Hindawi
Applied Computational Intelligence and Soft Computing
Volume 2019, Article ID 8342461, 13 pages
https://doi.org/10.1155/2019/8342461

*Research Article*

# Algo-Trading Strategy for Intraweek Foreign Exchange Speculation Based on Random Forest and Probit Regression

**Younes Chihab**,[1] **Zineb Bousbaa,**[2] **Marouane Chihab,**[3] **Omar Bencharef,**[2] **and Soumia Ziti**[3]

[1]*Department of Computer Sciences, Superior School of Technology, Ibn Toufail University, Kenitra, Morocco*
[2]*Department of Computer Sciences, Faculty of Sciences and Techniques, Cadi Ayyad University, Marrakesh, Morocco*
[3]*Department of Computer Sciences, Faculty of Sciences, Med V University, Rabat, Morocco*

Correspondence should be addressed to Younes Chihab; y.chihab@uca.ac.ma

Received 18 March 2019; Revised 7 July 2019; Accepted 25 July 2019; Published 27 August 2019

Academic Editor: Miin-Shen Yang



In the Forex market, the price of the currencies increases and decreases rapidly based on many economic and political factors such as commercial balance, the growth index, the inflation rate, and the employment indicators. Having a good strategy to buy and sell can make a profit from the above changes. A successful strategy in Forex should take into consideration the relation between benefits and risks. In this work, we propose an intraweek foreign exchange speculation strategy for currency markets based on a combination of technical indicators. This system has a two-level decision and is composed of the Probit regression model and rules discovery using Random Forest. There are two minimum requirements for a trading strategy: a rule to enter the market and a rule to exit it. Our proposed system, to enter the currency market, should validate two conditions. First, it should validate Random Forest access rules over the following week while in the second one the predicted value of the next day using Probit should be positive. To exit the currency market just one negative warning from Probit or Random Forest is enough. This system was used to develop dynamic portfolio trading systems. The profitability of the model was examined for USD/(EUR, JYN, BRP) variation within the period from January 2014 to January 2016. The proposed system allows improving the prediction accuracy. This indicates a good prediction of the behavior market and it helps to identify the good times to enter it or to leave it.

## 1. Introduction

The strong fluctuations in the financial markets make the stock market a risky area for investors. The various investment strategies on the stock market appear as a tool to collect more stock market shares. According to William O'Neil [1], the right strategy is to look for companies having rapid earning increases. The gains rise faster than those existing in the actual market [2]. However, the authors in this area have identified several strategic paths for investing in the stock market, passive and active strategies. The latter ensures simplicity to implement but with a long-term return. In the short term, speculation is considered as a balancing factor that regulates supply and demand, while leading to price equilibrium consistent with the real state of the economy. This strategy is based on buying when the price is lower than the average value and selling when the price is higher. In Forex case (the foreign exchange market), this hypothesis is not adopted by statisticians who consider that the temporary series of Forex are random. In reality, the unpredictability of these temporary series gives the impression that the variations are random. Nevertheless, we can notice the cyclical character of the Forex market [3] by using a large-scale analysis. The study and the analysis of past trends can help sometimes to predict the market movements.

Currently, speculators are considered as the first source of information on the state of the markets. Speculation techniques are improved constantly [4]. Actually there are two widely used approaches: first, a classic approach based on the technical indicators adopted in econometrics. The second approach is based on data mining algorithms. Used investment strategies in Forex market are numerous: day trading, trading news, swing trading, trend trading, carry trading, chart level trading, and technical indicators trading based on data mining algorithms. Most investors in the Forex market have not acted manually, but they generally sought the computer algorithms to opt for a strategy that it is simple or complex.

In this paper, we propose a secured investment strategy in two stages:

Firstly, we have opted for a temporal approach without any prescriptive hypothesis on financial market trends. This approach is based only on the observations made on the evolution of exchange rates and various temporary indicators. We have used Random Forest algorithm [5] to predict medium-term evolution. However, we have been able to establish regularities in the motivations of buying or selling decisions based on our speculative model. In a second step, we chose the Probit model [6] applied to Forex technical indicators. The two systems are combined to form an intraweek investment strategy. Our proposal will offer traders to create a trading strategy from varied indicators. This strategy is a means of confidence to determine the market direction.

## 2. Related Work

Developments in the algorithm trading have improved recently. Many new areas of research have been introduced and, most significantly, the combination of algorithms with the financial studies has made it possible to conduct research that would have been impossible only a few years ago. Nowadays, the electronic financial market has particularly progressed and the majority of transactions are done electronically. The electronic financial market has obtained an additional interest as a new area of research specially using trading algorithms and markets forecasting methods.

In literature, traditional trading systems implement only one specific strategy [8], whereas algorithmic trading is a method where a computer makes a specific investment instead of a human. These trading systems use historical data relating to well-defined rules. As the computer processing is necessary for forecasting methods in financial market, there are many advantages as well as pitfalls of this technical approach to trading and forecasting. Many studies suggest that algorithmic approaches are superior in comparison with traditional approaches. Since global markets are continuously evolving and becoming more interactive, the forecasting of financial markets and trading activity will play a more crucial role. The methods and techniques used to manage foreign exchange are more complex than ever before. For this reason, the researchers think that algorithm trading approach can make an investment more efficient at a lower price thanks to a quicker simultaneous analysis of many factors; moreover, the algorithms act independently of the psychological state of human. Other researchers think that this trading approach can also be less effective for several reasons. Firstly, making a good trading strategy is itself very complex due to the non-stationary, noisy, and deterministically unpredictable nature of the financial markets. However, machines cannot replace human intelligence or human critical aspect. In addition, there could also be problems with the calibration of the trading system, which would give incorrect timing of the buying and selling of an asset.

In the last decades, the growth of global trading markets made the foreign exchange market the largest and most lucrative of the financial markets. This foreign exchange market is open 24 hours/24 where participants buy and sell currency pairs. This monetary market is characterized by high liquidity, large volume of trade, and continuous transactions. Forex market is a volatile market with great uncertainty. However, foreign exchange investors are exposed to currency risk, which can seriously jeopardize international trade flows [9, 10]. These investors need to be aware of the uncertainty of this market and the major impact on their investment decisions. However, accurate forecasting of exchange rates could reduce this uncertainty and would be beneficial for both international trade flows and investor profits. As a result, exchange rate movements and predictability have been studied extensively in recent decades [11].

Among the most modern practical methods for predicting currency movements, using fundamental and technical analysis is of paramount significance.

There are multiple studies that have applied fundamental analysis to forecast currency exchange rate. Among these researches, we can quote, e.g., Meese and Rogoff 1983, Meese 1986, Baillie and Selover 1987, McNown and Wallace 1989, Baillie and Pecchenino 1991, Sarantis 1994, and Cushman, 2000 [7, 12–17]. In this paper, we concentrate our study mainly on technical analysis using data mining algorithms and technical indicators to predict future exchange rate values.

Poole (1967) and Dooley and Schafer (1976) were the pioneers to describe technical analysis [18, 19]. Poole (1967) indicated that the application of trading rules generates important benefits. Thanks to these rules, we can fix the buy or sell orders if the exchange rate increases or falls compared to the percentage already fixed. Dooley and Schafer (1976) also applied seven different filter rules on nine currencies. They concluded that using simple trading strategy based on information about past exchange rate fluctuations generated significant returns.

Recently, numerous advanced techniques have been widely applied to predict exchange rate fluctuations [20]. These techniques exploit the technological progress of computer tools. This progress permitted to manage the big data and to study the complex, nonlinear, and dynamic characteristics of the financial markets. Recent studies show that 80% to 90% of professionals and individual investors rely on at least some form of technical analysis [21–23].

