Makara Journal of Technology

Volume 22 | Issue 2 Article 1

8-2-2018

Internet Traffic Forecasting Model Using Self Organizing Map and Support Vector Regression Method

Enrico Laoh
Industrial Engineering Department, Faculty of Engineering, Universitas Indonesia, Depok 16424, Indonesia, enrico.laoh@ui.ac.id

Fakhrul Agustriwan
Industrial Engineering Department, Faculty of Engineering, Universitas Indonesia, Depok 16424, Indonesia

Chyntia Megawati
Department of Integerated System Engineering, College of Engineering, Ohio State University, Columbus, OH 43210, USA

Isti Surjandari
Industrial Engineering Department, Faculty of Engineering, Universitas Indonesia, Depok 16424, Indonesia

Recommended Citation

Laoh, Enrico; Agustriwan, Fakhrul; Megawati, Chyntia; and Surjandari, Isti (2018) "Internet Traffic Forecasting Model Using Self Organizing Map and Support Vector Regression Method," *Makara Journal of Technology*: Vol. 22: Iss. 2, Article 1.
DOI: 10.7454/mst.v22i2.3351
Available at: https://scholarhub.ui.ac.id/mjt/vol22/iss2/1

Makara J. Technol. 22/2 (2018), 60-65
doi: 10.7454/mst.v22i2.3351

Internet Traffic Forecasting Model Using Self Organizing Map and Support Vector Regression Method

Enrico Laoh[1]*, Fakhrul Agustriwan[1], Chyntia Megawati[2], and Isti Surjandari[1]

1. Industrial Engineering Department, Faculty of Engineering, Universitas Indonesia, Depok 16424, Indonesia
2. Department of Integerated System Engineering, College of Engineering, Ohio State University, Columbus, OH 43210, USA

e-mail: enrico.laoh@ui.ac.id

Abstract

Internet traffic forecasting is one of important aspect in order to fulfill the customer demand. So, the service quality of internet service provider (ISP) can be maintained at the good level. In this study self organizing map (SOM) and support vector regression (SVR) algorithm are used as forecasting method. SOM is first used to decompose the whole historical data of traffic internet into clusters, while SVR is used to build a forecasting model in each cluster. This method is used to forecast ISPs traffic internet in Jakarta and surrounding areas. The result of this study shows that SOM-SVR method gives more accurate result with smaller error value compared to that of the SVR method.

Abstrak

Model Perkiraan Trafik Internet Menggunakan Metode *Self Organizing Map* dan *Support Vector Regression*. Peramalan trafik internet merupakan salah satu hal yang penting agar permintaan konsumen dapat dipenuhi, sehingga kualitas pelayanan dari penyedia layanan dapat terjamin dengan baik. Pada penelitian ini, digunakan metode peramalan berupa kombinasi algoritma *self organizing map* (SOM) dan *support vector regression* (SVR). Metode SOM digunakan untuk membagi data historis trafik internet secara keseluruhan ke dalam beberapa klaster, sedangkan metode SVR digunakan untuk membentuk model peramalan pada setiap klaster yang terbentuk. Penelitian dilakukan dengan mengaplikasikan metode peramalan SOM-SVR pada data trafik penyedia jasa internet di Jakarta dan sekitarnya. Hasil peramalan pada penelitian ini menunjukkan bahwa model peramalan dengan metode SOM-SVR dapat memberikan prediksi yang lebih akurat terkait nilai error yang lebih kecil dibandingkan dengan metode SVR tunggal.

Keywords: forecasting, internet traffic, self organizing map, support vector regression

1. Introduction

Nowadays, the role of information and communication technology (ICT) increases rapidly. In line with that, the needs of internet also grow vividly that also impacted to the higher internet data traffic level. In order to face high internet traffic, internet service providers (ISPs) need to deliver and maintain its quality of service (QoS) optimally. One kind of approaches that ISPs can implement in maintaining QoS is through the internet traffic management [1]. By creating good internet traffic management, ISPs can be able to manage its traffic flows and improve its performance in dealing with varied and adverse traffic conditions [2], so as to achieve the defined service level agreement [3].

