An Analysis on Two Different Data Sets by using Ensemble of k-Nearest Neighbor Classifiers

NOR AZUANA RAMLI
School of Mathematical Sciences
Universiti Sains Malaysia
11800 Pulau Pinang,
MALAYSIA
ajue.ramli@gmail.com

MOHD TAHIR ISMAIL
School of Mathematical Sciences
Universiti Sains Malaysia
11800 Pulau Pinang,
MALAYSIA
mtahir@cs.usm.my

HOOY CHEE WOOI
School of Management
University Sains Malaysia
11800 Pulau Pinang,
MALAYSIA
cwhooy@usm.my

Abstract: An ensemble of classifiers is proposed in this study where we combined the k-nearest neighbor classifier with LAD Tree through stacking. Two different data sets which are macroeconomic and the International Risk Country Risk are used for application. Both of the data are taken for 27 countries from the first quarter of 1984 until the fourth quarter of 2011. Before we applied those two data sets on the ensemble of classifiers, we have done the multicollinearity test to see if the predictor variables are highly correlated. By using the variance inflation factor, the results showed that some predictor variables are highly correlated in the macroeconomic data set. To solve the multicollinearity problem, we used principal component analysis that is available in SPSS as it can reduce a set of factors. The remaining variables are used in the analysis of the proposed method to observe the prediction accuracy. It is proven that different type of data does affect the prediction accuracy, but then it is different for every country.

Key–Words: Currency crisis, principal component analysis, multicollinearity test, stacking, macroeconomic

1 Introduction

Currency crisis forecasting had been in the research field ever since 1970s. It is not something new but the academic literatures in this area are kept increasing year by year. There are different types of research area in forecasting currency crisis. At first before the empirical models were introduced, the research was mainly focused on finding explanations why this crisis occurred. Then, consequent to the evolving of empirical models, researchers and economists also started an analysis on the leading indicators. Now, the research in forecasting currency crisis becomes complex since it does not just involve with the empirical model but theoretical and indicators too. Due to this reason, currency crisis forecasting is regarded as a rather challenging task.

There are lots of empirical models involving different fields that already experimented by researchers in finding better predictability and accurate model to build an early warning system, which to be specified in forecasting currency crisis. That should be an answer for every new research on why we need a new model even we already have plenty much of them. So, the next questions will be what is wrong with the previous models and what are the improvements that could be done to overcome some disadvantages from previous models? As for the beginning, we will start with the most used methodology in developing an early warning system which is also the first empirical method that had been innovatively developed by Kaminsky et al. [10]. By the time signal approach was proposed as a method that can be used for early warning system in forecasting currency crisis, it looks like a capable model that can work for a long run.

A study conducted by Yolacan and Asma [20] stated that Kaminsky approach is one of the best track record but there is no comparison made in their study. Some studies that had been held recently proved the forecasting results from Kaminsky approach especially for out-of-sample data that had been applied in the Asian Financial Crisis 1997 case were quite unsatisfied. Berg and Patillo [1] compared Kaminsky approach with the panel logit model. The panel logit method had some advantages by showing a better in-sample prediction results through the quality probability score (QPS) and the log probability score (LPS) compared to Kaminsky approach.

To add another reason in the list, Edison [8] tested signalling approach method by using a broad range of countries, increase the number of indicators and even testing for regional differences in his study. Eventually, the results of the study still disappointing since there are too many false alarms. Even Berg and Patillo study showed better results than signaling approach, it does not mean that logit method has no disadvantages. Bussiere and Fratzscher [12] in their study found that Berg and Patillo models only moderately perform in predicting before crises while the best models should outperform in ex-ante and after crises. To overcome the weakness of Berg and Patillo models, they developed a new type of early warning system by using multinomial logit methods. Similar to previous research, this approach has no practical estimation results even if we use out-of-sample application on the datasets.

Subsequently, a few studies employed Markov switching methods to develop early warning systems in forecasting currency crisis. In most cases, researchers only define two types of crisis regimes which one with a higher volatility that had been conducted by Abiad [2], while the other one with a higher mean that had been done by Cerra and Saxena [16]. Although this approach is quite interesting, it has been shown by quite a lot of numbers of researches, for example one of them is Candelon et al. [3] that found its predicting abilities are lower compared to the study that had been conducted by Berg and Patillo. This means logit models outperform the Markov switching.

