Hybrid Information Gain Method and Bagging in Data Classification using Support Vector Machine

Immanuel H. G. Manurung, Tulus and Poltak Sihombing
Departement of Computer Science and Information Technology, University of Sumatera Utara, Jl. Dr. T. Mansur No.9, Medan, Indonesia

Keywords: Selection Attributes, SVM, Bagging, IG, *Fold Cross-Validation*

Abstract: The selection process is very influential attributes of dataset in SVM algorithm which tends to produce good accuracy on the results of the classification (classifier) SVM is not optimal. To reduce the effect of selection attributes on SVM classifiers, it is necessary to apply a combination of methods of feature selection algorithms that are Bootstrapping Aggregation (Bagging) and methods of Information Gain (IG). The application of the algorithm Bagging the feature selection is done to give weight to each feature are recommended, so that the found feature is a strong classifier, whereas IG focuses on identifying attributes and evaluate the impact of a beneficial features based on ranking the features that can be recommended to the classifier SVM in the process classification. Experiments implementation of Information Gain feature selection techniques that use attributes with election threshold level. The results showed that, the performance accuracy of SVM classifiers in dataset by combining IG before bagging process, by setting the value thresold >= 0.02 and a 10-fold cross-validation, show that with the implementation of information gain feature selection techniques can improve the performance of machine learning classification algorithm.

1 INTRODUCTION

1.1 Background

Data mining technology is one tool for data mining in large data bases and the specification level of complexity that has been widely used in many application domains such as banking, fraud detection and telecommunications fields .Some research has also been done using data mining techniques to gather information from a database, such as to analyze the performance (performance) of students in the learning process (Kalles & Pierrakeas, 2006) as well as to help teachers to manage classes diampunya (Agathe & Kalina, 2005) and it is possible to analyze and evaluate the academic data for know the quality of higher education (AlRadaideh et al, 2006).

Selection of the right features and the classifier is essential in improving the accuracy and computing in the classification. Feature selection technique is done to reduce the irrelevant features and reduce the dimensions of the features in the data. A feature selection techniques to reduce the dimension attributes. The dimensional reduction is done to get the attributes that are relevant and not excessive so as to speed up the classification process and can improve the accuracy of classification algorithms. (Arifin, 2015). Feature selection method used in this research is the Information gain. The method will perform computing process to obtain the attributes that most influence on the dataset. Dinakaran et al implement feature selection techniques Information Gain method of ranking the best features and classification Decision Tree algorithm J48 (Dinakaran et al, 2013). Ramaswami and Bhaskaran conduct a comparative study of five feature selection techniques and apply four classification algorithms in data mining education. The experimental results indicate that Information Gain feature selection techniques showed the best results (Ramaswami & Bhaskaran, 2009).

In this study, used Information Gain for feature selection and bagging to improve the accuracy of classification in machine learning. "Bagging is a method that can improve the results of the classification of machine learning algorithms (Breimann, 1994)". "This method formulated by Leo Breiman and the name is inferred from the phrase" Bootstrap aggregating "(Breimann, 1994)". Bagging

Manurung, I., Tulus, . and Sihombing, P.
Hybrid Information Gain Method and Bagging in Data Classification using Support Vector Machine.
DOI: 10.5220/0010040401940201
In *Proceedings of the 3rd International Conference of Computer, Environment, Agriculture, Social Science, Health Science, Engineering and Technology (ICEST 2018)*, pages 194-201
ISBN: 978-989-758-496-1

is a method based on the ensemble method, several studies applying meta-algorithm, these include the research for planning and environmental changes (Kang et al, 2016), research for the development of systems to track visual (Chang and Hsu, 2015), enhancing the performance related to research in the field of pharmacy (Galvan, 2015), improved performance in the field of voice processing (Chen, 2015),

Therefore, in this study the proposed Information Gain for feature selection algorithms and meta bagging to improve the classification performance of SVM method. Two stages in building a hybrid scheme, the first step is the selection of the features considered relevant for classification and there construct the second phase of the hybrid intelligent model scheme. Further features that are considered relevant will be used as input in the classification with SVM.

1.2 Problem Formulation

Issues to be addressed in this study is how much the value of accuracy obtained from the application of Information Gain for feature selection and bagging to improve the accuracy of classification in machine learning using Support Vector Machine (SVM) in the form of hybrid models.

