CrossMark

Stock market trend prediction using AHP and weighted kernel LS-SVM

Ivana Marković[1] · **Miloš Stojanović[2]** ·
Jelena Stanković[1] · **Milena Stanković[3]**

Abstract Nowadays, stock market trend prediction represents a challenging subject both in terms of the choice of the prediction model and in terms of constructing the set of features that model will use for prediction. To address both of these aspects, we propose a feature ranking and feature selection approach in combination with weighted kernel least squares support vector machines (LS-SVMs). We introduce the analytic hierarchy process (AHP) into the stock market and propose evaluation criteria which provide the prediction model with relevant knowledge of the underlying processes of the studied stock market. The feature weights obtained by the AHP method are used for feature ranking and selection, and used with the LS-SVMs through a weighted kernel. The test results indicate that the proposed model outperforms the benchmark models. In addition, the set of feature weights obtained by the proposed approach can also independently be incorporated into other kernel-based learners.

✉ Ivana Marković
ivana.markovic@eknfak.ni.ac.rs

Miloš Stojanović
milos.stojanovic@vtsnis.edu.rs

Jelena Stanković
jelenas@eknfak.ni.ac.rs

Milena Stanković
milena.stankovic@elfak.ni.ac.rs

[1] Faculty of Economics, University of Niš, Trg kralja Aleksandra Ujedinitelja 11, 18000 Niš, Serbia

[2] College of Applied Technical Sciences, Aleksandra Medvedeva 20, 18000 Niš, Serbia

[3] Faculty of Electronic Engineering, University of Niš, Aleksandra Medvedeva 14, 18000 Niš, Serbia

1 Introduction

The key to generating a high return on the stock market lies in how well we are able to successfully predict the future movement of financial asset prices (Huang et al. 2005). The stock market index as a hypothetical portfolio of selected stocks is commonly used to measure the performance of both the overall stock market and a particular sector. Consequently, a market trading strategy can be considered effective only if it relies on the precise prediction of the trend of change of the index value of that particular market (Kara et al. 2011; Wang and Choi 2013). Stock market trend prediction represents a challenge for science both in terms of the choice of methodology and in terms of the theoretical basis of its application.

To address these problems, machine learning models, among which the most popular were Artificial Neural Networks (ANNs) (Crone and Kourentzes 2010; Dai et al. 2012; Kara et al. 2011), Support Vector Machines (SVMs) (Huang et al. 2005; Lee 2009; Ni et al. 2011; Yu et al. 2005; Yuling et al. 2013) and Least Squares Support Vector Machines (LS-SVMs) (Chai et al. 2015; Marković et al. 2015; Yu et al. 2009), were the most frequently used alternatives to the classical statistical models in the area of financial forecasting during the last two decades. Due to the principles of the weak form of the efficient market hypothesis (EMH) (Fama 1970; Hawawini and Keim 1995), the behavior of financial asset prices is often governed by a random walk process; thus, the degree of accuracy of an approximate 60 % hit rate obtained in prediction using various machine learning techniques is

🖄 Springer

often considered a satisfactory result for stock market trend prediction (Atsalakis and Valavanis 2009b; Lahmiri 2011).

The determination of sufficient and necessary features is essential for training a good prediction model. If the number of features is insufficient, the prediction accuracy of the model will be poor, and the model may be prone to under-fitting (Stojanović et al. 2014). On the other hand, if we have too many features, the information that they provide for the model could be unnecessary or redundant. As a result, the model could have a poor generalization performance and may be prone to over-fitting. As stated by He et al. (2013) and Barak and Modarres (2015), the most important issue in the creation of a stock market prediction model is the selection of input features for predictors, where the selection of appropriate methods for feature subset selection is highly relevant. In Yuling et al. (2013) and Atsalakis and Valavanis (2009b), it was pointed out that a widely referred to prediction system consists of two parts: feature selection and prediction model design.

According to the selection strategies, feature subset selection can be performed using filter and wrapper approaches, as presented in Guyon and Elisseeff (2003). In the filter methods, the selection criterion uses a selection function which is independent of the learning algorithm used for model construction, as for example, different methods of variable ranking. On the other hand, in the wrapper methods, the selection criterion is based on evaluation measures according to their usefulness to a given learning algorithm. In this way, features that do not contribute to the prediction quality are discarded from the feature set.

In real-world data sets, it is common that different characteristics are less or more relevant to the given problem. However, most learning methods postulate that all the input features have equal relevance. In recent studies, feature weighting has become a very important issue, primarily in the area of clustering algorithms. To increase the effect of relevant features, a learning method that implies Mutual Information (MI) as a criterion, which assigns weights to the features in order to determine their relevance for a specific task, was proposed by Giveki (2012). Guo et al. (2008) introduced spectrally weighted kernels as a way of incorporating theoretical knowledge of the non-uniform information distribution into the machine learning method.

In numerous studies which focused on feature selection in the scope of financial time series and stock market trend prediction, input features are selected based on the analysis of the numerical values of financial assets, including index values, trading volume, financial ratios and technical indicators. For example, in Lee (2009), the F score and Supported Sequential Forward Search (F_SSFS) are combined, and the advantages of both filter and wrapper methods are used to select the optimal feature subset from the original feature set. In Ni et al. (2011), a fractal feature selection method is integrated with SVMs to predict the direction of the daily stock price index. Yu et al. (2005) used a hybrid data mining approach with a genetic algorithm (GA) as a feature selection method. A wide range of various feature selection algorithms, such as GA and sequential forward search, were studied by He et al. (2013). A comprehensive literature review on forecasting techniques can be found in the study of Atsalakis and Valavanis (2009b).

Several recent studies (Fung et al. 2002; Mittermayer 2004; Yoo et al. 2005; Zhai et al. 2007) have been based on qualitative data analysis. Overall, their use of event knowledge and time-series data led to increased accuracy of the prediction models. In Fung et al. (2002), Zhai et al. (2007) and Mittermayer (2004), forecasting of stock price trends is done within the framework of text mining techniques.

According to Atsalakis and Valavanis (2009a) and McNelis (2005), accurate stock market prediction should incorporate how stock market experts learn and process information. Thereby, stock trading is best described as a decision-making process influenced by dynamic market conditions and potential trading risk.

In the multi-criteria decision-making process, where it is necessary to both appropriately evaluate and rank the selected alternatives, the analytic hierarchy process (AHP) developed by Saaty (1999) has widely been applied. Evidence shows that AHP was introduced in several studies for feature weighting in combination with machine learning algorithms (Liu and Shih 2005; Liu et al. 2013; Wang and Zhang 2013). According to our best knowledge, despite their widespread use, there was insufficient evidence on the possibility of optimizing LS-SVMs in the field of stock market trend prediction through customized kernel weighting.

