Predicting the Outcome of a Football Game: A Comparative Analysis of Single and Ensemble Analytics Methods

Enes Eryarsoy
Istanbul Sehir University
eneseryarsoy@sehir.edu.tr

Dursun Delen
Oklahoma State University
dursun.delen@okstate.edu

Abstract

As analytical tools and techniques advance, increasingly large numbers of researchers apply these techniques on a variety of different sports. With nearly 4 billion followers, it is estimated that association football, or soccer, is the most popular sports for fans across the world by a large margin. The objective of this study is to develop a model to predict the outcomes of soccer (or association football) games (win-loss-draw), and determine factors that influence game outcomes. We used 10 years of comprehensive game-level data spanning the years 2007-2017 in the Turkish Super League, and tested a variety of classifiers to identify the most promising methods for outcome predictions.

1. Introduction

Value generating management of both structured and unstructured data in sports falls into the vast field of sport analytics. Application of predictive analytics has been used successfully on many different sports such as football [1], basketball [2, 3], cricket [4], rugby [5], and hockey [6]. Predicting outcomes of any sports game naturally is one of the most obvious objectives in sports analytics. However, outcome prediction of almost any sports game is a challenging task due to their dynamically changing and stochastic nature. Many stakeholders, such as odds traders, fans, or team managers are interested in deploying such methods not merely to predict the outcomes but even more to understand underlying driving factors behind success or failure.

Association football (for simplicity hereafter will be referred as "soccer") has not only been the world's most popular sport [7, 8, 9] but also, not surprisingly, the largest in the sports betting market. Due to advances in

analytics, during recent years there has been a surge of interest in predicting outcomes of a soccer game. Fans, sports gamblers, managers, researchers, software vendors are all interested in accurately predicting soccer games.

Even though game prediction in soccer, in its most basic format, has three outcomes (win-loss-draw), there are many other outcomes that may be of interest for prediction. Most popular soccer wagers involve not only the outcome, but also predicting the number of goals, number of corners, free kicks, or even cards. The club managers are also interested in analytics in order to grow their fan bases or their game audience.

However, we believe that the number of studies for soccer game outcome prediction is low due primarily to the following reasons:

i) Lack of data available: the majority of papers make use of limited data.

ii) Sports' dynamic nature: Soccer outcome prediction still remains in analytical backwaters because of the sport's highly dynamic nature. A "win-loss-draw" game is much more difficult to predict a "win-loss" game. Also, each tournament - such as playoffs, regular seasons, or European championship league- presents different modeling challenges.

iii) Authors' believe that most of the "Moneyballing" efforts are put by club managers, and sports gamblers who are not willing to publish their results.

The rest of this paper is organized as follows. In Section 2, we review the existing academic literature. In Section 3, we outline our methodology and our dataset. Section 4 is dedicated to our analysis and results.

2. Literature Review

Academic literature on soccer game outcome predictions can be roughly categorized based on i) the kind of data used, ii) prediction stage (i.e. during or

HİCSS

before the game, or even season), iii) the type of outcome to be predicted, and iv) technique employed.

i) Kind of data used: Most of the data used in sports analytics are structured data. The majority of studies use structured game/player data [10], or structured odds data based on past betting quotas [11]. However, there are also studies that make use of unstructured data, such as sentiment analysis on tweets [12], or Tumblr posts [13].

ii) Prediction stage: Studies also differ in terms of the prediction window used. While some studies focus on studying live data, such as player trajectories, to assess player performances [14, 15], others make use of the data from the first half of game to predict the outcome for the second half. This is due to the fact that the betting window is still open at halftime. Our study uses all data available up to the game week.

iii) The type of outcome to predict: Different studies attempt to predict different kinds of outcomes such as the number of goals scored [16, 17, 18], the outcome of the game directly in terms of "win-loss-draw" [19], the measurement of the efficiencies or inefficiencies of the betting market [20], soccer tipsters' behavior and performance [21].

iv) Techniques used in the study: Different studies used different techniques for outcome prediction. While the majority of earlier studies employed methods from statistics and probability, such as Poisson distributions [22], Markov Chain Monte Carlo iterative simulations [23] discrete choice regression models for "win-loss-draw" scenarios [18, 19, 21] Newer studies tend to use data mining methods such as Naïve Bayes [11, 24], Bayesian Belief Networks [25], Support Vector Machines [26], Neural and Genetic optimization [26] or combinations of various machine learning algorithms [26].

