Int. J. Nonlinear Anal. Appl.
Volume 12, Special Issue, Winter and Spring 2021, 971-980
ISSN: 2008-6822 (electronic)
http://dx.doi.org/10.22075/IJNAA.2021.5544

A Bayesian approach for major European football league match prediction

Nazim Razali[a], Aida Mustapha[b,*], Norwati Mustapha[c], Filipe M. Clemente[d]

[a]*Faculty of Computer Science and Information Technology, Universiti Tun Hussein Onn Malaysia, Malaysia.*
[b]*Faculty of Applied Sciences and Technology, Universiti Tun Hussein Onn Malaysia, Malaysia.*
[c]*Faculty of Computer Science and Information Technology, Universiti Putra Malaysia, Malaysia.*
[d]*School of Sport and Leisure, Viana do Castelo Polytechnic Institute, Portugal.*

(Communicated by Madjid Eshaghi Gordji)

Abstract

This paper presents a Bayesian Approach for Major European Football League match prediction. In this study, four variants of Bayesian approaches are investigated to observe the impact of different structural learning algorithms within the family of Bayesian Network which are Naive Bayes (NB), Tree Augmented Naive Bayes (TAN) and two General Bayesian Networks (GBN); K2 algorithm with BDeu scoring function (GBN-K2) and Hill Climbing algorithm with MDL scoring function (GBN-HC). The predictive performance of all Bayesian approaches is evaluated and compared based on football match results from five major European Football League consisting of three complete seasons of 1,140 matches. The results showed that GBN-HC gained 92.01% of accuracy while GBN-K2 and TAN produced comparable results with 91.86% and 91.94% accuracy, respectively. The lowest result was produced by NB, with only 72.78% accuracy. The results suggest that TAN requires further exploration in football prediction with its ability to cater the minimal dependency among attributes in a small-sized dataset.

Keywords: Football, Bayesian networks, Naive bayes, Tree augmented naive bayes and General bayesian networks

*Corresponding author

Email addresses: nazim.uthm@gmail.com (Nazim Razali), aidam@uthm.edu.my (Aida Mustapha), norwati@upm.edu.my (Norwati Mustapha), filipe.clemente5@gmail.com (Filipe M. Clemente)

Received: June 2021 *Accepted:* September 2021

1. Introduction

Bayesian approaches have been frequently employed to predict football match results with great accuracy as compared to statistical and machine learning approaches. A highly accurate match prediction model is vital to assist both managers and players making preparation so that if the opponent team appears to be stronger, tactics such as defensive mentality or an offside trap can be used [10, 28]. The literature has shown a wide range of football match prediction models based on Bayesian approaches such as in [19, 22, 2, 23, 6, 24, 5, 26, 7]. Although the researchers adopted Bayesian approaches in their own way with different set of data samples for modelling match prediction, the fundamental of Bayesian approaches remain.

[19]] tested Bayesian Networks (BN) with several machine learning approaches in a data-driven approach to football prediction, employing 30 attributes for the general model and 9 attributes for the expert model. The data used for both model included the individual player data including availability, position and key players and performance quality for home and away team. However, their model required input from football experts in order to construct the expert Bayesian model. [22] applied the Bayesian inference and rule-based reasoning to develop the intelligent football match predictions system called the Football Result Sports System (FRES). [6] also developed an intelligent system called pi-football which combines two attributes; objective information and subjective information to predict the football match outcome. In a more recent work, [7] developed a prediction model called Dolores by combining two methods, which are dynamic rating system and a hybrid Bayesian network that is able to incorporate the mixture of data containing both discrete and continuous forms. Dolores won second place in the 2017 Soccer Prediction Prediction Challenge for most accurate football prediction model [9].

However, the variety of attributes along with different structural learning algorithms used in previous match prediction models produced different Bayesian network structure, hence different accuracy performance in prediction. Naïve Bayes (NB), for example, is a constraint-based learning algorithm whereby the attributes are constrained with conditional independence assumption. NB rigidly assumes the target class attribute is independent from all other attributes. Tree Augmented Naive Bayes (TAN) extends NB but still maintain the conditional independence assumption of NB. TAN, on the other hand, is represented as a directed tree, with the root node chosen at random and the edges replaced with arcs. This resulted in TAN having a minimal dependency among the attributes. Different from NB and TAN, a General Bayesian Networks (GBN) is more dynamic in structure learning since it depends on the score or search based algorithm to construct its network.

