

PAPER • OPEN ACCESS

## A Review on Football Match Outcome Prediction using Bayesian Networks

To cite this article: Nazim Razali *et al* 2018 *J. Phys.: Conf. Ser.* **1020** 012004

View the [article online](#) for updates and enhancements.

### Related content

- [Predicting Football Matches Results using Bayesian Networks for English Premier League \(EPL\)](#)  
Nazim Razali, Aida Mustapha, Faiz Ahmad Yatim et al.
- [The research of elevator health diagnosis method based on Bayesian network](#)  
Chang Liu, Xinzheng Zhang, Xindong Liu et al.
- [A comprehensive probabilistic analysis model of oil pipelines network based on Bayesian network](#)  
C Zhang, T X Qin, B Jiang et al.



**IOP | ebooks™**

Bringing you innovative digital publishing with leading voices to create your essential collection of books in STEM research.

Start exploring the collection - download the first chapter of every title for free.

# A Review on Football Match Outcome Prediction using Bayesian Networks

Nazim Razali<sup>1</sup>, Aida Mustapha<sup>1</sup>, Sunariya Utama<sup>2</sup>, Roshidi Din<sup>2</sup>

<sup>1</sup> Faculty of Computer Science and Information Technology, Universiti Tun Hussein Onn Malaysia, 86400 Parit Raja, Batu Pahat, Johor, Malaysia

<sup>2</sup> School of Computing, College of Arts and Sciences, Universiti Utara Malaysia, 06010 Sintok, Kedah, Malaysia

E-mail: nazim.iium@gmail.com, aidam@uthm.edu.my, sunariya.utama@gmail.com, roshidi@uum.edu.my

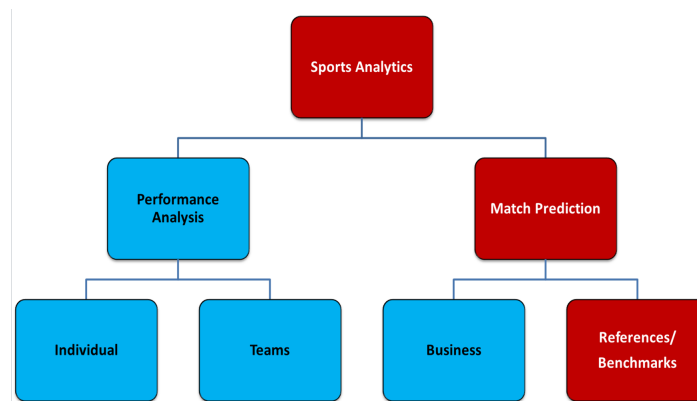
**Abstract.** Modelling association football or soccer prediction model has become a popular issue among scholars in the last few years. This paper is set to study a Bayesian networks (BNs) for football match outcome prediction. The main objective of this paper is to review the application of Bayesian networks (BNs) in football match prediction. This review may be useful for future researchers in developing new prediction model using BNs that incorporate several new features alongside with other existing features that contribute to increase the predictive performance of football prediction model.

## 1. Introduction

Association football (British English) or soccer (American English) is one of the most popular sports in the world. People of the world enjoy playing the sport or watching the matches played by amateur until professional players. The analytics for association football comprises into two parts. First, performance analysis can be classified according to two aspects, in terms of individual performance or team performance. Second is match prediction, which can also be classified into two perspectives, in terms of business purposes, generating profits for betting and bookmakers. The other part is benchmark for evaluating the chance of win, draw or lost for the team before the match begins. This help the managers and players with fresh insight in preparing the countermeasures such as using defensive mentality and offside trap if the opponents is stronger. Figure 1 shows that the categories of sports analytis.

The performance analysis is invaluable in discovering the principal of the observed outcome; win a match or score goals. Based on network theory analysis of football strategies, some of the teams participating in 2010 World Cup are analyzed in terms of team passes. The result produced from [1] were the closeness (how well connected are a player), betweenness (how dependent team to the player), page rank (how important are a player), and clustering. The limitation of this analysis is that it did not consider defensive strength of the opponents. It also did not focus on passing accuracy and shot analysis, which are fundamental in a football match. Otherwise, the match prediction obtains the profits from betting will not be touched in this research because Islam does not tolerate with gambling as Islam strictly prohibit its followers from gambling.