Internet data traffic forecasting is one aspect in internet traffic management which can be improved by the ISPs [4]. Historical data of traffic internet stores information about the internet usage patterns, and by doing forecasting, ISPs can predict the future use of the internet for various purposes. Short-term traffic forecasting can be used for load balancing, admission control, bandwidth allocation, congestion control, and detection of anomalies or abnormal traffic patterns. On the other hand, long-term traffic forecasting can be used for planning the bandwidth capacity development or arranging the optimum internet topology. Hence the offered access speeds will be able to meet the demand for the internet in the future [5].

Internet traffic historical data are basically massive, complex, and highly volatile. Using classical statistics or conventional forecasting methods will become ineffective and inaccurate forecasting results [6]. To deal with this issue, the development of several new algorithms are very important to analyze the data traffic with the large volume and extracting useful information accurately through data mining technology. One of the method, recently developed for doing forecasting, is combination of self-organizing map (SOM) and support vector regression (SVR) method.

SVR is a forecasting method included in the machine learning study, as a part of support vector machine (SVM). This method became popular in the time-series forecasting due to its good learning capacity in handling high-dimensional and nonlinear data. In case of internet traffic forecasting, SVR method has been proven to provide accurate results on some researches [7-9]. However, SVR was basically not able to capture the characteristics of the data that is non-stationary (its statistical distribution changes over time) [10]. Thus another kind of method is needed to be combined to deal with that characteristic.

SOM method is a popular method for classifying objects with high dimension attribute to decrease its dimension [11]. Through the SOM method, a set of time-series data can be decomposed into smaller groups of data with similar statistical distribution. SVR will then be applied to the formed groups of data to provide forecasting results.

The combination of methods SOM and SVR methods is purposed of this study to obtain accurate internet traffic forecasting model. The combination of SOM and SVR algorithm has been performed on several areas of research, particularly regarding electricity load forecasting and stock prices, and proven to provide more accurate forecasting results than other methods [12-14]. By performing this method on internet traffic data forecasting, ISPs can predict the condition of internet traffic in the future and then determine the right for both operational and planning of the optimal internet infrastructure development in order to meet the demand for internet access.

2. Experiment

Self-organizing map. Self-organizing map (SOM) which is often called self-organizing feature map (SOFM) or Kohonen map is a clustering method that was first introduced by Kohonen in 1982 [11]. This method is an unsupervised learning algorithm for clustering objects with high dimension attribute to a lower-dimensional space (usually one or two dimensions).

Basically, the network of SOM consists of two layers of nodes or artificial neurons, which are the output layer and input layer. The output layer is usually a two-dimensional lattice which acts as the distribution layer. Number of nodes in the input layer is equal to number of features related to the input. Each output node has the same number as the number of features on the input node. Relationship between the input and output layer of each node can be represented as a vector containing the number of input features.

Stages of SOM algorithm can be described as follows: 1) Initialization stage, which is setting the parameter of neighborhood or *R (t)* and learning rate or *η (t),* as well as determining the weight of each relationship with random values 0-1; 2) Competition stage, which is calculating the value of Euclidean distance between the input denoted as $x = \{x_i : i = 1,...., D)$ and the connection weights of input i to node j denoted as $w_j = \{w_{ji} : j = 1,...,N; i = 1,....,D)$ by the equation:

$$d_j(x) = \sum^{D}_{i=1}(x_i - w_{ji})^2 \quad (1)$$

Then the minimum *d* value of a node will be determined to specify the winning node; 3) Cooperation and adaptation stage, which is identifying all the nodes around the winning node that are included in the scope of the parameter *R(t)* to update the weights of each node by the equation:

$$w_j(t + 1) = w_j(t) + \eta(t)(i_l - w_j\,(t)) \quad (2)$$

4) Update the value of the two parameters by the equation:

$$\eta(t)e\frac{d}{R(t)} \quad (3)$$

5) Return to stage 2 to 4 and stop treatment when stopping criteria is reached

Support vector regression. Support vector regression (SVR) is closely related to the classification method of support vector machine (SVM) in terms of theory and implementation. SVM was originally developed by Vapnik [15] in the implementation of structural risk minimization principle which lowers the empirical based the limit of generalization error.