Following the classification above, this paper uses structured data to predict the outcome of a soccer game in terms of "win-loss-draw" before the game, employing a variety of machine learning algorithms.

3. Methodology and Dataset

In this research, we follow the most popular data mining framework called, CRISP-DM [27]. CRISP-DM has six sequential steps: (i) Understanding the domain and developing the goals for the study; (ii) identifying accessing and understanding the relevant data sources; (iii) pre-processing, cleaning, and transforming the relevant data; (iv) developing models using comparable analytical techniques; (5) evaluating and assessing the validity and utility of the models against each other and against the goals of the study; and (vi) deploying the models for use in decision making process. Following this framework, we were able to systematically conduct

this study and improved the likelihood of accurate and reliable results.

In order to assess the predictive power of the different models we employed, we used a cross-validation method, which is a widely used statistical technique often used to compare multiple models for their accuracies. Even though given a sufficiently large dataset, a single random split into two (or three, for neural network, for instance) may give enough accuracy, we chose cross-validation as each record represents a game, and may cover a valuable yet another different aspect. More details of this cross-validation are given in Section 3.3. A graphical depiction of the overall methodology employed in this study is shown in Figure 1.
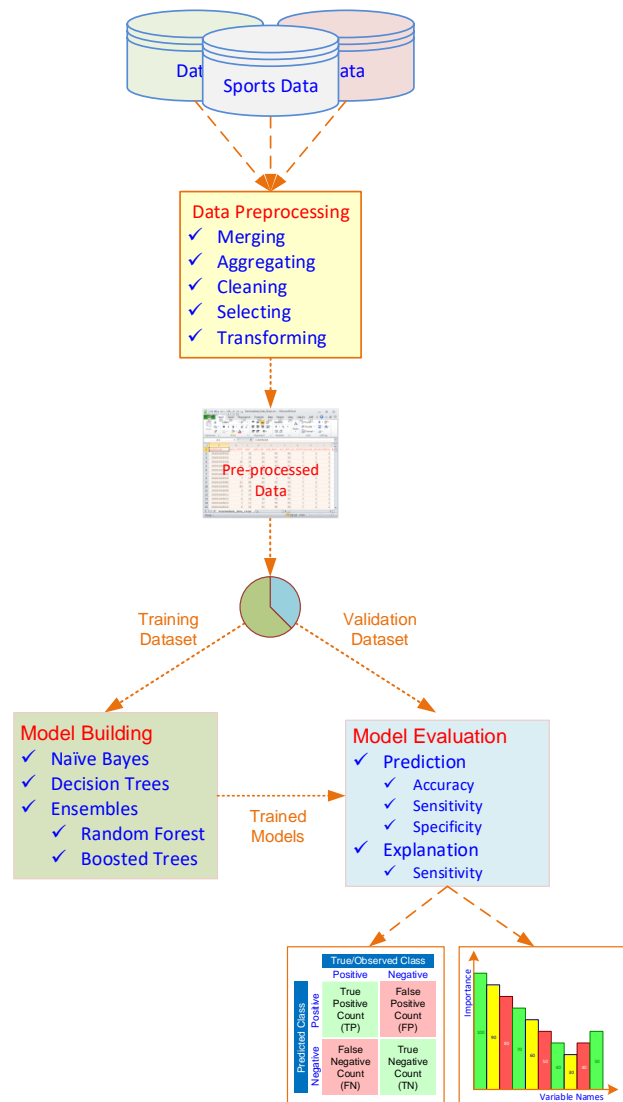

Figure 1: A Graphical Depiction of the Analytics Methodology

3.1. Data

The sample data for our study are collected from the Turkish Super League, using a variety of sources and means, including hand collection. The initial dataset included 3,060 game-level items of data, from 33 teams spanning a complete 10 seasons (2007-2017).

