



A Novel Hybrid Technique for Exchange Rate Forecasting

*Abbas Mohamed Salah, Tarik Ahmed Rashid & Shareef Maulod Shareef

**Software Engineering Department, College of Engineering, Salahadin University, Erbil, Iraq
studnet4ever@gmail.com
tarik.rashid@su.edu.krd
shareefshareef09@gmail.com*

Article info

Original: 27 Apr. 2015
Revised: 14 June 2015
Accepted: 25 June 2015
Published online:
20 Dec. 2015

Key Words:

*Exchange Rate,
Forecasting, e-
government, Feature
Extraction, Technical
and Fundamental
Analysis.*

Abstract

Exchange rate prices play a significant role in economic and financial systems. The precise forecasting of the exchange rate will aid to predict the situations of other aspects in the future. This offers useful information for economist traders to fulfill vital actions to eliminate risks that might lead to financial losses. One vital and common approach for forecasting is a quantitative technique. The quantitative technique is split into two approaches; fundamental and technical analysis. Both approaches have one different, which is the economic factor. The fundamental analysis uses economic factors while the other not, but both of them have drawbacks. In fundamental technique, the forecasters cannot understand the economic factors totally, the values and quantity of the factors unstable, while the technical analysis neglects the factors totally. This paper suggests a new technique which is appeared to be as a hybridized of the both techniques for forecasting exchange rate prices to overcome the issues mentioned above. The proposed technique makes new factors that have real values and have a real impact on the accurate result. It also derives the inputs and factors in the prediction models by a feature extraction technique which produces very accurate results.

Introduction

The significant role of government is to provide services to its citizen effectively and efficiently. This will be achieved by implementing e-government system in which uses technologies such as; Wide Area Networks, the Internet, and mobile computing which have the ability to transform relations with citizens, businesses, and other bodies of government. Information and Communications Technology (ICT) has the potential to improve business process and the value of governmental services, and also the relationship between government and business G2B e-government allows businesses to transact with each other more efficiently (B2B) and brings citizen closer to businesses (B2C), e-government aims to make the interaction between government and citizens (G2C), and inter-agency relationships (G2G) more friendly, convenient, transparent, and inexpensive [1]. Hence, business people can get benefits from forecasting technique to determine the gold price even not precisely, but it will help to plan and get a view of the big picture and to predict the future.

The exchange rate market is a complex and non-linear market, with a high degree of volatility characteristics. Thus, accurate forecasting in this field is not an easy task. As a result, literature provides a wide range of methods to predict economic activities. The range of these methods includes both the qualitative and the quantitative.

Qualitative approaches refer to those methodologies that absolutely depend on subjective approaches in the forecasting. In spite of value to forecasters, but the abilities of qualitative methods are changed from one forecaster to another. Since, they are fluctuating, so they cannot standardize in the prediction business. Therefore, they are ponderously depending on the forecasters' judgment, this is considered as a drawback of these approaches. Because of their trust on the judgment of "experts," and the less sensitive nature of qualitative methods, thus they are not very common by the researchers in the academic economic field [2].

Quantitative forecasting technique mainly involves the use of historical data, which is programmed into the future with the use of one or more mathematical approaches. It may, however, in some situations also be an intuitive or subjective prediction based on qualitative forecasting, which concentrated on subjective inputs that have been achieved from several sources like historical analogies or personal judgment. In such cases, quantitative forecasting is a combination of these two approaches and uses a mathematical model that is fitted by the forecaster's intuition. Quantitative category divided into two main techniques, which are the two major financial forecasting methodologies; Fundamental analysis and Technical analysis [3].

Fundamental analysis is the analyzing of macroeconomic and/or the microeconomic factors that affect the price exchange rate of a specific security in order to forecast its future movement [4]. The main macroeconomic factors that most described and effect on Iraqi dinar are: the interest rate, the rate of economic growth, the inflation rate, consumer spending, employment, and other economic indicators that have an impact on the currency market [5].

Technical analysis involves the forecasting financial price movement solely on financial historical data [6]. The accuracy of the data must be certain in order to minimize the risk associated with the exchange rate to find currency pairs price movement.

But, because of the financial forecasting field is a complex and non-linear field, therefore in this field each application, model or technique has its own drawback. For instance, the benefit of fundamental technique is that the researchers can enter the factors that may be occurring in the future. While the defect of this technique is that there are abundant economic factors that influence exchange rate but we cannot comprehend which factor truly is worked and which is not. As the other researchers have the same idea "The capital market is impossible to be efficient. Either stock market or exchange rate market, there is a lot of 'noise' in the market. Some 'cryptic information' play important role in exchange rate affecting but we have little sense of them. Information asymmetry made investor cannot analyses all factors that affect exchange rates. So they cannot build a specific bridge between the current price and the impact" as stated in [4].