This paper present a Bayesian Approach for Major European Football League match prediction. The goal of this research is to look into the impact of different structural learning algorithms within the family of Bayesian approaches via four Bayesian classifiers; NB, TAN, GBN with K2 structural learning algorithm and BDeu scoring function (GBN-K2), as well as a GBN with Hill Climbing structural learning algorithm and MDL scoring function (GBN-HC). The remainder of the paper is organized as follows. Section 2 describes the background of Bayesian approaches in classification. Section 3 describes the materials and methods used for the comparative experiment including the dataset, experimental setup, and the evaluation metrics. The experimental results are presented in Section 4, and Section 5 concludes with some suggestions for future research.

2. Bayesian network (BN)

Judea Pearl introduced the Bayesian Network (BN) in 1985, based on the work of Thomas Bayes. [16]. Bayesian network are commonly utilised in Artificial Intelligence (AI) and Statistics research since both disciplines aim to simulate real-world phenomena. The only distinction is that AI research prioritises knowledge-based approaches to achieving goals, whereas statisticians prioritise data-driven approaches to achieving goals [14]].

Belief Networks, Bayes Nets, probabilistic belief, Bayes Networks, and causal networks are all synonyms for Bayesian Networks (BN). It is used to simulate probabilistic relationships between random variables where the variables vary in an unpredictable or inexplicable way [15]. [25] stated at the outset of the probability theory reasoning process that BN are required to satisfy knowledge in situations where inferences are subjective, imprecise, and imperfect.

The probability theory is very beneficial for calculating and allows inferences to flow from hypothesis into proof (predictive) and evidence into hypothesis in two directions (inferential or diagnostic). Besides, BN is reliant on the availability of its network structures, known or unknown, in order to forecast outcomes, diagnose causal effects, and discover causal links. As a result, BN provides a quantitative and qualitative solution for inferential reasoning, decision making, and uncertainty. The network topology represents the qualitative component, while the network parameters, which are the conditional probability distributions of the nodes in the network, reflect the quantitative component. Equation 1 shows the basic principle of BN, which composed of network structure and its conditional probabilities.

$$P(X_1,\ldots,X_n)=\prod_{i=1}^{n}P(X_i|Pa_i) \qquad (1)$$

where $P_B(X_1,\ldots,X_n)$ is the joint probability distribution over a set of n random variables $X = (X,\ldots,X_n)$ and Pa_i is the parent of X_i in a Bayesian Networks. With the representation of joint distributions as a product of conditional distributions, the dependency relationship between the nodes in Bayesian Networks can be identified [1]. Figure 1 shows an example of BN structure.

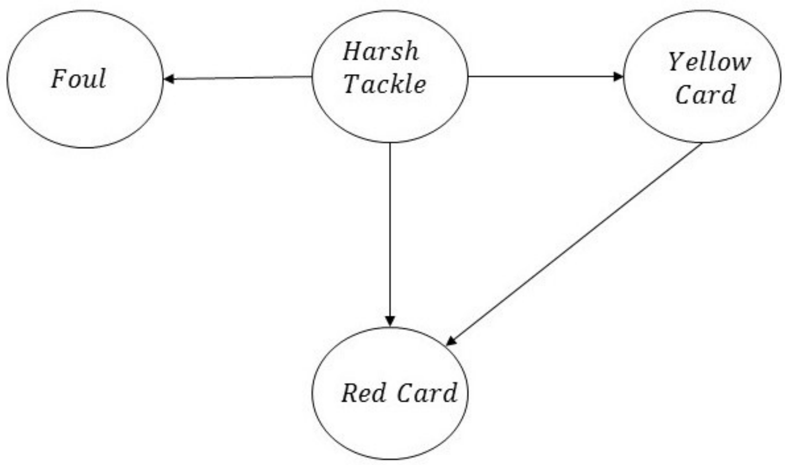

Figure 1: BN Network Structure in Football

3. Learning in bayesian network

Generally, learning a Bayesian Network (BN) from data are divided into two tasks: (1) structural learning and (2) parameter learning. The structure of a BN reveals the underlying probabilistic dependency through the relationships between attributes represented as nodes, as well as their conditional independencies resulting from the learned structure from the data [17]. In other words, structural learning has the role to identify and determine the topology of the network based on the attributes. Meanwhile, parameter learning is used to calculate the conditional probabilities or the numerical parameters for the attributes based on the network structure. According to [3], structural learning is able to offer interpretation of uncertain knowledge and accommodate attribute selection as well as missing data from the data viewpoints.