The regular soccer season in Turkish Super League lasts 34 weeks. Being the top flight of the country's football system, the three lowest performers at the end of the season are relegated to the 1st League. The top 4-5 performers get to represent the country in the UEFA Europe League (formerly called UEFA Cup), depending on the national team's performance in the UEFA League. Therefore, towards the end of the season, the teams at the bottom and top get more competitive while the team in the mid-section may get potentially reluctant, and prediction becomes more difficult. Also, during the first 5 weeks of the season, the transfer window is still open, and teams keep signing players. We, therefore, decided to filter out the first 5 and last 4 games of each league, and therefore included the games played during weeks 6 to 30. We also discarded data from 27 games where points were given according to "decision by referee".

The variables we used are provided in Table 1. Each record in Table 1 contains data about a game of a team.

Table 1. Description of the team and game-based variables used in this study

Variable	Cat	Explanation
Season	ID	Which season?
Week	ID	Which week of the season?
Team	ID	Which team?
WeekID	ID	Game ID for that week (1-9) for each week
CoachPosition	TM	If the coach was a former player, this is his position
LeagueStanding	TM	The current ranking in the league.
Exist10Year	TM	The team has been in the Super League consistently?
CoachAge	TM	Age of the coach
CoachNative	TM	Is the coach local?
MgtCntH	TM	How many different management teams during the season?
FormationConsistency	TM	How many different formations have been used so far during the season.
NumForeign	TM	Number of foreign players played in team
TeamValue	TM	Team value in local currency
Attend	GM	Number of spectators for the Home team.
Hour	GM	Game starting time
Formation	GM	Game formation (i.e. 4-2-3-1, 4-3-3, 4-4-2, 4-5-1…).
Home	GM	Did the team play as a home team?
PrcPossbPntsEarned	GM	Percentage of maximum possible points earned during the season
FrgnPlayerConsistency	GM	The consistency in the number of players of foreign nationals participating in the game
Plyr30PlusConsistency	GM	Number of players with 30+ of age
CoachConsistInSeas	GM	Same coach up to this week?
MgtConsistInSeas	GM	Same club management up to this week?
30PlsSubsUsedPrev	GM	Number of 30+ subs from bench included in the previous game
CoachCountInSeas	GM	How many different coaches up to that week?
PassesAttempted	GM	Weighted moving averages of passes attempted during recent games
PassesComplete	GM	Weighted moving averages of passes complete during recent games
Opportunities	GM	Weighted moving averages of number of opportunities during recent games
PercOpportunitiesScored	GM	Weighted moving averages of percent opportunities used during recent games
Assists	GM	Weighted moving averages of assists during recent games
ShotsOnTarget	GM	Weighted moving averages of shots on target during recent games
Shots	GM	Weighted moving averages of number of shots during recent games
Cross	GM	Weighted moving averages of number of crosses made during recent games
CrossComplete	GM	Weighted moving averages of number of crosses complete during recent games
Reception	GM	Weighted moving averages of number of receptions during recent games
Intercepts	GM	Weighted moving averages of number of intercepts during recent games

PosessLost	GM	Weighted moving averages of possessions lost during recent games
Corners	GM	Weighted moving averages of corner kicks awarded during recent games
Offside	GM	Weighted moving averages of number of offsides during recent games
FoulsCommitted	GM	Weighted moving averages of fouls committed during recent games
FoulsAwarded	GM	Weighted moving averages of fouls awarded during recent games
FreeKick	GM	Weighted moving averages of number of free kicks awarded during recent games
YellowCard	GM	Weighted moving averages of yellow cards during recent games
RedCard	GM	Weighted moving averages of red cards during recent games
PassIncomplete	GM	Weighted moving averages of passes incomplete during recent games
PercShotsOnTarget	GM	Weighted moving averages of percentage shots on target during recent games
PercClearCross	GM	Weighted moving averages of percentage of clear crosses during recent games
PercPassComplete	GM	Weighted moving averages of percentage of passes complete during recent games
AveragePlayerAges	GM	Weighted moving averages of average player ages played during recent games
NumPlayers30plus	GM	Weighted moving averages of number of 30+ players during recent games
SubValuePrevWeek	GM	Total values for players in the bench during the previous game
Scored	O1	Goals scored
Conceded	O1	Goals conceded
Result	O2	Win, Loss, or Draw

ID: Identifier variables; GM: game-related variables that change for each game. TM: team level variables that usually remain fixed throughout the season. O1: Output variable for regression models; O2: output variable for classification models.