While the researchers' idea in technical analysis is that the technical analysis provides a good result, since exchange rate price repeat itself in the future. But, the main drawback of the technical analysis is that completely neglected the economic factors and in the most cases it used for the short term horizon only. So, we are proposing the new technique, which is used to solve these issues between the both techniques. This technique is not neglected the factors completely as technical analysis and not get the unbalance factors as fundamental analysis, but makes the balance between the factors in the new form. The basic theory of the technique is done on the base of the theory of the neighbors as the many algorithms work on the same theory like, K-Nearest Neighbor algorithm, K-mean clustering, Hamming distance, Euclidean distance etc. [7-14]. Furthermore, if we show physically, the proposed technique is derived from foraging behavior of birds. When birds want to eat the food in the far or near regions, if the food is far, the birds use the wings and tail, and in the near region, they use just the wings to get the food [15]. It is clear birds' use wings to fly; while they use the tail to give them fly balancing (especially in the far away). Therefore, the wing is the inputs, the

tail is the factors, and the near region means the linear problem case and the far region means the non-linear problem case if compare to this technique. So, the main reasons that encourage us to make this technique are:

- 1) First to reduce the number techniques and methods that used during the preprocess data analysis like, de-noising, classification and feature extraction.
- 2) Second, to combine both traditional techniques in a new form for bettering forecasting results.
- 3) Third, make a new form of the factors that have the strongest relationship with the target and inputs in terms of more improving the forecasting result.

This paper is structured as follows: The details steps of the proposed technique are explained in Section 2. In Section 3, effecting factors on the forecasting result is detailed. Then, the benefits of the proposed technique for the prediction models in Section 4 are outlined. In Section 5, simulation and result which describes the empirical steps of the proposed technique is elaborated. The last section concludes the main points of the paper.

The Proposed Technique

This technique is absolutely new, it is the output of our thinking in the base of philosophy of new factors and it is not used by others before. The technique uses its own factors named PVER-PEF factors (later mentioned), as an effective factors in real exchange rate price. The technique also is a category of quantitative technique. The technique appears like a combination of fundamental and technical analysis. Since, in the beginning, it is technical analysis, then after joining the factors it will become the fundamental analysis, more details in later sections.

The proposed technique used in the preprocess analysis step, the main task of this technique is de-noising, classification and feature extraction.

The structure of the technique is shown in Figure 1, consist of the main three parts:-

- 1) Raw data
- 2) Calculations
- 3) Feature extraction

Details of the above parts are explained in the below sub sections:-

A. Raw data

The financial raw data are a combination of non-structural, missing and duplication data. In the first step, after collecting the data, then the pre-analysis process is necessary to get rid of the above issues and to reduce the force of tolerant that is often necessary by the preprocess technique to get good results. Then the proposed technique uses this data after applying pre-analysis process. Afterwards, apply many different processes and calculation to get high quality inputs to the prediction model.

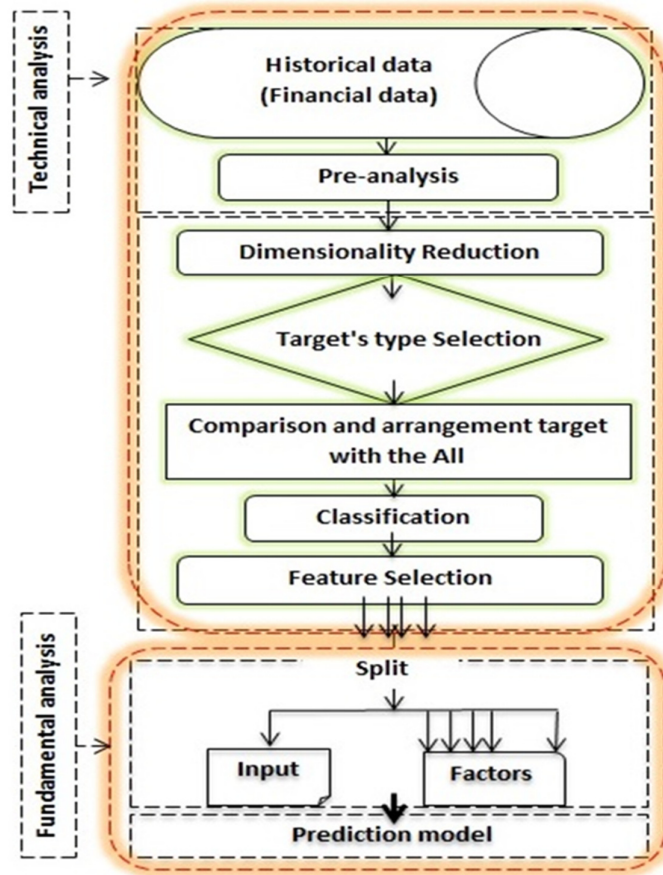


Figure-1: The overall proposed technique process

B. The computation part of the technique

The computations are the medium part in the structure of the proposed technique. This part is the main and very important part in this technique. Since, all the calculations that related to this technique will happen in this part as, de-nosing and classification. This part consists of a number of steps as follows:-

C. Dimensionality reduction of a raw data

Several applications in this field used multiple features in an attempt to get accurate prediction. If those features are complete and used to build up forecaster, then they operate in wider dimensions, and cause the learning process computationally and analytically to be complicated. Thus, the result often is on the extreme rise of prediction error. So, there is necessary to reduce the dimensionality of the raw data before forecasting. The purpose of this step is to reduce the dimensions and to show more difference between the months values of the historical data so that to easy process to the techniques succeeding steps. This reduction is done via “Get Average” for all months of all years for the analyzed data.

D. Target's type selection

The target is a forecaster-defined constant. In this step, after reducing dimensions via getting average for the all past months, then the target is selected. The role of the target in this technique is to classify the raw data into quality and non-quality data. There are two types of the target in the proposed technique, which are known and unknown targets, using each of these depends on the type of the prediction. In an in-sample prediction, the target is known target and, for out-sample the target is unknown as follows:-

i. Supervised target

Supervised target: is the target that is defined by the forecasters. In other words, it is the any month that the forecast will select so, this target well known as "forecaster target" (FT), and used as the basic input to perform the output in the prediction model.

ii. Unsupervised target

Unsupervised target: it is the supposed target. The selection depends on the expert of the forecaster. The forecaster should select the month as the target that it has similarity with the forecast month. However, often the target is the last month; since mostly the effecting of events repeats themselves in the different future months in same or different years. While a few times the sharply detachable happen. So, this target known as "supposed target" (ST).