Several algorithms have been used to forecast currency exchange rates as Random Forest, genetic algorithms, SVM, Neural Network [24–26], Linear Discriminant Analysis, Linear Regression, KNN, and Naive Bayesian Classifier. To improve accuracy, Booth et al. used the Random Forest algorithm for classification. It shows that a regency-weighted ensemble of random forests produces superior results when analyzed on a large sample of stocks from the DAX in terms of both profitability and prediction accuracy compared with other ensemble techniques [27]. In addition, Sorensen et al. show that CART decision trees perform better than single-factor models based on the same variables in picking stock portfolios [28]. Another one of the most used algorithms to improve testing accuracy is Support Vector Machine (SVM) which was proposed by Boser et al. [29]. Wang et al. in [30] demonstrated that the K-means SVM (KMSVM) algorithm can speed up the response time of classifiers by decreasing the number of support vectors while maintaining a compatible accuracy to SVM.

Some researchers have focused on neural networks to train algorithms. According to Shaoo et al., cascaded functional link artificial neural networks (CFLANN) perform better in FX markets [31].

Actually, some researchers suggest applying ensemble methods in order to improve the regression and classification performance. In [32], He and Shen have used a bootstrap method based on neural networks to construct multiple learning models and combined the output of these models to predict currency exchange rates.

Nowadays, few counted studies use Random Forests and Probit regression to predict exchange rate. According to Lv and Zhang [33], the RF algorithm showed its performance against the SVM method and the multiple linear regression method to accurately predict the Chinese Yuan.

In this paper, we will use a Random Forest classification algorithm and Probit regression. We combined this two algorithms to forecast currency exchange rate.

## 3. Trading Algorithms Approaches

Trading strategy is an important financial method. It can be defined as a set of instructions to make a profit and generate a positive return on its investment. Some trading strategies are not always outright profitable as standalone strategies. Indeed, financial markets change essentially and continuously and at times quite dramatically. One of the consequences of this transiency is that trading strategies that may have worked well for some time may die, sometimes quite abruptly. There are many factors that affect the trading strategy results and thus no universal model can predict everything well for all problems or even be a single best trading method for all situations.

However, technological advances gave rise to new types of trading such as the trading strategies based on data mining and machine learning. This strategy is based on algorithm trading and shows how it can execute complex analyses in real time and take the required decisions based on the strategy defined without human intervention and send the trade for execution automatically from the computer to the exchange. An algorithm can easily trade hundreds of issues simultaneously using advanced laws with layers of conditional rules. Algorithm trading seeks to identify typically quite ephemeral signals or trends by analyzing large volumes of diverse types of data. Many of these trading signals are so faint that they cannot be traded on their own. Thus, one combines a lot of such signals with nontrivial weights to amplify and enhance the overall signal and it becomes tradable on its own and profitable after trading costs.

Currently, foreign exchange market is the biggest and most liquid market in the world. However, a trading strategy using algorithmic trading has become an absolute must for survival both for the buy and sell sides. Due to the chaotic, noisy, and nonstationary nature of the data, major trader has had to migrate to the use of automated algorithmic trading in order to stay competitive. To make profit from each strategy, the majority of the research has focused on daily, weekly, or even monthly prediction.

The last decade has witnessed an increase in the reliance on artificial intelligence prediction in different markets such as the Foreign Exchange (Forex) market, including Random Forest (RF), Linear Regression, Genetic algorithms (GA), Artificial Neural Networks (ANN), and Support Vector Machine (SVM) approach to analysis and forecasting of financial time series.

*3.1. Artificial Neural Networks Approaches.* Neural Networks are a key topic in several papers in order germane to trading systems. Matas et al. [34] developed an algorithm that is based on neural networks and GARCH models to make predictions while operating in a heteroskedastic time series environment. Enam [35] experimented with the predictability of ANN on weekly FX data and concluded that, among other issues, one of the most critical issues to encounter when introducing such models is the structure of the data. Kamruzzaman [36] compared different ANN models, feeding them with technical indicators based on past Forex data, and concluded that a Scaled Conjugate Gradient based model achieved closer prediction compared to the other algorithms. C. Evans et al. [37] have introduced a prediction and decision making model based on Artificial Neural Networks (ANN) and Genetic Algorithms that produce profitable intraday Forex transactions. They used dataset for their research comprising 70 weeks of past currency rates of the 3 most traded currency pairs: GBP\USD, EUR\GBP, and EUR\USD. Their tests have confirmed that the daily Forex currency rates time series are not randomly distributed. Their proposed system has improved the prediction rate.

According to Omer Berat Sezer et al. [38], the combination of genetic algorithms and neural networks together in a stock trading system permits to get better results when compared with Buy & Hold and other trading systems for a wide range of stocks even for relatively longer periods. Their proposed system was based on the use of optimized technical analysis feature parameter values as input features for neural network stock trading system. They used genetic algorithms to optimize RSI parameters for uptrend and downtrend market conditions.

*3.2. Genetic Algorithms Approaches.* Genetic algorithms (GA), developed by Holland [39], are a type of optimization algorithms and they are used to find the maximum or minimum of a function. They are used in several applications such as automatic programming and machine learning. They are also well suited to modeling phenomena in economics, ecology, the human immune system, population genetics, and social systems.

In work [40], Subramanian present an approach to autonomous agent design that utilizes the genetic algorithm and genetic programming in order to identify optimal trading strategies. Their proposed system, based on the competing agents, recorded an average sharp ratio between 0.33 and 0.85. In paper [41], Hirabayashi has introduced a forecasting optimization model based on a genetic algorithm. This model searches for buying and selling rules that return the highest profits. They used Technical Indexes such as Moving Average (MA) and Relative Strength Index (RSI) to automatically generate trading rules. These rules are composed of a combination of Technical Indexes and their parameters and are used as the GA's genotype. In [42], Fuente et al. have presented a work which also attempts to optimize the timing of an automated trader. They use GA to develop trading rules for short time periods, using Technical Indexes, such as RSI, as GA's chromosome. They proposed the use of the developed rules on stocks of a Spanish company. However, the description in that work was too preliminary to allow for a comparison with our system to be made. Reference [43] presented a work which attempts to generate the buying and selling signals against 30 companies' stocks in Germany (DAX30). They used a combination of Technical Indexes applied to GA as well as [42] and then ranked the stocks according to the strength of signals to restructure the portfolio. However, they do not devise any criterion of profit cashing and loss cutting. In [44], R. Lakshman et al. proposed a buy-and-sell strategy by using genetic algorithm to predict Stock Market Index. Their results suggest that genetic algorithms are promising models that yield the highest profit among other comparable models.

*3.3. Support Vector Machine Approaches.* The Support Vector Machine (SVM) was introduced by Vapnik and coworkers in 1992. The SVM has been applied in many different fields of business, science, and industry to classify and recognize patterns. Models based on the Support Vector Machine (SVM) are among the most widely used techniques to forecast the movement direction of financial time series. It is becoming, more and more, an active learning method.

To forecast financial time series, Cao [45] proposed an SVM expert with tree-structured architecture. The obtained simulations results showed that the SVM expert had achieved significant improvement in the generalization performance in comparison with the single SVM model. In [46], Kyoung-jae compared SVM with back-propagation neural networks to predict the stock price index. The experimental results showed that SVM provided a promising alternative to stock market prediction. In paper [47], Y. Yuan presented a polynomial smooth support vector machine method to forecast the movement direction of financial time series. Their paper was based on a theoretic macroeconomic analysis. The result indicated that the machine learning methods are very important for forecasting research and the polynomial smooth support vector machine is a very powerful model. B.M. Henrique et al. [48] used Support Vector Regression (SVR) to predict stock prices for large and small capitalizations and in three different markets, employing prices with both daily and up-to-the-minute frequencies. They compared their proposed model with the random walk model proposed by the EMH. The obtained results manifested that the SVR has a good predictive power, especially when using a strategy of updating the model periodically.