Since there are some disadvantages with traditional statistical methods, researchers started to think outside the box by experimenting others method such as artificial neural network. Peltonen [15] was the first one that used artificial neural network and based on his results, its ability to signal currency crisis out-of-sample was poor. There are some other disadvantages in this method too such as the black box nature of artificial neural network makes it difficult for humans to understand how the networks predict the crisis. However, major disadvantage of this method lies in their knowledge representation and small changes in the training set or parameter selection can produce large changes in the predicted output. Thus, in this paper, we develop a novel method to overcome single classifiers and traditional statistical methods problem by ensemble three different classifiers.

As finding the right methodology is not the only problem that researchers have to face in developing better predictability and high accuracy early warning system, we also put our main focus in different type of indicators. Right choice of indicators to be used in the analysis and its relation to each other really play an important role in developing better early warning systems. It had been proven since the development of the first generation of currency crisis models. One way to make the right choice of indicators is by take considerations of the theoretical and empirical literature. As well, the tested indicators should consist of a wide range of macroeconomic and financial indicators. But how about political indicators? Do they count too? Bussiére and Mulder [13] had done an analysis and test due to the effect of political indicators on currency crisis. The results from their study supported their inference and it was verified that political indicators has a strong effect on currency crisis for countries with weak economic basics and low reserves. Even Kaminsky et al. had the same conclusion in their study where they stated that an effective warning system should consist of large diversity of indicators.

Therefore, in this paper, we use financial, economic and political indicators to compare the forecasting results with an analysis by using only macroeconomic indicators. This paper is organized as follows, where in the next section the building process of the nearest neighbour tree is briefly described here. Section 3 explains about research design which will be more on a dataset from indicators to sample of countries that used in this study. Multicollinearity test and data reduction will be on Section 4 before we discuss all the results. Last but not least, some concluding remarks are drawn in Section 6.

2 Model Building Processes

In this section, we briefly describe the model building process step by step starting with the introduction of the two methods that used in this model. Afterward, a stacking method and a final form of this model are presented.

2.1 k-nearest neighbour method

In machine learning, the k-nearest neighbour method (k-NN) is also known as lazy learning because of its training is held up to run time. This classifier is also one of the most straightforward and simplest since classification of the datasets is based on their nearest neighbours class. Since the conception by Fix and Hodges in 1951, this method has attracted many followers and continues to be studied by many researchers. Generally, we define the k-NN method by

$$f_n(x) = \begin{cases} 1, \text{ if } \sum_{i=1}^{n} w_{ni} I_{(Y_i=1)} > \sum_{i=1}^{n} w_{ni} I_{(Y_i=0)} \\ 0, \text{ otherwise,} \end{cases} \tag{1}$$

where $w_{ni} = 1/k$ if X_i is among the k-nearest neighbour of x, and w_{ni}=0 elsewhere.X_i is said to be the k-th nearest neighbour of x if the distance $||X_i - x||$ is the k-th smallest among $||X_1 - x||, \cdots, ||X_n - x||$. The value of k is usually small. When $k = 1$, the datasets are basically allocated to the class of its nearest neighbour.

In the beginning, the classifier was studied in 1951 by Fix and Hodges in the US Air Force School of Aviation Medicine. Then, Cover and Hart formalized the idea and invented the main properties of this method in 1967 [6]. They also described this classifier more properly and found the upper error bound of the method to be twice of Bayes' error probability. The k-NN classifier performance depends on the choice of a distance that is used. In our previous paper [14], we already investigated the performance of k-NN classifier based on different distance and different number of k. Therefore, we are just going to use the results that we got and applied in this experiment.

There are a few disadvantages in the k-NN such as it has too many choices of distance and it takes longer times to compute. But the major disadvantage of this classifier is it uses all the features equally in computing similarities. This can increase the percentage of classification errors especially when there is only a small subset of features that are useful for classification [17].

2.2 LAD Tree algorithms

Decision tree or tree-based method is one of the most famous predictor used in the pattern recognition or machine learning and it is frequently used as a base classifier in constructing an ensemble of classifiers. Tree-based method has a multistage decision process and the decision is made in the binary form at each stage. It has been called as a tree-based method since this method has nodes and branches and their nodes are designed as an internal or a terminal node. The difference between internal and terminal node is an internal node can be separated into two children while contradiction with a terminal node as it has no children at all. Additionally, a terminal node has a class label associated with it.