As can be seen, the choice of the prediction method and the determination of its parameters depend on knowing the properties of the underlying processes. The research presented in this paper is motivated by the work presented in Atsalakis and Valavanis (2009a), Guo et al. (2008), Liu et al. (2013), Omak et al. (2007) and Yuling et al. (2013). We proposed an approach concerned with decision making on stock trading, using AHP for feature ranking and feature selection. The contribution of the paper can be summarized as follow:

First, we propose criteria for AHP evaluation of the relevance of technical indicators through the construction of technical trading strategies as a measure of the success of each technical indicator relied on. In this way, we in essence provide the prediction model with a priori knowledge of the underlying processes of the observed stock market.

Second, the weights obtained by AHP are then used for technical indicator ranking and selection. Additionally, the obtained weights are integrated into the LS-SVM through a weighted kernel (WK).

Finally, the AHP-WK-LSSVM model is proposed for stock market trend prediction and tested with the following

data sets: the Belex15 index of the Belgrade Stock Exchange, S&P500 index of the US stock market and FTSE100 index of the London Stock Exchange. The obtained results are then compared with the benchmark results of commonly used classifiers and feature selection algorithms.

The test results indicate that the proposed approach outperforms most of benchmark models and that the set of feature weights obtained in our approach can also be incorporated into other kernel-based learners, such as SVMs.

The rest of this paper is organized as follows: Sect. 2 presents a brief overview of the theoretical preliminaries. Section 3 introduces feature evaluation criteria and the proposed algorithm for feature ranking and selection. Section 4 gives the data set analysis and presents the experimental results and discussions. Finally, Sect. 5 provides the conclusions.

2 Preliminaries

The following section provides an overview of the theoretical framework of Least Squares Support Vector Machines, weighted kernels and the Analytical Hierarchy Process.

2.1 Least squares support vector machines for binary classification

Least squares support vector machines are commonly used for function estimation and for solving non-linear classification problems (Suykens et al. 2002). Let us define the training set $\{x_k, y_k\}, k = 1, \ldots N$, where N represents the overall number of training examples, with the input of $x_k \in R^n$ and an output of $y_k \in \{-1, 1\}$. We can form a prediction model in the primal weight space using non-linear mapping $\phi(\cdot) : R^n \to R^{n_h}$ which maps the input feature space into a multi-dimensional space, defined as:

$$y(x) = \text{sign}[\omega^{\text{T}}\phi(x) + b] \tag{1}$$

where ω represents the weight vector and b defines the bias term.

The optimization problem is formed in the primal space:

$$\min_{\omega,b,e} J_p(\omega, e) = \frac{1}{2}\omega^{\text{T}}\omega + \frac{1}{2}\gamma\sum_{k=1}^{N} e_k^2 \tag{2}$$

with the following constraints:

$$y_k[\omega^{\text{T}}\phi(x_k) + b] = 1 - e_k, \quad k = 1, \ldots, N \tag{3}$$

where e_k are allowed errors during the formation of the prediction model, while γ is a parameter which assigns them with a relative weight.

After solving the optimization problem, the classification model in dual form can be represented as:

$$y(x) = \text{sign}\left[\sum_{k=1}^{N} \alpha_k y_k K(x, x_k) + b\right] \tag{4}$$

The dot product:

$$K(x, x_k) = \phi(x)^{\text{T}}\phi(x_k) \tag{5}$$

in (4) represents a kernel function, while α_k are Lagrange multipliers.

When using a radial basis function (RBF) defined by:

$$K(x, x_k) = e^{\frac{-\|(x-x_k)\|^2}{\sigma^2}} \tag{6}$$

the optimal parameter combination (γ, σ) should be established, where γ denotes the relative weights to allowed e_k errors during the training phase, and σ is a kernel parameter. For this purpose, a grid-search algorithm in combination with a k-fold cross-validation is a commonly used method (Arlot and Celisse 2010).

2.2 Weighted kernel LS-SVMs

In the following section, we present the basics of weighted kernels in relation to LS-SVM theory.

The weighted kernel function is defined as $K(\theta x_i, \theta x)$ where θ is a weight vector of data set features. Without presenting the complete mathematic derivation, which can be found in Xing et al. (2009) for SVMs and can be adopted for LS-SVMs, the classification model in dual form with feature weights is formulated in (7), with the note that feature weights were also included during the computation of α_k and b.

$$y(x) = \text{sign}\left[\sum_{k=1}^{N} \alpha_k y_k K(\theta x, \theta x_k) + b\right] \tag{7}$$

From (7), it can be seen that the defined weighted kernel is not dependent on the type of kernel function (Yao et al. 2006).

The proposed approach used to determine the weight vector $\theta = (\theta_1, \theta_2, \ldots, \theta_d)^{\text{T}}$ is based on the AHP method, and it will be introduced in detail in Sect. 3. However, as presented in Guo et al. (2008) and Xing et al. (2009), it should be noted here that the elements of the feature weight vector obey the following two conditions:

$$0 \leq \theta_i \leq 1 \qquad i = 1, \ldots, d$$

and

$$\sum_{i=1}^{d} \theta_i = 1 \tag{8}$$

The weighted RBF kernel in (6) can now be rewritten as:

$$K(x, x_k) = e^{\frac{-\|\Theta(x - x_k)\|^2}{\sigma^2}} \tag{9}$$

where $\Theta = \text{diag}[\theta_1, \theta_2, \ldots, \theta_n]$.

Further, as in the conventional RBF Kernel, the optimal parameter combination (γ, σ) should be established.

2.3 Basics of the analytic hierarchy process

AHP is a method of selection between sets of factors based on their relevance in terms of meeting even opposing criteria. The AHP calculation techniques are used on a designed pairwise comparison matrix to obtain the eigenvector which represents relative feature values for the obtained criterion. The pairwise comparison is represented using the Fundamental 1–9 Scale, as defined by Saaty (1999). The factors could be classified as Equal importance (denoted with 1), Weak importance of one over another (denoted with 3), Essential or strong importance (denoted with 5), Demonstrated—very strong importance (denoted with 7), and Absolute or extreme importance (denoted with 9). The remaining four scales are intermediate values. The successful application of AHP in various empirical data analysis, which is the result of the clarity of its underlying mathematical principles and its ability to evaluate decision-making consistency, has led to it being used on stock market data in this paper.

The AHP calculations can be summarized as follows: compare n elements, $A_1 \ldots A_n$ and determine the significance of A_i with respect to A_j by p_{ij} to form a reciprocal matrix $P = (p_{ij})_{nxn}$ with the implication that $p_{ij} = 1/p_{ji}$ for $i \neq j$ and $p_{ii} = 1$. For precisely measured data, the P_{ij} matrix is transitive and the eigenvector ω of the order n can be calculated such that $P\omega = \lambda\omega$, where λ is an eigenvalue. Referring to Coyle (2004), in practice, the first step is to provide an initial matrix for the pairwise criteria comparisons to obtain an eigenvector, referred to as the Relative Value Vector (RVV). Next, for each observed criterion, we need a pairwise comparison matrix (PCM) of how well the selected input features perform in terms of each evaluated criterion. Then, the evaluation of the Option Performance Matrix (OPM) enables us to present the observed features in terms of the selected criteria. The final step is the multiplication of the RVV and the OPM, to obtain the overall ranks.

