Classification of Liver Disease Diagnosis: A Comparative Study

Sina Bahramirad, Aida Mustapha, Maryam Eshraghi
Faculty of Computer Science and Information Technology, Universiti Putra Malaysia
43400 UPM Serdang, Selangor, Malaysia
Email: aida_m@upm.edu.my

***Abstract*—Medical Data Mining (MDM) is one of the most critical aspects of automated disease diagnosis and disease prediction. MDM involves developing data mining algorithms and techniques to analyze medical data. In recent years, liver disorders have excessively increased and liver diseases are becoming one of the most fatal diseases in several countries. In this study, two real liver patient datasets were investigated for building classification models in order to predict liver diagnosis. Eleven data mining classification algorithms were applied to the datasets and the performance of all classifiers are compared against each other in terms of accuracy, precision, and recall. Several investigations have also been carried out to improve performance of the classification models. Finally, the results shown promising methodology in diagnosing liver disease during the earlier stages.**

I. Introduction and Related Work

Medical data mining involves extracting and analysing the medical data in building certain prediction model to increase the accuracy of diagnosis in any specific disease. However, only few works in data mining investigate liver disorders, although this disease is aggressively increasing and becoming one of the most fatal diseases in some countries [1].

Ramana et al. [2] develop a classification model to predict liver disease diagnosis using five popular classification algorithms and evaluate the performance of each model in terms of accuracy, precision, sensitivity and specificity. The study shows that the performance of all classifiers are better in one dataset (AP Liver dataset) as opposed to the other (BUPA Liver dataset) due to highly significant attributes such as total count of bilirubin, direct bilirubin and indirect bilirubin in the AP dataset.

Further, they pursue this study in another work with consideration of improved Bayesian classification technique with bagging and boosting techniques for the liver diseases diagnosis in effort to enhance the performance of classifiers [3]. Subsequent work by the authors [4] repeat the same investigation using statistical approach via ANOVA and MANOVA analysis. The study indicates there are more significant differences in the groups with all existing attribute combinations except analysis on the feature SGPT (Alamine Aminotransferase) between non-liver patients of AP and BUPA datasets.

In [5], the work propose another classification model for liver disease diagnosis by using Modified Rotation Forest method as the baseline. Next, the baseline method is hybridized with a set of classification algorithms and feature selection techniques from different category of classification models. The chosen categories of selected classification algorithms are tree-based, statistical-based, neural networks-based, rule-based and lazy learners.

Various machine learning algorithms have been explored for liver disease diagnostic classification, such as case-based reasoning (CBR) and classification and regression tree (CART) [6], Naive-Bayes, K-Star and FT Tree [7], ensemble methods [8] of Artificial Neural Network (ANN), Learn by Example (LEM) and Fuzzy Logic [9]. ANN has also been used in building classification model to predict diagnosis of hepatitis virus [10]. This work uses both of neural network popular modes, supervised and unsupervised neural network models using various architectures. The results of supervised model are found to produce better performance than the unsupervised model.

In analyzing the performance of decision tree classifier in liver disease, Sug [11] suggest oversampling the data instances in the liver dataset to improve high error rate in the classification model. Another branch of approach is the use of probabilistic classifier [1]. In this work, they propose a network-based technique elicited from human experts and the numerical parameters were learned from a database of cases. However, in their model the system cannot judge automatically and it is dependent to an expert judgment.

The objective of this study is to improve the performance of classification model for liver diagnoses by using different classification algorithms that have not been atempted previously. The results are hoped to complement previous findings in achieving exhaustive comparison of accuracy percentage. To achieve this, the experiments in this paper will utilize similar dataset as the previous research, as well similar evaluation criteria, which is accuracy, precision and recall.

The rest of this paper is structured as follows. Section 2 presents the overview of liver patient dataset. Section 3 describes the classification experiments including the algorithms that are chosen to carry out the experiments. Section 4 present the results from classification experiments followed by sone discussion. Finally, Section 5 concludes with some indication for future work.