Tree-based methods are conceptually simple but they are known as powerful methods. In this paper, we chose to use the LAD Tree which is a learning algorithm that applies the logistic boosting algorithm in order to induce an alternating decision tree (AD Tree). This method was innovated by Holmes et al. [9] after they found the AD Tree can be transformed in several ways in order to map multiple class labels to two classes which is also a reason why we chose to use it since our study case is two classes problem.

From Holmes et al., the algorithm for LADTree to obtain a single directly induced tree is as follows:

1. Initialize by creating a root node with $F_j(x)$=0 and $P_j(x) = \frac{1}{J}\forall j$

2. Repeat for $m = 1, 2, \cdots, T$:

a. Repeat for $j = 1, 2, \cdots, J$:

Then, compute working responses and weights in the j-th class

$$z_{ij} = \frac{y'_{ij} - p_{ij}}{p_{ij}(1 - p_{ij})}, w_{ij} = \frac{y'_{ij} - p_{ij}}{z_{ij}}$$

b. Add the single test to the tree that best fits $f_{mj}(x)$ by a weighted least-squares fit of z_{ij} to x_i with weights w_{ij}

c. Add prediction nodes to tree by setting

$$f_{mj}(x) \leftarrow \frac{j-1}{J}(f_{mj}(x) - \frac{1}{J}\sum_{k=1}^{J} f_{mk}(x))$$

and

$$F_j(x) \leftarrow F_{ij}(x) + f_{mj}(x).$$

d. Update $p_j(x)$ by using equation below

$$p_j(x) = \frac{e^{F_j(x)}}{\sum_{k=1}^{J} e^{F_k(x)}}, \sum_{k=1}^{J} F_k(x) = 0 \quad (2)$$

3. Output the classifier $argmax_j F_j(x)$.

By using this algorithm, trees for the different classes are grown in parallel. Then, it is possible to merge all of them into a final model once every individual tree has been built. This merging process includes searching for identical tests on the same level of the tree. If there is a sign of existence of such test, then the test and its subtrees can be merged into one. The additive nature of the trees means that the prediction values for the same class can be added together when merged.

2.3 Stacking

Stacking is just another method in the ensembles and it is used to combine two or more classifiers generated by using different base classifiers, L_1,$\cdots$,L_N and a meta-level classifier on any data sets, D which consists of examples $d_i = (x_i, y_i)$, where x_i is a pair of feature vector and y_i is its classification. It was introduced by Wolpert [5] in 1992. Even though voting is more famous method than stacking when it comes to combine classifiers in machine learning, but in this study we would like to use a different approach than other studies that had been conducted. The difference between stacking with another global technique in ensembles which is voting is the ensemble by using stacking need not require the base classifiers to be linear since it is learned through a combiner system. Furthermore, there are two other advantages of using stacking which is trained rule is more flexible and less bias plus there is no need to normalize classifier outputs.

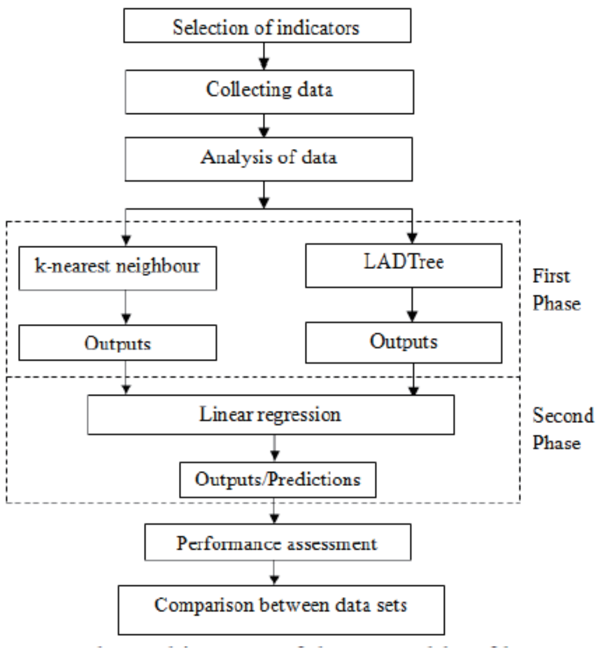

Figure 1: The architecture of the ensemble of k-nearest neighbour with LAD Tree by using stacking

Unlike voting, stacking has two phases in their system wherein the first phase, a set of base-level classifiers is generated. Then, a meta-level classifier is learned to combine the outputs from the base-level classifiers in the second phase. In stacking, the combiner model cannot be trained by using training data since the base-level classifiers possibly memorizing the training set. That is why stacking estimates and corrects whenever there is any bias in the base-level classifier.

