International Journal of Applied Engineering Research ISSN 0973-4562 Volume 13, Number 5 (2018) pp. 2780-2785

A Novel Hybrid Iterative Backward Feature Selection Framework for Intrusion Detection System

[1]R.Venkatarathinam, [2]Dr.V.Cyril Raj and [3]Dr.G.Victo Sudha George

[1]*Research Scholar, Department Of Computer Science and Engineering,* [2] *Dean Engineering and Technology,*
[3]*Associate Professor, Department Of Computer Science and Engineering,*
Dr. M.G.R. Educational and Research Institute University, Chennai-600095, Tamil Nadu, India
E-mail : [1]sudhajose72@gmail.com

Abstracts

The role of Intrusion Detection Systems (IDSs) in observing and analyzing the activities of computer systems is very much significant to detect security threats. But the redundant and uninformative features challenges the existing IDS. These features reduce the classification accuracy and also the slow down the process. In this paper a feature selection framework is proposed to identify a reduced set of most informative features from the training data set. The framework is a hybrid method for feature selection using mutual information based filter algorithm along with SVM-RFE. The evaluation results show that the proposed hybrid feature selection framework contributes more critical features for SVM based-IDS to achieve better accuracy and lower computational cost compared with other existing feature selection methods.

Keywords: Feature selection, Intrusion detection, Mutual information, support vector machines, 10 fold cross validation.

INTRODUCTION

Intrusion detection is the art of determining malicious network traffic patterns that are abnormal. In spite of the availability of the sophisticated security tools [1], attackers use complicated infiltration techniques and challenge the security personals. Hence there is a demand for efficient and reliable IDS to protect the computer network from vulnerabilities. The main purpose of these systems is to be perfect in detecting attacks with minimum false alarms. But to do those IDS should be able to handle large volume of network traffic and speed up the real time decision making.

In general an IDS handles huge volume of data with variety of traffic patterns. A set of features (or attributes) represent the characteristics of each pattern. They represent the pattern as point in a multi-dimensional feature space. A pattern might contain irrelevant and redundant features. Hence in practice, it is essential to remove those unwanted features to reduce the computational cost, speeds up classification process and the complexity of building a classifier. Thus, dimensionality reduction, such as feature extraction and feature selection, has been effectively used in the field of machine learning to overcome this difficulty. Feature extraction techniques transform the input features into a new feature set. On the other hand the most informative features are selected by Feature Selection (FS) algorithms from the given set of input features [2]. Here we focus on feature selection for dimensionality reduction. Feature Selection falls into three categories filter, wrapper and embedded method.

This paper presents a hybrid feature selection approach that uses the filter and the embedded technique to identify highly informative features. The filter method aims to reduce the computational cost of the embedded search by eliminating irrelevant and redundant features from the initial feature set. The embedded method is classifier specific. It chooses the highly class informative features that improve the classification accuracy. The objective of this work is to achieve both the high accuracy of embedded approaches and the efficiency of filter approaches.

From literature it is seen that Mutual Information (MI) can be used as a measure that very well quantify the amount of information shared between various features[3].Hence a filter-based feature selection method that rank features based on the theory of MI is formulated to assess the dependence between features and the output classes. This will aid the filter based search to eliminate uninformative features and present a ranked list of relevant features based on their importance. This incremental search method apply greedy search algorithm as their searching strategy. It is computationally attractive but "nesting effect" is problem faced by this method. Once a feature is discarded using the top-down approach, it cannot be added back. Therefore the final optimal subset may not contain all the best features. To cope with the "nesting effect" problem, 10 – fold CV SVM-RFE is further used to select the subset of the resultant features which improves the classification performance. 10-CV SVM-RFE involves an additional investigation step to find out whether a feature is really highly informative.

In order to examine the effectiveness of our proposed feature selection method, the selected feature subset is then passed through SVM classifier to build IDS. Experimental results presented for validation are obtained using different sets of NSL-KDD Cup 99 dataset that are commonly used in literature.

This paper is organized as follows: Section II explains the theory of mutual information for feature selection and the SVM-RFE. Section III brings in proposed hybrid feature selection algorithm. Section IV presents the experimental details and results. Finally, we conclude this paper by summarizing the work .

