International Journal of Physical Education, Sports and Health 2019; 6(3): 238-241

P-ISSN: 2394-1685
E-ISSN: 2394-1693
Impact Factor (ISRA): 5.38
IJPESH 2019; 6(3): 238-241

www.kheljournal.com
Received: 14-03-2019
Accepted: 18-04-2019

Dr. K Bijukumar
Associate Professor, Department of Physical Education, VTM NSS College, Dhanuvachapuram, Kerala, India

Correspondence
Dr. K Bijukumar
Associate Professor, Department of Physical Education, VTM NSS College, Dhanuvachapuram, Kerala, India

Artificial intelligence to predict sports performance

Dr. K Bijukumar

Abstract
This article presents the use of artificial intelligence for prediction of sporting out comes. This model is based on the historical data (data mining) collected from the teams as wells as individuals participating in the competition. Mathematical and statistical models are used in this, which is verified by a domain expert. Classification approach and Neural.
Network Approach are being explored in this article which are considered to be the most accurate prediction models in sports performance. A sample case study is also examined to understand more on the concept of prediction and artificial intelligence. This strategy is very useful for viewers, sponsors, commentators, sports analysts and team strategists.

Keywords: Artificial intelligence, performance prediction, data mining, classification approach, neural network approach

Introduction
This is a predictive analysis model for predicting sports performance. This predictive analysis strategy would be very useful for viewers, sponsors, and team strategists. This would also give insights to various sports analysts and commentators about the features that play a crucial role in the statistical analysis. This model is based on the historical data collected from the participating teams in the tournament, player performance indicators and opposition information. In addition, testing on the recent tournament data and measure the accuracy of predictions. This model can also be expanded according to the nature and requirement of the tournament once the final squads are announced. This model is based on the players who are participating with respect to the teams/squads in the recently concluded five or six tournaments.
One of the vital applications in sport that requires good predictive accuracy is match result prediction. Traditionally, the results of the matches are predicted using mathematical and statistical models that are often verified by a domain expert. Due to the specific nature of match-related features to different sports, results across different studies in this application can generally not be compared directly. Despite the increasing use of ML models for sport prediction, more accurate models are needed. This is due to the high volumes of betting on sport, and for sport managers seeking useful knowledge for modelling future matching strategies. Therefore, ML seems an appropriate methodology for sport prediction since it generates predictive models that can predict match results using predefined features in a historical dataset.

Data mining
Data Mining encompasses tools and techniques for the "extraction or 'mining' of knowledge from large amounts of data" (Han & Kamber, 2001). It is about finding patterns and relationships within data that can possibly result in new knowledge. Furthermore, these relationships can also result in predictors of future outcomes. The importance of data mining has been established for business applications, criminal investigations, bio-medicine, and more recently counter-terrorism. Data mining can be applied wherever there is an abundance of data available for and in need of analysis. According to previous performance, it makes prediction on performance that to be generated, it is generally used to major sporting events, predict future sports competitive levels is particular more important for athlete, so the sports

performance prediction becomes more and more important, but there are many kinds of modern prediction methods, from which neural network is more popular in contemporary prediction and analysis aspect.

Data mining presents building intelligent descriptive and predictive models incorporating the relationship between the description of a situation and a result related to the situation. Data mining methodology, techniques and tools help to develop a knowledge based system that can assist sport trainers in decision making. The data mining system is classified on the basis of functionalities. The major data mining functionalities are classification, prediction, clustering, association rule mining, and characterisation.

For example predicting the cricket world cup results, which is going to take place in 2019, utilize the data collected from every men's cricket world cup. From 1975 to the present, there have been 11 world cups (1975, 1979, 1983, 1987, 1992, 1996, 1999, 2003, 2007, 2011 and 2015) played so far. One thing to be noticed is that until 1983 world cup, each team played 60 overs each whereas from 1987 onwards, 50 overs. Also, run scoring has increased incredibly over the last few years that will be considered in our features as well.

In case of a team selection we can classify cricketers into three categories - performer, moderate and failure. We can collect data on cumulative player performance from 2007 onwards until the 2018 -2019 season. The neural network models will be progressively trained and tested using four sets of data. The trained neural network models will be applied to generate a forecast of the cricketer's near term performance. Based on the ratings generated and by applying heuristic rules we can recommend cricketers to be included in the World Cup 2019.

Features

While a great deal of sports data can be effective in data mining, lack of a general dataset forces the researchers to collect the required data, either manually or automatically, from valid sports websites. After data collection and adding new features to the existing ones, the accuracy and speed of predictions will depend on proper manual or automatic selection of the most significant, highly correlated features.

All these features except the number of ICC trophies won for the last 12 years is based solely on One-Day International (ODI) records. All the individual features are converted to a team statistic by taking the overall mean. Certain features are provided with a description of recent which basically means the period from 2015 world cup to present. Some features were also selected based on the location of the upcoming World Cup.