Due to the inconsistency of the decision-making process, the ω vector generally satisfies the equation $P\omega = \lambda_{\max}\omega$ and $\lambda_{\max} \geq n$. The relationship between $\lambda_{\max}$ and n determines the level of (in)consistency of the decisions, where equality between the two is an indication of consistency.

A Consistency Index (CI) is calculated as $(\lambda_{\max} - n)/(n - 1)$ and needs to be determined in relation to a corresponding Random consistency Index (RI) (Saaty 1999), which leads to the calculation of the Consistency Ratio (CR) as follows: CI/RI. It is established that a CR exceeding 0.1 indicates inconsistent decisions, while a CR of 0 indicates perfectly consistent decisions.

3 The proposed approach for feature ranking and selection

In this section, we explain the proposed feature selection procedure and the algorithm for determining feature weights by applying the analytic hierarchy process.

3.1 The proposed AHP evaluation criteria

First, we introduce AHP evaluation criteria for the assessment of the relevance of technical indicators, which in essence provides the model with a priori knowledge of the observed stock market. We suggest the construction of technical trading strategies as a measure of the success of each technical indicator relied on. A technical trading strategy is composed of a set of trading rules that are used to generate trading signals. In general, commonly used trading systems rely on one or two technical indicators that define the timing of trading signals (Kaufman 2003; Pauwels 2011). The AHP evaluation criteria are twofold. The first group consists of two criteria used to measure the economic relevance of the selected indicators: cumulative gross return, as a measure of stock market profitability, and systematic risk as a measure of market volatility. The third criterion represents a comparison of the trading signals generated with a trading strategy and the signals generated based on actual stock market index values, in relation to their achieved prediction accuracy.

3.1.1 Return evaluation

Returns on investments in the case of a specific stock market index were calculated as the differences between daily index values presented in national currency, multiplied by the generated trading signal for the current day. Gross returns were defined as the cumulative capital gains for a specified period of time, as follows:

$$R = \sum_{t=1}^{n} S_t * (\text{CP}_t - \text{CP}_{t-1}) \tag{10}$$

where S_t represents the trading signals generated by the trading strategy. The calculated return on investment value allows
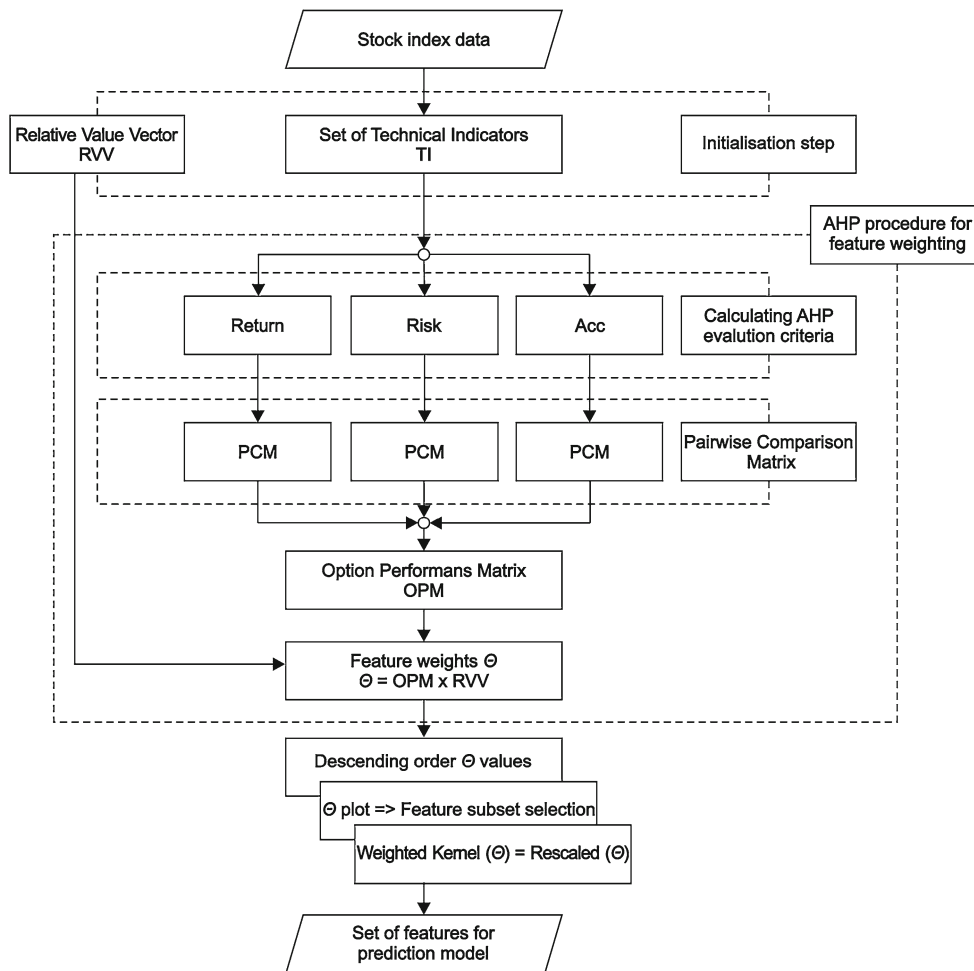
Fig. 1 Algorithm for feature ranking and selection based on AHP evaluation

us to compare the selected set of technical indicators. For the evaluation criteria, we created a relative weighting function which ascribes AHP scale values to the obtained returns, taking into consideration the min–max range of the resulting calculations. The same function is applied in the calculations of the following two criteria.

3.1.2 Risk evaluation

In this study, in addition to return, risk was introduced into stock market prediction as one of the evaluation criteria in the AHP analysis, since in stock trading the return is balanced with a proper level of risk (Barak and Modarres 2015; Rabin 2000). Systematic risk, in relation to return, is defined as:

$$\sigma = \sqrt{\frac{1}{n-1} \sum_{t=1}^{n} (R_t - \bar{R})^2} \qquad (11)$$

where $\bar{R}$ represents the mean value of the gross return R in a selected time period t.

3.1.3 Accuracy evaluation

As a general measure for the evaluation of the prediction effect, the Hit Ratio (HR) was used. HR was calculated based on the number of properly generated trading signals within the test group:

$$\mathrm{HR} = \frac{1}{m} \sum_{i=1}^{m} \mathrm{PO}_i \qquad (12)$$

where PO_i is the prediction output of the ith trading day, that is, S_t for the observed trading strategies. PO_i equals 1 if it is the actual value for the ith trading day; otherwise, PO_i equals 0, and m is the number of data in the used data set.

