AN ENHANCED BAYESIAN NETWORK PREDICTION MODEL FOR FOOTBALL MATCHES BASED ON PLAYER PERFORMANCE

MUHAMMAD NAZIM RAZALI

A thesis submitted in fulfillment of the requirement for the award of the
Master of Information Technology

Faculty of Computer Science and Information Technology
Universiti Tun Hussein Onn Malaysia

DECEMBER 2017

In the name of Allah, Most Gracious, The Most Merciful.
All praise is due to Allah. May Allah sends His peace and blessings upon the beloved Messenger (S.A.W) and upon his companions, his family, and all who follow his guidance until the Day of Judgement.
Special dedication to my beloved mother, late father, siblings, sister in law and niece, Pn. Hjh. Rose binti Hashim, Tn. Hj. Razali bin Hj. Atan, Mohd. Najib, Najihah, Nazirah, Faradila and Maryam. Thank you for your love, encouragement, prayers and everything since this journey begins as days by, passed by weeks went by, and years gone by, until the final stages. You guys are the main reason for me to always striving to do the best.
This thesis is dedicated to all of you.

ACKNOWLEDGEMENT

All praises to Allah for the strengths and His blessings in completing this thesis. At this moment of accomplishment, first of all, I would like to express my special appreciation to my supervisor, Assoc. Prof. Dr. Aida Mustapha and my co-supervisor, Dr. Mohamad Aizi Salamat for their guidance and constant support from the initial to the final stage, for their patience, motivation, enthusiasm, and immense knowledge. Their invaluable help of constructive comments and suggestions throughout the experimental and thesis works have contributed to the success of this research.

I take this opportunity to sincerely acknowledge the Universiti Tun Hussein Onn Malaysia (UTHM) and the Ministry of Higher Education Malaysia under the Research Acculturation Collaborative Effort (RACE) grant scheme Vot 1513 for providing financial assistance for my research work. I would also like to express my gratitude to the Centre for Graduate Studies (CGS), the Research, Innovation, Commercialisation and Consultancy Office (ORICC) and the Faculty of Computer Sciences and Information Technology (FSKTM) for their hospitality and support towards my postgraduate affairs. A credit is also due to the BayesFusion, LLC where the implementation of this work is based on GeNIe Modeler, a graphical interface for SMILE, a Bayesian inference engine developed at the Decision Systems Laboratory and available at `https://www.bayesfusion.com/`.

I am also extremely indebted to my beloved families, for their endless love, prayers and encouragements. My special gratitude is due to all my fellow friends, especially postgraduate members for the guidance and encouragement during my study. Thanks for the friendship and memories. Last but not least, it is a pleasure to thank all those who have helped either directly or indirectly. I would not have accomplished this work without the help and support from them. I really appreciate their helped. May Allah bless us with love, luck and prosperity. Insha'Allah. Ameen.

ABSTRACT

In sports analytics, existing researches have showed that the Bayesian networks (BN) approach has greatly contributed to predicting football match results with considerably high accuracy as compared to other classical statistical and machine learning approaches. However, existing prediction models rely solely on historical team features including the match statistical data as well as team statistical data, together with the historical features of team achievement such as ranking in FIFA, ranking in league and total number of points gained at the end of a season. There is no known work to date that has analysed individual player performance data as part of the parameters used to predict football match results. To address this gap, this research proposes a BN model for match prediction based on player performance data called the Player Performance (PP) model. To validate the performance of the proposed PP model, three existing prediction models were re-implemented and measured for prediction accuracy. The existing models are the General Individual (GI) model, Match Statistical (MS) model, and Team Statistical (TS) model. All BN models were constructed using the Tree Augmented Naive Bayes (TAN) for structural learning. The dataset used was data for the Arsenal Football Club in the English Premier League (EPL) for seasons 2014-2015 and 2015-2016. Apart from the proposed individual player performance data, the dataset includes individual player rating, absence or presence of players in a match, match statistics, and team statistics. Then, the PP model were re-constructed using other machine learning techniques such as k-Nearest Neighbour (kNN) and Decision Tree (DT) in order to compare with BN for prediction accuracy. The experimental results showed two fold; the proposed PP model using BN achieved a higher accuracy in predicting the outcomes for football matches with an overall average predictive accuracy of 63.76% compare to GI model, MS model and TS model as well as higher than PP model using kNN and DT by 1.64% and 6.02%.