3.1. Naïve bayes (NB)

The naive Bayes (NB) classifier is the most basic type of Bayesian classifier since it has a fixed network structure. [20] described NB in his work as a constraint-based learning algorithm, whereby there is no dependency between all the variables. Due to this assumption, NB neglect the effect or correlation of other variables on their relationship resulting in a network structure of one parent node as class node and all the remaining nodes as child nodes. In addition, no other connections are allowed in the network structure. Figure 2 shows the network structure for NB, whereby there is only one parent node connected to all child nodes and the child nodes do not have any relationship with other child node since NB assume independence assumption between child node to the parent.

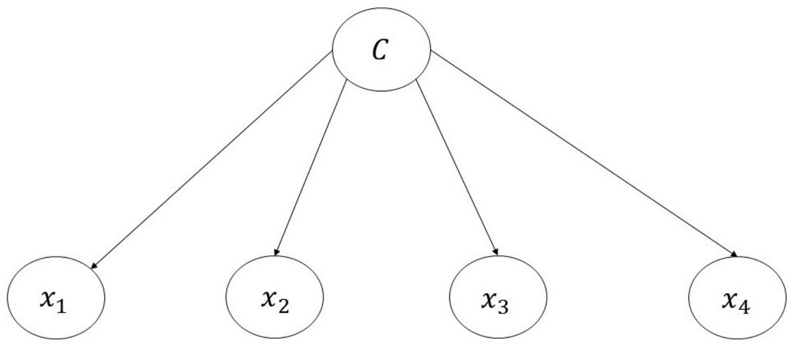

Figure 2: Network Structure for NB

In WEKA environment, implementation of the NB algorithm is based on flexible naïve Bayes by [18], which possesses the same characteristics with NB except on the capability to treat continuous attributes. The flexible NB uses kernel density estimation to model continuous attributes while the standard NB uses a single Gaussian, which is the common technique for handling continuous attributes. Although flexible NB increases the storage and computational complexity as compared to NB, it prevents the loss of information that happen under Gaussian assumptions. As the result, flexible NB offers better accuracy in a classification experiment as compared to a standard NB.

3.2. Tree augmented naïve bayes (TAN)

Tree Augmented Naive Bayes (TAN) was introduced by [13] as Bayesian conditional trees. TAN is an upgraded version of Naive Bayes (NB), whereby TAN assumes a minimal dependency between child nodes while still maintaining the basic structure of an NB. Due to this small information on the inter-relationship between the nodes, TAN has been proven to achieve higher classification accuracy than NB [1, 12, 21]. Figure 3 shows the network structure for TAN.

In the TAN model, each attribute is supposed to be reliant on its class and one other attribute from a future set. So, it is more realistic to use a TAN model because there is always some level of correlation in the data, no matter how small. TAN in Weka environment is based on [12], whereby the NB network is augmented with a tree generated by the maximum weight spanning tree.

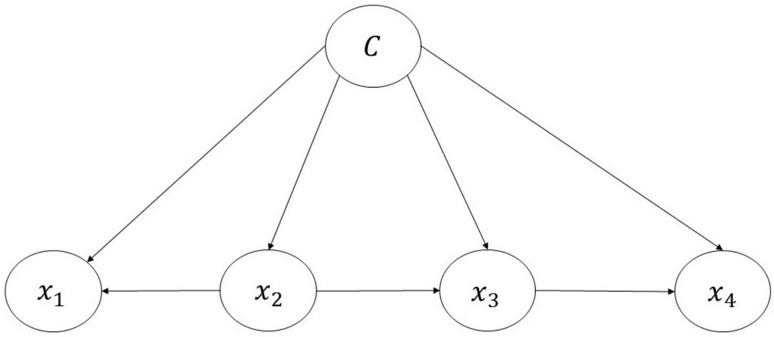

Figure 3: Network Structure for TAN

3.3. General bayesian network (GBN)

General Bayesian Networks (GBN) differs from NB and TAN since GBN offer flexibility on the manner it constructs its own structure. As shown in Figure 4, the parent nodes in GBN can be more than one since there are no restriction in forming the dependency between nodes.