2.4 Nearest neighbour tree

Before we chose nearest neighbour tree (NNT) as our final method to be used in predicting currency crisis, we had done some experiments involving single classifiers and combined classifiers such ensembles of support vector machines [21], logistic regression ensembles and neural network ensembles [22]. These combined classifiers were taken from previous study but the application is not under early warning system or currency crisis. Dietterich [18] had proved that combination of classifiers provide higher prediction accuracy than single classifiers. Chang [19] in his study showed that a combination of Taguchi method with nearest neighbour classification is better than neural network methods.

We chose a decision tree to be one of our classifiers because of its interpretability. Yet, it has a disadvantage which is high variance which we decided to cover by using k-nearest neighbour (k-NN) classifier. Figure 1 shows the architecture of our nearest neighbour tree model. Since we already explained about the k-nearest neighbour, LAD Tree and stacking, we will continue with a description of our model.

As can be seen in Figure 1, there are two phases of combination involves in the process. In the first phase, the base learners which are k-nearest neighbour and LAD Tree are going to analysis the input which will be either macroeconomic data set or the ICRG. Then, the outputs from both base learners will be combined and produced a linear equation for every set of data by using Linear Regression. This equation will be used to find the probability of crisis. The performance of this nearest neighbour tree already had been assessed and compared with other single classifiers and ensembles of classifiers from our previous study. Therefore, in this paper we will just compare the forecasting results between these two different data sets.

3 Research Design

3.1 Selection of indicators

There were lots of previous studies had been conducted related to find the suitable variables that could serve as leading indicators in forecasting currency crisis. This showed how crucial of selecting the suitable variables step in order to get an accurate forecasting result besides choosing the best method. In our analysis, we have included the International Country Risk Guide (ICRG) variables in order to compare with the macroeconomic indicators as the ICRG indicators never have been included as early warning system indicators before.

3.1.1 Macroeconomic indicators

The tested 13 macroeconomic indicators were selected based on the previous empirical literatures. Based on some literatures that we have done such as from Kaminsky et al. [10], it was cited that indicators like exports, imports, real effective exchange rate, terms of trade, GDP, the M2 money multiplier, international reserves and index of equity prices have the best track in giving a signal for a currency crisis. However, not all of these indicators data are available for some countries that have experienced a crisis. That is why in our study, we have chosen these 13 macroeconomic indicators based on the availability of the data too although there are certain indicators that its data are available for some countries but not available for others. That was one of the obstacles that we have to face in our study.

The 13 macroeconomic indicators that had been chosen were exports of goods and services, foreign direct investment (FDI), the M2 money multiplier, imports of goods and services, consumer price index (CPI), foreign exchange reserves (FER), general government final consumption (GC), industrial production index (IPI), producer price index (PPI), real effective exchange rate (REER), GDP per capita, terms of trade (TOT) and unemployment rate. Amongst all indicators, real effective exchange rate seems to play a very significant role since it is related to definition of currency crisis itself and currency devaluation is depending on the exchange rate regime in place. We selected the M2 money multiplier instead of the M1 money multiplier because when the Federal Reserve changes the money supply, it targets M2. M1 is the narrowest definition of money while M2 gives a broader definition of money (M1 and M2 are both measures of the money supply that the Federal Reserve Board can adjust when using active monetary policy).

We tried to make a difference in our research not just by using different methodology, but also by adding some indicators that are not usually included in the previous research involving predicting of currency crisis. Indicators like consumer price index, producer price index and industrial production index are red-hot economic indicators. Red-hot economic indicators mean that the indicators are getting so much attention in the financial markets. Consumer price index, producer price index and industrial production index are related to each other. These three indicators are important indicators because inflation affects everyone. Start with industrial production index and then producer price index to consumer price index. Producer price index measures changes in prices that manufactures and wholesalers pay for goods during various stages of production. If a business has to pay more for goods then consumers have to pay with much higher costs. Consumer price index determines how much consumers pay for goods and services, which definitely will affect the cost of doing business and causes havoc with personal and corporate investments then lastly influences the quality of life for retirees.