BASIC CONCEPTS AND ALGORITHM USED

The section explains the fundamentals of the theory of Mutual Information and also the SVM-RFE the embedded feature selection algorithm used.

A .Theory of Mutual Information

According to feature selection, the features having significant information about a class is termed as relevant and others are treated as uninformative features to the output class [4]. So it is essential to search for those informative features that hold as much information about the output class as possible. For this reason, information entropy and MI were proposed [5] to quantify the amount of information shared between two random variables. Assume $M=\{m_1,m_2,\ldots,m_k\}$and $N=\{n_1,n_2,\ldots,n_k\}$ are two random variables where k is the total number of samples. Equation (1) is used to compute the amount of knowledge called MI, shared between variable N and M.

$$I(M,N)=\sum_{m\in M}\sum_{n\in N} pd(m,n)\log\frac{pd(m,n)}{pd(m)p(n)} \quad (1)$$

where, pd(m,n) is a joint probability distribution , pd(m) the probability distribution of M and pd(n) the probability distribution of N.

The relation between two random variables is given by MI. The amount of MI is large if the two variables are closely related. When two variables are statistically independent the value of MI will be zero. It is noticed that the estimation of pdfs make the computation of MI difficult and the common estimation methods (e.g., histogram and kernel density estimations) are inclined to high-dimensional data [6]. So the estimator recommend by [7] is used here. This estimator relies on estimating information entropies from the data using an average distance of the k-nearest neighbors. It approximates MI between two random variables on a multi-dimensional data space by estimating the entropy based on the k-nearest neighbors technique.

B. Support Vector machine-Recursive Feature Elimination (SVM-RFE)

SVM-RFE (SVM-Recursive Feature Elimination) is the feature selection variant of SVM. Although simpler feature selection methods are existing [8], SVM-RFE is used as it has acknowledged good classification performance. Fundamentally, SVM-RFE is a multivariate iterative backward feature selection method in the sense that it considers feature interaction while evaluating the significance of features. During each iteration the left behind set of features are used to trains a linear SVM classifier, ranks them according to the squared values of feature weights. The feature will be removed.

A cost function J computed on training samples is used as an objective function. Expanding J in Taylor series to the second order using the OBD algorithm and neglecting the first order term at the optimum of J, yields

$$DJ(i) = \frac{1}{2}\frac{\partial^2 j}{\partial w_i^2} D(w_i)^2$$

Here $(wi)^2$ was used as the ranking criterion and we used LIBSVM (Library for Support Vector Machines) [9] with a linear kernel. This iterative Feature exclusion activity is carried out as long as all the features are eliminated or a preferred condition is met. We used SVM-RFE a soft margin based SVM using linear kernel.

PROPOSED HYBRID ITERATIVE BACKWARD FEATURE SELECTION FRAMEWORK (HIBFS)

In this section, we propose a hybrid feature selection approach that combines the advantages of both filter and embedded methods.

The framework of the proposed algorithm is shown in Fig. 1 which consists of two main stages: the first stage at which the mutual information is used for feature ranking and elimination, and the second stage which determines the optimal subset, and contributes maximum classification accuracy on training dataset. Suppose the total number of features considered in the dataset is k. The filtering process is applied selecting the features incrementally and eliminating any irrelevant and redundant features from the initial set. This phase will be continued until a specified number of features are selected. Then the embedded method is applied to find out the optimal feature subset to maximize the classification accuracy.

A. Filter method for feature pre-selection

The filter method plays an important role in the proposed hybrid method and is designed to eliminate irrelevant and redundent features. This helps the embedded method-based 10-CV SVM-RFE to decrease the searching range from the entire original feature space to the pre-selected features. The filter algorithm searches for relevant features by looking at the characteristics of each individual feature using MI as an evaluation criterion for the selection process. Choosing appropriate selection parameter is the problem faced by most of the existing MI based filter method. As a solution to this issue a new feature selection criterion is proposed which determines a feature that maximizes the term GF in equation (2). It selects a feature from a given input feature set to maximize I(C; fi) and to minimize the average redundancy RMR simultaneously. RMR, in equation (2), stands for the relative minimum redundancy of feature f_i against feature f_s.

