International Journal of Engineering, Mathematical and Physical Sciences
ISSN: 2517-9934
Vol:7, No:7, 2013

Designing Early Warning System: Prediction Accuracy of Currency Crisis by Using k-Nearest Neighbour Method

Nor Azuana Ramli, Mohd Tahir Ismail, and Hooy Chee Wooi

***Abstract*—**Developing a stable early warning system (EWS) model that is capable to give an accurate prediction is a challenging task. This paper introduces k-nearest neighbour (k-NN) method which never been applied in predicting currency crisis before with the aim of increasing the prediction accuracy. The proposed k-NN performance depends on the choice of a distance that is used where in our analysis; we take the Euclidean distance and the Manhattan as a consideration. For the comparison, we employ three other methods which are logistic regression analysis (logit), back-propagation neural network (NN) and sequential minimal optimization (SMO). The analysis using datasets from 8 countries and 13 macro-economic indicators for each country shows that the proposed k-NN method with k = 4 and Manhattan distance performs better than the other methods.

***Keywords*—**Currency crisis, k-nearest neighbour method, logit, neural network**.**

I. Introduction

CURRENCY crisis is like a disease in the economy and it's contagious. Therefore, detection of this disease at an early stage is needed so that crises like exchange-rate mechanism (ERM), Mexican peso, Asian Financial Crisis, Russian Flu, and others will not occur again. In order to detect currency crisis in the early stage, an early warning system (EWS) that can signal an early alarm is needed. The EWS basically set up an alert system on a list of warning indicators and monitor the dynamic of these indicators to ensure they do not exceed its warning limit.

Early warning system is not a new tool in the research field. In fact, it has long existed especially in detecting natural disasters, the spread of diseases, humanitarian emergencies, gross human rights violations, and economic crises. First EWS model in detecting currency crisis was invented by Krugman (1979) also Flood and Garber (1984). Their model was based on an adaptation of recurring discontinuous devaluations to interpret dynamics of the Mexican exchange rate in quarterly data from 1973 through 1982.

Nor Azuana Ramli is with the School of Mathematical Sciences, Universiti Sains Malaysia, Penang, Malaysia. (phone: +60173595498; e-mail: ajue.ramli@gmail.com).

Mohd Tahir Ismail, is with the School of Mathematical Sciences, Universiti Sains Malaysia, Penang, Malaysia. (e-mail: mtahir@cs.usm.my).

Hooy Chee Wooi is with the School of Management, Universiti Sains Malaysia, Penang, Malaysia. (e-mail: cwhooy@usm.my).

Then, researchers and economists decided to change the system from theoretical to empirical since the crisis cost was very high in terms of economic contraction, unemployment, and necessary financial restructuration process. There have been numerous studies in the literature on the leading indicators of currency crisis and one of the famous method is signaling approach which invented by Kaminsky et al. (1998).

Ever since that, EWS model in detecting currency crisis keeps developing until today. Several of methodological had been applied involving statistical and artificial intelligence in order to get better prediction results. The fundamental in designing a good early warning system is by using a model that is suitable with a reduction of cost and most importantly, accurate. Accuracy is an important element that must have in predicting or forecasting to measure how closed the results of our analysis to its actual value. In machine learning, accuracy is one of the most wanted measurements since the topic is widely studied.

In our study, we determine to apply the machine learning method in improving the accuracy of the model that will be used in designing our EWS later. There are various methods in machine learning such as a nearest neighbour method, linear discrimination, neural networks, kernel machines and others. The multilayer perceptrons which is also known as neural network is a very popular method and has been broadly used as EWS model in predicting currency crisis, for example Fratzscher (2003) and Kim et al. (2004). Meanwhile, support vector machine never been applied as EWS model but this method has been regularly used for bankruptcy prediction, such as by Jae and Young-Chan Lee (2005). The k-nearest neighbour method never been applied as EWS model either but it has been recurrently used for the prediction in biology, medicine, and other field.

We systematize this paper as; firstly, the method used and how we measure its performance will be explained in details in the Section II. Then, the results of the k-NN method will be presented in Section III and Section IV exemplify the comparison of k-NN with other methods which are logistic regression analysis (logit), back-propagation neural network (NN) and sequential minimal optimization (SMO). Last but not least, the conclusion of the experimental study and the discussion of potential future research will be summarized in the Section V.