3.1.2 The ICRG indicators

The International Country Risk Guide (ICRG) rating consists of 22 indicators in three different subcategories of risk which are financial, economic and political. There is a different index for each subcategory such that the Political Risk index is based on 100 points while Financial and Economic Risk on 50 points each. The risk is calculated by summation of all points from those three subcategories and then divided by two to get its weights which also known as composite country risk score. The range of the composite country risk score has different categorization where from 80 points and above has been categorized as very low risk and 49.9 points and below is the opposite.

The ICRG model is actually made to forecast financial, economic and political risk. It was created as early as in 1980 which exactly 34 years ago. The editors of International Reports who created this ICRG model generated a statistical model to compute risks which also come with an explanation of the analysis of the numbers and observe something that do not show explicitly by the numbers. Even the ICRG model already present a year after the first generation crisis model was created, however there is lack of research had been conducted to prove that maybe indicators in the ICRG model do work in forecasting currency crisis. Hence, our study applied the ICRG indicators with hope to see if there is any interaction between the occurrence of currency crisis with indicators of financial, economic and political risk.

3.2 Data and sample country

Most of the previous studies involving early warning system especially in a scope of finding which indicators give signals to the crises used large cross-country datasets. It is quite hard to see the effect on certain indicators if the data sets are too large. Therefore, we decided to cover datasets for 27 industrial and emerging countries only.

Table 1: List of countries and its scope

Data/Scope	Countries
Emerging market: Asia	Indonesia, Malaysia, the Philippines, South Korea, Thailand
Emerging market: Latin America	Argentina, Brazil, Bolivia, Chile, Ecuador, Mexico, Peru, Uruguay, Venezuela
Industrial	Denmark, Finland, Greece, Ireland, Italy, Norway, Portugal, Sweden, Spain
Emerging market: Countries in transition	Hungary
Middle East and Africa	Israel, South Africa, Turkey

Table 1 above shows the list of countries and all of these countries that were covered in our study had experienced currency crisis. The data sets taken for this study are from first quarter of 1984 to fourth quarter of 2011. All the data for macroeconomic indicators were downloaded under analysis via DataStream, while all the data for the ICRG indicators were downloaded through its website [24].

3.3 Performance assessment

3.3.1 Accuracy and false alarm

Accuracy can be measured in two different ways. First is by using percentage of accuracy = [(Correctly predicted data)/(Total testing data)]100. Another way is by using the results that we got from confusion matrix and then use the formula accuracy = 1 – error where error = total false positive and false negative/ total number. Confusion matrix for two-class problems is as shown in the Table 2. For a crisis that predicted to be occurred, if the prediction is also there is a crisis occurs, that means the prediction is true (which also means if a signal given for crisis, therefore that signal is a correct signal). On the contrary, if our prediction is no crisis occurs but there is a crisis that means our prediction is false negative.

Usually for certain software like WEKA [23], the confusion matrix is already included in the output or results of an experiment. Therefore, it is unnecessary to find the values for true positive, true negative, false positive and false negative. Yet, this table can be used as a guideline since we will use the results to find accuracy and false alarm. The percentage of false alarms can be found by using: [fp/p'] 100.

Table 2: Confusion matrix for two classes

Crisis prediction			
Crisis	Yes	No	Total
Yes	tp:true positive	fn:false negative	p
No	fp:false positive	tn:true negative	n
Total	p'	n'	N

3.3.2 Area under the ROC curve

The performance can also be assessed by using area under the receiver operating characteristic (ROC) curve. ROC curve is a plot where the true positive rate (y-axis) against the false positive (x-axis). The area under the ROC curve which abbreviate as AUC is a convenient way to compare performance of different type of classifiers or in our case, different type of data. A useful guidance on the classification for the accuracy of the AUC is shown as in Table 3 [7]. The area determines the capability of the test to suitably class the data with and without crisis.

Table 3: Classification for the accuracy of the AUC

Area Under the ROC Curve	Classification
0.90 – 1.00	Excellent
0.80 – 0.90	Good
0.70 - 0.80	Fair
0.60 - 0.70	Poor
0.50 - 0.60	Fail

4 Data Analysis

4.1 Multicollinearity test

There is a statistical phenomenon where two or more predictor variables in multiple regression model are highly correlated which means one can be linearly predicted from the others with non-trivial degree of accuracy and this phenomenon called multicollinearity. Multicollinearity can cause large forecasting error and make it difficult to assess the relative importance of individual variables in the model. Due to this reason, researchers do anything that they could to avoid or reduce multicollinearity. Multicollinearity is a problem between different macroeconomic indicators since they tend to move to the same direction. There are several methods that can be used to test for multicollinearity such as condition index, Farrar-Glauber test, construction of a correlation matrix among the explanatory variables, F-test and formal detection-tolerance or the variance inflation factor (VIF).

