

Predicting Basketball Results Using Cascading Algorithm

Jasmin A. Caliwag
Technological Institute of the Philippines
Quezon City, Metro Manila Philippines
(+63) 911-0964

jcaliwag.it@tip.edu.ph
Reynaldo E. Castillo
Technological Institute of the Philippines
Quezon City, Metro Manila Philippines
(+63) 911-0964
rcastillo.cs@tip.edu.ph

Maria Christina R. Aragon
Technological Institute of the Philippines
Quezon City, Metro Manila Philippines
(+63) 911-0964

christina.aragon@tip.edu.ph
Ellizer Mikko S. Colantes
Technological Institute of the Philippines
Quezon City, Metro Manila Philippines
(+63) 911-0964
ellizer01colantes01@gmail.com

ABSTRACT

Anybody can guess the winners of a basket game. The question is how big the chances are in predicting the real winners. Relying only on the experts' experiences and intuition could not discover all the value and potential of the collected data. Driven by the increasing comprehensive data in sports datasets and data mining technique successfully used in different areas, sports data mining technique emerges and enables us to find hidden knowledge to impact the sports industry. A more scientific approach is needed to use for these data that are collected. Some predictors based only on winning records and some based only on statistical records of both teams. There are also predictors which use both types of data, but the accuracy of applying different individual algorithms is only ranging about 60% – 70%. To achieve better prediction rates and deal with that complexity, a lot of machine learning methods have been implemented over these data. This paper presents an improved technique for predicting basketball game results implementing cascading algorithm. The researchers combined Naïve Bayes, Four Factor Analysis, and Fuzzy Logic Algorithms to predict basketball game result in an acceptable level of 69% - 70% accuracy. The researchers tested several times using data sets from NBA game Season 2015-2016, and the cascading algorithm result manages to reach 70% prediction accuracy. The result of this system can be used to assist basketball coaches in making plans for possible team developments. Also, the forecasted results can serve as an aid in building effective gameplay.

CCS Concepts

•Applied computing ~ Forecasting •Theory of computation ~ Bayesian analysis • Computing methodologies ~ Factor analysis •Computing methodologies ~ Vagueness and fuzzy logic.

Keywords

NBA; Four Factor Analysis; Naïve Bayes Model; Fuzzy Logic; Cascading Algorithm.

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. To copy otherwise, or republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee.

ICISS'18, April 27-29, 2018, Jeju, Republic of Korea.

© 2018 Association for Computing Machinery.

ACM ISBN 978-1-4503-6421-8/18/04\$15.00

<https://doi.org/10.1145/3209914.3209921>

1. INTRODUCTION

The sport like National Basketball Association (NBA), the men's professional basketball league in America [1], is very popular in the world wherein basketball enthusiasts and different basketball organizations collected data and transform these historical data into useful knowledge or information. This information is used to track players' and teams' records to formulate the best gameplay. Through the years, a lot of data and statistics have been collected based on NBA, and each day the data become richer and detailed. Predicting the result of a basketball game can be useful to assist in making plans for possible developments. It is an aid to planning and building effective strategy in the field of basketball games. Similarly, the team's coach may be able to create better strategies for plotting a winning plan, specifically a way for the coach to enhance their decision-making activities. Basketball coaches, analysts, and organizations which support the team can use the system as a coaching tool and decide on which type of training should their team use. In many instances, predicting the outcomes of sporting events has always been a challenging and attractive work [2]. However, "unexpected" events often occur in the world of sports. Most frequently basketball enthusiasts are upset about the basketball game result and of teams doing much better or worse than previously expected. Even for those with domain expertise, it can become difficult to tell which events were unexpected, and which are not [3]. Thus, a more scientific approach is needed to predict the future winning teams, and one of the most common ways is by applying the methods of a different algorithm. One of the main takeaways from this plot is that generally, prediction accuracy for a given season hovers between 60% and 70% [4]. In this study, the researchers aim to implement the cascading algorithm that will predict the outcome of the basketball game. Specifically, it aims to implement a basketball prediction using Naïve Bayes, Four Factor Analysis, and Fuzzy Logic Algorithm; test the accuracy of the combined algorithms' result using the historical record in predicting basketball games; simulate the data gathered to predict result in NBA 2015-2016 Regular Season ; and come up with an accuracy of at least 70%. The basis of the proposed study will be in last three recent games of the teams on NBA regular season 2015-2016. Factors such as fouls, substitution, and injuries are excluded in the system. Parameter such as field goals (fg), field goals attempt (fda), 3-points made (3p), free throws (ft), free throws attempt (fta), offensive rebound (oreb), defensive rebound (dreb), turnover (tov), and venue of the games and the status of the game in every date they played are needed for computation. Data used were available on the website of ESPN(Entertainment and Sports Programming Network).