*3.4. Random Forest Approaches.* The Random Forest was proposed by Breiman (2001). It is shown to generate accurate predictive models. It automatically identifies the important predictors, which is helpful when the data consists of a lot of variables and we are facing difficulties in deciding which of the variables need to be included in the model.

Random Forest is one among the applications of machine learning. It helps to understand financial markets. Random Forest has been used in several works in order to beat the market by forecasting changes in price. Predicting price fluctuations is very difficult area because of its impact on the investment. To predict future direction of stock movement, Khaidem et al., in [49], used a Random Forest classifier to build a predictive system. They used various parameters such as accuracy, precision, recall, and specificity to evaluate system robustness.

In paper [50], Patel et al. used a hybrid system to predict future values of Stock Market Index. They proposed two stage fusion approach involving Support Vector Regression (SVR) in the first stage. The second stage of the fusion approach used Artificial Neural Network (ANN), Random Forest (RF), and SVR resulting in SVR–ANN, SVR–RF, and SVR–SVR fusion prediction models. To validate the experimental results, they selected two indices, namely, CNX Nifty and S&P Bombay Stock Exchange (BSE) Sensex from Indian stock markets. Then, they compared prediction performance of their hybrid models with ANN, RF, and SVR. In [51], M. Kumar et al. used five algorithms to predict the direction of S&P CNX Nifty Market index of the NSE. They tested linear discriminant analysis (LDA), Logit, Artificial Neural Network (ANN), Random Forest, and SVM. The experiment's results indicated that Random Forest method outperforms the other methods.

In literature, a number of different methods have been applied in order to predict stock market returns. Despite these studies, definition and implementation of a stock market strategy remains a difficult problem to resolve.

## 4. Main Concepts

Regression's algorithms are not limited to the linear or logistic regression, in fact there are many forms of regressions and each one has its own importance and specific conditions where they are appropriate to apply [52].

*4.1. Probit Model.* Our choice is the Probit model, which is a type of regression where the dependent variable can take only two values, for our case increased (1) or decreased (0) value of currencies [53].

If we consider the binary choice model, the unobservable response variable $y_{it}^{*}$ can be written:

$$y_{it}^{*} = x_{it}\beta + \alpha_i + \vartheta_{it} \tag{1}$$

with $i = 1, \ldots, N$

and $t = 1, \ldots, T$.

The observed binary variable $y_{it}$ is defined by

$$y_{it} = 1 \quad if \;\; y_i^* \geq 0 \tag{2}$$

$$y_{it} = 0, \quad otherwise \tag{3}$$

where $\alpha_i$ the unobserved effect

and $\nu_{it}$ the general error term.

In the Probit model case, the cumulative distribution is a standard normal:

$$\begin{aligned}\Pr\left(y_{it} = 1 \mid x_i, \alpha_i\right) &= \Pr\left(y_{it} = 1 \mid x_{it}, \alpha_i\right) \\ &= \varphi\left(x_{it}\beta + \alpha_i\right)\end{aligned} \tag{4}$$

The first equality states that $x_{it}$ is assumed to be strictly exogenous conditional on $\alpha_i$. Another standard assumption is that the outcomes $y_{it} = y_{i1},\ldots,y_{iT}$ are independent conditional on ($x_i$, $\alpha_i$). Thus, the density of $y_{it}$ conditional on (xi, $\alpha$i) can be derived:

$$f\left(y_i, \ldots, y_t \mid x_i, \alpha_i; \beta_i\right) = \prod_{t=1}^{T} f\left(y_t \mid x_{it}, \alpha_i; \beta\right), \tag{5}$$

where $f(y_t \mid x_t, \alpha; \beta) = \varphi(x_t\beta + \alpha)^{yt}[1 - \varphi(x_t\beta + \alpha)^{1-y_t}]$.

In an arbitrary effect panel framework, the unobserved effect $\alpha_i$ conditional on xi is expected to be normally distributed with $\alpha_i|x_i \sim N(0, \sigma_\alpha^2)$.

The assumption of independency of outcomes (i.e., crisis and tranquil periods) is limited and it can be relaxed by using the formula:

$$\Pr\left(y_{it} = 1 \mid x_i\right) = \Pr\left(y_{it} = 1 \mid x_{it}\right) = \varphi\left(x_{it}\beta_\alpha\right) \tag{6}$$

where $\beta_\alpha = \beta/(1 + \sigma_\alpha^2)^{1/2}$ is estimated from pooled Probit of $y_{it}$ on $x_{it}$ using Huber/White robust standard errors, meaning that the coefficients are average partial effects.

*4.2. Random Forest.* Random Forest (RF)[54] is a machine learning algorithm, which is best defined as a "combination of tree predictors in which each tree depends on the values of a random vector sampled independently and with the same distribution for all trees in the forest" [55].

Random Forest is used for classification and regression; random decision forests correct for classic decision trees the problem of overfitting [56]. The algorithm of Random Forest combines the concepts of random subspaces and bagging. The decision tree forest algorithm performs learning on multiple decision trees driven on slightly different subsets of data.

Random Forest algorithm:

```
Input: description language; sample S
Begin
    Initialize to the empty tree; the root is the
    current node
    Repeat
        Decide if the current node is terminal
            If the node is terminal then
                Assign a class
            Else
                Select a test and
                create the subtree
            End if
        Move to the next node unexplored if
        there is one
    Until you get a decision tree
End
```

Decision trees provide effective methods that work well in practice. Decision trees have the advantage of being comprehensible to any user (if the size of the produced tree is reasonable) and to having an immediate translation in terms of decision rules.

*4.3. Preassessment of Classification Algorithm.* In order to justify our choice of using Random Forest classifier, we evaluate the results obtained over 90 days by Support Vector Machine (SVM), Artificial Neural Networks (ANN), and Random Forest algorithm. For this reason, we have used 490 days (EUR/USD currency pairs time series) for training data. In this preassessment we used a sample dataset composed of EUR/USD currency pairs time series from January to December 2015. The following Figures 1, 2 and 3 show the systems accuracy and performance.

Utilizing Artificial Neural Networks, Support Vector Machine, and Random Forest algorithms to build an algo-trading model for intraweek foreign exchange speculation, we can clearly notify that Random Forest shows better results than ANN and SVM regarding our case study.

*4.4. Investment Strategy.* The portfolio theory appeared in 1952 by Harry Markowitz [57]. This theory aims at the rational constitution of a portfolio arbitrage between the gains and the risks. Indeed, the risk of a portfolio can be correctly measured by the variance of its profitability. The question is how to maximize the gains while minimizing the risks.

Due to the volatility of the Forex market, there are three types of portfolios: high-frequency traders, long-term investors, and corporations. This paper proposes a trading for high-frequency traders who speculate small intraweek price fluctuations [37, 58].

In high-frequency trading strategy, we can separate between many types of traders [59]:

(i) Scalpers: Forex Scalpers perform transactions of very short duration and take their gain very quickly, even when the market continues to evolve in the direction of their speculation. Scalping requires a sufficient investment fund.

(ii) Swingers: Swing Trade means opening a position at the beginning of an uptrend and closing that position simply at the end of this trend, by using the pronounced tendencies.