3.2 Determining feature weights by AHP

The proposed approach for the selection of subsets of the features in accordance with the AHP evaluations is shown in Fig. 1.

After forming the initial set of technical indicators, the first step in the proposed algorithm is the calculation of the criterion values for AHP evaluation. For the technical indicators, calculate values of the evaluation criterion: return, systematic risk and prediction accuracy. The Relative Value Vector is calculated by the methods described in Sect. 2.3. Then three pairwise comparison matrices are constructed. The weights in the matrices reflect how the technical indicators perform in terms of each criterion. According to Sect. 2.3, we then create the Option Performance Matrix, and in the next step multiply the RVV and the OPM to obtain the overall feature weights. The weights (θ) determine the relative significance (ranking) of each input technical indicator candidate in relation to the criterion values. The following step is the sorting of the set of technical indicators in descending order according to θ values. The goal of this step is to find a feature subset that will be used for the prediction model. More precisely, if one plots the weights, the technical indicator that corresponds to the largest weight will add the most information to the prediction model. At some point the feature relevance will decrease, leading to what is known as an "angle" effect in the plot (see Fig. 3). The estimated feature weights for selected features should proportionally be rescaled in accordance with the constraints defined in (8). In the last step, kernel weighting is performed by feature multiplication with rescaled weights, within the input feature space.

4 Experimental results and discussion

This section presents the experimental results and discussion of applying the proposed approach. The goal of the experimental study is to compare the performance of the proposed feature ranking and selection approach in combination with weighted kernel LS-SVMs with benchmark models. The section begins with a description of the datasets used in the experiments, following the experimental setup. Then the results are presented, and finally a discussion of the results concludes the section.

The experiments were conducted on the data for the Belex15, S&P500 and FTSE100 stock market indexes. The value of indexes determines the price of the most liquid stocks traded on the regulated market of the observed markets. The series consists of six time-series values which are determined for each day: the closing price, the change in the value of the index in relation to the previous trading day in percentages, the opening price, highest price, lowest price and the trading volume. The data were divided into two groups. The first group consisted of records required for the model training, from 26 October 2005 to 31 December 2012. The Belex15 index training data set consisted of 1793 samples. The S&P500 training data set consisted of 1775 samples and the FTSE100 training set consisted of 1851 data samples. For
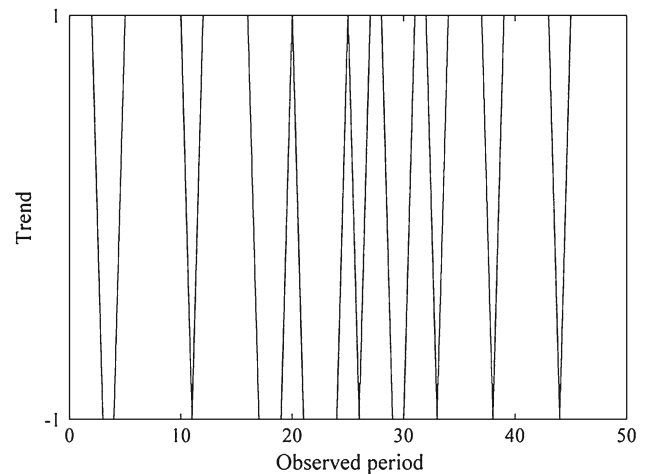

Fig. 2 Trend fluctuations

the model testing, data from 3 January 2013 to 31 December 2013 were used, a total of 252 days of trading for all the data series. The results are obtained for one-day-ahead predictions using data over an extended period of time, 1 trading year, and exceed most of the time horizons presented in the literature (Huang et al. 2005; Ni et al. 2011; Yuling et al. 2013).

The stock market trend prediction problem is commonly modeled as a two-class classification problem where the classes are labeled with -1 and 1. Class -1 indicates that the closing price of the current day is higher than the closing price of the following day. The second class indicates the opposite. Figure 2 shows the trend fluctuations.

From Fig. 2, it can be noticed that the trend fluctuates up and down repeatedly, rendering it challenging for prediction.

4.1 Experimental framework

We consider now the set of nine potential input features. In this study, we rely on the most commonly used technical indicators—Exponential Moving Average (EMA—the moving average of the closing price calculated using a smoothing factor to place a higher weight on recent closing prices), Relative Strength Index (RSI—the index that measures the speed and change of price movements), Stochastic Oscillator $\%K$ (an indicator that predicts the price turning points by comparing a security's closing price to its price range over a given time period), Stochastic Oscillator $\%D$ (the average of the last three $\%K$ values calculated daily), Moving Average Convergence–Divergence (MACD, the indicator that measures the strength and direction of the trend and momentum), ROC (Rate of Change, the indicator that shows the percentile change in the closing prices), Commodity Channel Index (CCI—an indicator used to detect cyclical movements in price change by measuring stock price variations from its statistical mean), and SAR (Parabolic Stop and Reverse—

Table 1 Descriptive statistics for the selected inputs features

Technical indicator	BELEX15				S&P500				FTSEl00			
	Min	Max	Mean	Std dev	Min	Max	Mean	Std dev	Min	Max	Mean	Std dev
ROC	−37.41	38.93	−0.10	7.16	−25.19	20.57	0.29	3.91	−22.83	15.88	0.17	3.83
CCI	−400.8	454.5	0.79	118.3	−396.8	286.42	26.32	108.4	−343.94	272.93	17.36	108.6
RSI	0	100	49.46	23.95	9.34	99.3	54.80	15.59	9.06	98.52	53.35	15.92
%K	0	100	48.97	33.32	0	100	61.29	31.54	0	100	58.19	31.1
%D	1.33	99.04	48.98	31.52	1.57	99.15	61.27	28.18	2.61	99.69	58.18	27.94
EMA_1	354.39	3304.6	1028.56	722.9	676.53	1842	1297.58	218.9	3512.1	6840.3	5701.21	674.1
EMA_{10}	378.61	3173.1	1029.60	721.3	713.78	1824	1296.32	216.4	3670.5	6725.8	5698.79	665.9
MACD	−147.2	230.14	−1.66	39.15	−77.2	25.81	1.94	14.98	−318.28	123.77	3.80	62.97
SAR	347.46	3335.2	1033.14	727.3	666.79	1813.6	1290.96	219	3460.7	6875.6	5676.33	689.3