II. Liver Patient Dataset

In this study, two liver patient datasets have been chosen for the purpose of building and testing the classification models. Both datasets are taken from University of California in Irvine (UCI) Machine Learning Repository [12]. The first dataset is from Andhra Pradesh state of India (AP dataset) and the

ISBN: 978-1-4673-5256-7/13/$31.00 ©2013 IEEE

TABLE I. ATTRIBUTES IN AP LIVER DATASET

Attribute	Type
Gender	Categorical
Age	Real number
Total_bilirubin	Real number
Direct_bilirubin	Real number
Indirect_bilirubin	Real number
Total_proteins	Real number
Albumin	Real number
Globulin	Real number
A/G ratio	Real number
SGPT	Integer
SGOT	Integer
ALP	Integer
Selector Field	Binomial (Class)

TABLE II. ATTRIBUTES IN BUPA LIVER DATASET

Attribute	Type
Attribute	Type
Mcv	Integer
Alkphos	Integer
SGPT	Integer
SGOT	Integer
Gammagt	Real number
Selector Field	Binomial (Class)

second dataset is from California state of USA (BUPA dataset). The AP dataset contains 583 records of liver patients of the particular state while the BUPA dataset contains 345 records.

A. AP Dataset

This dataset contains 416 liver patient records and 167 non-liver patient records. The dataset was collected from north east of Andhra Pradesh, India. The variable Selector is a class labeled by expert used to divide into groups (liver patient or otherwise). This dataset contains 441 male patient records and 142 female patient records [13]. Table I shows the details and attributes of the AP dataset.

B. BUPA Dataset

This dataset contains 345 records of liver patient from USA. There are 5 common attributes between BUPA and AP datasets. The first five variables are all blood tests which are thought to be sensitive to liver disorders resulting from excessive alcohol consumption [14]. Table II shows the details and attributes of the BUPA dataset.

In both datasets, the selector field is defined as class label and number. If the diagnosis is positive it comes under Class 1 category and if the diagnosis is negative it comes under Class 2 category.

III. CLASSIFICATION EXPERIMENTS

For the purpose of building the classification model for the liver patient diagnosis, a data mining tool called the RapidMiner [15] is used, which support plug-ins of Weka. This feature enable the RapidMiner to use all operations in the well-known Weka.

A. Data Pre-Processing

Because both datasets are rather matured, only a small data pre-processing is required. The AP dataset has 4 missing values, which are in the Albumin and Globulin Ratio attribute. We used a Replace missing value operator in RapidMiner to handle the missing values and we set the average value for replacement of those missing values. Next, we used the Remove Duplicate records operator in order to avoid any duplicate record in our datasets.

B. Classification Algorithms

- Logistic – This algorithm is used for building a multinomial logistic regression model with a ridge estimator and using the model afterward.
- Linear Logistic Regression (Simple Logistic) – This algorithm is very similar to Logistic, while the only difference is that Simple Logistic is used for building linear logistic regression models.
- Bayesian Logistic Regression – Another Logistic-based algorithm which is used for building Bayesian Logistic Regression model using both Gaussian and Laplace Priors.
- Logistic Model Trees (LMT) – This algorithm is a hybrid model of merging tree-based classification algorithms with Logistic regression models. The upper level of the tree is tree-based algorithms and the leaves are using Logistic-based regression models to do the classification. This algorithm can deal with binary and multi-class target variables, numeric and nominal attributes and missing values.
- Multilayer Perceptron – This algorithm uses back-propagation technique to perform the classification. This is similar to Neural Network operator but this algorithm is sourced from the Weka library.
- K-STAR – K* is an instance-based classifier, that is the class of a test instance based upon the class of those training instances similar to it as determined by some similarity function. It differs from other instance-based learners in that it uses an entropy-based distance function.
- RIPPER – This class implements a propositional rule learner, Repeated Incremental Pruning to Produce Error Reduction (RIPPER), which was proposed by William W. Cohen as an optimized version of IREP.
- Neural Net – This algorithm builds a model by implementing a feed-forward neural network trained by a back-propagation algorithm (multi-layer perceptron). This is the Neural Network classification operator in RapidMiner.
- Rule Induction – This algorithm learns a pruned set of rules with respect to the information gain. This operator works similar to RIPPER. Starting with the less prevalent classes, the algorithm iteratively grows and prunes rules until there are no positive examples left or the error rate is greater than 50%. In the growing phase, for each rule, conditions are greedily added to the rule until the rule is perfect (i.e. 100% accurate). The procedure tries every possible value of each attribute and selects the condition with highest information gain.
- Support Vector Machine (SVM) – This is a type of supervised learning model with associated learning

algorithms that analyze data and recognize patterns, used for classification and regression analysis. The basic SVM takes a set of input data and predicts, for each given input, which of two possible classes forms the output, making it a non-probabilistic binary linear classifier.

