

## A Conditional Fuzzy Inference Approach in Forecasting

Arman Hassanniakalager\*, Georgios Sermpinis\*\*, Charalampos Stasinakis\*\*\* and Thanos Verousis\*\*\*\*

This study introduces a Conditional Fuzzy inference (CF) approach in forecasting. The proposed approach is able to deduct Fuzzy Rules (FRs) conditional on a set of restrictions. This conditional rule selection discards weak rules and the generated forecasts are based only on the most powerful ones. Through this process, it is capable of achieving higher forecasting performance and improving the interpretability of the underlying system. The CF concept is applied in a series of forecasting exercises on stocks and football games datasets. Its performance is benchmarked against a Relevance Vector Machine (RVM), an Adaptive Neuro-Fuzzy Inference System (ANFIS), an Ordered Probit (OP), a Multilayer Perceptron Neural Network (MLP), a k-Nearest Neighbour (k-NN), a Decision Tree (DT) and a Support Vector Machine (SVM) model. The results demonstrate that the CF is providing higher statistical accuracy than its benchmarks.

**Keywords: Forecasting; Conditional Fuzzy Inference; Fuzzy Rules; Classification;**

\* School of Management, University of Bath, Wessex House, Bath BA2 7AY, United Kingdom  
(A.Hassanniakalager@Bath.ac.uk)

\*\* University of Glasgow Business School, University of Glasgow, Gilbert Scott Building, Glasgow G12 8QQ, United Kingdom (Georgios.Sermpinis@glasgow.ac.uk), Corresponding Author

\*\*\* University of Glasgow Business School, University of Glasgow, Gilbert Scott Building, Glasgow G12 8QQ, United Kingdom (Charalampos.Stasinakis@glasgow.ac.uk)

\*\*\*\* Essex Business School, University of Essex, SO4 3SQ, United Kingdom  
(av18527@essex.ac.uk)

## 1. Introduction

In this research a Conditional Fuzzy inference (CF) approach for forecasting is introduced. In CF a set of Fuzzy Rules (FRs) is generated in the in-sample and a power to each rule is assigned based on the rule's frequency and accuracy. Thereafter all rules are ranked and the strongest are applied in the out-of-sample. For each out-of-sample data point, membership functions are calculated and if certain conditions are met, a weighted average of the strongest FRs is estimated. These conditions are based on the in-sample and ensure that the rule is strong enough for out-of-sample estimation. If the conditions for a data point are not satisfied and no strong rule is close, then no forecast is generated for that point. Our CF process avoids weak rules and ensures that the generated forecasts are based on a weighted average of the most powerful FRs.

The proposed methodology is advantageous to a series of issues. Firstly, it is useful in problems where the practitioner or the researcher is interested only in strong signals and the risk of having a poor forecast is greater than having no forecast. For example, in financial trading, betting on a sport or in any other sensitive decision-making process, poor forecasts lead to financial losses. In these environments the underlying series are volatile and decision makers are risk averse; abstaining from the market is better than making decisions under uncertainty. Secondly, CF can improve the out-of-sample accuracy of the underlying system and offer transparency at the same time. This is beneficial in problems where complex models (such as machine learning) are necessary. Thirdly, the generated rules can be easily applied by non-experts as the number of rules applied is small and they are easily replicated. Lastly, the chosen rules do not suffer from over-fitting or under-fitting problems. In machine learning and complex models, it is common for the performance to be driven by extensive experimentation. In these cases, the generated forecasts can be due to over-specification and have no generalisation value. In CF the weak signals are dropped and the noise within the model is reduced.

An extensive empirical study is designed and implemented in order to test the merits of the proposed methodology. CF is applied to the two of the most popular forecasting exercises, namely stocks' and football games results prediction. Both forecasting exercises are characterized by their high return/risk nature and popularity amongst forecasters. Thus, it seems the perfect exercises for our CF approach. More specifically, the proposed model will forecast the result and the number of goals within football games at the three biggest football championships (the English Premier League, Italian Seria A and Spanish La Liga) and the direction of the Pfizer Inc. (PFE) and Goldman Sachs Group (GS) stocks. The forecasts are evaluated through a realistic betting exercise based on the Betbrain average odds for the case of football games and through a simple trading exercise for the case of stocks. For the case of the football games, CF is combined with a Relevance Vector Machine (RVM) that reduces the dimension of the input vector and a series of CF rules are extracted. For the stocks' prediction exercise, CF is applied directly to the dataset as the inputs vector has already small dimensions. Our CF forecasts are benchmarked against those generated by an Adaptive Neuro-Fuzzy Inference System (ANFIS) model (the most popular fuzzy extraction structure), a single RVM model, an Ordered Probit (OP) model, from a Multilayer Perceptron (MLP) neural network, a k-Nearest Neighbour (k-NN model), a Decision Tree (DT) and a Support Vector Machine (SVM) model. The comparison between CF and ANFIS reveals the benefits of the

proposed framework compared to an unconditional (ordinary) Fuzzy Inference Systems (FIS), whilst benchmarking the results with the single RVM will validate whether CF can increase the accuracy of the underlying system. OP, MLP, k-NN, DT and SVM are some of the most popular techniques in football games and stocks forecasting. Thus, the comparison with CF offers a benchmark close to the relevant literature.

Fuzzy Logic (FL) was introduced by Zadeh (1965). Its motivation is driven by the work and functioning of the human mind. Even though a tremendous amount of information presents itself to a human in any given situation – an amount that would ‘choke’ a typical computer – the human mind has the ability to discard the irrelevant elements and to concentrate only on the information that is relevant. The ability of the human mind to deal only with the part of the information that is relevant is connected with its possibility to process fuzzy information (Zadeh, 1983). In that way the amount of information the brain has to deal with is reduced to a manageable level and the decision process is faster, simpler and more effective. FL can be described as multi-value logic that allows for intermediate values instead of the conventional evaluations like yes/no, up/down, on/off. These linguistic expressions are called Fuzzy Rules (FRs). FRs are conditional statements in the “if-then” form. They are widely used in studies with systems whose actions are incomprehensible (Bellman and Zadeh, 1970). In these studies, researchers apply an FIS to the in-sample to generate a set of FRs that maps the interactions between variables (see amongst others Hruschka, 1988; Teodorović, 1994; Piramuthu, 1999; Kuo, 2001; Chang et al., 2008; and Gradojevic and Gençay, 2013). This set of FRs reveals the structure of the system and how inputs relate to the outputs. However, this structure is rarely retained to the same extent in the out-of-sample.

Several conditional approaches have been proposed in the literature aiming to improve the performance of FL. These modifications focus on different stages of FIS development. Earlier studies focus on the ‘inference’ stage and try to improve the quality of a trained FIS (see among others Fukami et al., 1980; Mizumoto, 1982; Sugeno, 1985). These techniques can be categorically termed as ‘conditional inference’ methods. The ANFIS is motivated by these works and provides a generalized double-path training algorithm which is shown to be superior to its precedent alternatives (Jang, 1993). An alternative conditional approach to FL can be amending the implementation (i.e. defuzzification) phase. The motivation here is to improve the performance of FL in practice rather than improving the quality of the FRs. In this case the inference phase is not affected, and the conditional approach is applied to an initially trained FIS. Therefore, any training/inference technique (either conditional or unconditional) can be used to generate an FIS. The algebra for such modifications is firstly introduced in Singpurwalla and Booker (2004). Kadri (2014) introduces an adaptive approach to defuzzification of FRs conditionally in signal-process. The CF introduced in this paper is motivated by the previous works on conditional defuzzification and provides a generalized straightforward procedure to improve the performance of any underlying FIS. In particular, CF proposes a conditional framework for implementing an FIS in a robust way based on a training dataset and the environment the dataset is placed. This allows us to adjust an FIS based on the application by looking at endogenous factors (training dataset) and exogenous ones (environment). These factors are introduced to an underlying FIS which could be trained conditionally or unconditionally. The aim of CF is to improve the performance of the underlying model no matter how the original model is trained. In terms

of the FIS structure, the part affected by CF is the defuzzification stage. The proposed changes are essential in problems with dynamic data (such as finance or economic series or football games results) where the 'input-output' relationship varies through time and the generated FRs in the in-sample cover only a part of the out-of-sample. This leads to a reduction of accuracy and to systems that are interpretable but inaccurate. Another problem is the strength of the generated in-sample FRs. Different rules have different strengths and different degrees of accuracy. Applying weak FRs in the out-of-sample leads to poor forecasts.

There are a few studies that apply FL to the context of sports and stocks forecasting. Rotshtein et al. (2005) propose a model for predicting the result of a football match from the previous results of both teams. Although they suggest that their model accounts for nonlinear dependencies through fuzzy knowledge, they are focusing on a very illiquid football betting market, the one of Finland, and they do not explore the betting profitability of their forecasts. Trawinski (2010) proposes a fuzzy model for extracting FRs in order to predict basketball game results. The author compares ten FR learning algorithms against a standard OLS, but they do not present robust empirical results or any betting application of football. Finally, Bastos and da Rosa (2013) propose a static and a dynamic Poisson-Gamma model to predict the outcome of World Cup football results based on the number of goals scored by each team. In their application, a fuzzy C-means algorithm is used for clustering. Nonetheless, they do not offer any betting application, while their forecasting exercise is limited to a football event occurring only every four years. Wang (2002) propose the use of a fuzzy grey model for intraday stock price forecasting. The authors suggest that the model is efficient in terms of statistical accuracy and user-friendly, but its shortcoming is that it is not adaptive. A FL approach for forecasting stock market trends in ASE and NYSE stock exchanges is also presented by Atsalakis and Valavanis (2009). Their empirical findings demonstrate the superiority of the proposed neuro-fuzzy approach compared to another thirteen models of the literature, while they challenge the weak form of the Efficient Market Hypothesis (EMH). Sun et al. (2015) forecast the Chinese stock index future prices using fuzzy sets and a multivariate fuzzy time series method. The authors combine traditional fuzzy time series models and a rough set method to extract strong rules and make superior predictions. Oztekin et al. (2016) apply ANFIS for forecasting stock price returns of the Istanbul Stock Exchange. Its performance is benchmarked with Support Vector Machines (SVMs) and Neural Networks (NNs). Although ANFIS is not outperforming the other models, the authors suggest that the combination of the three approaches is beneficial for such forecasting exercises. The merits of ANFIS are also validated by the recent work of Atsalakis et al. (2019), who apply successfully neuro-fuzzy techniques in predicting Bitcoin prices.

The majority of researchers in football games apply OP, probabilistic or machine learning methods. OP applications in football forecasting are common due to the ordered nature of the football result variable (Away team win, draw and Home team win). Kuypers (2000) applies OP in order to test how the betting market participants utilize available information and suggests that betting arbitrage is possible. Goddard and Asimakopoulou (2004) and Goddard (2005) apply OP to forecast English football outcomes based on teams' quality and past performance indicators. Their results indicate that this approach, which is followed also in our study, is robust and provides high forecasting performance. Other studies utilize SVM frameworks. Vlastakis et

al. (2008) show that SVM performs better than a Poisson model when applied to the task of predicting European football match scores. Gomes et al. (2016) and Martins et al. (2017) also utilize SVMs to forecast the number of football corners, goals and outcomes of Premier League and several other championships respectively. Recently, Baboota and Kaur (2018) show that SVM with radial basis kernel is efficient in predicting the rank probability scores for the English Premier League using teams' past performance indicators as inputs. In terms of probabilistic approaches, Dixon and Coles (1997) forecast the number of goals and the results of football games with forecasting systems based on the Poisson distribution. Crowder et al. (2002) estimate the football game results' probabilities with refinements of the independent Poisson model. Koopman and Lit (2015) and Angelini and De Angelis (2017) refine previously traditional Poisson models and improve upon the model of Dixon and Coles (1997). NNs have also been used in similar forecasting exercises. For example, Constantinou et al. (2012) suggest a pi-Bayesian Network (BN) for English Premier league match forecasting. Their model applies time-dependent data with weighted degrees of uncertainty and exhibits high statistical accuracy along with profitability over the publicly available odds. A similar approach is applied by Owramipur et al. (2013). Huang and Chen (2011) apply MLPs to forecast the match outcomes of 2006 World Cup, while Şahin and Erol (2018) suggest that MLP is superior in forecasting football matches' attendance compared to ANFIS.

Similar approaches are applied for stock market predictions. Hausman et al. (1992) show that OP is superior in capturing the stock price changes compared to simple linear regression approaches, while Purda (2007) utilize OP to analyse stock market reactions originating from anticipated and surprising bond rating changes. Yang and Parwada (2012) also incorporate OP to predict the direction of price movements of nine Australian stocks. Their average proportion of correct forecasts exceeding the seventy percent exhibits the success of OP in their forecasting application. Tay and Cao (2002) modify SVM to select fewer support vectors and their results show that SVM forecasting accuracy is improved when applied in future contract price prediction. Huang and Wu (2008) apply a Bayesian SVM in order to forecast returns of four major stock indices and they clearly exhibit the RVM success over other traditional SVM approaches. Gong et al. (2016) apply a hybrid SVM to predict stock return patterns, while Chen and Hao (2017) combine SVM and k-NN algorithms for stock index price forecasting. Both studies show the benefits of using SVM classifiers for financial forecasting. Guresen et al. (2011) apply a series of NNs to the task of forecasting NASDAQ stock prices and they find that classical MLP structures are superior in terms of statistical accuracy. Wang et al. (2011) also illustrate the merits of back-propagation NNs for similar financial forecasting exercise. Ticknor (2013) suggests that the Bayesian regularization of NNs is beneficial for obtaining more accurate forecasts of two US stocks. On the other hand, as shown by Patel et al. (2015) Bayesian classifiers struggle to outperform MLP and SVM structures when it comes to forecasting stocks' directional movements through technical indicators.

With the previous background in mind, our proposed methodology is benchmarked against some of the most popular models used in football games and stocks' prediction. None of the previous studies applies a conditional fuzzy framework similar to the one introduced in our study. The remainder of this paper is organised as follows: Section 2 introduces the concept of CF and describes in detail the methodology and the benchmarks applied.

Section 3 describes the dataset used and the empirical application to football betting. Section 4 presents the conclusion. Section 5 includes some remarks on the practical implications and future works while in the Appendix several technical details are explored. Finally, the mathematical background on ANFIS, RVM and OP and extra empirical results are placed on an online Appendix.

## **2. Methodology**

This section summarizes all the relevant information regarding the proposed methodology. Initially, the CF method is motivated. Then, its mathematical and technical description is explained in detail. Finally, short descriptions of the benchmarks are provided.

### **2.1 Conditional Fuzzy Inference**

#### **2.1.1 Motivation**

FL frameworks are traditionally applied in decision processes, where the level of uncertainty is high and a complete problem formulation is difficult. In FL, an FIS is applied and a set of FRs are extracted for prediction and decision-making (Sugeno, 1985)<sup>1</sup>. The most popular FIS is the ANFIS of Jang (1993). The mathematical formulation of ANFIS is presented in the online Appendix S.1.

The methods of defuzzification vary in the literature. For example, Jang (1993) classifies possible defuzzification functions in three main categories. In the first category a crisp output for each rule is estimated and then a weighted average of each rule's output forms the aggregate output as the fuzzy model's output. The applied weights are based on the firing strength of the rules and the output membership function. The second category of defuzzifiers involves applying a "max" operator to qualified fuzzy outputs that meet some criteria (e.g. minimum firing strength) and then the aggregate output is determined by a function such as the mean of maxima, the maximum criterion and so forth. The third category is the Sugeno approach (adopted in ANFIS), namely for each rule's output, a general linear model based on inputs is estimated and the aggregate output is the weighted average of all rules involved.

The conditional approach in the second category that is overlooked in the ANFIS structure can be promising. If all FRs extracted through the training process are not of the same quality, the user might prefer to use the FRs partially rather than entirely. The ANFIS adopts the double pass algorithm to optimise the rule specifications but may lead to overfitting and to undesirable performance for the previously unobserved points. In other words, the problem is that in the out-of-sample the membership grade may not be strong enough for some or all rules. This eventually leads to bad forecasts. This problem is common in machine learning where the out-of-sample performance is either considerably worse than the in-sample's one (over-fitting), or the model is inadequately trained (under-fitting).

---

<sup>1</sup> For the identification and parameterization of a FIS the interested reader should refer to studies such as Bezdek et al. (1984) and Jang (1993).

The proposed methodology in this study combines the merits of the second and the third defuzzification approach. A crisp output based on a simple linear model is highly favourable in case of large-scale problems. On the other hand, when a dataset comes with remarkable noise a selective procedure for the FRs aggregation can control the uncertainty. The next section explains the algorithm for the CF.

### 2.1.2 Algorithm

Let us define the firing strength for rule  $i = 1, \dots, M$  as:

$$\hat{w}_i^* = \prod_{k=1}^K \hat{\mu}_{ik}(x_k^*, mf) \quad (1)$$

where  $\hat{\mu}(\cdot)$  is the membership grade calculated based on a given membership function  $mf$  and a  $K$ -dimensional test data point with input vector  $x^* = [x_1^*, \dots, x_K^*]$ . The choice of the membership function can be central to the model's fit and its predictive ability. Based on Guillaume (2001), for most inputs of a fuzzy system, it can be intuitively accepted that as an observation places farther from the centre of a rule, the associated membership grade falls. Thus, it can be assumed that the membership function needs to come with a bell shape. The Gaussian membership function offers such smooth shape around centres of clusters and therefore it is a fitting one for our case. Such choice is suitable for forecasting applications like ours (see amongst others Wang and Mendel, 1992; Jang, 1993; Kuncheva, 2000; Cheng and Lee, 2001; and Akkoç, 2012). For football forecasting specifically, Vlastakis et al. (2008), Igiri (2015) and Baboota and Kaur (2018) that apply SVM to football match prediction, they also employ the Gaussian kernel to their models. We also experimented in the in-sample with the univariate/multivariate sigmoid, triangular and trapezoid membership functions. In all cases, the in-sample accuracy was worse than the one acquired with the Gaussian membership function. Under a Gaussian distribution assumption, the membership grade for the  $k$ -th element of rule  $i$  is given by:

$$\hat{\mu}_{ik}(x_k^*, mf) = \exp\left\{-\left(\frac{x_k^* - c_{ik}}{a_{ik}}\right)^2\right\} \quad (2)$$

where  $a_{ik}$  and  $c_{ik}$  are specified for the FRs through the training process.