**Figure 1.** Classification of sport analytics categories

Prediction in football has become an intriguing research problem, partly because of the difficulty, because there are many factors can influence the football matches outcome, such as teamwork, skills, weather, home advantage and many others. The challenge is also faced by the experts in football as well because it is very difficult to anticipate the actual results of football matches. Anything can be happened within 90 minutes of football match with none or extra time such as injury or sent off of player by red card. Luck also can be a factor that influence the results of football matches as the strong team are not necessarily win the match against the weak team. Due to the different factors that influence the football match, some research resort to Bayesian Networks (BNs) due its proven applicability in predicting the weather, sports, disease, technology and etc.

Bayesian Networks have been widely used in researches or studies in field of Artificial Intelligent (AI) and Statistics. The Artificial Intelligence and Statistics shared the same goal of modelling real-world phenomena and what the slightly difference between them is AI researches give priority to knowledge based approach for achieving objective study while statisticians give priority to data [2]. The remainder of this paper is organized as follows. Section II introduces the theoretical foundation of Bayesian Networks. Section III reviews related work in football prediction that applied Bayesian Networks. Section IV discusses the different implementation of BNs in match outcome prediction and finally Section V concludes with some indication of future works.

## 2. Bayesian Networks (BNs)

Bayesian networks (BNs) was known in many names such as belief networks, Bayes nets, probabilistic belief, Bayes network and causal network are actually belonging to the family of probabilistic graphical models. It is used to model probabilistic relations among random variables which mean the items involved vary in some random or unexplained manner [3]. [4] points out that BNs is needed to encounter the knowledge where the inferences drawn are subjective, uncertain and incomplete in the beginning reasoning process for probability theory. The probability theory is very useful in order to calculate and permits inferences to flow in two ways; from hypothesis to evidence (predictive) as well as from evidence to hypothesis (diagnostic). It showed that BNs can be used to predict the outcomes or diagnose causal effect (if structures are known).

In addition, BNs can be used to discover causal relationship (if structures are not known). Therefore, BN offered solution for inferential reasoning, decision and uncertainty using quantitative and qualitative components. The qualitative component is represented by the network structure while the quantitative component is represented by the network parameters, which are

the conditional probability distributions of the nodes in the network. The main idea of Bayesian network is actually come from the work of Thomas Bayes called the Bayes' theorem. The Bayes' theorem is represented in Equation 1 or Equation 2.

$$P(\text{cause}|\text{effect}) = \frac{P(\text{effect}|\text{cause}).P(\text{cause})}{P(\text{effect})} \quad (1)$$

where:

- $P(\text{cause}|\text{effect})$  is the conditional probability quantifies the relationship in causal direction.
- $P(\text{effect}|\text{cause})$  describes the diagnostic direction.
- $P(\text{effect})$  as the evidence of the unknown cause.
- $P(\text{cause})$  as the unknown variable that needed to be determined.

$$P(X|Y) = \frac{P(Y|X).P(X)}{P(Y)} \quad (2)$$

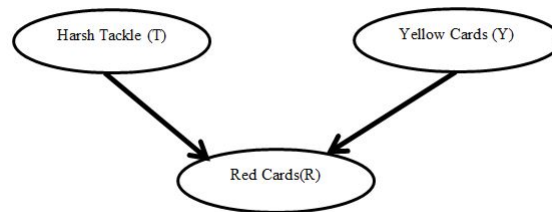
where:

- $P(X)$  is the prior probability or marginal probability of  $X$ .
- $P(X|Y)$  is the posterior probability or conditional probability of  $X$  given  $Y$ .
- $P(Y|X)$  is the conditional probability of  $Y$  given  $X$  (the likelihood of data  $Y$ ).
- $P(Y)$  is the prior probability or marginal probability of data  $Y$  (the evidence).