ABSTRAK

Kajian terdahulu dalam analisis sukan menunjukkan bahawa pendekatan Bayesian Networks (BN) telah banyak menyumbang kepada peramalan keputusan perlawanan bola sepak dengan kadar ketepatan yang sangat tinggi berbanding dengan pendekatan yang lain seperti pendekatan statistik yang klasik dan pendekatan pembelajaran mesin. Walau bagaimanapun, kebanyakan model ramalan yang sedia ada sangat bergantung kepada ciri-ciri sejarah pasukan termasuk data statistik perlawanan, data statistik pasukan dan ciri-ciri sejarah pencapaian pasukan seperti kedudukan pencapaian pasukan dalam FIFA, kedudukan pencapaian dalam liga dan jumlah bilangan mata yang diperolehi oleh pasukan pada akhir musim. Namun begitu, tiada kajian yang diketahui dalam menganalisis data prestasi pemain secara individu sebagai sebahagian daripada parameter dalam meramalkan keputusan perlawanan bola sepak. Untuk menangani jurang ini, kajian ini telah mencadangkan sebuah model BN yang berasaskan data prestasi pemain untuk meramal keputusan perlawanan yang diberi nama sebagai model Prestasi Pemain (PP). Untuk mengesahkan prestasi model PP yang dicadangkan itu, tiga model ramalan yang sedia ada telah dilaksanakan semula dan diukur ketepatan ramalannya. Model-model yang sedia ada ialah model yanng berasaskan Individu Umum(GI), model Statistik Perlawanan (MS), dan model Statistik Pasukan (TS). Semua model BN telah dibina menggunakan *Tree Augmented Naive Bayes (TAN)* untuk pembelajaran struktur. Sampel data yang digunakan diperolehi daripada Kelab Bolasepak Arsenal yang beraksi dalam Liga Perdana Inggeris (EPL) bagi musim 2014-2015 dan 2015-2016. Selain daripada data prestasi pemain secara individu yang dicadangkan, sampel data yang lain termasuk kehadiran pemain secara individu di dalam sesebuah perlawanan, data statistik perlawanan, dan data statistik pasukan turut digunakan. Kemudian, model PP telah dibina kembali dengan menggunakan teknik-teknik pembelajaran mesin seperti *Decision Tree (DT)* dan *k-Nearest Neighbour (kNN)*. Keputusan eksperimen menunjukkan 2 keputusan dimana model PP yang dicadangkan mencapai ketepatan yang lebih tinggi dalam meramalkan keputusan untuk perlawanan bola sepak dengan kadar purata ketepatan sebanyak 63.76% berbanding model GI, model MS dan model TS manakala Model PP berasaskan BN juga mencapai kadar purata ketepatan lebih tinggi daripada PP model berasasskan *Decision Tree* dan *k-Nearest Neighbour* dengan sebanyak 1.64% and 6.02%.

CONTENTS

LIST OF TABLES

LIST OF FIGURES

LIST OF ABBREVIATIONS

AFC	Asian Football Confederation
AI	Artificial Intelligence
Arsenal F.C	Arsenal Football Club
BN	Bayesian Networks
CPT	Conditional Probability Table
DAG	Directed Acyclic Graph
DT	Decision Tree
EM	Expected Maximization
EPL	English Premier League
FIFA	Internation Federation of Association Football
GI	General Individual Model
kNN	k-Nearest Neighbor
MLE	Maximum Likelihood Estimates
MLP	Multi Layer Perceptron Neural Networks
MS	Matchs Statistical Model
NN	Neural Networks
PP	Player Performance Model
PPI	Player Performance Index
TAN	Tree Augmented Naive Bayes
TS	Team Statistical Model
UEFA	Union of European Football Association
WEKA	Waikato Environment Knowledge Analysis

LIST OF APPENDICES

LIST OF PUBLICATIONS

Nazim Razali, Aida Mustapha, Faiz Ahmad Yatim, Ruhaya Ab Aziz. (2017). Predicting Player Position for Talent Identification in Association Football. In *Proceedings of the International Conference on Advances in Computing and Intelligent System* (ICACIS 2017), 6 May 2017, Melaka, Malaysia. (Indexed by Scopus)

Nazim Razali, Aida Mustapha, Faiz Ahmad Yatim, Ruhaya Ab Aziz. (2017). Predicting Football Matches Results using Bayesian Networks for English Premier League (EPL). In *Proceedings of the International Conference on Advances in Computing and Intelligent System* (ICACIS 2017), 6 May 2017, Melaka, Malaysia. (Indexed by Scopus)

