www.ijecs.in
International Journal Of Engineering And Computer Science ISSN:2319-7242
Volume 4 Issue 5 May 2015, Page No. 11945-11951

Comparative study and evaluation of various data classification techniques in data mining

Vivek Verma[1], Ram Nivas Giri[2]

[1]Research scholar
Department of computer science
Raipur institute of technology
Raipur, India
E-mail: vivekverma.exe@gmail.com

[2]Asst. professor
Department of computer science
Raipur institute of technology
Raipur, India

Abstract: ***Data Mining is knowledge discovery process in database designed to extract data from a dataset and transforms it in to desired data. data processing action is similarly acclimated in get of constant patterns and/or analytical relationships amid variables, and a new to validate the accusation by applying the detected patterns to new subsets of knowledge. Data categoryification is one in every of the info mining technique to map great amount of data set in to applicable class. Data categoryification is reasonably supervised learning that is employed to predict class for information input, wherever categories are predefined.Supervised learning is that part of automatic learning which focuses on modeling input/output relationship the goal of supervised learning is to identify an optimal mapping from input variables to some output variables, which is based on a sample of observations of the values of the variables. Data classification technique includes various applications like handwriting recognition, speech recognition, iris matching, text classification, computer vision, drug design etc. objective of this paper is to survey major techniques of data classification. Several major classification techniques are Artificial neural network, decision trees, k-nearest neighbor (KNN), support vector machine, navie-bayesian classifier, the aim of study to make comparative analysis of major data classification techniques.***

Keywords—data mining, data classification, decision tree, support vector machine, KNN

1. Introduction

Data classification [2]is kind of searching/deriving such kind of function which map data set D ={t_1,t_2,.....,t_n} to set of classes C= {c_1,c_2,c_3.....c_n} D→C. data classification is kind of supervised learning since we have to train our classification model in supervised manner in training phase we give informationabout predefine classes. In the wake of developing classification model we must perform testing of dataThere are three general steps for any classifier algorithm to classify data.

Step 1: Using a learning algorithm to extract rules from (create a model of) the training data. The training data are pre classified examples (class label is known for each example).

Step 2: Evaluate the rules on test data. Usually split known data into training data and test data

Step 3: Apply the principles to (classify) new data (illustrations with obscure class names).

Goals:

- create a model of data, clarify or better comprehend data.
- predict the class label of (classify) new examples.

So the data classification problem may be stated as follows:

Given a set of training data points along with associated training labels, determine the class label for an unlabeled test instance.

Various varieties of this issue can be characterized over diverse settings. Fabulous outlines on data classification may be found in. Classification algorithms ordinarily contain two stages:

• **Training Phase**: during training a model is constructed from the training instances.

• **Testing Phase**: In this phase, the model is used to assign a label to an unlabeled test instance.

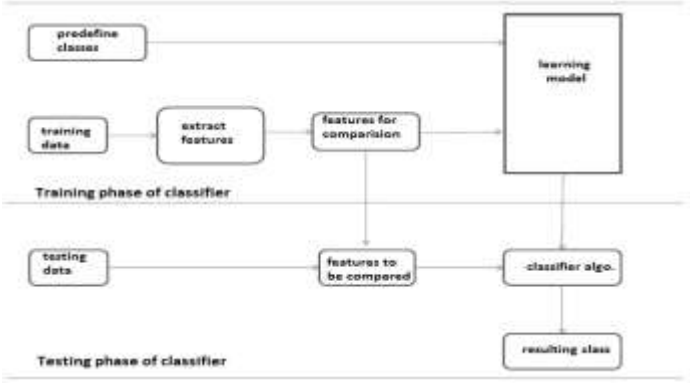

Figure:1 data classification task

The problem of data classification has numerous applications in a wide variety of mining applications. This is because the problem attempts to learn the relationship between a set of feature variables and a target variable of interest. Since many

practical problems can be expressed as associations between feature and target variables, this provides a broad range of applicability of this model.

