Compressive Study on Various Classification Techniques Used in Data Mining

Charan Singh Tejavath¹, Dr. R. P. Singh²

¹Research Scholar of Sri Satya Sai University of Technology and Medical Sciences. Sehore, Madhya Pradesh , India

²Research Supervisor for Sri Satya Sai University of Technology and Medical Sciences Sehore, Madhya Pradesh, India

ABSTRACT

Data Mining is a developing field which has pulled in an expansive number of data enterprises because of the colossal volume of data oversaw as of late. Productive data mining requires a decent comprehension of the data mining techniques to enhance business opportunity and to enhance the nature of administration gave. In light of such needs, this paper gives a survey of conventional classification techniques utilized for data mining. Classification is utilized to discover in which assemble every datum occasion is identified with a given dataset. It is utilized for ordering data into various classes as indicated by a few requirements. A few noteworthy sorts of classification calculations including C4.5, ID3, k-closest neighbor classifier, Naive Bayes, SVM, and ANN are utilized for classification. By and large, a classification system takes three methodologies Statistical, Machine Learning and Neural Network for classification. While considering these methodologies this paper gives a comprehensive review of various classification calculations and their highlights and confinements

Keywords: Classification Technique, Decision Tree Induction, K Nearest Neighbor Classifier, Bayesian Network, and Rule Based Classification. C4.5, ID3, ANN, Naive Bayes, SVM.

I. INTRODUCTION

Classification techniques in data mining are fit for handling a lot of data. It can be utilized to anticipate absolute class names and arranges data in light of preparing set and class names and it can be utilized for characterizing recently accessible data. The term could cover any setting in which some choice or conjecture is made based on by and by accessible data. Classification technique is perceived strategy for more than once settling on such choices in new circumstances. Here in the event that we expect that issue is a worry with the development of a technique that will be connected to a proceeding with the arrangement of cases in which each new case must be allocated to one of an arrangement of pre-characterized classes based on watched highlights of data. Production of a classification system from an arrangement of data for which the correct classes are referred to ahead of time is named as example acknowledgment or regulated learning. Settings in which a classification errand is essential to incorporate, for instance, allocating people to credit status based on budgetary and other individual data, and the underlying conclusion of a patient's ailment with a specific end goal to choose prompt treatment while anticipating flawless test outcomes. Probably the most basic issues emerging in science, industry, and business can be called as classification or choice issues. Three primary authentic strands of research can be distinguished: measurable, machine learning and neural system. All gatherings have a few targets in like manner. They have all endeavored to create methods that would have the capacity to deal with a

wide assortment of issues and to be to a great degree general utilized as a part of useful settings with demonstrated achievement.

Statistical procedure based approach

Two fundamental periods of work on classification can be distinguished inside the measurable group. The primary "established" stage focused on expansion of Fisher's initial work on direct separation. The second, "present day" stage focused on more adaptable classes of models a significant number of which endeavor to give a gauge of the joint appropriation of the highlights inside each class which can thusly give a classification manage [11]. Factual techniques are by and large portrayed by having an exact basic likelihood demonstrate which gives a likelihood of being in each class rather than only a classification. Additionally it is generally accepted that the techniques will be utilized by analysts and consequently some human inclusion is expected with respect to variable determination and change and general organizing of the issue.

Machine learning based approach

Machine Learning is, for the most part, secured programmed figuring methods in light of coherent or twofold activities that take in an assignment from a progression of illustrations. Here we are simply focusing on classification thus consideration has concentrated on choice tree approaches in which classification comes about because of a grouping of sensible advances. This classification comes about are fit for speaking to the most complex issue given adequate data. Different techniques, for example, hereditary calculations and inductive rationale methodology (ILP) are at present under dynamic change and its rule would enable us to manage more broad kinds of data including situations where the number and sort of characteristics may shift. Machine Learning approach expects to create arranging articulations sufficiently basic to be seen effectively by the human and must copy human thinking adequately to give understanding into the

choice procedure [11].Like factual methodologies, foundation information might be utilized as a part of advancement yet activity is accepted without human obstruction

II. LITERATURE OF REVIEW

P.D. Turney [8] proposed an unsupervised learning algorithm which classifies the sentiment as good or bad. Phrases extracted using POS taggers were subjected to PMI-IR algorithm which determined their semantic association. The entire review was categorized accordingly on the basis of the mean semantic orientation of all phrases in the document. An average of 77% accuracy was attained with this method. Time required for processing was one of the main stringent factors in this method. To increase the accuracy the author suggested the incorporation of supervised learning algorithms.

