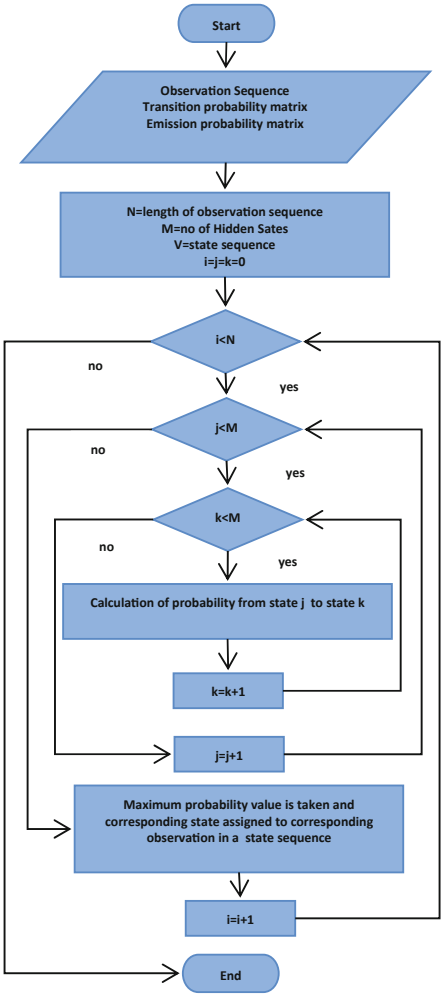

Fig. 6. Flow chart of Viterbi algorithm for finding hidden sequence.

relation between formats T20, ODI & Test in terms of runs or strike rate. It is observed that accuracy of the result depends on the length of observation sequence i.e. more matches the player plays more is the accuracy towards real outcome(100% probability) which concludes that experience of a player plays a crucial role in the system in terms of length of observation sequence.

Viterbi algorithm predicts the performance of the player in a particular emotional status(happy or stressed). Prediction results are again stored in the blockchain integrated IPFS which is accessible only by concerned authority and cannot be tampered in any way.

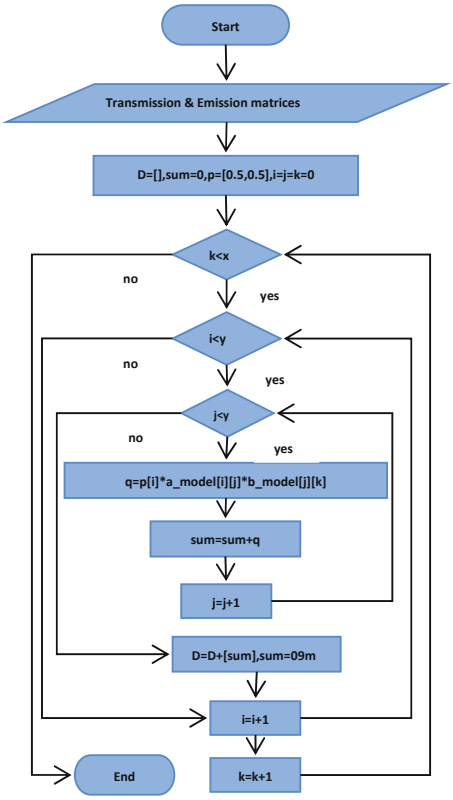

Fig. 7. Flow chart of prediction algorithm.

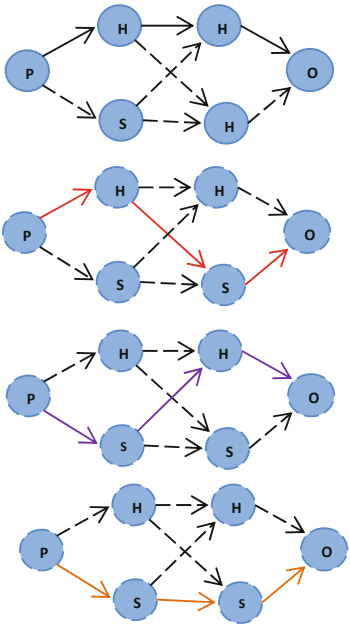

Fig. 8. All possible paths from initial state to observation state in prediction algorithm.