II. Materials and Methodology

The first step in setting up our framework is to define the requirements needed for. In developing EWS, suitable indicators need to be chosen. Based on a literature survey and availability of the data, 13 macroeconomic variables that have been chosen as our indicators are unemployment, consumer price index, export, import, foreign direct investment, real GDP per capita, terms of trade, money supply, real effective exchange rate, government consumption, industrial production index, producer price index and foreign exchange reserves. Then, historical data will be collected to date prior crisis. In our experiment, we take data from 1st quarter of 1980 to 3rd quarter of 2012 for the eight selected countries under analysis via DataStream. Before we employ the data into a k-NN method, a brief review of this classifier will be discussed in the sub-section.

A. 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. The data sets are consequently allocated to the class that's more similar and k must be a positive integer. The value of k is usually small. When k=1, the datasets are basically allocated to the class of its nearest neighbour.

At first, the classifier was studied in 1951 by Fix and Hodges in the US Air Force School of Aviation Medicine. Then, Cover and Hart (1967) formalized the idea and invented the main properties of this method. 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**.** There are four different types of distance in the k-NN but in our analysis, we only consider two types of distance that are commonly used.

1. The Euclidean Distance

The k-NN is normally set to default to the Euclidean distance. This distance is calculated between a test sample and the specified training samples. For an example, let x_i be an input sample with r features (x_{i1}, x_{i2}, …., x_{ir}), n be the total number of input samples (i = 1,2,….,n) and r the total number of features (j = 1,2,….,r). The Euclidean distance between sample x_i and x_k (k = 1, 2,…., n) is identified as

$$d(x_i, x_k) = \sqrt{(x_{i1} - x_{k1})^2 + (x_{i2} - x_{k2})^2 + .. + (x_{ir} - x_{kr})^2} \quad (1)$$

2. The Manhattan Distance

Manhattan distance also known as rectilinear distance, city block distance, Minkowski's L_1 distance or taxi cab metric. The Manhattan name is taken from the place itself where the grid layout of most streets and it is based on the shortest path that anything could take between two intersections in the area to have equality length to the intersections' distance.

If an attribute is numeric, thus the local distance function can be written as:

$$dist'(x, y) = \sum_{i=1}^{n} |x_i - y_i| \quad (2)$$

Manhattan distance refers to the global distance that is calculated as the sum of these local distances.

B. Algorithms and Computing

The feature of the training data is proportional to the results of the classification. Simultaneously, the choices of the number of k also playing the important role as for the different number of k will give the different outcome. In computing the high-dimensional data, the k-NN algorithm is a straightforward method among others. However, if the set of test, train, and data dimension are too large, the computational complexity will be massive and it takes a long time to operate. To improve the efficiency of the algorithm, we use optimization which is Linear Nearest Neighbor Searching as it can lower precision.

In our experiment, we use the latest version of WEKA to pre-analyze the prediction of the currency crisis and the accuracy of k-NN method for every k=1 to k=5 with different distance. WEKA is software that is written in Java language and this very well-known collection of machine learning software was developed at the University of Waikato, New Zealand.

C. Performance Measures

In measuring which distance gives better performance and should be taken as our model in designing EWS, we use the area under the Receiver Operating Characteristics (ROC). The performance of a classifier can be shown by using ROC as its illustrate a tradeoff between selectivity and sensitivity. However, the area under the ROC which known as AUC is a convenient way of comparing classifiers. The AUC is linked to the Gini coefficient, G_1 with the formula $G_1 = 2AUC$-1, where:

$$G_1 = 1 - \sum_{i=1}^{n} (X_i - X_{i-1})(Y_i + Y_{i-1}) \quad (3)$$

Accuracy is a measurement of the AUC. A useful guidance on the classification for the accuracy of the AUC is shown as in Table I. The area determines the capability of the test to suitably class the data with and without currency crisis.

TABLE I
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

III. Performance and Analysis

K-nearest neighbour classifier's distance computation parameter was first set to the Euclidean metric with Train Test

Split Maker. In our experiment, we did not use distance weight parameter and the training was set to 20% and the rest of the data for testing. Fig. 1 in the Appendix A shows area under the ROC curve for the Euclidean distance and the Manhattan distance in the same plot. Based on the plot for 8 countries, we found that the Manhattan distance outperform than the Euclidean distance and *k*=4 give the best accuracy among others.