1.3 Limitations

Boundary problem discussed in this research is to implement a meta model of machine learning bagging on classification Support Vector Machine (SVM) with Information Gain feature selection.

1.4 Research Objectives

The purpose of this study is to model meta bagging on classification machine learning Support Vector Machine (SVM), to improve the accuracy of each algorithm values and compare them.

1.5 Benefits Research

The benefits of this research are:

1. Can be used as reference material to expand horizons in research specifically in the classification of machine learning.
2. Can be used to provide the best information in the selection of attributes to get the attributes that most influence on the dataset.

2 LITERATURE

2.1 Data Mining

Data mining is a concept used to find hidden knowledge in the database. Data mining is a semi-automatic process that uses statistical techniques, mathematics, artificial intelligence, and machine learning to extract and identify potential knowledge and useful information that is stored in large databases. Data mining is part of the process of KDD (Knowledge Discovery in Databases), which consists of several stages such as the selection of data, pre-processing, transformation, data mining, and evaluating results. In terms of other data mining is also defined as the process to obtain useful information from large data base warehouse.

Data mining is a series of processes for adding additional value of a set of data in the form of knowledge that had been unknown to them manually. Said mining means that efforts to get a bit of valuables from a large number of basic material. Data mining is the process of finding patterns and relationships hidden in large amounts of data for the purpose of classification, estimation, prediction, association rule, clustering, description and visualization. In a simple data mining can be regarded as a process of filtering or "mining" knowledge from large amounts of data. The process and data filtering technique determines the quality of knowledge and information that will be obtained. Another term for data mining is Knowledge Discovery in Databases (KDD).

Data cleaning

Cleaning of the data was performed to remove noise and inconsistent data

Data integration.

Data integration process undertaken to menggabungkandata from various sources.

Data Selection.

Selection data is done to retrieve the relevant data, which will be used for process analysis in data mining.

Data transformation.

This process is done to transform the data into the proper form for in-mine.

Data Mining.

Data mining is the process for applying a metodeuntuk extract patterns in the data.

Evaluation pattern.

Evaluation of the pattern is needed to identify some interesting patterns representing knowledge.

Presentation of knowledge.

Represents the knowledge that had been dug up to the user to visualize such knowledge.

2.2 Support Vector Machine

Support Vector Machine(SVM) was developed by Boser, Guyon, Vapnik, and was first presented at the 1992 Annual Workshop on Computational Learning Theory. The basic concept of Support Vector Machine (SVM) is actually a harmonious combination of theories of computing has been around for decades before, such as margin hyperplane (Duda & Hart 1973, Cover 1965, Vapnik 1964, and others), the kernel was introduced by Aronszajn 1950, and likewise with the concepts supporting the other. However, until 1992, there has never been an attempt assembling these components.

Concept of Support Vector Machine (SVM) can be explained simply as an attempt to find the best hyperplane which serves as a separator of two classes in the input space. pattern that is a member of two classes: +1 and -1 and share alternative dividing lines (boundaries discrimination). Margin is the distance between the hyperplane to the nearest pattern of each kelas.Pattern closest is called a support vector. Attempts to locate this hyperplane is the core of the learning process on SVM (Christianini, 2000).

Initialization of training data is used, given the label in the form of $\{(y_i, x_i)\}_{i=1}^{N}$, with $y_i \in \{-1,+1\}, x_i \in R^n$ using formulations C-SVM. For the linear case, the algorithm is shown to find the best hyperplane to separate the data by minimizing the following function:

$$\tau(w,\xi) = \frac{1}{2}\|w\|^2 + C\sum_{i=1}^{N}\xi_i \qquad (2.1)$$

for $y_i(w \bullet x_i + b) \geq 1 - \xi_i$ and $\xi_i \geq 0$, Where C> 0 is a trade off of the constraints. Kernel function K data point (x, z): Rn x Rn→ R is the result of inner product of $\varphi(x) \bullet \varphi(z)$, Data mapping function $\varphi(x)$ and $\varphi(z)$ mentioned very difficult to find values of high yield dimensional inner product value is equal to K (x, z). Because of these difficulties, the system of direct control of the kernel function is used. Then to get the optimal hyperplane equation, can be used the following equation:

$$W(\alpha) = \sum_{i=1}^{N}\alpha_i - \frac{1}{2}\sum_{ij}\alpha_i\alpha_j y_i y_j K(x,x_i) \quad (2.2)$$

$$\text{for : } 0 \leq \alpha_i \leq C \text{ and } \sum_{i=1}^{N}\alpha_i y_i = 0 \qquad (2.3)$$

the function of the decision is $sign(f(x))$, Where :

$$f(x) = \sum_{i=1}^{m}\alpha_i y_i K(x,x_i) + b \qquad (2.4)$$

Note that Equation 3 only require access to the kernel function alone, without having to perform *mapping* data by function $\varphi(.)$, M is the number of support vectors. So this allows one to solve formulations in high-dimensional feature space with a highly efficient, step is called the kernel trick. For linear kernel, we can use the kernel functions $K(x,z) = x \bullet z$ And function hyperplane $f(x) = w \bullet x + b$, Where the vector w can be calculated by the formula $w = \sum_{i=1}^{m}\alpha_i y_i x_i$ and $b = -\frac{1}{2}\left(w.x^{+} + w.x^{-}\right)$, If in case of non-linear vector w can be calculated by the formula $w = \sum_{i=1}^{m}\alpha_i y_i \varphi(x_i)$ and the constant b by the formula $b = -\frac{1}{2}\left(w.\varphi(x^{+}) + w \bullet \varphi(x^{-})\right)$, Additive selected specifically for the kernel to use the formula $K(x,z) = \sum_{i=1}^{n} K_i(x,z)$ and $f(x)$ can be written as $f(x) = w \bullet \varphi(x) + b$,

2.3 Meta-algorithm

Meta-algorithm is an algorithm that uses another algorithm as a representative, and also an algorithm that has a sub-algorithms as variables and parameters can be changed. Examples include meta-algorithm is boosting, simulated annealing, bootstrap aggregating, AdaBoost, and random-restart hill climbing.

2.3.1 Algorithm *boosting*

Boosting is an meta-algorithm in machine learning to perform supervised learning. Boosting theory introduced by questions Kearns (1988): Can a set of weak learner creates a strong unity learner? Weak learner is a classifier that has little correlation with actual classification, while strong learner is a classifier which has a strong correlation with the actual classification.

Most of boosting algorithms follow a plan. Generally boosting occurs in iterations, incrementally adding weak learner into a strong learner. At each iteration, the weak learner to learn from a workout data. Then, the weak learner was added to the strong learner. After a weak learner is added, the data is then converted their respective weights. The data were misclassified will experience weight gain, and the data were classified correctly will experience a

reduction in weight. Therefore, the weak learner on the next iteration will be more focused on data misclassified by the weak learner that before.

Of the many variations of boosting algorithm, AdaBoost is the most famous in the history of its development, is the first algorithm that can adapt to the weak learner. However, there are some examples of other boosting algorithms such as LPBoost, TotalBoost, BrownBoost, MadaBoost, LogitBoost, and others. Most boosting algorithm can be incorporated into the framework AnyBoost.

2.3.2 Metode Bagging (Bootstrap Aggregating)

This combined method is a method models each of which is a combination of a lineup that includes a number of k-models of the M_1, M_2..., M_k, With the aim to create an increase in the combined models, M *. In general there are two methods of combination of models, namely Bagging and Boosting. Both can be used in the process of classification and prediction (Han and Kamber, 2006). Illustrations combined methods can be seen in Figure 2.1.

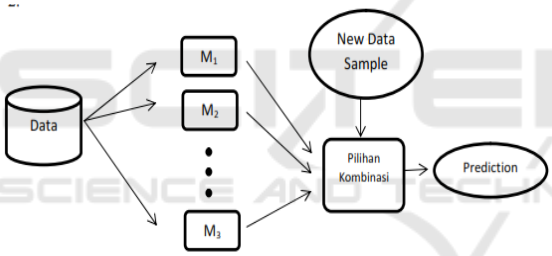

Figure 2.1: Illustration Model Combination Methods

Pada Figure 2, there are descriptions of the work steps of the method combination of models, where in the method aims to improve the accuracy of the model consisting of Bagging and Boosting method. This method produces a classification or prediction models, M_1, M_2..., M_k, Where voting strategy against emerging choice is a combination of strategies used to combine the predictions for the object that has been given the unknown category.