There are enormous amount of literature available for classification and vast categories of algorithm have been proposed. Here we only briefly review some major algorithm used for classification; more thorough accounts of various topics are covered in e.g.[13],[20],[26],[49]. Although classification has been used widely in many fields ranging from geographical information system to military and to medicine, image and other high-dimensional data classification, for example, face recognition, arguably attracts the most research attention among various pattern recognition or data mining tasks.in the year of 1993 Neri Merhav and chin-hui [21] presented "A Minimax Classification Approach with Application to Robust Speech Recognition" in which A minimax approach for robust classification of parametric information sources is studied and applied to isolated word speech recognition based on hidden Markov modeling. The goal is to reduce the sensitivity of speech recognition systems to a possible mismatch between the training and testing conditions. In the year of 1999 yuhong yang [23] present "Minimax Nonparametric Classification—Part I:Rates of Convergence" in this yuhhong and yang studies minimax aspects of nonparametric classification. We first study minimax estimation of the conditional probability of a class label, given the feature variable. in the year of 2002 Mohamed Afify, Olivier Siohan, and Chin-Hui Lee[22], derived upper and lower bounds on the mean of speech corrupted by additive noise. The bounds are derived in the log spectral domain. Also approximate bounds on the first and second order time derivatives, it is an application to minimax classification. In the year of 2006 Yakoub Bazi,and Farid Melgani[9] present an " Optimal SVM Classification System for Hyperspectral Remote Sensing Images".in the year of 2007 Xiaoou Li, Jair Cervantes, and Wen Yu[20]presents "Two-Stage SVM Classification for Large Data Sets via Randomly Reducing and Recovering Training Data". In the year of 2008 A. Mathurand ,G. M. Foody[32] presents "Multiclass and Binary SVM Classification: Implications for Training and Classification Users".In the year of 2009 Begüm Demir,and Sarp Ertürk[42]presents "improving svm classification accuracy using a hierarchical approach for hyperspectral images" in the year of 2010 Mark A. Davenport, Richard G. Baraniuk, and Clayton D. Scott[47] worked for "Tuning Support Vector Machines for Minimax and Neyman-Pearson Classification" in 2012 Mahesh Pal and Giles M. Foody [35] did "Evaluation of SVM, RVM and SMLR for AccurateImage Classification With Limited Ground Data" in 2014 Jair Cervantes, Xiaoou Li2, Wen[39] represents "SVM Classification for Large Data Sets by Considering Models ofClasses Distribution". In mar/2014 Sminu N.R , Jemimah Simon[28] represents "Feature Based Data Stream Classification (FBDC) and Novel Class Detection".

Rest of the paper organized in such a way that section 2 describe the actual problem of data classification section 3 represents various techniques of data classification. section 4 compares various data classification techniques and discuss the result.at last section 4 is conclusion part of the paper.

2. Problem definition

Data classification [1] is that the method of sorting and categorizing knowledge into varied varieties, forms or the other distinct category. Knowledge classification allows the separation and classification of knowledge per data set necessities for varied business or personal objectives. it's chiefly an information management method. The classification problem thus segments the unseen test instances into groups, as defined by the class label. While the segmentation of examples into groups is also done by clustering, there is a key difference between the two problems. In the case of clustering, the segmentation is done using similarities between the feature variables, with no prior understanding of the structure of the groups. In the case of classification, the segmentation is done on the basis of a training data set, which encodes knowledge about the structure of the groups in the form of a target variable. Thus, while the segmentations of the data are usually related to notions of similarity, as in clustering, significant deviations from the similarity-based segmentation may be achieved in practical settings. As a result, the classification problem is referred to as supervised learning, just as clustering is referred to as unsupervised learning. The supervision process often provides significant application specific utility, because the class labels may represent important properties of interest. We can define classification problem as follow

Given a database D={t1,t2,t3…,tn} of tuples and set of classes C={C1,C2,C3….Cm}.the classification problem is to define a mapping f:D→C. where each ti is assigned to one class.

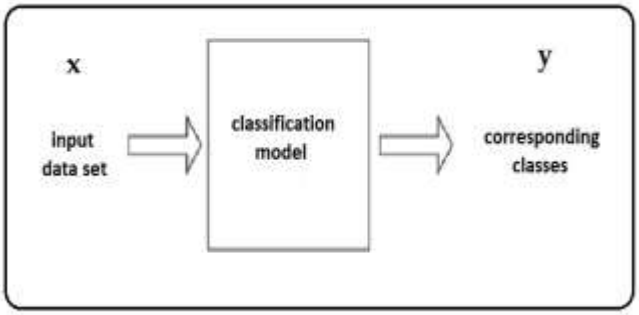

Figure :2 classification model

2. Methodology

A classification technique (or classifier) could be a systematic approach to assembling classification models from an input data set. Examples embody decision tree classifiers, rule-based classifiers, neural networks, support vector machines, and naive Bayes classifiers. Each technique employs a learning algorithm to identify a model that best fits the relationship between the attribute set and sophistication label of the input data. The model generated by a learning algorithmic program ought to each match the input data well and properly predict the category labels of records it has never seen before. Therefore, a key objective of the learning algorithm is to build models with sensible generalization capability; i.e., models that accurately predict the class labels of previously unknown records. Figure 1.4 shows a general approach for resolution classification issues. First, a training set consisting of records whose class labels are known must be provided. The training set is used to build a classification model, which is subsequently applied to the **test set**, which consists of records with unknown class labels. There square measure following common techniques for data classification.