Chien-Liang Liu et al. contrived a movie rating and review summarization system [9] in the mobile environment. The authors used latent semantic analysis (LSA) [9] to pinpoint various aspects or traits in the movie reviews while ignoring unsuitable sentences and then condensed the movie reviews highlighting the important features present. Based on the polarity of the words SVM was implemented to categorize the reviews. The summarized description of the review which contains only relevant aspects was then presented. Since this was implemented at sentence level the review had sentences obtained from different paragraphs causing a lack of fluency for the end users who read it. Even though this work could be extended too many languages, it suffered from limitations such as inability to deal with amalgamate data and inaptness to place the words in broader sense.

Jingbo Zhu et al. [10] fabricated an aspect-based opinion polling system where unstructured reviews were used as a dataset. For effective aspect identification, various terms related to a particular aspect were studied using multi-aspect bootstrapping

method. An aspect-based segmentation model was developed which sliced the compound sentences in the user reviews into multiple single-aspect units which were classified in an opinion poll. Experiments conducted on real Chinese restaurant gave them around 75.5% accuracy.

Hu and Liu [11] employed NL Processor linguistic parser which used POS tagging to split the sentence into its constituents. They have employed algorithms of their own for frequent feature identification, opinion word extraction and orientation determination and presented with the straightforward summary of the reviews. Future work in this domain would be monitoring customer reviews for determining strength of the opinions and expanding this work to concoct views expressed via verbs, nouns and adverbs.

Ali Harb et al. [12] developed a web mining approach which classified the reviews based on the extracted adjectives in context with the specific domain rather than using a broad and vague dictionary. Tree Tagger tool was used to split the sentence into its constituent parts and they employed Apriori Algorithm to evaluate the list of adjectives collected and to rank the adjectives accordingly. The document was categorized as positive if it contained more number of positive adjectives. Invert polarities were also considered while classification. The accuracy of their method was reduced due to the presence of low quality and noisy text which affected the extraction process.

Xuanjing Huang et al. [13] retrieved most frequently used features using the concept of frequent item set and used the adjectives near them as opinion words of that feature. Using Word Net dictionary these features were classified as positive or Comparison of Classification Techniques for Feature Oriented … 151 negative sentiments. It was found that this approach worked well data in a particular domain but was difficult to train models while dealing with reviews intermixed from multiple domains like in Twitter.

In their endeavor [14] Xing Fang and Justin Zhan aimed to resolve the problem of accurately classifying the polarity of the sentences by incorporating techniques such as POS tagging and classification models namely Naïve Bayesian, Random Forest, and Support Vector Machine. They have tested their methodology on Amazon Data Set at sentence and review level. The major limitation of the process proposed by them was that it doesn't classify reviews that purely contain implicit sentiments.

Nathalie Camelin et al. designed a system [15] which analyzed the opinions in a spoken message. A strategy was proposed for detecting those segments of the speech which contains opinions and inferring their polarity using Ada Boost algorithm accurately.
U. Krcadinac et al. modeled a sentence based emotion recognition system called as Synesketch [16] which individual sentences in the text as input and labeled it with respect to the six emotional types defined by Ekman [17]. In future they aim to improve their algorithm by adding POS tagging and improved negation detection to their work.

In [18] Dilara Torunoğlu et al. improved the efficiency of Naïve Bayes algorithm by making use of Laplacian Smoothing approach. They found that their approach even exceeded the efficiency of SVM classifier. The authors would like to expand their work to categorize the sentiments in Turkish Twitter datasets.