Bagging method is the theory proposed by Breiman (1996), which is based on the concept of a bootstrap theory and aggregating that combine both benefits into a single theory. Bootstrap is applied based on the theory of random sampling with replacement (Tibshirani and Efron, 1993). In effect, the model in engineering classification (classifier) that has been formed is likely to have a better performance. Understanding of the process of aggregating is incorporating some of the classifier.

Sometimes combined classifier gives better results compared to the classifier only, due to the incorporation of both the benefits of some of the classifier at the end Bagging penyelesaiannya.Oleh therefore helpful to help build or establish a better classifier on the training data samples. Here are the stages of implementation Bagging technique:

1. Perform as many Xb bootstrap replication of a number n m training data. Repeat these steps for 1,2, ..., B. Where m is the number of data that is taken from the training data, n is the sample size of the training data and B is the number of bootstrap replication is done.
2. By using a simple majority vote, been labeled the most widely emerged from the assessment results as a rule for making a final decision.

2.4 Hybrid Algorithm

Hybrid or Hybrid Algorithm algorithm is an algorithm that combines two or more methods into one or several models combined to produce items according to the user's wishes (Burke, R., 2007). Broadly speaking, this algorithm has two stages. The first stage is the algorithm generate node ordering and second stages of the algorithm constructs a structure.

3 RESEARCH METHODOLOGY

All data is primary data collected from UCI Machine Learning Repository, Research steps are generally depicted in the flow diagram Figure 3.1

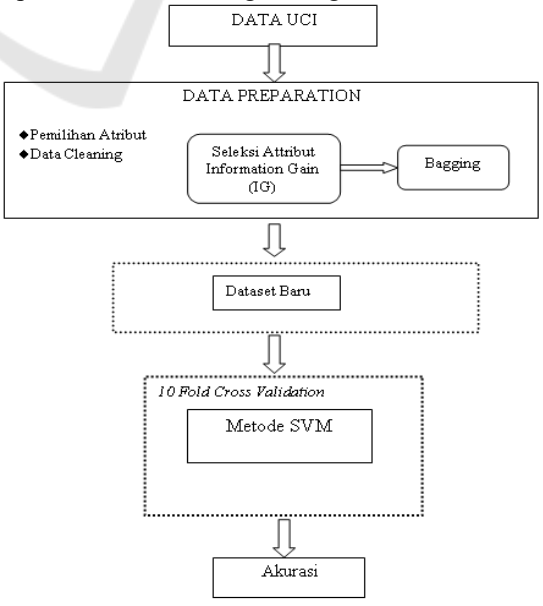

Figure 3.1 Block diagram of the stages of research

There are several stages in this study, the dataset collection, pre-processing, a single model, and hybrid models. After data collection, pre-processing is done data by eliminating the data contained missing value. In the single phase model of the result data preprocessing The classified using SVM. The final stage is to establish a scheme hybrid intelligent model to produce some hybrid combinations. At the stage of the model development hybrid, there is a feature selection to select features considered relevant as an input for classification.

3.1 Variable Data and Research

This study will use data on Diabetes Diagnosis of UCI Machine Learning Repository. Overall, the data Diabetes Diagnosis contains 9 attributes described in the following table.

Table 3.1.

No.	Attribute
1.	pregnancies
2.	PG Concentration
3.	diastolic BP
4.	Tri Fold Thick
5.	serum Ins
6.	BMI
7.	DP Function
9	Age
10.	diagnosis

4 RESULTS AND DISCUSSION

In this section explored further on the implementation of the proposed system architecture. In testing, the evaluation techniques used cross validation models to observe and analyze the performance measurement results. Tests were conducted at four methods of classification. In each method to do a comparison between the performance of the classification by application of meta-adaptive boosting algorithm and without the application of this meta-algorithm. To measure the performance of this classifier is used seven sizes, namely accuracy, Kappa, MAE, Precision, Recall, F-Measure, and the ROC.

4.1 Read Data Training

Training data used contains as many as 768 data line containing the seven attributes that have been described in the early chapters. To specify the data that will be analyzed then the first step is to read training data.The following are some of the training data will be used.

4.2 Testing Single Model SVM Method

From a probability value above 768 will be tested as much data as the data and completed using weka tools so that the resulting classification result as in Figure 4.2.

