T.C.

BAHÇEŞEHİR UNIVERSITY

CURRENCY CRISIS PREDICTION BY USING ROUGH SET THEORY

M.S. Thesis

SİBAR KAAN MANGA

İSTANBUL, 2011

T.C.

BAHÇEŞEHİR UNIVERSITY

THE GRADUATE SCHOOL OF NATURAL AND APPLIED SCIENCES
COMPUTER ENGINEERING

CURRENCY CRISIS PREDICTION BY USING ROUGH SET THEORY

M.S. Thesis

Sibar Kaan MANGA

Supervisor: Assoc. Prof. Dr. Adem KARAHOCA

İSTANBUL, 2011

T.C.
BAHÇEŞEHİR ÜNİVERSİTESİ
The Graduate School of Natural and Applied Sciences
Computer Engineering Graduate Program

Title of the M.S. Thesis : CURRENCY CRISIS PREDICTION BY USING ROUGH SET THEORY
Name/Last Name of the Student : Sibar Kaan Manga
Date of Thesis Defense : 31.01.2011

The thesis has been approved by The Graduate School of Natural and Applied Sciences.

Assoc. Prof. Dr. Tunç BOZBURA
Acting Director

This is to certify that we have read this thesis and that we find it fully adequate in scope, quality and content, as a thesis for the degree of Master of Science.

Examining Committee Members:

Assoc. Prof. Dr. Adem Karahoca (Supervisor) :

Asst. Prof. Dr. Yalçın Çekiç :

Asst. Prof. Dr. Alper Tunga :

ACKNOWLEDGEMENTS

I want to thank to my family, friends, classmates, colleagues and professors for their support, encouragement and help.

Sibar Kaan Manga

ABSTRACT

CURRENCY CRISIS PREDICTION BY USING ROUGH SET THEORY

Manga, Sibar Kaan

Computer Engineering Graduate Program

Supervisor: Assoc. Prof. Dr. Adem Karahoca

January, 2011, 61 pages

In this thesis, currency crisis and currency crisis prediction are studied for Turkish economy. The novelty of our study is using a new approach for predicting currency crisis; we used rough set theory for predicting currency crisis. While conventional approaches like “signal approach” and regression models are failing to predict currency crisis especially in out-of-sample experiments, by using rough set theory we made a decent improvement in prediction. We also tested our prediction model with various other methods that have been used or can be used for predicting currency crises. In our empirical experiments we used macroeconomic data from Turkey.

Keywords: Currency Crisis; Currency Crisis Prediction; Rough Set Theory; Data Mining; Data Mining Classifiers

ÖZET

YAKLAŞIMLI KÜME TEORİSİ KULLANARAK PARA KRİZLERİNİN TAHMİNLENMESİ

Manga, Sibar Kaan

Bilgisayar Mühendisliği Yüksek Lisans Programı

Danışman: Doç. Dr. Adem Karahoca

Ocak, 2011, 61 sayfa

Bu tez çalışmasında, günümüz ekonomilerinde çok önemli bir etken olan para krizleri ve para krizlerinin tahmini üzerine çalışılmıştır. Çalışmada daha önce yapılan benzer konulu çalışmalardan farklı olarak yaklaşımlı küme teorisini kullanılmıştır. Daha önceki çalışmalarda kullanılan sinyal yaklaşımı ve regresyon modelleri gibi geleneksel yöntemlerin özellikle örneklem dışı tahminlerde başarılı sonuçlar alamadığı bilinmektedir. Yaklaşımlı küme teorisini kullanarak, para krizi tahminleme işlemlerinde önemli gelişmeler kaydedilmiştir. Ayrıca yaklaşımlı küme modeli Türkiye'nin macroekonomik verilerini kullanarak test edilmiştir. Yaklaşımlı küme teorisi için yapılan amprik deney sonuçları ve farklı veri madenciliği metodları ile yapılan testlerin sonuçları birlikte değerlendirilmiştir.

Anahtar Kelimeler: Para Krizi; Para Krizi Tahminlemesi; Yaklaşımlı Küme Teorsi; Veri Madenciliği; Veri Madenciliği Sınıflandırma.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

LIST OF ABBREVIATIONS

Adaptive Neuro Fuzzy Inference System	: ANFIS
Artificial Intelligence	: A.I.
Artificial Neural Network	: A.N.N.
Currency Crises	: C.C.
Neural Network	: N.N.

1. INTRODUCTION

Currency crisis is in brief the sudden loss of a country's currency value against foreign currencies. But even the description of the currency crisis is simple; the underlying mechanisms that cause currency crisis are very complicated.

This sophisticated nature of currency crisis result in very hot topic of research of currency crisis and currency crisis prediction. This thesis is focused on currency crisis prediction with other methods then mostly used regression and logistic models but by means of data mining methods especially rough set theory. Rough set theory is a new approach in data mining. Rough set applications on financial forecasting and crisis predictions are not many but they have promising results for further researches like ours.

If we return to the currency crisis issue; currency crisis have various reasons, the loss of foreign exchange reserves, capital outflow, sovereign debt crisis, banking and financial sector problems can also cause currency crisis, even speculative attacks of some investors or manipulators can cause C.C. (Currency Crisis). Also the loss of competitiveness of a country's exports can force for a country into a currency crisis. But generally there is more than one cause in currency crisis.

The loss of value of one country's currency usually has a sudden devastating effect on the economy of the country. If people got panic during the swinging of exchange rates, people could try to change their local currency into a foreign currency. Such a move will increase the impact of the c.c. or people can try to withdraw their savings and deposits from banks in an effort to meet their demands causing imbalance and cash shortage of banks.

In this thesis our goal is to build a model to predict possible currency crisis by using macro-economic data of countries and the rough set methodology.

By predicting currency crisis we can get many benefits ranging from generating new regulations to prevent the currency crisis to giving investors time for hedging their investments. Governments can use the prediction mechanisms to project the result of their policies and can make changes on the policies to prevent currency crisis, the money authorities like national central banks can increase their savings of foreign currencies or calibrate interest rates. For investment funds and personal investors a currency crisis can have the possibility of large gains during the impact. As the market floats there can be various opportunities for selling or buying currencies of different kinds to make profit.

In this thesis the reason why we decide to use Macroeconomic date is basically macroeconomic data can be considered as the mirror of the country's economic condition and can hold sophisticated information within.

Rough set theory is chosen because of its high applicability and high performance in handling noisy and uncertain data.

There are many researches and literature in currency crisis as well as c.c. Prediction. Especially during the periods when currency crises are dense, many researchers work on this topic.

Currency crisis prediction is also a multidisciplinary field of research. In the past decades many researches of economists and mathematicians are done. Also there are many researches made by computer scientists and computational scientists. Computer scientists used AI or data mining tools and methodologies to predict currency crises. Some of those researches and some brief information about them are mentioned in the following literature review section.

2. LITERATURE REVIEW

In this section we will review the literature about currency crisis. We will review the methodologies and findings of the important researches on the topic, to get a better understanding of the studies, we will review the studies in two parts, and first part is to reflect the view of economists who work on the subject and in the second part the computer scientists and engineers approaches.

Currency crisis prediction and early warning systems are hot topics for a long time. The number of researches raised in number during 90's and late 90's.It is no coincidence that Asian, Russian and South American currency crushes happened around the same years.

There are many approaches to predict currency crisis. One of the novel currency crisis prediction approaches is signal approach. Kaminsky, Lizanda and Reinhart developed a signal based early warning system also known as KLR signal model. In this model they evaluate many economic indicators that they believe that indicators are acting differently before crisis. Every time an indicator exceeds a certain threshold that count as a signal of approaching currency crisis. They also tried to find which indicators are better or more useful in predicting the crisis period. These approaches got useful results and have potential for future expansions. The results also showed that with a broad band of indicators it is easier to predict crisis (Kaminsky et al., 1998).

Kaminsky continued researches on currency crisis. And they also worked on the existing models and categorized them in to 3 generations.

• The first generation of models focuses on fiscal and monetary causes of currency crisis.

• The second generation models aimed at EMS crisis. They focus on the effects of counter cyclical policies in mature economies and on self-fulfilling crisis with rumors unrelated to market fundamentals at the core of the crisis.

• The third model is based on moral hazard and imperfect information.

And stated all 3 generations have various short fallings and proposed a regression-tree model with indicators. So they combine many indicators to predict currency crisis. As a result of their researches, they found out that a regression tree model has good results in prediction but all crises are different in some way so a static method will fail to achieve accurate performance (Kaminsky, 2006).

Despite all the praise and promise of indicator signal approach many researchers criticized this approach. As an example Goldfajn and Valdes criticized KLR approach with data and evidence from Mexico and Thailand currency crisis. They showed that the signals were not accurate. In their work they also suggested they found a solid correlation with the stock market and real exchange rate. As conclusion they stated that currency crisis cannot be predicted because expectations cannot ve predicted (Goldfajn I. & Valdes R. O., 1998).

In general the researchers are focused on crisis and crisis free zones but the main goal is the crisis period but some researchers focused on crisis free zones so called safety zones. K. Osband and C.v.Rijchegens first they outlined the safety zones and try to figure out the crisis in the rest of the periods by doing so; they suggested they get better results. The results of their implementation and empirical work, they got a highly reliable model but they also state some short comings of their system like falsely labeling some vulnerable economic periods as safe because of the models statements (Osband and C.v.Rijckeghem, 2004).

Another research criticizing KLR is an empirical work on Argentinian currency crisis. Researchers implemented the indicators signal approach but cannot get positive results. And they suggested the expectations and political turbulence like non-macroeconomic variables are not included in indicators and those variables may have larger effect than expected.

There are also some works which evaluates KLR approach and proposed an alternative approach; a general probit-based model for predicting currency crisis. In their empirical studies they got better results but noted that their assumptions and the ignoring the lag of data may affect the out came and there must be more work on them (Berg A. & Pattile C., 1999).

