A Machine Learning Framework for Sport Result Prediction

Article · September 2017

DOI: 10.1016/j.aci.2017.09.005

A Machine Learning Framework for Sport Result Prediction

Rory P. Bunker[1], Fadi Fayez[2]

[1]Auckland University of Technology, Auckland, New Zealand
[2] Nelson Marlborough Institute of Technology, Auckland, New Zealand

Abstract

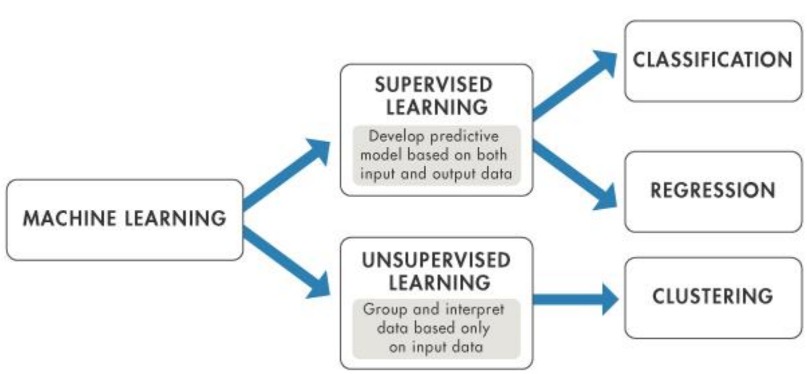
Machine learning (ML) is one of the intelligent methodologies that have shown promising results in the domains of classification and prediction. One of the expanding areas necessitating good predictive accuracy is sport prediction, due to the large monetary amounts involved in betting. In addition, club managers and owners are striving for classification models so that they can understand and formulate strategies needed to win matches. These models are based on numerous factors involved in the games, such as the results of historical matches, player performance indicators, and opposition information. This paper provides a critical analysis of the literature in ML, focusing on the application of Artificial Neural Network (ANN) to sport results prediction. In doing so, we identify the learning methodologies utilised, data sources, appropriate means of model evaluation, and specific challenges of predicting sport results. This then leads us to propose a novel sport prediction framework through which ML can be used as a learning strategy. Our research will hopefully be informative and of use to those performing future research in this application area.

Keywords: Machine learning, event forecasting, data mining, sport result prediction

1. Introduction

One of the common machine learning (ML) tasks, which involves predicting a target variable in previously unseen data, is classification (Mohammad, et al., 2015). The aim of classification is to predict a target variable (class) by building a classification model based on a training dataset, and then utilising that model to predict the value of the class of test data (Witten, et al., 2011). This type of data processing is called supervised learning since the data processing phase is guided toward the class variable while building the model. Some common applications for classification include loan approval, medical diagnoses, email filtering, among others (Abdelhamid & Thabtah, 2014).

Figure 1. Supervised Learning versus Unsupervised Learning (Mathworks, n.d.)



Sport prediction is usually treated as a classification problem, with one class (win, lose, or draw) to be predicted (Prasitio & Harlili, 2016). Although some researchers (e.g. Delen, et al., 2012), have also looked at the numeric prediction problem, where they predict the winning margin – a numeric value. In sport prediction, large numbers of features can be collected including the historical performance of the teams, results of matches, and data on players, to help different stakeholders understand the odds of winning or losing forthcoming matches. The decision of which team is likely to win is important because of the financial assets involved in the betting process; thus bookmakers, fans, and potential bidders are all interested in approximating the odds of a game in advance (Fernandez & Ulmer, 2014). Once a predicted result for the match is obtained, an additional problem is to then decide whether to bet on the match, given the bookmaker's odds. In addition, sport managers are striving to model appropriate strategies that can work well for assessing the potential opponent in a match (Nunes & Sousa, 2006). Therefore, the challenge of predicting sport results is something that has long been of interest to different stakeholders, including the media. The increasing amount of data related to sports that is now electronically (and often publically) available, has meant that there has been an increasing interest in developing intelligent models and prediction systems to forecast the results of matches.

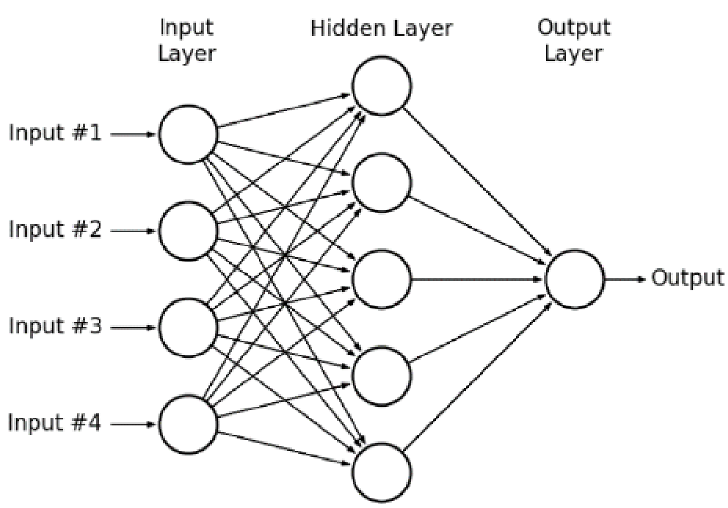
In this paper, we provide a critical survey of the literature on ML for sport result prediction, focusing on the use of neural network (NN) for this problem. Several studies in the statistical and operations research literature have previously considered sport results prediction, but the use of the NN paradigm for this purpose is a more recent area of study. The powerful NN technique has proven to be effective in deriving highly accurate classification models in other domains (Mohammad, et al., 2016). Discussions on the challenges that arise when using these intelligent models for sport results prediction is also provided. Our main contribution is that a CRISP-DM type framework for sport result prediction is proposed (SRP-CRISP-DM), based on the six steps of the standard CRISP-DM framework (Shearer, 2000). This paper serves researchers, sport fans, club managers, bookmakers, academics, and students who are interested in intelligent solutions based on NN for the challenging problem of sport results prediction. This paper will be of use to those who are interested in pursuing future research within this application domain.