3.2. Data Preprocessing

In the formulation of the dataset in Table 1, each row (or tuple) represents a team's performance during a game. Therefore, there are two rows, one for each home or away team in each game. We converted and reformulated Table 1, the "team"-based dataset, into a "game"-based dataset by representing each game with one record. When doing so, we calculated and used the differences between the measures of the home and away teams for each game. We then adjusted our target variables to be "Result", and "Score Difference", for the regression and classification tasks respectively.

The original dataset (Table 1) contained a significantly high amount of missing values (about 11%). These were mainly data that we were not able to collect, due to unavailability, and typically belonged to teams that were promoted to the top-flight league, or were relegated to the 1st league (lower league). Converting this "team" based dataset into "game" based dataset for analysis resulted in an increase in the number of missing values (17%), as some of the missing values were entered into difference calculations.

We concluded that the missing data is not MAR, and the number of complete instances was not sufficiently large enough (240 rows only) to perform a reliable imputation. We, therefore, for the sake of simplicity did not perform imputation.

3.3. Methods

In this study, we mainly used three popular prediction techniques, and we compared their performances against each other: Naïve Bayes, Decision Trees, and Ensemble models. Due to the high amount of missing values in our dataset, we could not employ some of the techniques such as Artificial Neural Networks, k-Nearest Neighbor, Logistics Regression or Support Vector Machines. Because of the non-random nature of the missing values in several of the critical independent variables, we could neither impute the missing values nor we could exclude them from the dataset. Therefore, we chose to use predictive analytics methods that allow and properly handles missing values in both training and testing sections of the final dataset.

Naïve Bayes

Naïve Bayes is one of the simplest and also one of the fastest techniques used in data mining and often performs surprisingly well [28]. However, it is known that the algorithm may underperform when its underlying independence assumption is not satisfied. The algorithm works well with almost any kind of dataset.

Decision Trees

The technique recursively separates observations into branches in order to construct a tree to achieve the highest possible prediction accuracy. In recursive separation, the technique may use different mathematical splitting criteria such as Information Gain, GINI index, or Chi-square statistics. We choose to use classification and regression trees (CART) which were initially developed by Breiman, Friedman, Olshen, and Stone [29]. This specific decision tree algorithm is capable of modeling both classification and regression-type problems and also performs well even with missing values.

Ensemble Methods

We used two ensemble methods, Gradient Boosting Trees [30] and Random Forest [29]. Both of these methods rely on a decision trees algorithm, CART in their standard setups. They handle missing data internally and automatically learn the best imputation value for missing values based on training loss. They also almost consistently outperform decision trees. While Gradient Boosted Trees uses very shallow regression trees and boosting to create ensemble, Random Forest creates a number of decision trees where each tree model is learned on different records and different columns recursively.

3.4. Evaluation criteria

A common way of evaluation is to split the data into two subsets for training, and testing (sometimes a third subset may be used as in the case of Neural Networks or Genetic Algorithms). Typically two-thirds of the dataset is used for model building and one third is left for testing. However such single splits often yield results that are prone to sampling bias regardless of the sampling technique. In order to minimize the bias many data scientists opt for cross-validation. In k-fold cross-validation, the dataset is split into k randomly split and mutually exclusive subsets of roughly the same size. Each model is trained and tested k times. Each time all (k-1) folds are used for learning and the last fold is used for testing. The cross-validation estimate of the accuracy is calculated by taking the average of the k individual accuracy measures. This usually increases total running times however the result is often less prone to overfitting and randomness. The overall accuracy (CV) of the cross-validation is calculated by taking the average accuracies of each fold (A_i) as in (1):

$$CV = \frac{1}{k} \sum_{i=1}^{k} A_i \qquad (1)$$

As cross-validation accuracy is influenced by the randomness in assigning records to folds, a common

practice is to stratify the folds so that the folds don't suffer from class imbalance problem. In our study, we set $k=10$.