There are works on early warning systems from other perspectives. A research on the lessons learned from previous works on KLR, logit model definition of crisis and approaches improve accuracy of prediction. By comparing many models they calculated that with the changing parameters, time scope or the definition of crisis, some different model performs better so there is no single best model, second they noted the contagion factor and third the out of sample work got disappointing results (Beckmann D. & Menhoff-Katja L., 2006).

One of the important properties of currency crisis is that currency crisis generally happens in emerging market economies, so studies especially on these countries carry on important role. One of this kind of research Kumar et al. Worked on a logit model with the data of 32 developing countries they also focused on out of example which is not common.as a result for this empirical work they got good results.

There are also some works done about extending the logit model researches. They used multinomial logit models for predicting financial crisis. They developed a new method that can distinguish crisis with post-crisis periods which previous early warning systems cannot. Their application of binominal logit model, they improved the accuracy of the crisis prediction considerably boy in-sample and out-of-sample experiments (Bussiere M. & Fratzscher M., 2006).

There are several researches focusing on emerging countries, one of those is BurKart and Coudert's work. They used a linear model called Fisher's linear discriminant analysis for predicting currency crisis. They also included contagion and indicators for banking sector problems which are known for their effect on causing currency crisis. They got important results and found out the importance of overvaluations in currency crisis.

Another approach for predicting currency crisis is by using future and option pricing. The reason behind this is the effect of expectations on currency crisis. Maltritz and

Eichler argued that traditional economic variables are the data of past but to accurately predict a future currency crisis they are not so capable. But the expectations which are priced in future and option markets are the key. In their research they used American depository receipts (ADR) and calculated the price spreads of the underlying stocks. When an origin or underlying stock and cross-listed, ADR prices differentiate this is considered as a possible correction in the money market or in the other words a currency crisis. In conclusion researchers get reasonable good results in predicting currency crisis (Maltritz D. & Eicher S., 2009).

Another research on financial crisis of emerging markets is done by Özlale and Metin-Özcan, this research is also using Turkey in empirical studies as we will do in this thesis. And they used extended kalman filter for prediction (Özlale Ü. & Metin-Özcan K., 2007).

2.1 COMPUTER SCIENCE LITERATURE ON CURRENCY CRISIS

In the previous section of the part there were the reviews of the some important works on currency crisis done by mostly by economists.

Now we continue with the literature review for currency crisis in a different aspect; the researches on currency crisis in the area of computer science.

Computer scientists interested in economical prediction like currency crisis prediction. They used many different kinds of data but the main difference is the methods they apply. Computer Scientists used artificial intelligence, data mining and like mathematical solutions to predict possible currency crisis.

After the researchers began to apply the currency crisis problem to AI Researchers began to try new ways for better adopting their technic and problem. One of these types of adaptions is the hybrid systems of neural networks and fuzzy logic. As we know the learning capacity of neural networks is very useful on this kind of problems. But to increase the effectiveness of NN researchers used fuzzy logic between the neural networks and knowledge base as an interface. They tested their new hybrid system in empirical studies. The results of their experiments were promising, their hybrid model got about 80 percent successes in predicting currency crisis (Lin et al, 2008).

There are other researches of ANN with fuzzy interface systems. One of them Lee et al.'s work for example specialize in predicting foreign exchange rates and make decisions. And use all this capacity in real time on money market (Lee V.C.S. & Wong H. T., 2007).

One of the other useful surveys of early warning models is done by Kim et al. They have two focuses in that research first they work on the training mechanism of their models and second the test and benchmark of several EWS like logistic discrimination , decision tress , support vector machine , neuro-fuzzy model and Artificial neural networks and they apply those methods mentioned on Korean data. In their testing process they found out decision tree and ANN fits best for Korean economy and SVM's over fitting caused problems in unstable, swiftly changing environments (Kim et al., 2004).

There are more specialized models for neuro-fuzzy interfaced neural networks the granular data mining. The use of granular data based granular neural networks has many benefits like handling the uncertain and noisy data of real life market data and economic data. In the research of Zhang et al. They proposed a new model called statical fuzzy interval neural networks SFINN. In conclusion including random factor by using statical estimation is very helpful in prediction (Zhang et al., 2007).

A deeper research in neural networks is done by L. Anstasakis et al. In their research they used both parametric and non-parametric self-organizing methods to overcame their individual difficulties and get a better model to predict exchange rates. Their method is called group method of data handling. GMDH by combining neural networks with GMDH the result is more successful prediction model then the previous two models (Anastasakis L. & Mort N., 2009).

Applying neural networks on time series data to obtain a currency crisis prediction model can be seen in various researches with some core different approaches. In the research of J. Yoa et al, the researchers focused on time series data and mining rules out from model. In conclusion the results were not 100 percent reliable especially model fails to explain yen while explaining both USD and British pound (Yao J. & L. Tan C., 2000).

One of the important researches for this thesis is done by Tay et al., In their work they used rough set theory as we do in this thesis. And they applied RST on predictions of economic and financial matters. In conclusion they stated there can be many works done to improve the performance of RST (Tay & Shen, 2002).

Some researchers use other data mining methods too, Sun and Shenoy used naïve bayes bayesin network for predicting banking sector crises, and they achieved comparable success to past researches in the literature (Sun & Shenoy, 2007).

Ahn et al. used support vector machine (Vapnik, 1998) to develop an early warning system for financial crises. They used Korean data for testing purposes (Ahn et al, 2011).

Many other researchers used various data mining methods for prediction, financial early warning systems and many more important issues. Some of these researches are; Lanine and Vennet used logistic regression for predicting the failure of Russian banks and financial establishments (Lanine & Vennet, 2006). Xu and Wang worked on multiple discriminant approach, support vector machines and logistic regression (Xu & Wang, 2009) and Bonafede and Giudici used Bayesian Networks' abilities in integrating different information and tried to contribute to risk calculation literature (Bonafede & Giudici, 2007).

3. MATERIAL AND METHODS

In this section the data and technics that are used in this thesis are introduced.

To develop and apply our methods we used the macro economic data of Turkey. And As method the “rough set theory” is used. We also used some other methodologies to test the success of rough set theory and benchmark proposes.

First we will discuss the macroeconomic variables which we used and then show the scope of those data.

Second we will review the theoretical background and the details of rough set theory and other methodologies, so we can compare the abilities, promises and drawbacks of the methods.

3.1 DATA USED IN THESIS

In this study various kinds of macro-economic data are used, so we can project and explain as many reasons as possible to predict the currency crisis as accurate as possible.

3.1.1 The Real Economy Confidence Index

These indexes are created by interviews with the businesses in the real economy sectors. This indicator is a solid mechanism about the expectations of the venture holders and investors and their attitude about investments when this index began to rise or decrease the number and amount of the investments will follow a similar direction. This indexes start at 100 point and move up or down as the confidence level changes, in our data set the max value of this confidence index is 121.2 and min. value is 52, the missing values are filled with 100 (Fig. 3.1).

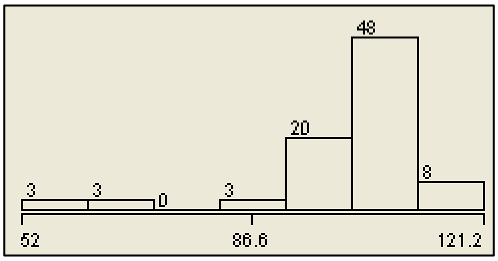


Figure 3.1: Real economy confidance index value distribution

3.1.2 The Amount of Bonds and Securities and Market Stocks That Foreign Settled People Hold

This data can also be considered as a trust indicator about the countries and its economic performance. Also we can note that as the world becomes more global or smaller, it became easier to move money because of this reasons, this index can change its attitude very suddenly and unexpectedly so the models with bigger time window cannot benefit from them fully.

This indicator shows the amount in million dollars, the missing data is filled with 0 the range of the data can be seen in Fig. 3.2 and we can see that the missing value is close to zero from the Fig 3.3.

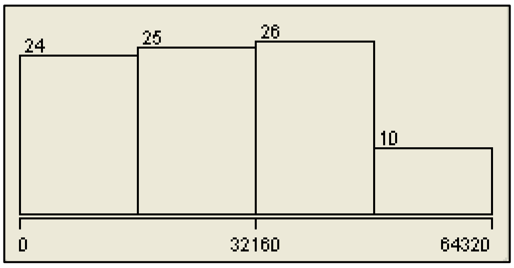


Figure 3.2: Distribution of the amount of bonds and securities and market stocks that foreign settled people hold

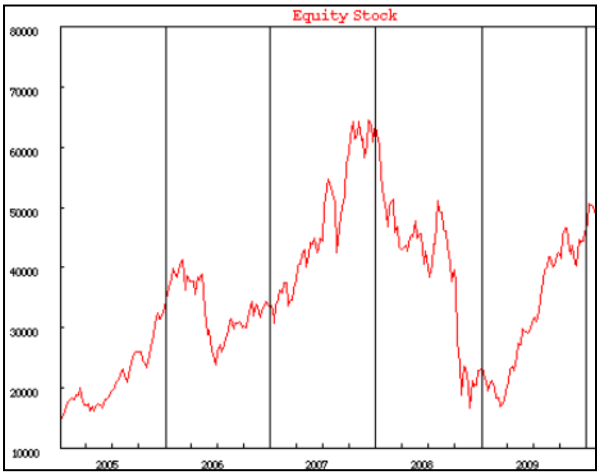


Figure 3.3: Graph amount vs time
Source: http://evds.tcmb.gov.tr/yeni/cbt-uk.html

3.1.3 The Ratio of Employment

The ratio of employment is one of the important measures of the economic position of the country although good economic performance won't always create new jobs because of the high utilization and automation in production, but the bad economic performance and expectations of fall in economic sense so the number of jobs loss are increase.

The missing values are filled with zero and the distribution of employment data can be seen in Fig 3.4.