This remainder of this paper is organised as follows. In section 2, studies that have used ANN exclusively, which was the key approach used in earlier research papers in the sport prediction application, are reviewed. Section 3 then provides critical discussion and observations on prior work in this application domain, in the context of the proposed SRP-CRISP-DM framework, conventional measures of model performance, and how we propose that model performance should be measured for the problem of sport results prediction. Finally, section 4 concludes the paper.

2. Literature Review and Critical Analysis

Artificial Neural Networks (ANNs) (Grossberg, 1988) are perhaps the most commonly applied approach among ML mechanisms to the sport result prediction problem. Thus, for this review, we focus on studies that have applied ANNs. An ANN usually contains interconnected components (neurons) that transform a set of inputs into a desired output (Witten, et al., 2011). See figure 2 for an example of an ANN structure. The power of ANN comes from the non-linearity of the hidden neurons in adjusting weights that contribute to the final decision. ANN output often relies on input features and other components associated with the network, such as these weights. The ANN model is constructed after processing the training dataset that contains the features used to build the ANN classification model. In other words, weights associated with interconnected components are continuously changing to accomplish high levels of predictive accuracy. These changes are performed by the ANN algorithm to fulfill the desired model's accuracy given earlier by the user. This may lead in some cases to the problem of overfitting, as well as wasting computing resources such as training time and memory (Mohammad, et al., 2014). An appealing feature of ANNs is that they are quite flexible in terms of how the class variable is defined e.g. whether it is probability of victory (e.g. McCabe & Travanthan, 2008), or whether two classes are used e.g. with home goals and away goals represented in the two different classes (e.g. Arabzad et al., 2014).

Figure 2. Example structure of an ANN with 4 input nodes in the input layer, 5 hidden nodes in the hidden layer and one output node in the output layer (Mohamed et al., 2015, p.252).



Purucker (1996) conducted one of the initial studies on predicting results in the National Football League (NFL) using an ANN model. Data from the first eight rounds of the competition and five features were used, consisting of yards gained, rushing yards gained, turnover margin, time of possession, and betting line odds. Unsupervised methods based on clustering were used to distinguish between good and poor teams. An ANN with backward-propagation (BP) was then used (Rumelhart, et al., 1986). Purucker achieved 61% accuracy compared with 72% accuracy of the domain experts. The BP algorithm was found to be the most effective approach. A limitation of this study is that only a relatively small number of features were used.

Kahn (2003) extended the work of Purucker (1996) and achieved greater accuracy, performing slightly better than experts in the NFL who were making predictions on the same games. Data on 208 matches in the 2003 season were collected. The features that were used were: total yardage differential, rushing yardage differential, turnover differential, away team indicator and home team indicator. There were two classes: away team outcome and home team outcome – a value of -1 indicating that the team lost the match, and a value of +1 indicating that the team won the match. The problem was treated as a classification problem. The first 192 matches were used as the training data set, and the remaining rounds (week 14 and 15) were used as the test set. Through testing, a network structure of 10-3-2 was found to be optimal. Accuracy of 75% was achieved across the week 14 and 15 matches. The results were compared to the predictions of eight sportscasters from ESPN.com. Across the same matches, the domain experts predicted an average of 63% of matches correctly.

McCabe & Trevathan (2008) attempted to predict results in four different sports: NFL (Rugby League), AFL (Australian Rules football), Super Rugby (Rugby Union), and English Premier League Football (EPL) using data back to the year 2002. A multi-layer perceptron, trained with BP and conjugative-gradient algorithms was used. The ANN had 20 nodes in the input layer, 10 nodes in the hidden layer, and 1 node in the output layer (20-10-1). Features that were used were the same across all the sports and attributes related to specific events within a rugby or soccer match were not considered. The average performance of the ANN algorithm in predicting results was around 67.5%, compared with expert tipster predictions that achieved around 60% to 65% accuracy.

ANN has also been applied by Davoodi & Khanteymoori (2010) to predict the results of horse races. The authors used data from 100 races at the Aqueduct Race Track held in New York during January of 2010. One ANN was used for each horse in the race, with the output being the finishing time of that horse. Eight features were used for the input nodes in each NN. These were horse weight, type of race, horse trainer, horse jockey, number of horses in race, race distance, track condition, and weather. This optimal network architecture (8-2-1), in terms of mean-squared error, consisted of four layers: an input layer (with eight input nodes), two hidden layers, and an output layer (with horse finishing time). Five different training algorithms were applied to the data: gradient-descent BP (BP), gradient-descent with a momentum parameter (BPM), Levenberg-Marquadt (LM), and conjugate gradient descent (CGD) (Rumelhart, et al., 1986; Sutton, 1986; Kanzow, et al., 2004; Yvan, 2000). It was found that with 400 epochs, the BPM (with momentum parameter of 0.7) and the BP algorithms were most effective at predicting the winner of the race, with BP obtaining an accuracy of 77%. However, the disadvantage of BP was that the training time was lengthy (LM had the shortest training time).

