

Survey of the Euro Currency Fluctuation by Using Data Mining

M. Baan, E. Saadati, and M. Nasiri

Department of Computer Engineering, Iran University of Science and Technology,
Narmak, 16846-13114, Tehran, Iran

baan3117@gmail.com, it.saadati@gmail.com, nasiri_m@iust.ac.ir

ABSTRACT

Data mining or Knowledge Discovery in Databases (KDD) is a new field in information technology that emerged because of progress in creation and maintenance of large databases by combining statistical and artificial intelligence methods with database management. Data mining is used to recognize hidden patterns and provide relevant information for decision making on complex problems where conventional methods are inefficient or too slow. Data mining can be used as a powerful tool to predict future trends and behaviors, and this prediction allows making proactive, knowledge-driven decisions in businesses. Since the automated prospective analyses offered by data mining move beyond the analyses of past events provided by retrospective tools, it can answer the business questions which are traditionally time consuming to resolve. Based on this great advantage, it provides more interest for the government, industry and commerce. In this paper we have used this tool to investigate the Euro currency fluctuation. For this investigation, we have three different algorithms: K, IBK and MLP and we have extracted Euro currency volatility by using the same criteria for all used algorithms. The used dataset has 21,084 records and is collected from daily price fluctuations in the Euro currency in the period of 10/2006 to 04/2010.*

KEYWORDS

Euro Currency Fluctuation, Data Mining, Stock Market, Knowledge Discovery in Databases.

1 Introduction

Data mining or knowledge discovery in databases (KDD) is a new science considering the countries progress in the field of IT and penetration of computer systems in the industry and creating large databases by government departments, banks and private sector need to use it is deeply felt. For example, data mining knowledge discovery and reliable information hidden in databases, or to better express, machine data analysis to find useful, new and reliable patterns, in large databases, called data mining. Data mining in small databases is most commonly used; patterns of results produced by the strategic decisions of small business firms can take advantage of many. Data mining application can be summed up in the following statement: Data mining gives information for intelligent decisions that you make regarding difficult dilemmas in the workplace. Neural networks have been used for prediction in various fields; as the extent of the prediction area and its issues, collecting all possible resources and articles published in this field makes us rigid.

Example applications of neural networks in forecasting in the period between 1996 to 2005 can be found in the Refs. [1]-[8]. More interested to study in other application areas of neural networks can be provided by the articles Lin and Fadlalla (2001) [9] and Wong (1998) [10] in the field of financial applications, Zhang and others (1998) [11] in the anticipated public domain (General Forecasting Area) Krycha and (1999) Wagner works in the field of management science.[12] other references in abundance are found to the global Internet network. As you can see, a very large number of diverse prediction issues are solved by neural networks. Diverse business fields including accounting domain (Financial predicted earnings, predicted earnings astronomy, predicted the bankruptcy and commercial losses). Finance (predicted direction, index, returns, risk, rate of change, futures, Stock and commodity prices, efficiency and ... in Stock market), Marketing (predicted selected customers, market share, classified, market trends) Economy (Forecasting Business Cycles) Record (Recession) Amount spent (buying) customers, inflation, industrial production of goods And bonds purchased by the production and utilization of America (demand predicted electricity consumption, traffic highway investment, the success of new product projects, project size, application or sale of goods and retail) International Trade (predicted performance projects with joint investment, currency exchange rate) Housing transactions (making massive demand predicted, price and housing), predicted processes related to the environment (predicted concentration levels of ozone and air quality) has been solved by neural networks [13] (Many facts mentioned above are classified among the prediction issues; But applications such bankruptcy prediction is not necessarily among prediction applications; Such matters like bankruptcy are not only important, but time is also matters. Hence, these issues can also be predicted among the issues considered; such as predicted probability of bankruptcy in the future.

2 Data Mining

The act of data mining is divided into several marked stages database. in this paper we confine to introduce and a brief description of each of these steps:

1. Define the problem.
2. Data recognition:
 - Data warehouse formation: This stage is where performed for the formation of continuous and integrated environments are performed in order to perform the next steps and data mining on it. In general, the data warehousing is continuous; collection and classified constantly changing and a new is dynamic that ready to explore.
 - Data Selection at this stage, in order to reduce costs of data mining operations, data that have been studied are selected from the database. Data mining aims to give results about them. Data conversion: is determined for data mining operations and should be performed on necessary transformations for certain data. These transformations may be very convenient and concise as byte to integer conversion or very complex and time consuming with high costs, such as defining new attributes and extracting or converting data from string values.
3. Pre-processing.
4. Explore in data: at this stage the data mining is done. At this stage data is being explored by using data mining techniques. What is extracted is hidden knowledge of them and modeling is done.