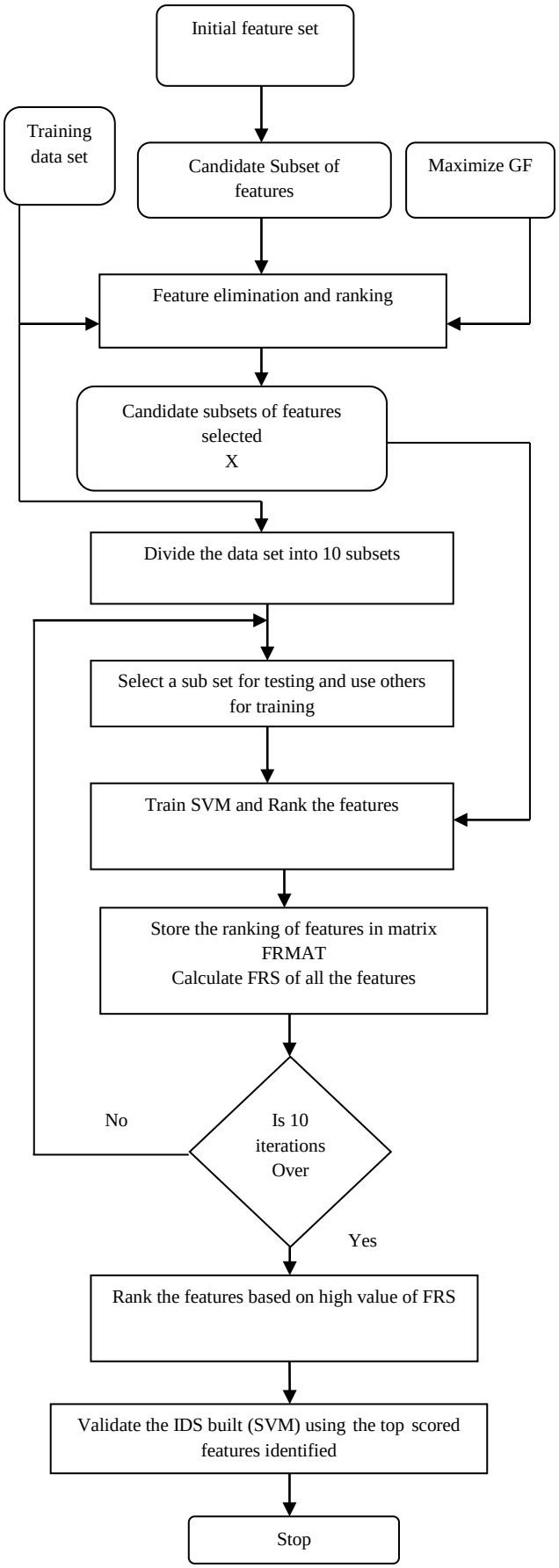

Figure 1. Proposed HIBFS Framework

Begin

Initialize : X= ϕ

$$GF = I(C, f_i) - \frac{1}{S} \sum_{f_s \in S} RMR \quad (2)$$

Algorithm 1: MI based filter method for feature pre-selection

Input: a training dataset with features F ={fi,i=1,….,n}

Output : X - the sebset of informative features

For each feature f ∈ Fdo

Compute GF using (2)

If (GF = 0 or GF < 0) then

F = F-{f_i}

Else

Rank the feature f_i based on the

value of GF

X=X ∪ {f_i}

End

End

Return X

The properties of the value of GF in (3) are as follows: If (GF = 0), then feature fi is irrelevant to the output class C and cannot provide any additional classification information. Likewise if (GF < 0), then feature fi is redundant to the output class C. Therefore, in both of the above cases fi can be discarded. On the other hand, if (GF > 0), then feature fi is relevant to the output class C and can provide some additional classification information. Hence, the current candidate feature fi should be included into X. Thus in the proposed filter approach,a numerical threshold value greater than zero is set for GF. The feature pre-selection process is given by algorithm 1.

B. 10-fold Cross Validation SVM-RFE

Once the filter method finishes its task, the next stage determine the optimal feature subsets, using embedded scheme that can produce the best classification performance. The proposed method is an iterative feature selection approach using SVM-RFE. Though from literature it is witnessed that SVM-RFE is a leading method for feature selection its performance can go down due its greedy nature. So to overcome the ill effect K-CV is adopted. The method of HIBFS is given as follows.