Tax and Joustra (2015) used Dutch football competition data from the past 13 years to predict the results of football matches. The authors were interested in how a model with betting odds alone compared with a hybrid model of both betting odds and other match features. Importantly, and something that has most often been missed in previous studies, they mentioned that cross validation is not appropriate for sport prediction because of the time-ordered nature of the data. A structured literature review from statistical and sport science papers was conducted to identify relevant features to include. Principal component analysis (PCA), sequential forward selection, ReliefF attribute evaluation, and correlation based feature subset selection were used (Jolliffe, 2002; Marcano-Cedeno, et al., 2010; Kononenko, et al., 1997; Hall, 1999). Nine classification algorithms were used in the experimentation, utilizing the machine learning software WEKA, namely naive Bayes, LogitBoost (with decision stumps), NN with BP, Random Forest, CHIRP, FURIA, DTNB, C4.5, and hyper pipes (Hall, et al., 2009; Wilkinson, et al., 2011; Hühn, & Hüllermeier, 2009; Hall & Frank, 2008; Quinlan, 1993). The highest performing classifiers on the full feature set were naive Bayes (used with a 3-component PCA), and the ANN (used with a 3 or 7-component PCA). Both achieved a classification accuracy of 54.7%. In a model including only betting odds features, the highest accuracy of 55.3% was achieved with the FURIA classifier, and was slightly higher than the model with the full feature set (although not statistically significant). In a hybrid model of the public data features with the betting odds features, LogitBoost with ReliefF attribute selection provided the highest classification accuracy of 56.1%. The difference between the public data model and the betting odds model was, however, not statistically significant according to McNemar's test. However, this did highlight that betting odds alone can be a reasonable predictor of match outcome.

In non-team sports, researchers have used machine learning models to predict the performance of the individual player.

Maszczyk et al. (2014) compared neural networks and non-linear regression to predict the distance of Javelin throws. The aim of the investigation was to identify the usefulness of neural networks as an athlete recruitment tool, and how this compared to the commonly used regression models. The data set consisted of 70 javelin throws - a training set consisting of 40 cases, a validation set consisting of 15 cases, and a test set consisting of 15 cases. Their initial statistical analysis using a correlation matrix and regression analysis found four significant predictors of Javelin throw length: cross step, specific power of the arms and the trunk, specific power of the abdominal muscles, and grip power. The numeric class variable used was the average distance of three throws from a full run-up after a 30 minute warm up. Through experimentation, the best architecture in terms of normalized root mean squared error, of the neural network was found to be 4-3-1 (four input neurons/variables, one hidden layer with three neurons, and one outcome). The javelin throws of 20 javelin throwers from the Polish national team were predicted using the models, and were compared with the actual length of the throws. Their results showed that the neural network models offered much higher quality of prediction than the nonlinear regression model. The absolute network error was found to be 16.77 m, versus the absolute regression error of 29.45 m.

Edelmann-Nusser et al (2002) investigated modeling the performance of an elite female swimmer in the finals of the 200 m backstroke at the Olympic Games in 2000 in Sydney. Data consisted of the performance output of 19 competitions in 200 m backstroke prior to the Olympics and data from the swimmer's training period - the last 4 weeks prior to the competition. An MLP with 10 input neurons, 2 hidden neurons and 1 output neuron was used. The results show that the MLP was accurate; the error of the prediction was only 0.05 s. The MLP was also compared with linear regression, which did not provide as accurate results. This paper (as well as Maszcyk et al. (2014)) highlights the potential usefulness for machine learning techniques to be used by high performance staff and analysts in professional sport for identifying the factors to focus on when developing training programs, not just purely for result prediction.

Wiseman (2016) predicted winning PGA golf score based on scores after round 1 of a competition. Note that they were predicting winning score, not tournament winner itself. The authors compared the performance of: linear regression, neural network regression, Bayesian linear regression, decision forest regression and boosted decision tree regression, in the Microsoft Azure service. The authors performed correlation matrix analysis of different features and selected Round 1 leading score, round 1 average score, course par, major event, course yardage and total prizemoney as the predictors. R-squared value and MSE were used to evaluate algorithm accuracy. Data from 2004 to 2015 was used to construct the models, and tournaments from 2016 were used to validate them. Linear regression and Bayesian linear regression were the best performing models on the 2016 data set, predicting the winning score to within 3 shots 67% of the time.

3. The Proposed Sport Result Prediction Intelligent Framework

We would argue that the use of a structured experimental approach to the problem of sport results prediction is useful to obtain the best possible results with a given data set. In this section, an intelligent architecture for sport results prediction is presented, proposing steps of a possible ML framework, and describing the characteristics of the data used for sport results prediction, and how this fits within the framework. Our framework (figure 4) focuses on result prediction for team sports rather than individual sport. Our Sport Result Prediction CRISP-DM framework or SRP-CRISP-DM framework consists of six main steps, based on the steps of the standard CRISP-DM (figure 3) framework (Shearer, 2000).