- Decision Trees
- Rule-Based Methods
- Probabilistic Methods
- SVM Classifiers

- Artificial Neural Networks

A. *Decision Tree*

Decision trees[49][48][46]create a hierarchical partitioning of the data, which relates the different partitions at the leaf level to the different classes. The hierarchical partitioning at each level is created with the use of a *split criterion*. The split criterion may either use a condition (or predicate) on a single attribute, or it may contain a condition on multiple attributes. The former is referred to as a univariate split, whereas the latter is referred to as a multivariate split. The overall approach is to try to recursively split the training data so as to maximize the discrimination among the different classes over different nodes. The discrimination among the different classes is maximized, when the level of skew among the different classes in a given node is maximized. Some of the earliest methods for decision tree construction include

- C4.5 (successor of ID3) [42],
- ID3 (Iterative Dichotomiser 3) [48],
- CART (Classification And Regression Tree) [22].
- CHAID(CHi-squared Automatic Interaction detector)
- MARS: extends decision trees to handle numerical data better
- Conditional instance tree(Statistics-based approach that uses non-parametric tests as splitting criteria, corrected for multiple testing to avoid overfitting. This approach results in unbiased predictor selection and does not require pruning.)

A detailed discussion of decision trees may be found in [46, 47, 48, 49] Decision tree builds classification or regression models within the type of a tree structure. It breaks down a dataset into smaller and smaller subsets while at the same time an associated decision tree is incrementally developed. The final result is a tree with decision nodes and leaf nodes. A decision node has two or more branches. Leaf node represents a classification or decision. The uppermost decision node in a tree which corresponds to the best predictor called root node Decision trees can handle both categorical and numerical data. Let the training data is meant to be a region of a transportation study concerning mode option to choose Bus, Car or Train.The data have 4 attributes . Attribute gender is binary kind, car ownership is quantitative integer (thus behave like nominal). Travel cost/km is quantitative of ratio type however in here we tend to place into ordinal kind (because quantitative knowledge has to be compelled to be split into qualitative data) and income level is also an ordinal kind.

Table 1: data set for decision tree[11]

ATTRIBUTES				CLASSES
GENDER	OWENERSHIP	TRAVEL COST	INCOME LEVEL	TRANSPORTATION MODE
Male	0	Cheap	Low	Bus
Male	1	Cheap	Medium	Bus
Female	1	Cheap	Medium	Train
Female	0	Cheap	Low	Bus
Male	1	Cheap	Medium	Bus
Male	0	Standard	Medium	Bus
Female	1	Standard	Medium	Train
Female	1	Expensive	High	Train
Male	2	Expensive	Medium	Car
Male	2	Expensive	High	Car

Based on above data set we can include decision tree as follow.

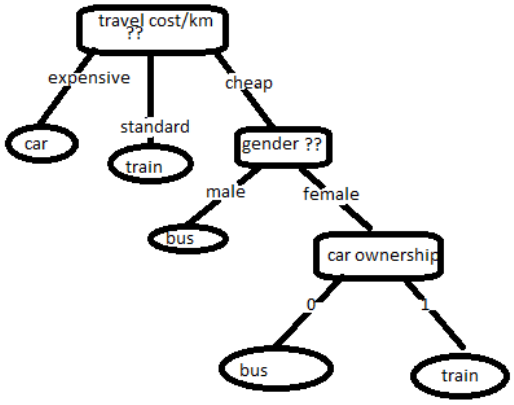

Figure 3: decision tree[11]

B. *Rule-Based Method*

Rule-based methods are closely related to decision trees, except that they do not create a strict hierarchical partitioning of the training data. Rather, overlaps are allowed so as to form greater robustness for the training model. Any path in an exceedingly decision tree could also be understood as a rule, that assigns a take a look at instance to a specific label. as an example lets contemplate following decision tree within which a case wherever the two measures (features) of the blood parameters of patients are used in order to assess the level of cardiovascular risk in the patient. The two measures are the C-Reactive Protein (CRP) level and Cholesterol level, which are well known parameters related to cardiovascular risk. for the case of the decision tree illustrated in Figure 1.7, the rightmost path corresponds to the following rule: CRP> 2 & Cholesterol > 200⇒ High Risk It is possible to create a set of disjoint rules from the different paths in the decision tree.