2. RESEARCH METHODOLOGY

The system architecture of the system developed mainly consists of the following components as shown in Figure 1. The users (see Figure 1) use computers in order to predict the basketball games then the computer gets all data sets needed for the calculation. The application flows starting from the analysing the records of data and giving the predicted results and suggestions of how to win the game. It also includes how the Fuzzy Logic, Four Factor Analysis, Naïve Bayes Algorithm works on the application. Naïve Bayes Algorithm and the Four Factor Analysis were utilized in computing and manipulating data sets. Next, the results were compared using the Fuzzy Logic which gave an overall result or the predicted winner of the game.

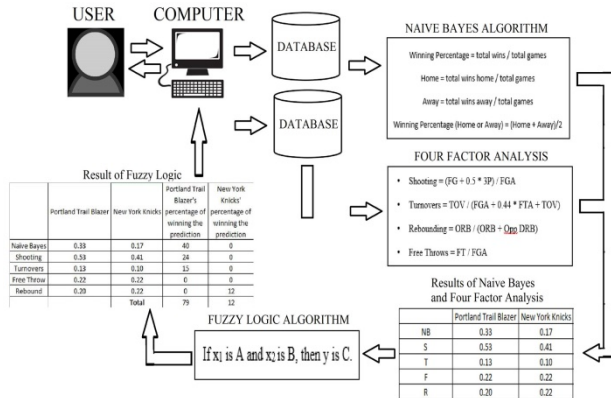


Figure 1. System Architecture.

3. ALGORITHM

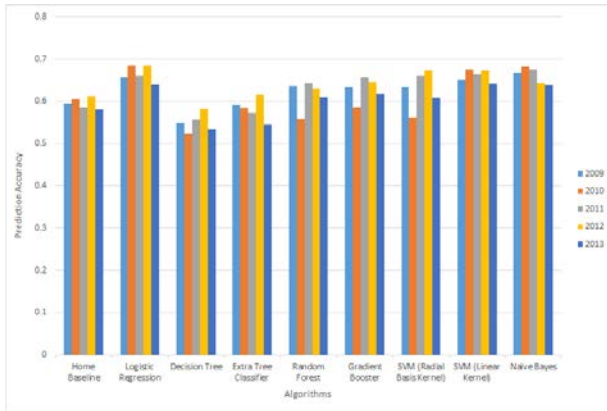


Figure 2 Algorithms used in predicting basketball games.

Figure 2 shows some of the algorithms used in predicting basketball games. Naive Bayes[5], Four Factor Analysis[6] and Fuzzy Logic Algorithm are few of the most commonly used algorithm in prediction. Many developers used these algorithms individually to predict the outcome of the game in NBA. These algorithms are used in analyzing the win-loss ratio and performance of the basketball teams. The researchers combined these three algorithms to formulate a new forecasting technique to predict the NBA basketball game.

3.1 Naive Bayes

In predicting using Naïve Bayes model, only the winning rate is needed as a feather. On this condition, the algorithm is effective and can predict which team has more ability to go to next round in playoffs [5]. The most straightforward way of describing

basketball teams in such a way that success in a match can be predicted relate to scoring points – either scoring points offensively or preventing the opponent's scoring defensively [11]. Naïve Bayes algorithm was used in the study to calculate the probability that team A will beat its opponent team B when they meet. The following equation was used:

$$\text{Winning Percentage} = \text{total wins} / \text{total games}$$

$$\text{Home} = \text{total wins home} / \text{total games}$$

$$\text{Away} = \text{total wins away} / \text{total games}$$

$$\text{Winning Percentage (Home or Away)} = (\text{Home} + \text{Away})/2$$

3.2 Four Factor Analysis

Four factors are box score derived metrics that correlate most closely with winning basketball games. These factors also identify a team's strategic strengths and weaknesses. Four factors can be applied to both team's offense and defense; hence it gives us eight factors [6]. These shows that predicting the winner in a game can also be related to the four factors that affect a basketball game. Although the Four Factors are not all equivalent in value, players can do better than their opponent in three of these factors and still lose. Players can be a good team at three factors and poor at the other and only end up with a mediocre team. Identifying the factors that are important then helps in identifying a strategy for constructing a successful team. The following are the weighted value of each factor and the formulas applied in the system [7], where the shooting was the most important factor, followed by turnovers, rebounding, and free throws and each factor measured as:

Weighted Values: Shooting (40%), Turnovers (25%), Rebounding (20%), Free Throws (15%)

Formulas for computing each factor:

$$\text{Shooting} = (\text{FG} + 0.5 * 3P) / \text{FGA}$$

$$\text{Turnovers} = \text{TOV} / (\text{FGA} + 0.44 * \text{FTA} + \text{TOV})$$

$$\text{Rebounding} = \text{ORB} / (\text{ORB} + \text{Opp DRB})$$

$$\text{Free Throws} = \text{FT} / \text{FGA}$$

Where:

FG = Field Goals

3P = 3 points made

FGA = Field Goals Attempted

TOV = Turnover

FTA = Free Throws Attempted

FT = Free Throws made

ORB = Offensive Rebound

Opp DRB = Opp Defensive Rebound

3.3 Fuzzy Logic Algorithm

An important concept in fuzzy logic is a fuzzy proposition. Fuzzy proposition connects variables with linguistic labels defined in variables. Fuzzy relations assembled from linguistic knowledge expressed as IF-THEN rules which leading to algorithms describing what action or output should be taken in terms of the currently observed information. This IF-THEN rule is widely used by the fuzzy inference system to compute the degree to which the input data matches the condition of a rule. The fuzzy rules are used for stabilizing a relationship between the input and the output [8][9][10]. The study used the Fuzzy rule that was defined as:

If variable IS property THEN action

From that rule, deriving it to an equation:

If x_1 is A and x_2 is B, then y is C.

Where:

x_1 = 1st team

x_2 = 2nd team

y = team to win between x_1 and x_2

A = computed data of the 1st input

C = winner

B = computed data of the 2nd input

4. RESULTS AND DISCUSSIONS

Cascading algorithm was the term used in the study pertaining to the three algorithms used in predicting the basketball game results. It is defined as the step by step or procedural use of different algorithms to come up with a better prediction in basketball. Several rounds of testing were done to ensure that the systems' goals were achieved. Comparison of results was done by predicting the games in 2015-2016 NBA regular season using the recent three games statistics data of the teams versus the real-time results. The researchers found out that a basketball team's win and statistical record of their last three recent games is more applicable to use rather than recent five or ten games. Distributing 40% for Naïve Bayes Algorithm and 60% for Four Factor Analysis for the comparison in Fuzzy Logic has better and more realistic result than other distribution [50%-50%, 30%-70%][7]. The prediction got 70% accuracy including the real-time prediction. In comparing the results, the study used the data of the computation in Naïve Bayes Algorithm and Four Factor Analysis which are the Winning Percentage, Shooting, Turnovers, Rebound, and Free Throws.

Table 1. Sample Computed Data of the Naïve Bayes Algorithm and Four Factor Analysis

	Portland Trail Blazer	New York Knicks
NB	0.33	0.17
S	0.53	0.41
T	0.13	0.10
F	0.22	0.22
R	0.20	0.22

Where: NB = Winning Percentage (Home or Away),
S = Shooting, T = Turnovers, F = Free Throws,
R = Rebound

Table 1 shows the Fuzzy Logic compares the outcome of each category and give points to the better category. From there, the team that has the higher total points has the chance of winning the game. In the study, 1 point was given to the team who will win in every category. Consequently, there are instances that a team got three points and the other team got two points but still the team with only two points win. To measure the results of the Fuzzy Logic Algorithm more realistic, the derived values for each factor are as follows:

NB = 40%, S = 24%, T = 15%, F = 9%, R = 12%.

Giving 40% for NB because winning percentage was the best factor among the five factors. The remaining 60% was distributed to the four factors by getting the 40% of 60 for S, 25% of 60 for T, 15% of 60 for F, and 20% of 60 for R pertaining to the measures given by Oliver[7] to the four factors.