(iii) Technical analyzers: technical analysis is a method to predict the fluctuations in the financial markets to get benefits by finding patterns and relationships in historical financial time series. The technical analysis of trends aims to determine when it is better to enter the market. For this, we consider that the Forex

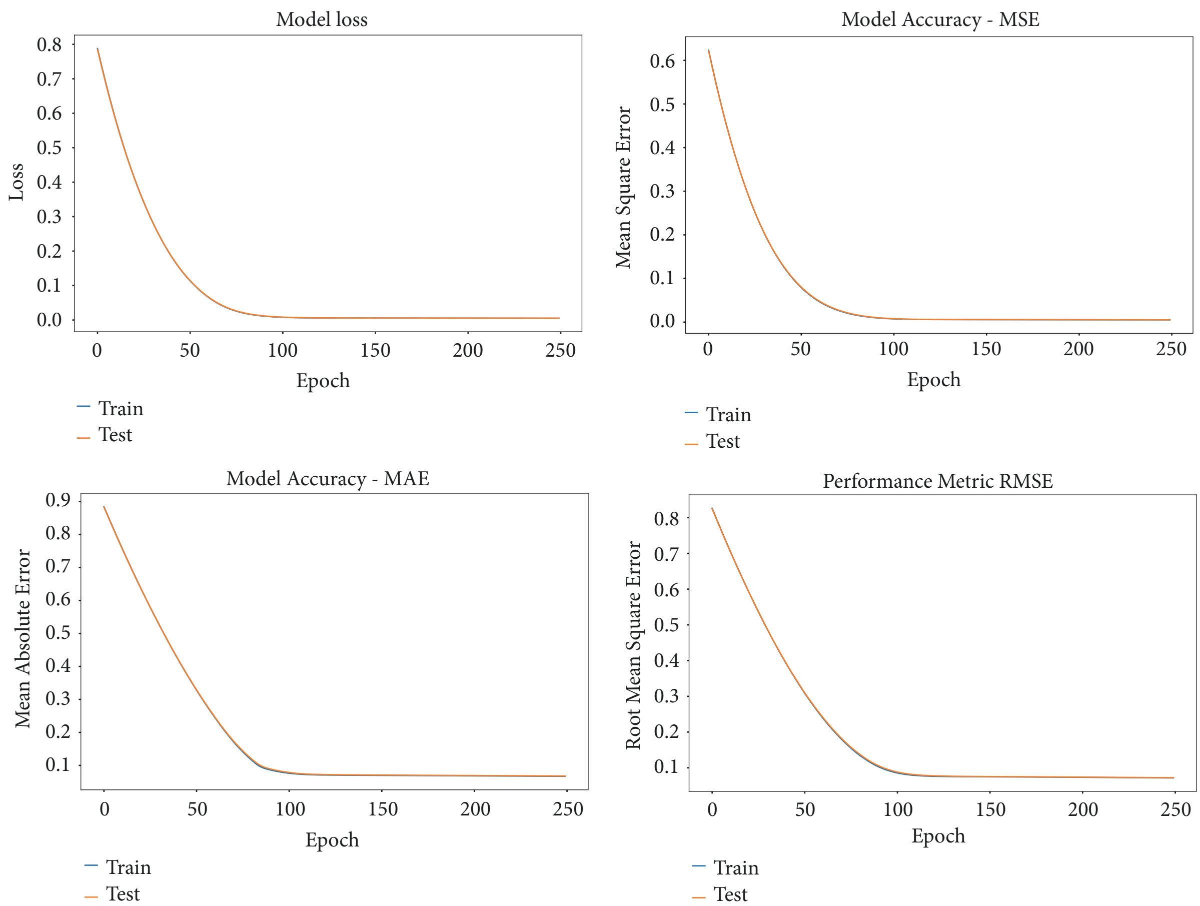


FIGURE 1: Model loss and MSE, MAE, and RMSE performance metrics of ANN (Test accuracy 0.519) using Keras package.

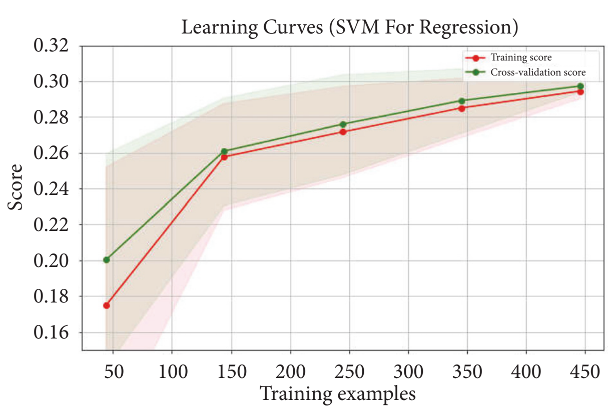


FIGURE 2: Cross validation score and MSE, MAE, and RMSE performance metrics of SVM classifier.

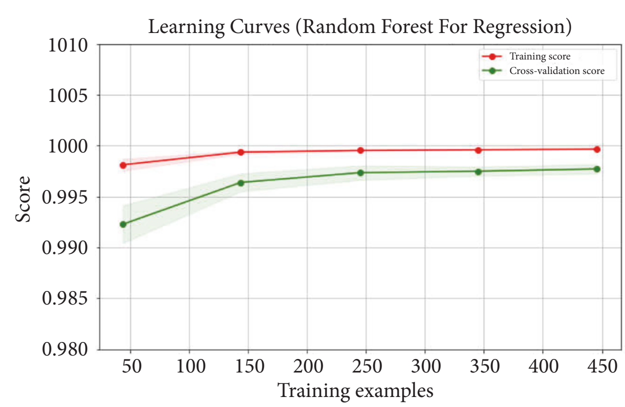


FIGURE 3: Cross validation score and MSE, MAE, and RMSE performance metrics of Random Forest classifier.

market follows a single direction over the long term. They use the monthly, weekly, and daily charts to accurately determine when a downturn may occur [60].

(iv) Fundamental analyzers: any market trend must be placed in its economic context and vice versa economic and political incidents directly influence market trends [61]. This analysis is based on the study of the economic and financial performance of a country in order to determine the real value of the market and the future evolution of its currency. This analysis is mainly based on economic information as well as important political events. Among indicators, we can quote the interest rate, the growth index, the inflation rate, the employment indicators, and the balance of trade. The observation and the evaluation of these indicators make it possible to know the state of the economy of a country. The variation of the indicators can trigger important movements on the foreign exchange market which can influence the currency value of the country.

With experience, traders combine those techniques to find a strategy to maximize their profit and they can include some unusual technique like the double-zero strategy.

This strategy is entirely based on the phenomena of psychological values. This means that the majority of traders tend to simplify stock prices by taking a position on round values. These values therefore very often become phenomena of support or resistance.

In our previous works we adopted Evans et al.'s [37] strategy, Figure 4; it starts by using historical data to train the decision system. Once trained, we used the system predictions to manage the buy, hold, and sell actions:

(1) If the system predicts a positive output, we buy.

(2) After buying, if the system predicts a positive evolution, we hold.

(3) After buying, if the system predicts a negative output, we sell.

We can add time limits and the process can be repeated.

## 5. Methodology and Results

*5.1. Data Sets.* We develop two datasets to train our system.

The first dataset utilized for this research encompasses 650 days of past currency rates of the EUR/USD currency pairs. An example of a currency pair is the EUR/USD. When trading the EUR/USD currency pair, US dollars are being sold to buy Euros; EUR is called the base currency and the USD is called the quote.

For each day, we use a time series composed of the 7 past days and the moving average of the last week and the last month.