IV. Comparison with Other Methods

Since we already have a suitable model for the k-NN method, it is needed to evaluate its performance by comparison with other methods. In order to achieve this mission, we employ logistic regression analysis (logit), back-propagation neural network (NN) and sequential minimal optimization (SMO). Logit and neural network are chosen because both of the methods have been traditionally used in currency crisis prediction. Additionally, sequential minimal optimization (SMO) has achieved good results in classification even never been applied to currency crisis prediction. All of these methods will be explained in the rest of this section before we present the results of the comparison.

A. Logistic Regression Analysis

Logistic regression analysis is (logit) is a technique that used to approximate the probability of an event by predicting a binary dependent outcome of a set of independent indicators. In our case of predicting currency crisis occur or not, the linear probability model represented it as:

$$P_r = E(Y = 1|X_r) = \omega_1 + \omega_2 X_r \tag{4}$$

where X is the indicators and Y=1 means that there is a crisis occurred.

B. Neural Network

Neural networks are compilation of highly interrelated processing nodes that work concurrently to solve specific problems. In machine learning, it is under multilayer perceptions topic and is a nonparametric estimator that can be used for classification and regression. This paper employs back propagation algorithm as it is widely used amongst others form of neural network. By considering the hidden units as inputs, the second layer is a perceptron and let say given the inputs as y_k. For the first-layer weights, w_{ij}, we use the chain rule to calculate the gradient:

$$\frac{\partial E}{\partial w_{ij}} = \frac{\partial E}{\partial z_i}\frac{\partial z_i}{\partial y_k}\frac{\partial y_k}{\partial w_{ij}} \tag{5}$$

where z_i is the output. Thus, it's named back propagation as if the error propagates from the output back to the inputs.

Same as k-NN, we will use WEKA to run the analysis on prediction of the currency crisis and for the neural network model, we set epoch fix to 500 with learning rate equal to 0.3. An epoch is a complete pass over all the patterns in the training set. Meanwhile in gradient descent, the learning rate determines whether it is needed to make a change in magnitude. Usually, magnitude is set between 0 and 1.0, mostly less than or equal to 0.2. In our experiment, we set it for 0.2.

C. Sequential Minimal Optimization

Sequential minimal optimization (SMO) is a high-speed and an attractive optimization algorithm that calculates a solution to the Langrangian dual optimization problem. This algorithm works by break down the global optimization problem into some smaller optimization problems and it takes benefit that the smallest possible working set for the Langrangian dual optimization problem is a set of two training points.

What makes SMO astonishing is we can solve the optimization subproblem over the two training instances analytically, and hence a call to the costly optimization library like Support Vector Machine isn't required. Consider the subproblem of Langrangian dual optimization using the Langrangian multipliers βi and β_j :

$$\max_{\beta_i,\beta_j} \varphi'(\beta_i, \beta_j), \tag{6}$$

subject to the constraints:

$$z_i\beta_i + z_j\beta_j = \tau, \tag{7}$$
$$D \geq \beta_i, \beta_j \geq 0, \tag{8}$$

with i, j = 1,…..,m and $i \neq j$. The constant, τ can be calculated as,

$$\tau = -\textstyle\sum_{k=1,k\neq i,k\neq j}^{m} z_k\beta_k \tag{9}$$

To be precise, the values of βi and β_j are constrained by all the other Langrangian multipliers of the global optimization problem. This optimization problem can be shortened by rewriting constraint (7) as,

$$\beta_j = \frac{1}{z_j}(\tau - z_i\beta_i) \tag{10}$$

Include equation (10) into the equation (6), we get

$$\max_{\beta_i,\beta_j}\varphi'(\beta_i, \beta_j) = \max_{\beta_i}\varphi'(\beta_i, \frac{1}{z_j}(\tau - z_i\beta_i)) = \max_{\beta_i}\gamma(\beta_i) \tag{11}$$

By that, our two variable optimization problem becomes an optimization problem over the single variable, β_i. The objective function of this single-variable optimization problem is symbolized as γ (β_i). In addition, φ' (β_i, β_j) is a function with a unique global maximum where β_i and β_j lies on the unique saddle point of the original Langrangian optimization problem. This means that γ (β_i) also has a unique maximum and the answer can get by differentiating γ with respect to β_i,

$$\frac{d\gamma}{d\beta_i} = 0, \quad (12)$$

and solving for β_i.