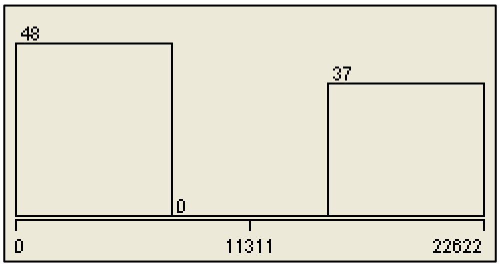


Figure 3.4: Distribution of the ratio of employment

3.1.4 Industrial Production Index

This index is another important indicator of economic performance of countries. As economy slows down and stocks increase, the production decrease and this starts a chain of events like loss of jobs, decrease in spending, loss of taxes as this trigger each other each of them hurt the economy of the country.

Industrial production index is set to 100 at 2005 the missing values are set to 100 too. General information about the index values are shown in Fig 3.5.This index is the relative increase in production to the base year.

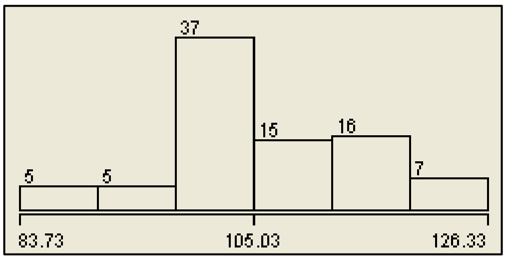


Figure 3.5: Distribution of industrial production index

3.1.5 The Number of New Opened, Closed and Bankrupt Business

Under normal conditions, there are many new businesses or some businesses fail to make money and close or go bankruptcy but in growing economies the number of new entrances are larger in number and a positive result will be seen from consolidating the numbers but when economic performance is bad the total number of businesses fall.

In this study we used the monthly data of newly opened and closed business numbers. The number of newly opened businesses range from 1692 to 6001 while the number of closed businesses ranges from 262 to 1942 as we can see from Fig. 3.6a and 3.6b.

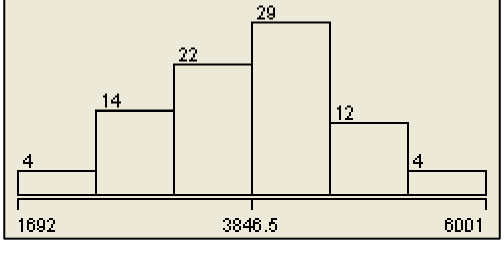


Figure 3.6a: Distributions of new opened businesses

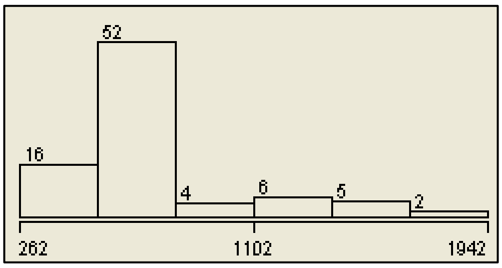


Figure 3.6b: Distribution of the closed businesses

3.1.6 Volume of Gold Market

Volume of gold market is an economic indicator unlike the other indicators mentioned before like production or employment indexes. During crisis or risky periods people invest into safe harbors like gold to protect their savings.

The unit of our data is USD/ONS. And there is a large gap between the top amount 1.5 billion and bottom 62 million (Fig 3.7).

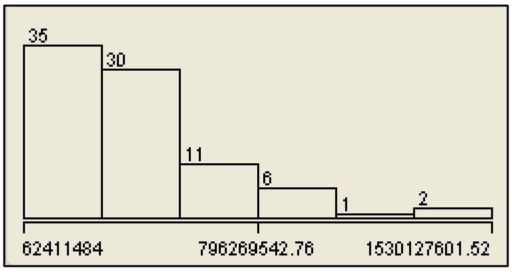


Figure 3.7: Distribution of volume of gold market

3.1.7 Value of the Stock Market

As economy paces the value of stocks will increase in expectations of higher profits or higher stock values and this will increase the new investors into stock markets. And with increasing investments, companies can have more funds to spend in new production and innovation investments.

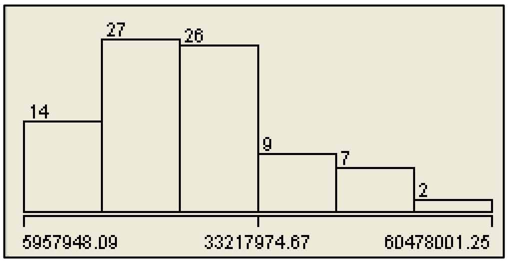


Figure 3.8: Distribution of value of the stock market

The stock market value data is used as continuous data except the data mining methods that we used like rough set and ID3. Data structure is shown in Fig. 3.8 and the change of value is large on different time periods (Fig. 3.9).

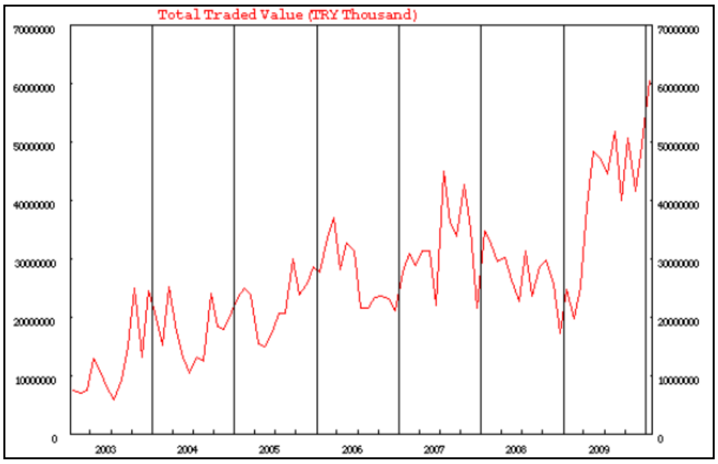


Figure 3.9: Value of stock market in time
Source: (http://evds.tcmb.gov.tr/yeni/cbt-uk.html)

3.1.8 Capacity Usage of Industry

Another indicator to consider the position of industry in order to determine the possible production ratio. Low capacity usage is a signal for bad economic period.

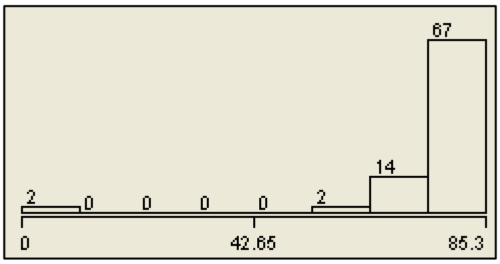


Figure 3.10: Distribution of capacity usage of industry

The missing values are filled with zero and most of the data is close to the max value 85.3 as we can see from Fig. 3.10.

3.1.9 Number of Strikes in Private Sector

This is a different indicator; this data can show the position of lower waged employees if there are many strikes the labors are unhappy with the current economic conditions or new regulations that they consider loss of their rights or earnings.

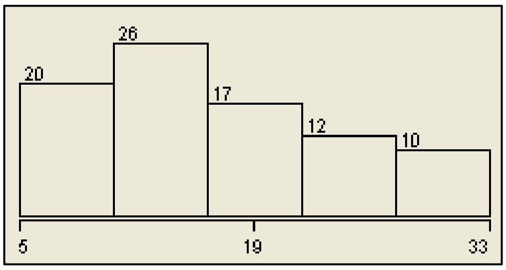


Figure 3.11: Distribution number of strikes in private sector

The number of strikes can be inspected and it can be seen that much data are clustered in relatively lower values as shown in Fig 3.11.

3.1.10 Export Index

Export index is the main source of a country's foreign currency reserves, a dramatic fall of exports if caused by loss of competitive advantage of cheap currency usually results in a currency crisis.

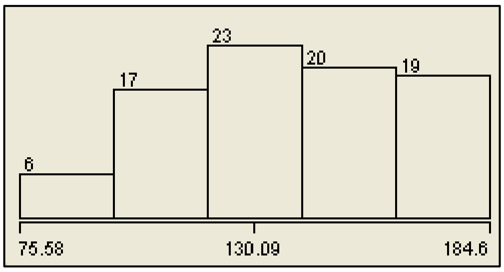


Figure 3.12: Distribution of Export index

The export index is started with 100 and the relative change is calculated. A value of 80 means the 80 percent of production of the base year. And we can see the max and min values of index from Fig 3.12.

3.1.11 The Interest Rate

the interest rates show the price of the money, if interest rates are low people are more likely to invest into new ventures or loan money from banks otherwise they will prefer the income of interest rates of deposited money in banks.

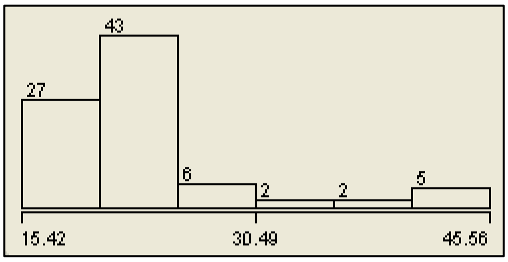


Figure 3.13: Distribution of the interest rate

Max and min values of interest rate can be seen in Fig 3.13; we can also inspect the number of months which has a certain interest rate.

3.1.12 Inflation Rate

A high inflation rate is always an indicator of problems of economies. High inflation can encourage the people to invest into foreign currencies.

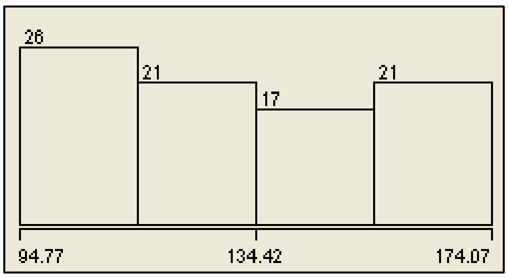


Figure 3.14: Disribution of inflation rate

Inflation rate is another index like variable. It starts at 100 and moves up and down with respect to base and previous inflations rates of months. In Fig 3.14 we can see the max and min values of inflation rate.