E. Comparison and arrangement

In the comparison step, after the selection of any target's type, and compare with the all months. Then arrange the all months by two directions left and right of the target value based on the nearest distance from the target. The right direction values are bigger than the target, increase by increasing farness from the target. The left direction values are smaller than the target, decreases by increasing farness from the target.

F. Classification

Data classification is the process of organizing data into categories for its most effective and efficient use. A well-planned data classification system makes essential data easy to find and retrieve. After putting the values around the target based on the distance, then the data are classified into the two parts: quality and non-quality data. The classification of the data based on the level relationship of the neighbors with the target. Indeed, the strong level relationship means that the data and the target have the similarity, so this helps the model easily generalise and classify. While the weak level relationship between them makes the model look complex and weak to generalise. Hence, in this technique specifying the range of quality and non-quality data depends on the model forecasting result.

G. Feature selection

In this step, after classified and arranged the data around the target by its two directions, then discarding the non-quality area (depend on the forecaster's expert to discard this range), then in the quality area, and more especially in the very strong target's neighbors, the technique selects the neighbor's values based on smallest distance from the target by the both directions. So, in this step, again the high quality data is classified into; high and higher quality.

H. Feature extraction

Feature extraction includes reducing the amount of resources that needed to describe a big set of data. Since, analysis with an enormous number of variables generally demands a wide amount of memory and computation power. Also, this makes the model overfit the training sample and generalise poorly to new samples. Feature extraction is a general name for methods of constructing joining of variables to take around these problems while still describing the data with adequate accuracy [15]. The best results are obtained if feature extraction is performed by expert application. So, after sorting the higher quality data and putted around the target. Then in this step split the target's neighbors into the two common parts of the prediction model, which are; the real inputs and factors as described below.

i. The real inputs

In the quantitative forecasting, the inputs cause the techniques to classify into two techniques (fundamental and technical analysis). Therefore, the technique's power is estimated via their inputs. In this technique, the inputs are two or three vector variables, which are the first left and right of the target. Selecting these inputs based on their nearness from the target. The small different inputs with the target mean the small error in forecasting result. Selecting the two variable vector in the left and right direction is that to tell the model to predict based on this range and to decrease the over and under fitting of the forecasting result. For instance (in the moving average model), if the target value is 3, left input 2 and the right is 5, in this case the dividing of the inputs is unbalancing. So, if we get just left one or right one, in such a case we feed the unbalance input to the model. Since, in the left part the value increases by one while in the right the value increases by two compared to the target's value. So, in such a case we can average between both (if the target type is FT) or triple (if the target type is ST) to get the result as one input, this is expressed in Equations (1) and (2).

$$Input=(I_L + I_R)/2 \dots\dots\dots (1)$$

or

$$Input=(I_L + I_R + ST)/3 \dots\dots\dots (2)$$

Where;

IL : The Left input vector

IR : The Right input vector

ST : The forecaster target

The result of the equation is one input, in this case the errors between the target and input is decreased.

ii. The factors

A factor is a set of fundamental information that influences a trade or an investment's value. Economic factors either directly or indirectly affect the complete economy and all of its sectors. The philosophy of the factors in the economic forecasting field is to decrease the error between the prediction results. But, as previously mentioned, the capital exchange rate market is impossible to be efficient, the market is full of noises and secret information that perform an important role in this market, while we have little sense of them. Lack of balance and proportion of the Information made forecasters unable to analyse main real factors that affect exchange rates. The factors from a researcher to another have changed in terms of quantity, value, weight, etc. Furthermore, the historical data in the quantitative technique are valued, while the factors are hypothesis. So, it is not easy to change these hypotheses to real value. Because of these reasons, the researchers cannot construct a specific bridge between the exchange rate price and the economic factors [4].

Whereas in the proposed technique, the factors are the combination of the real value of the exchange rate and "effected economic factors in the past data" which is abbreviated into PVER-PEF factors (Past Values Exchange Rate-Past Effected Factors). Hence, the factors are consisted of the neighbor's vectors' values of the target that begin directly after the inputs. The factors are the real values, they have a strong relationship with the input and target, and they make the result of the model balance between the under and over fitting result.

The main role of these factors is to draw an image of fluctuation of the input data (linearity and non-linearity) and to show the target's dimensions to the prediction model to reduce more diversity between the prediction result and the target during the process. The factors in this technique are in a specific range, and their quantity depends on the problem. The factors are differing from the inputs, more detail is demonstrated in Table 1, and the feature extraction division based on the following three points:-

- 1) The quantity
- 2) The distance (far and near to the target)
- 3) The frequency (continue or occasionally existing as model's input)

Table-1: The difference between inputs and factors in the proposed technique

Inputs and Factors in the proposed technique		
	Inputs	Factors
Quantity	They are one or two in all the cases	Equal and more than two, and their quantity changes (increases or decreases) to get more improvement
Distance	They have more similarity with the target	Less similarity with the target
Frequency	They are the main and ever exist as inputs to the models	Occasionally exist and depend on the problem

So, in the last step before feeding the feature extraction into the model, the technique collapses the averages of the inputs and factors into the original days of the month, which is often 28 , 29 or 30 days for each average month value, then feeding the feature extraction into the prediction model to forecast in the high quality way.