The second dataset is composed of a collection of Technical Indicators (TI). A technical indicator is a value or a mathematical formula used to analyze stock market securities in order to predict price movements. The dataset contains the most used TI [62]:

(i) Relatively Strength Index (RSI)

(ii) Stochastic oscillator

(iii) The average directional index (ADX)

(iv) Commodity Channel Index (CCI)

(v) High-Low

(vi) DMI (Directional Movement Index)

(vii) Moving average

*5.2. Speculation.* We start by evaluating Random Forest regression results over a simple dataset composed of a collection of time series concerning Euro/Dollar variation over two years. Figure 5 shows prediction outputs versus real outputs and Table 1 is related to the performance of results.

A first evaluation of Random Forest outputs shows that the algorithm performances are good in general with an MSE of $10^{-5}$. The number of trees does not affect effectively the system performance while the best results were found for 500 trees. When analyzing the number of errors when Random Forest predicts an uptrend in the next day and in reality it was a downtrend and also if we take into consideration the degree of risks we found in Forex, it is very risky to consider regression results over time series as a unique input to decision making.

In a second test, we used the first dataset composed of time series to train the Probit model in order to speculate next day values. The second dataset composed of technical indicators is used to train Random Forest to predict global trend of the next 7 days; this choice comes after many experiments.

The suitability of an estimated binary model can be evaluated by counting the number of true and false observations and by counting the number of observations equaling 1 or 0, for which the model assigns a correct predicted classification by treating any estimated probability above 0.8 (or, below 0.2), as an assignment of a prediction of 1 (or, of 0) [63, 64].

For the Random Forest evaluation, we consider a week with positive evolution, if its number of days showing an uptrend is more than 4. Tables 2 and 3 show classification results and Figure 6 shows a plotting example of predicted output versus real output using Probit regression.

From Tables 1 and 2 we can notice that both classifiers can give us a clear idea about the market trends in different ways. Probit binary regression shows an accuracy of 76,4% and 80,7% for Random Forest. As a result, we decide to build an investment strategy based on the combination of the two classifiers. In the next section, we are going to explain how we combined these two algorithms outputs to propose an efficient investment strategy.

*5.3. Our Investment Strategy.* To create an efficient strategy, we need to identify a personal risk profile, a realistic availability of time and resources, and a level of expectation during a trade. These characteristics of self-reflection can be identified by evaluating different strategies.

In Forex investments, the leverage is any technique involving the use of borrowed funds in the purchase of an asset. Online brokers offer their clients leverage. This tool actually allows the speculation with more money than the capital available in order to make the benefits more interesting. Currency exchange rate fluctuations are often very low. Without leverage, it would be very difficult to make profits, even with important investment capital. In our experiment, we used a leverage of 1:100, which is the equivalence of an investment of 1000$, and we can tread up to 100000$.

The investment sequences are presented in Figure 7. First, we will invest only in weeks with positive trends and in each positive week we will check for next day positive trend to trade. This way, we can reduce the number of false investment rates. The strategy can be described as follows.

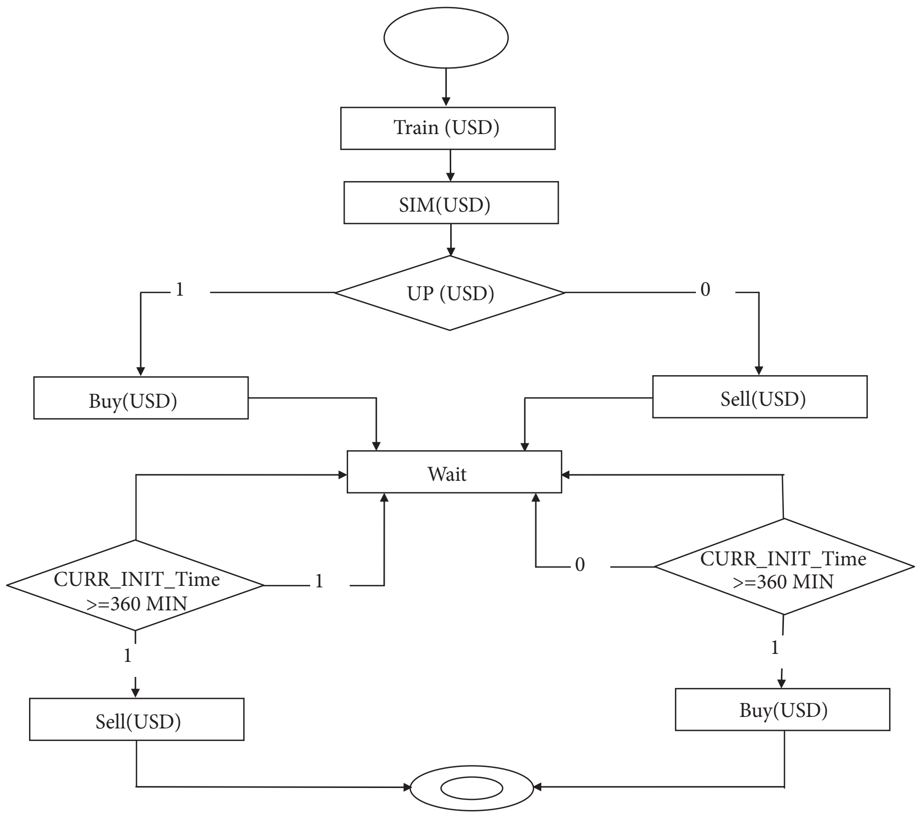


Figure 4: Investment strategy proposed in [7] for intraday foreign exchange.

Table 1: The effect of number of trees to MSE (Mean Square Error) and MAE over 70 days.

| Number of Trees | 100 | 300 | 500 | 700 |
|---|---|---|---|---|
| MSE | 1.257 $10^{-5}$ | 1.1102 $10^{-5}$ | 1.0922 $10^{-5}$ | 1.1129 $10^{-5}$ |
| MAE | 1,301 $10^{-4}$ | 1,189 $10^{-4}$ | 1,184 $10^{-4}$ | 1,191 $10^{-4}$ |
| % VAR explained | 99.78 | 99.81 | 99.81 | 99.80 |

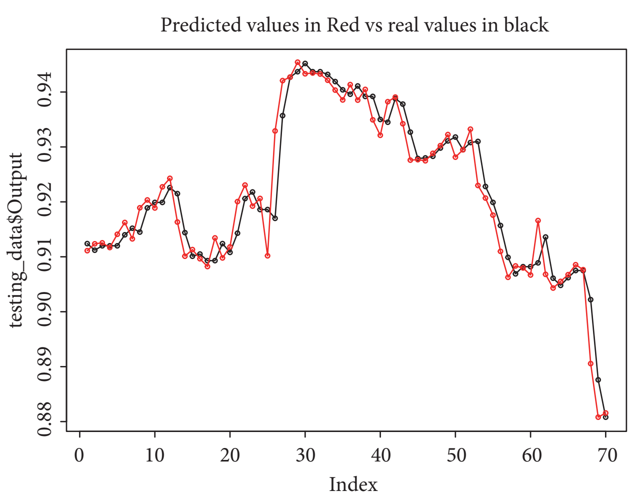


Figure 5: Predicted values versus real values (predicted values in red, real values in black); for Random Forest regression using: 500 tree and 8 variables tried for each split.

*Step 1.* For each currency we check for the week positive trend using the following rules:

(i) Based on technical indicators, we check the market status for one of these situations [65, 66]:

(a) The oversold situation: it is a situation where the price of an asset has fallen sharply to a level below its real value. It is a sign and probably the price should rebound.

(b) The overbought situation: it is a situation where the demand for a certain asset grows overly without justification. Probably it is an indication to sale.

