Prediction Analysis with Professional Basket Ball Teams Using Historical Data

Shruti Chopra[1], Vijay Kumar Joshi[2]

[1]Research Scholar, Department of Computer Science, Ludhiana College of Engineering and Technology, Katani, Ludhiana (Punjab), India

[2]Professor & Head Of Department, Ludhiana College of Engineering and Technology,Katani, Ludhiana(Punjab), India. Email:

Abstract: Predicting the accurate results of any game[1] is of high interest to teams, general public and the Sponsors of the game. This paper uses the NBA data todetermine the logic for predicting the accurate results of Basket Ball Matches.The developed model primarily does a feature selection using Chi Square Testand identifies twenty most significant features of the game. This is followed by creation and Implementation of Fuzzy Rules to calculate their Impact on Result.
Thereafter, we divide this data into Training and Testing Data Sets and employSVM [2] technique to obtain relativelyhigher prediction accuracy.The primary goal of this research paper is to devise out a model that gives a highly accurate prediction [3] of Basket Ball Match. As we have done a combination of Chi Square and SVM [4] Technique, the Investigation developed a Hybrid model of both these techniques and called it Hybrid Chi Square SVM Model(HCSSVM Model) for predicting high accuracy in game results.
Keywords: National Basket Ball Association, Support Vector Machine, Prediction Analysis, Chi-Square, Fuzzy Logic

I. INTRODUCTION

Prediction analysis is a field that entertain in extracting information[5] from a given dataset and use it to predict future results and further performance patterns. This technique is usually applied on past or present data whereas the predictionresult is in future. For Instance, recognizing traces after anydamage has been committed,or credit card theft, after it occurred. However, main key factor is, that closeness of results are directly proportional to principle of analysis and accuracy of supposal. Evaluating at a more granular level can be determined as Predictive analytics.Tsai[7] proved that in the field, a team's performance is directly proportional to effectiveness of three major positions (i.eGuard, Forward and Centre).Min et al.[13] provided a mechanism for forecasting Football Match resultsby making use of in-game time series, Bayesian Inference and a rule- based reaasoning. His mechanism provided high precision results.Wu[8] performed a study on BasketBall Matches held in High Schools and came to an observation that ranking of teams will always go down with the increase in personal fouls.The Ranking goes up with the acceleration in total rebounds and blocks between playing teams. Performance also goes up with increase in two-point shot hit rate.S trumbelj and Vracar[12] prepared a model for predicting basketball match result.Their results of model came out to be far accurate as compared with commonly used traditional methods.Kvan and Sokol[16] devised a combinational model using Markov chainfor predicting plus using Regressionfor forecasting basketball matches held in colleges. It proved instumentaland gave more precise outcomes as compared to other Mathematical models,existing prediction logics, tournament seeding etc.Mai[9] determined that players must reduce the personal as well as turnover mistakes known as fouls in game to win it.

Kvan and Sokol[10] devised a combinational model using Markov chainfor predicting plus using Regressionfor forecasting basketball matches held in colleges. It proved instumental
and gave more precise outcomes as compared to other Mathematical models,existing prediction logics, tournament seeding etc.Mai[14] determined that players must reduce the personal as well as turnover mistakes known as fouls in game to win it.Lee and Worthington[20] made use of data analysis to calculateaccuracy of National Basketball Association guards, their results were greatly remarkable.Chen and Sun[15] prepared a multiple-factor decision-making structure by combiningevaluation laboratory having analytical network process methodsas well as decision making attempts to decipher significant criteriaand importance of individual factor and the surrounding environmental impact on the matchLvankovic et al. [19] detertmined that with the increase in number of features, the classification accuracy is also increased.Liu[11] analysed that a higher level of Offensive capabilityproduces a higher Rank. He also claimed that Predition of Ranking can be easily calculated withthe average occurence of defensive rebounds and assits in gameCooper et al. [17] provided highly stable results to judge Basketball team's strength along with weaknesses by making use of data envelopmentstudy. There has always been a curiosity to predict results of Matches and lot ofResearch work have

been performed in Data Science to determine Models and logics that gives results of Game Win or Loss basis the Permutation and Combination of essential Game Parameters. These Models can be helpful instrengthening the teams and enhance the prospects of winning the Sport.