The CF proposes eligibility criteria for the defuzzifier function to discard rules with unsatisfactory membership grades from the set of rules that form the fuzzy output. This process involves two components: the criteria set-up and a satisfactory threshold ( $\Theta$ ). The criteria set-up determines the evaluation function for each rule and can take one of the forms offered in equations (3) and (4) below:

$$C_i^* = \begin{cases} 1, & \hat{w}_i^{*1/K} \geq \Theta \\ 0, & \text{Otherwise} \end{cases} \quad (3) \quad \text{or alternatively} \quad C_i^* = \begin{cases} 1, & \min_k(\hat{\mu}_{ik}(x_k^*, mf)) \geq \Theta \\ 0, & \text{Otherwise} \end{cases} \quad (4)$$

Where  $\hat{w}_{x^*,i}^{*1/K}$  is given by equation (1).

The criteria set-up can be defined as an average of membership grades for each rule compared to  $\Theta$  (equation (3)) or as the minimum membership grade compared with  $\Theta$  (equation (4)). Both procedures generate an indicator as to whether the rule should be applied or not. In the equations (3) and (4),  $C_i^*$  is the indicator to determine whether the rule  $i$  is qualified ( $C_i^* = 1$ ) as an eligible rule for the specific test observation to be

applied. Alternatively, the rule is considered as weak if  $C_i^* = 0$ . The other component to shape the eligibility criteria is the threshold parameter.  $\Theta$  determines the average power required for a rule to be considered as a strong rule. We propose defining the  $\Theta$  as:

$$\Theta = \max (w_\lambda^{1/K}, \Lambda) \quad (5)$$

Under this setting, the threshold level originates from two sources. Firstly, from the training dataset reflected in  $w_\lambda^{1/K}$  and secondly, from a general value allocated to membership grade that represents a strong rule ( $\Lambda$ ).  $\Theta$  is simply the larger of the two quantities. Alternative settings for  $\Theta$  depending on the nature of application might be used as well. In equation (5)  $w_\lambda^{1/K}$  is the  $K$ -th root of average firing strength of a rule  $\lambda$  that is the minimum acceptable level of firing strength based on the training dataset. In other words,  $w_\lambda^{1/K}$  is the endogenous threshold based on the in-sample. To set  $\lambda$  the following procedure is pursued:

Firstly, the arithmetic average membership grade for each rule  $i$  and input element  $k$  is computed over the in-sample dataset  $\bar{\mu}_{ik}(\cdot)$ . Secondly, the average firing strength  $w_i$  is given by

$$w_i = \prod_{k=1}^K \bar{\mu}_{ik}(\cdot) \quad (6)$$

Thirdly, the FRs are sorted in a descending order of average firing strength (sorted list). In the sorted list, the rules on the top are ones that data-points place closest to the centres of the FRs. An arbitrary percentile level for the top strongest rules is selected based on the complexity of the problem and the aims of the researcher (e.g. ten). Then, the 90<sup>th</sup> percentile rule on the sorted list is selected. The equivalent index (row) of the selected rule on the original list of rules is the  $\lambda$ .

Unlike  $\lambda$  and  $w_\lambda^{1/K}$  that are endogenous,  $\Lambda$  is a fixed quantity that determines the general quality of a strong rule. The membership grade ranges from zero to one. The closer the data-point gets to the centre of a rule, the higher will be the membership grade. Ideally, if the data-point and the rule centre overlap, the membership grade is equal to one. The further away the data-point gets, the lower the membership grade becomes. This implies that the FR is getting weaker. The choice of  $\Lambda$  depends on the problem under study and the practitioners' needs. In problems where uncertainty is high, and a wrong forecast can have a considerable impact,  $\Lambda$  should be set high (over 0.8). This will reduce on the one hand the uncertainty but on the other hand, it will also reduce the number of CF forecasts generated. In problems where wrong forecasts have a small effect on the utility function of the practitioner,  $\Lambda$  can be set lower. This will lead to more CF forecasts that retain some level of risk.

The combination of the endogenous and the exogenous threshold ensures that the applied rules for forecasting are correctly fitted. Overfitting and under-fitting are the most significant challenges in machine learning. If an FIS is over-fitted, the average membership grades, the average firing strengths and accordingly the  $w_\lambda^{1/K}$  are high. As the model is overly specified to match the training samples, once the out-of-sample data-points are fed the average membership grades for the new observations will fall below the  $w_\lambda^{1/K}$ . In the CF, the model drops these rules and looks for any remaining rule that can satisfy the  $w_\lambda^{1/K}$  threshold. This ensures that the model forecasts properly even if the original FIS is over-fit. If the CF is unable to identify for a single point any relevant



strong rule, then no forecast is generated. Similarly, in case of an under-fit model, the in-sample measure  $w_\lambda^{1/K}$  is low. However, there might be still some rules that are strong and fit enough for certain points in the out-of-sample.  $\Lambda$  ensures that under-fit rules with low  $w_\lambda^{1/K}$  are not applied.

The proposed CF modifies the last layer. The modification is able to grasp the strongest possible FRs and drop the mediocre and poor ones. In the CF model, the final layer outcome can be presented as:

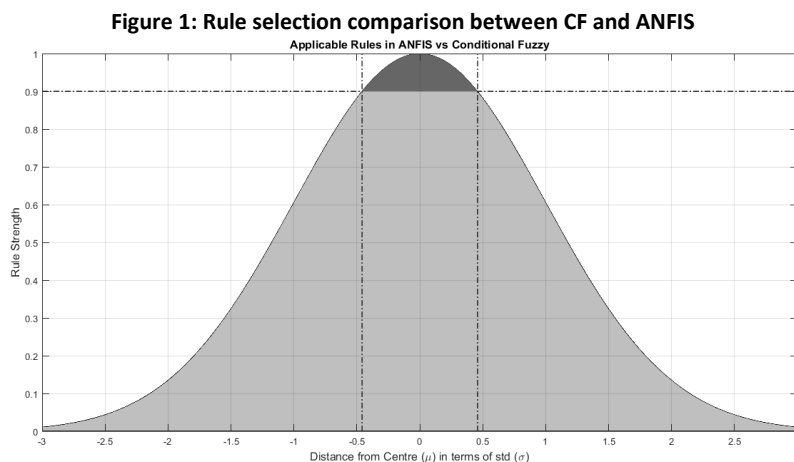
$$\hat{O}^* = \hat{O}_5(x^*, O_4, O_3, O_2, O_1, A, C, P, Q, R, \lambda, \Lambda) \quad (7)$$

As the equation implies the difference in CF compared to ANFIS is in the defuzzifier module, where the aggregate output is modified to include the eligibility criteria. The CF output  $\hat{O}^*$  is computed through a node function  $\hat{O}_5$  given the input vector  $x^*$ , ANFIS specification, and threshold parameters. The ANFIS specification include training nodes  $(O_1, \dots, O_4)$ , premise parameters set  $\{A, C\}$  and consequent parameters set  $\{P, Q, R\}$ . The CF threshold parameters are  $\{\lambda, \Lambda\}$ . We propose two defuzzification functions. Firstly, by calculating a weighted average of outputs for qualified FRs and secondly the output of the strongest qualified rule represented in equations (8) and (9) below:

$$O'_5 = \frac{\sum_{i=1}^M \hat{w}_i C_i f_i}{\sum_{i=1}^M \hat{w}_i C_i} \quad (8) \quad \text{or alternatively} \quad O'_5 = f_v(\cdot), \quad v = \arg \max_i(\hat{w}_i) \quad (9)$$

In equations (8) and (9),  $O'_5$  is the node function for the last layer of CF. For each rule  $i$  the firing strength  $\hat{w}_i$  is estimated by equation (1), while the conditional indicator  $C_i$  is given by equation (3) or (4) based on threshold parameters estimated in (5) and (6).  $f_i$  is the regression output estimated by the FR (traditionally as with any Sugeno fuzzy approach). Finally, in (9)  $v$  is the argument of maxima for the firing strength  $\hat{w}_i$ .

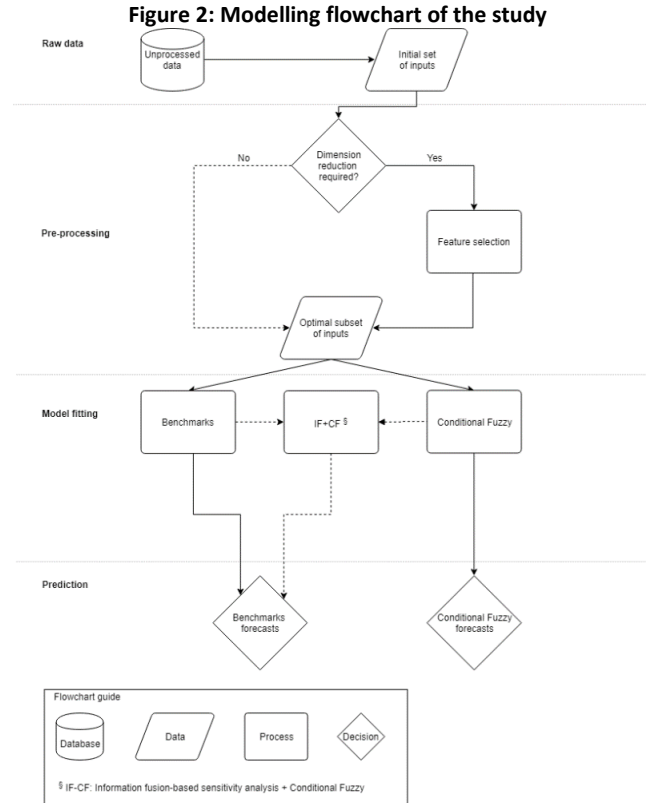
The above-mentioned modification is crucial, as it can provide predictions for points (observations) where strong rules are nearby and at the same time satisfy endogenous and exogenous threshold specifications. The final CF outcome is a weighted average only of the strongest rules. This attribute is innovative in FL, as it is able to offer interpretability of the final result, protection against substantial forecast errors and under- or over-fitting in the underlying decision-making system. To further demonstrate the novelty of the CF method, a comparative graph of the rules selected in the case of CF and ANFIS is presented in Figure 1.



Note: The distribution shows the selection of fuzzy rules with respect to the distance from centres ( $\mu$ ) in units of standard deviation ( $\sigma$ ) and the corresponding strength. In ANFIS (whole distribution) all fuzzy rules are accepted irrespective of the strength, but in CF (dark pick of the distribution) only the strongest rules are selected to for defuzzification stage.

From the Figure above, it is clear that CF selects rules that are on the pick of the distribution of the ANFIS selection and within small distances from the centre.

The flowchart for the modelling section of the paper is presented in Figure 2.



Note: The figure presents different steps how the empirical analysis of this paper is conducted. The solid-line arrows show the regular link between different blocks while the dashed-line arrows reflect the route for generating combined forecasts through information fusion-based sensitivity analysis.

Figure 2 shows that in the first stage, the system inputs and target series are generated from the raw datasets. The next step is to evaluate whether a dimension reduction is needed in the inputs' matrix. In case that a technique is applied, the output of the pre-processing stage is an optimal subset of inputs which explains best the dynamics in the target series. The RVM is used in the feature selection stage to extract the most informative input vectors (see Section S.2). Alternatively, if no dimension reduction technique is needed, the feature space remains unaffected. In the following stage (i.e. model fitting) the CF and the benchmarks are trained to make forecasts as individual models. Finally, the CF and the benchmark models are used to generate combined forecasts through the information fusion-based sensitivity analysis (IF) as discussed in Section 2.3.

In our application, we apply CF following the evaluation function of equation (3), while the CF output is calculated based on equation (8)<sup>2</sup>. Here it should be noted that the CF model synthesis is based on the algebra offered by Zadeh (1965), Jang (1993) and Singpurwalla and Booker (2004). The alternative definition of the

<sup>2</sup> The results of our empirical application are almost similar to both evaluation functions (equations (3) or (4)) and defuzzification functions (equations (8) or (9)).

indicator in equations (3) and (4) are based on AND and PRODUCT operators in the FL context, whereas the defuzzification methods in equations (8) and (9) are based on the discussion of Jang (1993). The theoretical analysis of this algebra is provided in Singpurwalla and Booker (2004).

In terms of applying CF, football betting is a high risk-high return exercise. A wrong bet usually results in the total loss of the invested capital. Thus,  $\Lambda$  is set high to 0.90. This will result in a stricter selection for CF and resembles the real-world practice where bettors are highly selective regarding the games on which they will bet. The setting for  $\Lambda$  is more lenient in stock market application, since the price changes are within a more limited range. In this case the  $\Lambda$  is chosen between 0.5, 0.75, and 0.9 based on number of rules generated. In the out-of-sample, for each point the membership grades for the FRs are estimated and based on the eligibility criteria, a signal might be generated. In case a signal is generated the relevant bet is placed. If a signal is not generated, we abstain from the market. A detailed illustrative example of the CF application is provided in Appendix A.

## 2.2 Benchmark Models

This approach is benchmarked against a basket of forecasting algorithms, namely a RVM, an ANFIS, an OP, an MLP, a k-NN, a DT and an SVM model. RVM is a sparse kernel model to tackle the problem of large-scale data-processing. The Bayesian learning framework used in RVM can generate precise forecasts while reducing the feature space to the most important (“relevant”) vector. RVM’s solid probabilistic approach allows the inference of the optimal hyperparameters and vectors’ weights from the data. In our proposed approach, RVM will serve for both feature selection and prediction for the football exercise and for prediction for the stocks exercise. The selected features, Relevance Vectors (RVs), are applied to ANFIS and to our CF approach. For the mathematical details of ANFIS and RVM, we refer the reader to the online Appendix (S.1 and S.2 respectively).

For the case of the football games exercises, we reduce the dimension of inputs vectors with an RVM and an IF approaches. The performance of all our models is better when RVs are selected as their inputs instead of the ones generated by the IF. Thus, in our application the inputs of the benchmark models (ANFIS, OP, MLP, k-NN, DT and SVM) and of our CF approach are the selected RVs from the RVM. Nevertheless, for the sake of comparison, we present also the performance of CF with the IF generated inputs vector (CF-IF)<sup>3</sup>. For the stocks’ prediction exercise, the input vector is small enough to be fed directly to CF. In this case, all models under study use the same set of inputs (ANFIS, OP, MLP, k-NN, DT and SVM).

The flowchart of Figure 2 makes apparent that our CF model needs to be benchmarked with ANFIS. With this comparison, we will be able to quantify if the modification of the last layer of ANFIS is successful and whether the conditional approach is beneficial. RVM can optimize the global parameters that affect the input variable space. Its Bayesian probabilistic approach is beneficial by firstly producing sparse solutions able to reduce the input space for other models and secondly optimal parameters that allow the RVM to forecast efficiently. This

---

<sup>3</sup> In the case of IF, we calculate the variable sensitivity scores by means of information-fusion as described in Online Appendix S.3. We also explored the utility of Principal Component Analysis (PCA) on the same task. It provides the worse sets of results from all those obtained by RVM or IF. For the sake of space, they are not presented in-text but are available upon request.

is very important, as traditional methods such as cross-validation are not able to achieve this (Tipping, 2001). OP originates from Logistic Regression (LR) models and tries to estimate the probability of each outcome for a dependent variable. Depending on the number and nature of the possible choices that the dependent variable can take, the choice of the LR type can be binomial, multinomial and ordinal. Given that the outcomes of football matches can be ordered as Home team winning, draw and Away team winning the game and the stocks' direction as higher or lower, the OP model is suitable for our application. The mathematical details of OP (as earlier with ANFIS and RVM) are provided in the online Appendix S.4.

The MLPs are the most popular and simple architecture of feed-forward Neural Networks. MLPs are consisted by three different types of layers. The input, the hidden and the output layer. The input layer includes the inputs nodes where each node corresponds to a different explanatory variable. The output layer is formed with one node and represents the output of the NN. The hidden layers separate the input from the output layers and define the amount of complexity the model is capable of fitting. Following the guidelines of Zhang et al. (1999) and the characteristics of our datasets, our MLP is consisted by three layers with one hidden layer with ten hidden nodes. k-NN is based on the idea that pieces of time series in the past have patterns that might resemble to pieces in the future. It locates similar patterns in terms of nearest neighbours and the Euclidean distance and use these patterns to predict the immediate future. The algorithm uses local information to forecast and makes no attempt to fit a model to the whole times series at once. Our algorithm is constructed based on the guidelines of Guégan and Huck (2005) and the characteristics of our datasets. DT is a machine learning technique that build classification models in the form of tree structures. The dataset breaks down in a smaller subsets while at the same time a decision tree is incrementally developed. In our application, we apply the CART procedure of Breiman et al. (1984). The DTs are easily to understand and interpret and any nonlinear relationships between their parameters do not affect the algorithm performance. SVM is a class of supervised machine learning model that has an identical functional form with RVM which does not use Bayesian inference. Compared to RVM, SVM always converge to global optima and require less training time. However, the classification accuracy of RVM is higher than the one of SVM (Rafi and Saikh, 2013). In this study, we design our SVM model based on the 10-fold cross validation approach of Oztekin et al. (2016). For the underlying mathematics of the MLP and the k-NN we refer the reader to Zhang et al. (1999) and Guégan and Huck (2005) respectively. A technical description of DT can be found in Breiman et al. (1984) and of SVM in Vapnik (1995).