In k-fold cross-validation, the original set of samples is portioned evenly into k groups. Of the k subgroups, a subgroup is kept for validating the model constructed from the training dataset and other k - 1 sub groups are used for training the classifier. This task is repeated k times which ensure that each group get a chance only once to be a validation set. Thus, if we are selecting d out of D features in this way, k different feature sets of dimension d may be chosen. Now, the features with high degree of occurrence [10, 11] might be used in the classification system in future.

In this work, 10-fold cross validation is used to find out the highly class informative features that can minimize the false rate and improve the predictive accuracy of the IDS built. The samples available in the dataset are divided into 10 groups of subsamples. Each time one subset of samples is eliminated from the input and the residual samples are used as training set. Thus 10 sets of ranked features are generated.

The steps to be followed are given below.

1. Partition the data set into k subgroups.
2. Select a subgroup as testing dataset and keep all the remaining samples as training set.
3. Train SVM-RFE using the training set and generates the ranking of all the features.
4. Store the ranking of the f of the features in the matrix FRMAT
5. Repeat steps 2 to 4 and k times.
6. Calculate the Final Ranking Score(FRS) of all the features and select the highly class discriminative features having high FRS values.
7. Validate the classification model using the selected high FRS scored features.

Pseudo Code for computing FRS of all the features

```
For i= 1 to n // for all the features
    FRS(i) =0
End For
For R=1 to n// for all the features
For S= 1 to N// for each iteration
FRS(FRMAT(S,R)=FRS(FRMAT(S,R)+(n+1)–R
End for
End For
```

EXPERIMENTS

This section Explain about the NSL-KDD Cup 99 dataset, experimental environments and presents the evaluation results for the proposed method.

A. SVM

Support Vector Machines (SVMs) are good candidate for intrusion detection systems which can provide real-time detection capability, deal with large dimensionality of data. SVMs map the training vectors in high dimensional feature space through nonlinear mapping and labeling each data vector by its class. The data is then classified by identifying a set of support vectors that generate a hyper plane in the feature spaceas given in Fig 2. Support Vector Machine has become one of the popular techniques for intrusion detection due to their good generalization ability and the potential to overcome the curse of dimensionality.

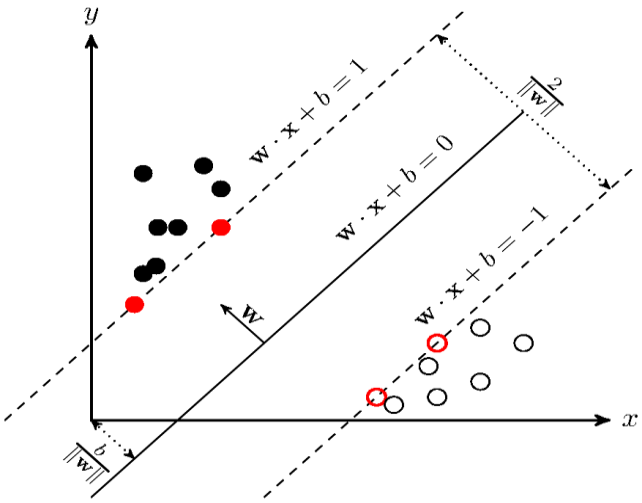

Figure 2. SVM Classifier

The SVM can select suitable setup parameters because it does not depend on traditional empirical risk such as neural networks [12]. One of the main advantage of using SVM for IDS is its speed, as the capability of detecting intrusions in real-time is very important. More over SVMs can work effectively on high dimensional data as the classification complexity is not influenced by the dimensionality of the feature space. The potential of dynamic training pattern updating is another merit of SVM. This will happen if there exist a new pattern at the time of training. [13]. Hence in this work SVM is used to build IDS to verify the predictive ability of the features identified by the proposed framework.