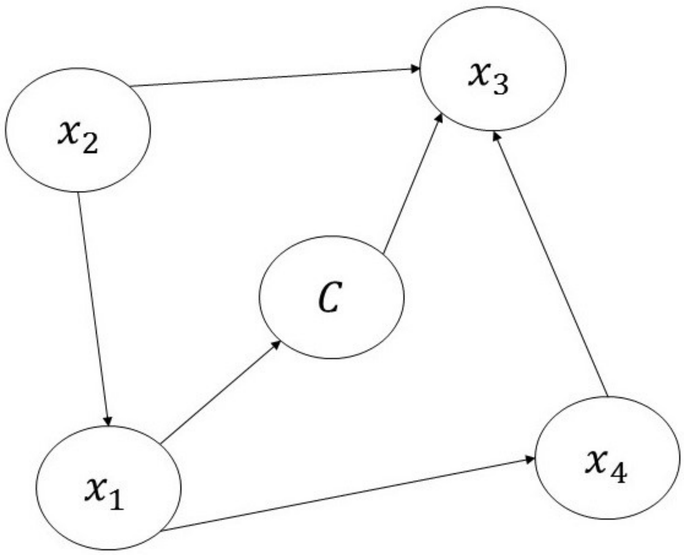

Figure 4: Network Structure for GBN

In this research, two types of searching procedure and scoring metrics have been chosen and implemented. Note that all GBN algorithm with the scoring metric are applicable in Weka environment. The first GBN is adopted from [21] with application of the K2 searching algorithm [8] and Bayesian Dirichlet BDeu scoring metric [4]. The second GBN applies the Hill Climbing searching procedure with the Minimum Description Length (MDL) scoring metric [13].

4. Materials and methods

This research will look into the impact of several structural learning algorithms from the Bayesian approach family on prediction football match outcomes. Following [27], prediction of football match outcome is carried out based on a classification methodology as shown in Figure 5. Four Bayesian classifiers are investigated, which are the Naive Bayes (NB), Tree Augmented Naive Bayes (TAN) and two General Bayesian Networks (GBN); K2 algorithm with BDeu scoring function (GBN-K2) and Hill Climbing algorithm with MDL scoring function (GBN-HC).

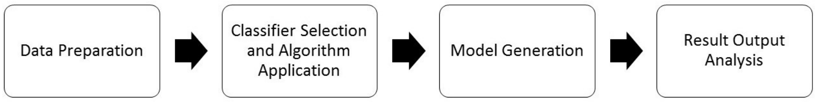

Figure 5: Steps involved in WEKA for Classification

Based on Figure 5, there are four steps involved; preparing the input data, applying the classification algorithm, generating the model and analyzing the prediction output. The classification algorithms focus on four variants of Bayesian approaches, all modeled and compared, which are Bayesian Networks (BN), Naïve Bayes (NB), Tree Augmented Naïve Bayes (TAN) and two General Bayesian Networks (GBN); K2 algorithm with BDeu scoring function (GBN-K2) and Hill Climbing algorithm with MDL scoring function (GBN-HC).

4.1. Dataset

All football European major league data were extracted from legitimate football website at http://www.football-data.co.uk. Top five major European football league including English Premier League (EPL), Spanish League (La Liga), German League (Bundesliga), French League (Ligue 1) and Italy League (Serie A) were chosen for three complete football seasons which from 2014-2015, 2015-2016 and 2016-2017. Each league has 20 competing football teams except for Bundesliga which only have 18. All four leagues; EPL, La Liga, Ligue 1 and Serie A have 380 matches played for the entire season, with Bundesliga having only 306 matches played in a season. In total there are 1,140 matches data used for each league and 924 matches data for Bundesliga in a complete three-season.

The collection of raw data extracted from football website contains 65 attributes for EPL and 64 for the remaining leagues. The EPL data has an extra attribute, which is the referee. However, from the total of 65 attributes, only 19 attributes for EPL and 18 attributes from La Liga, Bundesliga, Ligue 1 and Serie A were selected since the remaining attributes represent the odds from bookmaker companies. As shown in Table 1, the experiments used statistical data from every match, such as the number of shots fired and the number of fouls committed, including red and yellow cards.