III. CLASSIFICATION TECHNIQUES

Classification is a data mining strategy used to anticipate bunch participation for data occasions. There are numerous customary classification strategies like choice tree enlistment, k-closest neighbor classifier, Bayesian systems, bolster vector machines, control based classification, case-based thinking, fluffy rationale techniques, hereditary

calculation, harsh set approach et cetera. The essential contrast between the calculations relies upon whether they are apathetic students or anxious students. The choice tree classifiers, Bayesian classifier, bolster vector classifier are anxious students as they utilize preparing tulles to build the data show while closest neighbor classifiers are lethargic students as they hold up until a test tulle lands for classification to perform speculation. A concise talk about the choice tree calculation, Bayesian systems, Rule based classification, K Nearest neighbor and other classification techniques, their issues and late attempts to defeat these issues is done in this area.

Association Rule Mining

Association rule mining, a standout amongst the broadest and very much explored techniques of data mining.[3] It means to extricate unordinary relationship, visit examples, associations or causal structures among sets of things in the exchange databases or other data stores. [9] Association rules are generally utilized as a part of different territories, for example, media transmission systems, market and hazard administration, stock control and so forth. Different association mining techniques and calculations will be quickly presented and analyzed later. [10] Association rule mining is to discover association rules that fulfill the predefined least help and certainty from a given database. The issue is generally disintegrated into two sub problems. One is to discover those thing sets whose events surpass a predefined limit in the database; those thing sets are called incessant or huge item sets. The second issue is to produce association rules from those huge thing sets with the imperatives of insignificant certainty.

Bayesian Classification

A naive Bayes classifier expects that the nearness or nonappearance of a specific element is inconsequential to the nearness or nonattendance of some other element, given the class variable. [1]A naive Bayes classifier considers every one of these highlights to contribute freely to the likelihood that

this natural product is an apple, paying little mind to the nearness or nonappearance of alternate highlights. [8] Naive Bayes classifiers can be prepared productively in an administered learning setting. Favorable position of naive Bayes is that it just requires a little measure of preparing data to assess the parameters (means and changes of the factors) vital for classification. [14] Because free factors are accepted, just the differences of the factors for each class should be resolved and not the whole covariance network.

Decision Tree Classification

Decision tree classification approach is most useful in classification problems. [3] It is a flow chart like tree structure.
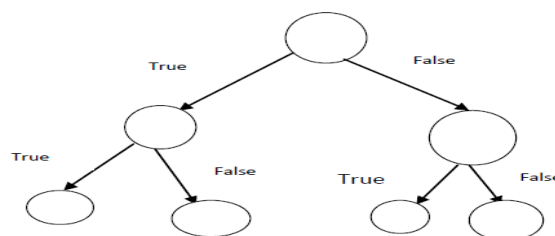

Figure1: Decision Tree

Figure 1

Trees are constructed in a top down recursive divide and conquer manner. [15] In this classification method used in different type algorithm to classify the data sets, the algorithms are: [1]

- ✓ ID3 (Iterative Dichotomies)
- ✓ C4.5 (a Successor of ID3)
- ✓ Classification and Regression Trees (CART)

The algorithm follows a top-down approach, which starts with a training set of tulles and their associated class labels.

Advantages

Rules can be generated that are easy to interpret and understand. It is scalable for large database because the tree size is independent of the database size. [3]Each tuple in the database must be filtered through the tree, and time is proportional to the height of the tree.

Disadvantages

It is does not handle continuous data. [3]Handling missing data is difficult because correct branches in tree could not be taken the labels.

Nearest Neighbor

Nearest neighbor classifiers are based on learning by analogy. The training samples are described by n dimensional numeric attributes []. Each sample represents a point in an n-dimensional space. In this way, all of the training samples are stored in an n-dimensional patter space. When given an unknown sample, a k-nearest neighbor classifier searches the pattern space for the k training samples that are closest to the unknown sample. "Closeness" is defined in terms of Euclidean distance. [9] The unknown sample is assigned the most common class among its k nearest neighbors. When k=1, the unknown sample is assigned the class of the training sample that is closest to it in pattern space. Nearest neighbor classifiers are instance-based or lazy learners in that they store all of the training samples and do not build a classifier until a new (unlabeled) sample needs to be classified. [15]

Neural Networks

Neural network speaks to a mind picture or image for Information processing. [1][3] These models are organically roused as opposed to a correct imitation of how the cerebrum really functions.[8] Neural networks have been appeared to be extremely gifted frameworks in numerous estimating applications and business classification applications because of their capacity to "learn" from the data, their nonparametric nature (i.e., no inflexible presumptions), and their capacity to sum up. Neural processing alludes to an example acknowledgment approach for machine learning. [11] The subsequent model of neural figuring is regularly called a fake neural network (ANN) or a neural network. Neural networks have been utilized as a part of numerous business applications for design acknowledgment, determining, expectation, and classification.