We then used three performance criteria to compare the prediction performances of the selected models. Where TP, TN, FP, and FN denote true positive (accurate prediction of wins), true negative (accurate prediction of losses), false positive (false prediction of losses as wins or draws), and false negatives (false prediction of wins as losses or draws). We then compute accuracy, sensitivity, and specificity (Equations 2, 3, and 4 respectively).

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

$$Sensitivity = \frac{TP}{TP + FN} \qquad (3)$$

$$Specificity = \frac{TN}{TN + FP} \qquad (4)$$

4. Results

Outcome prediction of soccer games presents more difficulties than most of the other sports where there is only one winner, or two classes ("win" and "loss"). After the initial analysis we decided to analyze data in two different ways:
   i) Win/Loss/Draw prediction
   ii) Points/No Point prediction.
The outcome frequencies for 10 seasons are given in Table 2.

Table 2: Win/Loss/Draw, and Points/No Point frequencies

Win or Loss	Draw
1636	541
Points	No Point
1558	619

4.1. Win/Loss/Draw Prediction

Due to the dynamic nature of soccer, it is difficult to separate classes especially when a game is not dominated by a team. The game could easily end with any of the three outcomes. However, as it the middle ground between "win" and "loss, "draw" intuitively is the most difficult to predict. Moreover, as Table 2 shows, the least frequent outcome is the draw, making the task even more difficult to learn (significantly fewer data points for learning). In order to remedy this, we used an oversampling method to enrich our training data

[31]. We also performed missing value imputation (single imputation, using mean and mode) to be able to run Neural Networks (2 hidden-layers, 10 nodes per layer), SVMs (with RBF), and kNN (k=5). Table 3 summarizes our results.

Table 3. Prediction results for the direct classification methodology with three classes

Prediction method (classification)	Confusion matrix		
	Loss	Win	Draw
Naive Bayes[*] Loss	611	365	41
Win	207	783	27
Draw	340	545	132

Acc.(%)	Sens.(%)	Spec.(%)
51.90%	60.10%	72.90%
	77.00%	55.00%
	13.00%	96.60%

	Loss	Win	Draw
Decision Trees[*] (CART) Loss	675	124	218
Win	117	622	278
Draw	214	278	525

Acc.(%)	Sens.(%)	Spec.(%)
59.70%	66.40%	83.70%
	61.20%	80.20%
	51.60%	75.60%

	Loss	Win	Draw
Neural Nets[*†] Loss	658	74	285
Win	101	616	300
Draw	258	327	432

Acc.(%)	Sens.(%)	Spec.(%)
55.91%	64.70%	82.35%
	60.57%	80.29%
	42.48%	71.24%

	Loss	Win	Draw
SVMs[*†] Loss	663	69	285
Win	108	618	291
Draw	267	237	513

Acc.(%)	Sens.(%)	Spec.(%)
58.80%	65.19%	81.56%
	60.77%	84.96%
	50.44%	71.68%

	Loss	Win	Draw
kNN(k=5)[*†] Loss	695	122	200
Win	306	330	381
Draw	199	116	702

Acc.(%)	Sens.(%)	Spec.(%)
58.80%	68.34%	75.17%
	32.45%	88.30%
	69.03%	71.44%

	Loss	Win	Draw
Gradient Boosted Tree[*] Loss	809	50	158
Win	70	744	203
Draw	126	172	719

Acc.(%)	Sens.(%)	Spec.(%)
74.50%	79.50%	90.40%
	73.20%	89.10%
	70.70%	82.30%

	Loss	Win	Draw
Random Forest[*] Loss	827	98	92
Win	100	774	143
Draw	131	211	675

Acc.(%)	Sens.(%)	Spec.(%)
74.60%	81.30%	88.60%
	76.10%	84.80%
	66.40%	88.40%

[*] Performed after synthetic minority class oversampling
[†] Performed after imputation