3.1.13 The Reserves of Central Banks

This data is one of the key data of our dataset because the general demand for foreign currencies is met by central banks. If reserves dry, people will not be willing to sell their foreign currency so the value of that currency increases this could trigger a currency crisis.

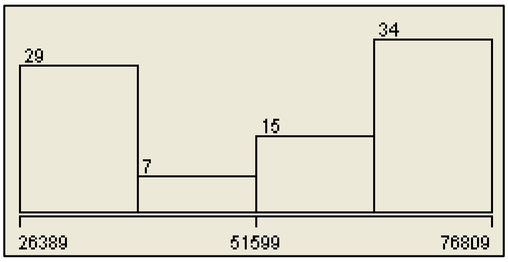


Figure 3.15: Distribution of The reserves of central banks

The maximum and minimum reserve values of central bank in foreign currency are shown in Fig. 3.15.

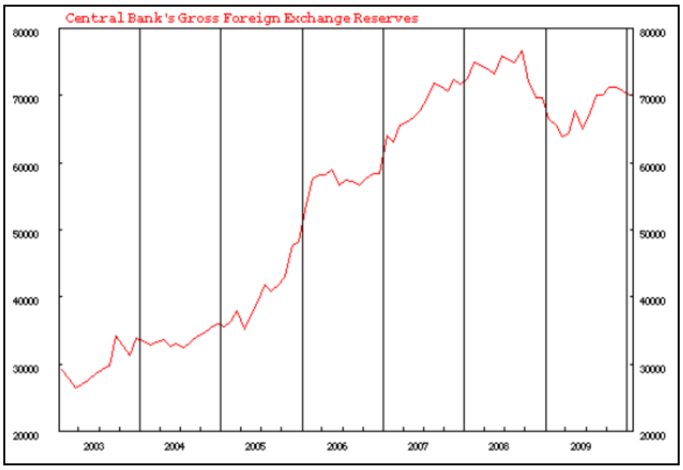


Figure 3.16: Central Bank's Foreign Exchange reserves
Source: (http://evds.tcmb.gov.tr/yeni/cbt-uk.html)

In figure 3.16 we can see the reserves change in time. We can see that during our scope, the reserves grow more than 2 times.

In this research we used the data of Turkey between 2003-2009 years; turkey is an emerging market showing descent performance in resent economic crisis of USA and EU. Turkey as a result of its geographical position is a mixture of western and eastern economies.

3.2 DATA SET

Our data set consist of the data mentioned above for the years between 2003 and 2009. There are many different approaches in defining currency crisis, one of those approaches is using the deviation of exchange rates like Kaminsky et al (1998), another approach is to set a value as crisis as used by Kumar et al(2003). We chose the second way and define currency crisis as an increase in USD exchange rate equal or above 4.5 percent for a month considered as a currency crisis. A lower value will produce too many crisis periods while a higher value will be producing too few crisis periods. Because of the nature of the data which almost all of our data set is announced at the end of every month, we lagged our decision attribute or crisis data for one month (for example: we used the data of March to predict crisis of April.) So the prediction became more realistic.

3.3 METHODOLOGY

In this thesis to implement our data mining technics, we choose to use RSES2, rossetta and Weka.

3.3.1 Rough set theory

Rough Set theory is one of the new areas of data mining. Rough set theory is developed by Zdzislaw Pawlak in the early 80's (Pawlak Z., 1982). In general it is an extension to set theory. The main idea of rough set analysis is induction of approximations of concepts. In other words, if we consider every member of a set has some knowledge about the set by decreasing the crispiness of one set, we can increase our ability to find out patterns hidden in the set.

Now let's have a deeper look at rough set theory and its mathematical background.

Information System (IS) can be shown as a pair (U, A) where U is a non-empty finite set of objects and A is a non-empty set of Attributes such that

a: U -> Va for every $a \in A$.

Va is called the value set of a.

A decision system, DS, is T = (U, A U {d}) and $d \notin A$, d is not an element of attributes but a decision attribute. A result for the attributes and the elements of A are called conditional attributes.

One of the most important properties of rough set theory is the indiscernibility

with and p<= A there is an associated equivalence relation IND (P) as seen in formula 3.1.

$$\text{IND(P)}=\{(x,y) \in U^2 | \text{ Va} \in \text{P, a(x)=a(y) }\} \qquad (3.1)$$

where IND(P) is called P-indiscernility relation.

The partion of U is a family of all equivalence classes of IND(P) and is denoted by formula 3.2.

A Partition of U is U/ IND(P) (3.2)

if (x,y) $\in$ IND(P) then x and y are indiscernible objects may be represented several times.

Table 3.1: Example data set

Object	Age	Doing Sports	Blood pressure
01	15-25	low	no
02	25-45	low	yes
03	25-45	medium	no
04	45-70	low	no
05	45-70	low	yes

In this example we can see the attributes are {age}, {doing sports}, {age - doing sports}

IND ({age})={{O1},{{O2}{O3}},{{O4},{O5} }

IND ({doing sports})) ={{{O1},{O2},{O4},{O5}},{O3}}

IND ({{age}, { doing sports}})={{O1},{O2},{O3},{{O4},{O5}}}

so for P=({age},{ doing sports})

the objects O4 and O5 cannot be distinguished from each other.

- An equivalence relation induces a partitioning of the universe
- Partitions can be used to build new subsets of universe
- Subsets that are most often of interest have the same value of decision attributes

It may happen, however that a concept such as high blood pressure cannot be defined in a crisp manner.

Approximation

Let T= (U, A) and let $B \in A$ and $X \subseteq U$ we can approximate X using the only information contained in B by constructing the C-lower and B-upper approximation of X, denoted $\underline{B}X$ and $\overline{B}X$ respectively, in formula 3.3 and 3.4

$$\underline{B}X=\{X \mid [X]_B \subseteq X\} \tag{3.3}$$

$$\overline{B}X=\{X \mid [X]_B \cap X \neq \emptyset\} \tag{3.4}$$

B-boundary region of X which is shown in formula 3.5 consist of object that we cannot classify X in B for sure.

$$BNB(X) = \overline{B}X - \underline{B}X \tag{3.5}$$

B-outside region of X, U- $\overline{B}X$ consist of objects that certainly classified as not belong to X.

A set is said to be rough if its boundary region is non-empty, otherwise the set is crisp.

Definability

In general upper and lower approximations are not equal; in such cases we say that target set X is indefinable or roughly definable on attribute set P. when upper and lower approximations are equal $\overline{P}X = \underline{P}X$ than X is definable on attribute set P.

There may be cases like these

• Set X is internally definable if $\underline{P}X \neq \emptyset$ and $\overline{P}X$ =U this means all objects in the universe can be defined to X in other words there are no objects we can definitely exclude from set X.

- Set X is externally definable if $\underline{P}X=\emptyset$ and $\overline{P}X \neq U$. in this case there are no objects which we can be certain about belonginess to set X , but there are objects we can definitely excluded from X

- Set X is totally indefinable $\underline{P}X=\emptyset$, $\overline{P}X = U$ neither there are no objects we can be certain to belong to set X nor there are no objects defiantly unbelong to X

Dispensable and Indispensable Attributes

Let $c \in C$. Attribute c is dispensable if $POS_C(D) = POS_{(C-\{c\})}(D)$ otherwise attribute c is indispensable in T.

Independence

T= (U, C, D) is independent if all $c \in C$ are indispensible in T.

Reduct and Core

The set of attributes $R \subseteq C$ is called a reduct of C, if T= (U, R, D) is independent and $POS_R(D) = POS_C(D)$

The set of all condition attributes in dispensable in T is denoted by Core (C) as shown in formula 3.6.

$$CORE(C) = \cap RED(C) \tag{3.6}$$

where RED(C) is he set of all reducts of C.

Accuracy of Approximations

Accuracy of approximation is shown in formula 3.7 where |X| denotes the cardinality of $X \neq \emptyset$.

$$\alpha C(V) = \left|\underline{B}(X)\right| \Big/ \left|\overline{B}(X)\right| \tag{3.7}$$

Obviously;

0<= α B<=1. If α B(X) = 1 then X is crisp with respect to 0, if $\alpha B(X) < 1$, X is rough with respect to B.

Those are the basic properties or rough sets.

Now let's study a small example on how rough sets can be used for decision making and prediction proposes.

There are various approaches for generating decision rules for rough sets some can be listed as,

filter approach: in this approach we try to choose the minimal subsets of attributes and select the higher ranked attributes.

The wrapper approach there are $2^n - 1$ possible subsets of attributes so the time complexity is high and can be used search and non-deterministic and heuristic methods in creating decision rules (International Rough set society, http://roughsets.home.pl/IRSS/rs-kdd/sld001.htm, http://roughsets.home.pl/IRSS/RSCTC_02_final_pliki/v3_document.htm).

There are extensions for rough set theory. The dominance -based rough set approach is built for assisting multi criteria decision analysis introduced by Greco Matarazzo and Slaminski in this extinction the indiscernibility relation is changed with a dominance relation.

Another extension is the decision theoretical rough sets is a probabilistic extension of rs classification createed by Yiyu Y.O. The extension makes use of loss functions to derive Alfa and Beta region parameters.

Rough sets can also be used with fuzzy sets.

