

# Integrating High Volume Financial Datasets to Achieve Profitable and Interpretable Short Term Trading with the FTSE100 Index

Thomas Amorgianiotis, Konstantinos Theofilatos, Sovan Mitra, Efstratios Georgopoulos, Spiros Likothanassis

► **To cite this version:**

Thomas Amorgianiotis, Konstantinos Theofilatos, Sovan Mitra, Efstratios Georgopoulos, Spiros Likothanassis. Integrating High Volume Financial Datasets to Achieve Profitable and Interpretable Short Term Trading with the FTSE100 Index. Lazaros Iliadis; Ilias Maglogiannis; Harris Papadopoulos; Spyros Sioutas; Christos Makris. 10th IFIP International Conference on Artificial Intelligence Applications and Innovations (AIAI), Sep 2014, Rhodes, Greece. Springer, IFIP Advances in Information and Communication Technology, AICT-437, pp.340-349, 2014, Artificial Intelligence Applications and Innovations. <10.1007/978-3-662-44722-2\_36>. <hal-01391061>

**HAL Id: hal-01391061**

**<https://hal.inria.fr/hal-01391061>**

Submitted on 2 Nov 2016

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



# Integrating high volume financial datasets to achieve profitable and interpretable short term trading with the FTSE100 index.

Thomas Amorgianiotis<sup>1</sup>, Konstantinos Theofilatos<sup>1</sup>, Sovan Mitra<sup>2</sup>, Efstratios F. Georgopoulos<sup>3</sup>, and Spiros Likothanassis<sup>1</sup>.

<sup>1</sup>Department of Computer Engineering and Informatics, University of Patras, Greece  
{amorgianio, theofilk, likothan}@ceid.upatras.gr

<sup>2</sup>Glasgow Caledonian University  
Sovan.Mitra@gcu.ac.uk

<sup>3</sup>Technological Educational Institute of Peloponnese, 24100, Kalamata, Greece  
sfg@teikal.gr

**Abstract.** During the financial crisis of 2009 traditional models have failed to provide satisfactory results. Lately many techniques have been proposed to overcome the deficiencies of traditional models but most of them deal with the examined financial indices as they are cut off from the rest global market. However, many late studies are indicating that such dependencies exist. The enormous number of the potential financial time series which could be integrated to trade a single financial index enables the characterization of this problem as a “big data” problem and raises the need for advanced dimensionality reduction techniques which should additionally be interpretable in order to extract meaningful conclusions. In the present paper, ESVM-Fuzzy Inference Trader is introduced. This technique is based on the hybrid methodology ESVM Fuzzy Inference which combines genetic algorithms and some deterministic methods to extract interpretable fuzzy rules from SVM classification models.

The ESVM-Fuzzy Inference Trader was applied to the task of modeling and trading the FTSE100 index using a plethora of inputs including the closing prices of various European indexes. Its experimental results were compared with a state of the art hybrid technique which combines genetic algorithm with Multilayer Perceptron Neural Networks and indicated the superiority of ESVM-Fuzzy Inference Trader. Moreover, the proposed method extracted a compact set of fuzzy trading rules which among others can be utilized to describe the dependencies between other financial indices and FTSE100 index.

**Keywords:** Trading Strategies, Financial Forecasting, FTSE100, Genetic Algorithms, Support Vector Machines, Fuzzy logic.

## 1 Introduction

Modeling and trading financial indices remains nowadays a very challenging and demanding problem for the scientific community because of their complexity and

their dynamic and noisy nature. The FTSE 100 index trades futures contracts were utilized as a case study. A future contract is a contract between two parties and its cash settlement is determined by calculating the difference between the traded price and the closing price of the index on the expiration day of the contract. Most traditional methods and more complex machine learning ones have recently failed to capture the complexity and the nonlinearities that exist in financial time series during the latest crisis period. Certain disadvantages have been identified on traditional modelling and trading methods including the difficulties in tuning the parameters of the algorithm, the disability of linear methods to provide good prediction results, the overfitting problem and the fact that modelling and trading are most of the times considered as different problems. Lately, several machine learning methods have been proposed to solve these problems [1]. Despite the encouraging results of new hybrid methodologies their performance could be further enhanced.