Table 2 Technical indicators and trading strategies

Technical indicator	Formula CP_t—closing price, $t = 1, 2, \ldots, n$, LP_N/HP_N—lowest/highest price in the past N days	Trading strategy signals, S_T
EMA	$EMA_t = CP_t * k + EMA_{t-1} * (1-k), k = 2/(N+1)$	$\begin{cases} 1 & \text{if } EMA_{1,t} > EMA_{10,t} \\ -1 & \text{if } EMA_{1,t} < EMA_{10,t} \end{cases}$
MACD	$MACD_t = EMA_{12,t} - EMA_{26,t}$	$\begin{cases} 1 & \text{if } MACD_t > EMA_{9,t} \\ -1 & \text{if } MACD_t < EMA_{9,t} \end{cases}$
RSI	$RSI_t = 100 - \frac{100}{1+RS_t}$ $RS_t = \frac{\sum_{n=t}^{t-d} \max(0, CP_n - CP_{n-1})}{\sum_{n=t}^{t-d} |\min(0, CP_n - CP_{n-1})|}$	$\begin{cases} 1 & \text{if } RSI_{t-1} \geq 30 \text{ and } RSI_t \geq 30 \\ -1 & \text{if } RSI_{t-1} \leq 70 \text{ and } RSI_t \leq 70 \end{cases}$
CCI	$CCI_t = (M_t - SM_t)/0.0015 D_t$ $M_t = HP_t + CP_t + LP_t$ $SM_t = \sum_{i=t-m+i}^{t} M_i/m$ $D_t = \sum_{i=t-m+i}^{t} |M_i - SM_t|/m$	$\begin{cases} 1 & \text{if } CCI_t > 100 \text{ or } CCI_t > -100 \\ -1 & \text{if } CCI_t < 100 \text{ or } CCI_t < -100 \end{cases}$
SO	$\%K_t = 100((CP_t - LP_{14})/(HP_{14} - LP_{14}))$ $\%D = \frac{1}{t} \sum_{t=1}^{3} \%K_t$	$\begin{cases} 1 & \text{if } \%D < 0.2 \text{ and } \%K_t > \%D \\ -1 & \text{if } \%D > 0.8 \text{and } \%K_t < \%D \end{cases}$
SAR	$SAR_{t+1} = SAR_t + \alpha(EP - SAR_t)$ EP is the extreme point α is the acceleration factor	$\begin{cases} 1 & \text{if } CP_t > SAR_t \\ -1 & \text{if } CP_t < SAR_t \end{cases}$
ROC	$ROC_t = 100((CP_t - CP_{t-n})/CP_{t-n})$	$\begin{cases} 1 & \text{if } ROC_t > 0 \\ -1 & \text{if } ROC_t < 0 \end{cases}$

an indicator which detects stock price trend direction and determines entry and exit points). Descriptive statistics for the selected indicators based on the available data sets were calculated, and are shown in Table 1.

The detailed procedure for calculating these indicators and the rules for generating trading signals are given in Table 2.

The first step is to provide an initial matrix for the criterion pairwise comparisons. The risk and return criteria are evaluated based on standard economic theory assumptions that investors are commonly averse to risk (Levy 2006; Lo 2007). Since the aim of this paper is to improve the precision of the prediction model, the third criterion is evaluated as the most significant one. For our calculations, we used a 4-year trading cycle sub-sample period starting from the beginning of 2009 and lasting until the end of 2012.

The eigenvector, the Relative Value Vector, is calculated by the methods described in Sect. 2.3. as RVV = (0.082, 0.236, 0.682)^T (Table 3). These three numbers correspond,

Table 3 Pairwise criteria comparison matrix

	Return	Risk	HR	RVV
Return	1	1/4	1/6	0.082
Risk	4	1	1/4	0.236
HR	6	4	1	0.682

$\lambda_{max} = 3.1078$, consistency ratio = 0.09297

respectively, to the relative values of each criterion of return, risk and accuracy. The result 0.682 means that the model values accuracy most of all; 0.236 shows that risk is valued less; and 0.082 shows that the model values return the least. The CR value is 0.09297, which is less than the value of the critical limit 0.1, and thus the model is consistent in its choices. Previously, the terms technical trading strategies and technical indicators were used. To simplify the notation for further

Table 4 Option performance matrix^T × RVV^T = feature weights $(\theta)^T$

	EMA	MACD	RSI	CCI	SO	ROC	SAR
BELEX15							
Return	0.159	0.239	0.056	0.134	0.020	0.093	0.299
Risk	0.021	0.100	0.274	0.122	0.283	0.100	0.100
HR	0.202	0.199	0.031	0.145	0.024	0.199	0.199
θ	0.156	0.179	0.090	0.139	0.085	0.167	0.184
S&P500							
Return	0.039	0.132	0.350	0.334	0.083	0.025	0.037
Risk	0.047	0.046	0.451	0.047	0.315	0.047	0.047
HR	0.209	0.209	0.022	0.234	0.023	0.135	0.168
θ	0.157	0.164	0.150	0.198	0.097	0.105	0.129
FTSE100							
Return	0.028	0.140	0.466	0.140	0.140	0.043	0.043
Risk	0.045	0.045	0.441	0.045	0.333	0.045	0.045
HR	0.189	0.193	0.021	0.193	0.022	0.193	0.189
θ	0.142	0.153	0.157	0.153	0.105	0.146	0.143

calculations, these two terms will be considered synonyms, although in fact a choice of technical indicators is made.

In the next step using three pairwise comparisons matrices, we compare the selected input features in terms of the gross return, systematic risk and prediction accuracy. Table 4 presents the summarized option performance matrix for the observed technical indicators.

Based on the final calculation, we obtained a decreasing order of feature weights and Fig. 3 shows a final summary of feature relevance.