Nazim Razali, Aida Mustapha, Sunariya Utama, Roshidi Din (2017). A Review on Football Match Outcome Prediction using Bayesian Networks. In *Proceedings of the 1st International Conference on Computing, Technology Science and Management in Sports* (ICoTSM 2017), 25 - 27 November 2017, Sarawak, Malaysia. (Indexed by Scopus)

Che Mohamad Firdaus Che Mohd Rosli, Mohd Zainuri Saringat, Nazim Razali, Aida Mustapha (2017). A Comparative Study of Data Mining Techniques on Football Match Prediction. In *Proceedings of the 1st International Conference on Computing, Technology Science and Management in Sports* (ICoTSM 2017), 25 - 27 November 2017, Sarawak, Malaysia. (Indexed by Scopus)

Mohamad Fahim Mohamed Nasir, Suriawati Suparjoh, Nazim Razali, Aida Mustapha (2017). P-Soccer: Soccer Games Application using Kinect. In *Proceedings of the 1st International Conference on Computing, Technology Science and Management in Sports* (ICoTSM 2017), 25 - 27 November 2017, Sarawak, Malaysia. (Indexed by Scopus)

LIST OF AWARDS

Nazim Razali, Fawwaz Zamir Mansor, Syahdan Mahad Hamzah. (2017). Pattern Analysis of Goals Scored for Association Football. *Silver Award*. Data Analytic Industry Challenge in 2017 CREST Industry Data Analytical Challenge (Southern Region).

Nazim Razali, Aida Mustapha, Sunariya Utama, Roshidi Din (2017). A Review on Football Match Outcome Prediction using Bayesian Networks. *Best Paper Award*. In *Proceedings of the 1st International Conference on Computing, Technology Science and Management in Sports* (ICoTSM 2017), 25 - 27 November 2017, Sarawak, Malaysia. (Indexed by Scopus)

CHAPTER 1

INTRODUCTION

1.1 Background of Study

Association football, more commonly known as football (British English) or soccer (American English), is one of the most popular sports in the world. People all around the globe from all ages enjoy playing or watching the sport, whether at the amateur or professional level. Nelson Mandela, the former president of South Africa (1918-2013), was quoted to having said that football has the power to inspire and unite people. Football enjoys great popularity and has a special place in the hearts of people. Every four years since 1930, there will be a prestigious international football competition between qualified nations called the FIFA World Cup. During that time, people all around the world without background discrimination will be discussing the performance of their team of choice. This has led to the rise of discussion and researches on sports analytics based on association football in order to measure every aspect of the match, whether at pre-match or post-match.

Research in sports analytics for association football advances in two directions; performance analysis and match prediction. Performance analysis depends on the performance level, whether as an individual or a team. For individual performance analysis, such analyses are invaluable in discovering the principal of the observed outcome; win a match or score goals. In a network theory analysis of football strategies, some of the teams which participated in the 2010 World Cup were analysed in terms of team passes. The results produced by Pena & Touchette (2012) were the closeness (how well connected is a player), betweenness (how dependent the team is on the player), page rank (how important a player is), and clustering. The limitation of the analysis is that it did not consider the defensive strength of the opponents. It also did not focus on passing accuracy and shot analysis, which are fundamental in a football match.

Team performance was evaluated in Cotta *et al.* (2013). In their research, the additional results produced were match speed (passes per minute) and offensive elaborateness (consecutive passes without losing the ball) in addition to the analysis by Pena & Touchette (2012). Goal scoring performance by season was also previously analysed by Castillo *et al.* (2011) using the Poisson Sampling and Negative Binomial Sampling which were then treated with the Bayesian approaches. The data used for this analysis were sourced from the Spanish football league from season 2000/2001 to season 2008/2009. The features used include the number of goals, minutes played, position, team rank, and goal scoring abilities obtained from the Poisson sampling. The results were then presented in terms of probability, whereby the higher the probability value, the better the player has performed for the season.

Meanwhile, match prediction is concerned with two aspects. The first is to serve as a benchmark for evaluating the chances of winning before the match begins. Match prediction is necessary to help the managers and players prepare countermeasures such using defensive mentality and offside trap if the opponent team is stronger than them. Second is in terms of profit generation for betting, which is beyond the scope of this research. The main focus for this research is to verify and justify whether the individual player performance attribute gives impact towards a match's outcome (win, lose and draw) more than the other attributes in a football match.