II. METHODOLOGY

A. Feature Selection using Chi-Square

The chi-square test is a statistical test of independence, majorly to conclude the dependency of features on final result. We can use Chi-Square to deducewhich feature holds significance in determining the result of game and which feature simply adds an overhead. We proceed by considering the target result and feature variables, that describe each sample of data,and calculate chi-square statistics between every feature variable and the target result. We decipher the dependency of variables on targetresult in form of a relationship formula. If the target result is not controlled by any specific feature variable, we can eliminate that feature. If they are dependent, the consideration of feature variable is very eminent in calculation of result.

In statisticsχ^2 test is implemented with below formula:

Two events M and N will be independent if:

$\qquad$ P(MN) = P(M)P(N)

or $\qquad$ P(M|N) = P(M) and P(N|M) = P(N)

In Feature Selection, two events are occurrence of terms and occurrence of class.

We rank terms wth respect to following quantity:

$$X^2(\mathbb{D}, t, c) = \sum_{e_t \in \{0,1\}} \sum_{e_c \in \{0,1\}} \frac{(N_{e_t e_c} - E_{e_t e_c})^2}{E_{e_t e_c}}$$

where e_t and e_c can have values 0,1 where 0 denotes that document is not in term or class respectively and 1 denotes that document is in term or class respectively .

And 'N' is the found frequency in 'D' and 'E' isthe expected frequency.

χ^2 calculates the deviation of found counts 'N' from expected counts 'E'. A higher

value of χ^2 indicates that the found and observed counts are alike. When two events are dependent, then there is direct proportionality between existence of the classand the existence of the term so it remarks the same as an essentialfeature.

B. Fuzzy Rules

These are the rules that categorize a set of values into user defined output following

some Inference Mechanisms. The Inference mechanisms follow such a waywhich helps in laying and processing the Fuzzy Rules.While designing the rules, every parameter value must be mapped to an output(should not be an empty set)in order to generate a useful output from a set of values given in the Input.

Inference Mechanisms

Inference mechanism uses the principle of fuzzy logic for synchronizing an input tooutput using fuzzy logic. It uses membership functions, logical operations and if-then rules.Mamdani and Sugeno are the common types of inference systems. They determine the outputs by their own ways.

Madani method

This method works like below:

Consider:

P1: if m is S1 and n is T1 then q is U1

P2: if m is S2 and n is T2 then q is U2

Result: q is U, where m equals m0 and n equals n0.Sugeno method

This method works like below:

Consider:

P1: if m is S1 and n is T1 then q is q1 = a1m1+b1n1

P2: if m is S2 and n is T2 then q is q2 = a2m2+b2n2

Result: q0, where m equals m0 and n equals n0.

C. *Support Vector Machines*

Support vector machines (SVMs) can be defined as acombination of associatedsteps that investigate data and identify results, utilised for grouping(machine learning)and regression analysis.[1]The standard SVM can be defined as non-probabilistic binary classifier or binary linear classifier. For each given parameter, it depicts the parameter is a memberof which two probable classes. Since the SVM is a classifier, when a group of training examples, each classified intoeither of the category, the SVM training algorithm build a model whichdepictsifthe new example falls into first category or the second. Graphically, SVM model represents points in space that areidentified into separate categories. Similarly, new examples are then represented on same spaceand their category is identified depending on their position. Along withdelivering linear classification, a non-linear classification can also be delivered bymaking use of the kernel trick, whichinternally maps their parametersamongst high-dimensional feature spaces.

III. A DIAGRAMMATICAL FLOW AND ITS EXPERIMENTAL RESULTS

The flowchart of developed model is shown in Fig 1 along with explanation of each step. First step starts with the Collection of Raw Data that has been done by extracting the data from NBA [6] Website. This data has been tuned finely and bought to a synchronized level so that the data types and values can be identified and processed by the system's logical formulas smoothly without changing any values. Once the data becomes recognizable by formulas, we begin the second step, which is, feature selection procedure, following Chi Square Algorithm to identify all the parameters of Basket Ball game that works to alter the result of match. This step is the most remarkable step as it decides which all parameters have to be taken into considerations. We identified 20 best features of the game from of 33 features and discarded the unimportantones.