5. Results Interpretation: At this stage, results and patterns offered by datamining tools investigated and useful results is determined. Also used in data mining, are genetic algorithms and neural networks. Neural networks are used due to their efficiency are useful in solving large complex problems. The application of genetic algorithms in data mining for search and build an optimal model among the models is achieved, Such that the initial models are placed on some chromosomes and competition over the transfer of traits to the next generation. The best and most worthy models are and most worthy to be presented to the user [14] [15].

3 Data collection

Our data set includes 21,084 records from 10/2006 to 4/2010 [16]; and have been gathered based on daily price fluctuations in the euro exchange forex Sponsored Europe. This includes the following classifications:

- Time, date: day, month, year: Each 24 record is a day.
- New Price in specific hours- Open: Euros Price at Arrival time in hours. For example, at 12:01 How much is the euro price.
- Out Price in an hour- Close: EUR Price per hour at the time of departure. For example, at 12:59 How much is the euro price.
- Lowest price per hour- Low: It is the Lowest Euros price in that time.
- The highest price at the time- High.
- Average price: Average price in that hour.

4 Data Pre-processing

In order to obtain better results when sorting through data, we applied the following changes:

- Add weekdays category to evaluate excitement of the first days of the week after weekend and holidays in the market. - Rounding the total data to get a better result.
- Add seasons Category to evaluate the market fluctuations in different seasons.
- Separated the day, month and year to better access to data.
- Open and Close is possible nearly equal and difference of them is in the decimal range. So to simplify the data we have decided to remove the exit price.
- Methods used in modeling.

5 Neural Network

Neural network have three concepts: 1 - Data Analysis System 2 - neurons or nerve cells 3 - neurons Labor Law Group, or Network. In a classical definition, neural networks, are a set of simple processing elements that are connected. Processing elements in neural networks are much easier than conventional processors with numerous differences. [17] Each neuron with a number of other neurons to connect directly and is independent and weight of the connections will determine their relationship, the data are placed in weights. Neural net- work has the following features:

1. Do Processing units.
2. No virtual memory Part And information are saved in a set of weights.

3. Loss of parts of networks and failure are not causes of the networks failure to resist the noise ratio and hardware failures.

Compared with artificial neural networks, the Vector Machine method is to support relatively new Methods, that have shown good performance than the perceptron neural networks in recent years. To find the solution equation Optimum Line for data by quadratic programming methods the known methods in solving problems are significant limitations is done. Support Vector Machine is basically a linear machine. From human neurons to artificial neurons: Setting aside some of the critical properties of neurons, and internal communications can be simulated a primary model of neurons by computer.

6 Neural Networks Structure

A neural network consists of components and weights of layers. Behavior of the network is dependent on the relationship between members. In general neural networks have three neuronal layers: 1. Input layer: the raw information is fed to the network. 2. Hidden layer: These layer performances are determined by relationship between input and weights and hidden layers. Weights between input and hidden units sets when a hidden unit that must be activated. 3. Output layer: Outputs Unit performances are depending on the hidden unit activity and weight between the hidden and output units.[18] There are also Single-layer and multi-layer networks that single-layer organization where all units have a connectivity layer, the most used has Greater computational potential than the multiple layers organization. Units in multi-layer networks are numbered by the layers (Instead of pursuing overall numbering). [17] Both layers of a network communicate with each other by weights, In fact connections. In neural networks are some types of connection or link weight: Pioneer: More links of this type in which signals are only in one direction. Does not exist any feedback from output to input (loop) . The output of each layer has no effect on the layer. Backward: [17] Peripherals: Output nodes of each layer are used as input nodes.

7 K* algorithm

An Instance-based Learner Using an Entropic Distance Measure selecting values for the parameters x and s , and a way of using the values returned by the distance measure to give a prediction. Choosing values for the arbitrary parameters for each dimension we have to choose values for the parameters x (for real attributes) and s (for symbolic attributes). The behavior of the distance measure as these parameters change is interesting. Consider the probability function for symbolic attributes as s changes.