Considering modelling and trading of the FTSE100 index, many researchers have been occupied with this problem. These methods are either based on simple traditional methods such as ARMA [2], or on machine learning techniques. The machine learning applications range from simple neural network techniques, such as Higher Order Neural Networks (HONN) [3] to more elaborate techniques such as the hybrid method combining Artificial Bee Colony Algorithm with Recurrent Neural Networks [4] and the hybrid methodologies which combine Genetic Algorithms with Support Vector Machines (SVM) [5].

Despite the promising methods of the aforementioned techniques most of them deal with the FTSE100 index as being independent and cut off from the global stock market. However, the reality is very different. As expected several articles indicated dependencies between the FTSE100 index and several other financial indices [6, 7] and these dependencies have not yet been studied thoroughly. Thus, in order to achieve optimal prediction and trading results using the FTSE100 index, several inputs from other financial indices should be utilized alongside with the traditional autoregressive and technical indicator inputs. The size of the universe of financial indices which could possibly be dependent with FTSE100 is so high that makes this problem relevant to the “big data” [8] topic which has gained the attention of the scientific community lately.

In the present paper, we have created an integrated dataset for modelling and daily trading with the FTSE100 index. This dataset, includes inputs and technical indicators from a variety of financial time-series including VIX, S&P500, DAXX, Euro Stoxx 50, and EURGBP exchange rate. The major problems for utilizing such integrated datasets for modelling and trading are: a) the difficulty to solve effectively the dimensionality reduction problem and b) the difficulty to achieve interpretability for extracting useful knowledge about the uncovered dependencies.

In the present study, we propose the ESVM Fuzzy Inference Trader to solve both these problems. This method is based on a Support Vector Machine methodology to achieve high performance predictions and it proposes an advanced technique to extract a meaningful compact set of fuzzy prediction rules. Moreover, a simple genetic algorithm finds the optimal feature subsets which should be used as inputs and optimizes the parameters of the overall modelling procedure. The final extracted inter-

pretable fuzzy rules presents extremely high statistical and trading performance surpassing another machine learning technique (the hybrid combination of genetic algorithms with Multilayer Perceptron Neural Networks) and enabled for the extraction of important conclusions about the dependencies of FTSE100 and other examined financial indices.

## 2 Dataset description

The dataset deployed in the present study incorporates information from 5 financial indexes and one exchange rate.

The basic financial index examined in the present paper is the FTSE100 index. The FTSE 100 index is a share index of the 100 companies listed in the London Stock Exchange having highest market capitalization. It is traded by futures contracts that are cash settled upon maturity of the contract with the value of the index fluctuating on a daily basis. The cash settlement of this index is simply determined by calculating the difference between the traded price and the closing price of the index on the expiration day of the contract.

The FTSE 100 daily time series is non-normal (Jarque-Bera statistics confirms this at the 99% confidence interval), containing slight skewness and high kurtosis. In the present study arithmetic returns were used and they are estimated using the following procedure: Given the price level  $P_1, P_2, \dots, P_t$ , the arithmetic return at time  $t$  is formed by:

$$R_t = \frac{P_t - P_{t-1}}{P_{t-1}} \quad (1)$$

Forecasting and trading the FTSE100 was attempted using an extended universe of inputs (50 inputs) containing both autoregressive inputs and technical indicators, such as moving averages of the FTSE 100 index, the VIX index, the S&P500 index, the DAX index, the Euro Stoxx 50 index and the Euro-GBP exchange rate. Specifically the technical indicators which were used by our model are for the FTSE 100 index the growth rate of the previous day, the maximum reduction rate of previous day, the seven previous days, the 5,15,30,50 and 252 days moving averages, the yesterday volume of transactions, the 5,15,30,50 and 252 days moving averages of volume of transactions, the highest close price for the past 30&252 days, the lowest close price for the past 30 and 252 days. The inputs computed for the VIX index, the S&P500 index, the DAX index, the Euro Stoxx 50 index and the Euro-GBP exchange rate are the 5,15,30,50 and 252 days moving average of their closing price.