By considering the Lagrangian dual optimization problem, we're capable to expand the algorithms from linear to the nonlinear separable training data case. The answer here is in the Lagrangian dual, where all data points in the input space show in the context of dot products. After all, we can substitute these dot products with suitable kernel functions by obtaining the benefit of the kernel trick.

One of the fascinating inferences of kernel functions is that the kernel trick can be applied to any linear classification algorithm which then suggests that it can be extended to the nonlinear case. In our analysis, we consider poly kernel as our kernel functions. The rest of the prediction accuracy is calculated by using WEKA.

D. Results for Comparison

How well does the k-NN method perform compared to other methods? A comparison of the percentage of accuracy, a percentage of the mean absolute error and percentage of false alarms for each classifier logit, NN, SMO and our proposed k-NN (k=4) for the 8 countries which are Mexico, Turkey, Hungary, Spain, Italy, Malaysia, South Africa and Brazil as shown in the Fig. 2 and table II respectively.

TABLE II
MAE AND FALSE ALARM FOR EACH CLASSIFIER

	Mean Absolute Error (%)			
Data set	k-NN	Logit	Neural Network	SMO
Mexico	13.09	12.51	15.44	13.33
Turkey	21.97	26.72	22.85	16.19
Hungary	9.12	7.72	12.23	8.57
Spain	16.13	16.19	99.6	16.19
Italy	16.13	12.5	12.52	17.14
Malaysia	25.71	21.01	20.94	26.67
South Africa	21.44	22.47	18.77	20.19
Brazil	14.26	21.95	15.53	21.9
Average	17.23	17.63	27.24	17.52
	False Alarm (%)			
Mexico	0	33.33	33.33	22.22
Turkey	0	35.29	35.29	0
Hungary	0	57.14	57.14	57.14
Spain	0	5.56	66.67	5.56
Italy	0	33.33	33.33	5.56
Malaysia	25	28.57	35.71	0
South Africa	9.52	28.57	23.81	0
Brazil	52.17	69.57	91.30	0
Average	10.84	36.42	47.07	11.31

From the comparison of the prediction accuracy of these classifiers histograms, we can't see which classifier outperform than others.

Therefore, by referring to the average of the mean absolute value and false alarm percentage, we can conclude that our proposed method k-NN is the most suitable model to use in designing EWS for currency crisis prediction. Besides providing minimize the number of mean absolute error, k-NN also compute less false alarm than others method.