Effecting Factors on the Forecasting Result

In the economic field, the prediction model is providing the forecasting result on the base of the financial data inputs and the factors. Therefore, the factors are participating as the main parts of the input model and have a big impact on the forecasting result. In this technique if the training model is not learning properly, then change their quantity (increase or decrease) number of factors .

The following illustrates the two different results of affecting the factors after feeding into the neural network model; the first one is without the factors, Figure 2 (a). If we look at the actual data (gray lines) and the training model (blue line), they are very different and the model generalises very poorly .

While the second one is with the factors, Figure 2 (b), but in this result, the training model according to the Figure, is a 100% fitted and the model forecasting has strong generalization ability. So, these visuals results can be taken as the strong evidence for PVER-PEF factors that have real influence on the forecasting result, and we proved it in practice, these factors have a vital impact on the prediction result (as mentioned in the next sections) in the term of improving and reducing error.

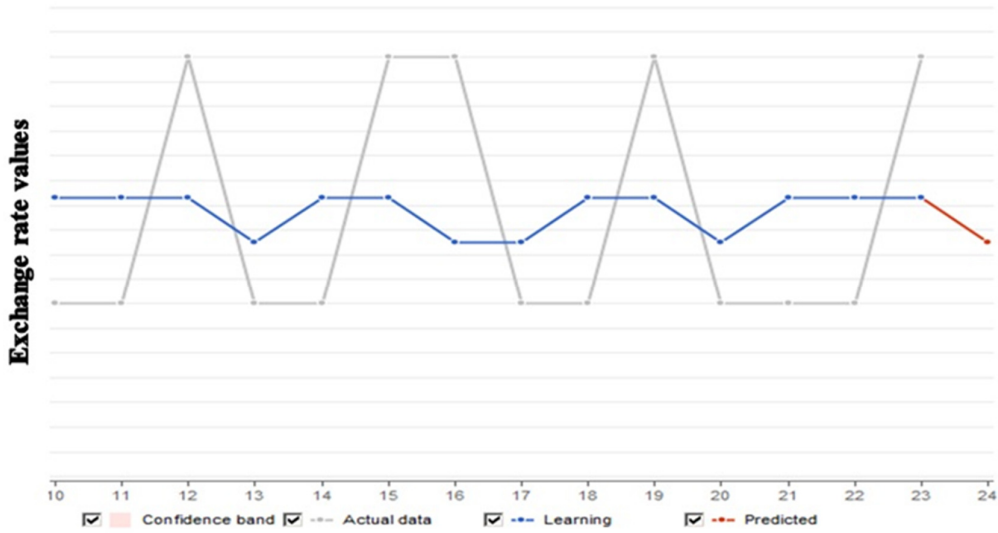


Figure-2 (a): Result without PERT factors

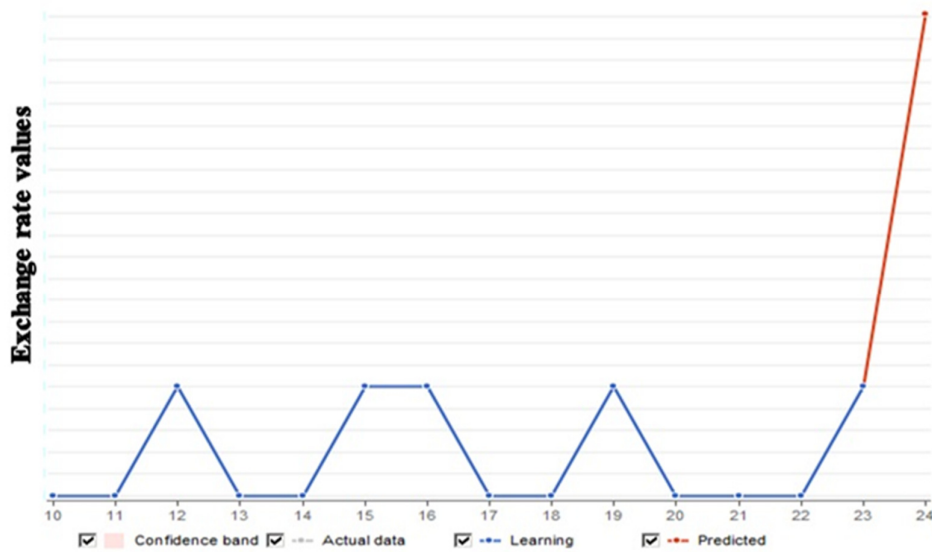


Figure-2 (b): Result with PERT factors

Benefit of This Technique for the Prediction Models

The raw data of exchange rate price are full of non-linearity, noisy, missing and duplicated. So, if we apply directly a prediction model to the raw data, our result has the big distance from the target, whereas if we apply some processes to this data the result will be more acceptable. Therefore, the advantage of the proposed technique to the prediction model's result is the three times increase in the result's quality, these processes can be described as follows :-

- 1) Classification; classify data into two parts; quality and non-quality data. Then, the technique splits the quality data into quality and highest quality data. Later in the highest quality data, select the inputs and factors, in this case the technique is appeared like fundamental analysis (but in the new form) because of joining factors in the prediction model, as shown in Figure 3. So, this technique, in

spite of combination of the both techniques in the new form, it also provides data classification and feature extraction for the prediction model, and this makes the prediction models easily predict, without necessary to complex model or add more hidden layers or increase the rate of training for the model. Therefore, the result is improved by one time.