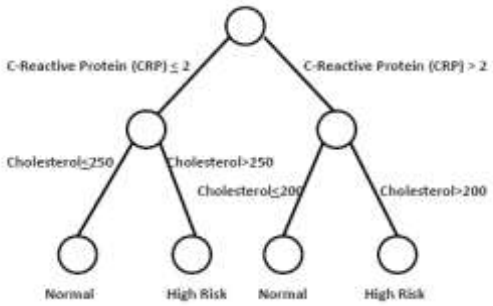

Figure :4 An example of Decision tree based classification

Rule-based classifiers can be viewed as more general models than decision tree models. While decision trees require the induced rule sets to be *non-overlapping*, this is not the case for rule-based classifiers.

C. *Probablistic Methods*

Probabilistic strategies [14]are the foremost basic among all data classification strategies. Probabilistic categoryification algorithms use statistical inference to seek out the simplest class for a given example.In addition to simply assigning the best class like other classification algorithms, probabilistic classification algorithms will output a corresponding posterior

probability of the test instance being a member of each of the possible classes. The posterior probability is defined as the probability after observing the specific characteristics of the test instance. On the other hand, the prior probability is simply the fraction of training records belonging to each particular class, with no knowledge of the test instance After getting the posterior probabilities, we have a tendency to use decision theory to work out category membership for every new instance.

the posterior probability of a particular class is estimated by determining the class-conditional probability and the prior class separately and then applying Bayes' theorem[14][11][8] to find the parameters. The most well-known among these is the Bayes classifier, which is known as a generative model. For ease in discussion, we will assume discrete feature values, though the approach can easily be applied to numerical attributes with the use of discretization methods. Consider a test instance with *d* different features, which have values $X =< x1 \dots\dots\dots xd >$respectively. It is desirable to determine the posterior probability that the class $Y(T)$ of the test instance T is i. In other words, we wish to determine the posterior probability $P(Y(T) = i|x1 \dots xd)$. Then, the Bayes rule can be used in order to derive the following:

$$P(Y(T)=i|x_1\dots\dots\dots\dots x_d)=P(Y(T=i).\frac{P(x1\dots\dots xd|Y(T)=i)}{P(x1\dots\dots\dots xd)}$$

Since the denominator is constant across all classes, and one only needs to determine the class with the maximum posterior probability, one can approximate the aforementioned expression as follows:

$P(Y(T) = i|x1 \dots xd) \propto P(Y(T) = i) \cdot P(x1 \dots xd|Y(T) = i)$

The key here is that the expression on the right can be evaluated more easily in a data-driven way, as long as the *naive Bayes assumption* is used for simplification. Specifically, in Equation1.2,the expression $P(Y(T) = i|x1 \dots xd)$ can be expressed as the product of the feature-wise conditional probabilities.

$$P(x1 \dots\dots xd|Y(T) = i) = \prod_{j=1}^{d} P(xj)|Y(T) = i)$$

This is referred to as *conditional independence*, and therefore the Bayes method is referred to as "naive." This simplification is crucial, because these individual probabilities can be estimated from the training data in a more robust way.

D. Support Vector Machine

SVM[28] methods use linear conditions in order to separate out the classes from one another. The idea is to use a linear condition that separates the two classes from each other as well as possible.Consider the medical example discussed earlier, where the risk of cardiovascular disease is related to diagnostic features from patients.

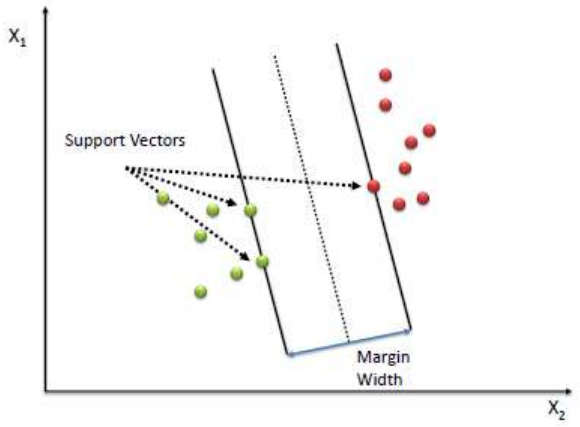

Figure:5 SVM classifier

In such a case, the split condition in the multivariate case might also be used as stand-alone condition for classification. This, a SVM classifier, is also thought-about single level decision tree with a really fastidiously chosen multivariate split condition. Clearly, since the effectiveness of the approach depends only on a single separating hyperplane, it is critical to define this separation carefully.