The human mind has diverted capacities for data preparing and critical thinking that cutting-edge PC can't contend with in numerous perspectives. [1]It has been the predicate that a model or a framework that is receptive or liberal and upheld by the outcomes from cerebrum investigate, with a structure like that of natural neural networks, could display comparable wise usefulness. In view of this base up figure, ANN (otherwise called connectionist models, parallel disseminated preparing models, neuromorphic frameworks, or essentially neural networks) have been produced as organically motivated and conceivable models for different assignments. [12] Biological neural networks are made out of numerous greatly interconnected crude organic neurons. Every neuron has axons and handwrites finger-like projections that empower the neuron to speak with its neighboring neurons by transmitting and getting electrical and substance signals. Pretty much taking after the structure of their partners, ANN is made out of interconnected, basic handling components called simulated neurons.ANN has some alluring characteristics like those of natural neural networks, for example, the capacities of learning, self-association, and adaptation to internal failure.

Rule Based Classification

In a rule-based classification show, the data is spoken to as though THEN rules organize. The if-part is known as the rule predecessor and the else part is the rule resulting. In the event that the condition in the rule precursor part is genuine, it implies that the rule covers the tulle and in the event that it is false then the rule does not cover the tulle. A rule is evaluated by its scope and precision. Rules scope is the level of tulles secured by the rule and its exactness is the level of accuracy in classification. The if-then rules can be extricated from the choice tree as the if-then classification rules are anything but difficult to decipher by people. These rules can likewise be removed from consecutive covering calculations.

The vast majority of the rule-construct classification techniques center in light of execution instead of interpretability of data. ROUSER is a rule-construct classification technique which centers with respect to human justifiable choice rules from data. It utilizes a harsh set way to deal with select trait esteem combine for the IF condition and it is said to be more effective contrasted with other rule-based classification strategies [11]

The fluffy rule-based classification framework is hard to manage because of the exponential development of the fluffy rule look space when the quantity of examples turns out to be high. This influences the learning to process more troublesome and it prompts issues of adaptability (as far as the time and memory devoured) and additionally intricacy (concerning the number of rules got and the number of factors incorporated into each rule). A fluffy association rule-based classification technique for high-dimensional issues has been proposed to acquire a precise and conservative fluffy rule-based classifier with a low computational cost [12].

SVM Algorithm

SVM have pulled in a lot of consideration in the most recent decade and effectively connected to different spaces applications. SVMs are regularly utilized for learning classification, relapse or positioning capacity. SVM depends on factual learning hypothesis and basic hazard minimization primary and have the point of determining the area of choice limits otherwise called hyper plane that delivers the ideal division of classes [1] [2] [3].Maximizing the edge and in this way making the biggest conceivable separation between the isolating hyper plane and the occasions on either side of it has been demonstrated to diminish an upper bound on the normal speculation mistake [8]. The productivity of SVM based classification isn't straightforwardly relied upon the measurement of characterized elements. In spite of the fact that SVM is the most powerful and exact classification procedure, there are a few issues.

The data investigation in SVM depends on raised quadratic programming, and it is computationally costly, as explaining quadratic programming strategies require expansive grid tasks and in addition tedious numerical calculations [4]. Preparing time for SVM scales quadratic ally in the number of illustrations, so inquires about endeavor all the ideal opportunity for more effective preparing algorithm[5], bringing about a few variation based calculation.