### **2.3 Information Fusion-Based Sensitivity Analysis (IF)**

For researchers that delve into intensive data mining tasks, performances of different sets of models are evaluated over different datasets. Therefore, identifying the best approach in terms of data mining methods is crucial, but at the same time it proves to be a challenge. One solution is to apply “composite forecasts” from multiple models in order to improve accuracy and obtain better results from a set of data mining models (Sevim et al., 2014). Information fusion is a technique that brings the above approach one step further, as it is able to account for and prioritize on information derived from multiple forecasting models. This technique is able to reduce bias and uncertainty that can be attributed to individual techniques (Oztekin et al.,2013; Oztekin et al.,

2016). Additionally, it offers a viable solution for coping with dimensionality issues common in classification studies (Bogaert, et al., 2018). The latter is achieved by applying IF for variable selection by creating a rank order of the importance of the input variables through the forecasting models involved.

Assuming the  $\check{Y}$  dependent variable and a set of independent variables  $\check{X} = \{\check{x}_1, \check{x}_2, \dots, \check{x}_N\}$ , an individual predictive model  $\check{f}$  can be expressed as:

$$\check{Y}_i = \check{F}_i(\check{x}_1, \check{x}_2, \dots, \check{x}_N) = \check{F}_i(\check{X}) \quad (10)$$

It should be noted that  $F$  is not constrained by any form. Considering  $\check{f}$  individual models and a linear combination as a fusion operation  $\check{F}$ , the infusion model can be expressed as:

$$\check{Y}_{fusion} = \check{F}(\check{Y}_1, \check{Y}_2, \dots, \check{Y}_i) = \sum_{i=1}^{\check{f}} \check{\alpha}_i \check{F}_i(\check{X}) = \check{\alpha}_1 \check{F}_1(\check{X}) + \check{\alpha}_2 \check{F}_2(\check{X}) + \dots + \check{\alpha}_i \check{F}_i(\check{X}) \quad (11)$$

Where  $\check{\alpha}_i$  is the weighting coefficient for each individual model and  $\sum_{i=1}^{\check{f}} \check{\alpha}_i = 1$ .

According to Oztekin et al. (2013), the weights are proportional to the forecasting performance of each model. For that reason, it is expected that information extracted from well performing classifiers should be found more important than information obtained from those that do not perform well. In our case, we follow Díez-Pastor et al. (2015) to calculate Area Under Curve (AUC) estimates that will be used as our  $\check{\alpha}_i$ . In order to calculate the AUC estimates we need an illustration of a confusion matrix as in the figure below.

**Figure S.1:** Confusion Matrix

	<b>P+</b>	<b>P-</b>
<b>O+</b>	<b>T+</b>	<b>F-</b>
<b>O-</b>	<b>F+</b>	<b>T-</b>

The confusion matrix can describe the binary classification task with the observed class  $\{O^+, O^-\}$  and predicted class  $\{P^+, P^-\}$  and its classification performance by means of true positives ( $T^+$ ), false positives ( $F^+$ ), false negatives ( $F^-$ ) and true negatives ( $T^-$ ). Obtaining these metrics allow us to calculate the AUC as below:

$$AUC = \int_0^1 \frac{T^+}{T^+ + F^-} d \frac{F^+}{F^+ + T^-}, \quad 0.5 \leq AUC \leq 1 \quad (12)$$

The boundaries of AUC indicate performances ranging from random classifiers to perfect prediction models. Different performance metrics can be used to approximate the weights. Most of those metrics are based on a pre-set threshold for the positive and negative classes. Given that such threshold setting is subjective, aggregate measures, such as the AUC, are more appropriate. We follow the guidelines of Bogaert et al. (2018) and apply five times two-fold cross validation (5x2cv) as proposed by Alpaydin (1998). In that way, we obtain median 5x2cv AUC values used as our  $\check{\alpha}_i$ . For the case of variable selection through IF, Bogaert et al. (2018) suggest that these estimates are appropriate to extract the rank order of the input variables by means of AUC-based permutation variable importance measures. The reason for that is that is two-fold. Firstly, this technique allows using information extracted from all forecasting models, while models with higher AUC obtain higher weight in the

fusion operator. Secondly, the use of the median 5x2cv AUC is allowing to associate better the sensitivity of a model to the change of specific variables. In other words, when the variable is more related with the response, the higher its mean decrease in AUC is expected to be. Rearranging equation (10) leads to the IF measure for the input  $\tilde{n}$  with  $\check{t}$  models expressed as:

$$\check{V}_{\tilde{n}}(fusion) = \sum_{\check{i}=1}^{\check{t}} \check{\alpha}_{\check{i}} \check{V}_{\check{i}\tilde{n}} = \check{\alpha}_1 \check{V}_{1\tilde{n}} + \check{\alpha}_2 \check{V}_{2\tilde{n}} + \dots + \check{\alpha}_{\check{t}} \check{V}_{\check{t}\tilde{n}} \quad (13)$$

Where  $\check{V}_{\check{i}\tilde{n}}$  stands for the variable importance measure for input  $\tilde{n}$  in model  $\check{i}$ . In our case the models are six, namely RVM, OP, MLP, k-NN, DT, and SVM<sup>4</sup>.

As a last step towards the construction of our final input features, we keep the variables that account for 90% of the cumulative sensitivity as described by the pseudo-Pareto analysis of Oztekin et al. (2013) and Bogaert et al. (2018). These input features are shown in Online Appendix S.3 (Tables S.1-S.9).

### 3. Case Study and Empirical Evaluation

In this section, the empirical applications are presented. Their purpose is to demonstrate the merits of the introduced CF concept and examine the statistical accuracy of the derived forecasts. For this purpose, CF is applied in a series of football games and stocks' datasets. Sections 3.1 and 3.2 presents the characteristics of the datasets and Sections 3.3 and 3.4 have the empirical performance of CF and its benchmarks.

#### 3.1 Football Dataset

The first empirical application is on football games forecasting. All models are applied to forecast the results and the number of goals of football games in the English Premier League, Italian Seria A and Spanish La Liga from 2005 to 2016. The dataset of this study is publicly available at <http://www.football-data.co.uk>. In football game result forecasting, there are three different outcomes (Home win, draw and Away win). A second football forecasting exercise is based on the number of goals, at which gamblers bet whether the total number of goals in a game will or will not exceed 2.5<sup>5</sup>. The models under study require a series of inputs that are bookmakers' odds and past teams' performance indicators. The BetBrain average odds are used in this study, which are the average odds from around 138 bookmakers and betting exchanges from across the globe. These inputs are summarized in Table 1.

**Table 1: Inputs Series**

Points of H team – Points of A team before the start of the game	Points of H in the last 1,2,3 games and from the start of the season	Points of A in the last 1,2,3 games and from the start of the season	Points of H in the last 1,2,3 games and from the start of the season when H plays at home	Points of A in the last 1,2,3 games and from the start of the season when A plays at away
Number of goals of H team in the last 1,2,3 games and	Number of goals of A team in the last 1,2,3 games and from the start of the season	Number of goals of H team in the last 1,2,3 games and from the	Number of goals of A team in the last 1,2,3 games and from the start of the season	Number of shots on target of H team in the last 1,2,3 games

<sup>4</sup> The reason for discarding the ANFIS and CF forecasts from IF procedure is that developing a fuzzy model with our entire set of 83 inputs would increase the number of fuzzy rules created may increase up to 2<sup>83</sup> which is computationally infeasible to process.

<sup>5</sup> An alternative approach based on the Asian handicap is also explored in online Appendix S.5.

from the start of the season		start of the season when H team plays at home	when A team plays away	and from the start of the season
Number of shots on target of H team in the last 1,2,3 games and from the start of the season	Number of shots on target of A team in the last 1,2,3 games and from the start of the season	Number of shots on target of H team in the last 1,2,3 games and from the start of the season when H team plays at home	Number of shots on target of H team in the last 1,2,3 and from the start of the season games when A team plays at away	Number of corner kicks of H team in the last 1,2,3 games and from the start of the season
Number of corner kicks of A team in the last 1,2,3 and from the start of the season games	Number of corner kicks of H team in the last 1,2,3 games and from the start of the season when H team plays at home	Number of corner kicks of A team in the last 1,2,3 games and from the start of the season when A team plays at away	H team booking points in the last 1,2,3 games and from the start of the season	A team booking points in the last 1,2, 3 games and from the start of the season
Betbrain average home win odds	Betbrain average draw odds	Betbrain average away win odds	Betbrain average over 2.5 goals odds	Betbrain average under 2.5 goals odds
Betbrain size of handicap (home team)	Betbrain average Asian handicap home team odds	Betbrain average Asian handicap away team odds		

*Note: There are six main categories of predictors. The first one is based on the odds offered by the bookies and the other five originate from the performance of the teams over the past games. The inputs categories are: (i) odds for match outcome, number of goals and Asian handicap size bets (home team winning, draw, away team winning, over 2.5 goals, under 2.5 goals, home team win, and away team win Asian handicaps – 8 total), (ii) Points achieved, (iii) Goals scored (iv) Corner kicks, (v) Shots on target, (vi) Booking points. For the last five categories, 15 different types of measures are introduced to make sure all types of developments are considered. Team H is the home team and team A is the away team. Booking points are 25 points per red card and 10 per yellow card in a game. The total number of inputs is 83.*

The previous literature on football forecasting employs only sub-sets of Table 1. This pool consists of 83 potential inputs. Feeding all these inputs to any predictive model increase the training time and causes overfitting. Let us consider a simple case of having just high/low outcomes for 83 inputs. The number of fuzzy rules created may increase up to  $2^{83}$ , which is impractical for both understanding the input-output interactions and forecasting. To deal with the dimensionality problem we use a probabilistic framework, namely RVM to select the most informative subset of potential predictors (see, online Appendix S.2). The RVM allows us to identify the best inputs without restricting our inputs to smaller sets and serves as the variable selection component in the pre-processing stage of Figure 2. Thus, it is not necessary to restrict our inputs to smaller sets or apply additional techniques to reduce the dimensions of our predictors' pool<sup>6</sup>.

The implemented forecasting exercises span the period of 2005 to 2016. The forecasts from the models are evaluated in terms of accuracy through the relevant BetBrain average odds, the average profit per bet, the proportional cumulative annualized return and the Kelly Criterion. The average profit per bet is defined as:

<sup>6</sup> For example, Dixon and Robinson (1998), Oberstone (2009) and Angelini and De Angelis (2017) use the number of goals scored in a match to improve forecasting accuracy of the final football outcome. Other studies consider the odds of home win/ draw/ away for football game predictions (see amongst others, Dixon and Coles, 1997; Crowder et al., 2002; Dobson and Goddard, 2003; Constantinou et al., 2012; and Boshnakov et al., 2017). The number of corner kicks implies offensive pressure and is considered a good proxy for higher scoring probability. Thus, Andersson et al. (2009) apply this variable in their football betting models. Oberstone (2011) and Martins et al. (2017) apply shots on target as another proxy of the offensive capacity of a team. Most of the inputs in Table 2 are based on the performance of the home/away team during the last three games. Incorporating teams' recent performance indicators (points of last three games, goals scored in the last three games, etc.) is crucial for the adaptiveness of the CF process and their utility are supported by Goddard and Asimakopoulos (2004), Goddard (2005) and Rotshtein et al. (2005). Here it should be noted that in the football betting literature it is well-accepted practice to use bookmaker odds for modelling and forecasting football outcomes (Goddard and Asimakopoulos, 2004; Štrumbelj, 2014; and Schumaker et al., 2016). Additionally, using fixed-odds, like ours from BetBrain, is considered advantageous as bettors know the final odds at the time of betting (Feess et al., 2016). This is why fixed-odds betting applications are widely found in the respective literature (see amongst others Forrest and Simmons, 2008; and Constantinou et al., 2012).

$$\frac{\sum_q x_q b_q - |Q|}{|Q|}, q \in Q \quad (14)$$

where  $Q$  is the set of games on the season that a bet is placed,  $x_q$  takes the value of 1 if the bet on game  $q$  is won based on the relevant forecast and 0 otherwise,  $b_q$  is the relevant BetBrain average odd and  $|Q|$  is the cardinality of set  $Q$ . The proportional cumulative annualized return is estimated simply by betting on each game always the 5% of the total capital which initially is 100 units. For each subsequent game, we continue to bet the 5% of the total remaining pot. The proportional cumulative annualized return is the accumulated return in the end of the season. This practice resembles the reality where gamblers bet a proportion of their wealth.

In any investment (such as in football betting) there are three main areas to cover: the investment strategy, the timing of the investment (whether to invest or not), and the size of the invested capital. The investment strategy can be guided by statistical models (RVM, ANFIS, CF, and OP in our case). The timing of the investment can also be guided by a conditional procedure like CF or the investor's preference. The optimal size of the investment or football bet can be determined through the Kelly criterion. Consider the case of having initial capital  $\mathcal{X}_0$ . The goal is to maximise the expected value of capital after  $n$  trial ( $\mathcal{X}_n$ ). Now suppose that a gambler is interested in a bet with win (loss) probability  $p$  ( $q$ ) and payoff  $b$  for every unit wager. The purpose is to maximise:

$$g(f) = E[\log(\mathcal{X}_n/\mathcal{X}_0)] = p \log(1 + bf) + q \log(1 - f) \quad (15)$$

where  $f$  denotes the fraction invested in the bet and  $g(f)$  is the growth based on the fraction invested in each bet. The optimal fraction ( $f^{opt}$ ) based on Kelly (1956) and Thorp (2008) is given by:

$$f^{opt} = \frac{bp - q}{b} \quad (16)$$

In order to apply the Kelly criterion, we need the corresponding probabilities  $p$ . These are readily available in the OP framework, as the estimated conditional probabilities of each outcome. RVM and CF have a different structure and the winning probabilities are not readily provided. However, the CF framework allows us to resemble the Kelly criterion and following a similar procedure to extract the optimal fractions  $f^{opt}$ <sup>7</sup>. The firing strength can determine how probable it is to use the specific rule  $i$  for bet  $k$ . By aggregating this measure and normalising it for the rules with similar output we provide an approximated conditional probability for each ordinal outcome.

More specifically, we associate the winning probability with the conditional probability of each outcome to happen. For a bet  $k$  to be placed the normalised conditional firing strength  $\tilde{w}_i^k$  for all the eligible FRs is estimated by:

$$\tilde{w}_i^k = \frac{C_i^k \hat{w}_i^k}{\sum_{i=1}^5 C_i^k \hat{w}_i^k}, \forall i \quad (17)$$

In the next step the  $\tilde{w}_i^k$ s are aggregated for the rules that fall under same ordinal outcome:

---

<sup>7</sup> Singpurwalla and Booker (2004) demonstrate that probability theory has a sufficiently rich structure for incorporating fuzzy sets within its framework.



$$\hat{\pi}_a^k = \sum_s \tilde{w}_s^k | f_s^k(\cdot) \approx a \quad (18)$$

where the  $f_s^k(\cdot)$  represents the predicted output of the rule  $s$ . The procedure to interpret the regression output  $f_s^k(\cdot)$  to a class label  $a$  is the same in OP. By setting  $p = \hat{\pi}_{k,a}$  in equation (16), the optimal fraction for each outcome of the bet is given. It should be noted that when the CF does not find any eligible rule for the bet, the winning probability is zero and the  $f^{opt}$  becomes negative. A bet is only placed in case of having a positive  $f^{opt}$ .

The in-sample consists of three football seasons and the out-of-sample by the following two seasons (i.e. in the first forecasting exercise for the Premiership, the football seasons 2006-2007, 2007-2008 and 2008-2009 act as in-sample and the seasons 2009-2010 and 2010-2011 act as out-of-sample). The estimation is not rolled forward from the first (2009-2010) to the second season (2010-2011) of the out-of-sample. Thus, the second out-of-sample season act as robustness to the models. In all seasons the first three Home and the first three Away games of a team are discarded from the exercise. Otherwise, if team A plays against team B and one of the two teams has fewer than three Home and three Away games, this game is excluded from the exercise (both as in-sample and out-of-sample data)<sup>8</sup>.

### 3.2 Stocks' Dataset

The dataset used in the second application is simpler, as it involves the returns of Pfizer Inc. (PFE) and Goldman Sachs Group (GS) stocks. Both stocks prices are publicly available at <https://finance.yahoo.com>. We investigate the stocks over the period of 10 years (2007-2016). In each case, five years and one year of weekly data are used as in-sample and out-of-sample period respectively. The target series is the direction of stock price change at the end of each week. As inputs, we employ the weekly returns of the last two months (8 weeks). In this case, there is no dimensionality issue that might impede our approach and thus the inputs vector is fed directly to CF. Our forecasts in this exercise are examined through two simple measures. The directional accuracy for each study period  $t$  is denoted by  $ACC_t$  and defined as:

$$ACC_t = \frac{\sum_{q_1} I_{q_1}(\hat{S}_{q_1} = S_{q_1})}{|Q_t|}, q_1 \in Q_t \quad (19)$$

where  $\hat{S}_{q_1}$  and  $S_{q_1}$  are the predicted and actual direction of change in case a prediction is made for the period  $q_1$ . In addition,  $I_{q_1}$  is the indicator equal to one if the prediction is correct and zero. The set of forecasted periods in each study period  $t$  is shown by  $Q_t$  and  $|Q_t|$  is the number of times a prediction is made over the period  $t$ .