Figure 3. The six phases of the traditional CRISP-DM Model (Shearer 2000, p.14)

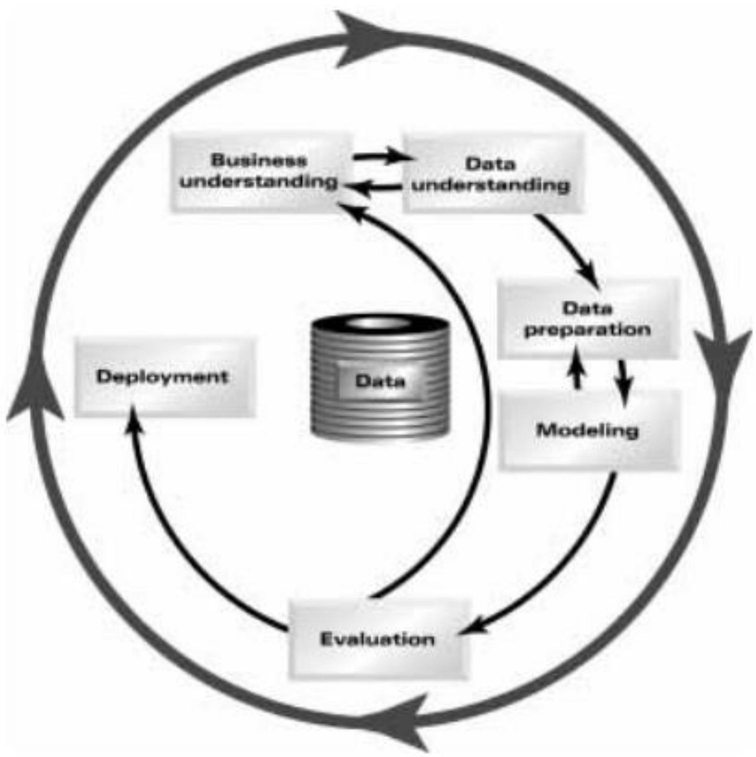
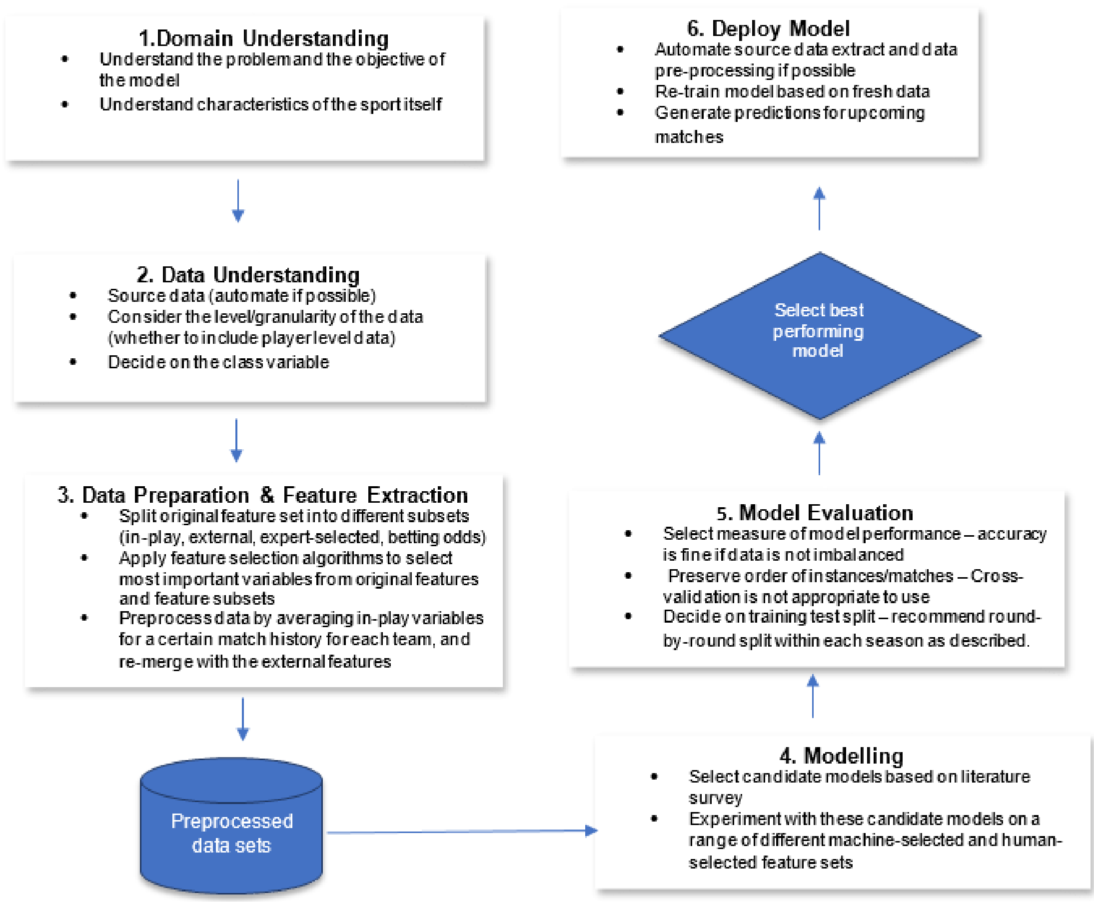


Figure 4. Steps of our proposed SRP-CRISP-DM framework



3.1. Domain Understanding

Domain understanding includes comprehending the problem, the goal of the modeling, and the specific characteristics of the sport itself. This involves having some understanding of how the sport is played and what factors are potentially involved in determining the outcome of matches. This could be obtained through personal knowledge of the sport or obtained by surveying existing literature or by consulting experts in the sport.