Classifying data is a common task in machine learning. Suppose some given data points each belong to one of two classes, and the goal is to decide which class a new Data point will be in. In the case of support vector machines, a data point is viewed as a {\display style p} p-dimensional vector (a list of {\display style p} p numbers), and we want to know whether we can separate such points with a {\display style (p-1)} (p-1)-dimensional hyper plane. This is called a linear classifier. There are many hyper planes that might classify the data. One reasonable choice as the best hyper plane is the one that represents the largest separation, or margin, between the two classes. So we choose the hyper plane so that the distance from it to the nearest data point on each side is maximized. If such a hyper plane exists, it is known as the maximum-margin hyper plane and the linear classifier it defines is known as a maximum margin classifier; or equivalently, the perception of optimal stability.
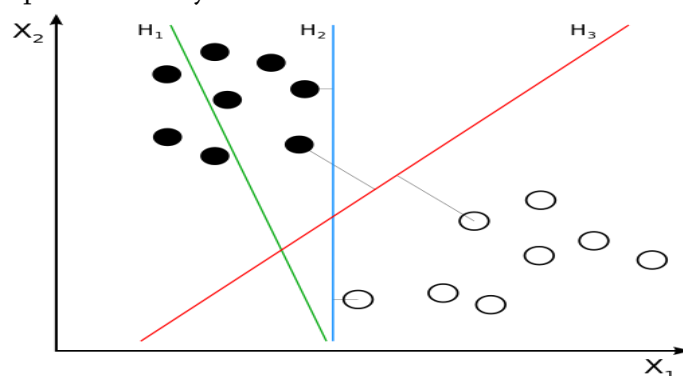

Figure 2. Svm Classification

SVM can likewise be stretched out to learn non-straight choice capacities by first anticipating the

information data onto a high-dimensional element space utilizing piece works and planning a direct classification issue in that component space [4]. The subsequent component space is considerably bigger than the span of the dataset which isn't conceivable to store in well-known PCs. Examination of this issues prompts a few decay based calculations. The fundamental thought of decay technique is to part the factors into two sections: an arrangement of free factors called as working set, which can be refreshed in every emphasis and set of settled factors, which are settled at a specific esteem briefly. This strategy is rehashed to the point that the end conditions are met [5]. Initially, the SVM was created for double classification, and it isn't easy to expand it for multi-class classification issue. The essential plan to apply multi-classification to SVM is to break down the multi-class issues into a few two-class issues that can be tended to straightforwardly utilize a few SVMs [6].

ANN Algorithm

Artificial neural networks (ANNs) are kinds of PC design enlivened by natural neural networks (Nervous frameworks of the cerebrum) and are utilized to inexact capacities that can rely upon an expansive number of sources of info and are for the most part obscure. Artificial neural networks are introduced as frameworks of interconnected "neurons" which can register esteems from inputs and are equipped for machine learning and also design acknowledgment because of their versatile nature.

An artificial neural network works by making associations between wide ranges of preparing components each relating to a solitary neuron in an organic cerebrum. These neurons might be really developed or reenacted by a computerized PC framework. Every neuron takes numerous information flags at that point in light of an interior weighting produces a solitary yield flag that is sent as contribution to another neuron. The neurons are unequivocally interconnected and sorted out into

various layers. The information layer gets the information and the yield layer delivers the last yield. When all is said in done, at least one shrouded layers are sandwiched in the middle of the two. This structure makes it difficult to gauge or know the correct stream of data.

Artificial neural networks ordinarily begin with randomized weights for every one of their neurons. This implies at first they should be prepared to tackle the specific issue for which they are proposed. A back-spread ANN is prepared by people to perform particular errands. Amid the preparation time frame, we can assess whether the ANN's yield is right by watching the example. On the off chance that it's right the neural weightings that delivered that yield are fortified; if the yield is inaccurate, those weightings capable can be decreased.

Actualized on a solitary PC, an artificial neural network is regularly slower than more conventional arrangements of calculations. The ANN's parallel nature enables it to be assembled utilizing various processors giving it an extraordinary speed advantage at almost no advancement cost. The parallel design enables ANNs to process a lot of data effectively in less time. When managing vast constant floods of data, for example, discourse acknowledgment or machine sensor data ANNs can work extensively speedier when contrasted with different calculations. An artificial neural network is helpful in an assortment of true applications, for example, visual example acknowledgment and discourse acknowledgment that arrangements with complex regularly fragmented data. Also, late projects for content to-discourse have used ANNs. Numerous penmanship examination programs, (for example, those utilized as a part of famous PDAs) are as of now utilizing ANNs.