We also evaluate our models and their forecasts through a simple trading application. We go or stay 'long' (buy) the underlying stock when the model's forecasted direction is high. Similarly, we go or stay 'short' (sell) the

---

<sup>8</sup> This is happening for two reasons. It is well-known amongst football enthusiasts that the behaviour of teams at the start of the season is volatile. This happens either due to changes to the roster of the team during the summer or due to the different training during that period (for example, a team that has qualification games for the European Championships is forced to start its preparation earlier than the rest). Secondly, this process ensures that all series are smooth. For example, in the first game of the season the input series Points of H team – Points of A team would have been equal to 0 and the rest of the inputs would have to be drawn from the previous season.

underlying stock when the model's forecasted direction is low. Based on the derived trading signals we estimated the annualized return ( $RET_t$ ) for each model which is defined as:

$$RET_t = N_t \left( \frac{\sum_{q_1} \hat{s}_{q_1} r_{q_1}}{|Q_t|} \right) \quad (20)$$

where  $r_{q_1}$  is the affiliated logarithmic return over the period  $q_1$ . The  $N_t$  is the number of new observations available in each study period  $t$ .

### 3.3 Empirical Results

In Section 3.3.1 we present the results of our CF procedure and its benchmarks applied to the task of forecasting the game result and the number of goals. For the sake of space, the results of our models for the case of the Asian handicap are available in the online Appendix S.5. In Section 3.3.2, we present the performance of all our models applied in the task of forecasting the direction of the PFE stock. The performance of our models for the GS stock are in Appendix S.7.

#### 3.3.1 Football Empirical Results.

The performance of our models in terms of accuracy ratio for the game result and the number of goals is presented in Tables 2-3. Additional measures of classification are provided in Tables S.19-S.20 of Section S.8. The number of games that CF is applied is presented in the online Appendix S.6.1. The Pesaran-Timmermann (1992) test for the aggregate performance of the models is also estimated. The null hypothesis of the test is that the relevant model is unable to classify correctly the underlying series<sup>9</sup>.

Table 2: Accuracy Ratios (Game Result)

Model	Championship	Out-of-sample	In-sample						Average
			2006-2009	2007-2010	2008-2011	2009-2012	2010-2013	2011-2014	
RVM-CF	Premiership	1	68.75%	75.00%	66.67%	61.54%	72.73%	66.65%	68.56%**
		2	63.60%	50.03%	62.50%	57.14%	67.41%	50.00%	58.45%**
	La -Liga	1	60.00%	48.33%	57.69%	50.04%	55.81%	46.55%	53.07%**
		2	63.60%	43.50%	56.68%	43.47%	44.68%	41.60%	48.92%**
	Seria A	1	53.33%	54.50%	61.53%	60.00%	57.45%	61.50%	58.05%**
		2	53.06%	52.63%	54.56%	66.10%	66.70%	53.57%	57.77%**
RVM	Premiership	1	46.05%	37.84%	38.25%	44.33%	35.02%	52.02%	42.25%**
		2	41.22%	37.58%	42.61%	40.23%	34.66%	46.32%	41.59%**
	La -Liga	1	40.33%	36.00%	39.38%	39.33%	35.67%	41.33%	38.68%**
		2	33.67%	36.99%	43.33%	35.33%	41.00%	37.00%	37.89%*
	Seria A	1	51.35%	50.34%	45.92%	44.96%	40.60%	39.08%	45.38%**
		2	47.97%	44.22%	46.22%	34.90%	39.44%	40.27%	42.17%**
ANFIS	Premiership	1	45.70%	40.20%	35.91%	39.18%	40.67%	39.33%	40.17%**
		2	35.81%	39.93%	37.46%	37.00%	36.67%	41.33%	38.03%**
	La -Liga	1	40.00%	36.67%	41.10%	35.33%	34.33%	40.33%	37.96%*
		2	38.33%	34.59%	41.00%	36.67%	40.00%	36.03%	37.77%*
	Seria A	1	42.91%	35.47%	42.18%	42.44%	40.27%	40.14%	40.57%**
		2	34.80%	41.16%	43.70%	36.24%	38.38%	41.61%	39.31%**

<sup>9</sup> We did not estimate the Pesaran-Timmermann (1992) test for each separate year for two reasons. Firstly, we are interested in the average performance of our models and not their individual accuracy for a specific year and championship. Secondly and more importantly, the CF approach selects a subsample of games and generates forecasts only for them. In some cases, the number of CF forecasts is too small for the test and given the test's assumptions this can cripple the test's efficiency (Pesaran and Timmermann, 1992).

OP	Premiership	1	46.70%	44,93%	43,60%	38,49%	46.67%	47,00%	46.69%**
		2	45.61%	47,65%	41,90%	43,33%	39.66%	39.32%	41.53%**
	La -Liga	1	47.67%	52.67%	52.40%	48.67%	42.67%	51.67%	49.29%**
		2	42.60%	49.66%	49.00%	47.33%	54.33%	44.33%	47.88%**
Seria A	1	47.30%	44.59%	48.30%	45.80%	46.64%	37.68%	45.05%**	
	2	47.97%	44.30%	46.72%	45.30%	41.55%	42.28%	44.69%**	
MLP	Premiership	1	37.80%	35.14%	30.54%	36.77%	22.67%	36.67%	33.26%**
		2	35.14%	31.88%	34.36%	30.00%	39.00%	30.67%	33.51%**
	La -Liga	1	39.67%	37.67%	41.44%	41.67%	36.67%	39.67%	39.46%**
		2	36.00%	38.01%	44.33%	39.67%	32.33%	35.33%	37.61%**
	Seria A	1	36.15%	37.84%	36.39%	36.55%	41.95%	36.62%	37.58%**
		2	40.20%	42.52%	44.12%	35.57%	36.97%	34.23%	38.93%**
k-NN	Premiership	1	50.52%	42.23%	44.63%	41.58%	47.00%	49.33%	45.88%**
		2	40.88%	48.99%	49.83%	42.67%	39.67%	48.00%	45.01%**
	La -Liga	1	49.67%	52.33%	45.21%	50.67%	49.67%	51.67%	49.87%**
		2	48.00%	48.63%	50.67%	47.33%	50.33%	44.67%	48.27%**
	Seria A	1	42.57%	43.58%	46.60%	50.84%	41.95%	36.97%	43.75%**
		2	45.61%	41.84%	43.70%	48.32%	34.86%	33.56%	41.31%
CF-IF	Premiership	1	34.78%	50.00%	30.77%	25.29%	48.57%	39.62%	38.17%
		2	32.56%	60.00%	34.00%	34.52%	36.07%	32.65%	38.30%
	La -Liga	1	30.77%	36.29%	40.91%	32.31%	52.17%	26.79%	36.54%
		2	52.86%	34.09%	40.00%	43.90%	47.06%	47.47%	44.23%
	Seria A	1	41.07%	35.96%	30.88%	52.00%	36.11%	40.00%	39.34%
		2	35.90%	18.52%	41.18%	47.71%	30.77%	42.86%	36.15%
DT	Premiership	1	38.83%	43.58%	39.93%	45.70%	38.00%	44.00%	41.68%**
		2	41.89%	40.27%	38.83%	39.67%	37.33%	41.33%	39.89%**
	La -Liga	1	43.33%	44.33%	39.38%	46.33%	47.67%	44.67%	44.29%**
		2	42.00%	46.58%	43.33%	45.00%	48.67%	43.00%	44.76%**
	Seria A	1	44.59%	45.61%	44.22%	44.54%	42.62%	40.85%	43.74%**
		2	42.91%	41.16%	47.06%	41.61%	39.08%	44.97%	42.8%**
SVM	Premiership	1	45.02%	38.85%	39.60%	35.40%	47.00%	39.00%	40.81%**
		2	42.57%	44.97%	36.43%	34.00%	43.00%	34.33%	39.22%**
	La -Liga	1	44.00%	33.00%	42.81%	40.67%	41.67%	37.67%	39.97%**
		2	44.67%	36.99%	45.33%	46.33%	46.00%	38.33%	42.94%**
	Seria A	1	33.78%	35.47%	46.94%	38.66%	39.60%	36.62%	38.51%**
		2	31.08%	39.12%	44.12%	40.27%	38.03%	42.28%	39.15%**

Note: The values in the table represent the out-of-sample accuracy ratios. Row 1 corresponds to the first year of the out-of-sample and row 2 to the second year of the out-of-sample. For example, for the first cell on the left corner 68.75% is the out-of-sample accuracy of CF for the 2009-2010 Premiership season and the 63.60% is the performance of the exact same model (same specification and rules) for the 2010-2011 season of the same championship. A random classifier provides 33.33% accuracy ratio in this example. \*\* and \* indicates that according to the Pesaran-Timmermann (1992) test, the forecasts are statistically accurate in classifying the football game result at the 99% and 95% level respectively. Online Appendix S.8 presents extra measures for the fitness of different models.

Table 3: Accuracy Ratios (Number of Goals)

Model	Championship	Out-of-sample	In-sample						Average
			2006-2009	2007-2010	2008-2011	2009-2012	2010-2013	2011-2014	
RVM-CF	Premiership	1	78.57%	73.68%	72.70%	62.50%	75.02%	69.23%	71.95%**
		2	75.71%	66.67%	71.43%	65.21%	70.59%	56.75%	67.73%**
	La -Liga	1	66.66%	65.93%	72.41%	72.02%	71.19%	62.50%	68.45%**
		2	60.87%	62.50%	62.61%	70.03%	72.70%	58.51%	64.54%**
RVM	Seria A	1	85.69%	83.33%	70.00%	64.29%	71.43%	72.97%	74.62%**
		2	60.00%	52.17%	57.41%	58.62%	75.00%	59.46%	60.44%**
	Premiership	1	50.17%	50.00%	55.37%	52.23%	51.67%	57.00%	52.74%**
		2	48.99%	55.03%	51.20%	51.67%	51.00%	53.00%	51.82%*
ANFIS	La -Liga	1	57.00%	53.33%	53.42%	54.67%	51.00%	60.33%	54.96%**
		2	50.67%	52.05%	52.34%	53.30%	48.33%	54.67%	51.89%
	Seria A	1	53.04%	58.45%	53.74%	48.32%	48.32%	56.34%	53.04%*
		2	55.74%	51.36%	54.62%	47.65%	54.58%	50.34%	52.38%
ANFIS	Premiership	1	53.26%	50.14%	51.68%	47.08%	49.00%	53.00%	50.69%
		2	46.96%	50.34%	48.80%	55.33%	43.67%	53.67%	49.80%
	La -Liga	1	52.33%	52.00%	54.45%	50.33%	58.00%	49.33%	52.74%*
		2	52.00%	47.33%	53.67%	45.33%	52.05%	47.00%	49.56%
	Seria A	1	50.34%	53.04%	51.70%	52.94%	54.03%	61.62%	53.95%**
		2							

		2	47.30%	53.74%	44.96%	51.68%	56.34%	49.33%	50.56%
<b>OP</b>	Premiership	1	49.83%	49.66%	46.98%	49.48%	49.33%	55.33%	50.10%
		2	48.99%	52.68%	46.39%	51.33%	57.33%	51.00%	51.29%
	La -Liga	1	53.33%	54.67%	54.79%	52.67%	56.00%	56.10%	54.59%**
		2	48.00%	53.42%	53.35%	56.67%	57.00%	55.67%	54.02%**
	Seria A	1	46.62%	56.76%	50.68%	55.46%	47.32%	51.06%	51.32%
		2	45.13%	49.12%	47.28%	45.30%	50.70%	50.67%	48.03%
<b>MLP</b>	Premiership	1	49.48%	46.96%	51.01%	52.58%	51.00%	53.67%	50.78%
		2	48.99%	51.34%	45.36%	55.00%	49.67%	53.67%	50.67%
	La -Liga	1	47.67%	49.67%	51.03%	55.67%	57.00%	53.00%	52.34%**
		2	50.67%	55.82%	47.67%	56.00%	51.33%	57.67%	53.19%
	Seria A	1	46.96%	58.11%	54.08%	52.52%	54.70%	52.11%	53.08%*
		2	52.36%	50.00%	55.46%	47.32%	53.87%	52.01%	51.84%
<b>k-NN</b>	Premiership	1	49.83%	43.92%	48.99%	52.58%	50.67%	53.00%	49.83%
		2	43.92%	45.97%	51.20%	51.00%	53.33%	55.67%	50.18%
	La -Liga	1	51.67%	52.33%	49.66%	48.67%	54.67%	51.67%	51.44%
		2	52.33%	53.42%	51.33%	54.33%	52.67%	49.33%	52.24%*
	Seria A	1	47.64%	51.69%	51.02%	49.16%	56.38%	55.28%	51.86%
		2	57.09%	50.34%	46.22%	53.02%	51.76%	52.01%	51.74%
<b>CF-IF</b>	Premiership	1	71.43%	48.98%	42.05%	59.52%	55.17%	51.92%	54.85%
		2	40.00%	51.72%	48.78%	49.38%	57.14%	38.46%	47.58%
	La -Liga	1	31.25%	50.00%	47.73%	55.17%	55.00%	42.62%	46.96%
		2	48.15%	42.08%	40.00%	47.62%	38.71%	51.43%	44.66%
	Seria A	1	56.41%	52.94%	54.55%	49.40%	57.89%	50.63%	53.64%
		2	44.44%	41.38%	49.38%	53.33%	52.38%	54.17%	49.18%
<b>DT</b>	Premiership	1	54.98%	51.01%	55.70%	43.30%	55.33%	47.67%	51.33%
		2	50.68%	53.36%	52.92%	55.00%	45.33%	54.33%	51.94%
	La -Liga	1	55.00%	51.00%	53.42%	52.33%	58.00%	52.00%	53.63%**
		2	48.67%	48.63%	53.67%	51.33%	54.33%	51.33%	51.33%
	Seria A	1	53.04%	55.74%	45.92%	52.10%	50.67%	56.34%	52.30%
		2	48.65%	52.72%	47.48%	51.01%	52.11%	53.02%	50.83%
<b>SVM</b>	Premiership	1	48.80%	46.62%	57.72%	49.14%	50.00%	50.00%	50.38%
		2	52.70%	55.37%	48.80%	54.33%	48.00%	44.67%	50.64%
	La -Liga	1	51.33%	50.33%	55.48%	52.33%	57.33%	49.33%	52.69%*
		2	50.33%	55.14%	54.00%	52.33%	53.00%	53.67%	53.08%*
	Seria A	1	49.66%	57.77%	46.60%	55.46%	52.01%	48.59%	51.68%
		2	43.92%	50.68%	51.26%	52.01%	54.23%	53.36%	50.91%

Note: The values in the table represent the out-of-sample accuracy ratios. Row 1 corresponds to the first year of the out-of-sample and row 2 to the second year of the out-of-sample. For example, for the first cell on the left corner, 78.57% is the out-of-sample accuracy of CF for the 2009-2010 Premiership season and the 75.71% is the performance of the exact same model (same specification and rules) for the 2010-2011 season of the same championship. A random classifier provides 50.00% accuracy ratio in this example. \*\* and \* indicates that according to the Pesaran-Timmermann (1992) test, the forecasts are statistically accurate in classifying the football game result at the 99% and 95% level respectively. Online Appendix S.8 presents additional measures of classification for all models.

From the tables above, we note that the RVM-CF approach is clearly improving the accuracy of the underlying system (RVM). It generates forecasts based on the strongest rules which in turn leads to an improved predictability. In most of exercises, RVM-CF is up to 10% more accurate than RVM. On the other hand, ANFIS leads to slightly poorer accuracy ratios than those obtained by RVM. A finding consistent with the related literature (see Martens et al., 2007). The MLP presents the worst performance in game result and close accuracies compared to ANFIS and RVM for over-under. The DT and SVM benchmarks also do not manage to outperform CF in both applications. The SVM as expected is providing less accurate results than RVM. Concerning the CF-IF approach, we note that its performance is significantly lower than the one generated by CF with the RVs as inputs. It seems that the Bayesian non-linear properties of RVM are superior in our exercise from the variable selection based on information-fusion sensitivity scores.

The PT test reveals that only the RVM-CF is capable of classifying accurately the underlying series even two seasons ahead in all exercises. Its benchmarks seem to work well in game result but lose power in over-under

exercise. We also note that the accuracies are higher for the Premiership. This might imply that the English championship has less noise or, in other words, is easier to forecast. The previously discussed accuracy ratios might seem promising, but they do not guarantee profitability. The risk in football gambling is almost always that the bookmakers' odds differ between the season and championships. Betting agencies naturally possess superior information and modify the odds by exploiting the bettor's cognitive biases in a way that mitigate their risk and increase their profitability (Forrest and Simmons, 2008). Tables 4 and 5 present the average profit per bet for the models, championships and seasons under study.