B. Data Set Used

For the evaluation NSL-KDD data set is used [14]. This data set is an advanced version of KDD-Cup that does not put up with issues such as redundancy and complexity level of data. Each NSL-KDD record contains 41 features and is labeled as either normal or an attack. The features in NSLKDD Cup 99 dataset are divided into three groups: (a) the basic input features of network connection including some flags in TCP connections, duration, prototype, number of bytes from source IP addresses or from destination IP addresses, and service, (b)

the content input features of network connections, and (c) the statistical input features that are computed either by a time window or a window of certain kind of connections. These features can be classified into three groups: connection based (9 features), content based (13 features) and time based (19 features). The attacks fall into four categories: Denial of service (DoS), Remote-to-Local (R2L), User-to-Root(U2R) and Probing.

The benefit of using NSL-KDD Cup 99 dataset is that the training and testing instances are reasonable and the experiments can be carried out on the total set of training and testing dataset without any random selection of dataset.

C. Experimental setup

The experiments are made on a 1.80 GHz Core (TM) 2 CPU personal computer with 2GB memory. The basic code for the SVM classifier is adopted fromWeka3.Weka3 is open source data mining software for machine learning applications [15]. The proposed algorithm was tested on NSL-KDD Cup 99 dataset. The intrusion detection model is built using SVM. The features selected by the proposed hybrid feature selection method are used to build the model. The 10-fold cross validation technique is used to verify the performance of the proposed hybrid feature selection approach. The overall dataset is divided into 10 subsets. Each time 9 sets are used for training and one set is used for testing. This is repeated 10 times. The performance is the average of the 10 cases.

D. Result and Discussion

The classification performance of the intrusion detection model constructed in combination with the proposed HIBFS is compared with the existing feature selection algorithms like MIFS and SVM-RFE The results given in table 1 clearly demonstrate that the classification performance of an IDS is enhanced by the proposed method. Moreover the proposed feature selection algorithm HIBFS shows promising results in terms of low computational cost and high classification results.

Table 1. Performance Comparison on Different Feature Selection Algorithms

Feature Selection Method Used	Detection Rate (DR)	False Positive Rate (FPR)	Accuracy
SVM-IDS + HIBFS	99.12	0.10	99.70
SVM-IDS + MIFS	95.26	0.25	97.42
SVM-IDS + SVM-RFE	98.29	0.34	98.15
SVM-IDS+ All features	92.16	0.45	95.69

Table 2. Average Computational Time For Constructing and Testing the Model

Type	Feature selection Method	Avg. Computational Time(Sec.)
Training Time	HIBFS	58.12
	MIFS	79.42
	All features	212.81
Testing Time	HIBFS	27.23
	MIFS	31.63
	All features	87.23

Table 2 gives the comparative study on average computational time for training and testing (in second) of the proposed hybrid feature selection model with MIFS and those using the entire 41 features. It is seen that the proposed hybrid method took less time compared with the other two. Also it is noticed that the proposed approach demonstrate the best average time of building and testing the model.

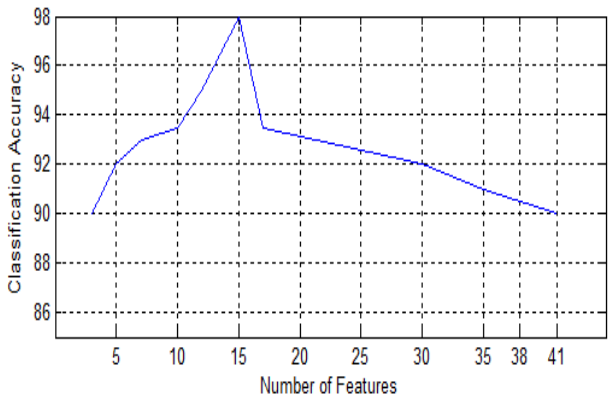

Figure 3. Classification accuracy on selected features of training data

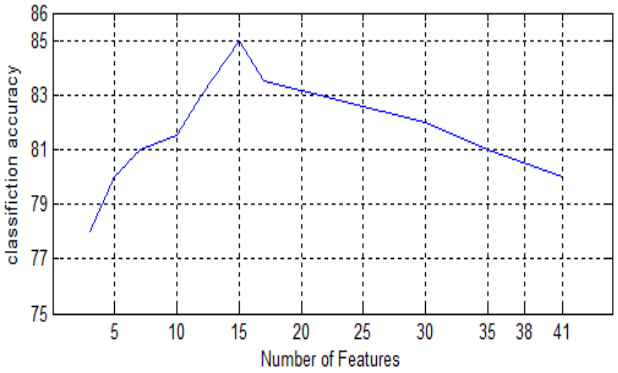

Figure 4. Classification accuracy on selected features of training data

Fig 3 and 4 depict the classification accuracy achieved by SVM classifier in the training phase and the testing phase for varied number of features. It is the average of the 10-fold Cross Validation. It is evidentially shown that accuracy is

more during tainting than testing. The inference is that the feature selection strategy has to be further strengthened.