- Classification and Regression Trees (CART) – This is a non-parametric decision tree learning technique that produces either classification or regression trees, depending on whether the dependent variable is categorical or numeric, respectively.

C. Assessment Criteria

For each algorithm, we measure the accuracy, precision and recall in order to compare the performance of different classification algorithms against each other. The measurement metrics are defined as follows:

- Accuracy – Accuracy of a classifier is the percentage of the test set tuples that are correctly classified by the classifier.

$$accuracy = \frac{TP + TN}{TP + FP + TN + FN} \quad (1)$$

- Precision – Precision is defined as the proportion of the true positives (TP) against all the positive results, which are both true positives (TP) and false positives (FP).

$$precision = \frac{TP}{TP + FP} \quad (2)$$

- Recall – Recall or Sensitivity is also referred as true positive (TP) rate i.e. the proportion of positive tuples that are correctly identified.

$$recall = \frac{TP}{TP + FN} \quad (3)$$

where TP is true positive rate, TN is true negative rate, FP is false positive rate and FN is false negative rate.

IV. Results and Discussion

Several algorithms were imported from Weka add-on for RapidMiner data mining tool in order to perform the classification experiment. In total, eleven different classification algorithms have been used to classify the data in both AP and BUPA datasets and the following results have been obtained. No special modifications have been done on the properties of each classifier operators in RapidMiner. Table III and Table IV show the results from the classification experiments using AP and BUPA datasets, respectively.

Figure 1 shows the comparison of performance across all eleven classifiers. Note there is a considerable difference between the percentage of Precision and especially of Recall in the AP dataset as compared to results from the BUPA dataset. The Recall percentage of AP dataset is very low, which may suggest that number of false negative tuples after applying the model to the testing sets was considerably high. This may be due to the presence of some anomalies in the AP dataset. Further investigation needs to be done for finding the reason of this problem. To solve this, the first attempt was to use the Optimization selector in selecting the best features of the datasets in order to obtain a better result. The second attempt was to use Bayesian Boosting.

TABLE III. Classification Results from AP Liver Dataset

Classification Algorithm	Accuracy	Precision	Recall
Logistic	73.39%	57.69%	22.73%
Linear Logistic Regression	72.10%	52.17%	18.18%
Gaussian Processes	70.82%	41.67%	7.58%
Logistic Model Trees	72.10%	52.17%	18.18%
Multilayer Perceptron	72.10%	51.22%	31.82%
K-STAR	63.52%	36.99%	40.91%
RIPPER	71.24%	48.15%	19.70%
Neural Net	66.52%	43.62%	62.12%
Rule Induction	68.67%	44.44%	42.42%
Support Vector Machine	71.24%	47.83%	16.67%
Classification and Regression Trees	70.39%	47.06%	36.36%

TABLE IV. Classification Results from BUPA Liver Dataset

Classification Algorithm	Accuracy	Precision	Recall
Logistic	67.39%	75.00%	68.67%
Linear Logistic Regression	69.57%	74.70%	74.70%
Gaussian Processes	73.91%	79.01%	77.11%
Logistic Model Trees	68.12%	73.49%	73.49%
Multilayer Perceptron	68.84%	76.32%	69.88%
K-STAR	59.42%	71.43%	54.22%
RIPPER	64.49%	71.25%	68.67%
Neural Net	73.91%	77.65%	79.52%
Rule Induction	64.49%	76.56%	59.04%
Support Vector Machine	69.23%	75.00%	75.00%
Classification and Regression Trees	66.35%	77.36%	64.06%