**Table 4: Average Profit per bet (Game Result)**

Model	Championship	Out-of-sample	In-sample						Average
			2006-2009	2007-2010	2008-2011	2009-2012	2010-2013	2011-2014	
RVM-CF	Premiership	1	17.71%	19.33%	25.07%	26.81%	29.92%	14.50%	22.21%
		2	3.59%	10.00%	9.69%	0.52%	-0.50%	2.16%	4.24%
	La -Liga	1	10.89%	12.16%	10.45%	16.61%	7.92%	3.70%	10.29%
		2	1.13%	3.10%	2.15%	0.78%	2.80%	-1.09%	1.48%
	Seria A	1	5.17%	12.10%	23.19%	12.04%	7.24%	16.24%	12.66%
		2	-2.61%	-10.00%	1.42%	3.24%	3.54%	1.26%	-0.53%
RVM	Premiership	1	7.10%	-8.76%	-9.81%	1.09%	-1.80%	5.23%	-1.16%
		2	-6.00%	-13.74%	9.57%	-7.51%	-8.34%	-0.58%	-4.43%
	La -Liga	1	-5.45%	-0.45%	-0.32%	-1.05%	1.92%	2.03%	-0.55%
		2	-13.23%	-11.30%	-0.49%	-14.80%	-6.23%	-6.09%	-8.69%
	Seria A	1	4.58%	4.19%	2.79%	3.16%	-7.39%	-1.22%	1.02%
		2	-5.05%	-3.34%	-11.85%	-14.02%	-3.26%	-8.61%	-7.69%
ANFIS	Premiership	1	-2.57%	-3.07%	-2.91%	8.26%	-15.17%	-12.23%	-4.62%
		2	-7.09%	-12.56%	-21.83%	-5.06%	-5.43%	-7.09%	-9.84%
	La -Liga	1	-5.08%	-9.58%	-6.00%	-17.29%	-11.82%	0.40%	-8.23%
		2	-7.72%	-9.65%	-10.67%	-21.71%	-5.19%	-5.26%	-10.03%
	Seria A	1	-1.70%	-1.15%	2.63%	-1.34%	-3.22%	0.88%	-0.65%
		2	-18.26%	-7.33%	0.42%	-16.40%	-7.38%	-1.74%	-8.45%
OP	Premiership	1	-8.19%	1.24%	-8.07%	-16.31%	7.26%	-6.78%	-5.14%
		2	-9.02%	-15.86%	-17.71%	-8.05%	-7.31%	-7.29%	-10.88%
	La -Liga	1	-6.75%	-0.29%	7.50%	-2.65%	-25.15%	-0.62%	-4.66%
		2	-11.30%	-6.80%	-5.22%	-7.18%	1.30%	-13.29%	-7.08%
	Seria A	1	-3.76%	1.33%	-5.87%	-1.18%	-0.15%	-14.86%	-4.08%
		2	-4.42%	-1.02%	17.71%	-3.88%	-7.67%	-19.38%	-3.11%
MLP	Premiership	1	-4.92%	-9.78%	-15.41%	4.01%	-31.52%	-5.79%	-10.57%
		2	-7.08%	-6.86%	-0.98%	-23.33%	-6.77%	2.56%	-7.08%
	La -Liga	1	-0.75%	-9.74%	-5.70%	-1.98%	-17.51%	-9.64%	-7.55%
		2	-12.84%	-13.98%	-0.95%	-12.63%	-9.33%	-4.81%	-9.09%
	Seria A	1	-14.94%	-14.53%	-10.39%	-8.63%	-5.51%	0.92%	-8.85%
		2	-13.65%	-1.66%	4.04%	-11.96%	1.39%	-17.78%	-6.60%
k-NN	Premiership	1	-0.11%	-7.48%	-6.99%	-9.12%	-4.44%	-4.42%	-5.43%
		2	-10.20%	-0.91%	2.05%	0.95%	-13.30%	5.35%	-2.68%
	La -Liga	1	2.25%	2.00%	-16.98%	0.33%	-4.61%	-5.66%	-3.78%
		2	-1.43%	-4.69%	1.16%	-10.77%	-6.38%	-15.10%	-6.20%
	Seria A	1	-11.58%	-2.71%	2.59%	2.71%	1.56%	1.81%	-0.94%
		2	0.62%	-3.58%	-15.21%	-1.00%	-14.60%	-21.34%	-9.18%
CF-IF	Premiership	1	-41.03%	1.46%	-6.74%	-41.38%	11.83%	-14.89%	-15.12%
		2	12.36%	-20.31%	-25.24%	-21.17%	3.43%	-15.99%	-11.15%
	La -Liga	1	-34.08%	3.44%	3.59%	-37.25%	35.39%	-10.11%	-6.50%
		2	25.49%	4.18%	-27.20%	48.49%	39.18%	13.15%	17.21%
	Seria A	1	-4.84%	-4.83%	-20.97%	11.52%	-8.22%	-9.70%	-6.17%
		2	-16.92%	-44.15%	2.45%	-0.23%	2.46%	-0.86%	-9.54%
DT	Premiership	1	-11.98%	-2.05%	-6.80%	7.08%	-8.10%	-1.19%	-3.84%
		2	-3.28%	-15.44%	-7.99%	-8.40%	-9.20%	-0.49%	-7.47%
	La -Liga	1	-4.75%	-10.24%	-16.73%	-3.66%	-3.27%	-14.36%	-8.83%
		2	-8.64%	0.79%	-6.94%	1.38%	-4.33%	-13.09%	-5.14%
	Seria A	1	-2.99%	-0.92%	-1.19%	-1.61%	-5.36%	-3.57%	-2.61%

		2	-6.70%	0.51%	4.45%	-0.85%	-2.38%	0.64%	-0.72%
SVM	Premiership	1	-8.54%	-14.43%	-2.01%	-14.20%	5.61%	-3.40%	-6.16%
		2	-6.93%	0.95%	-6.68%	-15.45%	-3.01%	-17.57%	-8.12%
	La -Liga	1	1.23%	-21.57%	-3.92%	-10.21%	-9.98%	-0.03%	-7.41%
		2	-0.97%	-7.25%	0.25%	-2.66%	-2.74%	-10.83%	-4.04%
	Seria A	1	-18.66%	-16.82%	-1.24%	-16.46%	-1.44%	-11.87%	-11.08%
		2	-16.89%	-1.26%	-6.48%	-6.32%	8.58%	-10.47%	-5.47%

Note: All values in the Table represent the average profit per season. Row 1 corresponds to the first year of the out-of-sample and row 2 to the second year of the out-of-sample. For example, for the first cell on the left corner, 17.71% is the average profit per bet of CF for the 2009-2010 Premiership season and the 3.59% is the performance of the exact same model (same specification and rules) for the 2010-2011 season of the same championship.

Table 5: Average Profit per bet (Number of Goals)

Model	Championship	Out-of-sample	In-sample						Average
			2006-2009	2007-2010	2008-2011	2009-2012	2010-2013	2011-2014	
RVM-CF	Premiership	1	17.44%	16.25%	9.14%	11.64%	28.78%	17.62%	16.81%
		2	2.80%	-1.08%	-0.46%	2.83%	7.07%	4.77%	2.66%
	La -Liga	1	13.41%	15.44%	19.49%	14.14%	14.65%	9.87%	14.50%
		2	0.42%	2.15%	5.14%	0.28%	3.33%	0.83%	2.03%
	Seria A	1	17.22%	10.28%	6.45%	20.93%	14.32%	19.03%	14.71%
		2	1.20%	-0.96%	-0.25%	6.97%	2.44%	3.27%	2.11%
RVM	Premiership	1	1.77%	0.14%	-0.42%	-3.76%	-4.80%	-8.84%	-2.65%
		2	0.95%	-1.20%	-4.88%	-4.03%	-12.51%	-15.76%	-6.24%
	La -Liga	1	1.70%	-1.96%	-0.95%	-3.00%	4.11%	-3.06%	-0.53%
		2	-6.16%	-5.54%	-2.28%	-5.45%	-5.78%	-6.49%	-5.28%
	Seria A	1	-1.47%	-0.51%	-3.46%	-6.13%	-0.02%	-1.35%	-2.16%
		2	-1.53%	-2.79%	-6.56%	-9.97%	-1.58%	-6.40%	-4.81%
ANFIS	Premiership	1	-6.57%	-5.43%	0.12%	-3.22%	-5.20%	-3.38%	-3.95%
		2	-9.84%	-6.64%	-8.55%	-6.35%	-5.95%	-4.18%	-6.92%
	La -Liga	1	-4.20%	-2.12%	0.78%	-2.92%	0.12%	-3.66%	-2.00%
		2	-7.13%	-3.56%	-4.02%	-5.26%	-3.92%	-7.46%	-5.23%
	Seria A	1	-2.84%	-2.48%	-4.16%	-2.42%	-2.60%	-0.32%	-2.47%
		2	-4.61%	-5.07%	-10.49%	-6.93%	-3.30%	-5.02%	-5.90%
OP	Premiership	1	-7.95%	-11.28%	-0.16%	-3.86%	-1.96%	3.15%	-3.68%
		2	-7.80%	0.43%	-15.66%	-6.35%	5.78%	-9.50%	-5.52%
	La -Liga	1	-0.80%	-8.90%	-0.18%	-10.24%	-1.81%	-5.10%	-4.51%
		2	-8.75%	-5.94%	-10.17%	5.07%	-12.75%	-4.20%	-6.12%
	Seria A	1	-13.14%	-0.66%	-4.76%	0.03%	-13.08%	-10.74%	-7.06%
		2	4.59%	-9.44%	2.22%	-14.16%	-1.59%	-4.86%	-3.87%
MLP	Premiership	1	-8.03%	-13.85%	-6.47%	-1.95%	-7.31%	-1.10%	-6.45%
		2	-7.31%	-6.33%	-18.37%	1.28%	-8.31%	-0.63%	-6.61%
	La -Liga	1	-13.34%	-10.16%	-8.95%	-1.53%	0.61%	-8.59%	-6.99%
		2	-7.59%	-2.79%	-15.21%	-1.63%	-9.69%	-0.58%	-6.25%
	Seria A	1	-16.32%	1.64%	-1.41%	-1.70%	1.45%	-5.56%	-3.65%
		2	-0.90%	-9.56%	0.30%	-7.76%	-2.83%	-6.76%	-4.58%
k-NN	Premiership	1	-9.24%	-15.35%	-7.52%	-0.43%	-6.01%	-3.26%	-6.97%
		2	-16.73%	-10.49%	-0.27%	-6.39%	0.15%	3.12%	-5.10%
	La -Liga	1	-6.50%	-1.67%	-9.52%	-12.46%	-0.49%	-5.94%	-6.10%
		2	-5.42%	-3.36%	-4.50%	-2.56%	-2.42%	-10.63%	-4.81%
	Seria A	1	-12.54%	-4.79%	-6.19%	-9.18%	4.44%	1.07%	-4.53%
		2	2.81%	-4.51%	-15.24%	-0.44%	-3.87%	-6.21%	-4.58%
CF-IF	Premiership	1	30.83%	-7.35%	-22.47%	17.14%	-1.83%	-2.37%	2.33%
		2	-20.00%	-3.38%	-8.49%	-5.90%	6.10%	-27.33%	-9.83%
	La -Liga	1	-39.38%	-2.03%	-13.57%	-0.83%	6.90%	-17.36%	-11.04%
		2	-6.72%	1.44%	-17.70%	-12.90%	-28.19%	-4.30%	-11.40%
	Seria A	1	3.74%	0.53%	4.09%	-5.36%	10.05%	-2.18%	1.81%
		2	-15.19%	-19.07%	-4.15%	0.02%	-4.10%	4.04%	-6.41%
DT	Premiership	1	1.44%	-6.23%	3.79%	-16.88%	4.27%	-10.35%	-3.99%
		2	-5.11%	-2.84%	-0.37%	0.10%	-14.99%	1.21%	-3.67%
	La -Liga	1	-0.29%	-5.93%	-3.03%	-6.92%	3.76%	-4.93%	-2.89%
		2	-9.90%	-12.08%	-2.99%	-7.23%	-1.53%	-8.45%	-7.03%
	Seria A	1	-2.34%	3.90%	-14.51%	-3.55%	-4.69%	3.86%	-2.89%
		2	-11.83%	-2.17%	-12.08%	-5.06%	-3.14%	-3.75%	-6.34%

<b>SVM</b>	Premiership	1	-8.68%	-14.95%	5.77%	-2.53%	-7.06%	-6.67%	-5.69%
		2	-0.92%	2.35%	-10.88%	0.27%	-9.99%	-16.56%	-5.95%
	La -Liga	1	-7.02%	-7.53%	3.30%	-2.77%	-2.17%	-10.61%	-4.47%
		2	-8.66%	-0.08%	-0.33%	-7.08%	-7.55%	-2.75%	-4.41%
	Seria A	1	-3.80%	1.91%	-7.68%	2.58%	-3.18%	-11.29%	-3.58%
		2	-13.38%	-9.26%	-1.85%	-3.82%	-2.34%	-3.31%	-5.66%

Note: All values in the Table represent the average profit per season. Row 1 corresponds to the first year of the out-of-sample and row 2 to the second year of the out-of-sample. For example, for the first cell on the left corner, 17.44% is the average profit per bet of CF for the 2009-2010 Premiership season and the 2.80% is the performance of the exact same model (same specification and rules) for the 2010-2011 season of the same championship.

The RVM-CF approach seems to clearly outperform its benchmarks and offers impressive profits. The average profit per bet of RVM-CF is clearly above 10% in the first year of the out-of-sample in all cases. This profit is substantially reduced in the second year of the out-of-sample but still remains positive. On the other hand, all other models under study seem to present negative profits per bet in most of the cases even at the first year of the out-of-sample. The CF-IF profit results remain far below the RVM-CF. Thus, the superiority of RVM variable selection is validated in terms of profitability too. If we compare Tables 4 and 5 with the associated accuracy results (Tables 2 and 3), we note that statistical accuracy is not synonymous with betting profitability. There are cases where the PT test indicates that the underlying model classifies accurately the underlying series and the related average profit per bet is negative.

The average profit per bet considers that the forecaster will invest the same amount of capital in all games irrespective of their previous performance. In reality, forecasters will probably modify their invested capital based on the previous record and follow a more adaptive strategy. Thus, in Tables 6 and 7 we present the proportional cumulative annualized return. This measure assumes that the forecaster always bets 5% of their total pot. So, for example, in the first game in any give exercise the gambler will bet the 5% of their total pot – if it is 100 units, this will be 5 units. If the forecaster wins the bet and their earnings are 4 units, their total pot would then be 104 units. Thus, in the next game the forecaster will bet 5.2 units. The proportional cumulative annualized return offers a more realistic approach to betting where participants modify the size of their bets based on their previous record.

**Table 6: Proportional Cumulative Annualized Return (Game Result)**

Model	Championship	Out-of-sample	In-sample						Average
			2006-2009	2007-2010	2008-2011	2009-2012	2010-2013	2011-2014	
<b>RVM-CF</b>	Premiership	1	40.26%	7.42%	10.89%	11.22%	19.58%	6.91%	16.05%
		2	38.08%	1.35%	1.52%	-4.63%	11.11%	-5.70%	6.96%
	La -Liga	1	-0.14%	14.58%	22.39%	18.14%	29.46%	10.04%	15.75%
		2	-2.52%	3.25%	11.81%	-0.48%	-9.60%	-5.85%	-0.57%
	Seria A	1	15.34%	10.95%	28.64%	32.30%	32.01%	10.17%	21.57%
		2	-1.63%	-32.41%	2.01%	-5.02%	-17.08%	-6.29%	-10.07%
<b>RVM</b>	Premiership	1	-15.14%	-15.74%	-20.17%	-29.53%	-12.47%	-17.24%	-18.38%
		2	-27.12%	-45.78%	-54.28%	-35.52%	-60.05%	-58.30%	-46.84%
	La -Liga	1	-33.45%	-20.64%	-19.25%	-28.65%	-17.53%	-21.02%	-23.42%
		2	-55.26%	-57.36%	-38.27%	-47.20%	-42.08%	-47.36%	-47.92%
	Seria A	1	0.15%	-14.21%	-7.50%	-12.37%	-32.55%	-47.32%	-18.97%
		2	-25.13%	-60.93%	-16.06%	-61.27%	-59.77%	-62.83%	-47.67%
<b>ANFIS</b>	Premiership	1	-5.26%	-10.25%	-24.07%	-30.17%	-9.52%	-16.29%	-15.93%
		2	-29.16%	-15.57%	-36.49%	-33.60%	-23.41%	-23.47%	-26.95%
	La -Liga	1	-32.55%	-33.24%	-32.24%	-16.47%	-10.26%	-45.07%	-28.31%
		2	-68.26%	-38.56%	-69.41%	-23.54%	-15.05%	-52.14%	-44.49%

	Seria A	1	-55.85%	-26.37%	-6.84%	-47.29%	-14.54%	-13.25%	-27.36%
		2	-60.24%	-48.70%	-21.58%	-55.28%	-67.78%	-33.88%	-47.91%
OP	Premiership	1	-52.66%	-23.85%	-36.44%	-25.68%	-15.22%	-40.28%	-32.36%
		2	-64.66%	-32.10%	-52.13%	-43.69%	-19.65%	-56.09%	-44.72%
	La -Liga	1	-15.14%	-36.60%	-35.06%	-59.07%	-22.18%	-28.46%	-32.75%
		2	-18.44%	-68.12%	-62.38%	60.57%	-56.87%	-55.41%	-33.44%
Seria A	1	-62.48%	-18.93%	-50.38%	-37.57%	-37.55%	-40.16%	-41.18%	
	2	-66.64%	-46.18%	-55.37%	-66.88%	-59.85%	-58.74%	-58.94%	
MLP	Premiership	1	-74.49%	-87.65%	-94.98%	-17.70%	-99.08%	-79.39%	-75.55%
		2	-82.26%	-83.75%	-59.58%	-98.43%	-80.21%	-42.43%	-74.44%
	La -Liga	1	-56.07%	-87.30%	-76.05%	-62.53%	-96.09%	-88.16%	-77.70%
		2	-91.90%	-92.74%	-52.73%	-92.22%	-88.23%	-77.60%	-82.57%
	Seria A	1	-93.52%	-93.37%	-88.56%	-78.34%	-75.31%	-44.88%	-79.00%
		2	-91.96%	-55.98%	-2.59%	-91.16%	-39.58%	-95.93%	-62.87%
k-NN	Premiership	1	-42.65%	-81.98%	-79.02%	-84.99%	-69.15%	-66.74%	-70.75%
		2	-87.78%	-47.25%	-16.42%	-45.13%	-91.86%	24.71%	-43.96%
	La -Liga	1	-17.47%	-12.96%	-94.67%	-33.41%	-68.68%	-73.52%	-50.12%
		2	-52.11%	-67.81%	-28.81%	-86.72%	-75.91%	-93.66%	-67.50%
	Seria A	1	-88.87%	-62.71%	-19.19%	-6.31%	-41.15%	-45.91%	-44.02%
		2	-38.15%	-69.63%	-88.30%	-48.46%	-93.58%	-97.40%	-72.58%
CF-IF	Premiership	1	-55.59%	-2.33%	-44.24%	-85.83%	14.63%	-38.35%	-35.29%
		2	4.89%	-36.39%	-51.09%	-65.86%	-0.35%	-61.33%	-35.02%
	La -Liga	1	-70.06%	-29.06%	-2.62%	-72.88%	38.37%	-47.00%	-30.54%
		2	-23.77%	-25.99%	-33.64%	-32.11%	-23.60%	-15.58%	-27.82%
	Seria A	1	-22.80%	-44.32%	-58.41%	10.81%	-20.44%	-6.18%	-23.56%
		2	-33.00%	-47.84%	-19.20%	-19.70%	-6.23%	-7.53%	-22.25%
DT	Premiership	1	-90.28%	-58.37%	-81.30%	47.17%	-85.51%	-54.17%	-53.74%
		2	-66.72%	-94.07%	-84.28%	-85.72%	-87.47%	-53.89%	-78.69%
	La -Liga	1	-71.98%	-86.90%	-95.17%	-66.26%	-63.23%	-93.25%	-79.46%
		2	-84.40%	-40.31%	-80.50%	-44.35%	-68.26%	-92.04%	-68.31%
	Seria A	1	-62.28%	-50.25%	-52.55%	-46.48%	-74.21%	-66.36%	-58.69%
		2	-77.97%	-48.24%	6.32%	-54.14%	-62.55%	-40.03%	-46.10%
SVM	Premiership	1	-82.41%	-92.97%	-64.35%	-93.62%	18.17%	-71.67%	-64.47%
		2	-78.53%	-37.90%	-81.67%	-95.39%	-64.81%	-96.12%	-75.74%
	La -Liga	1	-34.78%	-97.61%	-71.02%	-88.41%	-88.54%	-74.74%	-75.85%
		2	-50.70%	-84.09%	-45.31%	-62.05%	-62.66%	-91.07%	-65.98%
	Seria A	1	-95.91%	-95.45%	-49.32%	-90.75%	-58.43%	-89.89%	-79.96%
		2	-96.02%	-60.51%	-69.43%	-78.64%	52.94%	-87.66%	-56.55%

Note: All values in the Table represent the proportional cumulative annualized return per season. Row 1 corresponds to the first year of the out-of-sample and row 2 to the second year of the out-of-sample. For example, for the first cell on the left corner, 40.26% is the average profit per bet of CF for the 2009-2010 Premiership season and the 38.08% is the performance of the exact same model (same specification and rules) for the 2010-2011 season of the same championship.