Table 1: The feature and category description

Feature Number	Category	Feature Description
1	Team	Overall Win-Loss Ratio
2	Team	Win-Loss Ratio: Recent
3	Individual	Career Batting Average
4	Individual	Career Batting Strike Rate
5	Individual	Number of 100s scored in Total
6	Individual	Number of 50s scored in Total
7	Individual	Number of Boundaries scored in Total
8	Individual	Career Bowling Average
9	Individual	Career Bowling Strike Rate
10	Individual	Career Bowling Economy Rate
11	Individual	Number of wickets taken per innings > 2
12	Individual	Batting Average: Recent
13	Individual	Batting Strike Rate: Recent
14	Individual	Number of 100s: Recent
15	Individual	Number of 50s: Recent
16	Individual	Number of Boundaries: Recent
17	Individual	Bowling Average: Recent
18	Individual	Bowling Strike Rate: Recent
19	Individual	Bowling Economy Rate: Recent
20	Individual	Number of wickets taken per innings > 2: Recent
21	Individual	Number of ODI' s played
22	Individual	Number of World Cup Matches Played Before
23	Individual	Age
24	Team	Consolidated Average of Opening Batsmen in the Squad
25	Team	Consolidated Average of Middle Order Batsmen in the Squad
26	Team	Consolidated Batting and Bowling Averages of all-rounders in the Squad
27	Team	Consolidated Bowling Average of Spinners in the Squad
33	Team	Death Batting Strike Rate: Recent
34	Team	Death Batting Number of Boundaries: Recent
35	Team	Powerplay Bowling Average: Recent
36	Team	Powerplay Bowling Strike Rate: Recent
37	Team	Powerplay Bowling Economy Rate: Recent
38	Team	Death Bowling Average: Recent
39	Team	Death Bowling Strike Rate: Recent
40	Team	Death Bowling Number of Boundaries: Recent
41	Individual	Batting Average at World Cup location
42	Individual	Batting Strike rate at World Cup location
43	Individual	Number of 100s scored at World Cup location
44	Individual	Number of 50s scored at World Cup location

45	Individual	Bowling Average at World Cup location
46	Individual	Bowling Strike rate at World Cup location
47	Individual	Bowling Economy rate at World Cup location
48	Team	Win-loss ratio at World Cup location
49	Team	Number of ICC Trophies in last 12 years
50	Miscellaneous	Average runs scored in at World Cup location: Recent
51	Miscellaneous	Weather conditions: Recent
52	Team	Number of player affected by injury at Present
53	Team	Win-loss ratio of the Captain
54	Team	Number of bowler variations in the squad
55	Miscellaneous	Location of World cup
56	Team	Win/loss ratio while defending
57	Team	Win/loss ratio while chasing
58	Individual	Batting average in Big tournaments
59	Individual	Number of 50s scored in Big tournaments
60	Individual	Bowling average in Big tournaments
61	Individual	Number of wickets taken per innings > 2 in Big tournaments
62	Team	Ratio of number of matches won after winning the toss

In this article two different approaches for our predictive analysis are presented. A classifier approach and a neural network approach with hidden layers. Classifier approach would help us to identify the pattern whereas neural network would help us to identify the weights allocated for each feature.

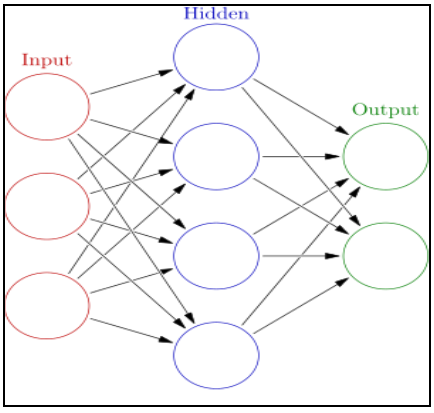
A. Classification Approach

The framework of ensemble classifier systems is established by combining numerous basic classifiers together to reduce the variance caused by a single training set and more expressive concept in classification than a single classifier. We utilize the 8 basic classifiers for this study. The number of basic classifiers are selected based on the leave one out fold validation of the training data. Ensemble classifier has proven to be effective for predictive analysis, hence we adopted the same for this research.

Logistic regression is a well-known tool for classification problems. Like linear regression, logistic regression depends on linear combination of features which is then mapped to a value between zero and one by the logistic function. Therefore, dependent variables should have a continuous value which is in turn a function of the odds of the events. Logistic regression involves two stages: first, estimating the odds of characteristics of each group, and second, determining cut-off points and categorizing the features accordingly.

B. Neural Network Approach

In the neural network approach, we utilize 12 hidden layers for this study. The number of hidden layers was chosen based on leave one out validation of the training data. Gradient descent back propagation method is utilized.