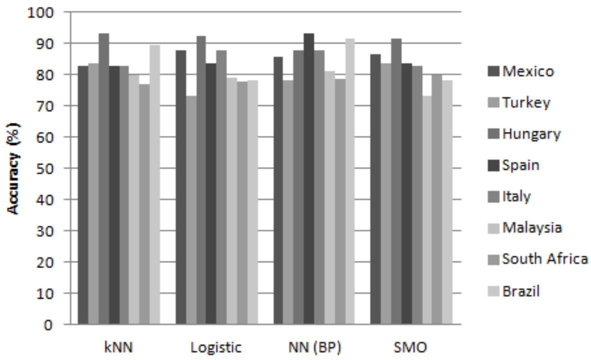


Fig. 2 Prediction accuracy for each classifier

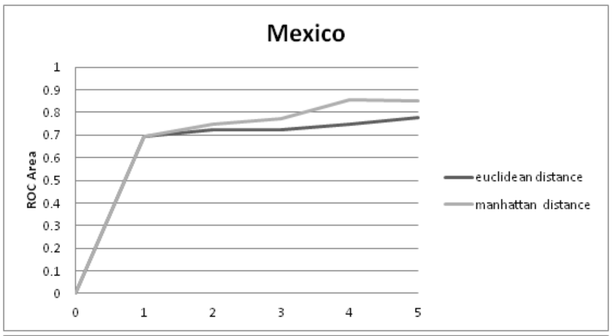
V. CONCLUSION

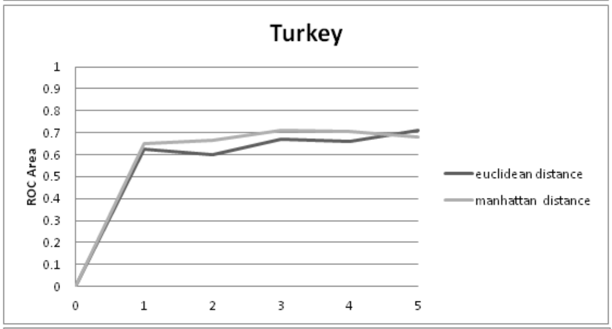
In this paper, we have employed four different types of methods in order to find the suitable model to use in our EWS. By using WEKA, experimentation of each method is done for currency crisis prediction based on 13 indicators for 8 countries. The comparison of the predictive ability of the various classifiers is done by using area under the ROC curve, percentage of prediction accuracy,a percentagee of the mean absolute error and percentage of false alarms. From the experiment that had been conducted, clearly k-NN with the Manhattan distance for k=4 is the most outperform method compared to k-NN with the Euclidean distance method, logit, sequential minimal optimization and backpropagation neural network.

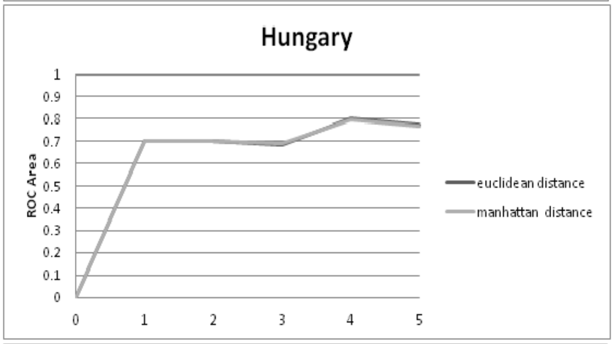
Even the average prediction accuracy of k-NN is only 84% compared to neural network that has a higher average prediction accuracy amongst other classifiers which is 85.5%, k-NN is still the model that more suitable than neural network since it will generate lowest mean absolute and false alarm. The prediction accuracy for the model maybe can improve if the k-NN classifiers combine with another classifier. It is something that can do in future research since our study only focus on single classifier.

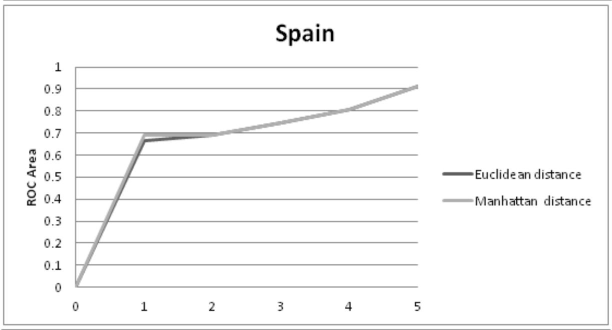
International Journal of Engineering, Mathematical and Physical Sciences
ISSN: 2517-9934
Vol:7, No:7, 2013

