Predictive Analysis of Sports Data Using Google Prediction API

V.Showmiyaa[1], S.Jambunathan[2]
[1, 2] Dept of Computer Science
[1, 2] Sri Krishna Arts and Science College, Coimbatore -641008

Abstract- *Sports events bring in a lot of people to watch the games. In india especially cricket. Cricket and now others sports are contributing to the Indian Economy in a high scale. The constructing an application that can extract process and analyze the data and then give the output to other people who plan on setting up any sports events at a given time. The data is collected from sources that supply with climatic data (Temperature data). Choosing the right place to conduct a sports event is the most critical assignments for any game especially for cricket in India. The factors like venue of the match, pitch, weather conditions first batting or fielding all play a vital role in predicting the winner of the match.[3] Suitable models are necessary to predict and data mining makes it possible to extract required information from the data files. The data's were analyzed using feed forward technique in neural networks. It is the most often used technique in analyzing nonlinear sports data. As a final result, we concluded*

Keywords- *Google prediction API, Neural network, sports data*

I. INTRODUCTION

Data mining and all its advantages, sports associations almost solely used knowledge and expertise of people responsible for scouting. Amount of data collected was increasing, so the aim was to find more practical methods to extract knowledge from raw data. In beginning, this led to hiring statistician who would enable better performances measurements for a given organization and therefore more correct decision-making. The next step was to find even more practical methods, that is, to start using data mining technique Prediction system works on the principles of machine learning. There are two types of machine learning namely supervised machine learning and unsupervised machine learning. In supervised machine learning we must train the machine by providing huge data sets and the outcomes. In this paper one such prediction methods is introduced [1] which is used to make predictions of the outcome of a cricket match using Google Prediction API.

Google API is a black box prediction technique. It is a form of supervised learning and hence it is required to provide huge data and train the models. Google prediction APIs make use of regression algorithms the numerical predictions have to be made. And Classifiers when the target output can assume only a limited set of values, either numbers or strings, based on the application content. This API can only account for relatable data. If the attributes are not related to one other, then a correct probability curve will not be drawn. By providing a CSV file of previous cricket matches and using appropriate queries to extract the required data and train the model, predictions can be made.

Need For Neural Network

It differs from for instance regression within the approach that regression models are heaps less complicated. Contemplate a neural network with just one output node, input nodes and hidden layers. If you're taking away the hidden layers of a neural network, you're left with solely input nodes and output nodes. The network then tries to predict the output nodes by exploitation solely the input nodes. This can be precisely the approach linear and supplying regression models attempt to predict values.

The hidden layer of the Neural Network is what makes the network smarter and additional correct than ancient prognostic tools, because it learns the approach a personality's can, it remembers past connections in information by incorporating this information within the algorithmic rule. Prognostic analytics have various doable applications, like analytical client relationship management, the detection of fraud in accounting, [10] project risk management or cross-selling in retail. Because the prices of computation power are getting cheaper because of Moore's law and can be cheaper within the future, prognostic analytics can become a additional viable tool.

II. RELATED WORK

Basket Ball Game Using Neural Network

Analysis of Basketball Games Using Neural Networks Data mining is a technology in data analysis with rising application in sports. Basketball is one of most popular sports. Due to this dynamic a large number of events happen during a game. Basketball statisticians have task to note as many of these events as possible, in order to provide their analysis. The author Ivankovic, M. Rackovi we used data from the First B basketball league for men in Serbia, for seasons 2005/06, 2006/07, 2007/08, 2008/09 and 2009/2010. During the five seasons total of 890 games were played. Data were collected for individual players, so it was necessary to adapt this in order to show statistics for a whole team. This data was analyzed using feed forward technique in neural networks which is the most often used technique in analyzing nonlinear sports data. As a final result, we concluded that the most important elements in basketball are two-point shots of under the hoop and defensive rebound.

Competition Results Prediction Model Based on Athlete Ability

Its offers formalized description of outcome prediction for sports competitions. It proposes a novel match prediction model and team model based study on existing technology. The team model adopts improved Bayesian network. It adopt an EM algorithm to learn player capability of each team to evaluate the ability value of player according to integrate match situation. [2]Then the scoring probability of team score expectation is computed combined with the case of players in presence. At last the score estimation of round prediction to match model is acquired. Zhao YijieDuring study of team model coefficients, we assume it obeys distribution of Logit and Probit and compare their effects. Experimental analysis adopts three seasons of real dataset in CBA league, and our model is compared to Bayesian network model. [9]The results show our model is more accurate than existing model in match process prediction the CBA regular season dataset of 10-14 season of, using part of the data set for training, and part of the data sets to do the test, to get score accuracy statistics of each attack round, compared to existing Trueskill algorithms. It is found that our model of round attack score prediction accuracy is higher than that of the algorithm Trueskill

Neural Network Model For Forecasting Education And Sports

Recurrent neural network is a network which provides feedback connections. The network is truly to have a more powerful approach than the typical neural network for learning given data. The current research is achieved to apply the simplest recurrent neural network model namely the