We chose to use VIF to detect multicollinearity in our data where Tolerance =$1 - R_j^2$, VIF = 1/Tolerance and R_j^2 is the coefficient of determination of a regression of explanatory j on all the other explanators. A VIF value that is higher than 10 indicates a multicollinearity. By using SPSS software, we found that there is multicollinearity in our data since the VIF values are greater than 10. Table 4 showed the results that we got from SPSS even when we kept repeating the analysis by changing the dependent variable. There are a few methods which can solve the multicollinearity problem but the choice of a remedial measure depends on the circumstances the researchers encounter. We chose to reduce the dimensionality of our data sets by using principal component analysis to overcome with this multicollinearity problem.

4.2 Dimensionality reduction

A reason for this dimension reduction step is to reduce the multicollinearity problem. In our experiment, principle component analysis (PCA) had been used as a statistical method for this data reduction pur-

Table 4: Multicollinearity results when terms of trade is taken as dependent variable

Indicator	Tolerance	VIF
Consumer price index	0.001	**718.047**
Foreign direct investment	0.784	1.276
Foreign exchange reserves	0.105	9.488
Government consumption	0.044	**22.570**
Imports	0.086	**11.600**
Industrial production index	0.015	**67.318**
Real effective exchange rate	0.012	**83.348**
Gross domestic product	0.025	**39.953**
Unemployment	0.102	9.819
Producer price index	0.002	**645.893**
Exports	0.025	**39.391**
M2 money multiplier	0.024	**41.227**

Table 5: Total variance explained

Component	Total eigenvalues	Percentage of variance	Percentage of Cumulative
1	7.566	58.203	58.203
2	1.834	14.110	72.313
3	1.334	10.262	82.575
4	0.862	6.629	89.204
5	0.689	5.302	94.506
6	0.255	1.962	96.468
7	0.144	1.106	97.574
8	0.103	0.793	98.367
9	0.090	0.694	99.060
10	0.063	0.487	99.547
11	0.036	0.277	99.825
12	0.020	0.154	99.979
13	0.003	0.021	100.000

pose. The advantage of this approach is simplicity and this type of dimension reduction is known as parsimonious summarization. By using SPSS software, we obtained the percentage of the variance in the data set as shown in the Table 5. This percentage can be computed by dividing eigenvalues for the components with total eigenvalues of the correlation matrix.

The cumulative initial eigenvalues of these components is 82.575 percent which means these components correspond to an information (variability) loss of 17.425 percent, compared with the whole set of parameters. After done with extracting the initial components, now we can finally finding the real significant of the principal indicators using the rotated component matrix. In our experiment, we chose to apply Varimax rotation with Kaiser Normalization. The outputs for the final results are shown in the Table 6 and 7. As can be seen in the tables below, with an orthogonal rotation such as the Varimax, the factors are not permitted to be correlated (they are orthogonal to each other). Oblique rotation such as Promax rotation on the other hand will produce factors that are correlated with one another.

5 Results and Discussion

From the previous analysis that we had done, the forecasting results that we obtained by using nearest neighbour tree are better than single classifiers and traditional statistical methods. Additionally, this method produced quite comparable results and had the highest area under ROC curve (AUC) compared to the other two ensembles of classifiers (k-NN with support vector machines and logistic regression ensembles) which already established in literatures. But how this method works on two different types of data? To answer that question, these two different data sets that we obtained from Datastream and the PRS website [24] were applied to our novel nearest neighbour tree method. The results are summarized as in Table 8 and Figure 2.

From Figure 2, it is clearly shown that this method produced the same percentage of accuracy for most countries except for Argentina, Bolivia, Greece, Norway, Peru, South Africa, Sweden and Turkey. From those nine countries, only three of them have the highest percentage of accuracy in forecasting currency crisis if the macroeconomic data set is used instead of the ICRG. If we take an average on percentage of accuracy for both data sets, the difference between them is slightly different with 0.11 percent. It is just a small percentage of difference to compromise if we compare the performance by using the percentage of false alarm and area under the ROC curve (AUC) where the percentage of false alarms by using the macroeconomic data is smaller than the ICRG data. Averagely speaking, forecasting results will be much better if we use the macroeconomic data based on Table 2 results. Having a high percentage of accuracy is important in forecasting currency crisis, but so do less percentage of false alarms and more than 0.8 in the area under the ROC curve.