Equation 1 and Equation 2 explained the Bayes' theorem both literally and technically. Upon observation of evidence, the prior probability of a hypothesis changes to a posterior probability. This can be illustrated in the following way; before observing the data  $Y/\text{effect}$ , the distribution for  $P(X)/\text{cause}$  would be known as the prior distribution because  $X/\text{cause}$  does not have any information about  $Y/\text{effect}$ . This distribution quantifies the parameter values before  $Y/\text{effect}$  is observed. However, once the data  $Y/\text{effect}$  has been observed, this prior distribution is updated to the posterior distribution  $P(X|Y)/P(\text{cause}|\text{effect})$ . The term "posterior" means that the probability is derived from data  $Y/\text{effect}$ . The task of computing this probability distribution for a set of query nodes, given the values for some evidence or observation nodes is how we reason with probabilistic systems like BNs or dynamic BNs.

### 2.1. Qualitative Component of Bayesian Networks

A qualitative component of Bayesian network (BN) is represented the network structure called directed acyclic graph (DAG). A directed acyclic graph implemented set of nodes that characterize random variables from the domain and directed edges connecting nodes to represent the conditional dependencies between nodes. Though, the directed edges unable on the form of any directed cycles. The Bayesian network in form of directed acyclic graph is stated in Figure 2 as follows. A qualitative component of Bayesian network (BN) is represented the network structure called directed acyclic graph (DAG). A directed acyclic graph implemented set of nodes that characterize random variables from the domain and directed edges connecting nodes to represent the conditional dependencies between nodes. Though, the directed edges unable on the form of any directed cycles. The Bayesian network in form of directed acyclic graph is stated in Figure 2 as follows.

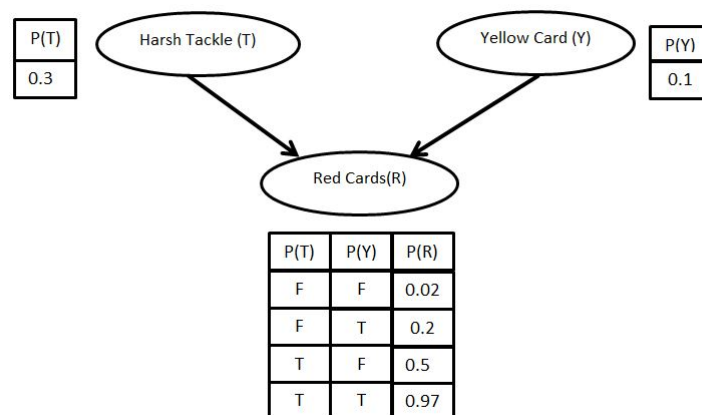


**Figure 2.** A simple Bayesian network in Directed Acyclic Graph

In Figure 2 showed there are two events can cause red card delivered by referee to the footballs player which are either due to harsh tackle or the second yellow card according on how the referee interpret the situation and judgement. Therefore, all three variables have two possibility which is *true(T)* and *false(F)*.

## 2.2. Quantitative Component of Bayesian Networks

A quantitative component of Bayesian network is used to encode the local influences among the random variables using probability distributions [5]. Thus, it is actually can be explained that a quantitative component of Bayesian network is in form of a set conditional probabilities for discrete data or probability density function for continuous data. The BN in form of conditional tables is stated in Figure 3.



**Figure 3.** A simple Bayesian network with conditional probability tables (CPT)

Based on Figure 3 presented the probabilities of two events; harsh tackle and yellow card affect the red card. Given the harsh tackle also give possibility for the player receiving yellow card or red card. It explained that each node in a Bayesian network must has a probability distribution table, where the probability distribution of the node is given respected to its parents, the equation shown as stated in Equation 3:

$$P(X_i | Parents(X_i)) \quad (3)$$

Parents ( $X_i$ ) means the all the nodes that have an arrow to node  $X_i$ . For example, yellow card and harsh tackle is a parent of red card because harsh tackle and yellow card can cause

red card or red card is a consequence of yellow card and harsh tackle. By multiplying all nodes given their parents, the joint distribution of the network is found as stated in Equation 4:

$$P(X_1, X_2, \dots, X_n) = \prod_{i=1}^n P(X_i | \text{Parents}(X_i)) \quad (4)$$