Most of the existing match prediction models are solely based on post-match team data such as team results, goals scored and goals conceded. Hence, not many of them acknowledge the real contribution of team performance, which are the players. The lack of a prediction model based on individual players' data is caused by the lack of the data itself. The expected outcome of this research is a model to predict football matches' outcomes based on individual performance data. It is also expected that the proposed model would have higher prediction accuracy than existing football prediction models because the model is sensitive to different team composition.

1.2 Problem Statement

Sports analytics, in general, focuses on match prediction or performance analysis, whether as an individual player or as a team including opponents. One particular analytics that draws wide attention to football is match prediction. Football match prediction can be classified into two models, which are ex-post model (after complete set of data for a season have been collected) and ex-ante model (before the match is played). For example, Maher (1982) created an ex-post model to estimate a team's capability for offensive and defensive strength. This means that he carried out an independent Poison Model based on the home team and away team scores in order to estimate both the offensive and defensive strength using a full set of data obtained for a season. The model was incapable of predicting matches that were not yet played. Dixon & Coles (1997) enhanced the Maher model through a few adjustments, making it into an ex-ante model so that predictions can take place before a match begins.

Predicting outcomes for future football matches is not easy because there are many important variables that need to be analysed before the match such as past meeting results with opponents and ranking of the team and its opponents. Yet, there are some important variables that cannot be observed until a match is played. For example, injuries, booked and psychological effects can only be observed during a match. Determination has a big influence over a match's outcome when one team dominate a match and play attack during the entire match. There are many uncertainties that need to be considered in the prediction of the outcomes of football matches.

Existing football match prediction models were developed using various approaches, such as the statistical approach, machine learning approach and Bayesian approach. Nonetheless, all of the models rely on team features such as match venue, team goals scored, team goals conceded, opponent goals scored, and opponent goals conceded in order to evaluate a team's outcomes. Other researches use the past features of a team's achievement such as its league ranking and points gained at the end of the season to construct the prediction model. This research scenario is common because the historical data for team features and team achievement are mostly readily accessible via newspapers or website compared to other types of historical data such as match features for each individual player.

To date, individual player data has not been fully capitalised as the main features in building a prediction model. Although there are researches using individual football player performance such as in (Duch *et al.*, 2010; McHale & Scarf, 2005; McHale *et al.*, 2012; Pena & Touchette, 2012), there is a wide range of individual player data such as a player's rating and the total number of shoots, passes, tackles and saves that have yet to be analysed. The closest works that used individual player data would be researches by Constantinou *et al.* (2012) and Joseph *et al.* (2006), but they only focused on the presence or absence of the selected first eleven players, the position of key players and the availability of three key players.

Langseth (2013) pointed out that harvesting more match data can generate even richer prediction models for football. He improvised the work of Rue & Salvesen (2000) by adding player statistic data and the total number of shots fired and shots on target which outperformed other traditional statistical football prediction model (Maher, 1982; Dixon & Coles, 1997; Rue & Salvesen, 2000) which are based on a team's past data. Other individual player data such as total number of shots, goals, assists, passes, tackles and saves cannot be ignored; hence it is formulated and incorporated in the match outcome prediction model.

1.3 Research Objectives

The research objectives are as follows:

i. To develop a prediction model for football matches outcome (win,lose and draw) using Bayesian networks (BN) based on individual player performance.

ii. To implement the Bayesian networks model based on (1) to predict the football match outcomes (win, lose and draw).

iii. To compare the accuracy in terms of football match outcomes in (2) against other machine learning techniques.

1.4 Scope of Study

This research is limited to performance data of individual players in the English Premier League (EPL) as well as historical data of football matches for the Arsenal Football Club (F.C) between seasons 2014-2015 and 2015-2016 as well as all 20 competing teams in EPL. Data was extracted from `https://www.premierleague.com/` `https://www.whoscored.com/` and `http://www.squawka.com/` (McHale *et al.*, 2012; Alberti *et al.*, 2013; Langseth, 2013; Constantinou & Fenton, 2017). This football websites were choose because the data have same data source which are supplied by Opta (`http://www.optasports.com/`), the world's leading sports data provider. There are 20 competing teams in the English Premier League and every single team faces the 19 other teams twice a season. Therefore, in total there will be 38 matches for a season played by Arsenal F.C.