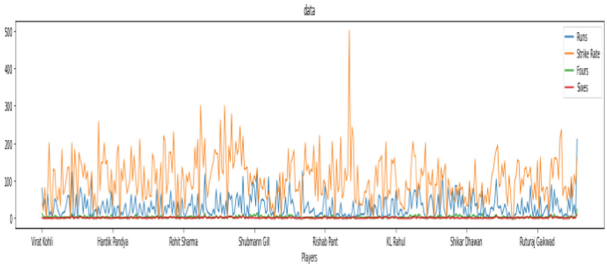

Fig. 9. Dataset.

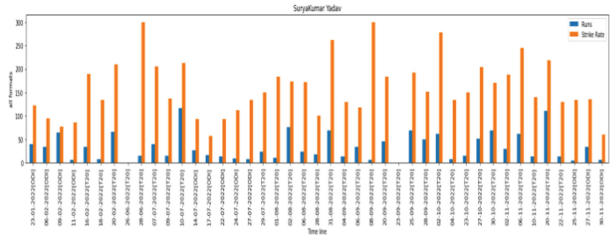

Fig. 10. Player data in all formats.

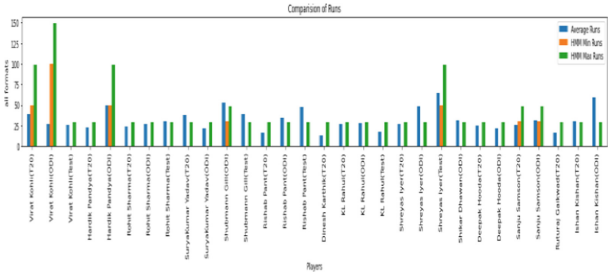

Fig. 11. HMM vs Average (Runs).

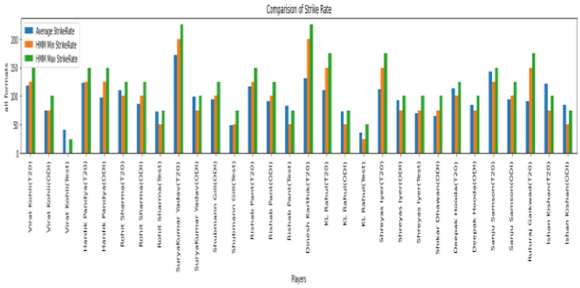

Fig. 12. HMM vs Average (Strike rate).

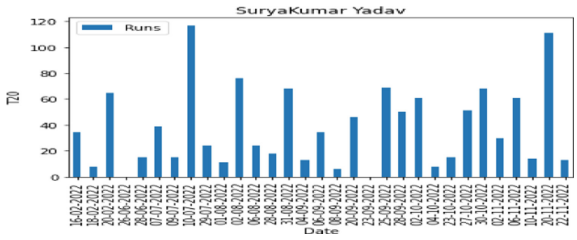

Fig. 13. Runs of a player in a required format.

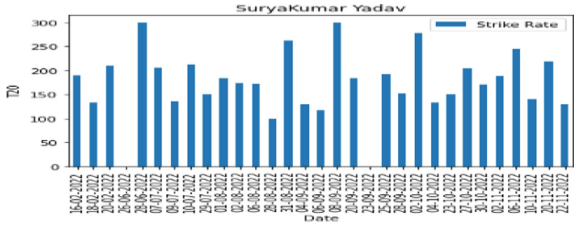

Fig. 14. Strike rate of a player in a required format.

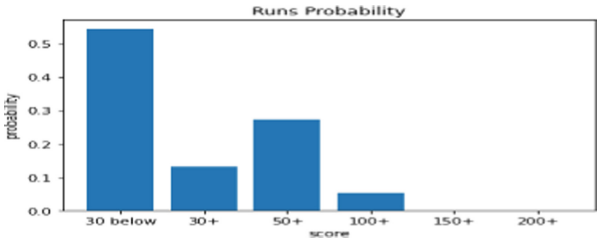

Fig. 15. Probability of runs of a player in a required format.

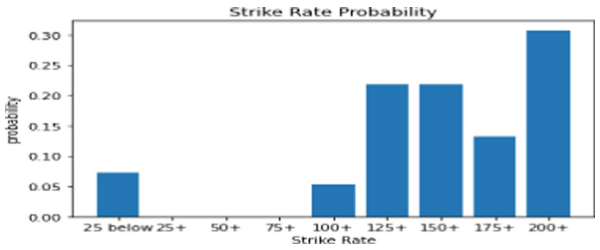

Fig. 16. Probability of strike rate of a player in a required format.