There also needs to be clarity regarding the objective of the model. It could be to predict results to compete with expert predictions, online competitions, or it could be to ultimately use the results of the model to bet on matches. If the predictive model is used for betting, there needs to also be consideration of which matches will be bet on. For example, there will likely be some betting odds threshold where, although the model predicts a victory for that team, the betting odds are so low that the return does not warrant betting on that match at all (e.g. if a team is paying $1.01 to win, meaning that a $100 bet placed would only return $1).

3.2. Data Understanding

Data for sport prediction is often able to be obtained online from publically available sources. Some prior studies have automated the data collection process, writing scripts that automatically extract the online data and then load it into some form of database. Some studies have also built an end-user interface, where users can input data for an upcoming match and the prediction is then generated.

The granularity/level of the data is something that needs to be considered. Previous studies have generally had training data that is at the match/team level. It is also possible to include player-level data, which contains statistics on the players that have played in each of the matches. Player level data will generally be contained in a separate data set that would then have to be transposed and joined with the match level data so that each match has certain player statistics as attributes in the data set. Including player level data would have the advantage that we can investigate whether specific players' actions or presence are important for the performance of the team in terms of whether they win or lose.

The definition of the class variable needs to also be considered. Most prior work has treated the sport prediction problem as a 2 or 3 class values classification problem (home win, away win) or (home win, draw, away win). Delen, et al. (2012), also considered the problem as a numeric prediction problem, using regression techniques to predict the points margin (home points minus away points), and then making a win-loss prediction based on the predicted points margin. The authors ultimately found that treating the problem as a classification resulted in superior results, but that is not to say that this would be the case on all sports or all data sets.

3.3. Data Preparation & Feature Extraction

3.3.1. Creating Feature Subsets

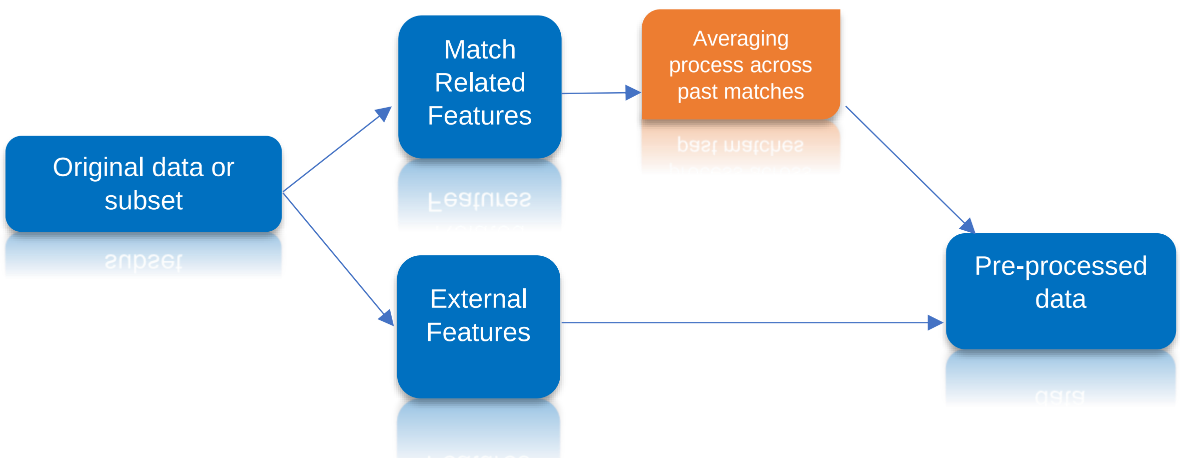
Features in sport result data can be divided into several different subsets. Miljković, et al. (2010), for example, split the features into match-related and standings features. Tax and Joustra (2015), considered how a hybrid model of betting odds and public data features compared with a feature set of betting odds alone. Hucaljuk and Rakipović (2011) used a separate expert-selected feature set against their own feature set, to investigate the value of expert opinion for feature selection. And of course, there are feature sets that are selected based on feature selection algorithms – either on the full feature set or a subset of the original feature set. Ideally, researchers should test several different feature selection approaches in conjunction with testing their candidate classification models. Then, as Tax and Joustra (2015) did, one can investigate which classifier and feature selection algorithm together produces the best classification accuracy.

Hucaljuk and Rakipović (2011) included an expert-selected feature set in addition to their initial feature set. Although this was not found to result in improved accuracy, this could depend on the characteristics of the sport, and perhaps the experts themselves. Another way that expert opinion can be used is in comparing the predictive accuracy of the predictive models, with the predictions of the experts. To incorporate expert opinion, one could either generate an expert-selected feature set to compare with machine-learned feature selection approaches, or alternatively, compare their model with expert predictions.