Table 6: Rotated component matrix by using Varimax rotation

Indicator	1	2	3
CPI	0.927	-0.093	-0.161
Exports	0.968	0.125	0.058
FDI	0.174	0.289	0.601
FER	0.915	0.190	0.237
GC	0.501	-0.489	0.620
Imports	0.897	0.167	0.268
IPI	-0.131	-0.584	-0.100
M2	0.882	-0.262	0.198
PPI	-0.581	0.157	0.707
REER	0.951	-0.022	0.224
GDP	0.934	0.056	-0.063
TOT	-0.122	0.879	0.103
Unemployment	0.932	-0.167	-0.232

Table 7: Component transformation matrix

Component	1	2	3
1	0.996	-0.050	0.078
2	-0.011	0.775	0.632
3	-0.092	-0.630	0.771

6 Conclusion

In this paper, we have introduced a new empirical method named nearest neighbour tree to be added in the forecasting currency crisis literatures. It is an innovative way, especially the number of research studies in an ensemble of classifiers involving forecasting currency crisis is still the least compare to forecasting bankruptcy, financial time series and stock market forecasting. The forecasting results that we obtained are technically better than single classifiers and traditional statistical methods since we had done an analysis on this method before. On top of that, this method produced quite comparable results and had the highest area under ROC curve (AUC) compared to the other two ensembles of classifiers (k-NN with support vector machines and logistic regression ensembles) which already established in literatures. When we tested this method on two different data sets as in this study, the results are not much different except for certain countries. Before we applied these two data sets, a multicollinearity test had been done and we found that there is a high correlated between predictor variables in macroeconomic data set. As this issue can affect the prediction accuracy, we did a data reduction by using principal component analysis.

Results that we obtained showed from those nine countries only three of them have the highest percentage of accuracy in forecasting currency crisis if the macroeconomic data set is used instead of the ICRG. As if the percentage of accuracy is taken averagely for both data sets, the difference is still not much compare to the less percentage of false alarms and highest number of areas under the ROC curve that we obtained if the forecasting is done by using macroeconomic indicators. To sum up this study that we have done, the choices of indicators do play an important role in forecasting currency crisis since it determines how accurate our prediction is. But then, this study proved that the selection of indicators is different for every country. It is as same results from Abiad's [2] theory where he stated that the different country's crisis is based on different indicators. For further research, we hope to see this method that we created to be applied in different area of study.

Table 8: Percentage of false alarm and area under the ROC curve (AUC) by using nearest neighbour tree on two different types of data sets for 27 countries

Country	False alarm (%)		AUC	
	Macro economic	ICRG	Macro economic	ICRG
Argentina	8.421	8.491	0.859	0.836
Bolivia	0	14.286	1	0.985
Brazil	4.808	4.808	0.916	0.918
Chile	0	0	0.645	0.645
Denmark	1.786	1.786	1	1
Ecuador	4.63	4.63	0.866	0.866
Finland	4.464	4.464	0.377	0.377
Greece	5.505	5.102	0.875	0.864
Hungary	3.636	3.636	0.789	0.789
Indonesia	1.942	1.942	0.898	0.898
Ireland	3.636	3.636	0.781	0.781
Israel	14.286	14.286	0.982	0.982
Italy	0	0	1	1
Malaysia	2	2	0.922	0.922
Mexico	5.66	5.66	0.883	0.883
Norway	3.604	3	0.709	0.628
Peru	3.636	3.03	0.825	0.981
Philippines	16.667	16.667	0.918	0.918
Portugal	3.571	3.571	0.484	0.484
South Africa	0	0	0.993	0.994
South Korea	1.852	1.852	0.753	0.753
Spain	4.505	4.505	0.78	0.78
Sweden	0	1.786	1	0.945
Thailand	0.917	0.917	0.98	0.98
Turkey	5.505	5.102	0.865	0.847
Uruguay	5.357	5.357	0.289	0.289
Venezuela	0	0	0.237	0.237
Average	3.94	4.463	0.801	0.799

Figure 2: Percentage of accuracy by using nearest neighbour tree on two different types of data sets for 27 countries

Acknowledgements We would like to acknowledge support for this project from Universiti Sains Malaysia (USM Short term grant 304/PMATHS/6313045).