The prediction results of the modeling techniques suggest ensemble tree models performed significantly better. However, the prediction accuracy for the "draw" class is proven to be more difficult. Figure 2 visualizes this phenomenon using Gradient Boosted Trees Regression results.
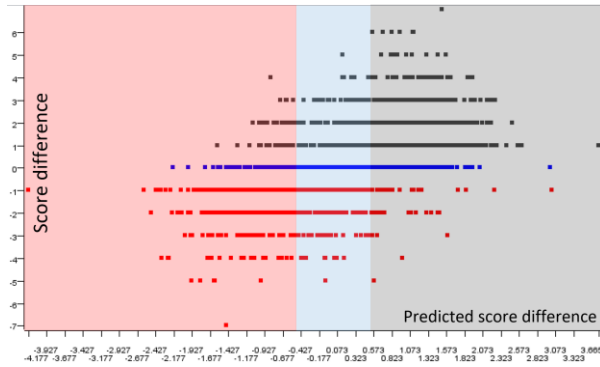
Figure 2. Highlighted area in "blue" represents games with "draw" prediction.

4.2. Points/NoPoint Prediction

In soccer, it is also important to predict whether a team will earn a point during a game. This makes the prediction task easier by reducing the possible number of outcomes to two: "Points" and "No Points". This prediction task is also relevant for soccer betting.

Following the same procedure (i.e. minority oversampling, and cross-validation) we report results for our top two performing algorithms in Table 4.

Table 4: Results for the two-class prediction task

Prediction method (classification)			
	Acc.(%)	Sens.(%)	Spec.(%)
Random Forest*	86.3%	88.1%	84.4%
		84.4%	88.1%
	Acc.(%)	Sens.(%)	Spec.(%)
Gradient Boosted Tree*	86.4%	89.0%	83.7%
		83.7%	89.0%

* With minority class oversampling

As the results indicate, ensemble-type prediction methods performed better. Among the four data mining methods, Naïve Bayes was the lowest performer. This may be due to dependencies in variables. Using a t-test, we also found that Gradient Boosting and Random Forest methods both outperformed Decision Tree (CART) and Naïve Bayes methods. However the performance difference between Gradient Boosting and Random Forest ensemble methods was not found to be statistically different at 0.05 alpha level for either of the "Win/Loss/Draw" and "Points/NoPoint" problems (α=0.11.2, and α=0.12 respectively).

4.3. Sensitivity Analysis

Algorithms are good at capturing non-trivial relationships and establish a relationship between input and output. While some algorithms are known as black-box algorithms, such as Neural Networks, others are transparent, such as decision trees. However, even when a transparent algorithm is used (for example tree structure of the decision tree) cannot always easily be interpreted. In the context of machine learning, sensitivity analysis refers to exclusive experimentation process to establish a possible cause and effect relationship between the input and the output variables [32].

Our sensitivity analysis is based on using our best performing algorithm (Random Forest). Random Forest is an algorithm based on decision trees. In a decision tree, variables that are used in earlier splits are considered more important. We use this characteristic of decision trees/random forests and compute the number of level-0 splits for each variable. This way we can to observe the impact of the variable on performance. The relative importance values are then tabulated normalized and graphically presented for the top 20 variables that are chosen for the first split in Figure 3.

As expected exclusion of these variables results in significant drops in classification performance.
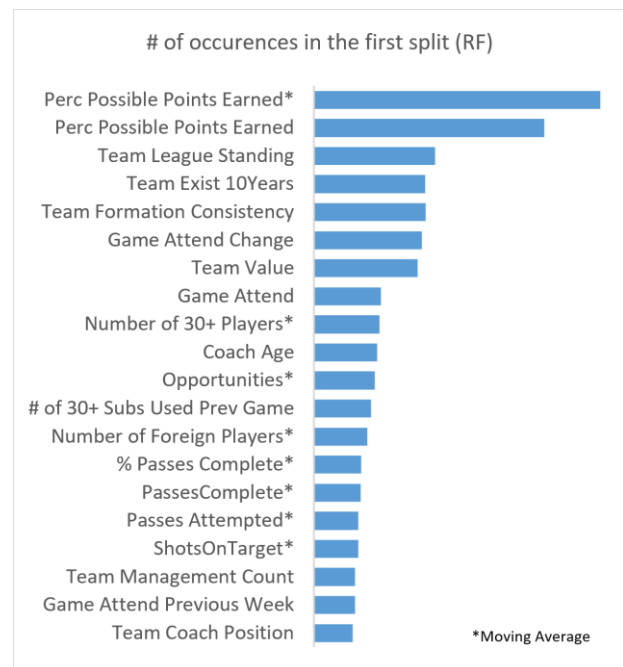