(c) Bullish Divergence: it is the price trend is bearish while the indicator indicates the opposite. The indicator indicates an increase in the price of the asset, while the asset continues to fall.

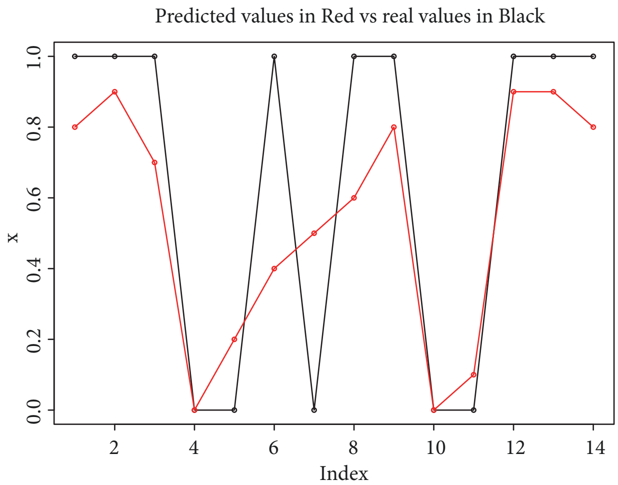


Figure 6: Predicted values versus real values (predicted values in red, real values in black); for Probit regression.

This can be analyzed by a possible reversal of the upward trend and by a future buy signal.

(d) Bearish Divergence: it is the opposite of the Bullish Divergence. This is analyzed by a reversal of the downtrend and by a sales signal to come.

(ii) If none of the situations mentioned above is detected, we take the Probit decision as final decision.

If there is a positive trend, go to step 2.

*Step 2.*

(i) For each day of the week

(ii) if there is a positive prediction buy and Go to step 3

(iii) End

*Step 3.*

(i) By considering a leverage of 1:100:

(ii) While (currency variations are between -10 pip and +5 pip and hold-time<420min) hold

(a) If the currency loses more than 10 pip then sell.

(b) If the currency wins more than 5 pip then sell

End

*5.4. Results and Discussion.* For a daily investment of 1000$ we fix

(i) a leverage of 1:100 over 17 weeks,

(ii) a maximum investment of 5000$+benefits over the 17 weeks,

(iii) a stop lose barrier of 20%, and

(iv) selling order for 10% of profit.

Table 2: Classification results for next day speculation using Probit mode for 109 days.

| Predicted value/ Real value | Positive Evolution | Negative Evolution |
|---|---|---|
| Positive Evolution | 39 | 13 |
| Negative Evolution | 8 | 49 |

Table 3: Classification of results for each week trend evolution results using Random Forest over 17 weeks.

| Predicted value/ Real value | Positive Evolution | Negative Evolution |
|---|---|---|
| Positive Evolution | 7 | 2 |
| Negative Evolution | 2 | 6 |

Tables 4, 5, and 6 show rates of return on three Forex pairs USD/Euro, USD/GBP, and USD/JAY over 17 weeks, using Probit binary regression, Random Forest, and the proposed system.

We tested our investments strategy over 17 weeks and two years data from January 2014 to January 2016 to train our algorithms. For final results we calculate the cumulated gain over 17 weeks.

The tables reveal that the proposed system demonstrates better results than Random Forest or Probit regression. To validate our model, we choose to evaluate its efficiency over three currency pairs and for the three pairs the proposed strategy shows the best results. In addition, the proposed system needs less investment to make more benefit.

The proposed model produces a quite promising profit with an average profit of 4.8% by week. We used only weeks with positive trends.

The proposed system allows us to reduce the number of daily investment without losing profit opportunity. We can notice a true positive rate, 78%. For the proposed system, it is 68% for Random Forest and 57% for Probit. The true positive measures the proportion of actual positives that are correctly identified.

We should clarify that the previous results influenced the currency pair global trend during the next six months. This means it is related to macroeconomic and political situation. For example, the results obtained by USD/GBP (only 2538$ as cumulative benefit) can explain the currency pair behavior. It was clear that we had a sideways trend. A sideways trend is a horizontal price movement. For the USD/EUR pair we can notice an uptrend for more than 8 weeks.

We also benefit from the fact that currency market is relatively stable and changes of more than even one percent are rare.

Simultaneously, an important issue that has not been mentioned so far is the trading cost. For each transaction, the currency market is a commission-free market. Instead of a commission, there is a pip spread. A pip spread is the difference between selling and buying price in the same moment. We consider a commission of 1 pip. From that fact and when using a leverage, we deduce that mostly some

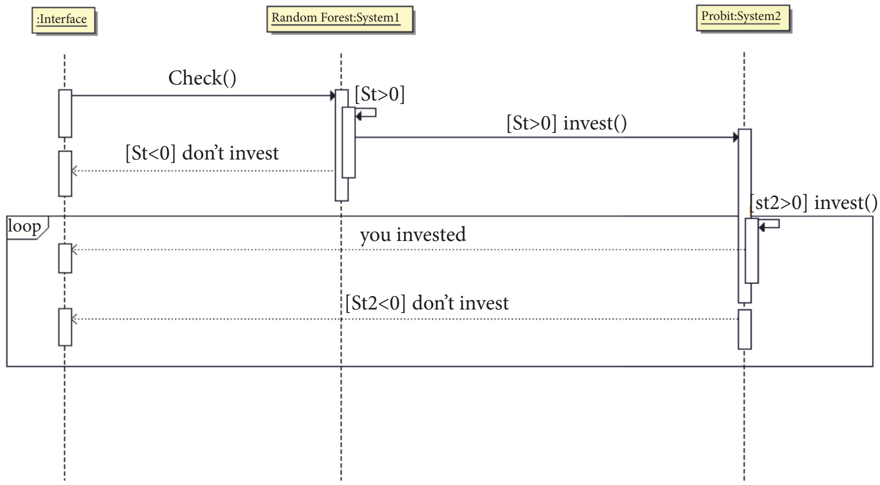


Figure 7: The sequences of the proposed investment strategy. With 'st': the Random Forest output and st2: the Probit model outputs.

Table 4: USD/EUR investment results using Random Forest, Probit model, and the proposed system over 17 weeks.

| | Number of daily investments/119 | Number of positive investments | Final results |
|---|---|---|---|
| Probit | 72 | 43 | -778$ |
| Random Forest | 74 | 51 | 2442$ |
| Proposed system | 63 | 49 | 4448$ |

Table 5: USD/GBP investment results using Random Forest, Probit model, and the proposed system over 17 weeks.

| | Number of daily investments/119 | Number of positive investments | Final results |
|---|---|---|---|
| Probit | 68 | 47 | 356$ |
| Random Forest | 71 | 58 | 2412$ |
| Proposed system | 54 | 42 | 2538$ |

Table 6: USD/JAY investment results using Random Forest, Probit model, and the proposed system over 17 weeks.

| | Number of daily investments/119 | Number of positive investments | Final results |
|---|---|---|---|
| Probit regression | 63 | 43 | 30$ |
| Random Forest | 61 | 46 | 1307$ |
| Proposed system | 51 | 41 | 3238$ |

currency pairs resulted in modest gains and some resulted in excessive losses; an excessive gain is really rare.

Generally, Forex traders act emotionally with fear and hope. Through this work, we presented a trading strategy that allows putting emotions aside, avoiding trading errors (greed, panic, or doubt) and not missing the trading opportunities. Clearly our strategy gives inputs and outputs signals when the predefined rules coincide. In this moment, our system is triggering regardless of sentiment and performance of the last losing or winning position.