A simple rough set algorithm example:

Assume that our information system is:

Table 3.2: Example information system

Object	Attribute A	Attribute B	Attribute C	Decision
01	No	Yes	High	Yes
02	No	Yes	Normal	Yes
03	No	No	Low	No
04	Yes	Yes	High	Yes
05	Yes	No	Normal	No
06	No	No	Low	No
07	Yes	Yes	High	Yes
08	No	Yes	High	No
09	No	No	Normal	No
10	Yes	Yes	Low	Yes

Table 3.3: Attributes

Attribute A	Yes, No
Attribute B	Yes, No
Attribute C	High, Normal, Low
Decision	Yes, No

Indiscernibility relationships:

INDA(AttributeA)={{01,02},{03,06,08,09},{04,07,10},{05}}

INDA(AttributeB)={{01,02,04,07,10},{03,05,06,09},{08}}

INDA(AttributeC)={{01,04,07},{08},{02},{05,09},{03,06},{10}}

Approximation

Lower Approximation of the set

$\underline{B}$Set={04,07,10} for yes

$\underline{B}$Set={03,05,06,09} for no

$\overline{B}$ - set upper approximation = {01, 02, 04, 07, 08, 10}

$\overline{B}$ - set = {01,02,03,05,06,08,09} for no

Boundry is {01,02,08}

With this information we can reduce our data set as follows:

03	No	No	Low	No

04	Yes	Yes	High	Yes

05	Yes	No	Normal	No

09	No	No	Normal	No

10	Yes	Yes	Low	Yes

As we continue our analysis

Attribute A and Attribute B with decision

Table 3.4: Attribute A and attribute B with decision

03	No	No	No
04	Yes	Yes	Yes
05	Yes	No	No
09	No	No	No
10	Yes	Yes	Yes

Attribute A and Attribute C with decision

Table 3.5: Attribute A and attribute C with decision

03	No	Low	No
04	Yes	High	Yes
05	Yes	Normal	No
09	No	Normal	No
10	Yes	Low	Yes

Attribute B and Attribute C with decision

Table 3.6: Attribute B and attribute C with decision

03	No	Low	No
04	Yes	High	Yes
05	No	Normal	No
09	No	Normal	No
10	Yes	Low	Yes

It can be seen that with only using Attribute B we can reduce the system without losing any information so the rule will be like this:

Table 3.7: Attribute B and decision

03	No	No
04	Yes	Yes
05	No	No
09	No	No
10	Yes	Yes

Attribute B and Decision

if Attribute B=Yes

then D=Yes

Else

D=No

3.3.2 ID3

ID3 (Iterative Dichotomiser 3) is an algorithm used to build a decision tree developed by Ross Quinlan (R. Quinlan, 1986).

Steps of ID3 algorithm can be listed as:

1. Take all unused attributes and count their entropy concerning test samples
2. Choose attribute for which entropy is minimum
3. Make node containing that attribute, and return to 1st step till there is no more attribute.

To test our data using WEKA's ID3 implementation (R. Quinlan, 1986), we first discretized the data by using RSES2's discretization tool. Then using the 66 percent of our data we build our decision tree (Fig. 3.17).

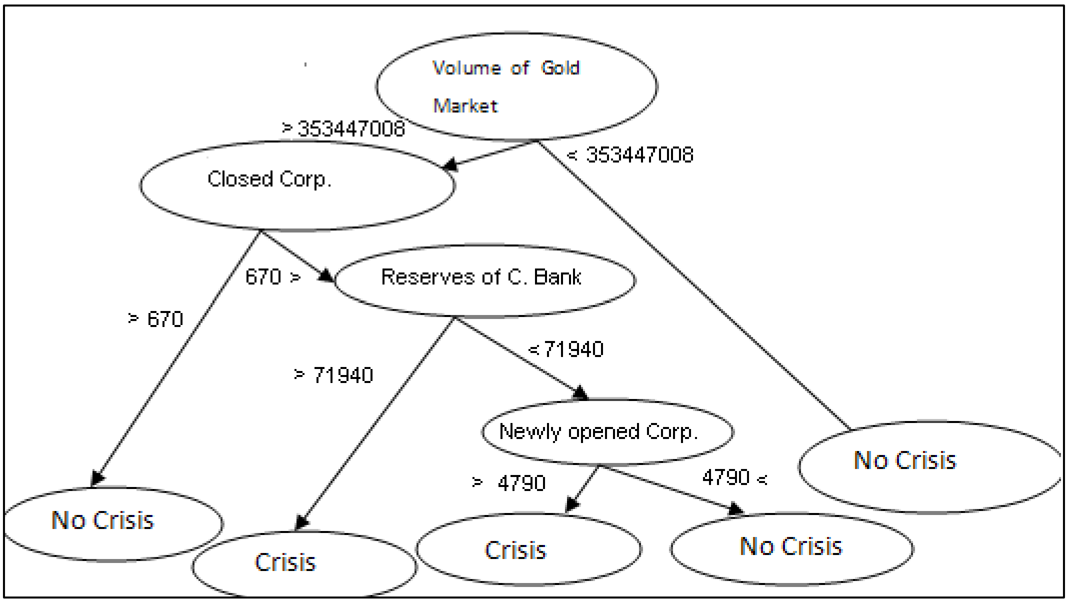


Figure 3.17: ID3 desicion tree

After finishing the decision tree, we check the remaining 34 percent of the data and label them as crisis and no-crisis periods.

3.3.3 BayesNet

Bayesian networks (BNs), also known as belief networks (or Bayes nets), are one of the probabilistic graphical models. These graphical structures are used to represent knowledge about an uncertain domain. Each node in the graph represents a random variable, while the edges between the nodes represent probabilistic dependencies between the corresponding random variables (Fig. 3.18) (Ben-Gal I., 2007).

These conditional dependencies in the graph are often estimated by using known statistical and computational methods. B.N.'s combine graph theory, probability theory, computer science, and statistics.

We used our continuous data set, Bayes Net can work with both continuous and discrete data, to train BayesNet, and build the probability tables and graph structure with 66 percent of our data which is called training data. Than we used the remaining 34 percent of our data as test data and note the prediction results.

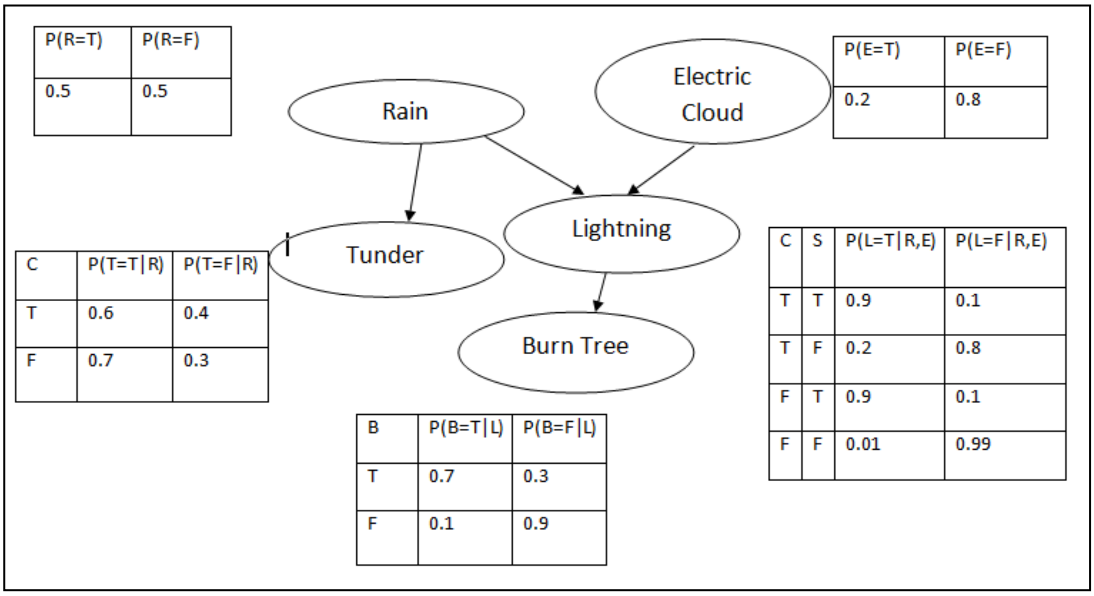


Figure 3.18: Bayesian Networks

3.3.4 Naive Bayes

Naïve Bayes classification is a simple probabilistic classification based on Bayes' theorem with strong (naive) independence assumptions. This model is also known as "independent feature model" (John & Langley, 1995).

```
Naive Bayes Classifier

                          Class
Attribute                     0               1
                         (0.91)          (0.09)
==============================================
Altin
  mean             401198058.3864 894666572.3864
  std. dev.        256433280.8613 417616246.8778
  weight sum                   52              4
  precision         26508639.1818  26508639.1818

enflasyon
  mean                   128.8129       140.1892
  std. dev.               22.9839        13.8839
  weight sum                   52              4
  precision                2.0541         2.0541
```

Figure 3.19: A part of naive bayes classifier

In naïve bayes method we first build the classifiers (as seen in Fig. 3.19) with our training data which consist of 66 percent of our data and then test the model using rest of the data, the test instances are classified according to their likely hood to the classifier.

3.3.5 Logistic Regression

Logistic regression is a type of linear regression, when the result of our variables form a binominal function, we use linear regression but when the function is a logit function like Fig. 3.20. We use logistic regression (Rouhani-Kalleh, O. 2006).

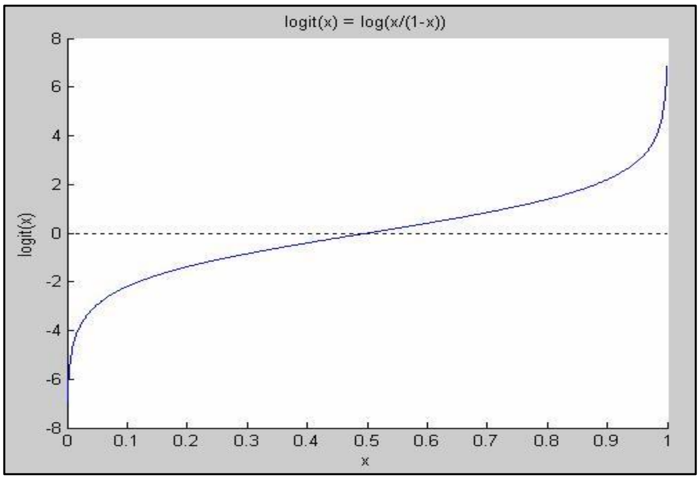


Figure 3.20: Logit function (source:http://omidrouhani.com/research/logisticregression/html/ logisticregression.htm#_Toc147483480 figure4.)