CONCLUSION

In this paper, for improving the predictive accuracy of the Intrusion Detection System a hybrid feature selection framework that takes advantage of both filter and the embedded method is proposed. The proposed framework has been evaluated using the well known NSL-KDD data set. Experiments conducted on the data set show evidence of promising results in terms of classification accuracy and low computational complexity. In addition, compared with other promising feature selection methods the proposed method has shown good comparable results in terms of detection, false positive and accuracy rate. Thus, the experimental results demonstrated that the proposed hybrid method achieved better performance in detecting intrusions.

Although the proposed hybrid feature selection algorithm HIBFS has shown hopeful performance, it could be further enhanced by optimizing the feature selection strategy. In future we are planning for ensemble feature selection to further improve the predictive accuracy.

REFERENCES

[1] C. C. Center, "Overview of attack trends," 2002.

[2] S. Cang and H. Yu, "Mutual information based input feature selection for classification problems," Decision Support Systems, 2012.

[3] M. S. Roulston, "Estimating the errors on measured entropy and mutual information," *Physica D: Nonlinear Phenomena*, vol. 125, no. 3, pp.285–294, 1999.

[4] Q. Wang, H.-D. Li, Q.-S. Xu, and Y.-Z. Liang, "Noise incorporated subwindow permutation analysis for informative gene selection using support vector machines," *Analyst*, vol. 136, no. 7, pp. 1456–1463, 2011.

[5] SHANNON-WEAVER, Mathematical theory of communication. University Illinois Press, 1963.

[6] F. Rossi, A. Lendasse, D. François, V. Wertz, and M. Verleysen, "Mutual information for the selection of relevant variables in spectrometric nonlinear modelling," *Chemometrics and intelligent laboratory systems*,vol. 80, no. 2, pp. 215–226, 2006.

[7] A. Kraskov, H. Stögbauer, and P. Grassberger, "Estimating mutual information," *Physical Review E*, vol. 69, no. 6, p. 066138, 2004.

[8] Tian Rui, Basu MK, Capriotti1 E. Contrast Rank: a new method for ranking putative cancer driver genes and classification of tumor samples. *Bioinformatics*. 2014; 30(17):i572–8.

[9] Chang CC, Lin CJ "LIBSVM: a Library for Support Vector Machines", 2001

[10] Chin Lynda, Hahn WC, Getz G, Meyerson M. Making sense of cancer genomic data. Genes & Dev. 2015; 25:534–55.

[11] Schuleruda H, Albregtsen F. Many are called, but few are chosen. Feature selection and error estimation in high dimensional spaces. *Comput Meth Programs Biomed*. 2004; 73:91–9.

[12] T.Shon, Y. Kim, C.Lee and J.Moon,(2005), A Machine Learning Framework for Network Anomaly Detection using SVM and GA, Proceedings of the 2005 IEEE.

[13] SandyaPeddabachigari, Ajith Abraham, CrinaGrosan, Johanson Thomas (2005). Modeling Intrusion Detection Systems using Hybrid Intelligent Systems.Journal of Network and Computer Applications.

[14] M. Tavallaee, E. Bagheri, W. Lu and A. A. Ghorbani, "A Detailed Analysis of the KDD CUP 99 Data Set", *In Proc. of the 2009 IEEE symposium on computational Intelligence in security and defense application (CISDA)*, Ottawa, ON, Canada, 2009, pp.1–6.

[15] Mark Hall, Eibe Frank, Geoffrey Holmes, Bernhard Pfahringer, Peter Reutemann, and Ian H. Witten, "The WEKA Data Mining Software: An Update," *SIGKDD Explorations*, Vol. 11, No. 1, 2009, http://www.cs.waikato.ac.nz/ml/we ka/, Accessed 26.08.2014