Elman recurrent neural network (ERNN) model to the consumer price index (CPI) of education, recreation, and sports data in Yogyakarta. [5] The pattern of CPI data can be categorized as a function of time period which tends to move upwards when the time period is increased, and jump at some points of the time period. This pattern is identified as similar to the pattern resulted by the function of the truncated polynomial spline regression model (TPSR). Wutsqa and Rosita Kusumawati Hence, this research considered ERNN model which the inputs such as in the TPSR model were established by taking into account the location of the knot or jump points. In addition, the ERNN model with a single input, a time period was also generated. The results demonstrated that the two models have high accuracy both in training and testing data. More importantly, it was found that the first model is more appropriate than the second one in testing data.

Neural Network Modeling For Sports Data Classification

The extraction of purposeful data by exploitation artificial neural network wherever the main target is upon developing new insights for sports performance and supporting deciding is crucial to achieve success. The aim of this text is to form a theoretical framework and structurally connect the sports and multi-layer artificial neural network domains through (a) describing sports as a fancy socio-technical system (b) identification of pre-processing scheme for classification (c) feature choice by exploitation data-driven valued tolerance magnitude relation methodology (d) style prophetical system model of sports performance employing a back propagation neural network. [8] This might permit distinguishing classifying, associate degreed statement performance levels for an enlarged information set. The author planned IvarsNamatevs and LudmilaAleksejeva system functions by exploitation sports dataset that consists of 4 sub-datasets revealing the coaching and competition objectives for prime performance. [4] Missing values of the sports dataset are handled by the mean imputation methodology. Special attention is given to information reduction supported the novel data-driven valued tolerance relation methodology for feature choice. The classifier of the sports performance analysis and deciding are going to be developed by employing a two step back propagation neural network. Finally, the classification accuracy and sensitivity are going to be evaluated.

Neural Network Predicting the Racing Time Of Cross Country Skiers

The learning methods and survey-based data for predicting the racing times of cross-country skiers. Specifically three popular types of artificial neural networks (ANN) including Multilayer Feed-Forward Artificial Neural

Network (MFANN), General Regression Neural Network (GRNN) and Radial Basis Function Neural Network (RBFNN) have been used for model development.[7] The utilized dataset is made up of samples related to 370 cross-country skiers with heterogeneous properties and include physiological variables such as gender, age, height, weight and body mass index (BMI) along with a rich set of survey based data. The result reveal that in general the three ANN-based method show comparable performance, and can be categorized as feasible tools to predict the racing time of cross-country skiers [6]with acceptable error rates. Furthermore, significant advantages such as the non-exercise-based usage and the applicability to a broader range of cross-country skiers make the prediction models proposed in this study easy-to-use and more valuable.
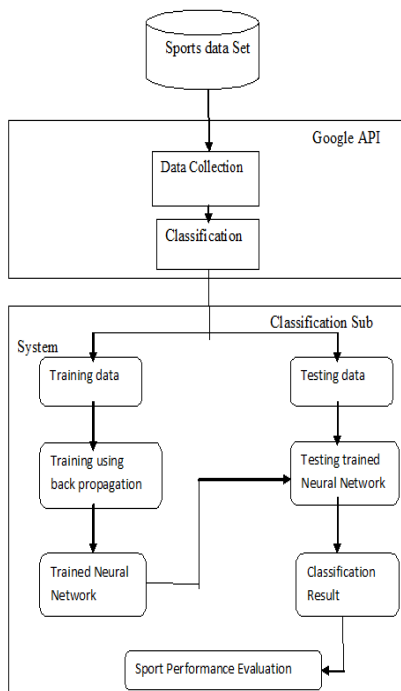
III. RESEARCH METHODOLOGY

I.

Figure 1: Predictive Analysis Neural Network method

Data Collection

The data collection is the initial step. Data sets of previous cricket matches are available in many websites and we collected the data from www.cricksheet.org and www.espncricinfo.com. The files are in CSV format and this is the required format for Google API. The data set will have all the information of a cricket match such as team names, gender, season, date, venue, city, toss winner, toss decision,

player of the match, umpires, TV umpire, winner, winner runs, total score and ball to ball information .

Classification

Classification technique is the one that does most of the job by identifying the type of the class an instance belongs to in machine learning. As Google Prediction API is a supervised learning, classifiers are used by it to train data models by giving the data of already correctly identified instances. By analyzing this training data set, the class type of newer instances is predicted.

Neural Network

Neural network method is not commonly used for data mining tasks because they often produce incomprehensible models and require long training times. However, using two approaches of neural network learning algorithms: (a) rule extraction and (b) easy-to-understand networks, one would be able to produce comprehensible models that do not require

The excessive training times for several decades researchers from many scientific disciplines areas have been putting their efforts for designing Neural Networks predictive models considering the following problems pattern classification, clustering and categorization, function approximation, optimization, content-addressable memory, and control Neural Networks represent one of the most successful identification techniques used to model nonlinear dynamics, complexity and randomness of systems. NNs represent a modern branch of automatic control theory, which has existed for several years and suggests an alternative solution to this problem.