After obtaining the feature weights, we performed feature selection by analyzing the results shown in Fig. 3, as described in Sect. 3.2. It can be noticed from Fig. 3 that the indicator weights eventually significantly decrease after the second ranked indicator for the Belex15 and S&P500 index, and that for the FTSE100 the decrease is significant after the third indicator. As a result, we selected the first two ranked indicators as input features for the prediction model for the Belex15 and S&P500, and the first two rescaled weights to be incorporated into the LS-SVM kernel. For the FTSE100,
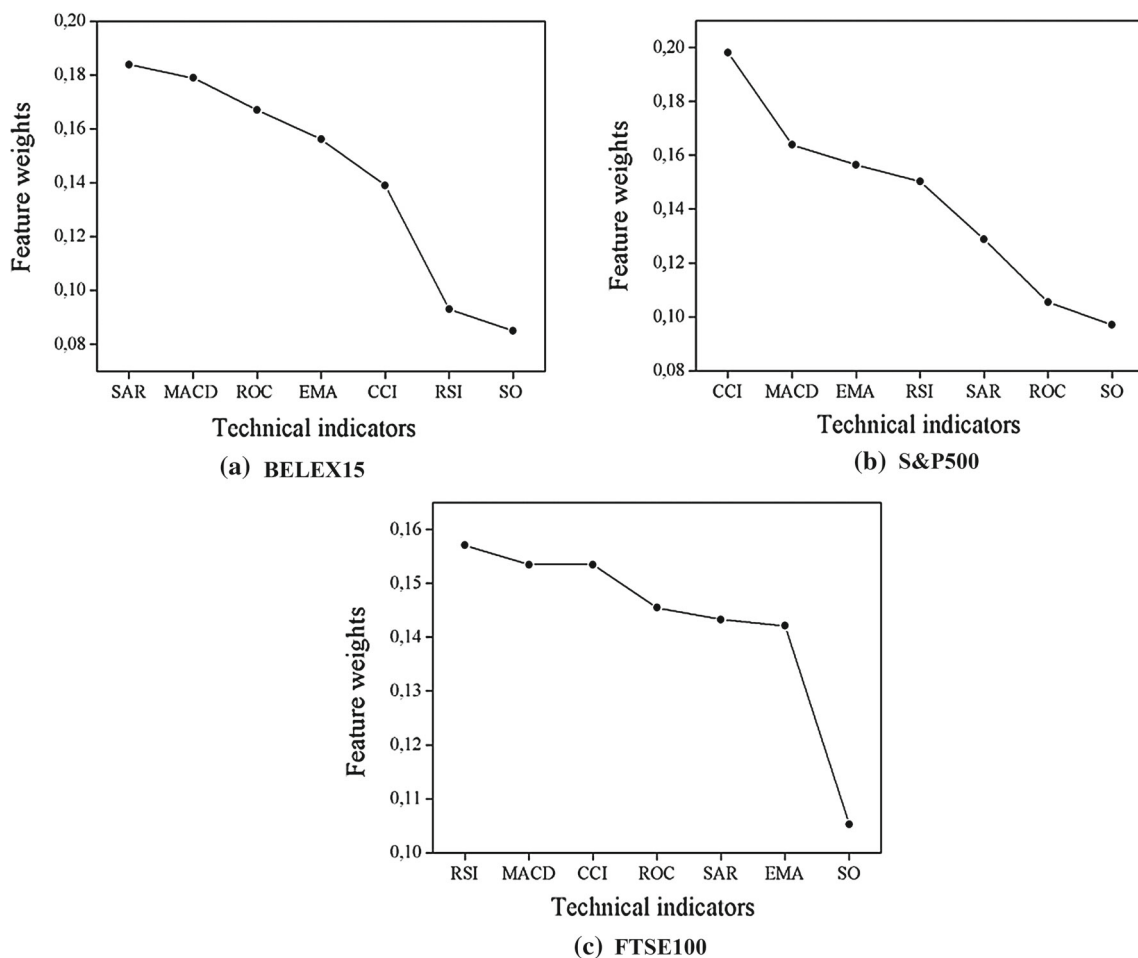

Fig. 3 Decreasing order of the obtained feature weights

Table 5 Comparison of the feature ranking and selection approaches

Selection approach	Selected features	Number of features
BELEX15		
AHP	SAR, MACD	2
MI	$\%K$, $\%D$	2
RF	CCI, RSI,$\%K$, $\%D$	4
LDC	CCI, RSI, $\%K$, $\%D$, MACD	5
S&P500		
AHP	CCI, MACD	2
MI	$\%K$, $\%D$, EMA_1, EMA_{10}, SAR	5
RF	ROC, CCI, RSI, $\%K$, $\%D$	5
LDC	$\%K$, $\%D$, MACD, SAR	4
FTSE100		
AHP	RSI, MACD CCI	3
MI	CCI, $\%K$, $\%D$	3
RF	ROC, CCI, RSI, $\%K$, $\%D$, MACD	6
LDC	CCI, $\%K$, $\%D$	3

Table 6 The performance comparison of the individual prediction models

Prediction model	BELEX15	S&P500	FTSE100
LS-SVM	53.57	53.97	51.58
MI-LS-SVM	52.38	53.18	50.79
RF-LS-SVM	52.38	51.19	51.19
LDC-LS-SVM	53.17	51.98	51.19
AHP-WK-LS-SVM*	61.11	57.14	57.54
SVM	53.98	53.57	52.78
AHP-WK-SVM	57.54	54.76	55.56
ANN	53.97	58.33	54.37
RF	50.00	51.19	53.57
RW*	50.00	48.41	50.00

* Statistically different models Nemenyi post-hoc $p \leq 0.05$

we selected the first three ranked indicators. To form the LS-SVM models, LS-SVMlab (Brabanter et al. 2011) was used.

4.2 Experimental evaluation

To assess the increase in the accuracy of the proposed model and its contribution to forecasting research, the accuracy of the model is compared with the results of other classification algorithms: Random Forest (RF) (Breiman 2001), Linear SVM (Chang and Lin 2011) and artificial neural networks (ANN). For the Random Forest, we used 1000 trees and a number of features for each split were set to the square root of the features dimensionality. For the Linear SVM soft, margin parameter C is fixed to $C = 1$. For the ANN, we used two hidden layers with 100 neurons. In addition, we compared the proposed feature selection strategy with several feature selection approaches: mutual information (MI) with forward–backward selection (Gómez-Verdejo et al. 2009), random forest (RF) for feature selection (Genuer 2010), and a linear discriminant classifier (LDC) with sequential forward selection (He et al. 2013).

First, Table 5 presents the comparison of selected features according to different feature selection methods.

From Table 5, it can be seen that according to the feature selection approach, both the number and the set of the selected input features vary. Thus, for testing purposes, we built 10 different models, denoted with the abbreviations of one of the above-mentioned feature selection approaches and used prediction models. Accordingly, the MI-LS-SVM is a

model trained with features selected by MI. The RF-LS-SVM is an LS-SVM model trained with features selected by Random Forest. The LDC-LS-SVM is an LS-SVM model trained with features selected with a linear discriminant classifier. The AHP-WK-LS-SVM implements the proposed approach for feature selection and weighted kernel. The AHP-WK-SVM model incorporates weights obtained from AHP into the SVM kernel. The Random Walk (RW) model uses the current value to predict the future value, assuming that the value in the following period (y_{t+1}) will be equal to the current value (y_t). The hit rate values in percentages for observed data sets according to the initial split of approximately 90 % for training and 10 % for test data are shown in Table 6. All of the benchmark prediction models used the same experimental setups across the data series, that is, the same training and test sets for each experimental data set. All of the models are built within the Matlab Tollboxes by using additional libraries where necessary LS-SVMlab (Brabanter et al. 2011), Lib-SVM (Chang and Lin 2011) and MILCA-MI (Kraskov et al. 2004).