VIX is a measure of the implied volatility of S&P 500 index options and it is used as an advanced volatility indicator which is able to measure the risk of the market. The S&P 500 consists of the 500 large companies having common stock listed on the NYSE or NASDAQ. The DAX index consists of the 30 major German companies trading on the Frankfurt stock Exchange. The Euro Stoxx 50 is made up of fifty of the largest and most liquid stocks of Eurozone. It was first introduced on February 26, 1988. Historical data of the five examined indexes were downloaded from the finance yahoo site (<http://finance.yahoo.com/>) and the Euro-GBP exchange rate was down-

loaded from the global-view site (<http://www.global-view.com/forex-trading-tools/forex-history/>).

Our dataset consists of a three-year in-sample dataset (02/01/2009 - 31/12/2011) and a two-year out-of-sample dataset (02/01/2012 - 31/12/2013). Specifically the in-sample dataset consists of 757 rows and 50 columns and the out-of-sample dataset consists of 502 rows and 50 columns. Each row represents a day and each column the 49 input features and the output. This dataset includes inputs and technical indicators from a variety of financial time-series including VIX, S&P500, DAXX, Euro Stoxx 50, and EUR-GBP exchange rate. The ESVM fuzzy inference trader tries to find the dependencies between the above financial indices and the FTSE100. FTSE100 index can also be dependent on an enormous number of indices. From this point of view, the problem of modeling and forecasting the FTSE100 can be seen as a big data problem. The ESVM fuzzy inference trader can find these dependencies and eliminate the financial time series that seem to have no dependencies with the FTSE100 index and this is how we deal with the big data problem. This is done by the feature selection technique of the proposed evolutionary algorithm which is explained at the next section. Due to the market's non stationary nature and its continuing, we deployed a sliding window approach. In particular, the out of sample dataset (testing dataset) was divided in three-month sub-datasets and in order to forecast and trade each individual sub-dataset the algorithms were trained using as in-sample dataset (training dataset) the past three years.

### 3 Method description

#### 3.1 GAMLP methodology

The first machine learning methodology which was applied in the present study is a hybrid combination of Genetic Algorithms and Multi-Layer Perceptron Neural Networks. It has been successfully applied in forecasting and trading the DJIE index [9] and the FTSE100 index [10]. This technique deploys an adaptive evolutionary algorithm to optimize the inputs, the structure and the parameters of MLP NNs. At this work three layered MLP NNs were utilized. The first is the input layer, the second is the hidden layer and the third is the output layer. The fitness function which was utilized by the evolutionary algorithm is described below:

$$Fitness = annualized\ return - MSE - 0.001 * \#selected\_inputs \quad (2)$$

where  $\#selected\_inputs$  is the number of inputs selected by every iteration of the evolutionary algorithm.

This fitness function enables the extraction of profitable trading strategies while at the same time selecting only the most compact feature subset to be used as inputs.

The deployed evolutionary algorithm utilized an adaptive mutation rate operator which decreases the mutation probability as the evolutionary process progresses. Moreover, when stagnation effect is diagnosed the mutation probability is increased to escape local optimal solutions. The adaptive mutation rate, enables the overall al-

gorithm to present advanced exploration properties in its initial steps and gradually reinforcing its exploitation properties.

### 3.2 ESVM Fuzzy Inference Trader

Support Vector Machines (SVM) is a method which belongs to the category of supervised learning models and it is nowadays considered as state-of-the-art in the field of machine learning. SVM attempts to locate the optimal hyperplane which can classify the points at categories and presents the largest distance between the points of different categories. SVMs do not require prior knowledge of the problem and they are general purpose classifiers. Non linear SVMs use a kernel function to project the input space to a space of higher dimensions in hope of transforming non-linear separable problems to linear separable ones. Through the years, there have been proposed several kernel functions. At this work we used the radial basis function because it is the most commonly used kernel function in non-linear classification problems [11]. The main limitations of the SVM models are their black-box nature which restricts its interpretability, and the dependence of their performance to the selection of their inputs and parameters.