Where application of the chain rule of probability theory shown in Equation 3 allowed factorization of joint probabilities as in Equation 4,

$$\begin{aligned} P(X_1, X_2, \dots, X_n) &= P(X_1) \times P(X_2 | X_1), \dots, \times P(X_n | X_1, \dots, X_{n-1}) \\ &= \prod_{i=1}^n P(X_i | \text{Parents}(X_i)) \end{aligned} \quad (5)$$

Given probability 0.3 for  $P(T)$  means that it is 30% chance of tackle committed by player considered harsh but there is 70% chance the tackle committed is considered legal. Furthermore, 97% chance of receiving red card when harsh tackle was committed and the player already received yellow card once. It is because the nodes representing a red card are conditioned on the nodes for harsh tackle and yellow card. Its probabilities are given respected to its parents nodes. Therefore, only 3% chance that the player will not gained red card. This might be the result when the referee making wrong decision or there are some other events that might influence the judgement of referee such as the referee failed to interpret the situation.

### 3. Application of Bayesian Networks in Football Prediction

Application of Bayesian networks (BNs) in football prediction is all started by [6] that improvised the Poisson distribution of [7] and [8] for predicting football match outcome using dynamic Bayesian networks (DBNs). [7] present the strength of defensive and offensive strength of home and away team with goal scoring distribution using independence Poisson distribution. However, Maher model just an ex-ante model which means use actual result for expressing the defensive and offensive of team and cannot predict the future. [8] enhanced the Maher model by implementation of bivariate Poisson distribution for football prediction. Next, [6] applied the concept of [8] in Bayesian way where it can incorporate more uncertainty or time-variation parameter such as psychological effect.

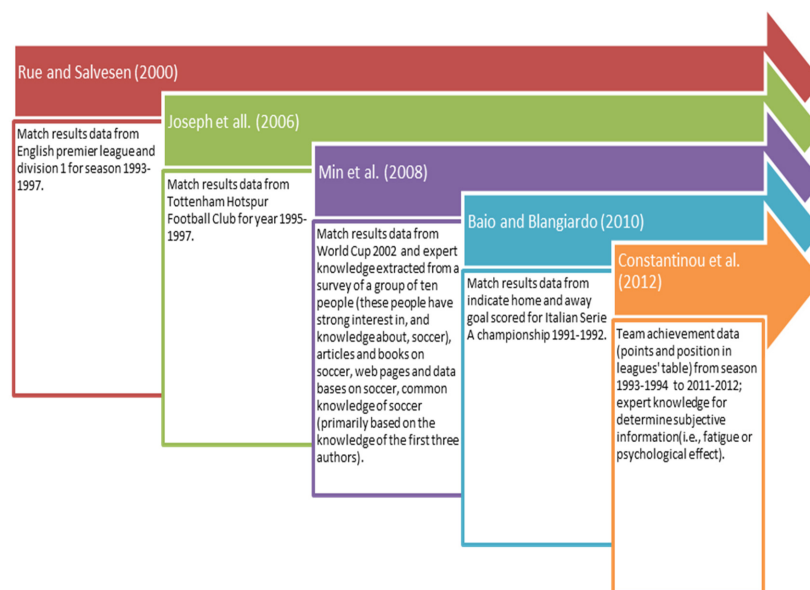
[9] proposed an expert Bayesian Networks (BNs) model and compared with four alternative machine learning (ML) technique, which are naive BNs, data driven BNs, K-Nearest Neighbor and Decision Tree. The efforts of the study was restricted to all league matches by Tottenham Hotspur F.C. during seasons 1995/1996 and 1996/1997. The features used in expert BNs was presence of three players, the position of a player who always play, quality of opposing team in 3-point scale, venue, attacking force quality, overall team quality, and performance given their own and opposition quality. The experiment was repetitive with various machine learning approaches using the same set of features.

