CRICKET SCORE PREDICTION SYSTEM (CSPS) USING CLUSTERING ALGORITHM

Prof. Preeti Satao[1], Ashutosh Tripathi[2], Jayesh Vankar[3], Bhavesh Vaje[4], Vinay Varekar[5]
[1]Asst. Prof. Department of CS, [2,3,4,5]Student BE Comps
Email: Preeti.satao@mctrgit.ac.in[1], Ashu.tripathi426@gmail.com[2], Jayeshvankar54@gmail.com[3], Bhavesh.vaje94@gmail.com[4], vinay.varekar@gmail.com[5]

Abstract
Cricket is a popular team sport played internationally. It has tremendous spectator support and the masses show great interest in predicting the outcome of games both in their one-day international as well as the modern T-20 format. The game is governed by complex rules and scoring system. Accurate prediction of winning or losing a match faces significant challenges. Multiple parameters, including cricketing skills and performances, match venues can significantly affect the outcome of a game. These diverse parameters, along with their interdependence and variance create a non-trivial challenge to create an accurate prediction of a game. In this paper, we build a prediction system that takes in historical match data, player performance as well as the scores predicted by spectator, and predicts future match events culminating in a victory or loss. Our system predicts match outcome by analyzing pre-stored match data using simple but effective K-means clustering algorithm. We describe our system and algorithms and finally present quantitative results, demonstrating the performance of our algorithms in predicting the number of runs scored, one of the most important determinants of match outcome.
Keywords: Sports prediction, K-means, analysis, clustering

I. INTRODUCTION

Cricket has the second largest viewership for any sport and generates an immensely passionate following among the supporters. There is huge commercial interest in player performance prediction. This has motivated many analysis of individual and team performance, as well as prediction of future games, across all formats of the game. Currently, strategists rely on a combination of player experience, team constitution and “cricketing sense" for making instantaneous strategic decisions. We choose to focus our testing and evaluation on the most popular format, namely T-20 cricket played for Indian Premier League (IPL).

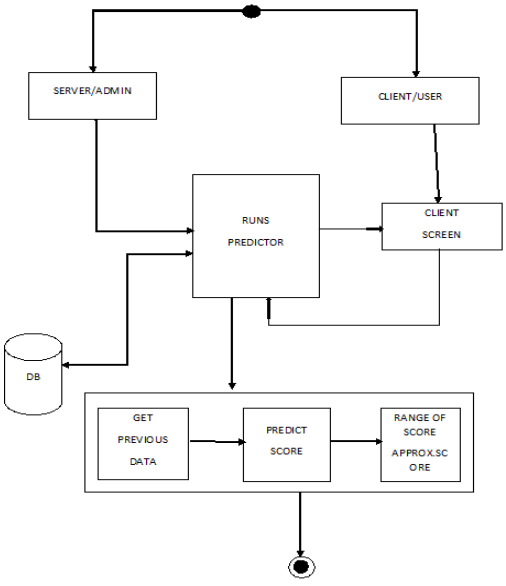

Fig.1 Block Diagram

By using unsupervised learning algorithms, our approach learns a number of features from T-20 cricket dataset which consists of complete records of all games played since the beginning of IPL in the year 2009. For every match that is

being played our system predicts the player's score by checking scores from database and also taking into account the scores predicted by fans watching the match. The system outputs a range of probable score that the player will make on that particular round.

Data about player's previous performance is stored in MySQL database classified with respect to the ground. The business logic is written in Java.

II. DESIGN OF CSPS

A. Class Diagram

The class diagram for the proposed system describes the system in terms of classes, attributes, operations, and their associations. In UML, classes and objects are shown by boxes composed of three compartments. Top compartment displays the name of the class or object. The Centre compartment displays its attributes; the bottom compartment displays its operations.

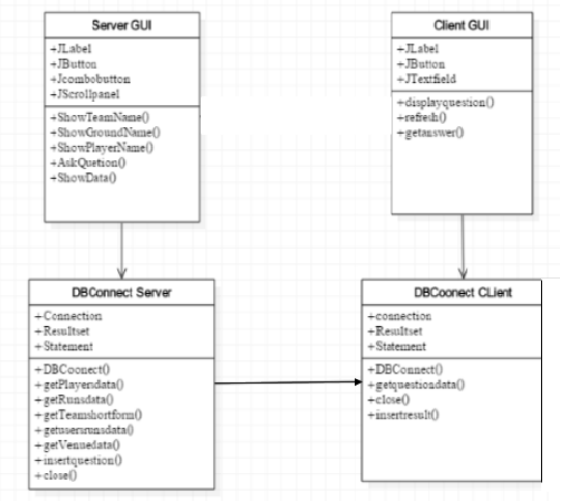

Fig.2 Class diagram

B. Data Flow Diagram

A data flow diagram is a graphical representation of the flow of data of CSPS. It helps us to understand the flow of system.