8 IBK algorithm

IBK is an implementation of the k-nearest-neighbors classifier that employs the distance metric discussed. If more than one neighbor is selected, the predictions of the neighbors can be weighted according to their distance to the test instance.[19]

9 Decision Tree; Association Rules

Two different formulas have been implemented for conversion between weight and distance. Number of training samples that are kept by the classifier may be limited to window size by setting options. When new samples are added and old samples deleted the old samples remain as the total number of samples in training set size [20] [21].

10 Results and Discussion

In this paper we used three classification algorithms to forecast the euro currency fluctuation. Multi-layer Perceptron procedure neural network, K* and IBK were the models which used to forecast the euro currency fluctuation. Five layers show input that has duty to get raw information which feed the networks. Table 1 compares same criteria of each algorithm: This index is the root mean square approximation. Root Mean Square Error index for good models is 0.05 or less. The models which there RMSEA is 0.1 have a week processes. This index is a relative absolute error. Relative absolute error index is better to have a higher value. This index is expressed as a percentage.

	IBK	K*	MLP
Root Squared Error	0.051	0.42	0.52
Relative Absolute Error	1.078	82.75	100.1189

Table 1. Compares criteria of RSE and RAE

Also association rules and classification rules which induction of decision trees used for conclusion. Some simple of association rules and classification rule are shown below follows C5 algorithm in Clementine software. For example, analyses of the 5 roll are discussed below:

First roll:

If first month

And first day, then price decreases.

If second day and the average price is less than 1.745 and on Monday and Thursday, the price decreases.

if Tuesday and

the average price is smaller than 432 or less then price decreases. the average price is greater than 432, the price increases.

on Wednesday, the price increases. on Friday, the price decreases.

the average price is greater than 1.745, the price increases.

on day nine, on Monday, Tuesday, Wednesday, Thursday and Friday, the prices decrease

Second roll:

on days 10, 15, 17, 20, 21, 22, 29, the price increases.

If on first month, 9.48 percent done shopping that 18.01 percent are on Monday. Trading volume in January was higher. It seems after the holidays ensure the relative movement at the end of the year causes the market itself shows good mobility. 9.48 percent for trades made in January, confirmed it. Statistics show that on Mondays in January, more trading takes place during the week than the rest of. On Fridays we also have a relative increase in trading volume and price fluctuation. It is well because it could be time to declare the U.S. unemployment statistics attributed. The listed statistics affected more on the trend of equality exchange between dollars to euro rate. Since it shows the strength of the U.S. economy and future prospects in the short term. Successful transaction is not so simple and easy and needs several components such as extensive knowledge, understanding market conditions and confidence and composure. In currency markets, the timing and on time entering into the trade are the most important factor in successful deal but sometimes determine the right time to deal is unknown. Never expect that every transaction could earn profits. Trading in foreign exchange markets based on conjecture and estimates and can be caused losses. However, these transactions can be exciting, teaser and even addictive. Whatever you have more dependence on your money and investment, making decisions with a comfortable mind about it would be more difficult. Your money is worth so with the money that you need to live, you should never be traded. Before the transaction you should know what the market situation. Whether the process has upside, downside, or is neutral. Whether this trend is strong or weak? And did so before the start of the process or process is new? Obtain the clear and accurate picture of market position, cause to the successful transaction. Many traders attempted to transaction without specifying the time out of the transaction. Of course not doubt that the first goal is profit; however the trader should be focused his mind on exactly what to predict the market movements. For the transaction, carefully considered and determined on the anticipated market moves in a certain period of time is essential. One of the cases in this area should be considered is out the deal time. Importance of this issue is that your mind would be ready to make it. Although specifying the exact time of departure of the transaction is not possible, but to specify the time before entering the trade is very important. If your number of transactions is high during the day, technical analysis on the daily charts is less important and is better to use 30- minute charts or an hour ones. Moreover, you should know start and end of working hours of financial and economic centers of the world. You should keep in mind these times when doing the trading. Since hours, volatility, liquidity and market movements are noticeably changed. Can be speculated synchronize on with the potential of market But it may be too early or too late to do. Attention to the time of transaction can be effective in the result. The news will be announced in the market, such as CPI (Consumer Price Index), announced retail sales or the central banks decide to increase interest rates, can stabilize the previous market movement. Due to the timing transaction means that know what to expect before the transaction and you can specify them in advance. Technical analysis can help you to detect when and how the prices will be changed. If you have doubt about doing a deal correctly and you are not sure, do not enter the market. Generally, Adjusting and measuring transactions in the way to re-enter the market and trading with other currencies are more rational. In short, with large quantities that may destroy your account do not deal. And with a slang words; do not put all your eggs in one basket. What the majority of market plan for situation and movement or would do in future are called market trends. Which means you will successful if you go to the right direction for trading