3.3.2. Data Preprocessing: Match Features versus External Features

There can be a distinction made between 'match-related' and 'external' features. Match-related features relate to actual events within the sport's match. For example, in football these could be meters gained, passes made, and so on. External features do not relate to events within the match, that is are external to the match itself (e.g. recent form, travel, players available for the match, etc.). This distinction is important for data preprocessing purposes. External features are known prior to the upcoming match to be played. For example, we know the distance that both teams have travelled and we know both teams' recent form leading into the upcoming match. Match-related features however, are not known until the match has been played. Thus, we only know an average of these features for a certain number of past matches for these teams. For example, we would know the average passes made per match by both teams prior to the match, but do not know the actual passes made in the upcoming match until after it has been played. This means that only past average statistics for these features can be used to predict an upcoming match. Therefore, match-related features should undergo a separate averaging process before being re-merged with the external features. Buursma (2011) followed this process, and found, through experimentation, that using an average across the past 20 matches resulted in the best classification accuracy.

Figure 5. Match-Related statistics should go through an averaging process across a certain number of historical matches for each team, and then be re-merged with the external match features.



3.4. Modelling

The first step in the modelling process is to select which candidate models will be used in the experimentation. This would involve a review of past literature, and identification of commonly applied predictive models that have previously been successful. Each model can then be trialed on each feature subset, and subsets that have been selected by feature selection algorithms. Experimentation with these different feature selection methods and classification models will identify the best combination of classifier and feature selection technique.

3.5. Sport Prediction Model Evaluation

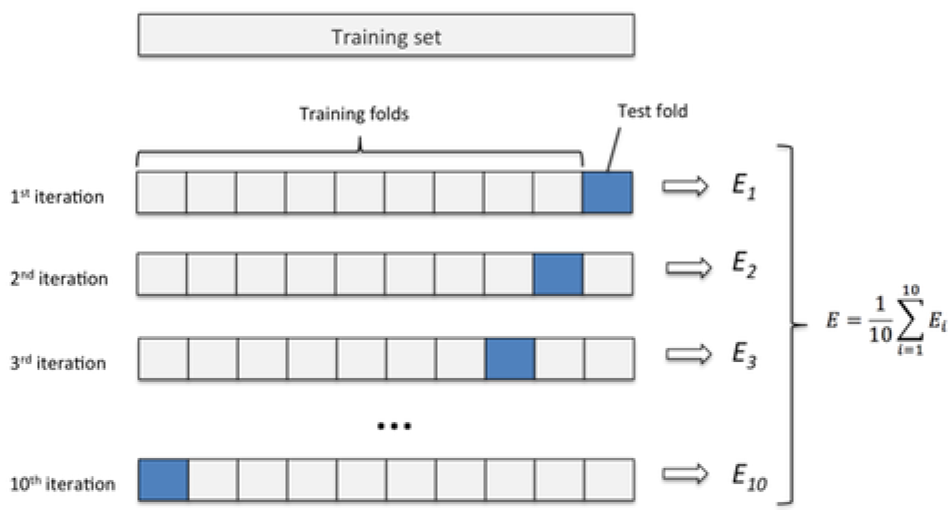
3.5.1. Measuring Model Performance

To evaluate model performance, one would classify match results into home wins, away wins and draws (if the sport has draws) and then look at the number of matches that the model has correctly identified, using a standard classification matrix. There is unlikely to be a great degree of imbalance in the class values for the dataset, although given the commonly observed home advantage phenomenon, one is likely to see a slight skew in favor of home wins. In this case, classification accuracy is a reasonable measure of evaluation. In cases where the data is highly imbalanced, ROC curve evaluation may be more appropriate.

3.5.2. Training and Testing

As has been mentioned, it is important to preserve the order of the training data for the sport prediction problem, so that upcoming matches are predicted based on past matches only. Cross-validation (figure 6) generally involves shuffling the order of the instances and therefore is not an appropriate means of splitting the data into training and testing, for the sport result prediction problem. A held-out training test split is more appropriate, with the order of the instances being preserved. Machine learning software such as WEKA provide the option to preserve the order of instances.

Figure 6. Diagrammatic representation of 10-fold Cross-Validation (Raschka, n.d.)



An appropriate training-test split needs to be decided on. This may depend on the amount of data that the researcher has on hand – whether they have only one season of data, or multiple seasons. Usually professional sport competitions are organized in rounds, with teams playing matches over the weekend. Teams usually play one match in each round unless they have a 'bye'. In the case where one season of data is on hand, the number of rounds that will be used for training the model, and the number of rounds that will be used for testing the model needs to be determined. For example, in a data set with 10 rounds of data, the first 7 rounds of the competition could be used for training the model and the last 3 rounds of the competition could be used for testing the model. However, to obtain a more realistic measure of model performance, round 1 could be used as training to test on round 2, round 1 & 2 could be used as training to test on round 3, round 1 – 3 could then be used as training to test on round 4, and so on. So, within a season which contains a certain number of competition rounds, we use rounds 1 to n-1 to train our model, and use round n as the test data set, for each round n in N, where N is the total number of rounds in the competition. We thus obtain a classification accuracy for each of these training/test splits, and take an average of the accuracies to give an overall measure of model performance.

Rather than round-by-round prediction, another possibility is to update the training data set after every match has been played. In this case, all past matches up to the current match as training data, and the upcoming match as the training data (i.e. only having that one record as the training data). This is essentially like order-preserved leave-one-out cross-validation. This match-by-match approach is probably not necessary unless teams play more than one match over the same competition round.