Table 2. Computed results by comparing both teams

	Portland Trail Blazer	New York Knicks	Portland Trail Blazer's Percentage in Winning the Prediction	New York Knicks' Percentage in Winning the Prediction
NB	0.33	0.17	40	0
S	0.53	0.41	24	0
T	0.13	0.10	15	0
F	0.22	0.22	0	0
R	0.20	0.22	0	12
TOTAL			79	12

Table 2 shows that Portland Trail Blazer got 79% chance to win rather than New York Knicks with only 12% to win. F was not computed because they just tied so it will be void. Using these

values, the prediction was better than giving 1 point to each category. The study tried to give the higher percentage of 50% and a lower percentage of 30% but it resulted to a lesser prediction accuracy.

4.1 Project Structure

The Naive Bayes and four factor analysis are the ones responsible for computing the win-loss record and statistic record of the teams. While fuzzy logic is used to compare the output of in the Naive Bayes and four factor analysis. The output shows the computed data of both teams and the predicted winner result of the game.

Figure 3. Enter inputs.

The user (see Figure 3) needs to enter the first and second team then ask to enter when the exact date of the match that will be predicting so the system can get the stats of the team of their last 3 recent games.

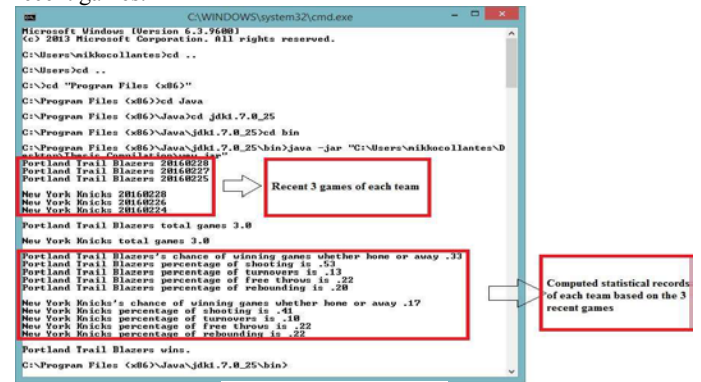


Figure 4. Capabilities.

After inserting the two teams and date, the output shows (see Figure 4) the dates of the last 3 recent games of both teams. Then the system shows the computed statistic record of the teams during their last 3 recent games and the predicted winning team.

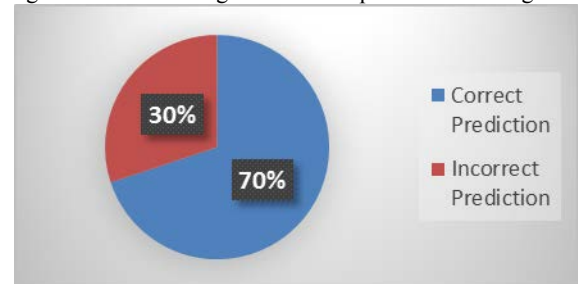


Figure 5. Accuracy Chart of the Basketball Prediction using Cascading Algorithm.

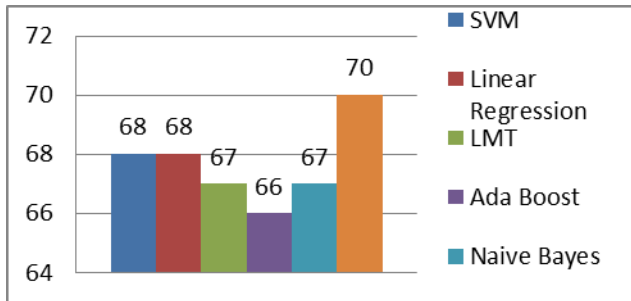


Figure 6. Accuracy Chart.

The researchers also compared the accuracy of the cascading algorithm in other existing approaches such as Support Vector Machine (SVM), Linear Regression, Ada Boost, Logistic Model tree, and Naive Bayes using data sets of 2015-2016 NBA regular seasons. These algorithms are commonly used to predict basketball outcomes because they always show a high accuracy result. In this chart (see Figure 6), you can see the cascading algorithm got the highest accuracy reaching 70 % accuracy compared to the other algorithms.

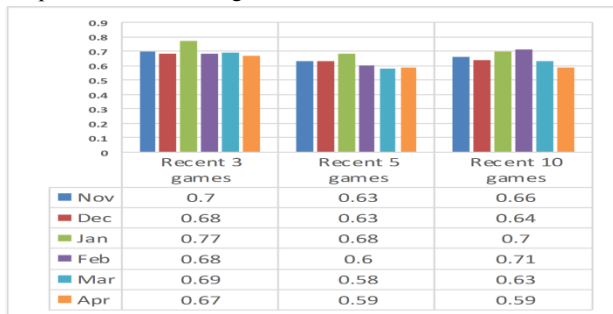


Figure 7. Accuracy based on different set of Recent Games.