- 2) De-noising; the noisy data, which makes the model generalise poorly, also it will remove and does not allow them to be fed into the model, as shown in Figure 4. Therefore, the result is improved by two times.
- 3) Feature extraction; despite that the remaining data is high quality (H_data) data, but the technique will also split the H_data into the *high* (H_data) and ($H+_data$) higher quality data. Then the technique divides the higher quality ($H+_data$) data into inputs and factors. This feature extraction makes the model generalise easily, thus, the result is improved by three times.

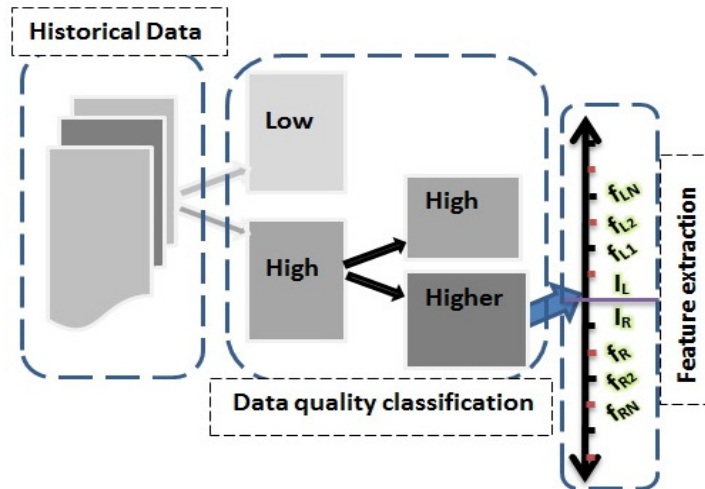


Figure-3: Classification and feature extraction

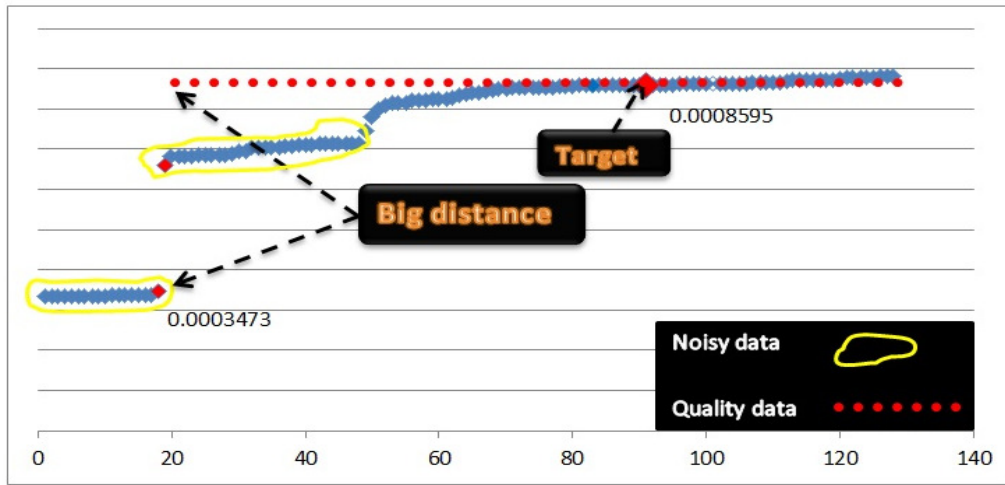


Figure-4: Noisy data by proposed technique

Simulation and Result

In last section of this paper, we want to apply the proposed technique for the historical exchange rate price for the Iraq (2003-2013), and then feed the output of the technique to the prediction model, and evaluate the result via mean absolute error (MAE) to provide the ability of the technique as follows.

Step one: The collect data for this paper are taken from the sources (<http://www.exchangerates.org.uk>) and (<http://www.oanda.com>) .The range of this data is from 2003 to 2013. The pre-analyses step is done to clean the data from duplication, missing, un-structured and distortion. Let raw data be R after pre-analysis phase. So, $R = \{A, B, C, \dots, \alpha\}$, the symbol of α is the number of the years in the raw data.

Hence α are 11 since it is from 2003 to 2013. So, $\alpha = K$ (K index in English alphabet is 11), therefore, the raw data in this paper is in the following Equations:-

$$R = \{A, B, C, \dots, K\}. \dots\dots\dots(3)$$

Where,

$$A = \{A1, A2, A3, \dots, An\} \dots\dots\dots(4)$$

$$B = \{B1, B2, B3, \dots, Bn\} \dots\dots\dots(5)$$

$$C = \{C1, C2, C3, \dots, Cn\} \dots\dots\dots(6)$$

$$K = \{K1, K2, K3, \dots, Kn\} \dots\dots\dots(7)$$

The n : is the number of months in each year in the historical data, the n value by default is 12.

And more,

$$A1 = \{a1, a2, a3, \dots, a\beta\}, \dots\dots\dots(8)$$

$$B1 = \{b1, b2, b3, \dots, b\beta\} \dots\dots\dots(9)$$

β is the number of days in the months of the year. Indeed, β is 28, 29, 30 or 31 according to the month in the English calendar. So, if we multiply 11 (years) X 12 (Months) = 132 months (Vectors values).