To overcome the disadvantages of the SVMs we combined them with a Genetic algorithm and a fuzzy rule based system. The genetic algorithm is used for the feature selection problem in an embedded framework and for the optimization of the SVM's parameters. Fuzzy logic is more close to the linguistic language and domain experts can interpret fuzzy rule-based classifiers more easily. Fuzzy logic has the inherent advantage of hiding the imprecise knowledge through fuzziness. Their main disadvantage is overfitting when the dataset includes non linear data.

The ESVM Fuzzy Inference Trader approach deploys an advanced technique to extract interpretable fuzzy rules from SVM classification models.

The problem of extracting fuzzy rules from trained classification SVMs models has been of interest to many researchers through the last decade. Our model is based at the work of [12]. The proposed methodology initially deploys the technique which was proposed by [13]. This is an algorithm to implicitly transform the SVM classification model to a set of Support Vector Fuzzy Inference (SVFI) rules which are proved to be equivalent to the initial classification model.

SVFI rules are in the following general form:

Rule k: if  $P_1^k$  **and**  $P_2^k$  **and**...  $P_N^k$  then  $c_k$

where  $P_i^k, i = 1, \dots, N$  are fuzzy clauses, having the form where  $x_i$  is *CloseToSV*( $k, i$ ). These fuzzy clauses examine the membership of the  $i$ -th input value in the  $i$ -th fuzzy set of the  $k$ -th support vector. The support vectors are the training samples that are selected by the SVM algorithm to define the final classification hyperplane. A Gaussian function of the form  $\mu_i^k(x_i) = \exp(-\frac{1}{2}(\frac{x_i^k - x_i}{\sigma_k})^2)$ , estimates the membership function by quantifying the distance of the inputs component  $x_i$  from the value  $x_i^k$  of the  $i$ th component of  $SV_k$ . The parameters  $\sigma_k$  are real constant numbers ( $\sigma_k \in \mathbb{R}$ ).

Because SVFI rules are large and do not include linguistic clauses they are hard to be interpreted by domain experts. The ESVM Fuzzy Inference Trader applies a series of filtering and merging rules to obtain a more compact set of rules. In particular, fuzzy clauses with membership values less than a predefined threshold ( $\beta$ ) and even whole fuzzy rules whose total membership value is less than another predefined threshold ( $\delta$ ) are discarded. As a second step, in the produced more compact set of rules *CloseToSV(k, i)* clauses are replaced with linguistic clauses (low, medium, high) which are represented as Gaussian functions with centers placed in the minimum, median and maximum value of each feature.

To fine tune the whole procedure of training SVMs and extracting interpretable fuzzy rules from them, the ESVM Fuzzy Inference Trader algorithm deploys a simple Genetic Algorithm. The GA is used to select the optimal feature subset, and find the optimal values for the regularization parameter C of the SVM, the parameter gamma for the RBF Kernel function, the threshold  $\beta$  for discarding a fuzzy clause, the threshold  $\delta$  for discarding a fuzzy rule and the variations of the linguistic clauses' Gaussian function (low, medium, high).

The chromosome of the deployed GA consisted of input genes and parameter genes. The input genes take values 0 or 1 and force the classifier to use a specific feature as input if the feature value for this input is 1. The parameter genes are represented as double variables. The basic variation operators which were applied are the one point crossover and mutation operators. The crossover probability was set to 90% as this is the most important operator in GAs. For the mutation operator we used the binary mutation operator for the feature genes and the Gaussian mutation for the other genes as they are real valued genes. The binary mutation randomly alters a gene value from 0 to 1 and opposite. The Gaussian mutation operator adds a random number in a randomly selected gene. This random number is taken from the Gaussian distribution using as center the zero value and as width the interval of allowed values for this gene divided by 10.