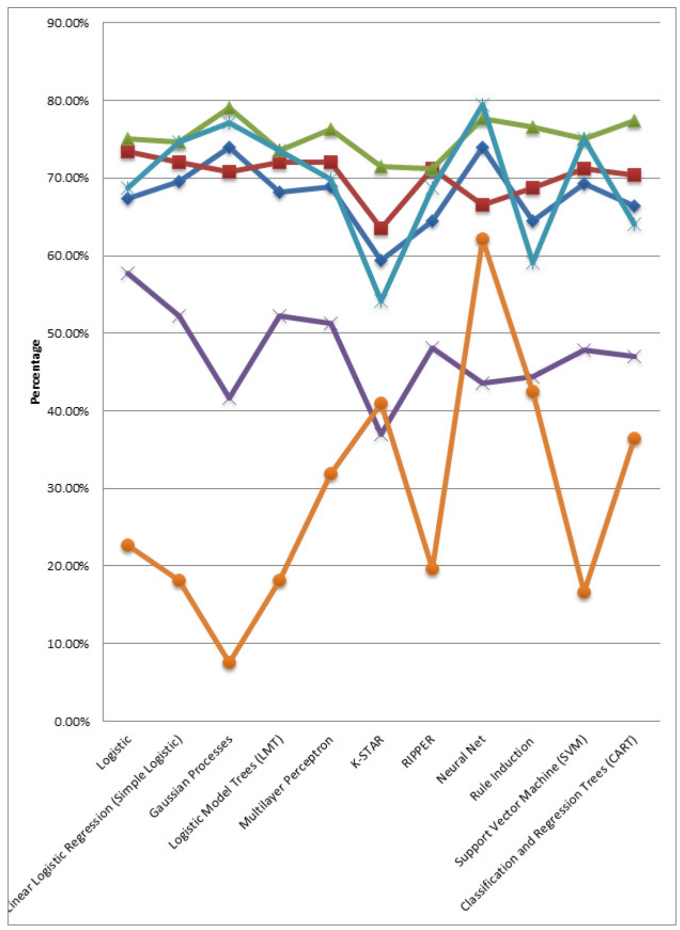
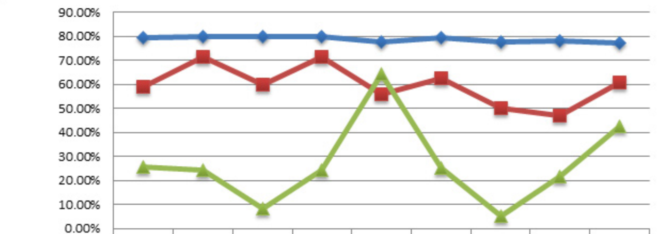
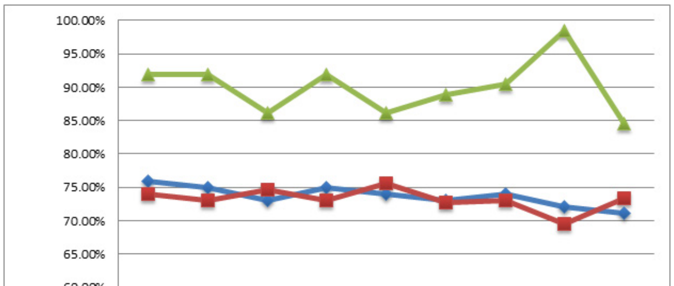


Fig. 1. Comparison of Performance



	Logistic	Linear Logistic Regression (Simple Logistic)	Gaussian Processes	Logistic Model Trees (LMT)	K-STAR	RIPPER	Rule Induction	Support Vector Machine (SVM)	Classification and Regression Trees (CART)
Accuracy	79.43%	80.00%	80.00%	80.00%	77.71%	79.43%	77.71%	78.29%	77.14%
Precision	58.82%	71.43%	60.00%	71.43%	55.77%	62.50%	50.00%	47.06%	60.61%
Recall	25.64%	24.39%	8.33%	24.39%	64.44%	25.00%	5.13%	21.62%	42.55%

Fig. 2. Brute Force Optimization on AP Dataset



	Logistic	Linear Logistic Regression (Simple Logistic)	Gaussian Processes	Logistic Model Trees (LMT)	K-STAR	RIPPER	Rule Induction	Support Vector Machine (SVM)	Classification and Regression Trees (CART)
Accuracy	75.96%	75.00%	73.08%	75.00%	74.04%	73.08%	74.04%	72.12%	71.15%
Precision	74.03%	73.08%	74.67%	73.08%	75.68%	72.73%	73.08%	69.57%	73.33%
Recall	91.94%	91.94%	86.15%	91.94%	86.15%	88.89%	90.48%	98.46%	84.62%