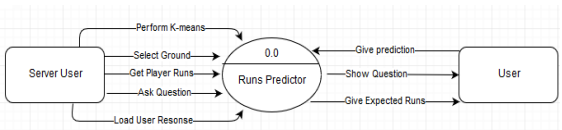

Fig.3 DFD Level 0

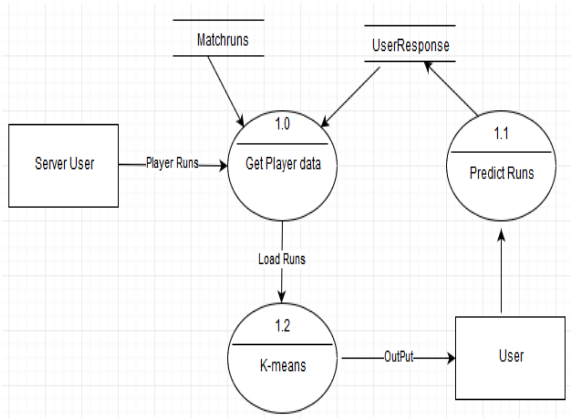

Fig.4 DFD Level 1

From the above Data Flow Diagrams we get a detailed understanding of how the flow of each function will take place. DFD level 0 gives the overview of CSPS. Whereas DFD Level 1 of the different functions give a more in depth knowledge of the methods and database usage. The DFD give us a good visualization of the steps involved in functions of the user and the database that is connected to the various methods.

III. TECHNOLOGIES USED

A. JAVA

Java programming language is concise which makes it easy to use and learn. Java Virtual Machine (JVM) enables java to be executed in any environment and platform making it portable. Web applications and applets can be accessed in a secure way using Java. It is object oriented and supports multithreading. Java supports cross platform optimized code called bytecode which are faster to execute. Hence it gives high performance.

B. MySQL

MySQL is an open source database management system suitable for relational data. MySQL is a popular choice of database for use in web applications. It has high availability and can also run on cloud computing platforms. Administration of MySQL is handled with the use of phpMyAdmin which is a free and open source tool for web browser.

IV. IMPLEMENTATION

For the purpose of predicting score made by a batsman we make use of clustering algorithm called k-means. It is a simple but effective algorithm. K-means clustering aims to partition n observations into k clusters in which every observation is put into the cluster which has the nearest mean, serving as a prototype of the cluster. The algorithm is as follows:

Step 1: Begin with a decision on the value of k being the number of clusters.

Step 2: Put any initial partition that classifies the data into k clusters. You may assign the training samples randomly or systematically as the following:

- Take the first k training sample as single-element clusters.
- Assign each of the remaining (N-k) training sample to the cluster with the nearest centroid. After each assignment, recompute the centroid of the gaining cluster.

Step 3: Take every sample in the sequence, compute its distance from centroid of each of the clusters. If sample is not in the cluster with the closest centroid currently, switch this sample to that cluster and update the centroid of the cluster accepting the new sample and the cluster losing the sample.

Step 4: Repeat step 3 until convergence is achieved, that is until a pass through the training sample causes no new assignments.

V. RESULTS

A. SERVER SIDE

The system starts with providing drop-down options to select particular IPL Team. The list of players gets populated automatically based on team selection. There is also an option to select venue of match. On clicking fetch button, pre-stored data from database is retrieved and divided into two clusters based on random cluster center assumptions by the system.

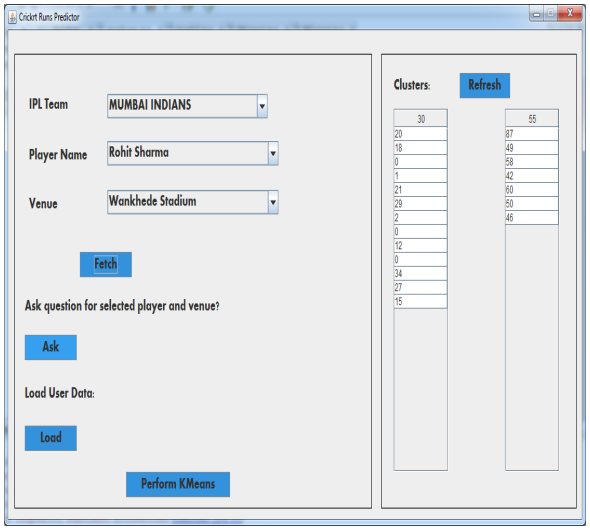

Fig.5 Server Side UI

B. PROMPT TO USER

The next step of the system is to ask the user to enter an estimated score for the respective player.