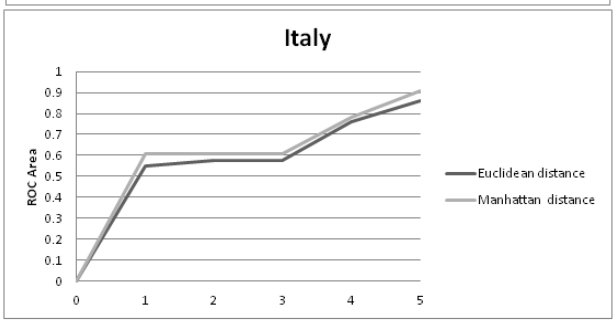
APPENDIX

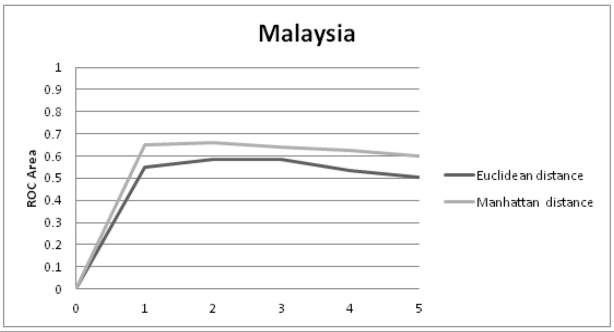


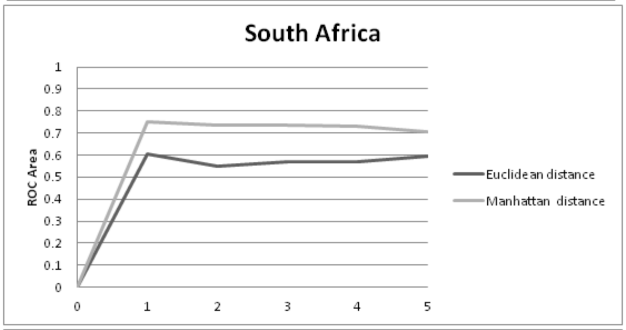












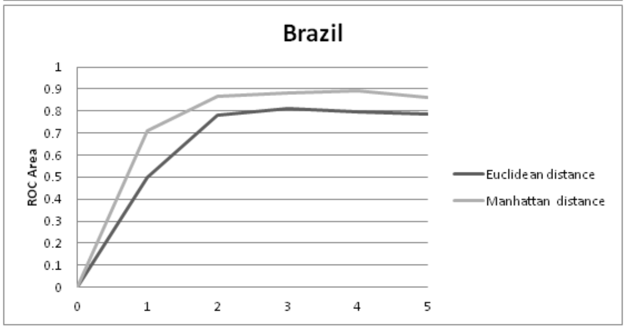


Fig. 1 Area under the ROC curve for the Euclidean distance and the Manhattan distance for 8 countries

REFERENCES

[1] M. Gassner, B. Brabec, “Nearest neighbour models for local and regional avalanche forecasting,” *Natural Hazards and Earth System Sciences*, Vol. 2, pp. 247-253, 2002.

[2] Quansheng Kuang, Lei Zhao, “A practical GPU based kNN algorithm,” in *Proc. 2nd International Computer Science and Computational Technology*, China, 2009, pp. 151-155.

[3] Javier Arroyo, Carlos Maté, “Forecasting histogram time series with k-nearest neighbours methods,” *International Journal of Forecasting*, vol. 25, pp. 192-207, 2009.

[4] Madhavi Pradhan, “Design of Classifier for Detection of Diabetes using Neural Network and Fuzzy k-Nearest Neighbor Algorithm,” *International Journal of Computational Engineering Research*, vol. 2, no. 5, pp. 1384-1387, Sept. 2012.

[5] Jieh-Haur Chen, “KNN based knowledge-sharing model for severe change order disputes in construction,” *Automation in Construction*, vol. 17, pp. 773-779, 2008.

[6] Jae H. Min, Young-Chan Lee, “Bankruptcy prediction using support vector machine with optimal choice of kernel function parameters,” *Journal Expert Systems with Applications*, vol.28, pp. 603-614, 2005.

[7] James C. Bezdek, Siew K. Chuah, “Generalized k-nearest neighbour rules,” *Fuzzy Sets and Systems*, vol.18, no.3, pp. 237-256, 1986.

[8] Shixin Yu et al., “Genetic feature selection combined with composite fuzzy nearest neighbor classifiers for hyperspectral satellite imagery,” *Pattern Recognition Letters*, vol. 23, pp. 183-190, 2002.

[9] Lamartine Almeida Teixeira, Adriano Lorena, "A method for automatic stock trading combining technical analysis and nearest neighbor classification," *Expert Systems with Applications*, vol. 37, pp. 6885-6890, 2010.
[10] S. Arya et al., "An optimal algorithm for approximate nearest neighbor searching fixed dimensions," *Journal of the ACM*, vol. 45, no. 6, pp. 891-923, 1998.
[11] K. Q.. Weinberger, L.K. Saul, "Distance metric learning for large margin nearest neighbor classification," *Journal of Machine Learning Research*, vol.10, pp. 207-244, 2009.
[12] J. Goldberger et al., "Neighbourhood components analysis," in *Proceedings of the Conference on Neural Information Processing Systems*, 2004.
[13] Lutz Hamel, *Knowledge Discovery with Support Vector Machines*. New Jersey: A John-Wiley & Sons, 2009.
[14] Ethem Alpaydin, *Introduction to Machine Learning,* 2nd ed. Massachusetts: The MIT Press, 2010.
[15] Trevor Hastie, Robert Tibshirani, Jerome Friedman, *The Elements of Statistical Learning: Data Mining, Inference, and Prediction,* 2nd ed. New York: Springer-Verlag, 2008.
[16] C. M. Bishop, *Pattern Recognition and Machine Learning.* New York: Springer-Verlag, 2007.
[17] WEKA – www.cs.waikato.ac/nz/ml/weka