Logistic regression is one of the best known models for prediction purposes. Many researchers used this model for predicting currency crisis that is why we also use this approach for testing and benchmarking.

As we know that logistic regression is a function, we build the Coefficients and odd ratio in order to classify our test objects. Then we calculate which class has a higher possibility. We mark our test object to the class with greater possibility (le Cessie et al, 1992).

3.3.6 JRip

JRip is a rule learner algorithm, Repeated Incremental Pruning to Produce Error Reduction (RIPPER), which is proposed by William W. Cohen as an optimized version of IREP (Cohen, 1995).

We used our data set to build rules and then implement the rules to the test set to get results.

3.3.7 Conjuctive Rule

This method implements a single conjunctive rule learner that can predict for numeric and nominal class labels.

A rule consists of antecedents "AND"‘ed together and the consequent (class value) for the classification/regression. In this case, the consequent is the distribution of the available classes (or numeric value) in the dataset. If the test instance is not covered by this rule, then it's predicted using the default class distributions/value of the data not covered by the rule in the training data. This learner selects an antecedent by computing the Information Gain of each antecedent and prunes the generated rule using Reduced Error Pruning (REP).

For classification, the Information of one antecedent is the weighted average of the entropies of both the data covered and not covered by the rule. For regression, the Information is the weighted average of the mean-squared errors of both the data covered and not covered by the rule.

In pruning, weighted average of accuracy rate of the pruning data is used for classification while the weighted average of the mean-squared errors of the pruning data is used for regression (WEKA software guide.).

3.3.8 SMO

SMO is an Implementation of John C. Platt's sequential minimal optimization algorithm for training a support vector classifier using polynomial or RBF kernels. This implementation globally replaces all missing values and transforms nominal attributes in our training data into binary ones. It also normalizes all attributes by default. Multi-class problems are solved using pairwise classification. In the multi-class case the predicted probabilities will be coupled using Hastie and Tibshirani's pairwise coupling method (J. Platt, 1998).

In WEKA we first run our training and let the SMO to normalize and weight the attributes than we use our test set 34 percent of our total data to get prediction results for comparing with other methods.

4. FINDINGS

The implementation of Rough Set Theory in this study can be handled in four phases. First is the discretization of our information system so the continuous values are changed into discrete values.

Second phase is reduction. In this phase information system is reduced into a simpler form, so we can mine the rules much more effectively. The attributes that don't affect the out came which are independent of the decision attribute can be removed.

Third phase is mining the rules. We use our reduced information system to mine rules. At this point too general rules can cause much more false negative or false positive outcome but a too specialized rule may never be triggered.

The fourth and the last phase is the test of our test data the remaining 34 percent from the information system after using the 66 percent in training.

4.1 DISCRETIZATION PHASE

In this phase the data set is discetized and all the continues data are converted into discrete data. First the cuts (Fig. 4.1) are produced. Using the cuts the data is converted.

Cut set: IS_currency1

(1-14)	Attribute	Size	Descrip...
1	Altin	1	3.53447008E8
2	enflasyon	1	106.5
3	Grev	1	12.5
4	ihracat	0	*
5	imkb	0	*
6	mevduat_faiz	0	*
7	rezerv	1	71940.0
8	ydbk	0	*
9	istihdam	0	*
10	kapasite	0	*
11	reel_sektor	0	*
12	new_corp	1	4789.5
13	close_corp	1	669.5
14	sanayi	0	*

Figure 4.1: Cut set

Table: IS_currency1_0.34

29 / 15	Altin	enflas...	Grev	ihracat	imkb	mevd...	rezerv	ydbk	istihd...	kapas...	reel_s...	new_...	close...	sanayi	crisis
0:1	62411484	100	20	99	9364933	34	29787	0	0	79	106	2425	286	100	0
0:2	83810942	101	21	105	14882451	31	34151	0	0	83	106	2784	393	100	0
0:3	78596030	104	16	103	19905932	26	33516	0	0	77	104	4109	928	100	0
0:4	359775830	106	22	113	17450017	22	33717	0	0	76	110	3645	454	100	1
0:5	195552470	107	22	113	13245101	22	32687	0	0	84	107	3376	507	100	0
0:6	268306912	108	26	103	12558118	23	33324	0	0	81	106	3092	520	100	0
0:7	300306343	109	27	124	23977632	22	34195	0	0	84	108	3190	561	100	0
0:8	166677034	112	28	126	18462867	22	34696	0	0	82	107	3123	572	100	0

Figure 4.2: Test set (continuous data)

As you can see from Fig. 4.2, our data set have continuous data and after the cuts are used table X is converted into Fig. 4.3.

Table: IS_currency1_0.34D

29 / 15	Altin	enflas...	Grev	rezerv	new_...	close...	crisis
O:1	"(-Inf,3.534...	"(-Inf,106.5)"	"(12.5,Inf)"	"(-Inf,71940...	"(-Inf,4789....	"(-Inf,669.5)"	0
O:2	"(-Inf,3.534...	"(-Inf,106.5)"	"(12.5,Inf)"	"(-Inf,71940...	"(-Inf,4789....	"(-Inf,669.5)"	0
O:3	"(-Inf,3.534...	"(-Inf,106.5)"	"(12.5,Inf)"	"(-Inf,71940...	"(-Inf,4789....	"(669.5,Inf)"	0
O:4	"(3.534470...	"(-Inf,106.5)"	"(12.5,Inf)"	"(-Inf,71940...	"(-Inf,4789....	"(-Inf,669.5)"	1
O:5	"(-Inf,3.534...	"(106.5,Inf)"	"(12.5,Inf)"	"(-Inf,71940...	"(-Inf,4789....	"(-Inf,669.5)"	0
O:6	"(-Inf,3.534...	"(106.5,Inf)"	"(12.5,Inf)"	"(-Inf,71940...	"(-Inf,4789....	"(-Inf,669.5)"	0

Figure 4.3: Test set (discrete data)

4.2 REDUCTION

In this phase the data are reduced into subsets of attributes that still carry the same knowledge. In our implementation 8 out of 14 attributes are removed because they dont have any effect on the out came of our prediction. Our model propose that only volume of gold market, inflation, number of strikes, central bank reserves, newly opened businesses and closed busineses are enough for prediction for this data set.

An important point is that this removed attributes are highly depended on the discretization methods and the discretized sets structure. Our reduced set is like Fig. 4.4

Table: IS_currency1_0.34D

29 / 15	Altin	enflas...	Grev	ihracat	imkb	mevd...	rezerv	ydbk	istihd...	kapas...	reel_s...	new_...	close...	sanayi	crisis
0:1	"(-Inf,3.534...	"(-Inf,106.5)"	"(12.5,Inf)"	*	*	*	"(-Inf,71940...	*	*	*	*	"(-Inf,4789....	"(-Inf,669.5)"	*	0
0:2	"(-Inf,3.534...	"(-Inf,106.5)"	"(12.5,Inf)"	*	*	*	"(-Inf,71940...	*	*	*	*	"(-Inf,4789....	"(-Inf,669.5)"	*	0
0:3	"(-Inf,3.534...	"(-Inf,106.5)"	"(12.5,Inf)"	*	*	*	"(-Inf,71940...	*	*	*	*	"(-Inf,4789....	"(669.5,Inf)"	*	0
0:4	"(3.534470...	"(-Inf,106.5)"	"(12.5,Inf)"	*	*	*	"(-Inf,71940...	*	*	*	*	"(-Inf,4789....	"(-Inf,669.5)"	*	1
0:5	"(-Inf,3.534...	"(106.5,Inf)"	"(12.5,Inf)"	*	*	*	"(-Inf,71940...	*	*	*	*	"(-Inf,4789....	"(-Inf,669.5)"	*	0
0:6	"(-Inf,3.534...	"(106.5,Inf)"	"(12.5,Inf)"	*	*	*	"(-Inf,71940...	*	*	*	*	"(-Inf,4789....	"(-Inf,669.5)"	*	0
0:7	"(-Inf,3.534...	"(106.5,Inf)"	"(12.5,Inf)"	*	*	*	"(-Inf,71940...	*	*	*	*	"(-Inf,4789....	"(-Inf,669.5)"	*	0
0:8	"(-Inf,3.534...	"(106.5,Inf)"	"(12.5,Inf)"	*	*	*	"(-Inf,71940...	*	*	*	*	"(-Inf,4789....	"(-Inf,669.5)"	*	0
0:9	"(-Inf,3.534...	"(106.5,Inf)"	"(12.5,Inf)"	*	*	*	"(-Inf,71940...	*	*	*	*	"(-Inf,4789....	"(-Inf,669.5)"	*	0
0:10	"(-Inf,3.534...	"(106.5,Inf)"	"(12.5,Inf)"	*	*	*	"(-Inf,71940...	*	*	*	*	"(-Inf,4789....	"(669.5,Inf)"	*	0
0:11	"(3.534470...	"(106.5,Inf)"	"(12.5,Inf)"	*	*	*	"(-Inf,71940...	*	*	*	*	"(-Inf,4789....	"(-Inf,669.5)"	*	0
0:12	"(3.534470...	"(106.5,Inf)"	"(12.5,Inf)"	*	*	*	"(-Inf,71940...	*	*	*	*	"(-Inf,4789....	"(-Inf,669.5)"	*	0

Figure 4.4: Final set before mining rules

4.3 MINING RULES

In this phase we produce the rules for our prediction.