The fitness function which guides the evolutionary process of ESVM Fuzzy Inference System is described below:

$$fitness = g1 * annualized\_return + g2 * geometric\_mean + g3 * inter\_sv\_ratio + g4 * sel\_feat\_ratio \quad (3)$$

where *annualized\_return* is the annualized return of the extracted trading strategy, geometric mean is the geometric mean of sensitivity and specificity metrics,

$$inter\_sv\_ratio = 1 - \frac{\#Interpretable\_Rules}{\#SVFI\_Rules} \quad \text{and} \quad sel\_feat\_ratio = 1 - \frac{\#Selected\_Features}{\#Total\_Features}$$

The parameters  $g1$ ,  $g2$ ,  $g3$  and  $g4$  were set to 2, 1, 0.5 and 0.25 respectively to reflect the importance of each goal. In particular the most important goal is to extract profitable trading strategies and the second most important goal is to achieve high classification performance. Secondary goals are the reduced number of fuzzy rules and selected inputs in order to increase the interpretability of the final set of fuzzy rules. The population of the genetic algorithm was set to 30 and the generations to 500.

The ESVM Fuzzy Inference Trading System also calculates a regression prediction for every data sample by calculating its distance from the classification hyperplane.

## 4 Experimental Results

In this section we present the results of the studied models in the problem of trading the FTSE100 index. The trading strategy for the examined models is simple and identical for both of them: go or stay long when the forecast return is above zero and go or stay short when the forecast return is below zero. Additionally, a confirmation filter was used which disables changing positions when the forecasted value is below the threshold 0.001.

Because of the stochastic nature of the proposed methodologies a simple run is not enough to measure their performance. This is the reason why ten runs were executed and the mean results are presented in the next table. When using the proposed trading strategy traders could utilize the execution which provided the highest in sample performance or use a more elaborate technique to combine the independent traders as proposed in [14].

**Table 1.** Trading Results

	<b>GA-MLP</b>	<b>ESVM Fuzzy Inference Trader</b>
<i>Annualized Return (excluding costs)</i>	15,41%	25,70%
<i>Positions Taken (annualized)</i>	54	40
<i>Transaction Costs</i>	5,40%	3,95%
<i>Annualized Return (including costs)</i>	10,01%	21,75%
<i>Annualized Volatility</i>	12,71%	9,75%
<i>Information Ratio (including costs)</i>	0,79	2,23
<i>Maximum Drawdown</i>	-12,88	-12,78%
<i>Correct Directional Change</i>	53,43%	59,34%

From Table 3 it is easily observed that both methodologies presented highly profitable trading results even when transaction costs were considered. This is mainly attributed to the extended universe of candidate inputs which was utilized and the sliding window approach which enables the model to be recalculated every three months. Furthermore, both deployed machine learning methods utilize an advanced feature selection mechanism which enables them to ignore redundant feature and extract more compact and simple prediction models.

The ESVM Fuzzy Inference Trader, clearly outperformed GAMLN approach in all examined metrics [15]. The main reason for this fact is the advanced generalization properties that the SVM provides to the overall methodology of ESVM Fuzzy Inference Trader. Moreover, the prediction model of ESVM Fuzzy Inference Trader could be further analyzed by studying the extracted interpretable fuzzy rules sets. The ten



most highly rated, in terms of aggregated weight, rules which came up during the sliding window runs of ESVM Fuzzy Inference Trader are provided in Table 2. From this table it is shown that traditional autoregressive inputs were not so significant in the prediction models which were extracted from the ESVM Fuzzy Inference Trader. Important roles seem to play as expected the different moving averages technical indicators. Moreover, other not traditional inputs such as past volumes of transactions, highest value of the indexes, lowest values of the indexes and external inputs such as VIX, DAX, S&P500 and EURGBP related inputs seem to be present in the final prediction models. With a closer examination of the extracted fuzzy rules low values of the moving averages of VIX past values seem to be an indicator for positive predictions. This is explained from the fact that low volatility in the market which is indicated by low values of the VIX index may have positive effects on the FTSE100 index. It is noteworthy that both DAX and S&P500 present negative effects on the FTSE100 outcome. In particular the extracted fuzzy rules predict positive values when negative values for the moving averages of these indexes are observed. EURGBP has similar action on the extracted fuzzy rules as when high values are observed for its moving averages the fuzzy rules decide to predict negative values.