Support vector machines are typically outlined for binary classification issues. Therefore, the class variable yi for the ith training instance Xi is assumed to be drawn from {−1,+1}. The most important criterion, which is commonly used for SVM classification, is that of the maximum margin hyperplane. In order to understand this point, consider the case of linearly separable data illustrated in Figure . Two possible separating hyperplanes, with their corresponding support vectors and margins have been illustrated in the figure. it's evident that one in every of the separating hyperplanes contains a abundant larger margin than the opposite, and is so additional fascinating owing to its bigger generality for unseen check examples. Therefore, one in every of the necessary criteria for support vector machines is to attain most margin separation of the hyperplanes.

E. Neural Network

Neural networks[4] plan to simulate biological systems, like the human brain. within the human brain. In the human brain,neurons square measure connected are connected to one another via points, that square measure named as synapses. In biological systems, learning is performed by changing the strength of the synaptic connections, in response to impulses. This biological analogy is preserved in a Artificial neural network. The fundamental computation unit in an artificial neural network is a neuron or unit.These units may be organized in numerous forms of architectures by connections between them. the foremost basic design of the neural network could be a perceptron, that contains a group of input nodes and an output node. The output unit receives a group of inputs from the input units. There square measure d completely different input units, that is strictly adequate the spatial property of the underlying data. The data is is assumed to be numerical. Categorical information might have to be reworked to binary representations, and thus the amount of inputs could also be larger. The output node is associated with a set of weights W, which are used in order to compute a function f (·) of its inputs. Each component of the weight vector is associated with a

connection from the input unit to the output unit. The weights can be viewed as the analogue of the synaptic strengths in biological systems. In the case of a perceptron architecture, the input nodes do not perform any computations. They simply transmit the input attribute forward. Computations are performed only at the output nodes in the basic perceptron architecture. The output node uses its weight vector along with the input attribute values in order to compute a function of the inputs.

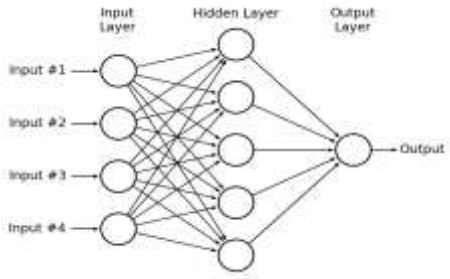

Figure 6: Single and multilayer neural networks.

The training process proceeds in two phases, one of which is in the forward direction, and the other is in the backward direction.

1. Forward Phase: In the forward phase, the activation function is repeatedly applied to propagate the inputs from the neural network in the forward direction. Since the final output is supposed to match the class label, the final output at the output layer provides an error value, depending on the training label value. This error is then used to update the weights of the output layer, and propagate the weight updates backwards in the next phase.

2. Backpropagation Phase: In the backward phase, the errors are propagated backwards through the neural network layers. This leads to the updating of the weights in the neurons of the different layers. The gradients at the previous layers are learned as a function of the errors and weights in the layer ahead of it. The learning rate λ plays an important role in regulating the rate of learning. In practice, any arbitrary function can be approximated well by a neural network. The price of this generality is that neural networks are often quite slow in practice. They are also sensitive to noise, and can sometimes overfit the training data. The previous discussion assumed only binary labels. It is possible to create a k-label neural network, by either using a multiclass "one-versus-all" meta-algorithm, or by creating a neural network architecture in which the number of output nodes is equal to the number of class labels. Each output represents prediction to a particular label value. A number of implementations of neural network methods have been studied in , and many of these implementations are designed in the context of text data. It should be pointed out that both neural networks and SVM classifiers use a linear model that is quite similar. The main difference between the two is in how the optimal linear hyperplane is determined.

3. Result & discussion

The efficiency of classification algorithm is that how correctly it classify the input data.[Qiang Cheng],[Hongbo Zhou],[Jie Cheng], and [Huiqing Li][17] presents comparison between various data classification technique, comparison is made by taking various data sets and applying classification technique on data. following table shows comparison amid techniques

Table 2: efficiency of various technique

Data set	GMD	SVM(P)	SVM(R)	KNN(E)	KNN(C)	RF
Iris	94.67%	97.76%	94.76%	96.00%	97.34%	98.00%
Wine	97.28%	97.76	98.24%	83.73%	97.28%	92.22%
Pen digits	98.85%	97.61%	99.52%	99.31%	-	-
Optical pen	97.40%	97.45%	96.83%	98.72%	93.83%	95.22%
UCI mulFeat	80.54%	97.32%	88.75%	96.33%	93.00%	96.11%
Gcm	99.45%	99.45%	99.44%	87.30%	82.88%	75.56%