The unimportant features are those, whose presence does not make any difference in final result. The third step is to generate rules to check the impact of these parameters on final result. Fuzzy rules have been generated on which, the data set is fed, and the rules calculate the impact of these parameter on result. Now these obtained results have been fed to a well trained SVM model where the entire data set was classified into training and testing data sets.
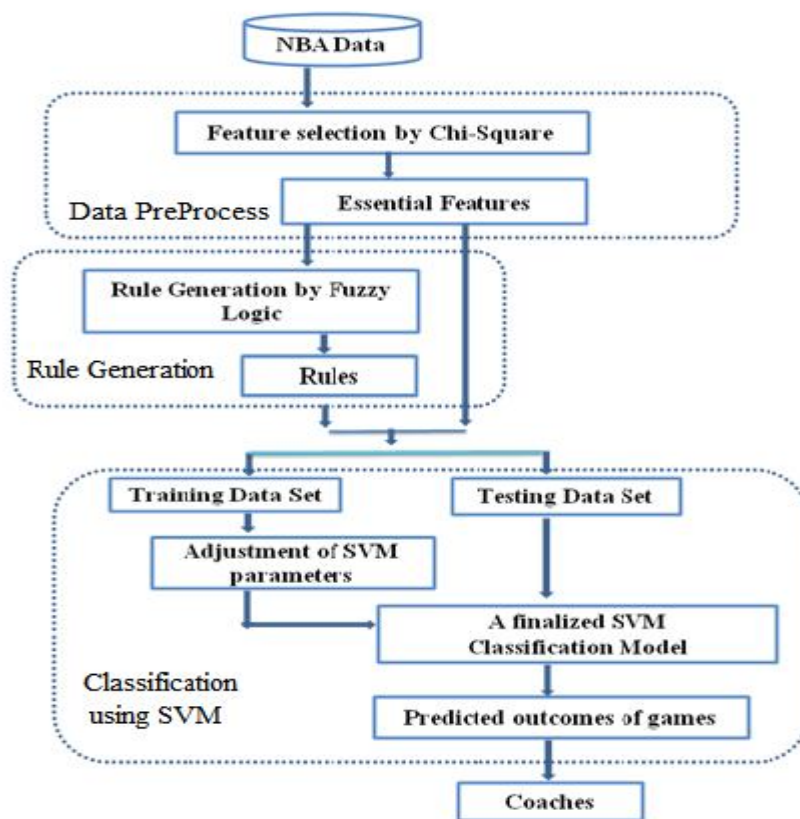

Figure 1

A. Experimental Results

Table I: SELECTED ATTRIBUTES AFTER FEATURE SELECTION

S. No	Abbreviation	Description
1	FG	Field Goals
2	3P	Three-point field goals
3	FT	Free throws
4	FTA	Free throw attempts
5	DRB	Defensive rebounds
6	TRB	Total rebounds
7	AST	Assists
8	STL	Steals
9	BLK	Blocks
10	TOV	Turnovers
11	PF	Personal fouls
12	PTS	Points
13	ORB%	Offensive Rebound Percentage
14	TRB%	Total Rebound Percentage
15	AST%	Assist Percentage
16	STL%	Steal Percentage
17	BLK%	Block Percentage
18	TOV%	Turnover Percentage
19	ORtg	Offensive Rating
20	DRtg	Defensive Rating

Fuzzy Rules to derive better performance:

if value of scaled points=0 then significant array=0

if value of scaled points is greater than 0 and less than 0.25 then significant array=0.1

if value of scaled points is greater than 0.25 and less than 0.50 then significant array=0.2

if value of scaled points is greater than 0.50 and less than 0.75 then significant array=0.3

else significant array=0.4

With the above approach a relatively better performance result has been measured as highlighted in Table IV. The HCSSVM model proved to be very constructive in measuring the accuracy of Basket Ball Result.

B. Experimental Comparison in results of CFS and Chi Square Algorithm:

table ii: experimental results of testing accuracy with chi square

With Chi Square	Testing Accuracy %
CS1	88.7
CS2	89.95
CS3	88.22
CS4	89.62
CS5	88.4