6 Conclusion

The introduction of modern day technology like blockchain networks and machine learning models have made selection of the talented players easy. In the proposed model the performance of batsmen is predicted by HMM based on the performance of the player during last one year. Further,integration of blockchain and machine learning models can be used for future training plans of a player like providing suitable environment, planning different drill sessions for the improvement and a hidden sequence is predicted which is useful for the further analysis of a player. This model also finds its applications in selection of bowlers, wicket keepers and fielders. This system can also be applied for different sports.

Acknowledgement. We would like to express our sincere gratitude to Prof. V. Venkata Ramana, Vice Chancellor and Prof. Satish Kumar Peddapelli, Director, Rajiv Gandhi University of Knowledge Technologies, Basar for their motivation and providing all facilities to do this project. We owe a particular debt of gratitude to Dr. Rekha P, Assistant Professor and Dr. Hemalatha T, Assistant Professor, Center for Wireless networks & Applications(WNA), Amrita Vishwa Vidyapeetham, Amritapuri for their valuable guidance and significant suggestions.

References

1. Cao, Ping, Guoqing Zhu, Qingguo Zhang, Fan Wang, Yuwen Liu, and Ruichao Mo. "Blockchain-enabled HMM model for sports performance prediction." IEEE Access 9 (2021): 40255–40262.

2. Krishna, G. Sai Radha, and P. Rekha. “Food Supply Chain Traceability System using Blockchain Technology.” In 2022 8th International Conference on Signal Processing and Communication (ICSC), pp. 370–375. IEEE, 2022.
3. Guan, Fei, Jie Wu, and Weiwei Cui. “A method of fault alarm recognition based on HMM model.” In 2016 Prognostics and System Health Management Conference (PHM-Chengdu), pp. 1–4. IEEE, 2016.
4. Luo, Mengxiao, Youling Yu, Weisheng Xu, Qiang Zhang, and Jun Xia. “Analysis and Prediction on Sports Performance of Primary School Students.” In 2020 Chinese Automation Congress (CAC), pp. 1041–1045. IEEE, 2020.
5. Morciano, Gianluca, Andrea Zingoni, Andrea Morachioli, and Giuseppe Calabrò. “Machine Learning prediction of the expected performance of football player during training.“ In 2022 IEEE International Conference on Metrology for Extended Reality, Artificial Intelligence and Neural Engineering (MetroXRAINE), pp. 574–578. IEEE, 2022.
6. Zhang, Jingwen, Fanggang Wang, Zhangdui Zhong, and Shilian Wang. “Continuous phase modulation classification via Baum-Welch algorithm.” IEEE Communications letters 22, no. 7 (2018): 1390–1393.
7. Turin, William. “Unidirectional and parallel Baum-Welch algorithms.” IEEE Transactions on Speech and Audio Processing 6, no. 6 (1998): 516–523.
8. Forney, G. David. “The viterbi algorithm.” Proceedings of the IEEE 61, no. 3 (1973): 268–278.
9. Lou, H-L. “Implementing the Viterbi algorithm.” IEEE Signal processing magazine 12, no. 5 (1995): 42–52.
10. Mitani, Tatsuo, and Akira Otsuka. “Traceability in permissioned blockchain.” IEEE Access 8 (2020): 21573–21588.
11. Shahbazi, Zeinab, and Yung-Cheol Byun. “Blockchain-based event detection and trust verification using natural language processing and machine learning.” IEEE Access 10 (2021): 5790–5800.
12. Shynu, P. G., Varun G. Menon, R. Lakshmana Kumar, Seifedine Kadry, and Yunyoung Nam. “Blockchain-based secure healthcare application for diabetic-cardio disease prediction in fog computing.” IEEE Access 9 (2021): 45706–45720.
13. Abdul Basit, Muhammad Bux Alvi, Fawward Hassan Jaskani, Majdah Alvi, Kashif H.Memon, Rehan Alishah.”ICC T20 Cricket World Cup 2020 Winner Prediction Using Machine Learning Techniques.” IEEE, 2020.
14. Tejinder Singh, Visha Singala, Parteek Bhatia. “Score and Winning Prediction in Cricket through Data Mining.” IEEE, 2015.