Some papers have used multiple seasons of data. A common approach has been to use earlier seasons as training data, to predict the later seasons as the test data set. For example, Cao (2012) used seasons up from 2005/2006 to 2009/2010 were used as the training data, and the 2010/2011 season was used as the test data, to evaluate models that predict basketball match results. Prior seasons may not be relevant to predict matches in future seasons, particularly in sports where team rosters and strengths can change significantly from year to year. This approach may not give a reliable picture of model performance (although this could be mitigated to some extent if player level data is included, and so player changes would be captured from season to season). We would argue that, although more computationally intensive and laborious, our round-by-round training test split approach mentioned above should be used *within each season*. An average model classification accuracy could then be produced for each season, and a plot could be shown of model accuracy by season.

3.6. Model Deployment

Ideally, one can automate the process so that new round data is obtained from the web, and added to the match database (or otherwise added to the database manually by the end-user). The training data and test data are then adjusted, the model is retrained with the new training data, and new matches are predicted. Predictions are then returned to the end user. The learning model in the proposed architecture could also be online and dynamically receiving input data prior to the match beginning (external features) and while the match is played (match features). It also should be incremental in the way that the training data set is continuously updated, and thus the classifier would keep changing to reflect those of the learning environment.

4. Conclusions

One of the vital applications in sport that requires good predictive accuracy is match result prediction. Traditionally, the results of the matches are predicted using mathematical and statistical models that are often verified by a domain expert. Due to the specific nature of match-related features to different sports, results across different studies in this application can generally not be compared directly.

Despite the increasing use of ML models for sport prediction, more accurate models are needed. This is due to the high volumes of betting on sport, and for sport managers seeking useful knowledge for modelling future matching strategies. Therefore, ML seems an appropriate methodology for sport prediction since it generates predictive models that can predict match results using predefined features in a historical dataset.

This article critically analyses some recent research on sport prediction that have used ANN, and following this, we proposed a sport result prediction 'SRP-CRISP-DM' framework for the complex problem of sport result prediction. Moreover, challenges facing the sport prediction application were shown to pinpoint future work for scholars in this important application. Future studies concerning ML in sport result prediction research will hopefully be benefitted by this study.

References

1. Abdelhamid N., and Thabtah F. (2014). Associative Classification Approaches: Review and Comparison. *Journal of Information and Knowledge Management (JIKM)*, 13(3).
2. Agrawal R., and Srikant R. (1994). Fast algorithms for mining association rules in large databases. Proceedings of the 20th International Conference on Very Large Data Bases, VLDB, pages 487-499, Santiago, Chile, September 1994.
3. Altman, N. S. (1992). An introduction to kernel and nearest-neighbor nonparametric regression. *The American Statistician*, 46 (3): 175–185.
4. Arabzad, A. C., Araghi, M. E. T., & Soheil, S. N. (2014). Football Match Results Prediction Using Artificial Neural Networks: The Case of Iran Pro League. *International Journal of Applied Research on Industrial Engineering*, *1*(3), 159-179.
5. Bouckaert, R. R. (2004). Bayesian network classifiers in Weka. (Working paper series. University of Waikato, Department of Computer Science. No. 14/2004). Hamilton, New Zealand: University of Waikato.
6. Cao, C. (2012). Sports data mining technology used in basketball outcome prediction. Master's Thesis, Dublin Institute of Technology, Ireland.
7. Cortes C. and Vapnik V. (1995). Support-Vector Networks. *Machine Learning*, 20(3): 273 – 297.
8. Davoodi, E., and Khanteymoori, A. (2010). Horse racing prediction using artificial neural networks. *Recent Advances in Neural Networks, Fuzzy Systems & Evolutionary Computing*, 2010: 155–160.
9. Delen, D., Cogdell, D., & Kasap, N. (2012). A comparative analysis of data mining methods in predicting NCAA bowl outcomes. *International Journal of Forecasting*, *28*(2), 543-552.
10. Duda, R. O. and Hart P. E. (1973). *Pattern Classification and Scene Analysis*. New York: John Wiley & Sons.
11. Edelmann-Nusser, J., Hohmann, A., & Henneberg, B. (2002). Modeling and prediction of competitive performance in swimming upon neural networks. *European Journal of Sport Science*, *2*(2), 1-10.
12. Fernandez, M., and Ulmer, B. (2014). Predicting Soccer Match Results in the English Premier League.
13. Friedman J., Hastie T., and Tibshirani R. (2000) Additive logistic regression: a statistical view of boosting. *Annals of Statistics*, 28(2): 337–407.
14. Grossberg S. (1988). Nonlinear neural networks: Principles, mechanisms, and architectures. *Neural Networks*, 17–61.
15. Hall, M. A. (1999). Correlation-based Feature Subset Selection for Machine Learning. PhD dissertation, Department of Computer Science, University of Waikato.
16. Hall, M. and Frank, E. (2008). Combining naive Bayes and decision tables. Proc 21st Florida Artificial Intelligence Research Society Conference, Miami, Florida. AAAI Press.
17. Hall M., Frank E., Holmes G., Pfahringer B., Reutemann P., and Witten I. (2009). The WEKA data mining software: an update. *SIGKDD Explorations*, 11(1):10-18.
18. Hühn, J., & Hüllermeier, E. (2009). FURIA: an algorithm for unordered fuzzy rule induction. *Data Mining and Knowledge Discovery*, *19*(3), 293-319.
19. Hofmann, M., and Klinkenberg, R. (Eds.). (2013). RapidMiner: Data mining use cases and business analytics applications. Chapman & Hall/CRC Data Mining and Knowledge Discovery Series, CRC Press.
20. Hucaljuk J., and Rakipović A. (2011). Predicting football scores using machine learning techniques. In *MIPRO, 2011 Proceedings of the 34th International Convention, IEEE*, 1623–1627.