market. It should be mansion that this removal is a very simple and basic image as a process may take in any time and move in the opposite direction. Technical and fundamental analysis can determine when to start the process and whether this trend is strong or weak. Market expectations are indicating tendencies that most traders and analysts of the market and the news will be announced in the near future. If they expect that interest rates rise then it will. A lot of changes in market movements will not be seen because the market already announced this change to react and prepare for it but if the news is announced unlike forecasts, market will inevitably react strongly to it. Attractive Association Rules Extracted:

- If it is 13th day and 3.15 percent of the purchase is done then 28.87 percent is on Tuesday.
- If it is 4th month and 7.57 percent of the purchase is done then 22.63 percent is on Wednesday.
- If it is 5th month and 3.75 percent of the purchase is done then 22.88 percent is on Friday.
- If it is 12th day and 3.38 percent of the purchase is done then 22.88 percent is on Friday.
- If it is first month and 9.48 percent of the purchase is done then 18.01 percent is on Monday.
- If it is 3th day and 3.28 percent of the purchase is done then 16.64 percent is on Friday.
- If it is 12th day and 3.38 percent of the purchase is done then 16.69 percent is on Wednesday.
- If it is 4th day and 3.39 percent of the purchase is done then 16.81 percent is on Tuesday.
- If it is 11th day and 3.39 percent of the purchase is done then 16.81 percent is on Tuesday.
- If it is 3th month and 5.01 percent of the purchase is done then 17.98 percent is on Thursday.
- If it is second month and 4.57 percent of the purchase is done then 17.96 percent is on Wednesday.
- If it is first month and 4.46 percent of the purchase is done then 17.75 percent is on Friday.
- If it is second month and 9.14 percent of the purchase is done then 20.29 percent is on Friday.
- If it is first month and 9.48 percent of the purchase is done then 47.07 percent of price decreases.

11 Conclusion

K* algorithm has a less Root Squared error. MLP algorithm has e more Relative absolute error than the rest. In order to make a better understanding in the classification, classification techniques based on the Association Rules, i.e. the association classification. The main purpose of classification is a prediction in terms of class. While the discovery of association rules describe relationships between items in transaction database. In association classification, classifier made section on a subset of association rules called association rules classification. In association rules classification, after every law is a class attributes. Classifier studies a rule or set of compatible laws with the object to predict an object tag. The results of work performed in this field show that association classification act better than Machine learning Classification algorithms [22] However, association rules algorithms has challenges such as determine the minimum values, For extracted association rules because first, the algorithm produces a large number of laws and storage, retrieval, pruning and sorting of these laws is difficult. Also, to find the best subset of rules for build- ing strong and accurate classification is

challengeable. [22] However, association rules algorithms has challenges such as determine the minimum values for extracted association rules because first, the algorithm produces a large number of laws and storage, retrieval, pruning and sorting of these laws is difficult. Also, to find the best subset of rules for building strong and accurate classification is challengeable. [9] In recent years a number of association classification such as CPAR (Han J.; Pei J. Yin Y.; 2000), CMAR (Li W.; Han J.; Pei J.; 2001), MCAR (Thabtah, F.; Cowling P.; 2005), MMAC (Thabtah F.; Cowling P.; 2004) is presented. These algorithms are used various ways in the discovery, ranking, pruning, prediction and evaluation of interest. To construct classifier with association rules algorithms, first found complete set of association rules from training data set and select a subset to make a classifier. Selecting such a subset has different methods. For example CBA (Liu B.; Hsu W.; Ma Y.; 1998[23] and CMAR algorithms, this method occurs of selection based on coverage heuristic method. In this method we evaluate complete set of CARs on educational data sets and considered laws that cover certain number of educational data. When the classifier made, it evaluate the power of predictive the test data for predicting class labels. Several AC techniques are presented in recent years. [22] These methods make different approaches in discover frequent item collection, extraction rules, classification rules, sorting rules; pruning waste and harmful laws - that leads to the wrong classification and classification of new educational samples. The first finding from this study is to understand the complexity of mechanisms in stock price changes. Neural networks models in recent research got success to predict the indicators. Neural network designed to predict the index of input data interrupt such interruption of economic factors, along with some of their interrupt has a better performance than neural networks that have only input index. But this situation not found when the interrupt input was removed from the index. This shows that macroeconomic variables associated with stock market indices in this study does not determine a relationship and indicators applied the most effective from their historical values. Partially adding macroeconomic variables to the model increases power of provider distribution models, and no decisive role. Contents stated above indicate that psychological climate prevailing stock market price changes. Stock market is not yet prices to determine principally but Chartyst theory is based on yesterday price changes to determine today price changes.