From Table 6 it can be observed that in terms of the hit rate, the proposed AHP-WK-LS-SVM prediction model significantly outperforms all the benchmark models for the BELEX15 and FTSE100 data sets. In comparison with ANN, the AHP-WK-LS-SVM obtained hit rate is slightly lower for the S&P500 index, around 1 % less, but significantly higher for the FTSE100 and Belex15, more than 3 and 7 % respectively. Besides the AHP-WK-LS-SVM model, we tried to incorporate weights obtained from AHP into the SVM kernel. From Table 6, it can also be noted that the AHP-WK-SVM model significantly improves the SVM model, by 3 % for the BELEX15 and FTSE100, and more than 1 % for S&P500.

For comparing multiple models on multiple data sets, a two-stage procedure is recommended (Dešmar 2006). First, applying Friedman's test, to test whether the compared models have significant general differences in the

Table 7 Prediction performance depending on the number of training instances (given in %)

Prediction model	75			50			40			30		
	BELEX15	S&P500	FTSE100	BELEX15	S&P500	FTSE100	BELEX15	S&P500	FTSE100	BELEX15	S&P500	FTSE100
LS-SVM	54.71	52.77	49.33	55.47	52.31	49.28	56.38	52.10	48.29	57.62	50.92	48.27
MI-LS-SVM	55.49	52.96	49.33	53.72	51.92	49.47	55.66	51.03	48.92	57.20	50.63	49.21
RF-LS-SVM	55.49	53.16	49.14	54.79	51.82	49.01	55.75	50.69	48.85	56.78	50.63	49.69
LDC-LS-SVM	55.88	51.78	49.53	54.59	51.92	49.47	55.91	51.76	48.92	57.62	51.99	48.87
AHP-WK-LS-SVM	59.61	55.34	53.33	57.63	55.07	51.67	57.62	55.39	50.83	58.81	54.79	51.05
SVM	56.08	51.38	48.38	57.24	51.92	48.05	57.21	50.62	47.19	58.11	50.07	47.65
AHP-WK-SVM	58.04	52.77	53.33	57.24	53.89	49.38	57.37	50.95	47.42	58.18	52.68	50.64
ANN	50.39	55.54	52.00	50.69	55.86	50.62	50.20	56.14	51.38	52.72	55.08	50.24
RF	50.00	51.19	53.57	54.31	53.84	49.71	54.21	52.36	49.76	57.13	51.48	49.72

performance, and if the null hypothesis is rejected, at the second stage applying some post-hoc test. Friedman's test is a nonparametric test which is designed to detect differences among two or more groups. Applying Friedman's test, a *p* value of 0.0057 is obtained. Thus, the null hypothesis is rejected at the 5 % significance level, which indicates statistically significant differences in the mean ranks among the compared models. For the post-hoc test, the Nemenyi test was used, which indicates no significant differences at the 0.05 significance level between the obtained prediction models, except between AHP-WK-LS-SVM and RW.

Finally, we compared the accuracy of the proposed prediction model with other benchmark classifiers, depending on the number of training instances, and the results are shown in Table 7.

Based on the results presented in Table 7 it can be seen that for all the splits and series, our proposed AHP-based features ranking and selection approach improved the LS-SVM and SVM prediction accuracy. For BELEX15 and FTSE100 data series, the proposed model had the highest hit rate among all the benchmark models, while for S&P500 the recorded hit rate was only slightly lower than that of the ANN and same trend was noted for all the splits.

5 Conclusion

One possible approach for improving stock market trend prediction is presented in this paper. The proposed methodology is based on the concept of AHP analysis for feature ranking and selection. In addition, we used a weighted kernel to increase the generalization performance of the LS-SVM prediction model, where the kernel is weighted based on the feature relevance obtained by the conducted AHP analysis. The influence of the weighted kernel and feature selection led to a significant increase in the prediction accuracy. In addition, the set of feature weights obtained by the proposed

approach can also independently be incorporated into other kernel-based learners, beside LS-SVMs.

The improvement in hit rates obtained on the test sets that contain data for 1 trading year can be considered a significant improvement, considering the fact that the stock market trend is predicted for the purpose of the optimization of investment strategies on the financial markets. Thus, percent increase in model precision can lead to a gain in terms of profit, since it results in greater return and a decrease in the risk involved in trading. Therefore, future improvements will focus on the study of criteria relevant to investors with different preferences regarding risk. Also, further work should include the formation of an ensemble model, where the outputs from several models would be combined into a final model by some aggregating scheme.

References

Arlot S, Celisse A (2010) A survey of cross-validation procedures for model selection. Stat Surv 4:40–79

Atsalakis SG, Valavanis PK (2009) Forecasting stock market short-term trends using a neuro-fuzzy based methodology. Expert Syst Appl 36(7):10696–10700. doi:10.1016/j.eswa.2009.02.043

Atsalakis SG, Valavanis PK (2009b) Surveying stock market forecasting techniques—Part II: Soft computing methods. Expert Syst Appl 36(3):5932–5941. doi:10.1016/j.eswa.2008.07.006

Barak S, Modarres M (2015) Developing an approach to evaluate stocks by forecasting effective features with data mining methods. Expert Syst Appl 42(3):1325–1339. doi:10.1016/j.eswa.2014.09.026

De Brabanter K, Karsmakers P, Ojeda F, Alzate C, De Brabanter J, Pelckmans K, De Moor B, Vandewalle J, Suykens JAK (2011) LS-SVMlab toolbox user's guide version 1.8. http://www.esat.kuleuven.be/sista/lssvmlab/

Breiman L (2001) Random forests. Mach Learn 45(1):5–32

Coyle G (2004) Practical strategy: structured tools and techniques. Pearson, New York

Crone SF, Kourentzes N (2010) Feature selection for time series prediction—a combined filter and wrapper approach for neural networks. Neurocomputing 73(10–12):1923–1936. doi:10.1016/j.neucom.2010.01.017

Chai J, Du J, Lai KK, Lee YP (2015) A hybrid least square support vector machine model with parameters optimization for stock forecasting. Math Probl Eng (article ID 231394, 7 pages). doi:10.1155/2015/231394

Chang C-C, Lin C-J (2011) LIBSVM: a library for support vector machines. ACM Trans Intell Syst Technol 2(3):27:1–27:27

Dai W, Wu J-Y, Lu C-J (2012) Combining nonlinear independent component analysis and neural network for the prediction of Asian stock market indexes. Expert Syst Appl 39(4):4444–4452. doi:10.1016/j.eswa.2011.09.145

Dešmar J (2006) Statistical comparisons of classifiers over multiple data sets. J Mach Learn Res 7:1–30

Fama E (1970) Efficient capital markets: a review of theory and empirical work. J Finance 25:383–417

Fung GPC, Yu JX, Lam W (2002) News sensitive stock trend prediction. In: Advances in knowledge discovery and data mining. Springer, Berlin, pp 481–493. doi:10.1007/3-540-47887-6_48

Genuer R, Poggi J-M, Tuleau-Malot C (2010) Variable selection using random forests. Pattern Recognit Lett 31(14):2225–2236