Step two: Raw data dimensionality reduction

As previously mentioned, we make a suitable environment for succeeding proposed technique's steps via get average to the all months. The following are the calculation process to get average for each month. We will change the month name after getting average by adding a small dash for the each alphabet symbol as expressed in the following Equations:

$$\hat{A}_1 = \frac{a_1 + a_2 + \dots + a_\beta}{\beta} \dots\dots\dots (10)$$

$$\hat{B}_1 = \frac{b_1 + b_2 + \dots + b_\beta}{\beta} \dots\dots\dots (11)$$

$$\hat{K}_1 = \frac{k_1 + k_2 + \dots + k_\beta}{\beta} \dots\dots\dots (12)$$

$$\hat{A} = \{\hat{A}_1, \hat{A}_2, \hat{A}_3, \dots, \hat{A}_n\} \dots\dots\dots (13)$$

$$\hat{B} = \{\hat{B}_1, \hat{B}_2, \hat{B}_3, \dots, \hat{B}_n\} \dots\dots\dots (14)$$

$$\hat{K} = \{\hat{K}_1, \hat{K}_2, \hat{K}_3, \dots, \hat{K}_n\} \dots\dots\dots (15)$$

Finally, the raw dimensions are reduced to \hat{R} .

$$\hat{R} = \{\hat{A}, \hat{B}, \hat{C}, \dots, \hat{K}\} \dots\dots\dots (16)$$

See Figure 5, which shows the movement of the price in the past.

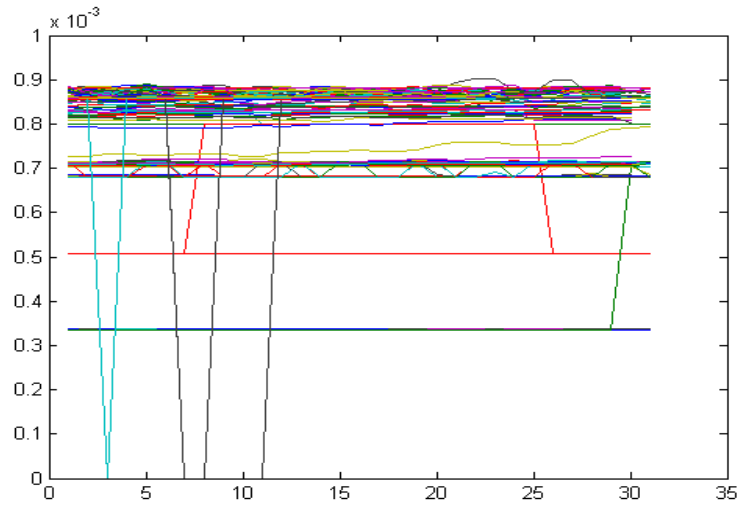


Figure-5: Observed exchange rate data 132 months

Step three: Target selection

The target's type in this paper is select target (ST), so the target is Aug-2013, it means \hat{K}_8 of the proposed technique. Hence the target is Aug-2013, so we should drop three months of 9 to 12 in the year of 2013, thus $132-4=128$ months remain to the rest steps.

Step four: comparison and arrangement

In the comparison step, after the selection the ST target \hat{K}_8 , and compare with the all $(\{\hat{A}_1, \hat{A}_2, \hat{A}_3, \dots, \hat{A}_n\}, \{\hat{B}_1, \hat{B}_2, \hat{B}_3, \dots, \hat{B}_n\}, \dots, \{\hat{K}_1, \hat{K}_2, \hat{K}_3, \dots, \hat{K}_n\})$. Then arrange them by two directions left and right of the \hat{K}_8 value based on nearest distance from it, Table 2 is illustrating the result of this step which is 40 neighbors of the \hat{K}_8 (we can't display 128 months). If we look at the table, the right direction values are bigger than the target, increases by increase far from the target. And vice versa this applied for the left direction values. Furthermore, if we note, the bold black line which is \hat{K}_8 in the Figure 6, it is in the center of the color lines (which are $(\{\hat{A}_1, \hat{A}_2, \hat{A}_3, \dots, \hat{A}_n\}, \{\hat{B}_1, \hat{B}_2, \hat{B}_3, \dots, \hat{B}_n\}, \dots, \{\hat{K}_1, \hat{K}_2, \hat{K}_3, \dots, \hat{K}_n\})$). The lines are arranged around the \hat{K}_8 according to their distances. So, if we compare this output (See Figure 6) with Figure 5, it is show us that this technique did big deal despite of the succeeding steps.

Table-2: A sample result of target and its neighbors in proposed technique

Factors										Input	Target	Input	Factors																											
I10	I12	I8	I5	I7	I6	J1	J8	J9	J7	J10	J6	J3	F9	K1	J5	J11	K2	J4	K8	K7	K4	I3	K6	K3	I4	I2	I1	F7	H12	H4	H11	K5	F10	H9	H10	H6	H7	H5	F11	I10
0.000854619	0.000855238	0.000855261	0.000855565	0.00085581	0.000856	0.000856682	0.000858174	0.0008583	0.000858474	0.000858478	0.000858647	0.000858682	0.0008587	0.000858739	0.000859	0.000859045	0.0008591	0.000859238	0.0008595	0.000860182	0.000860484	0.0008608	0.000860952	0.0008611	0.000861214	0.000861871	0.000862355	0.000863452	0.000863467	0.000863633	0.000864304	0.000864548	0.0008651	0.000867097	0.000867133	0.000867323	0.000867645	0.0008677	0.000854619	