[p.nancy] [R.geethamani][8] presents a ―comparative study of various data mining classification algorithms ― on the dataset ―social side of the internet‖ for two set. For sub set 1, the features selected by Feature ranking and for sub set 2 relief filtering. The features selected by feature reduction techniques are chosen as input attributes with necessary class variables as target attribute and various classifications algorithms were executed for all selected features one by one and there error rates are tabled below. In this research their conclusion was Rnd Tree performed well for their taken dataset. The following table 1 describes the comparison of various classifications algorithms with error rate

Table 3: error rate of various classification technique

CLASSIFICATION TECHNIQUES	ERROR RATES	
	FACEBOOK	TWITTER
C4.5	0.0860	0.1042
C-RT	0.1798	0.1976
ID3	0.1650	0.2097
KNN	0.1871	0.2097
SVM(P)	0.0865	0.0709
SVM(R)	0.0714	0.0714

4. Conclusion

The goal of classification algorithms is to generate more certain, precise and accurate system results. Numerous methods have been suggested for the creation of ensemble of classifiers. Classification methods are typically strong in modeling interactions. Several of the classification methods produce a set of interacting logic that best predict the phenotype. However, a straightforward application of classification methods to large numbers of markers has a potential risk picking up randomly associated markers. But still it is difficult to recommend any one technique as superior to others as the choice of a dataset. Finally, there is no single classification algorithms is best for all kind of dataset. Classification algorithms are specific in their problem domain.

REFERENCES

[1] Sundar.C, M.Chitradevi and Dr.G.Geetharamani ―"Classification of Cardiotocogram Data using Neural Network based Machine Learning Technique"‖ International Journal of Computer Applications (0975 – 888) Volume 47– No.14, June 2012

[2] Smitha .T, V. Sundaram ―Comparative Study Of Data Mining Algorithms For High Dimensional Data Analysis‖ International Journal Of Advances In Engineering & Technology, Sept 2012.IJAET ISSN: 2231-1963

[3] Sminu N.R , Jemimah Simon " Feature Based Data Stream Classification (FBDC) and Novel Class Detection" *International Journal of Engineering Research and Applications (IJERA) ISSN: 2248-9622 International Conference on Humming Bird (01st March 2014)*

[4]. Al-Ghoneim and B. V. K. V. Kumar, "Learning ranks with neural networks", *Proc. SPIE Applications Science Artificial Neural Networks*, vol. 2492, pp.446 -464 1995

[5] Andrew Secker1, Matthew N. Davies2, Alex A. Freitas1, Jon Timmis3, Miguel Mendao3, Darren R. Flower2 ―An Experimental Comparison of

Classification Algorithms for the Hierarchical Prediction of Protein Function‖ [http://www.cs.kent.ac.uk/projects/biasprofs]

[6] Koliastasis C and D.K. Despotis ―Rules for Comparing Predictive Data Mining Algorithms by Error Rate‖ OPSEARCH, VOL. 41, No. 3, 2004 Operational Research Society of India

[7] Matthew N. Anyanwu, Sajjan G. Shiva ―Comparative Analysis of Serial Decision Tree Classification Algorithms‖ International Journal of Computer Science and Security, (IJCSS) Volume (3) : Issue (3) page number 230

[8] Mrs.P.Nancy, R. Geetha Ramani and Shomona Gracia Jacob ―A Comparison on Performance of Data Mining Algorithms in Classification of Social Network Data‖ International Journal of Computer Applications (0975 – 8887) Volume 32– No.8, October 2011

[9] Yakoub Bazi, *Student Member, IEEE*, and Farid Melgani, *Senior Member,IEEE"* Toward an Optimal SVM Classification System for Hyperspectral Remote Sensing Images*"* IEEE TRANSACTIONS ON GEOSCIENCE AND REMOTE SENSING, VOL. 44, NO. 11, NOVEMBER 2006

[10] Aman Kumar Sharma Suruchi Sahni ― Comparative Study of Classification Algorithms for Spam Email Data Analysis‖ International Journal on Computer Science and Engineering (IJCSE) ISSN : 0975-3397 Vol. 3 No. 5 May 2011 1890-1895.