References

1. Dhar and Chou (2001) et al. (1996); earnings surprises,
2. Yang (1999), Zhang et al. (1999), Mckee and Greenstein; Business failure, bankruptcy, or financial health.
3. Kohzadi et al. (1996), Yao et al. (2000); Stock and commodity prices.
4. Agrawal and Schorling (1996), West et al. (1997), Aiken and Bsai (1999), Wang (1999), Jiang et al. (2000), Vellido et al. (1999) ; Selected customers, market share, packaging market and market trends.
5. Chiang et al. (1996), Indro et al. (1999); Investment Performance.
6. Wang and Leu (1996), Wittkemper and Steiner (1996), Desai and Bharati (1998), Saad et al. (1998), Qi (1999), Leung et al. (2000b), Chen et al. (2003) ; Predicted direction, Index, returns, risk, Rate of change, Futures stocks and commodity prices, efficiency and ... Stock market
7. Pang-Ning Tan, Michael Steinbach, Vipin Kumar; Introduction to Data Mining, Pearson Addison Wesley (2005).
8. Ian H. Witten and Eibe Frank, Morgan Kaufmann; Data Mining: Practical Machine Learning Tools and Techniques (2005).
9. Fadlalla, A., Lin, C. (2001). An Analysis of the Application of Neural Networks in Finance. Interfaces, Vol. 33, No. 4, 112.

10. Wong, B. and Selvi, Y.,(1998), Neural network applications in finance: A review and analysis of literature (1990-1996), *Information & Management*, 34, pp. 129- 139
11. Zhang and Hu (1998), Leung et al. (2000a), Nag and Mitra (2000); Currency exchange rate.
12. Krycha,K.,andU.Wagner.(1999).Applicationsof artificial neural networks in man- agement science: a survey, In: *Journal of Retailing and Consumer Services*, 6,185-203.
13. G. Peter Zhang, (2004). *Neural Networks in Business Forecasting*, Idea Group Publishing, USA.
14. Han J.; Pei J. Yin Y.;(2000). Mining frequent patterns without candidate gener- ation. In *Proceedings of the 2000 ACM SIGMOD International Conference on Management of Data*. Dallas, TX: ACM Press, pp. 112.
15. Li W.; Han J.; Pei J.; (2001). CMAR: Accurate and e ficient classification based on multiple-class association rule. In *Proceedings of the International Conference on Data Mining (ICDM01)*, San Jose, CA, pp. 369376.
16. Used Data from: <http://ch.saxobank.com>
17. Arezoo Aghaie, Ali Saeedi, (2009) , *Using Bayesian Networks for Bankruptcy Pre- diction: Empirical Evidence from Iranian Companies*, ISBN: 978-0-7695-3595-1
18. Hassanali,Saeid, Teimoriasl, 2007, Tehran stock exchang eindex prediction using artificial neural networks, *The Iranian Accounting and Auditing Review*,ISSN 1024- 8161
19. M.Nasiri, B.Minaei, A.Hadian, 2007,Comparison-Find Closestdistance, SV- MandC.5, Third ConferenceonData Mining, Iran University of Science and Technology
20. Thabtah, F.; Cowling P.; Peng Y.; (2005). MCAR: Multi-class classification based on association rule approach. In *Proceeding of the 3rd IEEE International Conference on Computer Systems and Applications*, Cairo, Egypt, pp. 17.
21. Thabtah F.; Cowling P.; Peng Y.; (2004). MMAC: A new multi-class, multi-label associative classification approach. In *Proceedings of the 4th IEEE International Conference on Data Mining (ICDM04)*, Brighton, UK, pp. 217224.
22. Pourhassan, Atabak, Minaei-Bidgoli, 2007, 3, *Classification using the combinedas- sociation rulesand classification*, Amir Kabir University.
23. Liu B.; Hsu W.; Ma Y.; (1998) Integrating classification and association rule mining. In *Proceedings of the International conference on Knowledge Discovery and Data Mining*. New York, NY: AAAI Press, pp. 8086.