Giveki D, Salimi H, Bahmanyar G, Khademian Y (2012) Automatic detection of diabetes diagnosis using feature weighted support vector machines based on mutual information and modified Cuckoo search. arXiv:1201.2173

Gómez-Verdejo V, Verleysen M, Fleury J (2009) Information-theoretic feature selection for functional data classification. Neurocomputing 72(16–18):3580–3589. doi:10.1016/j.neucom.2008.12.035

Guo B, Gunn SR, Damper RI (2008) Customizing kernel functions for SVM-based hyperspectral image classification. IEEE Trans Image Process 17(4):622–629. doi:10.1109/TIP.2008.918955

Guyon I, Elisseeff A (2003) An introduction to variable and feature selection. J Mach Learn Res 3:1157–1182

Hawawini G, Keim DB (1995) On the predictability of common stock returns: world-wide evidence. In: Handbooks in operations research and management science, vol 9. North-Holland, Amsterdam, pp 497–544

He Y, Fataliyev K, Wang L (2013) ICONIP 2013, Part II, LNCS 8227. In: Lee M et al (eds) Feature selection for stock market analysis. Springer, Berlin, pp 737–744

Huang W, Nakamori Y, Wang S-Y (2005) Forecasting stock market movement direction with support vector machine. Comput Oper Res 32(10):2513–2522. doi:10.1016/j.cor.2004.03.016

Kara Y, Boyacioglu MA, Baykan ÖK (2011) Predicting direction of stock price index movement using artificial neural networks and support vector machines: The sample of the Istanbul Stock Exchange. Expert Syst Appl 38(5):5311–5319. doi:10.1016/j.eswa.2010.10.027

Kaufman PJ (2003) A short course in technical trading. Wiley, New York

Kraskov A, Stogbauer H, Grassberger P (2004) Estimating mutual information. Phys Rev E 69(6):066138

Lahmiri S (2011) A comparison of PNN and SVM for stock market trend prediction using economic and technical information. Int J Comput Appl 29:24–30

Lee M-C (2009) Using support vector machine with a hybrid feature selection method to the stock trend prediction. Expert Syst Appl 36(8):10896–10904. doi:10.1016/j.eswa.2009.02.038

Levy H (2006) Stochastic dominance investment decision making under uncertainty, 2nd edn. Springer, New York

Liu D-R, Shih Y-Y (2005) Integrating AHP and data mining for product recommendation based on customer lifetime value. Inf Manag 42(3):387–400. doi:10.1016/j.im.2004.01.008

Liu D, Tian Z, Luo B, Xia J (2013) Feature ranking in intrusion detection by hybrid algorithm with support vector machine and analytic hierarchy process. Int J Digit Content Technol Appl (JDCTA) 7(7):1005–1013. doi:10.4156/jdcta.vol7.issue7.119

Lo AW (2007) The new palgrave: a dictionary of economics. In: Blume L, Durlauf S (eds) Efficient markets hypothesis. Palgrave Macmillan, Basingstoke

Marković I, Stojanović M, Božić M, Stanković J (2015) ICT innovations. In: 2014 proceedings of the advances in intelligent systems and computing. In: Bogdanova AM, Gjorgjevik D (eds) Stock market trend prediction based on the LS-SVM model update algorithm. Springer, New York, pp 105—114

Mittermayer MA (2004) Forecasting intraday stock price trends with text mining techniques. Proc Hawai Int Conf Syst Sci. doi:10.1109/HICSS.2004.1265201

McNelis PD (2005) Neural networks in finance: gaining predictive edge in the market. Elsevier, New York

Ni L-P, Ni ZW, Gao YZ (2011) Stock trend prediction based on fractal feature selection and support vector machine. Expert Syst Appl 38(5):5569–5576. doi:10.1016/j.eswa.2010.10.079

Omak EC, Polat K, Gunes S, Arslan A (2007) A new medical decision making system: least square support vector machine (LSSVM) with fuzzy weighting pre-processing. Expert Syst Appl 32(2):409–414. doi:10.1016/j.eswa.2005.12.001

Pauwels S, Inghelbrecht K, Heyman P, Marius D (2011) Technical trading rules in emerging stock markets. World Acad Sci Eng Technol 5:11–20

Rabin M (2000) Risk aversion and expected-utility theory: a calibration theorem. Econometrica 68:1281–1292

Saaty TL (1999) Monográfico: Problemas complejos de decisión. II. Basic theory of the analytic hierarchy process: how to make a decision. Rev R Acad Cienc Exact Fis Nat (Esp) 93:395–423

Stojanović BM, Božić MM, Stanković MM, Stajić ZP (2014) A methodology for training set instance selection using mutual information in time series prediction. Neurocomputing 141(2):236–245. doi:10.1016/j.neucom.2014.03.006

Suykens JAK, Van Gestel T, Brabanter J De, Moor B De, Vandewalle J (2002) Least squares support vector machines. World Scientific, Singapore

Wang Y, Choi I-C (2013) Market index and stock price direction prediction using machine learning techniques: an empirical study on the KOSPI and HIS. arXiv:1309.7119

Wang D, Zhang H (2013) Group AHP and K-means cluster for a new segmentation of brand customer. Int J Adv Comput Technol (IJACT) 5:213–221

Xing H, Ha M, Hu B, Tian D (2009) Linear feature-weighted support vector machine. Fuzzy Inf Eng 1(3):289–305. doi:10.1007/s12543-009-0022-0

Yao J, Zhao S, Fan L (2006) Enhanced support vector machine model for intrusion detection. Rough Sets Knowl Technol LNCS 4062:538–543. doi:10.1007/11795131_78

Yoo P D, Kim MH, Jan T (2005) Machine learning techniques and use of event information for stock market prediction: a survey and evaluation. In: Computational intelligence for modelling, control and automation and international conference on intelligent agents, web technologies and internet commerce, pp 835–841. doi:10.1109/CIMCA.2005.1631572

Yu L, Wang S, Lai KK (2005) WINE 2005, LNCS 3828 In: Deng X, Ye Y (eds) Mining stock market tendency using GA-based support vector machines. Springer, Berlin, pp 336–345

Yu L, Chen H, Wang S, Lai KK (2009) Evolving least squares support vector machines for stock market trend mining. IEEE Trans Evolut Comput 13(1):87–102. doi:10.1109/TEVC.2008.928176

Yuling L, Guo H, Hu J (2013) An SVM-based approach for stock market trend prediction. Neural Netw (IJCNN) (IEEE Press, New York) 1– 7. doi:10.1109/IJCNN.2013.6706743

Zhai Y, Hsu A, Halgamuge SK (2007) ISNN 2007, Part III, LNCS 4493. In: Liu D et al (eds) Combining news and technical indicators in daily stock price trends prediction. Springer, Berlin, pp 1087–1096. doi:10.1007/3-540-47887-6_48