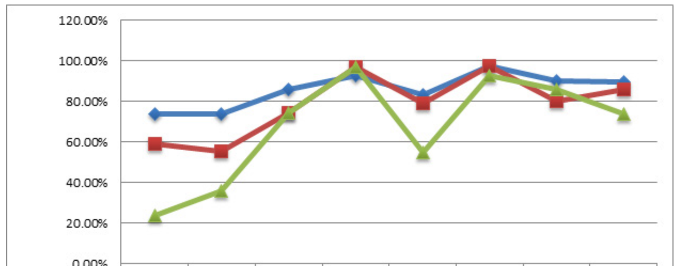
Fig. 3. Brute Force Optimization on BUPA Dataset

A. Brute Force Optimization

Several different attempts have been done for improving the obtained results of this experiment. We tried to reduce some of the features using the Chi-Square Attribute Evaluation but the results deteriorated. Finally, after testing different optimization technique in feature selections, we concluded that we get the best results using the Brute Force technique for feature selection. It was obvious that for every different algorithm that we have chosen we get a different set of attributes that have been chosen by Brute Force operator to test the model. However, we eliminated the Neural Network and Multilayer Perceptron algorithms from this experiment because of their extreme processing load. Figure 2 and Figure 3 show the optimized feature selection with Brute Force technique on AP and BUPA datasets, respectively.

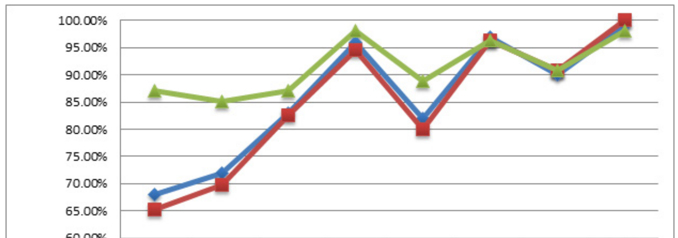
B. Bayesian Boosting

Bayesian Booster operator trains an ensemble of classifiers for Boolean target attributes. In each iteration, the training set is re-weighted, so that previously discovered patterns and other kinds of prior knowledge are sampled out. In this step, we used this operator to improve the performance of our model and the final results are very interesting. For the purpose of testing the model we have applied a chunk of the datasets (150 instances for AP and 100 instances for BUPA dataset). Figure 4 and Figure 5 show the optimized feature selection with Bayesian Boosting technique on AP and BUPA datasets, respectively.



	Logistic	Linear Logistic Regression (Simple Logistic)	Logistic Model Trees (LMT)	K-STAR	RIPPER	Rule Induction	Support Vector Machine (SVM)	Classification and Regression Trees (CART)
Accuracy	74.00%	74.00%	86.00%	92.67%	83.33%	97.33%	90.00%	89.33%
Precision	58.82%	55.56%	74.42%	96.97%	79.31%	97.50%	80.00%	86.11%
Recall	23.81%	35.71%	74.42%	96.97%	54.76%	92.86%	85.71%	73.81%

Fig. 4. Bayesian Boosting Optimization on AP Dataset



	Logistic	Linear Logistic Regression (Simple Logistic)	Logistic Model Trees (LMT)	K-STAR	RIPPER	Rule Induction	Support Vector Machine (SVM)	Classification and Regression Trees (CART)
Accuracy	68.00%	72.00%	83.00%	96.00%	82.00%	97.00%	90.00%	99.00%
Precision	65.28%	69.70%	82.46%	94.64%	80.00%	96.36%	90.74%	99.98%
Recall	87.04%	85.19%	87.04%	98.15%	88.89%	96.36%	90.74%	98.15%

Fig. 5. Bayesian Boosting Optimization on BUPA Dataset

C. Discussion

We obtained the best results by using the Bayesian Boost techniques for improving the performance. The best results in boosting techniques were with the tree-based classification algorithms. In this attempt, we also obtained our best results using the CART algorithm. By comparing our final performance results with the previous results [2], we can conclude that using Optimize feature selection and also different techniques of boosting, we were able to make sufficient improvement in the classification of this datasets.

V. Conclusion

In this study, several classification algorithms that have not been tested before on BUPA and AP datasets of Liver patients from UCI repository [12] have been considered for building classification model and assessing the performance of all classification models in terms of accuracy, precision and recall. In the first phase, the obtained results were insufficient, therefore we improved the results by performing optimization in selecting the attributes. Next, we also employed Bayesian Boosting in order to further improve the results. The results

ISBN: 978-1-4673-5256-7/13/$31.00

were then compared with previous works, which used the same datasets.