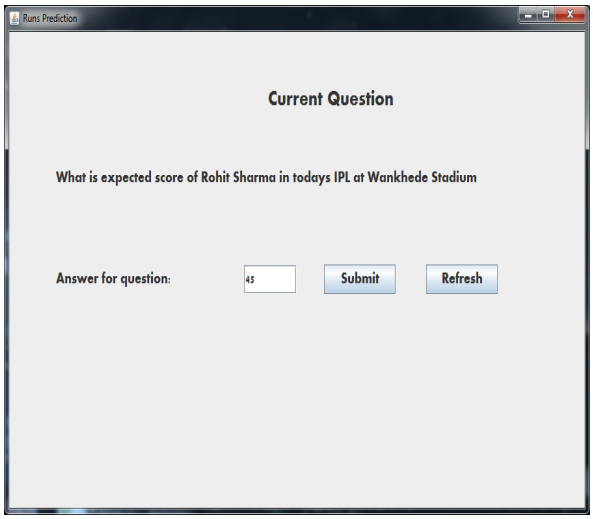

Fig.6 Prompt to user

C. PERFORMING KMEANS

Estimated data collected from multiple users is thus stored in database. On clicking Load button this data is retrieved and clustered along with past data.

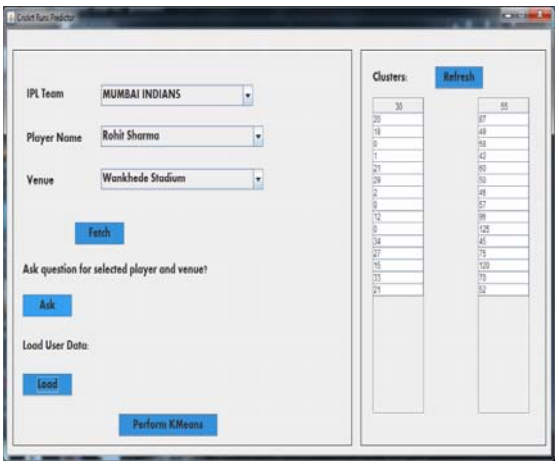

Fig.7 K-means algorithm

D. PREDICTED SCORE

After the final iteration, the cluster which contains maximum number of entries, both past as well as expected data, is selected. The cluster center of that particular cluster is the expected score and also result of the system.

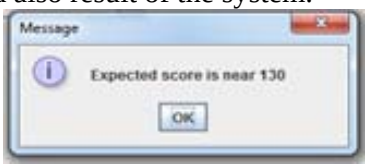

Fig.8 Result

VI. CONCLUSION

This system is essential for making strategic decisions. It is a holistic approach as it takes in current input from user. The database maintained is updated on every prediction and hence is up-to-date. The system work efficiently with a massive dataset of two thousand rows. Our system focuses on only the performance of the player and is very quick at processing due to clustering of data. Currently the system processes and predicts data for IPL held on Indian cricket grounds. In future, the system can be upgraded to encompass ODI and test match formats as well as on grounds around the world.

REFERENCES

[1] Dost Muhammad Khan, Nawaz Mohamudally
"An Agent Oriented Approach for Implementation of the Range Method of Initial Centroids in K-Means Clustering Data Mining Algorithm" International Journal of Information Processing and Management
Volume 1, Number 1, July 2010

[2] Anil K. Jain
"Data Clustering: 50 Years beyond K-Means" Department of Computer Science & Engineering Michigan State University

[3] S. Jatimiko, R. Refianti, A.B. Mutiara, R. Waryati
"Analysis Data of Student's GPA and Travelling time using Clustering Algorithm affinity propagation and k-means"
Journal of Theoretical and Applied Information Technology
30th June 2012. Vol. 40 No.2

[4] Vignesh Veppur Sankaranarayanan, Junaed Sattar, Laks V. S. Lakshmanan "Auto-play: A Data Mining Approach to ODI Cricket Simulation and Prediction" Department of Computer Science University of British Columbia

[5]Tapas Kanungo, David M. Mount, Nathan S. Netanyahu, Christine D. Piatko, Ruth Silverman, Angela Y. Wu "An Efficient k-Means Clustering Algorithm: Analysis and Implementation" IEEE TRANSACTIONS ON PATTERN ANALYSIS AND MACHINE INTELLIGENCE
VOL. 24, NO. 7, JULY 2002