The last three, five, ten recent games were compared to identify the highest accuracy. During these months, comparing the last three recent games to the other sets, in most of the months, the accuracy level during January is higher than the 70 % expected rating (see Figure 7).

Table 3. Recent 5 Games

Recent 5 Games	Correct Prediction	Total Games	Accuracy Percentage
November	160	254	63%
December	140	228	63%
January	157	230	68%
February	98	164	60%
March	137	237	58%
April	61	103	59%
TOTAL	756	1216	62%

Table 3 shows the number of correct prediction, total games and accuracy percentage per month using the last five recent games which have a total of 62% accuracy.

Table 4. Recent 10 Games

Recent 10 Games	Correct Prediction	Total Games	Accuracy Percentage
November	168	254	66%
December	146	228	64%
January	162	230	70%
February	112	164	71%
March	164	237	63%
April	69	103	59%
TOTAL	852	1216	66%

Table 4 shows the number of correct prediction, total games and accuracy percentage per month using the last ten recent games which have a total of 66% accuracy.

Table 5. Recent 3 Games

Recent 3 Games	Correct Prediction	Total Games	Accuracy Percentage
November	177	254	70%
December	154	228	68%
January	176	230	77%
February	112	164	68%
March	164	237	69%
April	69	103	67%
TOTAL	852	1216	70%

Table 5 shows the number of correct prediction, total games and accuracy percentage per months using the last three recent games which have a total of 70% accuracy higher than the last five and ten recent games.

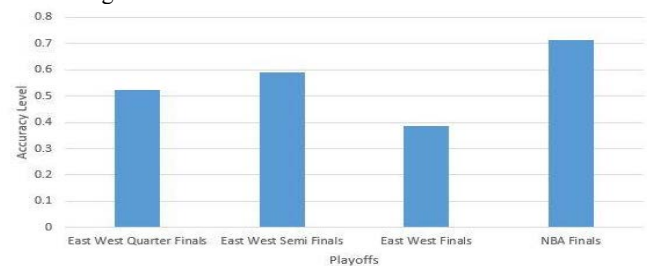


Figure 8. Playoffs Accuracy.

Figure 8 shows the accuracy percentage of the playoffs in both east and west teams. The NBA finals range an accuracy level of 70%.

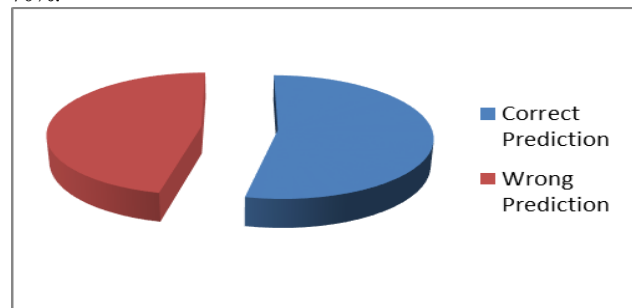


Figure 9. Playoffs Accuracy level

Figure 9 shows the percentage of correct and wrong prediction during NBA playoffs where it got a total of 46 correct predictions out of 86 games and 40 wrong predictions out of 86 games with an accuracy of 53%.

Table 6. Recent 3 Games Accuracy

3 Recent Games	Correct Prediction	Total Games	Accuracy Percentage
East West Quarter Finals	23	44	52
East West Semi Finals	13	22	59
East West Finals	5	13	38
NBA Finals	5	7	71
TOTAL	46	86	53%

Table 6 shows the number of correct prediction, total games, and accuracy percentage during NBA playoffs using three recent games which have a total of 53% accuracy level.

4.2 Project Evaluation

The study used different types of evaluation tools for non-technical and technical users. IT professionals used the technical evaluation form while the basketball enthusiast, coaches, and analysts used the non-technical sheet. From the two categories, the researchers used the purposive random sampling in getting the population of the survey by randomly asked different people who are capable of testing our system. Basketball enthusiast classified as the people that are the fan of the basketball. Basketball coaches and analysts classified as the people that know the rule of the basketball game and willing to contribute their knowledge and opinion about the game. The evaluation tools characteristics are based on the ISO 9126 international standard for the evaluation of software[12]. This statistical formula was used to get the results or the average of the target users.

$$A = \frac{S}{N}$$

Where:

A = the average of the results

N = the number of items or the number of terms

S = the sum of the result or the sum of number being averaged

The survey used the scaling basis from 1 to 4 wherein 1 is the lowest and 4 is the highest. The following are the equivalent rating for each score and its interpretation (see Table 7).