Table 7: Proportional Cumulative Annualized Return (Number of Goals)

Model	Championship	Out-of-sample	In-sample						Average
			2006-2009	2007-2010	2008-2011	2009-2012	2010-2013	2011-2014	
RVM-CF	Premiership	1	7.29%	24.58%	19.03%	6.68%	3.81%	33.46%	15.81%
		2	0.85%	8.04%	-1.69%	-3.42%	-2.03%	-6.21%	-0.74%
	La -Liga	1	10.24%	12.48%	5.87%	33.61%	9.56%	12.57%	14.06%
		2	-9.22%	4.65%	-8.56%	4.18%	2.55%	0.87%	-0.92%
	Seria A	1	9.17%	10.66%	11.54%	14.03%	25.19%	11.86%	13.74%
		2	3.63%	0.57%	3.12%	7.35%	-1.25%	2.63%	2.68%
RVM	Premiership	1	-21.14%	-15.58%	-31.61%	-58.47%	-14.87%	-18.40%	-26.68%
		2	-56.24%	-41.28%	-59.35%	-65.55%	-40.16%	-51.81%	-52.40%
	La -Liga	1	-26.24%	-16.38%	-46.72%	-57.43%	-32.55%	-30.04%	-34.89%
		2	-41.02%	-38.45%	-63.23%	-58.06%	-59.67%	-60.12%	-53.43%
	Seria A	1	-30.25%	-29.03%	-32.19%	-41.13%	-10.36%	-37.35%	-30.05%
		2	-65.02%	-50.14%	-48.21%	-56.01%	-40.62%	-46.18%	-51.03%
ANFIS	Premiership	1	-42.14%	-39.43%	-41.02%	-23.48%	-66.86%	-23.17%	-39.35%



	La -Liga	2	-49.56%	-51.17%	-49.52%	-49.08%	-70.35%	-49.61%	-53.22%	
		1	-61.63%	-60.78%	-19.22%	-40.79%	-23.95%	-42.25%	-41.44%	
	Seria A	2	-63.28%	-61.36%	-53.24%	-67.59%	-59.31%	-52.22%	-59.50%	
		1	-52.78%	-55.90%	-56.32%	-50.63%	-57.05%	-54.00%	-54.45%	
	OP	Premiership	2	-59.55%	-58.37%	-60.24%	-62.47%	-59.88%	-60.12%	-60.11%
			1	-39.45%	-49.56%	-55.32%	-58.48%	-47.22%	-52.17%	-50.37%
La -Liga		2	-60.17%	-58.16%	-59.42%	-62.21%	-55.66%	-63.14%	-59.79%	
		1	-48.87%	-30.45%	-43.60%	-49.62%	-43.46%	-63.46%	-46.58%	
Seria A		2	-65.24%	-48.80%	-66.75%	-63.96%	-60.17%	-62.85%	-61.30%	
		1	-41.13%	-32.87%	-63.55%	-60.32%	-57.08%	-58.45%	-52.23%	
MLP	Premiership	2	-63.48%	-65.17%	-65.22%	-66.57%	-63.42%	-65.01%	-64.81%	
		1	-77.51%	-90.43%	-72.68%	-45.79%	-75.36%	-39.49%	-66.88%	
	La -Liga	2	-76.23%	-71.34%	-94.72%	-13.07%	-80.01%	-34.67%	-61.67%	
		1	-90.61%	-84.27%	-80.29%	-40.46%	-19.38%	-79.69%	-65.78%	
	Seria A	2	-77.35%	-49.73%	-92.39%	-42.49%	-83.27%	-31.34%	-62.76%	
		1	-93.59%	-3.88%	-40.79%	-37.48%	-11.78%	-66.50%	-42.34%	
k-NN	Premiership	2	-36.91%	-82.05%	-19.14%	-77.75%	-49.88%	-73.48%	-56.53%	
		1	-81.17%	-92.76%	-77.47%	-32.76%	-71.13%	-55.57%	-68.48%	
	La -Liga	2	-93.90%	-85.03%	-33.13%	-72.74%	-27.86%	15.09%	-49.60%	
		1	-73.03%	-45.28%	-83.06%	-89.12%	-33.63%	-71.41%	-65.92%	
	Seria A	2	-67.94%	-55.63%	-63.77%	-51.58%	-51.68%	-85.80%	-62.73%	
		1	-88.46%	-64.39%	-71.27%	-74.32%	38.80%	-14.08%	-45.62%	
CF-IF	Premiership	2	11.60%	-63.98%	-87.61%	-33.17%	-58.16%	-70.35%	-50.28%	
		1	83.57%	-21.02%	-66.06%	84.76%	-5.45%	-11.47%	10.72%	
	La -Liga	2	-12.82%	-13.47%	-35.67%	-28.43%	-11.03%	-53.14%	-25.76%	
		1	-48.93%	-21.64%	-29.18%	-8.27%	4.60%	-45.64%	-24.84%	
	Seria A	2	-22.12%	-60.52%	-9.80%	-27.08%	-37.76%	-29.16%	-31.07%	
		1	6.42%	-3.09%	1.97%	-27.52%	7.82%	-16.76%	-5.19%	
DT	Premiership	2	-21.13%	-26.81%	-23.48%	-11.06%	-6.36%	4.54%	-14.05%	
		1	-9.96%	-71.35%	26.21%	-93.74%	33.57%	-85.56%	-33.47%	
	La -Liga	2	-66.39%	-52.64%	-32.89%	-26.06%	-92.73%	-13.14%	-47.31%	
		1	-30.59%	-70.39%	-52.81%	-74.56%	27.96%	-66.39%	-44.46%	
	Seria A	2	-83.89%	-87.51%	-53.64%	-75.65%	-43.09%	-80.00%	-70.63%	
		1	-48.08%	28.49%	-91.69%	-49.48%	-64.70%	26.89%	-33.09%	
SVM	Premiership	2	-88.06%	-47.45%	-81.75%	-66.65%	-53.44%	-58.30%	-65.94%	
		1	-79.45%	-91.68%	71.25%	-52.31%	-75.01%	-74.56%	-50.29%	
	La -Liga	2	-38.31%	1.68%	-85.14%	-25.55%	-84.26%	-94.38%	-54.33%	
		1	-74.56%	-76.70%	15.04%	-53.13%	-45.87%	-85.57%	-53.46%	
	Seria A	2	-80.36%	-28.57%	-32.61%	-74.61%	-76.53%	-52.17%	-57.48%	
		1	-60.71%	-1.22%	-78.12%	4.98%	-55.30%	-85.34%	-45.95%	
		2	-90.70%	-81.08%	-40.10%	-60.30%	-46.50%	-55.22%	-62.32%	

Note: All values in the Table represent the proportional cumulative annualized return per season. Row 1 corresponds to the first year of the out-of-sample and row 2 to the second year of the out-of-sample. For example, for the first cell on the left corner, 7.29% is the average profit per bet of CF for the 2009-2010 Premiership season and the 0.85% is the performance of the exact same model (same specification and rules) for the 2010-2011 season of the same championship.

We note that the pattern of the RVM-CF's profitability is similar to the one obtained by the previous metric. This profitability varies throughout the seasons and the championships but remains positive in the first year of the out-of-sample. Conversely, now the average profitability of RVM-CF in the second year of the out-of-sample is not always positive. The other models present a consistent negative performance in all exercises. Especially, the traditional MLP, k-NN and DT appear to perform worse than the rest of the benchmarks, while SVM performance is far below the RVM one especially in the case of game result. In this metric, the information-fusion benchmark outperforms the other benchmarks, but it does not manage to achieve positive returns as in RVM-CF. The Kelly criterion allows us to bet based on the probability of a favourable outcome. The proportion of capital dedicated

for each game is based on the probability of winning the bet based on RVM-CF or OP. Tables 8 and 9 present the average profit per bet of the RVM-CF<sup>10</sup> and the OP based on the Kelly criterion respectively.

**Table 8: Kelly Criterion (RVM-CF)**

Exercise	Championship	Out-of-sample	In-sample						Average
			2006-2009	2007-2010	2008-2011	2009-2012	2010-2013	2011-2014	
Game Result	Premiership	1	17.03%	20.27%	23.93%	25.90%	27.88%	15.34%	21.73%
		2	5.91%	12.45%	11.67%	4.88%	3.33%	6.16%	7.40%
	La -Liga	1	12.13%	15.27%	10.38%	17.57%	8.35%	4.91%	11.44%
		2	4.49%	7.66%	5.02%	6.32%	5.29%	3.45%	5.37%
	Seria A	1	6.17%	11.13%	24.61%	11.93%	8.10%	15.77%	12.95%
		2	2.78%	-4.50%	3.12%	8.12%	6.34%	6.25%	3.69%
Number of Goals	Premiership	1	16.58%	17.31%	10.64%	12.23%	27.07%	19.84%	17.28%
		2	3.51%	2.37%	4.27%	4.59%	9.68%	6.54%	5.16%
	La -Liga	1	12.60%	15.86%	18.32%	15.91%	14.67%	11.32%	14.78%
		2	2.69%	3.23%	5.67%	4.25%	6.87%	3.85%	4.43%
	Seria A	1	17.54%	11.32%	6.03%	19.67%	15.41%	19.43%	14.90%
		2	5.20%	2.19%	3.14%	7.37%	6.42%	6.44%	5.13%

Note: All values in the Table represent the average profit per season of CF with the Kelly criterion. Row 1 corresponds to the first year of the out-of-sample and row 2 to the second year of the out-of-sample. For example, for the first cell on the left corner, 17.03% is the average profit per bet of CF for the 2009-2010 Premiership season and the 5.91% is the performance of the exact same model (same specification and rules) for the 2010-2011 season of the same championship.

**Table 9: Kelly Criterion (OP)**

Exercise	Championship	Out-of-sample	In-sample						Average
			2006-2009	2007-2010	2008-2011	2009-2012	2010-2013	2011-2014	
Game Result	Premiership	1	0.53%	2.35%	1.59%	-3.08%	7.35%	-1.71%	1.17%
		2	-0.92%	-2.15%	-3.62%	-2.29%	-3.00%	-0.35%	-2.06%
	La -Liga	1	3.90%	0.41%	8.38%	-1.53%	-10.05%	-0.09%	0.17%
		2	1.45%	-4.17%	-3.84%	-5.16%	2.99%	-2.69%	-1.90%
	Seria A	1	-0.33%	2.61%	0.84%	0.36%	1.35%	-2.50%	0.39%
		2	-3.71%	-0.64%	-6.35%	-0.61%	-3.80%	-1.06%	-2.70%
Number of Goals	Premiership	1	-1.88%	-0.43%	0.78%	-1.03%	0.10%	5.48%	0.50%
		2	-0.76%	1.08%	-2.27%	0.39%	-1.27%	-1.68%	-0.75%
	La -Liga	1	-0.08%	1.38%	1.37%	-1.91%	-1.00%	-1.95%	-0.37%
		2	-0.61%	0.58%	-2.28%	-3.15%	-1.32%	-0.50%	-1.21%
	Seria A	1	4.26%	-0.08%	-2.58%	1.05%	-13.08%	6.58%	-0.64%
		2	-1.26%	-6.92%	-6.75%	-14.05%	-1.55%	-1.58%	-5.35%

Note: All values in the Table represent the average profit per season of OP with the Kelly criterion. Row 1 corresponds to the first year of the out-of-sample and row 2 to the second year of the out-of-sample. For example, for the first cell on the left corner, 0.53% is the average profit per bet of CF for the 2009-2010 Premiership season and the -0.92% is the performance of the exact same model (same specification and rules) for the 2010-2011 season of the same championship.

From Tables 8 and 9 we observe an increase in the profitability in most exercises of RVM-CF and OP with the Kelly criterion. This increase is obvious in the cases where our models were presenting before negative results, namely the second out-of-sample year of RVM-CF and all OP exercises. In first year of the out-of-sample for RVM-CF, where RVM-CF had already positive profit per bet, the results remain similar.

As discussed earlier we argue that bookmakers' odds are biased to exploit the bettors' behaviour. The nominator of the Kelly fraction (equation 12) is the so called "edge" or expected return of the bet. The edge considers both the correct prediction probability and the odds. When the odds are biased the edge decreases. Thorp (2008) argues that Kelly criterion may need millions of trials to dominate other strategies in case of having a low edge. On the other hand, Maclean et al. (2010) find Kelly criterion to be very risky in the short term and argue that

<sup>10</sup> Given the superior performance of RVM-CF over CF-IF, we apply the Kelly criterion only to RVM-CF.

despite its promising long-run growth properties, it may lead to poor return outcomes. In our exercise, Kelly improves the betting performance in case of having enough trials, such as in OP where a bet is placed for each game. Similarly, when RVM-CF loses power and the forecasts deteriorate (as in the second year of out-of-sample), Kelly can reduce the exposure to the risk and control the size of losses. On the other hand, for RVM-CF in the first year of out-of-sample, where the model produces already significant profits, the Kelly criterion seems to have no effect on our results. This happens because the number of games where RVM-CF is applied is small and our model produce significant positive results that are not affected by the odds' biases. At last, we note that in most cases, our models are more accurate and profitable in the English Premiership. We assume that the English championship has less noise or in other words, it is easier to be forecasted<sup>11</sup>. At last, we demonstrate the benefits of the Kelly criterion and how it can be translated within the RVM-CF context. Its probabilistic nature seems highly beneficial in football betting in cases where the underlying model has a low power.

### 3.3.2 Stocks Empirical Results

The performance of our models in terms of accuracy ratio and annualized excess return for the PFE stock are presented in Table 10. Additional measures of classification are provided in Table S.22 of Online Appendix S.8. Online Appendix S.6.2 presents the number of forecasts obtained by CF. The Pesaran-Timmermann (1992) test for the aggregate performance of the models is also estimated.