1.5 Significance of Study

According to the FIFA's Big Count in 2016 through FIFA official website (`http://www.fifa.com/worldfootball/bigcount/allplayers.html`), as of 3 May 2016, Malaysia has 585,730 players in which 9,930 are registered and the remaining 575,800 are unregistered players. This makes up less than one percent of the 265,000,000 football players across the world. Malaysia also hosts 110 clubs and a total of 11,810 officials in football matches. Thus, the findings of the study will contribute greatly to sports development especially for association football or soccer in our country with need a touch in order to improve our teams ranking, please our football fans and cool down the criticism towards our football national team. The proposed model based on player performance is hoped to help coaches evaluate and estimate their players' performance individually in order to create a more accurate network structure between players and the match outcomes (win, lose and draw). As a result, these findings are invaluable for developing and forming teams with high winning probabilities despite detect non-performing players as well as new talents.

In the state of the art for computer science, the football prediction model using Bayesian networks may provide an analysis on the network of probability in term of Directed Acyclic Graph (DAG) and Conditional Probability Table (CPT) accordingly to total number of data row and the data features. Besides, this research will show the comparative accuracy results between Bayesian networks and other machine learning techniques such as k-Nearest Neighbour (kNN) and Decision Tree (DT) in prediction. It is not just helping the football team to predict the match outcome but also help assist team to identify what data features that correlates in football matches which may contribute to match outcomes. As the result, it may help managerial team staffs and coaches a fresh insight for organising better tactics and forming a better first eleven squad for each match.

From the economic perspective, team managers from clubs are able to identify and acquire excellent players undervalued by the market, hence minimising the cost of purchasing talent. Furthermore, by acquiring excellent players, the club's improved performance on the field would lead to improved revenue from stadium attendance and merchandise sold, thus leading to national economic growth. Besides, this new prediction model can be used to set benchmarks which may assist managers and players to act with more caution and prepare alternative options to beat opponents that are stronger than them such as setting up offside traps, implement defensive mentality and play counter attack if the chance come.

1.6 Organisation of the Thesis

This chapter presents the basis of using sports analytics in prediction of association football matches by explaining the background of study, research motivation, problem statement, research objectives, scope of study and significance of study. The rest of the chapters in this thesis are organised as follows:-

- Chapter 2 discusses in general the overview of Bayesian Networks (BN) and existing association football prediction models, which are divided into three approaches; statistical approaches, machine learning approaches and Bayesian approaches. Later, a comparative analysis of all prediction models will summarise the entire prediction model.
- Chapter 3 discusses about the research methodology and the proposed research prediction model based on player performance using the Bayesian networks. The chapter also includes a description of the indicator of player performance which will be used in this research. It also discusses the important processes in the prediction model, the dataset used and a description and implementation of existing prediction models, which will be executed using BN and re-constructed prediction based on player performance using k-Nearest Neighbour (kNN) and Decision Tree (DT).
- Chapter 4 discusses the results of the new prediction model based on player performance and compares the accuracy of the developed model with some other existing model in term of features and techniques.
- Chapter 5 finally concludes the research, explains the limitations of the proposed model and provides suggestions and recommendations for future works.

CHAPTER 2

LITERATURE REVIEW

2.1 Introduction

In 1956, Artificial Intelligence (AI) began its first move when it was proposed by John McCarthy at a conference in Dartmouth. Although, AI was only officially introduced in 1956, AI has actually been discussed by researchers several years before that. For example, a simple question, “Can machines think?” had been raised by Turing (1950) in his paper entitled “Computing Machinery and Intelligence”. AI is a combination of two terms, machine and human being, which form an intelligent mechanisms to deal with real world problem. Human intelligence in a machine can be in the form of a robot, computer or any other electronic equipment (Nehra, 2015). These intelligent mechanisms contribute greatly to human being in terms of problem-solving, knowledge, reasoning, learning, communication, robotics and uncertainty. As the time flows, the AI field keeps expanding and developing whether in practical perspective as well in theoretical perspective. Even in sports analytics, the machine learning which parts of AI have been use for assisting the football team to achieve their objective (Joseph *et al.*, 2006).

In sports, researches on football analytics have grown rapidly due to the popularity of the sport itself. Sports analytics can be classified into two categories which are performance-related (team and individual) or predictive analysis. Football analytics is crucial to clubs management teams, managers, coaches and players themselves in order to help them win more matches and become better footballers. Football in particular is a sport full of uncertainty. Anything can happen within 90

REFERENCES

Abdalla Alfaki, I. M. (2014). Assessment and Dynamic Modeling of the Size of Technology Transfer. *Journal of the Knowledge Economy*, *7*(2), 600–612.