Given a set of training data $\{(x_1, y_1)\ (xi, y_i)\} \subset X \times R$, where *X* symbolizes the space of input pattern. SVR goal is to find a function f(x) which does not deviate more than ε towards targets y_i for all training data, and at the same time, it is as flat as possible. Linear function *f*(*x*) is formulated as follows:

$$F(x) = w^T + b \text{ with } w \in X, b \in R \quad (4)$$

The flatness on equation (4) means smaller value of ||w||. This problem can be formulated as follows:

$$\min \frac{1}{2}\|w\|2$$

$$s.t.\begin{cases} \text{yi T} - \text{wT}x\text{i} - b \le \varepsilon \\ \text{w xi} + \text{b} - \text{yi} \le \varepsilon \end{cases} \quad (5)$$

However, the objects cannot always be separated linearly. To overcome this, a non-negative slack variables, ξi, ξi*, is introduced to deal with infeasible constraints of the optimization problem in equation (5). The new formulation for this problem is:

$$\min \frac{1}{2}\|\text{w}\|^2 + C\sum_{i=1}^{i}(\xi_i + \xi_i^*)$$
$$s.t.\begin{cases} \text{yi} - \text{wTxi} - \text{b} \le \varepsilon + \xi\text{i} \\ \text{wT}x\text{i} + b - \text{y} \le \varepsilon + \xi * \text{i} \\ \xi\text{i}, \xi * \text{i} \ge 0 \end{cases} \quad (6)$$

The constant C determines the trade-off of error margin between the flatness of *f(x)* and the number of deviations that exceeds ε and still be tolerated. To allow the SVR to predict a nonlinear condition, input data mapping into a feature space is used. It is denoted by:

$$\Phi: \mathbb{R}^n \to F$$
$$x \to \Phi(x) \quad (7)$$

The decision function can be calculated by the inner product of $\Phi(x)^{T} \Phi(x_i)$ without explicitly mapping x into a higher dimension. $\Phi(x)^{T} \Phi(x_i)$ is then called the kernel function $K(x, z) \equiv \Phi(x)^{T} \Phi(z)$.

The combination of SOM and SVR algorithm. The combination of SOM and SVR adopted the 'divide and conquer' approach that was first developed by Jacobs, Jordan, Nowlan and Hinton [16]. This principle is used to deal with complex problems, by dividing a complex problem into smaller and simpler problems to be overcome, so that the real problem could be more easily resolved. This model can be applied to perform better forecasting.

In the first stage, SOM decomposes the whole training data into several clusters, in which data with the same statistical distribution are grouped in a cluster, so that the non-stationary and nonlinearity of this historical data data can be handled. In the second stage, SVR model is developed from each cluster. By this clustering process, SVR can predict future values more accurately because the characteristics and patterns identification of the data can be more easily performed in each cluster. To obtain an accurate model of the SVR, it is also required an appropriate kernel function and loss function [13]. The parameters of SVR, such as parameter γ of kernel function, the parameter ε of loss function, and the soft margin constant *C* (for example, the penalty parameter) must be optimized to achieve a high degree of accuracy. In determining the γ parameter the trial and error technique is adopted, while for parameter ε and C the v-fold cross validation is used In general, the combination of SOM and SVR algorithm is shown in Figure 1.

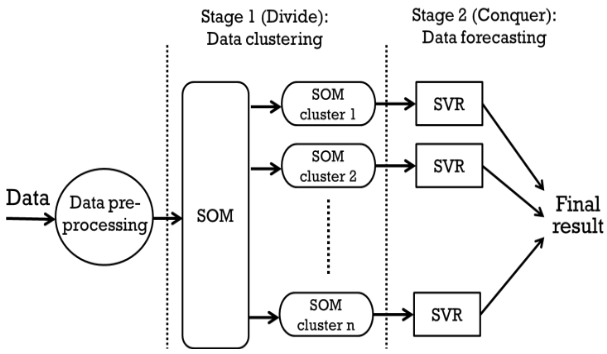

Figure 1. The Combination of SOM and SVR

Data Sets. This study is conducted by using historical data of internet traffic on Jabodetabek region. Jabodetabek it self is divided into six operational areas or clusters in accordance to the policy established, which are Central Jakarta (*Jakarta Pusat* or Jakpus), South Jakarta 1 (*Jakarta Selatan 1* or Jaksel 1), South Jakarta 2 (*Jakarta Selatan 2* or Jaksel 2), East Jakarta 1 (*Jakarta Timur 1* or Jaktim 1), East Jakarta 2 (*Jakarta Timur 1* or Jaktim 2), and North Jakarta (*Jakarta Utara* or Jakut). Thus, the number of data sets used in this study is six in accordance with the number of those areas.