From the experiment, we found that the performance results on AP dataset was slightly better than the BUPA dataset in terms of Accuracy. The reason can be due to the difference between the number of instances or the number of attributes that are existing only in AP dataset. In some algorithms, the results of BUPA were better than in AP. The significant difference between the results on both datasets were in Precision and especially in Recall, whereby the percentage of Recall was very low in AP dataset. Also, the percentage for Precision was considerably lower in the AP dataset as compared to Precision and Recall from the BUPA dataset. Further studies need to be done for the purpose of inspecting the reason of considerable differences in Precision and Recall from both datasets.

REFERENCES

[1] A. Onisko, M. J. Druzdzel, and H. Wasyluk, A Bayesian Network Model for Diagnosis of Liver Disorders. In *Proceedings of the Eleventh Conference on Biocybernetics and Biomedical Engineering*, Vol. 2, pp. 842-846, 1999.

[2] B. V. Ramana, M. S. P. Babu and N. B. Venkateswarlu, A Critical Study of Selected Classification Algorithms for Liver Disease Diagnosis. *International Journal of Database Management Systems*, 3(2):101-114, 2011.

[3] B. V. Ramana, M. S. P. Babu and N. B. Venkateswarlu, A Critical Evaluation of Bayesian Classifier for Liver Diagnosis using Bagging and Boosting Methods. *International Journal of Engineering Science and Technology*, 3(4):3422-3426, 2011.

[4] B. V. Ramana, M. S. P. Babu and N. B. Venkateswarlu, A Critical Comparative Study of Liver Patients from USA and INDIA: An Exploratory Analysis. *International Journal of Computer Science Issues*, 9(3):506-516, 2012.

[5] B. V. Ramana, M. S. P. Babu and N. B. Venkateswarlu, Liver Classification Using Modified Rotation Forest. *International Journal of Engineering Research and Development*, 6(1):17-24, 2012.

[6] R. H. Lin, An Intelligent Model for Liver Disease Diagnosis. *Artificial Intelligence in Medicine*, 47(1):53-62, 2009.

[7] P. Rajeswari and G. Reena, Analysis of Liver Disorder using Data mining Algorithm. *Global Journal of Computer Science and Technology*, 10(14):48-52, 2010.

[8] B. Chen and H. X. Zhang, An Approach of Multiple Classifiers Ensemble based on Feature Selection. In *Fifth International Conference on Fuzzy Systems and Knowledge Discovery*, Vol. 2, pp. 390-394, 2008.

[9] S. Karthik, A. Priyadarishini, J. Anuradha and B. K. Tripathy, Classification and Rule Extraction using Rough Set for Diagnosis of Liver Disease and Its Types. *Advanced Applied Scientific Research*, 2(3):334-345, 2011.

[10] S. Ansari, I. Shafi, A. Ansari, J. Ahmad and S. I. Shah, Diagnosis of Liver Disease Induced by Hepatitis Virus using Artificial Neural Networks. In *2011 IEEE International Multitopic Conference (INMIC)*, pp. 8-12, 2011.

[11] H. Sug, Performance Comparison of Different Over-Sampling Rates of Decision Trees for the Class of Higher Error Rate in the Liver Data Set. *International Journal of Mathematics and Computers in Simulation*, 2(6):282-289, 2012.

[12] A. Frank and A. Asuncion, *UCI Machine Learning Repository* [http://archive.ics.uci.edu/ml]. Irvine, CA: University of California, School of Information and Computer Science, 2010.

[13] ILPD (Indian Liver Patient Dataset) Dataset. UCI Repository of Machine Learning Databases. Available in http://archive.ics.uci.edu/ml/datasets/ILPD, last accessed: 15 December 2012.

[14] BUPA Liver Disorders Dataset. UCI Repository of Machine Learning Databases. Available in http://archive.ics.uci.edu/ml/datasets/Liver+Disorders, last accessed: 15 December 2012.

[15] I. Mierswa, M. Wurst, R. Klinkenberg, M. Scholz and T. Euler, YALE: Rapid Prototyping for Complex Data Mining Tasks. In *Proceedings of the 12th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (KDD-2006)*, pp. 935-940, 2006.

ISBN: 978-1-4673-5256-7/13/$31.00 ©2013 IEEE