Figure 3. Variable importance values

5. Discussion, conclusion, and future research

The results of this study once again show that the prediction of soccer outcomes is not straightforward. We were able to achieve over 74% accuracy in "Win/Loss/Draw", and over 86% accuracy in "Points/NoPoint" type of classification problems. Perhaps, one of the shortcomings of this study was due to the significant amount of missing values (not at random). However, for the sake of inclusion of other methods (Neural Networks, kNN, and SVMs, we also performed missing value imputation in its most basic form (mean imputation). However, as expected these algorithms were not able to outperform ensemble methods. This limited our options for algorithms. We believe more comprehensive datasets may yield even better results. The ensemble methods outperformed the other two methods we used. Within ensemble trees, gradient boosted trees slightly outperformed random trees, however, the difference was not significant, and also should not be generalized beyond the scope of the study.

In soccer games, there are many different aspects to be studied from the different stakeholders. A coach, given the week and the opponent, may try to evaluate impacts of different formations, or different players included in the game. A team manager may try to understand what kind of coach would be suitable for his team, or what drives fans to stadiums. A sports gambler can attack the same problem with different target variables, such as the number of cards, corners etc. We also believe that a more carefully collected dataset (i.e. with fewer missing values), as well as a richer dataset in terms of variables, will help in predicting outcomes more accurately.

6. References

[1] D. Delen, D. Cogdell, N. Kasap, "A comparative analysis of data mining methods in predicting NCAA bowl outcomes", International Journal of Forecasting, Vol. 28, Issue 2, (2012), pp. 543-552.

[2] M.J. Lopez, G.J. Matthews, "Building an NCAA men's basketball predictive model and quantifying its success", Journal of Quantitative Analysis in Sports, Vol 11, Issue 1 (2015), pp. 5-12.

[3] P. Vračar, E. Štrumbelj, I. Kononenko, "Modeling basketball play-by-play data", Expert Systems with Applications, Vol 44 (2016), pp 58-66.

[4] M. Asif, I. G. McHale, "In-play forecasting of win probability in One-Day International cricket: A dynamic logistic regression model", "International Journal of Forecasting", Vol 32, Issue 1 (2016), pp 34-43.

[5] J. Carbone, T. Corke, F. Moisiadis, "The rugby league prediction model: using an ELO-based approach to predict the outcome of the national rugby league (NRL) matches", International Education Scientific Research Journal, Vol 2, Issue 5 (2016), pp 26-31.

[6] G. Wei, L.T. Saaty, R. Whitaker, "Expert System for Ice Hockey Game Prediction: Data Mining with Human Judgment", International Journal of Information Technology & Decision Making, Vol 15, Issue 4, pp. 763-789.

[7] E.G. Dunning, A. Joseph, R.E. Maguire, "The Sports Process: A Comparative and developmental approach" (1996), pp.129, Champaign: Human Kinetics.

[8] E. Dunning, "Sports Matters: Sociological Studies of Sport, Violence and Civilisation" (1999). London: Routledge.

[9] A.C. Constantinou, N.E. Fenton, M. Neil, "Profiting from an inefficient association football gambling market: Prediction, risk and uncertainty using Bayesian networks", Knowledge-Based Systems, Vol 50 (2013), pp. 60-86.

[10] R. Baboota, H. Kaur, "Predictive analysis and modeling football results using machine learning approach for English Premier League", International Journal of Forecasting, (2018), Forthcoming.

[11] S. Dobrovec, "Predicting sports results using latent features: A case study", 38th International Convention on Information and Communication Technology, Electronics and Microelectronics (MIPRO), 25-29 May 2015, pp. 1267 – 1272.