Alamaniotis, M., Bargiotas, D., & Tsoukalas, L. H. (2016). Towards smart energy systems: application of kernel machine regression for medium term electricity load forecasting. *SpringerPlus*, *5*(1), 1–15.

Alberti, G., Iaia, M. F., Arcelli, E., Cavaggioni, L., & Rampinini, E. (2013). Goal scoring patterns in major European soccer leagues. *Sport Sciences for Health*, *9*(3), 151–153.

Armatas, V., & Pollard, R. (2014). Home advantage in Greek football. *European Journal of Sport Science*, *14*(2), 116–122.

Armatas, V., Yiannakos, A., Papadopoulou, S., & Skoufas, D. (2009). Evaluation of goals scored in top rankings soccer matches: Greek Superleague 2006-07. *Serbian Journal of Sports Sciences*, *3*(1), 39–43.

Baio, G., & Blangiardo, M. (2010). Bayesian hierarchical model for the prediction of football results. *Journal of Applied Statistics*, pp. 1–13.

Baker, R. D., & McHale, I. G. (2013). Forecasting exact scores in National Football League games. *International Journal of Forecasting*, *29*(1), 122–130.

Boulier, B. L., & Stekler, H. (2003). Predicting the outcomes of National Football League games. *International Journal of Forecasting*, *19*(2), 257–270.

Bro, R., Kamstrup-Nielsen, M. H., Engelsen, S. B., Savorani, F., Rasmussen, M. A., Hansen, L., Olsen, A., Tjønneland, A., & Dragsted, L. O. (2015). Forecasting individual breast cancer risk using plasma metabolomics and biocontours. *Metabolomics : Official journal of the Metabolomic Society*, *11*(5), 1376–1380.

Castillo, S., Rodriquezi, A., & Perez-Sanchez, J. (2011). Expected number of goals depending on intrinsic and extrinsic factors of a football player. An application to professional Spanish football league. *European Journal of Sport Science*, pp. 1–12.

Catal, C. (2012). Performance evaluation metrics for software fault prediction studies. *Acta Polytechnica Hungarica*, *9*(4), 193–206.

Cattelan, M., Varin, C., & Firth, D. (2013). Dynamic Bradley Terry modelling of sports tournaments. *Journal of the Royal Statistical Society: Series C (Applied Statistics)*, *62*(1), 135–150.

Constantinou, A., & Fenton, N. (2017). Towards smart-data: Improving predictive accuracy in long-term football team performance. *Knowledge-Based Systems*, *124*, 93–104.

Constantinou, A. C., & Fenton, N. E. (2013). Determining the level of ability of football teams by dynamic ratings based on the relative discrepancies in scores between adversaries. *Journal of Quantitative Analysis in Sports*, *9*(1), 37–50.

Constantinou, A. C., Fenton, N. E., & Neil, M. (2012). Pi-football: A Bayesian network model for forecasting Association Football match outcomes. *Knowledge-Based Systems*, *36*, 322–339.

Constantinou, A. C., Fenton, N. E., & Neil, M. (2013). Knowledge-based systems profiting from an inefficient association football gambling market: Prediction, risk and uncertainty using Bayesian networks. *Knowledge-Based Systems*, *50*, 60–86.

Cotta, C., Mora, A., Merelo, J.-J., & Merelo-Molina, C. (2013). A Network Analysis of the 2010 FIFA World Cup Champion Team Play. *Journal of Systems Science and Complexity*, *26*(1), 21–42.

Crowder, M., Dixon, M., Ledford, A., & Robinson, M. (2002). Dynamic modelling and prediction of English Football League matches for betting. *Journal of the Royal Statistical Society Series D: The Statistician*, *51*(2), 157–168.

Dixon, M. J., & Coles, S. G. (1997). Modelling Association Football Scores and Inefficiencies in the Football Betting Market. *Journal of the Royal Statistical*

Society. Series C: Applied Statistics, *46*(2), 265–280.

Do, C. B., & Batzoglou, S. (2008). What is the expectation maximization algorithm? *Nature Biotechnology*, *26*, 897–899.

Duch, J., Waitzman, J. S., & Nunes Amaral, L. A. (2010). Quantifying the performance of individual players in a team activity. *PLoS ONE*, *5*(6), 1–7.