4.2. Experimental setup

The classification experiment is then performed using the data mining tool, Waikato Environment for Knowledge Analysis (WEKA) [11]. Weka is a Java-based set of data mining tools and machine learning algorithms for data pre-processing, classification, regression, clustering, association rules, and visualization. Additionally, it is well-suited for constructing novel machine learning techniques, as the algorithms may be applied directly to a dataset or invoked via Java code. It is free software that released under the GNU General Public License. Implementation for all the Bayesian algorithms used in this research are fully available in Weka. Table 2 show the hardware and software specification used in the experiment.

4.3. Evaluation metrics

Accuracy has been selected as an assessment criterion for analyzing the performance of each Bayesian classifier to measure and evaluate the prediction performance of all four algorithms within the Bayesian approach; GBN-K2, GBN-HC, TAN and NB. Weka software provide the confusion

Table 1: Sample of Dataset from the European League

Attribute	**Sample Values**	**Data Type**
Home Team	Burnley	-
Away Team	Swansea	-
Full Time Home Goal (FTHG)	0	Numeric
Full Tima Away Goal (FTAG)	1	Numeric
Full Time Result (FTR)	A	Nominal
Half Time Home Goal (HTHG)	0	Numeric
Half Time Away Goal (HTAG)	0	Numeric
Half Time Result (HTR)	D	Nominal
Referee	JMoss	Nominal
Home Shot (HS)	10	Numeric
Away Shot (AS)	17	Numeric
Home Shot on Target (HST)	3	Numeric
Away Shot on Target (AST)	9	Numeric
Home Foul (HF)	10	Numeric
Away Foul (AF)	14	Numeric
Home Corner (HC)	7	Numeric
Away Corner (AC)	4	Numeric
Home Yellow Card (HY)	3	Numeric
Away Yellow Card (AY)	2	Numeric
Home Red Card (HR)	0	Numeric
Away Red Card (AR)	0	Numeric

Table 2: Hardware and software specification for experimental setup

Hard Specification	
Central Processing Unit (CPU)	Intel Core i5-4200M
Random Access Memory (RAM)	8 GB DDR3
Graphics Processing Unit (GPU)	nVIDIA GeForce GT740M
Hard Drive	1000 GB
Software Specification	
Data Mining Tool	Waikato Environment for Knowledge Analysis (WEKA)
Operating System (OS)	Window 8 (64 bit)
Documentation Tool	LaTeX and Microsoft Word 365

matrix and its analysis as output in term of accuracy, recall, f-measure and precision as shown in Table 3. The accuracy performance of each algorithm was validated using a 10-fold cross validation method based on the work of [1] and [13]. Three complete seasons of each football league are chosen to assess each Bayesian classifier's performance consistency: GBN-K2, GBN-HC, TAN and NB.

Columns denote the predicted class, whereas rows denote the actual class target. The win result is denoted by the label YES, while the loss result is denoted by the label NO. As a result, the accurate predictions are represented by the diagonal elements (TN, TP) in Table 2, while the erroneous predictions are represented by the other elements (FN, FP). For example, there are two outcomes in football match which are win and lose. Corrected win result prediction is TP, and corrected lose

Table 3: Confusion Matrix

	NO (Prediction)	YES (Prediction)
NO (Actual)	True Negative (TN)	False Positive (FP)
YES (Actual)	False Negative (FN)	True Positive (TP)

result prediction is TN, whereas incorrected win result prediction is FP, and incorrected loss result prediction is FN. This test result is added to the FP in the table if a target class is expected to win (YES) despite being a loss (NO) target class. As a result, number FP is increased by 1. Thus, accuracy in the confusion matrix can defined as in Equation 2:

$$Accuracy = \frac{TP + TN}{FP + TP + TN + FN} \qquad (2)$$

5. Results and discussion

The experimental findings for each Bayesian classifier are compared in terms of accuracy performance as the assessment measure, as described in [12, 21] and [1]. Table 4 shows the overall classification result across three seasons of European Major League achieved by TAN, NB, GBN-K2, and GBN-HC. The overall average accuracy results showed that GBN-K2 (91.86%), GBN-HC (92.01%) and TAN (91.94%) are comparable however, extremely outperformed the NB (72.78%). It seems NB result are the worst during Ligue 1 season 2014-2015 with only 69.21% accuracy while the best accuracy was obtained by GBN-HC (97.37%) for La Liga season 2014-2015. In terms of football data, the results showed that NB is not practical to represent the causal relationship due to the fact that NB discard the relationships between nodes, making the network structure under-represented. Besides, it also showed that constraint-based learning such as NB seem illogical because there must be little dependency between some child nodes as well as TAN.