[10] proposed framework called FRES (Football Result Sports System) for Bayesian inference and rule-based reasoning. There are twenty seven features used in this model, which were sourced from match statistics. The features were used to construct four Bayesian networks for offense, defense, possession, and fatigue. It is found that the produced a Relative Rank Error of 5.920 and Root Mean Square Error of 5.306 during forecasting the result of World Cup 2002. The model considered dynamically in order to changes team strategies throughout the entire match, which in turn influenced the prediction results.

Then, [11] proposed a Bayesian hierarchical model with aim to guesstimate the appearances that bring a team win and lose a match, predict the outcome of a particular match and test

the models predictive strength. The model test used the data Italian Series A Championship 2007-2008. However, the model have drawback called over shrinkage. The over shrinkage occurred when the grand mean of the overall observation has been pulled by some of the extreme occurrences such as there are few teams that have very good performance and few teams that struggling for relegation due to ineffective performance faced each other in a match. It is to changeling in order to estimate and capture each team performance because the strong team may overestimate the weak team.

Other Bayesian network model for forecasting match was is presented in [12]. The data used for training were the English Premier League season 1993/1994 to 2009/2010. The two types of features used the objective information and subjective information. The objective information is the total points previous five season and total points of current season. Whereas the subjective information are the expert belief of team strength, key player and first team player availability, previous form, motivation, team spirit, managerial issue, head-to-head bias, days since last match, first team player rested during last match, previous match toughness, and national team participant. The accuracy of the model was evaluated using Rank Probability Score (RPS) on forecast that make use objective information only and forecast that make use of objective and subjective information. Based on the forecast of English Premier League season 2010/2011 matches, objective and subjective information forecast possess better accuracy than objective information only forecast. Figure 4 summarized the dataset used for prediction model using Bayesian approaches.



**Figure 4.** The dataset used by previous researchers for BNs football prediction model

Based on Figure 4, it is discovered that the datasets used for implementing the football prediction using Bayesian approaches generally from match statistical data or team statistical data and even from expert judgment of the club for football league in the several season.

#### 4. Discussion

The reviews have shown that existing models rely heavily on historical data for team features (e.g. total number of shots, the number of goals scored or conceded by a team in match and etc.), matches features (e.g. results of previous meeting between team 1 and team 2, results of past matches for team 1 and etc.) and team achievement (ranking of a team in league or



world, total number of points gained by team during or end of the season and etc.) which are mostly readily accessible via FIFA Association website and news broadcasting. However, some of them used expert knowledge or specific datasets which they need to get the help or service from football expertise or need to buy from sports data provider such as Opta (<http://www.optasports.com/>) and STATS (<https://www.stats.com/>) in order to increase their prediction model accuracy. Thus, this expert type of data gained from expertise sometimes can be biased toward prediction model itself because every expert has different view and opinion which are subjective and may affect the data reliability. Besides, sports data gained from data provider can be expensive if a lot of data was needed for modelling prediction model (learning state).

[13] point out that harvesting more match data can generate even richer prediction model for football. He improvised the works of [6] by adding the player statistic data, the total number of shots fired and shots on target which outperformed other traditional statistical football prediction model [7, 8, 6] which based on the team past results data. As for football sports, the use of Bayesian approaches in predicting match outcomes are gaining momentum due its capability for predicting uncertainty and reasoning over time. This is particularly important considering the dynamic nature of football matches within 90 minutes. The pi-rating system [12], which was developed based on Bayesian approach outperformed the Elo rating model [14] and its improved version in [15] because the rating models were becoming more dependent towards the bookmaker odds for prediction. As the prediction model improved time by time, they become more complex in how they treat data in order to incorporate more features or variables and associate dynamically with time like the work of [10]

In summary, the research in match outcome prediction for football thrives in two major directions; the statistical approach and the Bayesian approach. Contributions in machine learning approach are expected to grow further in the future as the fact that sports analytics in football requires more features from various perspectives such as the players' strength and team abilities on top of the past match statistics. The performance of existing machine learning approach is also fair as shown by the work of [16] (Fuzzy model) and [17] (Neural Network model), whereby the prediction accuracy for both machine learning techniques is reasonably good for both international and domestic football competition. However, [9] showed that the Bayesian approach outperformed the  $k$ -nearest neighbor and decision tree algorithms under the machine learning approach.