table iii: experimental results of testing accuracy with cfs

With CFS	Testing Accuracy %
CS1	85
CS2	87.5
CS3	86.25
CS4	86.25
CS5	81.25

Table Iv: Comparison In Results Of Cfs And Chi-Square Algorithm

Algorithm used	Average Testing Accuracy
CFS	85.25%
Chi Square	89.074%
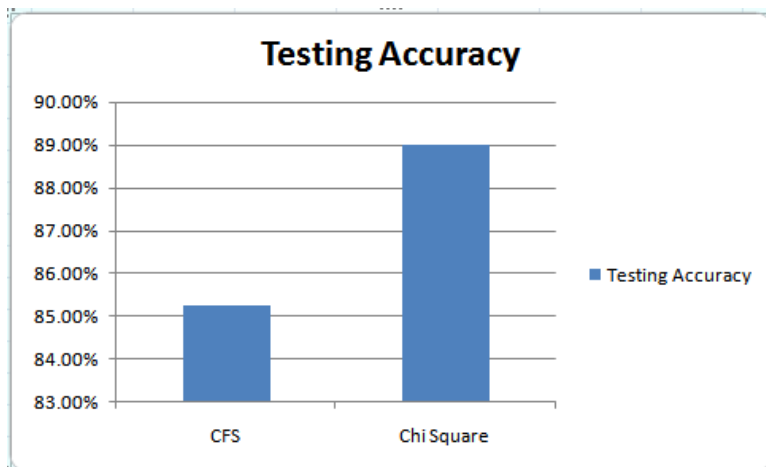
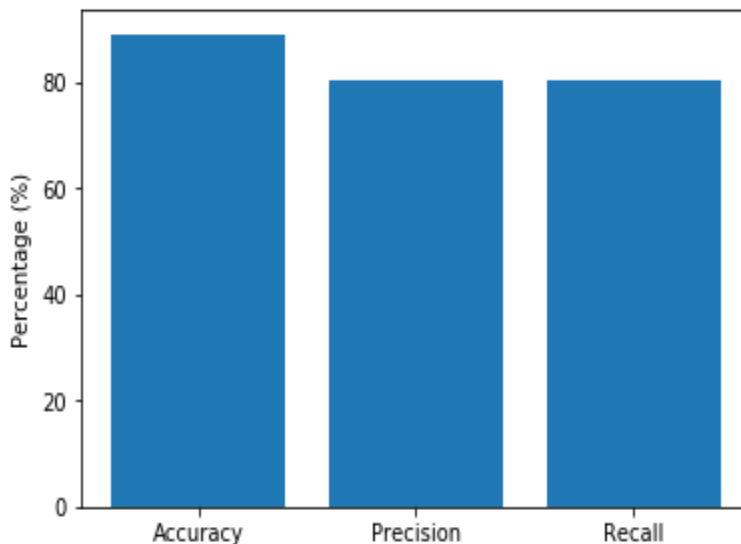Graph 1 :A graph comparing the testing Accuracy of CFS and Chi Square Method
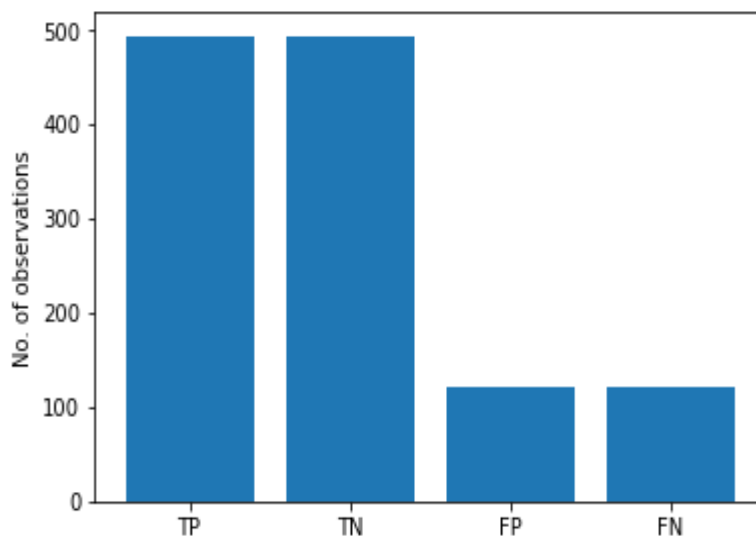
Graph 2 :A graph measuring the number of True/False Observations in Postive and Negative Scenarios

TP - True Positive
TN - True Negative
FP - False Positive
FN- False Negative

00 - True Positive Results
01 - False Negative Results
10 - True Negative Results
11 - True Negative Results

Graph 3 : Measuring the Percentage of Accuracy, Precision and Recall obtained with HCSSVM approach