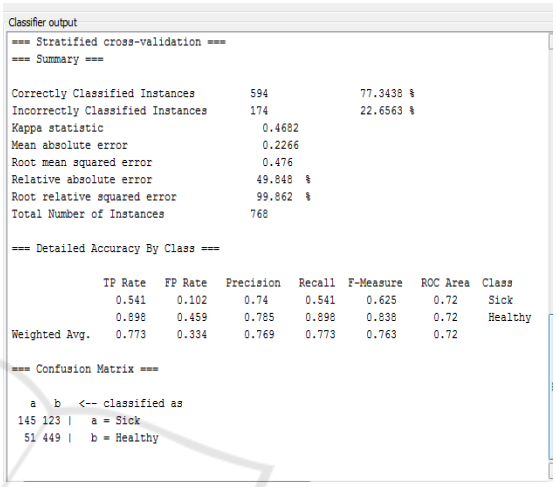

Figure 4.4: Testing Results with SVM method

a. Accuracy obtained is 77.3438% of total Correctly *classified instances* as many as 594.
b. Number of instances was incorrectly classified as much as 174 or 22.6563%.
c. The results of the root of the mean squared error is 0.476.

4.3 Testing Results with Information Gain

From the calculation of the value of the determined value overlapping provisions of the discriminant (discriminant ability). Average - average value of the provisions of the discriminant (discriminant ability) of each feature on each - each class is divided by the number of changes made reference in calculating the suspension of each feature. Where the features with the smallest score is the highest rating in ranking and can be recommended in the classification process. The following ranking of feature selection using the IG shown in Table 4.2.

Table 4.2: Rating Each Attribute

Ranked	Attribute	Score
1	PG Concentration	0.1901
2	BMI	0.0749
3	Age	0.0725
4	serum Ins	0.0595
5	Tri Fold Thick	0.0443
6	pregnancies	0.0392
7	DP Function	0.0208
8	diastolic BP	0014

In the third experiment, started by applying feature selection using the IG. The results of ranking the features of the IG, the percentage of moderately by 70% the results of the second experiment had the highest accuracy value will be recommended to the bagging algorithm. From the selection by bagging features that the Hx value equal to +1 is decent features recommended and the value is equal to -1 Hx is a feature that is not recommended in classification. Here are the results of measurement with IG before bagging shown in Table 4.3.

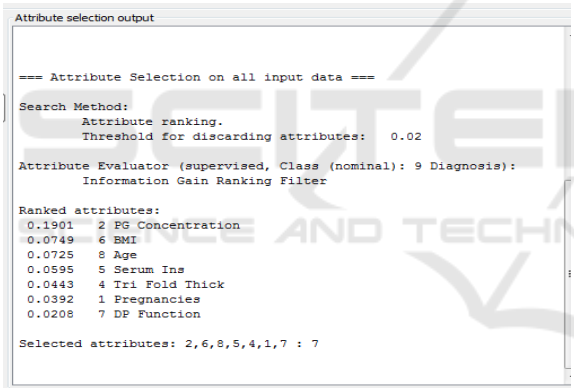

Figure 4.5: Selection Attributes with the Information Gain

Information gain feature selection techniques applied by using perangkingan attributes by using the selected threshold level features by using threshold >= 0,02.

4.4 Testing with Bagging - SVM

In the fourth experiment begins by applying feature selection using bagging to perform weighting on seven features recommended so that a strong classifier and further evaluation of the impact of a beneficial features based on ranking features using IG. The measurement results of ranking the fourth feature selection in the experiment are shown in Table 4.5.