**Table 2.** Ten most significant extracted fuzzy rules

a/a	Fuzzy Rule	Aggregated Weight - Significance
1	IF 15_days_moving_average_of_volume_of_transactions = Low & yearly_moving_average_of_volume_of_transactions = High & highest_close_price_for_the_past_30_days = High & lowest_close_price_for_the_past_252_days = Low & vix_index_moving_5_days_average = Low & vix_moving_15_days_average = Low & vix_moving_30_days_average = Low & dax_moving_252_days_average = Low THEN class = Positive Prediction	1.948
2	IF maximum_growth_rate_of_the_previous_day = Low & yesterday_volume_of_transactions = Low & 5_days_moving_average_of_volume_of_transactions = Low & 15_days_moving_average_of_volume_of_transactions = Low & highest_close_price_for_the_past_30_days = High THEN class = Negative Prediction	1.333
3	IF vix_index_moving_5_days_average = Low & eur_gbp_moving_252_days_average = High THEN class = Negative Prediction	1.207
4	IF maximum_growth_rate_of_the_previous_day = Low & moving_average_50_days_of_FTSE100 = High & dax_moving_30_days_average = Low & sp500_moving_30_days_average = Low & sp500_moving_50_days_average = Low THEN class = Positive Prediction	1.125
5	IF vix_index_moving_5_days_average = Low & vix_moving_15_days_average = Low & vix_moving_30_days_average = Low THEN class = Positive Prediction	1.087
6	IF maximum_growth_rate_of_the_previous_day = Low & yesterday_volume_of_transactions = Low & 15_days_moving_average_of_volume_of_transactions = Low & highest_close_price_for_the_past_30_days = High THEN class = Negative Prediction	1.077
7	IF maximum_growth_rate_of_the_previous_day = Low THEN class = Positive Prediction	1.008

8	IF maximum_growth_rate_of_the_previous_day = Low & vix_index_moving_5_days_average = High & st_50_moving_30_days_average = Low THEN class = Positive Prediction	1.000
9	IF maximum_growth_rate_of_the_previous_day = Low THEN class = Negative Prediction	0.986
10	IF maximum_growth_rate_of_the_previous_day = Low & seven_previous_day_return = Low & vix_index_moving_5_days_average = High & st_50_moving_30_days_average = Low THEN class = Positive Prediction	0.934

## 5 Conclusions

In the present paper, the problem of modelling and daily trading of the FTSE100 index was approached for the first time with as a big data problem. In specific an extended universe of 50 potential inputs were integrated to a single dataset. These inputs were related to simple autoregressive inputs, and technical indicators of the FTSE100 index and several other financial time-series.

The proposed novel machine learning algorithm for solving the aforementioned problem was the ESVM Fuzzy Inference Trader. It is a hybrid methodology which combines genetic algorithms, Support Vector Machines, Fuzzy systems and some deterministic techniques to extract a compact and interpretable set of linguistic fuzzy prediction rules. The extracted fuzzy rules presented extremely high statistical and trading performance and outperformed one other machine learning approach which was utilized even when transactions costs were taken into account. Further analysis of the extracted interpretable fuzzy rules revealed that traditional autoregressive inputs are not significant for trading the FTSE100 index. Moreover, other technical indicators, such as moving averages, of FTSE100 and of other examined time series are the most important inputs for forecasting and trading the FTSE100 on a daily bases. Considering the role of the other financial time series which were studied, both the moving averages of VIX, S&P500, DAX and EURGBP were found to be inversely related with the FTSE100 index.