**Table 10: Models' performance (PFE stock)**

In-sample	Out-of-sample	OP	RVM	MLP	k-NN	DT	SVM	ANFIS	CF
2002-2006	2007	47.17% (-3.91%)	49.06% (12.24%)	54.72% (20.54%)	47.17% (9.26%)	58.49% (13.86%)	49.06% (11.41%)	49.06% (-8.6%)	56.52% (2.91%)
2003-2007	2008	46.15% (-5.73%)	50.00% (12.4%)	50.00% (0.47%)	59.62% (18.34%)	46.15% (-5.55%)	40.38% (-24.37%)	53.85% (10.02%)	50.00% (35.4%)
2004-2008	2009	61.54% (0.97%)	61.54% (21.67%)	46.15% (-12.34%)	40.38% (-22.68%)	53.85% (13.13%)	46.15% (-2.27%)	34.62% (-24.78%)	45.95% (-4.69%)
2005-2009	2010	59.62% (30.41%)	51.92% (15.94%)	53.85% (-15.72%)	46.15% (6.79%)	55.77% (11.71%)	55.77% (-5.6%)	59.62% (29.33%)	62.75% (33.76%)
2006-2010	2011	61.54% (35.35%)	36.54% (-19.35%)	50.00% (-1.81%)	36.54% (-22.59%)	53.85% (11.36%)	59.62% (21.29%)	36.54% (-23.59%)	43.14% (-16.53%)
2007-2011	2012	50.94% (10.75%)	60.38% (16.14%)	41.51% (-25.72%)	41.51% (-16.93%)	60.38% (5.97%)	52.83% (16.8%)	45.28% (3.06%)	61.54% (44.71%)
2008-2012	2013	57.69% (16.93%)	57.69% (20.80%)	48.08% (2.39%)	42.31% (-19.05%)	44.23% (-22.41%)	55.77% (17.08%)	61.54% (15.63%)	64.00% (16.82%)
2009-2013	2014	59.62% (9.98%)	53.85% (2.03%)	57.69% (11.38%)	53.85% (-2.58%)	59.62% (12.34%)	48.08% (3.03%)	61.54% (26.44%)	61.22% (25.68%)
2010-2014	2015	48.08% (5.3%)	51.92% (-1.11%)	59.62% (19.35%)	51.92% (0.25%)	44.23% (-7.95%)	48.08% (3.49%)	53.85% (17.05%)	52.08% (16.11%)
2011-2015	2016	44.23% (2.39%)	53.85% (10.36%)	46.15% (-2.29%)	59.62% (4.91%)	51.92% (-7.76%)	36.54% (-0.58%)	59.62% (3.31%)	52.50% (2.30%)
<b>Average</b>		<b>53.66%</b> <b>(10.24%)</b>	<b>52.67%</b> <b>(9.11%)</b>	<b>50.78%</b> <b>(-0.38%)</b>	<b>47.91%</b> <b>(-4.43%)</b>	<b>52.85%</b> <b>(2.47%)</b>	<b>49.23%</b> <b>(4.03%)</b>	<b>51.55%</b> <b>(4.79%)</b>	<b>54.97%*</b> <b>(15.65%)</b>

*Note: The first two columns report the in-sample and out-of-sample years. The cells reporting the models' performance present the accuracy and the annualized excess return in the parenthesis. The average performances over the out-of-sample periods of 2010-2016 are highlighted in bold. \* indicates that according to the Pesaran-Timmermann (1992) test the aggregate forecasts are statistically accurate in identifying the directional change of PFE stock price at 95% level. Online Appendix S.8 presents additional measures of classification for all models.*

<sup>11</sup> It is well known amongst football fans, that in the English Premiership, teams' competition is higher. The number of teams that fight for the championship or to avoid relegation is higher than the one in La Liga and Seria A. Also, some of the biggest bookmakers originate from UK while football betting in the English Premiership is widely prevalent in Europe and Asia.

The results the finance application further validate the success of CF. We observe that with CF is achieving on average the highest accuracies and profitability. CF presents more than 5% and 6% higher excess returns from the second (OP) and third (RVM) best model respectively. MLP and k-NN present on average negative profitability over the ten-year period, while k-NN and SVM does not manage to exceed the 50% accuracy benchmark for this case. Following the pattern of previous results, SVM's profitability falls far behind RVM's one. Although ANFIS is achieving high profitability in specific years, on average the annualized excess returns are around 5% lower than OP and RVM. Finally, DT overcomes the accuracy benchmark, but its profitability is the lowest after the negative ones of MLP and k-NN. The relevant GS stock results confirm the previous results and are presented in Online Appendix S.7.

#### **4. Conclusions**

This study introduces a CF concept in forecasting and demonstrates its utility. CF generates a set of FRs in the in-sample and estimates their average firing strength. These rules are ranked and applied in the out-of-sample. CF generates forecasting signals at points where strong rules are nearby and satisfy an endogenous and an exogenous threshold. The forecasting signal is a weighted average of the strongest rules. CF is useful in OR problems where uncertainty is high and poor forecasts are associated with substantial losses. It can offer transparency, protection against under and over-fitting while at the same time improves the forecasting accuracy of the underlying system.

In order to examine the CF performance, two forecasting exercises are designed. The first exercise is based on the game result, Asian handicap and number of goals of football games in the Premiership, La Liga and Seria A championships. CF is combined with RVM and generates forecasts in six consecutive exercises. These forecasts are evaluated in terms of statistical accuracy and betting profitability. CF presents higher statistical accuracy and betting profitability than a RVM, ANFIS, OP, MLP, k-NN, DT, SVM and a CF-IF model where the CF inputs vector is based on Information Fusion-based approach. We note that although forecasting accurately the game result and number of goals does not seem a strenuous task with our models, profitability is only obtained through CF. Our procedure outperforms not only the most popular fuzzy approach (ANFIS) but also improves the predictability of the underlying system. This contrasts the common dilemma with developing transparent decision systems that sacrifice accuracy for interpretability (Martens, 2007). The introduction of the Kelly ratio can further improve the profitability of CF and limit the losses of OP in most cases. The second application aims to forecast the direction of stock price change of two stocks. In this exercise, the inputs vector is small enough to be fed directly to CF. Our results validate the ones of the previous exercise, with CF providing the most accurate forecasts and the highest annualized excess returns on average.

Both empirical applications presented above have demonstrated the merits of CF, since CF's generated forecasts are accurate and profitable. These forecasts are generated from a transparent process, which also reveals the factors that determine the target series. This is important given that we applied CF to two independent forecasting exercises of different nature. The first set of results contributes to the literature of the football

betting market efficiency, since we note that it is possible to generate consistent profits with CF. In the stock market case, some benchmarks (OP, RVM and ANFIS) have a time-varying profitability. Nonetheless, CF is the only model that is able to translate consistently forecasting accuracy to profitability across all periods.

## 5. Implications, Drawbacks and future works

CF can be a useful tool for practitioners and academics who deal with OR problems that demand complex non-linear techniques. In areas such as Medicine, Finance and Economics it can improve researchers' understanding of the underlying nature of the series. Where the general public is concerned, in areas such as financial trading or football betting, it can offer decision rules that are simple and interpretable. It is an attractive alternative to the numerous unconditional fuzzy inference approaches that dominate the literature. On the other hand, CF needs large datasets as it is applied in only a subset of forecasts. This prohibits its application in managerial problems where the practitioner wants a forecast for every sample or the available dataset is short.

The current study introduces CF as a robust forecasting technique and evaluates its performance through two applications. Despite the demonstrated merits, room exists for further improvements in forecast accuracy. Future research in CF could enhance its methodology as well as its application scope. One possible pathway is applying the conditional framework proposed for CF to other machine learning techniques such as SVM and MLP. This line of inquiry has much relevance for the evaluation of cost-sensitive datasets for which errors result in severe losses (e.g. football betting). Another area to explore is using alternative forecast combination techniques, as was conducted using IF technique in this survey to select the best performing model dynamically. A Bayesian treatment such as dynamic model averaging (Raftery et al., 2010) could be explored in further research as an alternative to the IF. In terms of application, future studies could deploy the CF in other fields of practice to study how it could assist decision-making processes.

### Acknowledgments

We would like to thank the three anonymous reviewers, Jason Laws, Frank McGroarty, Hans-Jörg von Mettenheim, Neil Kellard, Hans-Georg Zimmermann, Philip Newall, as well as the participants at the Operations Research seminar of the University of Oviedo (2014), the 27<sup>th</sup> European Conference on Operational Research (2015), the 23<sup>rd</sup> Forecasting Financial Markets conference (2015), the Computational Operations Research workshop (2016), and the 3<sup>rd</sup> Quantitative Finance and Risk Analysis conference (2017) for their helpful comments.

### REFERENCES

- Akkoç, S., 2012. An empirical comparison of conventional techniques, neural networks and the three stage hybrid Adaptive Neuro Fuzzy Inference System (ANFIS) model for credit scoring analysis: The case of Turkish credit card data. *European Journal of Operational Research*, 222(1), pp.168-178.
- Alpaydm, E., 1999. Combined  $5 \times 2$  cv F test for comparing supervised classification learning algorithms. *Neural computation*, 11(8), pp.1885-1892.
- Andersson, P., Memmert, D. and Popowicz, E. 2009. Forecasting outcomes of the World Cup 2006 in football: Performance and confidence of bettors and laypeople. *Psychology of Sport and Exercise*, 10(1), 116-123.
- Angelini, G. and De Angelis, L., 2017. PARX model for football match predictions. *Journal of Forecasting*, 36(7), pp.795-807.

- Atsalakis, G.S., Atsalaki, I.G., Pasiouras, F. and Zopounidis, C., 2019. Bitcoin price forecasting with neuro-fuzzy techniques. *European Journal of Operational Research* (forthcoming).
- Atsalakis, G.S. and Valavanis, K.P., 2009. Forecasting stock market short-term trends using a neuro-fuzzy based methodology. *Expert systems with Applications*, 36(7), pp.10696-10707
- Baboota, R. and Kaur, H. 2018. Predictive analysis and modelling football results using machine learning approach for English Premier League. *International Journal of Forecasting*, (forthcoming)
- Bastos, L.S. and da Rosa, J.M.C., 2013. Predicting probabilities for the 2010 FIFA World Cup games using a Poisson-Gamma model. *Journal of Applied Statistics*, 40(7), pp.1533-1544.
- Bellman, R.E. and Zadeh, L.A., 1970. Decision-making in a fuzzy environment. *Management Science*, 17(4), pp. B141-B164.
- Bezdek, J.C., Ehrlich, R. and Full, W., 1984. FCM: The fuzzy c-means clustering algorithm. *Computers & Geosciences*, 10(2-3), pp.191-203.
- Bogaert, M., Ballings, M. and Van den Poel, D., 2018. Evaluating the importance of different communication types in romantic tie prediction on social media. *Annals of Operations Research*, 263(1-2), pp.501-527.
- Boshnakov, G., Kharrat, T. and McHale, I. G. 2017. A bivariate Weibull count model for forecasting association football scores. *International Journal of Forecasting*, 33(2), 458-466.
- Breiman, L., Friedman, J. H., Olshen, R. A. and Stone, C. J., 1984. Classification and regression trees. *Wadsworth & Brooks/Cole Advanced Books & Software, Monterey, CA.*
- Chang, P.C., Liu, C.H. and Lai, R.K., 2008. A fuzzy case-based reasoning model for sales forecasting in print circuit board industries. *Expert Systems with Applications*, 34(3), pp.2049-2058.
- Chen, Y. and Hao, Y., 2017. A feature weighted support vector machine and K-nearest neighbor algorithm for stock market indices prediction. *Expert Systems with Applications*, 80, pp.340-355.
- Cheng, C.B. and Lee, E.S., 2001. Switching regression analysis by fuzzy adaptive network. *European Journal of Operational Research*, 128(3), pp.647-663.
- Constantinou, A.C., Fenton, N.E. and Neil, M., 2012. pi-football: A Bayesian network model for forecasting Association Football match outcomes. *Knowledge-Based Systems*, 36, pp.322-339.
- Crowder, M., Dixon, M., Ledford, A. and 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), pp.157-168.
- Dixon, M.J. and 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), pp.265-280.
- Dixon, M. and Robinson, M. 1998. A birth process model for association football matches. *Journal of the Royal Statistical Society: Series D (The Statistician)*, 47(3), 523-538.
- Díez-Pastor, J.F., Rodríguez, J.J., García-Osorio, C. and Kuncheva, L.I., 2015. Random balance: ensembles of variable priors classifiers for imbalanced data. *Knowledge-Based Systems*, 85, pp.96-111.
- Dobson, S. and Goddard, J., 2003. Persistence in sequences of football match results: A Monte Carlo analysis. *European Journal of Operational Research*, 148(2), pp.247-256.
- Feess, E., Müller, H. and Schumacher, C. 2016. Estimating risk preferences of bettors with different bet sizes. *European Journal of Operational Research*, 249(3), 1102-1112.
- Forrest, D. and Simmons, R., 2008. Sentiment in the betting market on Spanish football. *Applied Economics*, 40(1), pp.119-126.
- Fukami, S., Mizumoto, M. and Tanaka, K., 1980. Some considerations on fuzzy conditional inference. *Fuzzy sets and Systems*, 4(3), pp.243-273.
- Goddard, J. and Asimakopoulos, I., 2004. Forecasting football results and the efficiency of fixed-odds betting. *Journal of Forecasting*, 23(1), pp.51-66.
- Goddard, J., 2005. Regression models for forecasting goals and match results in association football. *International Journal of forecasting*, 21(2), pp.331-340.

- Gomes, J., Portela, F. and Santos, M.F., 2016. Pervasive Decision Support to predict football corners and goals by means of data mining. In *New Advances in Information Systems and Technologies* (pp. 547-556). Springer, Cham.
- Gong, X., Si, Y.W., Fong, S. and Biuk-Aghai, R.P., 2016. Financial time series pattern matching with extended UCR Suite and Support Vector Machine. *Expert Systems with Applications*, 55, pp.284-296.
- Gradojevic, N. and Gençay, R., 2013. Fuzzy logic, trading uncertainty and technical trading. *Journal of Banking & Finance*, 37(2), pp.578-586.
- Guégan, D. and Huck, N., 2005. On the use of nearest neighbors in finance. *Finance*, 26(2), pp.67-86.
- Guresen, E., Kayakutlu, G. and Daim, T.U., 2011. Using artificial neural network models in stock market index prediction. *Expert Systems with Applications*, 38(8), pp.10389-10397.
- Guillaume, S., 2001. Designing fuzzy inference systems from data: An interpretability-oriented review. *IEEE transactions on fuzzy systems*, 9(3), pp.426-443.
- Hausman, J.A., Lo, A.W. and MacKinlay, A.C., 1992. An ordered probit analysis of transaction stock prices. *Journal of Financial Economics*, 31(3), pp.319-379.
- Hruschka, H., 1988. Use of fuzzy relations in rule-based decision support systems for business planning problems. *European journal of operational research*, 34(3), pp.326-335.
- Huang, K.Y. and Chen, K.J., 2011. Multilayer perceptron for prediction of 2006 world cup football game. *Advances in Artificial Neural Systems*, 2011 (11), p.1-8.
- Huang, S.C. and Wu, T.K., 2008. Combining wavelet-based feature extractions with relevance vector machines for stock index forecasting. *Expert Systems*, 25(2), pp.133-149.
- Igiri, C. P. 2015. Support Vector Machine—Based Prediction System for a Football Match Result. *IOSR Journal of Computer Engineering (IOSR-JCE)*, 17(3), pp. 21-26.
- Jang, J.S., 1993. ANFIS: adaptive-network-based fuzzy inference system. *IEEE transactions on systems, man, and cybernetics*, 23(3), pp.665-685.
- Kadri, M.B., 2014. Disturbance rejection using fuzzy model free adaptive control (FMFAC) with adaptive conditional defuzzification threshold. *Journal of the Franklin Institute*, 351(5), pp.3013-3031.
- Kelly, J.L., 1956. A new interpretation of information rate. *Bell Labs Technical Journal*, 35(4), pp.917-926.
- Koopman, S.J. and 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), pp.167-186.
- Kuncheva, L.I., 2000. How good are fuzzy if-then classifiers?. *IEEE Transactions on Systems, Man, and Cybernetics, Part B (Cybernetics)*, 30(4), pp.501-509.
- Kuo, R.J., 2001. A sales forecasting system based on fuzzy neural network with initial weights generated by genetic algorithm. *European Journal of Operational Research*, 129(3), pp.496-517.
- Kuypers, T., 2000. Information and efficiency: an empirical study of a fixed odds betting market. *Applied Economics*, 32(11), pp.1353-1363.
- Maclean, L.C., Thorp, E.O. and Ziemba, W.T., 2010. Long-term capital growth: the good and bad properties of the Kelly and fractional Kelly capital growth criteria. *Quantitative Finance*, 10(7), pp.681-687.
- Martens, D., Baesens, B., Van Gestel, T. and Vanthienen, J., 2007. Comprehensible credit scoring models using rule extraction from support vector machines. *European journal of operational research*, 183(3), pp.1466-1476.
- Martins, R. G., Martins, A. S., Neves, L. A., Lima, L. V., Flores, E. L., and do Nascimento, M. Z. 2017. Exploring polynomial classifier to predict match results in football championships. *Expert Systems with Applications*, 83, 79-93.
- Min, B., Kim, J., Choe, C., Eom, H. and McKay, R.B., 2008. A compound framework for sports results prediction: A football case study. *Knowledge-Based Systems*, 21(7), pp.551-562.
- Mizumoto, M. and Zimmermann, H.J., 1982. Comparison of fuzzy reasoning methods. *Fuzzy sets and systems*, 8(3), pp.253-283.

Oberstone, J. 2009. Differentiating the top English premier league football clubs from the rest of the pack: Identifying the keys to success. *Journal of Quantitative Analysis in Sports*, 5(3).

Oberstone, J. 2011. Comparing team performance of the English premier league, Serie A, and La Liga for the 2008-2009 season. *Journal of Quantitative Analysis in Sports*, 7(1).

Oztekin, A., Delen, D., Turkyilmaz, A. and Zaim, S., 2013. A machine learning-based usability evaluation method for eLearning systems. *Decision Support Systems*, 56, pp.63-73.

Oztekin, A., Kizilaslan, R., Freund, S. and Iseri, A., 2016. A data analytic approach to forecasting daily stock returns in an emerging market. *European Journal of Operational Research*, 253(3), pp.697-710.

Owramipur, F., Eskandarian, P. and Mozneb, F.S., 2013. Football result prediction with Bayesian network in Spanish League-Barcelona team. *International Journal of Computer Theory and Engineering*, 5(5), p.812.

Patel, J., Shah, S., Thakkar, P. and Kotecha, K., 2015. Predicting stock and stock price index movement using trend deterministic data preparation and machine learning techniques. *Expert Systems with Applications*, 42(1), pp.259-268.

Pesaran, M.H. and Timmermann, A., 1992. A simple nonparametric test of predictive performance. *Journal of Business & Economic Statistics*, 10(4), pp.461-465.

Piramuthu, S., 1999. Financial credit-risk evaluation with neural and neurofuzzy systems. *European Journal of Operational Research*, 112(2), pp.310-321.