[11]http://people.revoledu.com/kardi/tutorial/DecisionTree/what-is-decision-tree.htm

[12] Muskan Kukreja, Stephen Albert Johnston and Phillip Stafford ―Comparative study of classification algorithms for immune signaturing data‖ Kukreja et al. BMC Bioinformatics 2012, 13:139 http://www.biomedcentral.com/1471-2105/13/139

[13] Mahendra Tiwari, Manu Bhai Jha, OmPrakash Yadav ―Performance analysis of Data Mining algorithms in Weka‖ IOSR Journal of Computer Engineering (IOSRJCE) ISSN: 2278-0661, ISBN: 2278-8727 Volume 6, Issue 3 (Sep-Oct. 2012), PP 32-41 www.iosrjournals.org

[14] N. Sivaram,K. Ramar ―Applicability of Clustering and Classification Algorithms for Recruitment Data Mining‖ International Journal of Computer Applications (0975 – 8887) Volume 4 – No.5, July 2010

[15] Anurag Upadhayay , Suneet Shukla and Sudsanshu Kumar ―Empirical Comparison by data mining Classification algorithms (C 4.5 & C 5.0) for thyroid cancer data set‖ Anurag Upadhayay et al, International Journal of Computer Science & Communication Networks,Vol 3(1), 64-68 ISSN:2249-5789

[16] Eric bauer, Ron kohavi ―An Empirical Comparison of Voting Classification Algorithms: Bagging, Boosting, and Variants‖ Mining and Visualization, Silicon Graphics Inc. 2011 N. Shoreline Blvd, Mountain View, CA. 94043

[17]Qiang Cheng, Hongbo Zhou, Member, IEEE, Jie Cheng, and Huiqing Li-"A Minimax Framework for Classification with Applications to Images and High Dimensional Data" ieee transactions on pattern analysis and machine intelligence, vol. 36, no. 11, november 2014

[18] Venkatadri.M, Lokanatha C. Reddy ―A Comparative Study On Decision Tree classification Algorithms In Data Mining ―International Journal Of Computer Applications Engineering, Technology And Sciences (Ij-Ca-Ets) Issn: 0974-3596 April '10 –Sept '2010 , Volume 2 : Issue 2 , Page:24

[19] Neslihan Dogan, Zuhal Tanrikulu ― A Comparative Framework for Evaluating Classification Algorithms‖ Proceedings of the World Congress on Engineering 2010 Vol I WCE 2010, June 30 - July 2, 2010, London, U.K. ISBN: 978-988-17012-9-9

[20] Xiaoou Li, Jair Cervantes, and Wen Yu—" Two-Stage SVM Classification for Large Data Sets via Randomly Reducing and Recovering Training Data" 1-4244-0991-8/07/$25.00/©2007 IEEE

[21] Neri Merhav, *Member, IEEE,* and Chin-Hui Lee, *Member, IEEE*―"A Minimax Classification Approach with Application to Robust Speech Recognition",1993.

[22] Mohamed Afify Olivier Siohan, *Member, IEEE*, and Chin-Hui Lee,*Fellow,"* Upper and Lower Bounds on the Mean of NoisySpeech: Application to Minimax Classification*"* IEEE TRANSACTIONS ON SPEECH AND AUDIO PROCESSING, VOL. 10, NO. 2, FEBRUARY 2002

[23] Yuhong Yang—" Minimax Nonparametric Classification—Part I:Rates of Convergence" IEEE TRANSACTIONS ON INFORMATION THEORY, VOL. 45, NO. 7, NOVEMBER 1999

[24] Karthikeyani, Parvin Begum, K Tajudin and Shahina Begam. ―Comparative of Data Mining Classification Algorithm (CDMCA) in Diabetes Disease Prediction. International Journal of Computer Applications‖ 60(12):26-31, December 2012. Published by Foundation of Computer Science.

[25] G. Th. Papadopoulos K. Chandramouli, V. Mezaris, I. Kompatsiaris, E. Izquierdo3 and M.G. Strintzis ―A Comparative Study of Classification Techniques for Knowledge-Assisted Image Analysis ―978-0-7695-3130-4/08 ,2008 IEEE.

[26] Qiang Cheng, Hongbo Zhou, Member, IEEE, Jie Cheng, and Huiqing Li, "A Minimax Framework for Classification with Applications to Images and High Dimensional Data," IEEE transactions on pattern analysis and machine intelligence, vol. 36, no. 11, november 2014

[27]Zonghu Wang, Zhijing Liu, 2010. Graph Based K-NN Text Classification**.** In Proc. Of International Conference on Electrical and control engineering. pp 1092-1095.