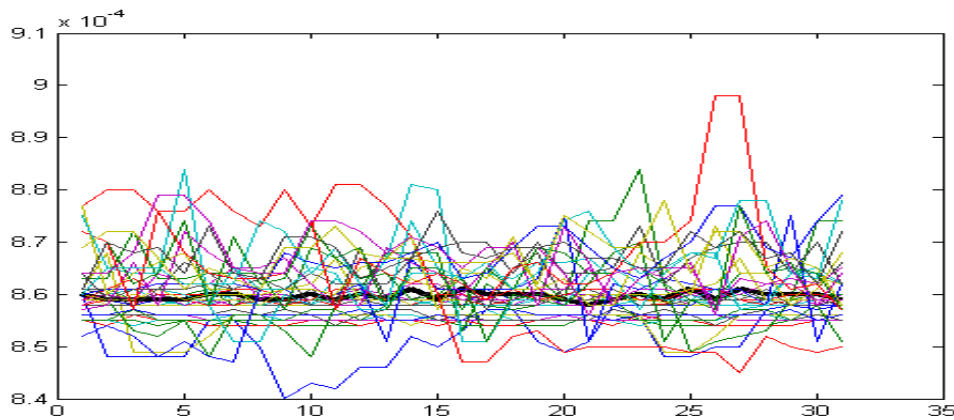


Figure-6: Sorted some months based on distance from target

Step Five: Classification

After putting the values around the target based on their distances. Then the data are classified into the two parts: quality and non-quality. The range of the quality data is a custom range, depends on the forecaster's idea. For instance, in this paper, we get just 40 neighbors around \hat{K}_8 (as shown in Table 2). The classification of the data based on the level relationship of the neighbors with the target. Indeed, the strong level relationship means that the data and the target have a big similarity (high quality), so this helps the model easily generalise and classify. While the weak level relationship (low quality) between them make the model get complex and weaker to generalise.

Step Six: Feature selection

In this step, after discarding the non-quality area (2003 - 2006 and more), then in the quality area, the technique selects the neighbor's values based on smallest distance from the \hat{K}_8 by the both directions. So, in this step, again the process of classifying the quality data into high and higher quality is seen. Table 3 shows a higher quality data, and this table is produced from Table 2 which is a high quality data. These values are in the shape of linear are arranged around of \hat{K}_8 . Figure 7 displays higher qualities of the target's neighbors.

Table-3: Higher quality data produced from the quality

Factors					Input	Target	Input	Factors				
0.0008587	0.000858739	0.000859	0.000859045	0.0008591	0.000859238	0.0008595	0.000859565	0.000860182	0.000860484	0.0008608	0.000860952	0.0008611
F9	K1	J5	J11	K2	J4	K8	K7	K4	I3	K6	K3	I4
Sep-08	Jan-13	May-12	Nov-12	Feb-13	Apr-12	Aug-13	Jul-13	Apr-13	Mar-11	Jun-13	Mar-13	Apr-11

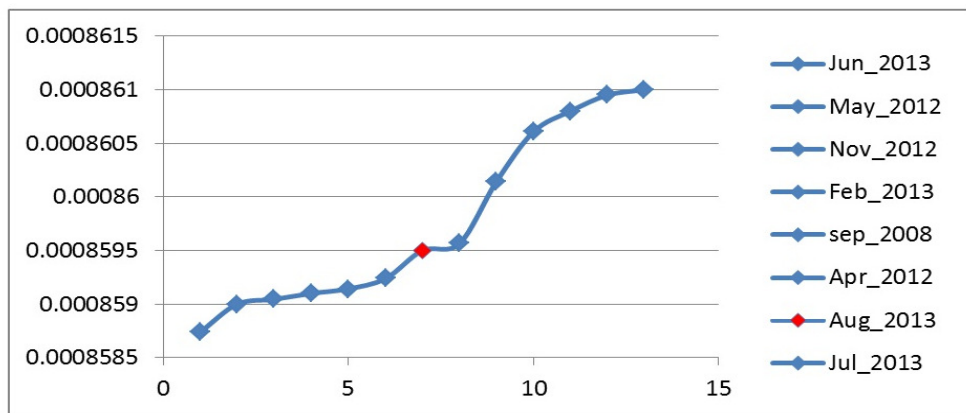


Figure-7: Higher qualities of the target's neighbors

Step seven: Feature extraction

Feature extraction includes reducing the amount of resources that needed to describe a big set of data. The quality of the forecasting result depends on the feature extraction. So, after sorting the higher quality data and putting them around the target, then in this step, split the Strong's neighbors of K_8 into; Inputs and Factors. The inputs are the $J4$ and $K7$ in the case of ST target, but in the case of the FT target, we get also the FT value ($K8$) as an input to join with the both standard inputs to the prediction model, as shown in Figure 8 (a) and (b). The factors in the left part are $\{F9, K1, J5, J11, \text{ and } K2\}$ and in the right part, are $\{K4, I3, K6, K3, I4\}$.