The historical data is taken in the period of March 2014 until February 2015, wich consist of 365 data on each set of data from each operational area. The interval of internet traffic measurement on these historical data is every single day. The internet traffic is measured in gigabytes (Gb).

Data Pre-processing. In this pre-processing, some steps are done to the data, which are: 1) *Determining the Data Attribute*. By using machine learning principles on fore-casting, some relevant data attributes can be added. So, the learning phase of implemented method will provide more accurate forecasting results. These attributes, acted as the independent variable, are considered to give effect to the data that will be forecasted and the forecasted data will be act as the dependent variable. In this study, the attributes that are used are: (a) Time series attribute, which is the numerical attribute in the form of number that describes the sequence of each data in the historical data set used; (b) Day attribute, which is the categorical attribute to identify the day of each data in the historical data set used; (c) Day description attribute, which is the categorical attribute to identify the day description of each data in the historical data set used, whether weekday, weekend, or national holiday; (d) Month attribute, which is the categorical attribute to identify the month of each data in the historical data set used; (e) Historical traffic vector attribute, which are the numerical attributes that describes the traffic conditions or magnitude of k-day before an internet traffic data occurs. In this study, it is set that k = 7. 2) *Changing the*

Variables with Categorical Attribute into Variables with Binary Numbers. At this stage, each level of the categorical attribute is converted into a separate variable consisting of binary numbers [0,1]. This change was made because SVR method can only be performed to variables that contain numbers; 3) *Continuous Data Transformation*. On this phase, the values of the variables with numerical attribute or continuous data are transformed into a range of 0 to 1.

The transformation is done by using the scaling formula as follows:

$$x' = \frac{x - min_a}{max_a - min_a} \tag{8}$$

where x is the original value of the data, x' is the transformed value, min_a is the minimum value of *a* variable, while max_a is the maximum value of *a* variable.

Continuous data transformation through scaling has two advantages, which are to prevent a variable with a larger range of values dominate other variables that have a smaller range of values as well as to simplify the computation process [10].

Dividing each data set into training and test data. The last stage of the data pre-processing is dividing each set of historical data into training and test data. Training data are used in testing the parameters to get the best model, while the test data are used to evaluate the accuracy of the obtained model in predicting a value. In this study, the data is divided 75% into training data (269 data), while the rest into test data (89 data). Using 75% training and 25% testing aims to facilitate enough data for learning and to make sure that the model will not be over fitted.

SOM implementation. The input variables in this phase are all independent variables, which are the time series number variable, Monday variable to Sunday variable, weekday variable, weekend variable, national holidays variable, January variable to December variable, and D-7 traffic data variable to D-1 traffic data variable. The total number of input variable is 30 variables.

On implementing this SOM method, the initial map size needs to be set first. Next, trial and error is performed in determining the size of this initial map. The minimum amount of data on each cluster is 30 data to make SOM method gives a more accurate result [12]. Because each set of data consists of 365 data, the initial map size being to be tested in the process of trial and error is the size of 2x1 (maximum 2 clusters), 2x2 (maximum 4 clusters), and 2x3 (maximum 6 clusters).
Besides the initial map size, there are two parameters whose values should be determined first, which are the neighborhood parameter or *R (t)* and learning rate parameter or *η (t)*. Determination of these parameters is also done by trial and error, in which the parameter *R (t)* to be tested is 1, 2, 3, and 4, while the parameter *η (t)* to be tested is 0.1, 0.2, 0.3, 0.4, and 0.5. Training cycle is set at 1000 times. The value of initial map size, parameter *η (t),* and parameter *R (t)* that are chosen to form the clustering model are the parameters which give the smallest error value.

SVR implementation. In contrast to the implementation of SOM method that only input variables should be determined, the input and output variables of SVR method needs to be determined. Input variables are all independent variable or variables used to predict the value of the dependent variable (30 variables), while the output variable is the dependent variable or variable whose value wants to be predicted, which in this case is the internet traffic data variables (one variable).

Kernel function, used for SVR method, is radial basis function (RBF). This kernel function is chosen because its ability to analyze the high dimensional data and requires only one parameter, the gamma parameter (γ), thus it can simplify the parameter optimization process [13]. In addition to this parameter, the other parameters that also need to be determined are the parameters of SVR itself, which are the epsilon parameter (ε) and soft margin constant parameter (*C*).