The results presented in this work show the benefits of our system compared to a simple use of regression or classification using Random Forest. Taking into account the obtained results, using a combination of classification and regression trees can be implemented as a successful algorithmic trading system. Our results indicate that further research on the consecutive combination of many algorithms for Forex portfolio management is useful. This combination helps traders to determine the moment when we can buy or sell the currency pair.

## 6. Conclusion

In Forex there are many currency pairs and many trading people and each pair is different from the other, and each person thinks in his own way. Finding the best trading strategy is really a complex preoccupation. In order to find an adequate solution, we have presented in this study a new strategy based on two data mining algorithms. Our approach was to introduce a prediction and decision model that produces profitable intraweek investment strategy. The proposed strategy allows improving trading results in intraweek high-frequency trading. The results of the performed tests have demonstrated considerable advantage of our system versus a simple use of regression or classification using Random Forest. Such results are promising for research on consecutive combination of many algorithms to Forex portfolio management.

It is concluded that algorithmic trading based on combination of classification and Probit regression can be effective in improving the prediction accuracy. This combination helps to identify the good times to buy or to sell currency pairs. The proposed system, based on this combination, helps traders to take profit from the many opportunities on the Forex market.

## Data Availability

The data used to support the findings of this study are available from the corresponding author upon request.

## Conflicts of Interest

The authors declare that they have no conflicts of interest.

## References


[1] R. M. CALIFF, "Failure of Simple Clinical Measurements to Predict Perfusion Status after Intravenous Thrombolysis," *Annals of Internal Medicine*, vol. 108, no. 5, p. 658, 1988.

[2] V. Vajda, "Could a Trader Using Only "Old" Technical Indicator be Successful at the Forex Market?" *Procedia Economics and Finance*, vol. 15, pp. 318–325, 2014.

[3] S. Borenstein, "Report of the market simulation group on competitive supply/demand balance in the California allowance market and the potential for market manipulation," *Institute at Haas Working Paper*, vol. 251, 2014.

[4] R. Wolf, "Regulation of supply and demand for maternal nutrients in mammals by imprinted genes," in *The Journal of Physiology*, vol. 541, pp. 35–44, 2003.

[5] L. Breiman, "Random forests," *Machine Learning*, vol. 45, no. 1, pp. 5–32, 2001.

[6] J. A. Dubin and D. Rivers, "Selection bias in linear regression, logit and probit models," *Sociological Methods & Research*, vol. 18, no. 2-3, pp. 360–390, 1989.

[7] R. T. Baillie and D. D. Selover, "Cointegration and models of exchange rate determination," *International Journal of Forecasting*, vol. 3, no. 1, pp. 43–51, 1987.

[8] G. Fusai and A. Roncoroni, *Implementing models in quantitative finance: methods and cases*, Springer Science & Business Media, Springer, Berlin, 2008.

[9] P. Hooper and S. W. Kohlhagen, "The effect of exchange rate uncertainty on the prices and volume of international trade," *Journal of International Economics*, vol. 8, no. 4, pp. 483–511, 1978.

[10] J. Yao, Y. Li, and C. L. Tan, "Option price forecasting using neural networks," *Omega* , vol. 28, no. 4, pp. 455–466, 2000.

[11] M. R. El Shazly and H. E. El Shazly, "Comparing the forecasting performance of neural networks and forward exchange rates," *Journal of Multinational Financial Management*, vol. 7, no. 4, pp. 345–356, 1997.

[12] R. A. Meese and K. Rogoff, "Empirical exchange rate models of the seventies: do they fit out of sample?" *Journal of International Economics*, vol. 14, no. 1-2, pp. 3–24, 1983.

[13] R. A. Meese, "Testing for Bubbles in Exchange Markets: A Case of Sparkling Rates?" *Journal of Political Economy*, vol. 94, no. 2, pp. 345–373, 1986.

[14] R. McNown and M. Wallace, "Co-integration tests for long run equilibrium in the monetary exchange rate model," *Economics Letters*, vol. 31, no. 3, pp. 263–267, 1989.

[15] R. T. Baillie and R. A. Pecchenino, "The search for equilibrium relationships in international finance: the case of the monetary model," *Journal of International Money and Finance*, vol. 10, no. 4, pp. 582–593, 1991.

[16] N. Sarantis, "The monetary exchange rate model in the long run: An empirical investigation," *Weltwirtschaftliches Archiv*, vol. 130, no. 4, pp. 698–711, 1994.

[17] D. O. Cushman, "The failure of the monetary exchange rate model for the Canadian-U.S. dollar," *Canadian Journal of Economics/Revue Canadienne d'Economique*, vol. 33, no. 3, pp. 591–603, 2000.

[18] W. Poole, "Speculative Prices as Random Walks: An Analysis of Ten Time Series of Flexible Exchange Rates," *Southern Economic Journal*, vol. 33, no. 4, article no 468, 1967.

[19] M. P. Dooley and J. Shafer, "Analysis of short-run exchange rate behavior: March 1973 to November 1981," *Floating exchange rates and the state of world trade and payments*, pp. 43–70, 1973.

[20] R. Moscinski and D. Zakrzewska, "Building an Efficient Evolutionary Algorithm for Forex Market Predictions," in *Intelligent Data Engineering and Automated Learning – IDEAL 2015*, vol. 9375 of *Lecture Notes in Computer Science*, pp. 352–360, Springer International Publishing, Cham, 2015.

[21] M. P. Taylor and H. Allen, "The use of technical analysis in the foreign exchange market," *Journal of International Money and Finance*, vol. 11, no. 3, pp. 304–314, 1992.

[22] L. Menkhoff, "The use of technical analysis by fund managers: International evidence," *Journal of Banking & Finance*, vol. 34, no. 11, pp. 2573–2586, 2010.

[23] Y. Lui and D. Mole, "The use of fundamental and technical analyses by foreign exchange dealers: Hong Kong evidence," *Journal of International Money and Finance*, vol. 17, no. 3, pp. 535–545, 1998.

[24] D. Sahoo, K. Patra A, S. Mishra, and M. R. Senapati, "Techniques for time series prediction," *International Journal of Research Science and Management*, vol. 2, pp. 6–13, 2015.

[25] C. Panda and V. Narasimhan, "Forecasting exchange rate better with artificial neural network," *Journal of Policy Modeling*, vol. 29, no. 2, pp. 227–236, 2007.

[26] H. Ni and H. Yin, "Exchange rate prediction using hybrid neural networks and trading indicators," *Neurocomputing*, vol. 72, no. 13–15, pp. 2815–2823, 2009.

[27] A. Booth, E. Gerding, and F. McGroarty, "Automated trading with performance weighted random forests and seasonality," *Expert Systems with Applications*, vol. 41, no. 8, pp. 3651–3661, 2014.

[28] E. H. Sorensen, K. L. Miller, and C. K. Ooi, "The decision tree approach to stock selection: An evolving tree model performs the best," *The Journal of Portfolio Management*, vol. 27, no. 1, pp. 42–51, 2000.

[29] B. E. Boser, I. M. Guyon, and V. N. Vapnik, "Training algorithm for optimal margin classifiers," in *Proceedings of the 5th Annual ACM Workshop on Computational Learning Theory (COLT '92)*, pp. 144–152, July 1992.

[30] J. Wang, X. Wu, and C. Zhang, "Support vector machines based on K-means clustering for real-time business intelligence systems," *International Journal of Business Intelligence and Data Mining*, vol. 1, no. 1, pp. 54–64, 2005.

[31] S. Patra, "Techniques for time series prediction," *International Journal of Research Science and Management*, vol. 2, pp. 6–13, 2015.

[32] H. He and X. Shen, "Bootstrap methods for foreign currency exchange rates prediction," in *Proceedings of the 2007 International Joint Conference on Neural Networks, IJCNN 2007*, pp. 1272–1277, USA, August 2007.