Efron, B. (2010). The Jackknife, the Bootstrap and other resampling plans. *In CBMS-NSF regional conference series in applied mathematics 1982. Philadelphia, PA: Society for industrial and Applied Mathematics (SIAM).*

Friedman, N. (1998). The Bayesian structural EM algorithm. *In Proceedings of the 14th Conference on Uncertainty in Artificial Intelligence, UAI-98.*

Friedman, N., Geiger, D., & Goldszmidt, M. (1997). Bayesian network classifier. *Machine learning*, *29*(2), 131–163.

Gerrard, B. (2014). Sports Analytics: A Guide for Coaches, Managers and Other Decision Makers, B.C. Alamar. Columbia University Press, New York (2013). *Sport Management Review*, *17*(2), 240–241.

Gilsdorf, K. F., & Sukhatme, V. A. (2008). Testing Rosen's Sequential Elimination Tournament Model: Incentives and Player Performance in Professional Tennis. *Journal of Sports Economics*, *9*(3), 287–303.

Glickman, M. (1995). The glicko system. *Boston University*, pp. 1–5.

Goddard, J., & Asimakopoulos, I. (2004). Forecasting Football Results and the Efficiency of Fixed-odds Betting. *Journal of Forecasting*, *23*(1), 51–66.

Heaton, J. (2013). Quantifying the performance of individual players in a team activity. *Forecasting & Futurism 7*, pp. 6–10.

Heckerman, D., Geiger, D., & Chickering, D. (1995). Learning Bayesian Networks: The Combination of Knowledge and Statistical Data. *Machine Learning*, *20*(3), 197–243.

Hesar, A. S., Tabatabaee, H., & Jalali, M. (2012). Structure Learning of Bayesian Networks Using Heuristic Methods. *International Conference on Information and Knowledge Management*, *45*, 246–250.

Holmes, D., & Jain, L. E. (2008). Innovations in Bayesian Networks: Theory and

Application. *Studies in Computational Intelligence 156. Berlin: Springer-Verlag Berlin Heidelberg*.

Huang, K.-y., Member, S., & Chang, W.-l. (2010). A Neural Network Method for Prediction of 2006 World Cup. *The 2010 International Joint Conference on Neural Networks (IJCNN)*, pp. 18–23.

Hvattum, L. M., & Arntzen, H. (2010). Using ELO ratings for match result prediction in association football. *International Journal of Forecasting*, *26*(3), 460–470.

Joseph, A., Fenton, N. E., & Neil, M. (2006). Predicting football results using Bayesian nets and other machine learning techniques. *Knowledge-Based Systems*, *19*(7), 544–553.

Karlis, D., & Ntzoufras, I. (2009). Bayesian modelling of football outcomes: Using the Skellam's distribution for the goal difference. *IMA Journal of Management Mathematics*, *20*(2), 133–145.

Kersting, K., & Raedt, L. (2000). Bayesian Logic Programs (Report No. 151).

Koopman, S. J., & Lit, R. (2015). A dynamic bivariate Poisson model for analysing and forecasting match results in the English Premier League. *Journal of the Royal Statistical Society: Series A (Statistics in Society)*, *178*(1), 167–186.

Lago-Peñas, C., Gómez-ruano, M., & Megías-navarro, D. (2016). Home advantage in football : Examining the effect of scoring first on match outcome in the five major European leagues. *International Journal of Performance Analysis in Sport*, *16*(2), 411–421.

Langseth, H. (2013). Beating the bookie : A look at statistical models for prediction of football matches. *Frontiers in Artificial Intelligence and Applications*, *257*, 165–174.

Lasek, J., Szlavik, Z., & Bhulai, S. (2013). The predictive power of ranking systems in association football. *International Journal of Applied Pattern Recognition*, *1*(1), 27–46.

Leitner, C., Zeileis, A., & Hornik, K. (2010). Forecasting sports tournaments by ratings of (prob)abilities: A comparison for the EURO 2008. *International Journal of Forecasting*, *26*(3), 471–481.

Li, E. Y., Tung, C.-Y., & Chang, S.-H. (2016). The wisdom of crowds in action: Forecasting epidemic diseases with a web-based prediction market system. *International Journal of Medical Informatics*, *92*, 35–43.

Madden, M. (2009). On the Classification Performance of TAN and General Bayesian Networks. *In: Bramer M., Petridis M., Coenen F. (eds) Research and Development in Intelligent Systems XXV. Springer, London*.