In this study, the determination of optimal value of parameter γ, parameter C, and parameter ε are shown in Table 1. With these parameters, the best SVR forecasting model will be selected, which is a model that gives the smallest error value. After obtaining the best forecasting model for a cluster, this forecasting model is then used to predict the value of output variable, or in the other word, the internet traffic data variable from test data included in that cluster. Then the forecasting results of each cluster are combined as one final result of a data set.

Error Value Measurement. The measurement of error value in this study uses the formula of mean average percentage error (MAPE), mean absolute error (MAE)

Table 1. Determination of SVR Parameters

Parameter	Determination Method	Tested Value
γ	Trial & Error	[2, 1, 0.5, 0.1, 0.01, 0.001, 0.0001]
ε	Cross validation	0,1 to 1 with the incremental of 0,1
C	Cross validation	1 to 100 with the incremental of 1

Table 2. The Formula Used for Error Value Measurement

Error Value	Formula

Measurement		
MAPE	$\frac{\sum_{i=1}^{n} \lvert A_i - F_i \rvert / A_i}{n} \times 100$	(9)
MAE	$\frac{\sum_{i=1}^{n} \lvert A_i - F_i \rvert}{n}$	(10)
RMSE	$\sqrt{\frac{\sum_{i=1}^{n} (A_i - F_i)^2}{n}}$	(11)

and root mean square error (RMSE) as shown in Table 2, where A_i is the actual value of the i^{th} data, F_i is the prediction results of the i^{th} data, and n is the number of data samples in the data set. The use of both of the tree metrics is to show how well the error reduced using the proposed method by satisfying all of the error parameters.

3. Result and Discussion

The error values of forecasting results by using SOMSVR method are shown in Table 3. To find out how accurate the forecasting results of this method, forecasting by using single SVR method to the whole data without any clustering are also conducted. The error values of forecasting results with single SVR method are shown in Table 4.

From both tables, it can be seen that the average error value, in which MAPE, MAE, and RMSE of forecasting results by using SOM-SVR method are smaller than the one by using a single SVR method. So, the result shows that the forecasting model with a combination of SOM-SVR method is more accurate than the single SVR method for internet traffic forecasting in this study. In addition, when the error value compared one by one on each set of historical data, the result also show that the value of MAE and RMSE of forecasting results from SOM-SVR method implementation is smaller than the single SVR method on all data sets. For the value of MAPE, the forecasting results from SOM-SVR method implementation also give smaller values than the single SVR method for almost all data sets, except the data sets of Jaktim 1 and Jakut.

Basically, the smaller error value than the other does not always mean the difference between the two error values is statistically significant. Therefore, the mean significance test is also conducted by using paired ttest to differences of MAPE, MAE and RMSE from each forecasting results obtained from the two types of methods applied. The results of the paired t-test are shown in Table 5.

From the paired t-test results, we can conclude that in the case of internet traffic forecasting in this study, the application of SOM and SVR can reduce the value of

Table 3. Error Value of Forecasting Results from the Implementation of SOM-SVR Method

Data Set	MAPE	MAE	RMSE
Jakpus	3.26%	265.49	341.12
Jaksel 1	2.96%	196.17	247.62
Jaksel 2	3.08%	301.81	372.1
Jaktim 1	3.23%	225.41	290.02
Jaktim 2	2.83%	161.78	214.61
Jakut	2.63%	210.96	278.11
Average	3.00%	226.94	290.60
Standard Deviation	0.24%	50.08	58.26

Table 4. Error Value of Forecasting Results from the Implementation of Single SVR Method

Data Set	MAPE	MAE	RMSE
Jakpus	3.72%	294.69	381.17
Jaksel 1	3.18%	207.03	255.87
Jaksel 2	3.12%	310.5	407.9
Jaktim 1	3.16%	226.86	309.18
Jaktim 2	3.21%	188.04	259.77
Jakut	2.62%	215.38	299.36
Average	3.17%	240.42	318.88
Standard Deviation	0.35%	50.05	62.84

Table 5. Results of Paired t-test for the Error Values Difference from the Implementation of SOM-SVR and Single

Error Measurement	P-Value	Significance
MAPE	0.114	Not Significant
MAE	0.035	Significant
RMSE	0.005	Significant

MAE and RMSE of forecasting results significantly compared with the application of single SVR, but the combination of the method does not reduce MAPE significantly. One possible cause of insignificant differences of MAPE values is the much smaller value range of MAPE compared to the value range of MAE and RMSE. So, the decline in the value of MAPE is not as big as MAE or RMSE. Nevertheless, the combination of SOM and SVR algorithm in this study generally is able to give smaller error values than single SVR method and the obtained forecasting model is also more accurate.