Sample Neural Network

As Artificial Neural Network (ANN) mimic a biological neural network, they consist of a number of interconnected neurons (processing elements) in particular layers. Neurons in each layer have weighted connections with neurons in the previous and next layer. An ANN comprises at least one input layer, one output layer, and some hidden layers if necessary. During the learning phase, an ANN processes a training dataset and seeks proper weights for the network to correctly classify all training data

Decision trees are powerful and common classification and prediction tools. In contrast to ANN which only provide the final prediction and keep the path hidden in the network, decision trees result in a set of rules that clarifies the final prediction. A decision tree poses questions about data features and classifies the data accordingly. Each question is the subset of a node and each interior node points to a child node for each possible answer to the question. Therefore, the posed questions will form a hierarchy and finally a tree.

A sample case study taken from Sportskeeda, a sports news website

2015 Cricket World Cup

At first, we validate our approach by estimating the probabilities of winning the World Cup of these 10 teams for the 2015 world cup and match with the actual 2015 world cup results. We estimate the probabilities based on the data collected from 1975-2011 world cups for team data and from 2007 to 2018 for individual performance data. Despite the 2015 world cup being played among 14 different countries, we focus on the results of these 10 teams. Table -2 appended below lists the probabilities for the 2015 world cup based on both classifier and neural network approaches along with the actual result.

Table 2: Classifier and Neural Network Predictions for 2015 Cricket World Cup

Team	Classifier Probability	Neural Network Prediction	Actual Result
Afghanistan	1%	1%	Group Stage
Australia	28.5%	25.1%	Winners
Bangladesh	5%	2%	Quarter Finals
England	6.5%	5%	Group Stage
India	16.0%	12.4%	Semi Finals
New Zealand	12.5%	16.1%	Finalists
Pakistan	9%	11%	Quarter Finals
South Africa	12.0%	15.4%	Semi Finals
Sri Lanka	7%	9%	Quarter Finals
West Indies	2.5%	3%	Quarter Finals

2019 Men's cricket world cup

Now, we predict the 2019 world cup results based on the data collected from 1975-2015 world cups and individual performance data from 2007 to 2018. Table-3 appended below presents the probabilities based on the classification approaches based on the data collected until a specific date (July 2018).

Table 3: Classifier and Neural Network Predictions for 2019 Cricket World Cup

Team	Classifier Probability	Neural Network Prediction
Afghanis-tan	0.5%	1%
Australia	10%	6.1%
Bangladesh	3%	2%
England	21.0%	18.8%
India	18.0%	20.8%
New Zealand	14.5%	15.7%
Pakistan	17.5%	19.0%
South Africa	8.5%	10.6%
Sri Lanka	3.5%	3%
West Indies	3.5%	3%

Conclusion

Defending champions Australia gets a relatively lower probability due to their poor performance at England recently. Top 2 contenders for the world cup according to the classifier are England and India. Classifier approach predicts that the England cricket team has the highest probability of winning. This could be because the previous world cups were won by host nations and due to excellent record by their team in last few years at their home ground (world cup location). Meanwhile, the Neural Network approach predicts India and Pakistan as the top two contenders, this could be due to their excellent performance in the ICC champion's trophy 2017. Also, India won the champions trophy 2013 and reached the finals in 2017 which happened in England, hence India gets a relatively higher probability by both Neural Network and Classifier approaches. However, their middle order performance is considerably low when compared to England, hence Classifier predicts England as the winner whereas the Neural Network still believes India has higher chances.

Some of the other interesting notes about this research: If team India plays without the so called super stars in this upcoming world cup, India's chances of winning the world cup reduces to 2% whereas if South Africa plays with AB de Villiers, their chances go up to 18%.

References

1. Abdelhamid N, Ayesh A, Thabtah F, Ahmadi S, HadiMAC W. A multiclass associative classification algorithm J. Info. Know. Mgmt. (JIKM). 2012; 11(2):125001-1-1250011-10.
2. https://www.sportskeeda.com/cricket/predictions-for-cricket-world-cup-2019-trophy-winners.
3. ByunghoMin Jinhyuck KimChongyoun Choe Hyeonsang Eom R.I. (Bob) McKay A compound framework for sports results prediction: A football case study.
4. Mohamed H, Negm A, Zahran M, Saavedra OC. Assessment of artificial neural network for bathymetry estimation using High Resolution Satellite imagery in Shallow Lakes: case study El Burullus Lake, in: International Water Technology Conference, 2015.
5. Qabajeh I, Thabtah F, Chiclana F. A dynamic rule-induction method for classification in data mining. J. Manage. Anal. 2015; 2(3):233-253.
6. Grossberg S. Nonlinear neural networks: principles, mechanisms, and architectures Neural Networks. 1988, 17-61.