[33] W. Lv and R. Zhang, "A regression model on effective exchange rate of RMB based on Random Forest," in *Proceedings of the 2nd International Conference on E-Business and E-Government, ICEE 2011*, pp. 4781–4783, China, May 2011.

[34] J. Matías, M. Febrero-Bande, W. González-Manteiga, and J. Reboredo, "Boosting GARCH and neural networks for the prediction of heteroskedastic time series," *Mathematical and Computer Modelling*, vol. 51, no. 3-4, pp. 256–271, 2010.

[35] A. Emam, "Optimal artificial neural network topology for foreign exchange forecasting," in *Proceedings of the 46th Annual Southeast Regional Conference on XX, ACM-SE 46*, pp. 69–73, USA, March 2009.

[36] J. Kamruzzaman and A. S. Ruhul, "ANN-based forecasting of foreign currency exchange rates," *Neural Information Processing-Letters and Reviews*, pp. 49–58, 2004.

[37] C. Evans, K. Pappas, and F. Xhafa, "Utilizing artificial neural networks and genetic algorithms to build an algo-trading model for intra-day foreign exchange speculation," *Mathematical and Computer Modelling*, vol. 58, no. 5-6, pp. 1249–1266, 2013.

[38] O. B. Sezer, M. Ozbayoglu, and E. Dogdu, "A deep neural-network based stock trading system based on evolutionary optimized technical analysis parameters," in *Proceedings of the Complex Adaptive Systems Conference with Theme: Engineering Cyber Physical Systems, CAS 2017*, vol. 114, pp. 473–480, USA, November 2017.

[39] J. H. Holland, *Adaptation in Natural and Artificial Systems: An Introductory Analysis with Applications to Biology, Control, and Artificial Intelligence*, University of Michigan Press, Oxford, UK, 1975.

[40] H. Subramanian, S. Ramamoorthy, P. Stone, and B. J. Kuipers, "Designing safe, profitable automated stock trading agents using evolutionary algorithms," in *Proceedings of the the 8th annual conference*, p. 1777, Seattle, Washington, USA, July 2006.

[41] A. Hirabayashi, C. Arahna, and H. Iba, "Optimization of the trading rules in forex using genetic algorithms," 2009, http://www.dollar.biz.uiovwa.edu.

[42] D. de la Fuente, A. Garrido, J. Laviada, and A. Gómez, "Genetic algorithms to optimise the time to make stock market investment," in *Proceedings of the the 8th annual conference*, p. 1857, Seattle, Washington, USA, July 2006.

[43] C. Schoreels, B. Logan, and J. M. Garibaldi, "Agent based genetic algorithm employing financial technical analysis for making trading decisions using historical equity market data," in *Proceedings of the IEEE/WIC/ACM International Conference on Intelligent Agent Technology. IAT 2004*, pp. 421–424, China, September 2004.

[44] R. Naik, "Prediction of stock market index using genetic algorithm," *Computer Engineering and Intelligent Systems*, pp. 162–171, 2012.

[45] F. E. Tay and L. J. Cao, "Improved financial time series forecasting by combining Support Vector Machines with self-organizing feature map," *Intelligent Data Analysis*, vol. 5, no. 4, pp. 339–354, 2001.

[46] K.-J. Kim, "Financial time series forecasting using support vector machines," *Neurocomputing*, vol. 55, no. 1-2, pp. 307–319, 2003.

[47] Y. Yuan, "Forecasting the movement direction of exchange rate with polynomial smooth support vector machine," *Mathematical and Computer Modelling*, vol. 57, no. 3-4, pp. 932–944, 2013.

[48] B. M. Henrique, V. A. Sobreiro, and H. Kimura, "Stock price prediction using support vector regression on daily and up to the minute prices," *The Journal of Finance and Data Science*, vol. 4, no. 3, pp. 183–201, 2018.

[49] S. Basak, S. Kar, S. Saha, L. Khaidem, and S. R. Dey, "Predicting the direction of stock market prices using tree-based classifiers," *The North American Journal of Economics and Finance*, vol. 47, pp. 552–567, 2019.

[50] J. Patel, S. Shah, P. Thakkar, and K. Kotecha, "Predicting stock market index using fusion of machine learning techniques," *Expert Systems with Applications*, vol. 42, no. 4, pp. 2162–2172, 2015.

[51] M. Kumar and T. M., "Forecasting stock index movement: a comparison of support vector machines and random forest," *SSRN Electronic Journal*, 2006.

[52] A. Geraci, D. Fabbri, and C. Monfardini, "Testing exogeneity of multinomial regressors in count data models: does two-stage residual inclusion work?" *Journal of Econometric Methods*, vol. 7, no. 1, 2018.

[53] C. Daganzo, *Multinomial Probit: The Theory and Its Application to Demand Forecasting*, Elsevier, 2014.

[54] A. Peltonen Tuomas, "Are emerging market currency crises predictable?-A test," 2006.

[55] G. Miner, R. Nisbet, and J. Elder, "Handbook of Statistical Analysis and Data Mining Applications," *Handbook of Statistical Analysis and Data Mining Applications*, 2009.

[56] T. Hastie, R. Tibshirani, and J. Friedman, *The Elements of Statistical Learning*, Springer, New York, NY, USA, 2008.

[57] M. Markowitz Harry, "Foundations of portfolio theory," *Journal of Finance*, pp. 469–477, 1991.

[58] J. Korczak, M. Hernes, and M. Bac, "Collective Intelligence Supporting Trading Decisions on FOREX Market," *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics): Preface*, vol. 10448, pp. 113–122, 2017.

[59] J. Korczak, M. Hernes, and M. Bac, "Fundamental analysis in the multi-Agent trading system," in *Proceedings of the Federated Conference on Computer Science and Information Systems, FedCSIS 2016*, pp. 1169–1174, Poland, September 2016.

[60] Y. Cheung and M. D. Chinn, "Currency traders and exchange rate dynamics: a survey of the US market," *Journal of International Money and Finance*, vol. 20, no. 4, pp. 439–471, 2001.

[61] A. V. Rutkauskas, A. Miečinskiene, and V. Stasytyte, "Investment decisions modelling along sustainable development concept on financial markets," *Technological and Economic Development of Economy*, vol. 14, no. 3, pp. 417–427, 2008.

[62] Y. L. Yong, D. C. Ngo, and Y. Lee, "Technical Indicators for forex forecasting: a preliminary study," in *Advances in Swarm*

*and Computational Intelligence*, vol. 9142 of *Lecture Notes in Computer Science*, pp. 87–97, Springer International Publishing, 2015.

[63] M. Ozturk, I. H. Toroslu, and G. Fidan, "Heuristic based trading system on Forex data using technical indicator rules," *Applied Soft Computing*, vol. 43, pp. 170–186, 2016.

[64] D. Pradeepkumar and V. Ravi, "FOREX rate prediction using chaos and quantile regression random forest," in *Proceedings of the 3rd International Conference on Recent Advances in Information Technology, RAIT 2016*, pp. 517–522, India, March 2016.

[65] Y. Chihab, Z. Bousbaa, H. Jamali, and O. Bencharef, "An approach based on heterogeneous multiagent system for stock market speculation," *Journal of Theoretical and Applied Information Technology*, vol. 97, no. 3, pp. 835–845, 2019.

[66] R. M. C. Pinto and J. C. M. Silva, "Strategic methods for automated trading in Forex," in *Proceedings of the 12th International Conference on Intelligent Systems Design and Applications, ISDA 2012*, pp. 34–39, India, November 2012.