4. Conclusions

In short, through data processing and analysis towards internet traffic historical data of Jabodetabek region, it is showed that the SOM-SVR method gives better forecasting results in terms of the error value (MAPE, MAE and RMSE) than the single SVR method. In addition, the difference between the MAE and RMSE value of forecasting results with SOMSVR and single SVR method is statistically significant. For MAPE value, the error value difference of forecasting results with these two methods is not statistically significant.

In the future, any other related attributes to internet traffic can be added to the forecasting model. In addition, the SOM-SVR method also can be combined with optimization algorithms, such as genetic algorithm, in order to determine the optimal value of SOM and SVR parameters.

References

[1] A. Hardiansyah, S.I. Lestarininganti, T.N. Nizar, Universitas Komputer Indonesia, Bandung, 2013. [In Indonesian]

[2] 3G Americas, Traffic Management Technique for Mobile Broadband Network, 3G Americas, Washington DC, 2011.

[3] H. Abrahamsson, Internet Traffic Management, Malardalen University Press, Vasteras, 2008.

[4] G.R. Ash, Traffic Engineering and QoS Optimization of Integrated Voice & Data Networks, Morgan Kaufmann, San Fransisco, 2006.

[5] K. Papagiannaki, N. Taft, Z.L. Zhang, C. Diot, IEEE Trans. Neural Netw. 16 (2005) 1110.

[6] D. Prangchumpol, Int. Schol. Sci. Res. Innovation. 7/7 (2013) 1141.

[7] M. Lippi, M. Bertini, P. Frasconi, IEEE Trans. Intell. Transp. Syst. 14 (2012) 871.

[8] P. Bermolen, D. Rossi, Computer Network. 53 (2009) 191.

[9] X. Liu, X. Fang, Z. Qin, C. Ye, M Xie, J. Netw. Syst. Manag. 19/4 (2011) 427.

[10] C.W. Hsu, C.C. Chang, C.J. Lin. A Practical Guide to Support Vector Classification, Department of Computer Science, 2003. Available at: http://www. csie.ntu.edu.tw/~cjlin/papers/ guide/guide.pdf.

[11] T. Kohonen, Self-organizing Maps, Springer, New York, 2011.

[12] S.H. Hsu, J.P.A. Hsieh, T.C. Chih, K.C. Hsu, Expert Syst. Appl. 36 (2009) 7947.

[13] C.L. Huang, C.Y. Tsai, Expert Syst. Appl. 36 (2009) 1529.

[14] J. Nagi, K.S. Yap, S.K. Tiong, S.K. Ahmed, The 2nd International Power Engineering and Optimization Conference, Shah Alam, Selangor, Malaysia, 2008.

[15] V. Vapnik, The Nature of Statistical Learning Theory, Springer-Verlag, New York, 1995.

[16] R. Jacobs, M. Jordan, S. Nowlan, G. Hinton, Neural Comput. 3 (1991) 79.

[17] E. Alpaydin, Introduction to Machine Learning, MIT Press, Cambridge, 2004.

[18] R.F. de Brito, A.L. Olivera, IEEE World Congress on Computational Intelligence, Brisbane, QLD, Australia., 2012, pp.1-7.

[19] H. Feng, Y. Shu, S. Wang, M. Ma, Proceedings of the IEEE International Conference on Communications, Istanbul, Turkey, 2006, pp. 597-602.

[20] J. He, Rethinking Traffic Management: Design of Optimizable Networks, Princeton University, New Jersey, 2008.

[21] J. He, J. Rexford, M. Chiang, Algorithms for Next Generation Networks, 2010, pp.18.

[22] N.V. Kalyankar, Network Traffic Management, J. Comput. 1/1 (2009) 191.

[23] R. Kohavi, F. Provost, Glossary of Terms: Machine Learning, Kluwer Academic, Boston, 1998.