Table 7. Decision Criteria

Scale	Range	Interpretation
4	3.25-4.00	Very Acceptable
3	2.50-3.24	Acceptable
2	1.75-2.49	Slightly Not Acceptable
1	1.00-1.74	Strongly Not Acceptable

Table 8. Summary Evaluation for Technical Users Results

Criteria	Mean	Interpretation
Efficiency	2.89	Acceptable
Maintainability	3.33	Very Acceptable
Portability	2.83	Acceptable
Mean Average	3.02	Acceptable

The efficiency of the system got a rating of 2.89. It means that the system is efficient in a way that it is working well depending on the algorithm and statistical data used. While the maintainability got a rating of 3.33, thus the system is easy to maintain. In terms of portability, it got a rating of 2.83 and can be interpreted that the system is portable with every change that comes with the statistical data given. As a whole, that evaluation shows that there is 3.02 rating which can be interpreted that the system is good to deploy or use in predicting (see Table 8).

Table 9. Summary Evaluation for Nontechnical Users Results (Basketball Enthusiasts, Coaches and Analysts)

Criteria	Enthusiasts		Coaches & Analysts		Mean	Interpretation
	Mean	Interpretation	Mean	Interpretation		
Functionality	3.44	Very Acceptable	3.53	Very Acceptable	3.44	Very Acceptable
Reliability	3.64	Very Acceptable	3.67	Very Acceptable	3.64	Very Acceptable
Mean Average					3.54	Very Acceptable

With overall rating of 3.54 from the non-technical side, the system's result can be accepted as a valid forecast.

5. CONCLUSION

The researchers conclude that the implemented cascading algorithm is capable of predicting the outcome of a basketball game in NBA basketball matches regular season 2015-2016. With the used of Naive Bayes, Four Factor Analysis, and Fuzzy Logic algorithms the implemented cascading algorithm manages to reach 70% acceptable prediction accuracy which is better than other basketball prediction commonly used such as SVM and Logistic Regression model.

6. REFERENCES

- [1] Torres, R. A. (2013). Prediction of NBA games based on Machine Learning, 2. Retrieved from http://homepages.cae.wisc.edu/~ece539/fall13/project/AmorimTorres_pro.pdf
- [2] Cao C. (2012). Sports Data Mining Technology Used in Basketball Outcome Prediction. Retrieved from <http://arrow.dit.ie/cgi/viewcontent.cgi?article=1040&context=scschcomdis>
- [3] Puranmalka,K. (2012). Modelling the NBA to Make Better Predictions. Retrieved from <http://hdl.handle.net/1721.1/85464>
- [4] Richardson L. (2015) How Predictable is NBA?, Nylon Calculus Retrieved from nyloncalculus.com/2015/02/02/predictable-nba/
- [5] Wei, N. (2012). Predicting the outcome of NBA playoffs using the Naïve Bayes Algorithms. Retrieved from www.few.vu.nl/nl/
- [6] NBAstuffers. (2007). Team Evaluation Metrics. Retrieved from <https://www.nbastuffer.com/analytics101/four-factors/>
- [7] Oliver, D. (2004). Roboscout and the Four Factors of Basketball Success. Basketball on Paper. Retrieved from http://www.rawbw.com/~deano/articles/20040601_roboscout.htm
- [8] Cruz, P. , Molina, A., MscCleery, B. (2016). Fuzzy Logic Type 1 and Type 2 Based on LabVIEW FPGA. DOI: 10.1007/978-3-319-26656-5
- [9] Bai, Y.& Wang, D. (2006). Advanced Fuzzy Logic Technologies in Industrial Application. Pp 17-36. DOI: 10.1007/978-1-84628-469-4
- [10] Trawinski, K. (2010). "A fuzzy classification system for prediction of the results of the basketball games",IEEE International Conference on Fuzzy Systems, Barcelona, Spain,pp.1-7
- [11] Shi,Z., Moorthy,S.,& Zimmermann,A. (2013). Predicting NCAAAB match outcomes using ML techniques – some results and lessons learned. Retrieved from www.math.smith.edu/~bbaumer/pub
- [12] ISO 9126 Software Quality Characteristics. Retrieved from www.sqa.net/iso9126.html