[28] air Cervantes1, Xiaoou Li2, Wen Yu3.—" SVM Classification for Large Data Sets by Considering Models of Classes Distribution" IEEE TRANSACTIONS ON NEURAL NETWORKS AND LEARNING SYSTEMS 2014

[29] Mark A. Davenport, Student Member, IEEE, Richard G. Baraniuk, Fellow, IEEE, and Clayton D. Scott, Member, IEEE,"Tuning Support Vector Machines for Minimax and Neyman-Pearson Classification". Cambridge, MA, USA: Cambridge Univ. Press, 2010.

[30] L. Breiman, "Random forests," Mach. Learn., vol. 45, no. 1, pp. 5–32, 2001.

[31] M. Dash and H. Liu. Feature selection for classification. Intelligent data analysis,1(1-4):131–156, 1997.

[32] A. Mathur and G. M. Foody, *Member, IEEE—"* Multiclass and Binary SVM Classification: Implications for Training and Classification Users*"* IEEE GEOSCIENCE AND REMOTE SENSING LETTERS, VOL. 5, NO. 2, APRIL 2008

[33] P. Domingos and M. J. Pazzani. On the optimality of the simple Bayesian classifier under zero-one loss. *Machine Learning*, 29(2–3):103–130, 1997.

[34] Q. Cheng, H. Zhou, and J. Cheng, "The Fisher-Markov selector: Fast selecting maximally separable feature subset for multi-class classification with applications to high-dimensional data," IEEE Trans. Pattern Anal. Mach. Intell., vol. 33, no. 6, pp. 1217–1233, Jun. 2011.

[35] Mahesh Pal and Giles M. Foody, *Senior Member, IEEE—"* Evaluation of SVM, RVM and SMLR for AccurateImage Classification With Limited Ground Data*"* IEEE JOURNAL OF SELECTED TOPICS IN APPLIED EARTH OBSERVATIONS AND REMOTE SENSING, VOL. 5, NO. 5, OCTOBER 2012.

[36] J. Dem_sar, "Statistical comparisons of classifiers over multiple data sets," J. Mach. Learn. Res., vol. 7, pp. 1–30, 2006. [12] D. Donoho, "Compressed sensing," IEEE Trans. Inf. Theory, vol. 52, no. 4, pp. 1289–1306, Apr. 2006.

[37] R. Duda, P. Hart, and D. Stork, Pattern Classification, 2nd ed. Hoboken, NJ, USA: Wiley, 2001.

[38] L. El Ghaoui and H. Lebret, "Robust solutions to least-squares problems with uncertain data," SIAM J. Matrix Anal. Appl., vol. 18, pp. 1035–1064, 1997.

[39] L. Fei-Fei, R. Fergus, and P. Perona, "One-shot learning of object categories," IEEE Trans. Pattern Anal. Mach. Intell., vol. 28, no. 4, pp. 594–611, 2006.

[40] R. Fisher, "The use of multiple measurements in taxonomic problems," Ann. Eugenics, vol. 7, pp. 179–188, 1936.

[41] Cem Tekin, Mihaela van der Schaar,"Distributed Online Big Data Classification Using Context Information", Fifty-first Annual Allerton Conference Allerton House, UIUC, Illinois, USA October 2 - 3, 2013.

[42] Begüm Demir,and Sarp Ertürk—" improving svm classification accuracy using a hierarchical approach for hyperspectral images" 978-1-4244-5654-3/09/$26.00 ©2009 IEEE

[43] Y. Fujikoshi, V. Ulyanov, and R. Shimizu, Multivariate Statistics:High-Dimensional and Large-Sample Approximations. Hoboken, NJ, USA: Wiley, 2010.

[44] K. Fukunaga, Introduction to Statistical Pattern Recognition. San Diego, CA, USA: Academic, 1990.

[45] A. Georghiades, P. Belhumeur, and D. Kriegman, "From few to many: Illumination cone models for face recognition under variable lighting and pose," IEEE Trans. Pattern Anal. Mach. Intell., vol. 23, no. 6, pp. 643–660, Jun. 2001.

[46] L. Breiman. *Classification and regression trees*. CRC Press, 1993.

[47] Mark A. Davenport, Student Member, IEEE,Richard G. Baraniuk, Fellow, IEEE, and Clayton D. Scott, Member, IEEE—"Tuning Support Vector Machines for Minimax and Neyman-Pearson Classification" IEEE TRANSACTIONS ON PATTERN ANALYSIS AND MACHINE INTELLIGENCE, VOL. 32, NO. 10, OCTOBER 2010

[48] J. R. Quinlan. Induction of decision trees, *Machine Learning*, 1(1):81–106, 1986.

[49] S. K. Murthy. Automatic construction of decision trees from data: A multi disciplinary survey, *Data Mining and Knowledge Discovery*, 2(4):345–389, 1998.