IV. CONCLUSION

Examining basketball match outcome is a stimulating job for the coaches and management of

basketball matches. Our study devised a HCSSVM model whichanalyzes the match's results in NBA. In contrast with last studies in determining the outcomes of basketballmatches, the prediction closeness retrieved from HCSSVM model is marginallylarge. HCSSVM model works on backward and forward reasoning functions.The forward reasoning is used for determining outcomeof games.The backward reasoning helps in giving the recommendations forcoaches, to enhance which play strategyto follow, so as towin game.With the Importance of associated features, coaches can easilydepict the chance of winning the match, and use given methodsor a rotation policy for a better outcome. Hence, the HCSSVM approach isbetter as well as promising substitutetoexamine basketballgame results. More researchcan be performed to judge which basketball parameters can be utilized toenhance the performance of Basketball match. Along with this, by elaborating theHCSSVM approach, as an intelligent information system, eminentfeatures can be picked and used for training players, deciding players' salaries and facilitation of basketballteams. For Instance, enlisting which all features and to how muchextent it requires improvement for a win, could be the future scope of this studyAdditionally, the developedHCSSVM approach is solely predicting the win or loss ofmatch. To predict the points madein game would be an interestinginvestigation. Furthermore, researches could be performed with statistics regarding other games, like cricket, volleyballetcso as to check the flexibility ofthe HCSSVM approach.

REFERENCES

[1] M. Bruckner, C. Kanzow, and T. Scheffer, "Static prediction games foradversarial learning problems," J. Mach. Learn. Res., vol. 13, pp. 2617–2654, September 2012

[2] S. Abe, "Analysis of support vector machines," in Proc. 12th IEEEWorkshop on Neural Networks for Signal Processing, 2002, pp. 89–98

[3] RenatorAmorim Torres. Prediction of nba games based on machinelearning methods. University of Wisconsin, Madison, 2013.

[4] C. J. C. Burges and D. J. Crisp, "Uniqueness of the SVM solution," inNIPS, S. A. Solla, T. K. Leen, and K.-R. Muller, Eds. The MIT Press, 1999, pp. 223–229.

[5] Chenjie Cao. Sports data mining technology used in basketball outcomeprediction. Master's thesis, Dublin Institute of Technology, 2012.

[6] http://stats.nba.com.

[7] Tsai YC (2004) Analysis of the relationship between the techniquesof offence and defense of basketball players and the ranksof their teams in the High School Basketball League. Thesis forMaster of Graduate Institute of Physical Education NationalTaiwan Sport University, Taiwan, R.O.C.

[8] Wu SS (2005) Analysis of the offensive and defensive strategiesin the 2004 High School Basketball League Competitions. Thesisfor Master of Graduate Institute of Sport Coaching ScienceChinese Culture University, pp. 1–53

[9]   Mai YH (2004) The analysis of the statistics of Asian women's basketball championships in Sandai. Thesis for Master of GraduateInstitute of Coaching Science National Taiwan SportUniversity, Taiwan, R.O.C.

[10] Kvan P, Sokol JS (2006) A logistic regression/Markov chainmodel for NCAA basketball. Naval Res Logist 53:1–23

[11] Liu SH (2006) Analysis of the offensive and defensivenumbers in the 2005 female teams of high school basketballleague competitions. Thesis for Master of Graduate Instituteof Physical Education Taipei Municipal University ofEducation

[12] S trumbelj E, Vracar P (2012) Simulating a basketball match witha homogeneous Markov model and forecasting the outcome. Int JForecast 28:532–550

[13] Min B, Kim J, Choe C, Eom H, McKay RI (2008) A compoundframework for sports results prediction: a football case study.Knowl Based Syst 21:551–570

[14] Mai YH (2004) The analysis of the statistics of Asian women'sbasketball championships in Sandai. Thesis for Master of GraduateInstitute of Coaching Science National Taiwan SportUniversity, Taiwan, R.O.C.

[15] Chen T, Sun K-S (2012) Exploring the strategy to improve seniorcitizens' participations on recreational sports. Knowl Based Syst26:86–99

[16] Kvan P, Sokol JS (2006) A logistic regression Markov chainmodel for NCAA basketball. Naval Res Logist 53:1–22

[17] Cooper WW, Ruiz JL, Sirvent I (2009) Selecting non-zeroweights to evaluate effectiveness of basketball players with DEA.Eur J Oper Res 195:563–574

[18] Liu SH (2006) Analysis of the offensive and defensive numbers in the 2005 female teams of high school basketballleague competitions. Thesis for Master of Graduate Instituteof Physical Education Taipei Municipal University ofEducation

[19] Lvankovic Z, Rackovic M, Markoski B, Radosav D, Ivkovic M(2010) Analysis of basketball games using neural networks. In:IEEE International Symposium on Computational Intelligenceand Informatics, Budapest, Hungary (pp. 251–256)

[20] Lee BL, Worthington AC (2013) Note on the 'Linsanity' ofmeasuring the relative efficiency of National Basketball Associationguards. Appl Econ 45:4193–4202

INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

Call : 08813907089 ○ (24*7 Support on Whatsapp)