Rule set: IS_currency1_0.66D

(1-11)	Match	Decision rules
1	46	(rezerv="(-Inf,71940.0)")&(new_corp="(-Inf,4789.5)")=>(crisis={0[46]})
2	30	(Altin="(-Inf,3.53447008E8)")=>(crisis={0[30]})
3	28	(close_corp="(669.5,Inf)")=>(crisis={0[28]})
4	16	(Grev="(-Inf,12.5)")=>(crisis={0[16]})
5	12	(enflasyon="(-Inf,106.5)")=>(crisis={0[12]})
6	2	(new_corp="(4789.5,Inf)")&(close_corp="(-Inf,669.5)")=>(crisis={1[2]})
7	2	(Altin="(3.53447008E8,Inf)")&(Grev="(12.5,Inf)")&(new_corp="(4789.5,Inf)")=>(crisis={1[2]})
8	2	(rezerv="(71940.0,Inf)")&(new_corp="(4789.5,Inf)")=>(crisis={0[2]})
9	2	(Grev="(12.5,Inf)")&(rezerv="(71940.0,Inf)")=>(crisis={1[2]})
10	2	(rezerv="(71940.0,Inf)")&(close_corp="(-Inf,669.5)")=>(crisis={1[2]})
11	2	(Altin="(3.53447008E8,Inf)")&(rezerv="(71940.0,Inf)")&(new_corp="(-Inf,4789.5)")=>(crisis={1[2]})

Figure 4.5: Generated rules

In Fig. 4.5 we can see the rules generated with RSES2 Software.

RSES2 generated 11 rules for our data set, those rules can be listed as below;

- *Rezerves ="(-Inf,71940.0)"&new_corp="(-Inf,4789.5)"=>crisis=0 (46)*

Our first rule states that when central bank rezerves in foreign currency is lower than 71940 and the number of newly openned corporations are lower then 4790, there will be no crises. There were 46 objects in our training set supporting this rule. The lower reserves may show the lesser risk estimation and smaller number of corporations may show a more stable trade life so a currency crises is not expected.

- *Gold="(-Inf,3.53447008E8)")=>(crisis=0[30]) 30*

When the volume of gold market is lower than 353447008, there won't be crises. As the volume of gold market is lesser, the money of investers are on other instruments like stock market or local currency. People are expecting a stable economic condition and 30 of our training objects support this rule.

- *close_corp="(669.5,Inf)")=>(crisis=0[28]) 28*

If the number of closed corporations is more than 669 then there will be no crisis. This is an unexpected rule and almost 50 percent of our training set, 28 objects support this rule. This may have two meaning; first number of closing businesses may increase before a currency crises and reach lowest value before actual crises happen or second this increase in the number of closing businesses happen after the hit of crises.

- *Strike="(-Inf,12.5)")=>(crisis=0[16]) 16*

If the number of closed corporations is less than 13 then there will be no crisis. This rule is an expected rule. If the number of strikes is lower, this can be interpreted as working classes are fine with current situations or at least don't expect worst conditions.

- *inflation="(-Inf,106.5)")=>(crisis=0[12]) 12*

If the inflation index is less than 106.5 than there won't be a crisis. With lower in inflation rate the currency stay more stable than a high inflation period.

- *new_corp="(4789.5,Inf)")&(close_corp="(-Inf,669.5)")=>(crisis=1[2]) 2*

when new corporations are more than 4789 and closed corporations are less than 670 there is a possibility of crisis. This rule may seem to be confusing but actually it states that all weakly financed or non-profitable corporations are already closed before the crises and there appears opportunity for new corporations as old closed corporations leave the market.

- *Gold="(3.53447008E8,Inf)")&(Strike="(12.5,Inf)")&(new_corp="(4789.5,Inf)")=>(crisis=1[2]) 2*

When the volume of gold market is larger then 353447008 and numer of strikes are more than 12 and newly opened corparation number is larger then 4789 there is a possibility of crisis. As volume of gold market is increaseing with strikes and a similar relation with the newly opened corporations like previous rule, this objects indicate a currency crisis.

- *rezerv="(71940.0,Inf)")&(new_corp="(4789.5,Inf)")=>(crisis=0[2]) 2*

If reserves of central bank is more than 71940 and number of newly opened business is bigger than 4789, there is no crisis. With the help of high reserves and new dynamic entries to the economy the curreny crisis is not possible for this rule.

- *Strike="(12.5,Inf)")&(rezerv="(71940.0,Inf)")=>(crisis=1[2]) 2*

When the number of strikes is bigger than 12 and the reserves of central bank is bigger than 71940 there is possibility of crisis. Bigger number of strikes shows the bad conditions of workers and bigger reserves show the high risk and the inevitable currency crises.

- *rezerv="(71940.0,Inf)")&(close_corp="(-Inf,669.5)")=>(crisis=1[2]) 2*

when reserves of central bank is bigger than 71940 and number of closed corporations are smaller than can 670 there is chance for currency crisis. Bigger reserves show the high risk and low number of closed businesses show that there is an already damaged economy and a currency crisis is near.

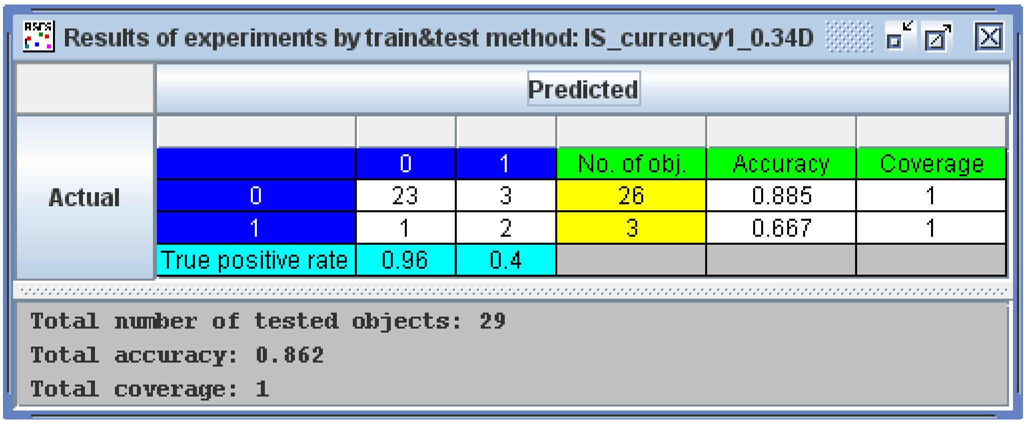
- *Gold="(3.53447008E8,Inf)")&(rezerv="(71940.0,Inf)")&(new_corp="(-Inf,4789.5)")=>(crisis=1[2]) 2*

When the volume of gold market is larger than 353447008 and reserves are larger than 71940 and newly opened corporations are lower than 4790 there is a possibility for currency crisis. The high volume of gold market indicates that people seek shelter from risk; similarly governments are tempted to keep larger reserves in potential expectation of crises. And the low number of newly opened businesses shows a slowing economy.

Depending on the software and discretization rules can be changed but the goal is to produce rules as general as possible to no let any decision undifined and as specify as possible to give minimum false negatives and positives.

4.4 TESTING OUR MODEL

In this final step we use the rules that were produced in previous phases. Then we compare the outcame of our model with the real values. The results in our experiment


Results of experiments by train&test method: IS_currency1_0.34D

	Predicted					
Actual		0	1	No. of obj.	Accuracy	Coverage
	0	23	3	26	0.885	1
	1	1	2	3	0.667	1
	True positive rate	0.96	0.4			

Total number of tested objects: 29
Total accuracy: 0.862
Total coverage: 1

Figure 4.6: RSES2 results

can be seen in Fig. 4.6. As we can see out of 29 objects we correctly spotted 24. And we were able to find 2 out of 3 crisis periods.

4.5 COMPARING RESULTS

In this section we compare the results from the previous sections, we can again remember that we used 66 percent of our data as training set and used the remaining 34 percent for testing proposes.

Here are some formulas (4.1, 4.2, 4.3) we used in creating the comparison, Table 4.1.

$$\text{Sensitivity} = TP / (TP + FN) \quad (4.1)$$

Specificity = TN / (TN + FP) (4.2)

Accuracy = (TP + TN) / (TP + FP + TN + FN) (4.3)

Table 4.1: Comparing the results of methods

Method	ROC	RMSE	Sensitivity"1"	Correctness
BayesNet	0.462	0.377	0%	82.75%
NaiveBayes	0.672	0.409	33.3%	82.75%
Logistic Reg.	0.752	0.308	33.3%	89.65%
SMO	0.5	0.321	0%	89.65%
ConjunctiveRule	0.5	0.304	0%	89.65%
JRip	0.5	0.316	0%	89.65%
Id3 (disc.RSES2)	0.813	0.267	66.7 %	92.85%
RSES2-RST	0.609	0.376	66.7%	86.2%
Rosetta_RST	0.613	0.183	33.3%	89.2%

As we can see from the Table 4.1, Rough set theory achieved well especially predicting crisis periods while Id3 also gets similar results. In general although the correctness value of rough set theory is lower than JRip, Logistic Reg. or SMO like methods, its ability to find crisis periods are much higher. About 10 percent of our test set is consist of crisis periods so a 90 percent correct method may not have any use if it finds no crisis periods. As we can see from the Table 4.1 only id3 algorithm can achieve the same level of crisis prediction with rough set theory implementation RSES2, id3 also get a little more correct results than RSES2.

If we compare the rules that were generated with RSES2 in section 4.3 with the decision tree of Id3 algorithm shown in Fig. 3.17, we may see the difference. Five rules can be generated from id3 algorithm while RSES2 generated 11 rules. 2 out of 5 rules in id3 indicate a crisis while 5 out of 11 rules indicate crisis in RSES2. When we investigate into the rules of both methods, we notice that they are similar for crisis periods like the volume of cold market or closed corporation numbers, but the rules of RSES2 are a little more general while id3 check more variable before reaching a decision attribute. This can be caused by our relatively small database maybe a larger database which contains data of 10 years or more, RSES2 maybe have more precision.

We used two software tools to use rough set theory both are open source tools, the rossetta and the RSES2, as we can see rses2 is performing better than rosetta this may be seen as an implementation issue in generating rules.