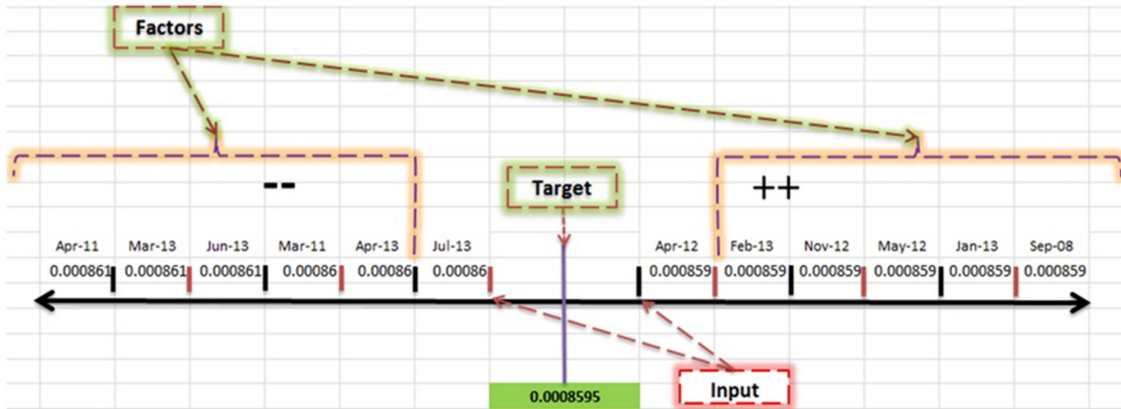


Figure-8 (a): FT target, inputs and factors

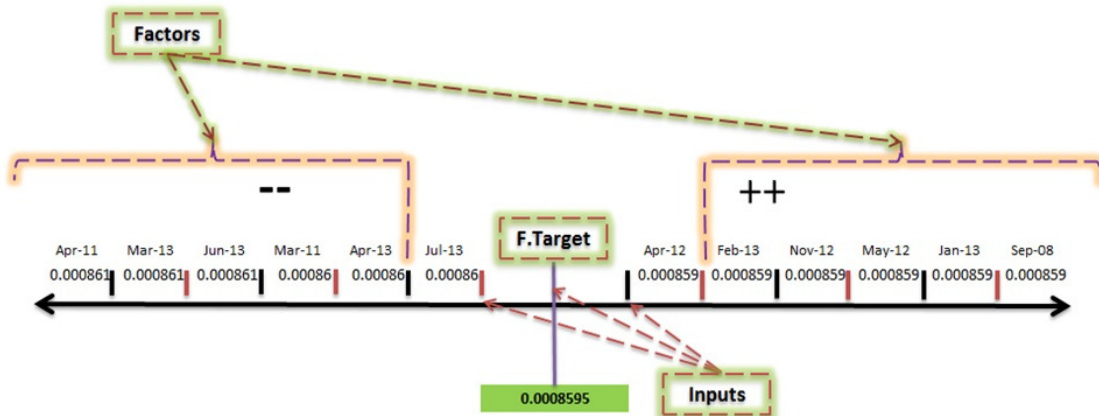


Figure-8 (b): ST target, inputs and factors

Then, in the last step, collapse inputs and factors to the normal days, and then feed the selected data into the prediction model to get a high quality forecast. To compare the in-sample volatility forecasting performance, the standard statistical criteria, mean absolute error (MAE), can be used to measure the forecasting accuracy. It expressed via Equation (17).

$$MAE = \frac{1}{N} \sum_{t=1}^N |A_t^2 - F_t^2| \dots \dots \dots (17)$$

- [3]. Granvik , A. R., "*Forecasting Exchange Rates*", Msc Thesis, International Business, (2010).
- [4]. Liu , B., Wan H., Chen, X., "*Exchange Rate Forecasting Method Based on Particle Swarm Optimization And Probabilistic Neural Network Model*", IEEE, (2011).
- [5]. Eddelbuttel, D., Mccurdy, T.H., "*The Impact Of News On Foreign Exchange Rates*", Evidence from High Frequency Data, Manuscript, Rotman School Of Management, University Of Toronto. Forex News, Data Gathered Since Sep, (2008).
- [6]. Alamili, M., "Exchange Rate Prediction Using Support Vector Machines", Msc Thesis, January, 2011.
- [7]. Meng L., "*Exchange Rate Forecasting Based On Neural Network With Revised Weight*", IEEE, 2011.
- [8]. Bhatia N., Vandana SSCS., "*Survey Of Nearest Neighbor Techniques*", International Journal Of Computer Science And Information Security, Vol. 8, No. 2, (2010).
- [9]. Anguilli, F., "*Fast Condensed Nearest Neighbor Rule*", Proceedings Of The 22nd International Conference On Machine Learning, Bonn, Germany, (2005).
- [10]. Kanungo, T., Mount D. M., Netanyahu N. S., Piatko C. D., Silverman R., Wu A. Y., "*An Efficient K-Means Clustering Algorithm: Analysis And Implementation*", Ieee Transactions On Pattern Analysis And Machine Intelligence, Vol. 24, No. 7, July (2002).
- [11]. Hamming, R. W., "*Error Detecting And Error Correcting Codes*", Bell System Technical Journal Vol. (29), No. 2, pp. 147–160, (1950).
- [12]. Deza, E. , Deza, M. M., "*Encyclopedia Of Distances*", Springer. pp. 94, (2009).
- [13]. Jenjira Tipyan, "*Quantitative Models For Forecasting Vehicle Fuel Prices In Thailand*", Computer Science And Information Engineering, 2009 Wri World Congress, Vol (2), (2009).
- [14]. Huang B.Q. , Kechadi T.-M. , Buckley B., Kiernan G. , Keogh E. , Rashid T. , "*A New Feature Set With New Window Techniques For Customer Churn Prediction*", Expert Systems With Applications, Expert Systems With Applications, Vol. (37), pp. 3657–3665, (2010).
- [15]. Isabelle G., "*An Introduction To Variable And Feature Selection*", pp. 1157-1182, (2003).