This review has set two potential avenues for future work on the topic of football match prediction. The first is to model the rich data based model which may enhance the capabilities of predictive accuracy. The second is to make the prediction models to be more dynamic and sensitive to real-time changes during the match in order to cater uncertain events such as injury, fouls, and substitutions (team as well as opponents). For example, a match prediction model should be able to put face values on each individual player at specific point of time on an off the match.

## 5. Conclusions

According to FIFA Big Count in 2016 through FIFA official website (<http://www.fifa.com/worldfootball/bigcount/allplayers.html>), Malaysia as of 3 May 2016 has 585,730 players in which 9,930 are registered and the remaining 575,800 are unregistered players. This makes less than one percent from the 265,000,000 football players across the world. Malaysia also hosts 110 clubs and a total of 11,810 officials in football matches. The findings of the study will contribute greatly on sports development especially in association football or soccer in our country with need a touch in order to improve our teams ranking, pleasing our football fans and shut the criticize towards our football national team. The analysis provided is hoped to help coaches to evaluate and estimate their players performance, hence a more accurate network



structure between players and the match outcomes (win, draw and loss). As the results, this findings are invaluable for developing and forming a team with the highest winning probability.

From the economic perspective, the improvement of club performance on the field would lead to improved revenue from stadium attendance and merchandise sold, thus lead to national economic growth. Besides, a new BNs prediction model can be used to set benchmark which may assists the managers and players acts with more cautions and prepare the alternative way to beat the opponent that stronger than them such as setting up offside trap, implement defensive mentality and play counter attack.

### Acknowledgements

This research is funded by the Ministry of Higher Education Malaysia (MOHE) under the Research Acculturation Collaborative Effort (RACE) Grant Scheme Vot 1513.

### References

- [1] Pena J and Touchette H 2012 *Sports Physics: Proc. 2012 Euromech Physics of Sports Conference* 517–528
- [2] Heckerman D, Geiger D and Chickering D 1995 *Machine Learning* **20** 197–243
- [3] Heaton J 2013 *Forecasting & Futurism* **7** 6–10
- [4] Pearl J 1985 Bayesian Networks: A Model of Self-Activated Memory for Evidential Reasoning (No: CSD-850021)
- [5] Kersting K and Raedt L 2000 Bayesian Logic Programs (Report No. 151)
- [6] Rue H and Salvesen O 2000 *Journal of the Royal Statistical Society: Series D (The Statistician)* **49** 399–418
- [7] Maher M 1982 *Statistica Neerlandica* **36** 109–118
- [8] Dixon M J and Coles S G 1997 *Journal of the Royal Statistical Society. Series C: Applied Statistics* **46** 265–280
- [9] Joseph A, Fenton N E and Neil M 2006 *Knowledge-Based Systems* **19** 544–553
- [10] Min B, Kim J, Choe C, Eom H and (Bob) McKay R I 2008 *Knowledge-Based Systems* **21** 551–562
- [11] Baio G and Blangiardo M 2010 *Journal of Applied Statistics* 1–13
- [12] Constantinou A C, Fenton N E and Neil M 2012 *Knowledge-Based Systems* **36** 322–339
- [13] Langseth H 2013 *Frontiers in A.I. and Applications* **257** 165–174
- [14] Hvattum L M and Arntzen H 2010 *International Journal of Forecasting* **26** 460–470
- [15] Leitner C, Zeileis A and Hornik K 2010 *International Journal of Forecasting* **26** 471–481
- [16] Rotshtein A P, Posner M and Rakityanskaya A B 2005 *Cybernetics and Systems Analysis* **41** 619–630
- [17] Huang K y, Member S and Chang W l 2010 *The 2010 International Joint Conference on Neural Networks (IJCNN)* 18–23