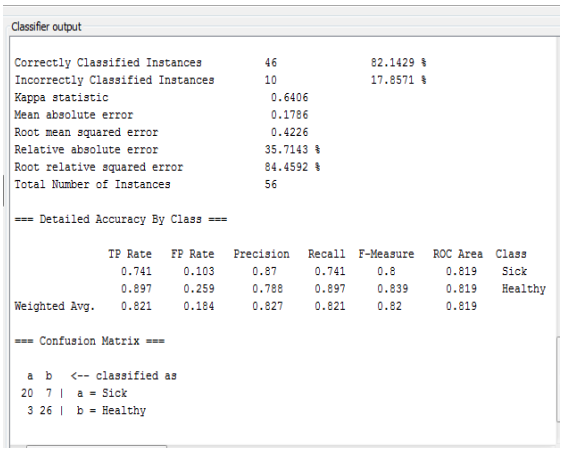

Figure 4.6: Testing Results with bagging-SVM models

In the figure indicates that the value of the highest accuracy by applying feature selection IG after bagging in a 10-fold cross-validation is equal to 82.14%. From the results of measurements performed some experiments can be concluded that the increase in performance accuracy, precision and sensitivity (recall) SVM models, are affected by the application of feature selection in the SVM classifier and the determination of the number-fold cross validation. That is, if the value-fold cross validation the greater the value of accuracy, precision and sensitivity (recall) tends to increase and will get better reliability SVM models.

4.5 Results Discussion

In the experimental study, there are two models: Single classification results using SVM for hybrid models while there are 2 classification accuracy results, respectively for hybrid models with bagging classifier. For hybrid models bagging scheme can improve accuracy. Accuracy using only one classifier was 77.34% and increased accuracy using hybrid scheme with bagging-SVM amounted to 81.14%. Because the hybrid scheme has fewer features in addition to mengkatkan accuracy, computation process will be faster. This shows that by using more features does not improve accuracy, but by choosing the features that are considered relevant so as to reduce the size of features with the right method and the right classifier will improve accuracy and speed up the computation.

5 CONCLUSION

5.1 Conclusion

Based on the results of experimental studies hybrid model with bagging method for classifying a number of conclusions as follows:

1. Bagging highly adaptive method with the features, because each iteration in bagging an election classifier that has the smallest error. bagging choose the best feature in every iteration. So with many or few features that are used or with any data, bagging would classify properly.
2. The research concluded that the hybrid scheme with classifier bagging on classification has been proven to improve accuracy and speed up the computation. By using a single classifier accuracy of 77.34% increased by 81.14% using a hybrid scheme.

5.2 Suggestions

For further research, bagging metaalgoritme will be developed that are not too sensitive to outliers (data outliers), so that optimal performance of metaalgoritme over again.

REFERENCES

Agathe, and Kalina, 2005, *the Educational Data Mining: A Case Study,* Pole UniversitaireLéonard de Vinci, France

Al-Radaideh, QA, Al-Shawakfa, EM, and AlNajjar, MI, 2006, *the Student DataUsing Mining Decision Trees, The 2006* International Arab Conference on Information Technology (ACIT'2006).

Birant, D. 2011.*Comparison of Decision Tree Algorithms for Predicting Potential Air Pollutant Emissions Models with Data Mining.* Journal of Environmental Informatics.Vol. 17 Issue 1, p46-53. 8p.

Christianini, N. & Taylor, SJ 2000.*An Introduction to Support Vector Machine and other Kernel-Based Learning methods,* Cambridge University Press.

Chang and Hsu, *"Development of a Visual Compressive Trackng System Enhanced by Adaptive Boosting,"* in 41st Annual Conference of the IEEE Industrial Electronics Society (IECON 2015), pp. 3678-3682, 2015.

Chen, Y. Li, and X. Xu, *"Rotating Target Classification base on Micro-doppler Features Using a Modified Adaptive Boosting Algorithm,"* International Conference on Computers, Communications, adn Systems, pp. 236-240, 2015.

Dekker, W. Gerben., Et.al. (2009). *Predicting Students Drop Out: ACase Study.* Proceedings of the 2 nd International Conference on EducationalData mining. 41-50.

Deb, AK, Member, Student, Member, senior, Gopal, M., & Chandra, S. (2007). *SVM-Based Tree-Type Neural Networks as a Critic in Adaptive Critic Designs for Control.* IEEE, 18 (4), 1016-1030.

DPK Muhammad Yunus, *Prediction Model Design Graduate Student With Decision Tree algorithm*, Matrix, vol. 2, no. 13, pp. 1-5, 2015.

Erdogan, SZ, Timor, M. (2005) *.A Data Mining Applications in Student Database.* Journal of Aeronautics andSpace Technologies. Vol 2 (2) .53-57.

Galvan, *"Integrase Inhibition Using Differential Evolution-Binary Particle Swarm Optimization and Non-Linear AdaptiveBoosting Random Forest Regression,"* 16th International Conference on Information Reuse and Integration, pp. 485-490, 2015.