Meanwhile, GBN has potential to be improved further since it has good graphical structure for data representation and do not have any restriction to connect the attributes and the target class. Because the network structure in GBN appears similar to an experts' network structure, GBN models can be tuned with expert knowledge, in other words, a knowledge integration process in football match prediction or betting for further research. At the same time, GBN is adjustable in number of their parent and the random in order. In the future, this research may focus on improving TAN for match or game prediction in both single and team sports since TAN assumes there a minimal dependency on its class and attributes.

6. Conclusions

This paper was set to investigate the effect of different structural learning algorithms within the family of Bayesian approaches. There are several forms of structural learning for Bayesian approaches including four algorithms discussed in this paper; Naïve Bayes (NB), Tree Augmented Naïve Bayes (TAN), General Bayesian Network with Hill Climbing algorithm (GBN-HC) and General Bayesian Network with K2 algorithm (GBN-K2). All the structural learning algorithms were implemented and compared in terms of accuracy performance using three seasons of five major European football leagues; English Premier League (EPL), Spanish League (La Liga), German League (Bundesliga), French League (Ligue 1) and Italy League (Serie A).

The results showed that GBN-HC structural learning algorithm produced the best performance in overall average accuracy but is comparable with GBN-K2 and TAN algorithms across all five

Table 4: Overall Accuracy Results across Three Seasons of the European Major League

Data Sample (Season)	Football League Name	NB (%)	TAN (%)	GBN-K2 (%)	GBN-HC (%)
2014-2015	EPL	71.32	91.32	93.95	93.68
	La Liga	74.47	96.84	96.05	97.37
	Bundesliga	71.57	92.16	89.54	91.18
	Ligue 1	69.21	90.26	88.42	89.47
	Serie A	72.11	83.95	85.26	86.05
2015-2016	EPL	71.05	82.37	81.58	85
	La Liga	74.74	95.79	95.79	96.84
	Bundesliga	72.22	93.46	93.79	92.48
	Ligue 1	71.58	93.42	93.68	92.37
	Serie A	72.89	93.16	93.42	93.68
2016-2017	EPL	78.42	94.74	95.26	93.42
	La Liga	74.47	92.89	94.21	93.42
	Bundesliga	70.26	91.18	92.16	91.18
	Ligue 1	72.89	92.63	89.74	90.53
	Serie A	74.47	95.79	95	93.42
Average Accuracy Percentage	EPL	73.6	89.48	90.26	90.7
	La Liga	74.56	95.17	95.35	95.88
	Bundesliga	71.35	92.27	91.83	91.61
	Ligue 1	71.23	92.1	90.61	90.79
	Serie A	73.16	90.47	91.23	91.05
Overall Average Accuracy Percentage		72.78	91.94	91.86	**92.01**

major European football leagues. Eventhough the performance of both variations of GBN and TAN are comparable, GBN can be improved by incorporating more information and tuning by expert reference since the learned network structure is similar to the network structure in which produced by expert knowledge. In the future, this work will explore further capabilities of TAN in football prediction in order to cater the minimal dependency among attributes in a small-sized dataset as well as other Bayesian structural learning algorithm. Thus, this research may benefited to assist the team manager or coaches in choosing the best eleven players and build up sensible tactical which have high probabilities of winning rate.

Acknowledgement

Communication of this research is made possible through monetary assistance by Universiti Tun Hussein Onn Malaysia and the UTHM Publisher's Office via Publication Fund E15216.

References

[1] S. L. Ang, H. C. Ong, and H. C. Low, *Classification Using the General Bayesian Network*," Pertanika J. Sci. Technol., 24 (1) (2016) 205–211.

[2] G. Baio and M. Blangiardo, *Bayesian hierarchical model for the prediction of football result*s, J. Appl. Stat., (2010) 1–13.