NNs have been effectively used for approximating complex nonlinear functions. However, they cannot cope well with feature interaction. NNs are treated as black box learning and it is difficult for humans to understand or interpret the classification explicitly.

IV. RESULT AND DISCUSSION

The proposed model will give results based on the data provided previously. The better the data model is trained, the better the results will be. The outcome will be of the fixed team. So when the result is "Win", it means the fixed team wins the match, if the result is "Lose" then the fixed team will lose the match.

Comparison of Results

The training data model is separated from the testing model. After the training is done, the testing data is used to check the accuracy of the prediction system. The result of the cricket matches between India and other teams and the prediction of our prediction system. In this research paper, decision support system is developed for predicting the prediction is done based on sports database.

Also the results of the prediction can be stored and used as a reference when the actual outcome of the match is known and then the correctness percentage of the prediction can be calculated.
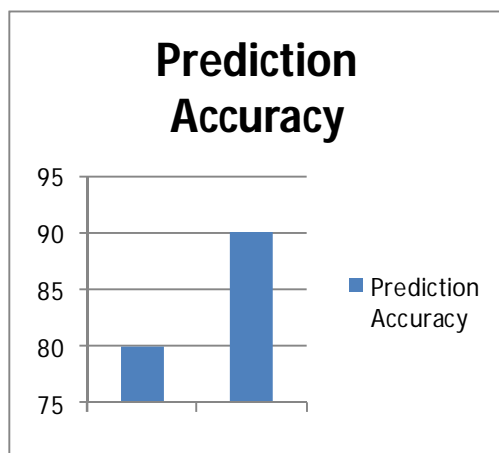
Figure 2: Accuracy of prediction in linear Regression and Neural Network

Table :1 Linear Regression VS Neural Networ

Algorithm name	Accuracy
Neural Network	90
Linear Regression	80

V. CONCLUSION AND FUTURE WORK

The aim of this study has been to point out the pertinency of socio-technical system theory and structurally connect the sports and multi-layer neural network domains. A narrative review of information mining for sports performance is given. to hide all aspects of sports performance analysis, the task needs information pre-processing and classification. Future research should empirically validate and determine the most suitable classifier for sports performance evaluation, develop a comprehensible system for complex nonlinear sports performance, consider and choose the number of neurons in hidden layer.

REFERENCES

[1] D. Jyothsna and K. Srikanth , "Analyzing and Predicting outcome of IPL Cricket Data", International Journal of Innovative Research in Science, Engineering and Technology,2019:8(4)

[2] Passi, Kalpdrum& Pandey, Niravkumar. (2018) "Predicting Players' Performance in One Day International Cricket Matches Using MachineLearning" 2018: 111-126.

[3] P. Wickramasinghe et. al, "Predicting the performance of batsmen in test cricket," Journal of Human Sport & Exercise", vol. 9, no. 4, pp. 744-751, May 2014.

[4] Tejinder Singh, Vishal Singla, Prateek Bhatia, Thapar University Punjab "Score and Winning Prediction in Cricket through Data Mining" 2015 International Conference on Soft Computing Techniques and Implementations (ICSCTI) Department of ECE, FET, MRIU, Faridabad, India, Oct 8-10, 2015

[5] H. I. Witten, and E. Frank, "Data Mining: Practical machine learning tools and techniques, 2nd Edition". *Morgan Kaufmann*, San Francisco, 2005.

[6] O. Sandbakk and H.-C. Holmberg, "A Reappraisal of Success Factors for Olympic Cross-Country Skiing," InternationalJournal of Sports Physiology and Performance, vol. 9, no. 1, pp. 117–121, 2014.

[7] L. Hultin, W. Zhu, K. Pribanic, J. Terry, and F. DP Heil, "Predicting Cross Country Skiing Performance: A Comparison Of Two Different Upper Body Power Egometers," InternationalJournal of Exercise Science: Conference Proceedings, vol. 8, no. 4, Aug. 2016.

[8] L. Grinčikaitė-Samuolė, "Attitudes of High-Level Female Sprinters Towards Factors Influencing the Training System," Baltic Journal ofSports and Health Science, vol. 100, issue 1, pp. 10–16, 2016

[9] Kirshners, I. Polaka and L. Aleksejeva, "Gastric Cancer Risk Analysis in Unhealthy Habits Data with Classification Algorithms," *InformationTechnology and Management Science*, vol. 18, pp. 97–102, Dec. 2015.

[10] R. Janssen, P. Spronck, P. Dibbets and A. Arntz, "Frequency Ratio: A Method for Dealing with Missing Values within Nearest Neighbour Search," *Journal of System Integration*, vol. 6, no. 3, 2015.