Purda, L.D., 2007. Stock market reaction to anticipated versus surprise rating changes. *Journal of Financial Research*, 30(2), pp.301-320.

Rafi, M. and Shaikh, M. S., 2013. A comparison of SVM and RVM for Document Classification. *arXiv preprint arXiv:1301.2785*.

Raftery, A.E., Kárný, M. and Ettlér, P., 2010. Online prediction under model uncertainty via dynamic model averaging: Application to a cold rolling mill. *Technometrics*, 52(1), pp.52-66.

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

Şahin, M. and Erol, R., 2018. Prediction of Attendance Demand in European Football Games: Comparison of ANFIS, Fuzzy Logic, and ANN. *Computational intelligence and neuroscience*, 2018.

Schumaker, R. P., Jarmoszko, A. T. and Labeledz Jr, C. S. 2016. Predicting wins and spread in the Premier League using a sentiment analysis of twitter. *Decision Support Systems*, 88, 76-84.

Sevim, C., Oztekin, A., Bali, O., Gumus, S. and Guresen, E., 2014. Developing an early warning system to predict currency crises. *European Journal of Operational Research*, 237(3), pp.1095

Singpurwalla, N.D. and Booker, J.M., 2004. Membership functions and probability measures of fuzzy sets. *Journal of the American Statistical Association*, 99(467), pp.867-877.

Štrumbelj, E. 2014. On determining probability forecasts from betting odds. *International journal of forecasting*, 30(4), 934-943.

Sugeno, M., 1985. Industrial applications of fuzzy control. *Elsevier Science, New York, NY, USA*.

Sun, B., Guo, H., Karimi, H.R., Ge, Y. and Xiong, S., 2015. Prediction of stock index futures prices based on fuzzy sets and multivariate fuzzy time series. *Neurocomputing*, 151(3), pp.1528-1536.

Tay, F.E. and Cao, L.J., 2002. Modified support vector machines in financial time series forecasting. *Neurocomputing*, 48(1-4), pp.847-861.

Teodorović, D., 1994. Fuzzy sets theory applications in traffic and transportation. *European Journal of Operational Research*, 74(3), pp.379-390.

Thorp, E. O. (2008). The Kelly criterion in blackjack sports betting, and the stock market. In Zenios, S.A and Ziemba, W.T. (Eds) *Handbook of asset and liability management*, 385-428.

Ticknor, J.L., 2013. A Bayesian regularized artificial neural network for stock market forecasting. *Expert Systems with Applications*, 40(14), pp.5501-5506.

Tipping, M.E., 2001. Sparse Bayesian learning and the relevance vector machine. *Journal of machine learning research*, 1(Jun), pp.211-244.



Trawinski, K., 2010. A fuzzy classification system for prediction of the results of the basketball games. In *IEEE International Conference on Fuzzy Systems (FUZZ), 2010* (pp. 1-7). IEEE.

Vapnik, V. N. 1995. The nature of statistical learning theory. *Springer, New York, NY, USA*.

Vlastakis, N., Dotsis, G. and Markellos, R.N., 2008. Nonlinear modelling of European football scores using support vector machines. *Applied Economics*, 40(1), pp.111-118.

Wang, Y.F., 2002. Predicting stock price using fuzzy grey prediction system. *Expert systems with applications*, 22(1), pp.33-38

Wang, L.X. and Mendel, J.M., 1992. Fuzzy basis functions, universal approximation, and orthogonal least-squares learning. *IEEE Transactions on Neural Networks*, 3(5), pp.807-814.

Wang, J.Z., Wang, J.J., Zhang, Z.G. and Guo, S.P., 2011. Forecasting stock indices with back propagation neural network. *Expert Systems with Applications*, 38(11), pp.14346-14355.

Yang, J.W. and Parwada, J., 2012. Predicting stock price movements: an ordered probit analysis on the Australian Securities Exchange. *Quantitative Finance*, 12(5), pp.791-804.

Zadeh, L.A., 1965. Fuzzy sets. *Information and control*, 8(3), pp.338-353.

Zadeh, L.A., 1983. The role of fuzzy logic in the management of uncertainty in expert systems. *Fuzzy sets and systems*, 11(1), pp.199-227.

Zhang, G., Hu, M.Y., Patuwo, B.E. and Indro, D.C., 1999. Artificial neural networks in bankruptcy prediction: General framework and cross-validation analysis. *European journal of operational research*, 116(1), pp.16-32.

Zhong, X. and Enke, D., 2017. A comprehensive cluster and classification mining procedure for daily stock market return forecasting. *Neurocomputing*, 267, pp.152-168.

## Appendix

### A. CF illustration example

In this appendix, we present a detail CF example for the Premiership, the first forecasting exercise and the game result as a target. The in-sample is the seasons 2006 to 2007, 2007 to 2008 and 2008 to 2009 and the out-of-sample is the season 2009 to 2010.

As mentioned above, the first step is to collect the generated RVs when feeding all the inputs of Table 1 into RVM. For the specific case, a set of 12 RVs are selected and based on these, the RVM generates a series of in-sample forecasts. The RV set is presented below in Table A.1.

**Table A.1: Relevance Vectors (RVs)**

Betbrain average over 2.5 goals odds (BbAv>2.5)	Betbrain size of handicap (home team) (BBAHh)	Betbrain average Asian handicap home team odds (BbAvAHH)	Betbrain average Asian handicap away team odds (BbAvAHH)	Points of H in the last 3 games when H plays at home (PtH3H)
Points of A in the last 3 games when A plays away (PtA3A)	Number of shots on target of H team in the last 1 game (StH1)	Number of shots on target of A team in the last 1 games (StA1)	Number of corner kicks on target of H team in the last 3 game (CkH3)	Number of corner kicks on target of H team in the last 2 game (CkH2)
Number of corner kicks on target of H team in the last 1 game plays home (CkH1H)	Number of corner kicks on target of A team in the last 2 game plays away (CkA2A)			

*Note: Team H is the home team and team A is the away team. The parentheses represent the relevant abbreviation of the RV. In total, 12 RVs are selected.*

Based on these forecasts, a set of FRs is derived. In this exercise, 13 rules are generated which cover the whole in-sample. The rules have the form:

If ( $BbAv > 2.5$  is cluster  $A_{i,1}$ ) and ( $BbAHh$  is cluster  $A_{i,2}$ ) and ( $BbAvAHH$  is cluster  $A_{i,3}$ ) and ( $BbAvAHA$  is cluster  $A_{i,4}$ ) and ( $PtH3H$  is cluster  $A_{i,5}$ ) and ( $PtA3A$  is cluster  $A_{i,6}$ ) and ( $StH1$  is cluster  $A_{i,7}$ ) and ( $StA1$  is cluster  $A_{i,8}$ ) and ( $CkH3$  is cluster  $A_{i,9}$ ) and ( $CkH2$  is cluster  $A_{i,10}$ ) and ( $CkH1H$  is cluster  $A_{i,11}$ ) and ( $CkA2A$  is cluster  $A_{i,12}$ ), then the output is the result of regression  $\delta$ .

where  $A_{i,k}$  is the cluster specified for input element  $k$  of the  $i$ -th rule. The FRs are specified by the premise and consequent parameters. The premise parameters determine the clusters specification: the centres ( $c_i$ ) and the standard deviations ( $\sigma_i$ ) for each rule. These are presented in Table A.2.

**Table A.2: Cluster characteristics for the generated rules**

Rule \ Input variable	1	2	3	4	5	6	7	8	9	10	11	12	13
<b>BbAv&gt;2.5</b>	2.14 (0.17)	2.12 (0.13)	1.96 (0.12)	2.03 (0.15)	2.00 (0.17)	2.00 (0.17)	2.09 (0.15)	1.75 (0.16)	2.02 (0.15)	1.77 (0.16)	2.01 (0.14)	2.03 (0.20)	2.13 (0.10)
<b>BbAHh</b>	0.00 (0.80)	-0.50 (0.79)	0.01 (0.81)	0.00 (0.80)	-0.01 (0.82)	0.01 (0.80)	-0.01 (0.79)	-1.50 (0.79)	-0.01 (0.78)	-1.50 (0.79)	-0.50 (0.80)	-0.75 (0.80)	-0.25 (0.81)
<b>BbAvAHH</b>	1.83 (1.42)	1.97 (1.42)	1.82 (1.42)	1.94 (1.42)	1.52 (1.42)	2.05 (1.42)	1.58 (1.42)	1.79 (1.42)	1.70 (1.42)	2.03 (1.43)	1.81 (1.42)	1.84 (1.42)	1.88 (1.42)
<b>BbAvAHA</b>	1.99 (1.72)	1.89 (1.71)	1.96 (1.67)	1.81 (1.72)	2.36 (1.71)	1.74 (1.72)	2.29 (1.73)	2.08 (1.72)	2.10 (1.70)	1.83 (1.74)	2.08 (1.72)	2.04 (1.66)	1.98 (1.69)
<b>PtH3H</b>	4.28 (1.59)	6.23 (1.59)	3.02 (1.60)	4.66 (1.59)	0.99 (1.59)	6.99 (1.59)	3.32 (1.61)	6.44 (1.62)	6.68 (1.66)	6.23 (1.59)	8.99 (1.53)	3.99 (1.54)	4.00 (1.59)
<b>PtA3A</b>	2.99 (1.58)	1.00 (1.58)	1.99 (1.59)	6.00 (1.59)	1.01 (1.59)	5.01 (1.59)	5.00 (1.59)	3.00 (1.58)	2.99 (1.58)	0.01 (1.59)	0.99 (1.59)	0.01 (1.59)	3.01 (1.59)
<b>StH1</b>	4.00 (3.88)	2.99 (3.87)	6.00 (3.88)	4.22 (3.89)	6.11 (3.89)	6.57 (3.89)	4.99 (3.89)	8.12 (3.90)	7.32 (3.88)	9.23 (3.85)	3.99 (3.88)	4.32 (3.89)	4.82 (3.89)
<b>StA1</b>	5.99 (3.71)	6.00 (3.71)	3.00 (3.71)	6.01 (3.71)	8.23 (3.71)	9.01 (3.72)	6.02 (3.71)	10.11 (3.71)	3.00 (3.71)	5.03 (3.71)	7.00 (3.71)	4.01 (3.71)	5.01 (3.72)
<b>CkH3</b>	13.00 (6.19)	15.12 (6.19)	19.02 (6.19)	13.48 (6.19)	13.20 (6.19)	17.81 (6.19)	19.82 (6.19)	16.94 (6.19)	16.29 (6.19)	19.31 (6.19)	9.40 (6.18)	9.54 (6.19)	21.39 (6.19)
<b>CkH2</b>	7.74 (4.72)	11.69 (4.27)	14.39 (4.35)	10.12 (4.71)	8.79 (4.68)	9.10 (4.70)	13.21 (4.77)	12.31 (4.72)	11.59 (4.71)	10.30 (4.78)	5.30 (4.90)	5.99 (4.72)	16 (4.75)
<b>CkH1H</b>	3.99 (3.36)	6.59 (3.37)	6.34 (3.32)	6.03 (3.33)	5.29 (3.35)	4.99 (3.34)	10.01 (3.36)	8.32 (3.36)	6.03 (3.36)	6.53 (3.36)	2.99 (3.39)	4.42 (3.36)	12.23 (3.37)
<b>CkA2A</b>	7.99 (4.06)	11.01 (4.07)	10.20 (4.06)	7.01 (4.07)	9.11 (4.06)	10.39 (4.05)	12.90 (4.04)	8.39 (4.07)	8.20 (4.08)	8.87 (4.07)	9.68 (4.05)	7.62 (4.04)	4.07 (4.05)

Note: The values in the Table represent the centre (standard deviation) of the relevant cluster. For instance, consider the first input ( $BbAv > 2.5$ ); to determine which rule an observation belongs to, the membership grade is calculated by the membership function of  $\exp(-(x_{*,1} - 2.14)^2 / (2 \times 0.17^2))$  where  $x_{*,1}$  is the given odd for the Betbrain average for greater than 2.5 goals for the match. The firing strength (weight) of each rule is the product of the membership grades for all inputs.

The output of each rule is associated with a regression  $\delta$  specified with consequent parameters. The parameters of these regressions for each rule are presented in Table A.3.

**Table A.3: Regression coefficients for the generated rules**

Rule Coefficient	1	2	3	4	5	6	7	8	9	10	11	12	13
Intercept	<b>-2.71</b>	-9.44	10.12	2.95	-10.84	0.16	-4.96	2.46	10.57	-8.08	-12.30	15.94	3.46
<b>BbAv&gt;2.5</b>	<b>1.04</b>	1.68	-2.44	-0.18	0.86	-0.31	0.68	-0.99	0.26	1.12	-0.82	-0.45	4.49
<b>BbAHh</b>	<b>-0.76</b>	-0.20	-0.50	-0.26	-1.06	-0.74	-0.39	-0.43	-0.72	-0.70	-0.40	-1.76	-3.10
<b>BbAvAHH</b>	<b>1.07</b>	2.36	-2.39	-0.96	2.64	-0.25	0.46	-0.19	-4.48	1.55	4.10	-6.55	-3.70
<b>BbAvAHA</b>	<b>0.48</b>	1.06	-1.06	-1.10	1.54	0.12	1.2	-0.35	-0.65	0.95	1.54	-2.01	-1.61
<b>PtH3H</b>	<b>-0.04</b>	-0.14	0.16	0.07	0.04	-0.09	-0.02	-0.09	-0.19	0.07	0.42	0.04	-0.34
<b>PtA3A</b>	<b>0.21</b>	-0.44	0.11	0.09	0.14	0.05	0.23	0.16	0.02	0.17	0.37	-0.74	0.10
<b>StH1</b>	<b>-0.14</b>	0.08	0.05	-0.01	0.02	-0.02	0.01	0.01	-0.00	0.01	-0.14	0.06	-0.12
<b>StA1</b>	<b>-0.09</b>	0.12	-0.16	0.01	-0.04	0.03	-0.03	0.08	0.16	0.06	-0.08	0.21	-0.11

CkH3	<b>-0.20</b>	0.05	0.14	0.02	0.08	0.03	-0.02	-0.07	0.01	-0.06	-0.04	0.03	-0.08
CkH2	<b>-0.02</b>	-0.03	-0.13	0.04	0.03	-0.06	0.04	0.11	0.04	-0.02	0.02	-0.06	0.03
CkH1H	<b>0.15</b>	-0.01	-0.06	-0.08	0.05	0.07	0.01	-0.04	-0.15	0.16	0.15	-0.12	0.06
CkA2A	<b>0.07</b>	-0.09	0.05	-0.03	-0.02	0.04	-0.05	-0.03	0.03	0.03	0.06	0.08	-0.005

From the table, it is easy to extract the regression specification for the first rule  $\delta_1^*$  :

$$\begin{aligned} \delta_1^* = & -2.71 + 1.04x_{1,1}^* - 0.76x_{1,2}^* + 1.07x_{1,3}^* + 0.48x_{1,4}^* - 0.04x_{1,5}^* + 0.21x_{1,6}^* \\ & - 0.14x_{1,7}^* - 0.09x_{1,8}^* - 0.20x_{1,9}^* - 0.02x_{1,10}^* + 0.15x_{1,11}^* + 0.07x_{1,12}^* \end{aligned} \quad (A.1)$$

where  $x^* = [x_1^*, \dots, x_{12}^*]$  is the vector of observed values for  $K = 12$  RVs.

The next step is to evaluate the strength of the generated rules. To determine the CF threshold, the algorithm of Section 2.2.2 is followed. Firstly, all rules are evaluated based on the Gaussian membership function. For each of these rules, the average firing strength ( $w^{1/12}$ ) is estimated. The FRs are sorted based on the average firing strength over the in-sample in a descending order (the sorted list). The 90<sup>th</sup> percentile of the sorted list determines the endogenous threshold. In this case, the rule number one is the 90<sup>th</sup> percentile of the sorted list ( $\lambda = 1$ ). Thus, the endogenous threshold level is set to  $w_1^{1/12} = 0.85$ .

As the general threshold level (0.90) is greater than the endogenous one (0.85), the effective threshold ( $\Theta$ ) is set to 0.90. Given the specifications of the FRs and the effective threshold, the performance of the CF model can now be evaluated. At each test point, average membership grade for each rule is calculated by plugging the premise parameters of Table A.1. Comparison of the membership grade with the  $\Theta$  in the evaluation function of equation (3) determines the signal. Based on equation (A.1), the evaluation function becomes:

$$C_i^* = \begin{cases} 1, & \hat{w}_i^{*1/12} \geq 0.9 \\ 0, & \text{Otherwise} \end{cases} \quad (A.2)$$

If the geometric mean of the membership grades of all the inputs (12 elements) is greater than 0.9, the rule is considered strong and qualifies for decision-making and a decision is made based on the CF model. The weighted average of the regressions for the qualified rules determines the CF output (equation (8)). Based on the equation (8), we have:

$$\hat{O}^* = \frac{\sum_{i=1}^{13} \hat{w}_i C_i^* f_i}{\sum_{i=1}^{13} \hat{w}_i C_i^*} \quad (A.3)$$

In the game result forecasting, the match outcome ( $\hat{O}^*$ ) is a weighted average of the regressions specified for each rule (see Table A.2). The weight for each rule's regression is normalized amount of  $\hat{w}_i C_i^*$  as of equation (12). Finally, for betting purposes the estimated value must be interpreted and classified ( $\hat{O}_{adj}^*$ ) under three labels of "win (+1)", "draw (0)" or "lose (-1)". The regression outputs are classified according to the following transfer function:

$$\hat{O}_{adj}^* = \begin{cases} 1, & 0.5 < \hat{O}^* < \infty \\ 0, & -0.5 < \hat{O}^* \leq 0.5 \\ -1, & -\infty < \hat{O}^* \leq -0.5 \end{cases} \quad (A.4)$$