References:

[1] A. Berg and C. Patillo, Predicting Currency Crises: The Indicators Approach and an Alternative, *Journal of International Money and Finance*, Vol.18, 1999, pp. 561-586.

[2] A. G. Abiad, Early Warning Systems: A Survey and a Regime-Switching Approach, *IMF Working Paper*, 2003.

[3] Candelon et al., How to Evaluate an Early Warning System? Towards a Unified Statistical Framework for Assessing Financial Crises Forecasting Methods, *IMF Economic Review*, Vol. 60, No.1, 2012, pp. 75-113.

[4] C. M. Bishop, *Pattern Recognition and Machine Learning*, New York: Springer-Verlag, 2007.

[5] D. H. Wolpert, Stacked Generalization, *Neural Networks*, Vol. 5, No.2, 1992, pp. 241-259.

[6] Ethem Alpaydin, *Introduction to Machine Learning 2nd Edition*, Massachusetts: The MIT Press, 2010.

[7] G. T.Thomas-http://gim.unmc.edu/dxtests/ROC3.htm

[8] H. J. Edison, Do Indicators of Financial Crises Work? An Evaluation of an Early Warning System, *International Journal of Finance and Economics*, Vol. 8, 2003, pp. 11-53.

[9] Holmes et al., Multiclass Alternating Decision Trees, *Lecturer Notes in Computer Science*, No. 2430, 2002, pp. 161-172.

[10] Kaminsky et al., Leading Indicators of Currency Crises, *IMF Working Paper*, 1998.

[11] L. I. Kuncheva, *Combining Pattern Classifiers: Methods and Algorithms*, New Jersey: John Wiley and Sons, 2004.

[12] M. Bussiere and M. Fratzscher, Towards a New Early Warning System of Financial Crises, *ECB Working Paper*, No. 145, 2002.

[13] M. Bussiere and C. Mulder, Political Instability and Economic Vulnerability, *IMF Working Paper*, 1999.

[14] Nor Azuana Ramli, Mohd Tahir Ismail and Hooy Chee Wooi, Designing Early Warning System: Prediction Accuracy of Currency Crisis by using k-nearest Neighbour Method, *World Academy of Science, Engineering and Technology*, Vol. 79, 2013, pp. 1023-1028.

[15] T. A. Peltonen, Are Emerging Market Currency Crises Predictable? A Test, *ECB Working Paper*, No. 571, 2006.

[16] V. Cerra and S. C. Saxena, Contagion, Monsoons, and Domestic Turnoil in Indonesia: A Case Study in the Asian Currency Crisis, *IMF Working Paper*, 2000.

[17] J. Kim, B-S. Kim and S. Savarese, Comparing Image Classification Methods: K-Nearest Neighbor and Support-Vector-Machines, *Applied Mathematics in Electrical and Computer Engineering*, pp. 133-138. Retrieved from: www.wseas.us/e.../MATHCC-18.pdf

[18] T. G. Dietterich, Machine-learning Research: Four Current Directions, *Al-Magazine*, Vol. 18, No. 4, 1997, pp. 97-136.

[19] H. Y. Chang, An Experimental Comparison of Two Methods Enterprise Voluntary Turnover Prediction Methods, *Computers and Simulation in Modern Science*, pp. 12-16. Retrieved from: www.wseas.us/e-library/.../CSMS02.pdf

[20] S. Yolacan and E. Asma, Information about Currency Crisis Given by KLR Approach, *4th WSEAS International Conference on Non-linear Analysis, Non-linear Systems and Chaos,* 2005, pp. 125-128.

[21] Lima et al., Ensembles of Support Vector Machines for Classification Tasks with Reduced Training Sets. Retrieved from: www.wseas.us/e-library/.../449-257.pdf

[22] H. Sallam, C.S. Regazzoni, I. Talkhan and A. Atiya, Evolvable Neural Networks Ensembles for Accidents Diagnosis, *12th WSEAS International Conference on Computers*, 2008, pp. 470-475.

[23] WEKA- www.cs.waikato.ac/nz/ml/weka

[24] The PRS Group - www.prsgroup.com