[12] R.P. Schumaker, A.T. Jarmoszko, C.S. Labedz, "Predicting wins and spread in the Premier League using a sentiment analysis of twitter", Decision Support Systems, Vol 88 (2016), pp. 76-84.

[13] V. Radosavljevic et al., "Large-scale World Cup 2014 outcome prediction based on Tumblr posts", KDD Workshop on Large-Scale Sports Analytics: Sydney, Australia, 2014.

[14] T. D'Orazio, C. Guaragnella, M. Leo, A. Distante, "A new algorithm for ball recognition using circle Hough transform and neural classifier", Pattern Recognition (2004), Vol 37, pp. 393–408.

[15] Y.L. Kang, J.H. Lim, M.S. Kankanhalli, C.S. Xu, Q.Tian, "Goal detection in soccer video using audio/visual keywords", IEEE International Conference on Image Processing (ICIP), Singapore, 24–27 October 2004, pp. 1629–1632.

[16] M.J. Maher, "Modeling association football scores", Statistica Neerlandica (1982), Vol 36, pp.109-118.

[17] B. Gianluca, M. Blangiardo, "Bayesian hierarchical model for the prediction of football results", Journal of Applied Statistics (2010), Vol 37, Issue 2, pp. 253-264.

[18] D. Karlis, I. Ntzoufras, "Bayesian modelling of football outcomes: Using the Skellams distribution for the goal

difference". IMA Journal of Management Mathematics (2008), pp. 229–244.

[19] J. Goddard, "Regression models for forecasting goals and match results in association football", International Journal of Forecasting (2005), Vol 21, Issue 2, pp.331-340.

[20] M.J. Dixon, S.G.Coles, "Modelling association football scores and inefficiencies in the football betting market", Journal of the Royal Statistical Society: Series C (Applied Statistics) (1997), Vol 46, pp. 265–280.

[21] D. Forrest, R. Simmons, "Forecasting sport: the behaviour and performance of football tipsters", International Journal of Forecasting (2000), Vol 16, Issue 3, pp.317-331.

[22] S. J. Koopman, R. Lit, "A dynamic bivariate Poisson model for analysing and forecasting match results in the English Premier League", Journal of the Royal Statisitical Society: Series A, Vol 178, Issue 1, pp.167-186.

[23] M. Crowder, M. Dixon, A. Ledford, M. Robinson, "Dynamic modelling and prediction of English football league matches for betting", The Statistician (2002), Vol 51, pp.157–168.

[24] F. Godin, et al., "Beating the bookmakers: leveraging statistics and twitter microposts for predicting soccer results", KDD Workshop on Large-Scale Sports Analytics. Sydney, Australia, 2014.

[25] M. Byungho, J. Kim, C. Choe, H. Eom, R.I. McKay, "A compound framework for sports results prediction: A football case study", Knowldege Based Systems, Vol 21, pp.551-562.

[26] A.P. Rotshtein, M. Posner, A.B. Rakityanskaya, "Football Predictions Based on a Fuzzy Model with Genetic and Neural Tuning", Cybernetics and Systems Analysis (2005), Vol 41, Issue 4, pp. 619-630.

[27] C. Shearer, "The CRISP-DM model: the new blueprint for data mining", Journal of Data Warehousing (2000), Vol 5, pp.13–22.

[28] D.J. Hand, K. Yu, "Idiots Bayes—not so stupid after all?", International Statatistical Review (2001), Vol 69, pp.385–398.

[29] L. Breiman, J. Friedman, R. Olshen, C. Stone, "Classification and Regression Trees" (1984), Wadsworth.

[30] J.H. Friedman, "Greedy function approximation: a gradient boosting machine". Technical Report, Department of Statistics, Stanford University (1999).

[31] N. Chawla, K.W. Bowyer, L.O. Hall, W.P. Kegelmeyer, "SMOTE: Synthetic Minority Over-sampling Technique", Journal of Artificial Intelligence Research (2002), Vol 16, 321-357.

[32] G. Davis, "Sensitivity analysis in neural net solutions". IEEE Trans. Syst., Man, Cybern. (1989), Vol 19, pp.1078–1082.