[3] G. Bielza and P. Larrañaga, *Discrete bayesian network classifiers: A survey*, ACM Comput. Surv., 47 (1) (2014).
[4] W. Buntine, *Theory refinement on Bayesian networks*, Proc. Seventh Annu. Conf. Uncertain. Artif. Intell., (1991) 52–60.
[5] A. C. Constantinou, N. E. Fenton, and M. Neil, *Knowledge-Based Systems Profiting from an inefficient association football gambling market: Prediction , risk and uncertainty using Bayesian networks*, Knowledge-Based Syst., 50 (2013) 60–86.
[6] A. C. Constantinou, N. E. Fenton, and M. Neil, *Pi-football: A Bayesian network model for forecasting Association Football match outcomes*, Knowledge-Based Syst., 36 (2012) 322–339.
[7] A. C. Constantinou, *Dolores: a model that predicts football match outcomes from all over the world*, Mach. Learn., (2019) 1–27.
[8] G. F. Cooper and E. Herskovits, *A Bayesian method for the induction of probabilistic networks from data*, Mach. Learn., 9 (1992) 309–347.
[9] W. Dubitzky, P. Lopes, J. Davis, and D. Berrar, "The Open International Soccer Database for machine learning," Mach. Learn., 108 (1) (2019) 9–28.
[10] J. Fernandez, D. Medina, M. A. Gomez, and R. Gavalda, *From training to match performance: A predictive and explanatory study on novel tracking data*, IEEE Int. Conf. Data Min. Work., (2017) 136–143.
[11] E. Frank, M. A. Hall, and I. H. Witten, *The WEKA Workbench. Online Appendix for Data Mining: Practical Machine Learning Tools and Techniques*, Morgan Kaufmann, Fourth Ed., (2016).
[12] N. Friedman, M. Geiger, and M. Goldszmidt, *Bayesian network classifier*, Mach. Learn., 29 (2) (1997) 131–163.
[13] D. Geiger, *An entropy-based learning algorithm of Bayesian conditional trees*, Proc. Eighth Annu. Conf. Uncertain. Artif. Intell., (1992) 92–97.
[14] D. Heckerman, D. Geiger, and D. M. Chickering, *Learning Bayesian Networks: The Combination of Knowledge and Statistical Data*, Mach. Learn., 20 (3) (1995) 197–243.
[15] J. Heaton, *Quantifying the performance of individual players in a team activity*, Forecast. Futur., 7 (2013) 6–10.
[16] D. E. Holmes and L. C. (eds) Jain, *Innovations in Bayesian Networks: Theory and Applications*, Stud. Comput. Intell. Springer, 156 (2008).
[17] J. Ji, C. Yang, J. Liu, J. Liu, and B. Yin, *A comparative study on swarm intelligence for structure learning of Bayesian networks*, Soft Comput., 21 (22)(2017) 6713–6738.
[18] G. H. John and L. P., *Estimating Continuous Distributions in Bayesian Classifiers*, Elev. Conf. Uncertain. Artif. Intell., (1995) 338–345.
[19] A. Joseph, N. E. Fenton, and M. Neil, *Predicting football results using Bayesian nets and other machine learning techniques*, Knowledge-Based Syst., 19 (7) (2006) 544–553.
[20] A. S. Hesar, T. H., and M. Jalali, *Structure Learning of Bayesian Networks Using Heuristic Methods*, Pertanika J. Sci. Technol., 45 (2012) 246–250.
[21] M. G. Madden, *On the classification performance of TAN and General Bayesian Networks*, Knowledge-Based Syst., 22 (2) (2009) 295–489.
[22] B. Min, J. Kim, C. Choe, H. Eom, and R. I. (Bob) McKay, *A compound framework for sports results prediction: A football case study*, Knowledge-Based Syst., 21 (7) (2008) 551–562, 2008.
[23] A. Owen, *Dynamic Bayesian forecasting models of football match outcomes with estimation of the evolution variance parameter*, IMA J. Manag. Math., 22 (2) pp. 99–113, 2011.
[24] F. Owramipur, P. Eskandarian, and F. S. Mozneb, "Football Result Prediction with Bayesian Network in Spanish League-Barcelona Team," Int. J. Comput. Theory Eng., 5 (5) (2013) 812–815.
[25] J. Pearl, *Bayesian Networks: A Model of Self-Activated Memory for Evidential Reasoning*, Rep. No CSD-850021, (1985).
[26] N. Razali, A. Mustapha, S. Utama, and R. Din, *A Review on Football Match Outcome Prediction using Bayesian Networks*, J. Phys. Conf. Ser., 1020 (1) (2018).
[27] S. Singhal and M. Jena, *A study on WEKA tool for data preprocessing, classification and clustering*, Int. J. Innov. Technol. Explor. Eng., 2 (6) (2013) 250–253.
[28] P. Tufekci, *Prediction of football match results in Turkish super league games*, IEEE Int. Conf. Data Min. Work., 427 (2016) 515–526.