The present study could be further extended using additional time series for generating even more potential inputs. The ESVM Fuzzy Inference system is selecting only the more relevant inputs and its performance is not affected by the number of the initial potential features used. For this reason, the incorporation of even more inputs will only affect positively the performance of the overall method and will additionally reveal even more dependencies of the FTSE100 with the rest of the global market. Moreover, The proposed algorithmic technique can be applied to extract intraday trading strategies which are even closer to the big data problem definition.

## References

- [1] Dunis, C., Likothanassis, S., Karathanasopoulos, A., Sermpinis, G., and Theofilatos, K.: Computational Intelligent Techniques for Trading and Investment, Taylor and Francis (Series: Advances in Experimental and Computable Economics), ISBN: 978-0-415-63680-3, (2013).

- [2] Chen, C.C., Tsay, W.J.: A Markov regime-switching ARMA approach for hedging stock indices. *Journal of Futures Markets*. 31(2), 165–191 (2011).
- [3] Sermpinis, G., Laws, J., Dunis, C.L.: Modelling and trading the realised volatility of the FTSE100 futures with higher order neural networks. *European Journal of Finance*, 19(3), 165-179 (2011).
- [4] Hsieh, T.J., Hsiao H.F., Yeh W.C.: Forecasting stock markets using wavelet transforms and recurrent neural networks: An integrated system based on artificial bee colony algorithm. *Applied Soft Computing*. 11(2), 2510–2525 (2011).
- [5] Dunis, C., Likothanassis, S., Karathanasopoulos, A., Sermpinis, G. and Theofilatos. K.: A hybrid genetic algorithm–support vector machine approach in the task of forecasting and trading. *Journal of Asset Management*, 14, 52–71 (2012).
- [6] Doman, M., and Doman, R.: Dependencies between Stock Markets During the Period Including Late-2000s Financial Crisis. *Procedia Economics and Finance* 1, 108-117 (2012).
- [7] Graham, M., Kiviaho, J., and Nikkinen, J.: Short term and long-term dependencies of the S&P500 index and commodity prices, 13(4), 583-592 (2013).
- [8] Boyd D., Crawford K.: CRITICAL QUESTIONS FOR BIG DATA, *Information, Communication & Society*, 15(5), 662-679(2012).
- [9] Theofilatos, K., Karathanasopoulos, A., Sermpinis, G., Amorgianiotis, T., Georgopoulos, E., Likothanassis, S.: Modelling and Trading the DJIA Financial Index Using Neural Networks Optimized with Adaptive Evolutionary Algorithms. *Engineering Applications of Neural Networks Communications in Computer and Information Science*, 311, 453-462 (2012).
- [10] Theofilatos, K., Amorgianiotis, T., Karathanasopoulos, A., Sermpinis, G., Georgopoulos, E., Likothanassis, S.: Advanced short-term forecasting and trading deploying neural networks optimized with an adaptive evolutionary algorithm. In *Computational Intelligence techniques for trading and investment*, Routledge, 133-145 (2014).
- [11] Theofilatos, K., Pylarinos, D., Likothanassis, S., Melidis, D., Siderakis, K., Thalassinakis, E., Mavroudi, S.: A Hybrid Support Vector Fuzzy Inference System for the Classification of Leakage Current Waveforms Portraying Discharges. *Electric Power Components and Systems*, 42(2), 180-189 (2014).
- [12] Papadimitriou, S., and Terzidis, K.: Efficient and interpretable fuzzy classifiers from data with support vector learning. *Intelligent Data Analysis*, 9(6), 527–550(2005).
- [13] Chen, Y., and Wang, J.: Support vector learning for fuzzy rule based Classification. *IEEE Transactions on Fuzzy Systems*, 11(6), 716–728(2003).
- [14] Pendaraki, K., Spanoudakis, N.: Constructing Portfolios Using Argumentation Based Decision Making and Performance Persistence of Mutual Funds. In the 2nd International Symposium & 24th National Conference on Operational Research (EEEE 2013), 24-29 (2013).
- [15] Lindemann, A., Dunis, C. L. and Lisboa, P.: Probability distributions, trading strategies and leverage: an application of Gaussian mixture models. *J. Forecast.*, 23: 559–585(2004).