5. CONCLUSION

In this study, we present a new approach for currency crisis prediction. The implementation of "Rough Set Theory" in currency crisis, as we noticed in the results of our empirical study in the findings, result in an increase in our capability for both predicting crisis periods accurately and handling missing or noisy data.

As our work also proves the rough set theory is a very robust and capable technic. But there are also some disadvantages;

First of all the data set must be preprocessed and the data must be discretized. This is a very important step and directly affect the out came of the model. Too many different objects in an attribute decrease the roughness of the set and lower the ability of predictions but with a large enough data set this will increase the precision. But lower number of objects increases the roughness and eases the prediction.

Second with the increase of the number of attributes the data need to make prediction increase considerably and the time needed to find of rules and reducts increase.

In the light of our results both currency crisis prediction and rough set theory will remain to be hot topics in the future. With the USA and EU debt problems and possible East Asian real estate balloons; the currency crisis will be the fear of many countries especially with weaker fundamentals or the emerging economies.

The rough set theory has an increasing literature but there still much to study. Implementation of this powerful method to currency crisis like hot topics will attract interest of researchers and will increase the literature considerably.

REFERENCES

Books

Bazan J. G., Szczuka M. S. & Wróblewski J. 2002. A new version of rough set exploration system. In: James J. Alpigini, James F. Peters, Andrzej Skowron, Ning Zhong, redaktorzy, Third International Conference on Rough Sets and Current Trends in Computing RSCTC, volume 2475, *Lecture notes in artificial intelligenc*e,pp. 397-404, Malvern, PA,. Springer-Verlag.

Ben-Gal I., 2007. Bayesian Networks, in Ruggeri F., Faltin F. & Kenett R., *Encyclopedia of statistics in quality & reliability*, Wiley & Sons

Rissino S. & Lambert-Torres G., 2009,Rough Set Theory – Fundamental Concepts, Principals, Data Extraction, and Applications, *Data mining and knowledge discovery in real life applications*, ISBN 987-3-90263-53-0, pp. 438

Periodicals

Ahn J. J., Oha K. J., Kim T. Y., Kim D. H., 2011, Usefulness of support vector machine to develop an early warning system for financial crisis *Expert Systems with Applications* Vol. **38** 2966–2973

Alvarez-Plata P. & Schrooten M., 2004, Misleading indicators? The Argentinean currency crisis, *Journal of Policy Modeling* Vol. **26** 587-603

Anastasakis L. & Mort N., 2009, Exchange rate forecasting using combined parametric and nonparametric self-organizing modeling approach, *Expert Systems with applications* Vol. **36** 12001-12011

Beckmann D., Menkhoff L. & Sawischlewski K., 2006, Robust lessons about practical early warning systems, *Journal of Policy Modeling* Vol. **28** 163-193

Berg A. & Pattillo C, 1999, Predicting currency crises: The indicators approach and an alternative, *Journal of International Money and Finance* Vol. **18** 561-586

Bonafede C.E., Giudici P., 2007, Bayesian Networks for enterprise risk assessment, *Physica A: Statistical Mechanics and its Applications*, Volume **382**, Issue 1, 1, Pages 22-28

Burkart O. & Coudert V., 2002, Leading indicators of currency crises for emerging countries, *Emerging Markets Review* Vol. **3** 107-133

Bussiere M. & Fratzscher M., 2006, Towards a new early warning system of financial crisis, *Journal of International Money and Finance* Vol. **25** 953-973

le Cessie, S. and van Houwelingen, J.C., 1992. Ridge Estimators in Logistic Regression. *Applied Statistics,* Vol. **41**, No. 1, pp. 191-201.

Dimitris A.I., Slowinski R., Susmaga R. & Zopounidis C., 1999, Business failure prediction using rough sets, *European Journal of Operational Research* Vol. **114** 263-280

Goldfajn I. & Valdes R. O., 1998, Currenct Account Sustainability, Are currency crises predictable*?* *European Economic Review* Vol. **42**, 873-885

Hall M., Frank E., Holmes G., Pfahringer B., Reutemann P., Witten I. H., 2009, The WEKA Data Mining Software: An Update; *SIGKDD Explorations,* Volume **11**, Issue 1.

John G. H. and Langley P. , 1995. Estimating Continuous Distributions in Bayesian Classifiers. *Proceedings of the Eleventh Conference on Uncertainty in Artificial Intelligence*. pp. 338-345. Morgan Kaufmann, San Mateo.

Kaminsky G., Lizondo S. & Reinhart C. M. , 1998, Leading Indicators of Currency Crises, *Staff Papers – International Monatary Fund*, Vol. **45**, No. 1 (Mar 1998), pp. 1-48

Kaminsky G. L., 2006, Currency Crises: Are they all the same? *Journal of International Money and Finance* Vol. **25**, 503-527

Kim T. Y., Oh K.J., Sohn I. & Hwang C., 2004, Usefulness of artificial neurl networks for early warning system of economic crisis, *Expert Systems with applications* Vol. **26**, 583-590

Kumar M., Moorthy U. & Perraudin W., 2003, Predicting emerging market currency crashes, *Journal of Emprical Finance* Vol. **10**, 427-454

Lanine G., Vennet R. V., 2006, Failure prediction in the Russian bank sector with logit and trait recognition models, *Expert Systems with Applications Volume* **30**, Issue 3, Pages 463-478

Lee V.C.S. & Wong H. T., 2007, a multivariate neuro-fuzzy system for foreign currency risk management decision making, *Neurocomputing* Vol. **70**, 942-951

Lin C.-S., Khan. H. A., Chang R.-Y. & Wang Y.-C., 2008, A new approach to modeling early warning systems for currency crises: Can a machine learning fuzzy expert system predict the currency crises effectively?, *Journal of International Money and Finance* Vol. **27**, 1098-1121

Maltritz D. & Eicher S., 2009, Currency crisis prediction using ADR market data: An options-based appoach, *International Journal of Forecasting* (2009), doi:10.1016/j.ijforecast.2009.05.28

Osband K. & Rijckeghem C. v., 2000, Safety From Currency Craches, *IMF Staff Papers* Vol. **47** No.2 pp. 238-258

Özlale Ü. & Metin-Özcan K., 2007, An alternative method to measure the likelyhood of financial crisis in an emerging market, *Physica A* Vol. **381**, 329-337

Pawlak Z. 1982, Rough Sets. *International Journal of Computer and Information Sciences* Vol. **11**(5) 34-356

Quinlan R., 1986. Induction of decision trees. *Machine Learning*. Vol.**1**, No.1, pp. 81-106.

Platt J. 1998. Fast Training of Support Vector Machines using Sequential Minimal Optimization. Advances in Kernel Methods - Support Vector Learning, B. Schoelkopf, C. Burges, and A. Smola, eds., MIT Press. S.S. Keerthi, S.K. Shevade, C. Bhattacharyya, K.R.K. Murthy, Improvements to Platt's SMO Algorithm for SVM Classifier Design. *Neural Computation*, Vol. **13**(3), pp 637-649, 2001.

Sun L. & Shenoy P. P. 2007, Computing, Artificial Intelligence and Information Management, Using Bayesian networks for bankruptcy prediction: Some methodological issues. , *European Journal of Operational Research* Vol. **180**, 738–753

Xu X. & Wang Y. 2009. Financial failure prediction using efficiency as a predictor, *Expert Systems with Applications* Vol. **36**, 366–373

Vapnik, V. 1998. Statistical learning theory. New York: Wiley.

Yao J. & Tan C. L., 2000, A case study on using neural networks to perform technical forecasting of forex, *neurocomputing* Vol. **34**, 79-98

Tay F. E. H. & Shen L., 2002, Economic and financial prediction using rough sets model, *European Journal of Operational Research* Vol. **141**, 641-659

Zhang Y. Q. & Wan X., 2007, Statical fuzzy interval neural networks for currency exchange rate time series prediction, *Applied Soft Computing* Vol. **7**, 149-1156

Øhrn A. & Komorowski J. ,1997, ROSETTA: A Rough Set Toolkit for Analysis of Data, Proc. *Third International Joint Conference on Information Sciences, Fifth International Workshop on Rough Sets and Soft Computing* (RSSC'97), Durham, NC, USA, March 1-5, Vol. **3**, pp. 403-407.

Internet and Other Publications

Rouhani-Kalleh, O., 2006.Analysis, Theory and Design of Logistic Regression Classifiers Used For Very Large Scale Data Mining. *Master Thesis*. (http://omidrouhani.com/research/logisticregression/html/logisticregression.htm #_Toc147483480)

International Rough set society, http://roughsets.home.pl/IRSS/rs-kdd/sld001.htm, http://roughsets.home.pl/IRSS/RSCTC_02_final_pliki/v3_document.htm

Internet site of central bank of Turkey, http://www.tcmb.com.tr

Electronic Data Delivery System of central bank of Turkey; http://evds.tcmb.gov.tr/ yeni/cbt-uk.html

WEKA classifier guide; http://weka.sourceforge.net/doc/weka/classifiers/rules/ ConjunctiveRule.html

CURRICULUM VITAE

Name/Last Name	: Sibar Kaan Manga
Adress	: Eğitim Mah. Muratpaşa Cad. No:19 Ziverbey-Kadıköy/ İstanbul
Place/Date of Birth	: Ankara, 1984
Foreign Languages	: English, German, French
Education	

Primary School	: Uşak Gazi Mustafa Kemal İlk Okulu	1995
Secandory and Highschool	: İzmir Bornova Anadolu Lisesi	2002
Undergraduate	: Ege Üni. Computer Eng, İZMİR	2006
Master	: Bahçeşehir Üni., İstanbul Natural and Applied Sciences Computer Engineering Graduate Program	2011

Professional Work Exp.	: 2007 - … Istanbul Stock Exchange Takasbank Inc., ISTANBUL Software Specialist