21. Jolliffe, I. T. (2002). *Principal Component Analysis*, second edition. Berlin: Springer-Verlag.
22. Joseph A., Fenton N.E., and Neil, M. (2006). Predicting football results using Bayesian nets and other machine learning techniques. *Knowledge-Based Systems*, 19:544–553.
23. Kahn, J. (2003). Neural network prediction of NFL football games. *World Wide Web electronic publication* (2003), 9–15.
24. Kanzow C., Yamashita N., and Fukushima M. (2004). Levenberg-Marquardt methods with strong local convergence properties for solving nonlinear equations with convex constraints. *JCAM*, 172(2):375-97.
25. Kononenko, Igor, et al. (1997). Overcoming the myopia of inductive learning algorithms with RELIEFF. *Applied Intelligence*, 7(1): 39-55.
26. Marcano-Cedeno A, Quintanilla-Dominguez J, Cortina-Januchs M, and Andina D (2010). Feature selection using sequential forward selection and classification applying artificial metaplasticity neural network. In: 36th annual conference on IEEE industrial electronics society, pp 2845–2850
27. Maszczyk, A., Gołaś, A., Pietraszewski, P., Roczniok, R., Zając, A., & Stanula, A. (2014). Application of neural and regression models in sports results prediction. Procedia-Social and Behavioral Sciences, 117, 482-487.
28. Mathworks (n.d.). machinelearning_supervisedunsupervised.png. Retrieved from https://de.mathworks.com/help/stats/machinelearning_supervisedunsupervised.png
29. McCabe A. and Trevathan J. (2008). Artificial intelligence in sports prediction. In *Information Technology: New Generations*, ITNG 2008. Fifth International Conference on, IEEE, 1194–1197.
30. Miljković, D., Gajić, L., Kovačević, A., & Konjović, Z. (2010). The use of data mining for basketball matches outcomes prediction. In *Intelligent Systems and Informatics (SISY), 2010 8th International Symposium on* (pp. 309-312). IEEE.
31. Mohammad R., Thabtah F., and McCluskey L. (2015). Tutorial and critical analysis of phishing websites methods. *Computer Science Review Journal*, 17: 1–24.
32. Mohamed, H., Negm, A., Zahran, M., & Saavedra, O. C. (2015). Assessment of artificial neural network for bathymetry estimation using High Resolution Satellite imagery in Shallow Lakes: case study El Burullus Lake. In *International water technology conference*.
33. Nunes, S., and Sousa, N. (2006). Applying Data Mining Techniques to Football Data from European Championships. *Proceedings of Conferência de Metodologias de Investigação Científica*, 4-16.
34. Purucker M. C. (1996). Neural network quarterbacking. *IEEE Potentials*, 15: 9–15.
35. Prasitio D., and Harlili D. (2016). Predicting football match results with logistic regression. Proceedings of the 2016 International Conference On Advanced Informatics: Concepts, Theory And Application (ICAICTA). 16-19 Aug. 2016, Penang, Malaysia.
36. Quinlan, J. (1993). *C4.5: Programs for machine learning*. San Mateo, CA: Morgan Kaufmann.
37. Raschka (n.d.). k-fold.png. Retrieved from https://sebastianraschka.com/images/faq/evaluate-a-model/k-fold.png
38. Rumelhart, D. E., Hinton, G. E., Williams, R. J. (1986). Learning representations by back-propagating errors. *Nature,* 323 (6088): 533–536.
39. Shearer, C. (2000). The CRISP-DM model: the new blueprint for data mining. *Journal of data warehousing*, *5*(4), 13-22.
40. Sutton, R. S. (1986). Two problems with backpropagation and other steepest-descent learning procedures for networks. Proc. 8th Annual Conf. Cognitive Science Society
41. Thabtah F., Hammoud S. and Abdeljaber H. (2015). Parallel Associative Classification Data Mining Frameworks Based Mapreduce. To Appear in *Journal of Parallel Processing Letter*. March 2015. World Scientific.

42. Thabtah F., Mohammad R. M., McCluskey L. (2016). A dynamic self-structuring neural network model to combat phishing. Neural Networks (IJCNN), 2016 International Joint Conference, pp. 4221-4226. Vancouver, Canada, 2016.
43. Wilkinson, L., Anand, A., and Tuan, D. (2011). CHIRP: a new classifier based on composite hypercubes on iterated random projections. In *Proceedings of the 17th ACM SIGKDD international conference on Knowledge discovery and data mining (KDD),* 11:6–14.
44. Wiseman, O. (2016). *Using Machine Learning to Predict the Winning Score of Professional Golf Events on the PGA Tour* (Doctoral dissertation, Dublin, National College of Ireland).
45. Witten, I. H., Frank, E., and Hall, M. A. (2011). *Data mining: Practical machine learning tools and techniques*. Burlington, MA: Morgan Kaufmann.
46. Yvan, N. (2000). Flexible Conjugate Gradients. *SIAM Journal on Scientific Computing*, 22 (4): 1444.