Maher, M. (1982). Modelling association football scores. *Statistica Neerlandica*, *36*(3), 109–118.

Manner, H. (2016). Modeling and forecasting the outcomes of NBA basketball games. *Journal of Quantitative Analysis in Sports*, *12*(1), 31–41.

Matsuo, Y. (2003). Prediction, Forecasting, and Chance Discovery. *Chance Discovery*, pp. 30–43.

McHale, I., & Scarf, P. (2005). Ranking football players. *Significance*, *2*(2), 54–57.

McHale, I. G., Scarf, P. A., & Folker, D. E. (2012). On the development of a soccer player performance rating system for the english premier league. *Interfaces*, *42*(4), 339–351.

Min, B., Kim, J., Choe, C., Eom, H., & (Bob) McKay, R. I. (2008). A compound framework for sports results prediction: A football case study. *Knowledge-Based Systems*, *21*(7), 551–562.

Nehra, E. (2015). Artificial Intelligence in modern time. *International Journal of Science, Technology and Management*, *4*(1), 112–118.

Owen, A. (2011). Dynamic Bayesian forecasting models of football match outcomes with estimation of the evolution variance parameter. *IMA Journal of Management Mathematics*, *22*(2), 99–113.

Pan, B., Demiryurek, U., Gupta, C., & Shahabi, C. (2014). Forecasting spatiotemporal impact of traffic incidents for next-generation navigation systems. *Knowledge and Information Systems*, *45*(1), 75–104.

Pearl, J. (1985). Bayesian Networks: A Model of Self-Activated Memory for Evidential Reasoning (No: CSD-850021).

Pena, J., & Touchette, H. (2012). A network analysis of football strategies. *Sports Physics: Proc. 2012 Euromech Physics of Sports Conference*, pp. 517–528.

Rotshtein, A. P., Posner, M., & Rakityanskaya, A. B. (2005). Football predictions based on a fuzzy model with genetic and neural tuning. *Cybernetics and Systems Analysis*, *41*(4), 619–630.

Rue, H., & Salvesen, O. (2000). Prediction and retrospective analysis of soccer matches in a league. *Journal of the Royal Statistical Society: Series D (The Statistician)*, *49*(3), 399–418.

Russell, S., John, B., & Kanazawa, K. (1995). Local learning in probabilistic networks with hidden variables. *In Proceedings of the 14th International Joint Conference on Artificial Intelligence, Montreal*, pp. 1146–1152.

Scheibehenne, B., & Bröder, A. (2007). Predicting Wimbledon 2005 tennis results by mere player name recognition. *International Journal of Forecasting*, *23*(3), 415–426.

Seckin, A., & Pollard, R. (2008). Home advantage in turkish professional soccer. *Perceptual and Motor Skills*, *107*(1), 51–54.

Seni, G., & Elder, J. (2010). Ensemble Methods in Data Mining: Improving Accuracy Through Combining Predictions . *Synthesis Lectures on Data Mining and Knowledge Discovery*.

Shao, H., & Deng, X. (2016). Short-term wind power forecasting using model structure selection and data fusion techniques. *International Journal of Electrical Power & Energy Systems*, *83*, 79–86.

Sismanis, Y. (2010). How I won the ”Chess Ratings - Elo vs the Rest of the World” Competition. *arXiv:1012.4571 [cs.LG]*, pp. 1–8.

Stefani, R. (1977). Football and Basketball Predictions Using Least Squares. *IEEE Transactions on Systems, Man, and Cybernetics*, *7*(2), 117–121.

Stefani, R. (1980). Improved least squares football, basketball, and soccer predictions. *IEEE Transactions on Systems, Man, and Cybernetics*, *10*(2), 116–123.

Tucker, W., Mellalieu, S. D., James, N., & Taylor, J. B. (2005). Game Location Effects in Professional Soccer: A Case Study. *International Journal of Performance*

Analysis in Sport, *5*(2), 23–35.

Turing, A. (1950). Computing machinery and intelligence. *Mind*, *59*, 433–460.

Wang, J., Song, Y., Liu, F., & Hou, R. (2016). Analysis and application of forecasting models in wind power integration: A review of multi-step-ahead wind speed forecasting models. *Renewable and Sustainable Energy Reviews*, *60*, 960–981.

Yu, T., Xiang, L., & Wu, D. (2016). Grey system and BP neural network model for industrial economic forecasting. *Recent Patents on Computer Science*, *9*(1), 40–45.