Han, J., et al. *Data Mining: Concepts and Techniques 2nd Edition,* Morgan Kaufmann Publishers, 2006.

Han J, Kamber M. 2001. *Data Mining: Concepts and Techniques. Simon Fraser University,* Morgan Kaufmann Publishers.

Han J and Kamber M. *Data Mining: Concept and Techniques.* New York: Morgan Kaufmann Publishers; 2006.

Hassan, *"Biomedical Signal Processing and Control Computeraided obstructive sleep apnea detection using inverse Gaussian normal parameters and adaptive boosting,"* Biomed. Signal Process. Control, vol. 29, pp. 22-30, 2016.

Huang, Z., & Shyu, M.-ling. (2010) .k-*NN LS-SVM Based Framework for Long-Term Time Series Prediction*. System, 4-6, 69-74.

Jingbo Yuan and Ding Shunlin, *"Research And Improvement On Association Rule Algorithm Base On FP-Growth,"* 2012.

Kalles, D., & Pierrakeas, C. 2006, *Analyzing Student Performance in Distance Learning with Genetic Algorithms and decsion Trees,* Hellenic Open University.

Kusumadewi Sri, *Classification of Nutritional Status Using Naïve Bayesian Classification*, COMMIT, vol.3 no. 1, pp. 6-11, 2009.

Kraipeerapun, P., & Fung, CC (2008) *Performance of Interval Neutrosophic*. Comparing Sets and Neural Networks with Binary Support Vector Machines for Classification Problems. Learning, 34-37.

Larose, Daniel T. 2005*. Discovering Knowledge in Data: An Introduction to Data Mining.* John Willey & Sons, Inc.

Lassibille, G., Gomez, LN (2007). *Why Do Higher EducationStudents Drop Out? Evidencefrom Spain.Education Economics*.Vol 16 (1) .89-105.

Luan, J. (2002*). Data Mining and Its Applications in Higher Education.* New Directions for Institutional Research.Vol 133.17-36.

Maimon, O. and Last, M. 2000. *Knowledge Discovery and Data Mining, The Info-FuzzyNetwork (IFN) Methodology.* Dordrecht: Kluwer Academic.

Mark A. Hall and Geoffrey *Holmes.Benchmarking Attribute Discrete Class Selection Techniques for Data Mining.* IEEE Transactions on Knowledge and Data Engineering, 15, no. 3, June 2003

Mujib Ridwan, Suryono Hadi, M.Sarosa, *Application of Data Mining for Evaluation of Student Academic Performance Algorithm Naïve Bayes,* EECCIS, vol 7 no. 1, pp. 59-64, 2013.

Nugroho, Satrio, A. 2003.*Support Vector Machine Theory and Applications in Bioinformatics.*

Pintowati, W., & Otok, W. (2012) .Pemodelan Poverty in East Java with Multivariate Adaptive Approach, 1 (1), 2-7.

Sartono, B. et al. 2003. *Analysis of Multiple Variables*. Bogor Agricultural University, Bogor.

Sri Wahyunigsih, D. (2012). *Modeling Farmer On Crops With Transfer Functions and Multivariete Pendekaran Regresion Adaptive Multi Variate*. ITS, 1-11.

Turban, R., Rainer, R. and Potter, R 2005.*'Introduction to Information Technology* '. USA: John Wiley & Sons, Inc.

Tan, P, et al. 2006. *Introduction to Data Mining. Boston*: Pearsion Education

Thomas E. 2004. *Data Mining: Definition and Decision Tree Examples,* e-book

Wang, T. Lu, and Z. Xiong, "*Adaptive Boosting for Image Denoising: Beyond Low-Rank Representation and Sparse Coding,*" 23rd International Conference on Pattern Recognition (ICPR), 2016.

Xia, L., & Pan, H. (2010). *Inferential Estimation of Polypropylene Melt Index Stacked Using Neural Networks Based on Absolute Error Criteria. Control*, 216-218.

Dinakaran, S. Dr. PRJ Thangaiah, *"Role of Attribute Selection in Classification Algorithms"*, International Journal of Scientific & Engineering Research, vol. 4, issue 6, pp. 67-71. June, 2013.

Ramaswami.M, Bhaskaran.R, "*A Study on Feature Selection in Educational Data Mining Techniques*", Journal of Computing, vol. 1, Issue 1, pp. 7-11. December 2009.