Alternative methods of classification

Other than the previously mentioned classification strategies other established techniques additionally

exits. To give some examples of them, classification by back spread, harsh set approach, bolster vector machines, hereditary calculations, and classification by association rule investigation, fluffy set and some more. The strategies for classification are picked in light of the client and data classification needs. A portion of the issues looked by these techniques is inspected beneath.

The back spread is a neural network learning calculation and it learns by iteratively preparing a data set of preparing tulles. It looks at the network expectation of each tulle with class marks. Back Propagation networks are perfect for straightforward Pattern Recognition and Mapping Tasks and they are utilized for classification of data. One of the notable issues in back engendering strategy is the issue experienced with neighborhood maxima [16].Many varieties to back proliferation technique have been proposed to conquer the issues looked because of developing network estimate.

Bolster vector machines were intended for paired classification of data i.e. grouping data into two classes. A few research works are done to stretch out these paired classification techniques to multiclass classification. As explaining multiclass classification is computationally costly numerous relative investigations to enhance this system has been done [17]. The help machine vector is likewise utilized for classification of straight and nonlinear data. It changes the data in a higher measurement from where it can discover a hyper plane for partition of data utilizing preparing tulles called bolster vector.

The hereditary calculations, harsh set approach, and fluffy sets are calculations which are not utilized frequently. However, their rationales are connected to other classification techniques. The Fuzzy set hypothesis can be utilized as a part of frameworks which perform rule-based classification. The unpleasant sets can be utilized for property determination or highlight decrease where traits that

don't contribute to a classification can be distinguished and expelled.

Advantages and Disadvantages of Classification Algorithm

Sr. No	Algorithm	Features	Limitations
1	C4.5 Algorithm	-Build Models can be easily interpreted. -Easy to implement. -Can use both discrete and continuous values -Deals with noise.	-Small variation in data can lead to different decision trees. -Does not work very well on a small training data set. -Overfitting.
2	ID3 Algorithm	-It produces the more accuracy result than the C4.5 algorithm. -Detection rate is increase and space consumption is reduced.	-Requires large searching time. -Sometimes it may generate very long rules which are very hard to prune. -Requires large amount of memory to store tree.
3	K-Nearestneighbor Algorithm	-Classes need not be linearly separable. -Zero cost of the learning process. -Sometimes it is Robust with regard to noisy training data. -Well suited for multimodal classes.	-Time to find the nearest Neighbours in a large training data set can be excessive. -It is Sensitive to noisy or irrelevant attributes. -Performance of algorithm depends on the number of dimensions used.
4	Naive Bayes Algorithm	-Simple to implement. -Great Computational efficiency and classification rate. -It predicts accurate results for most of the classification and prediction problems.	-The precision of algorithm decreases if the amount of data is less. -For obtaining good results it requires a very large number of records.
5	Support vector-machine Algorithm	-High accuracy. -Work well even if data is not linearly separable in the base feature space.	-Speed and size requirement both in training and testing is more. -High complexity and extensive memory requirements for classification in many cases.
6	Artificial Neural Network Algorithm	-It is easy to use, with few parameters to adjust. -A neural network learns and reprogramming is not needed. -Easy to implement. -Applicable to a wide range of problems in real life.	-Requires high processing time if neural network is large. -Difficult to know how many neurons and layers are necessary. -Learning can be slow.

Associative classification uses association mining techniques for mining data. Classification rule mining and association rule mining are two important data mining techniques. Association rule mining is one which finds the rules in the database that satisfy some minimum constraints whereas classification rule mining aims to discover a small set of rules in the database to form an accurate classifier. Both classification rule mining and association rule mining are vital to practical applications. It would be of value to the user if the two mining techniques can be integrated. Research works have been done to integrate the two frameworks, called associative classification [13].

IV. CONCLUSION

This paper centers around different classification techniques (factual and machine learning based) utilized as a part of data mining and an examination on every one of them. Data mining can be utilized as a part of a wide region that incorporates techniques from different fields including machine learning, Network interruption identification, spam sifting, artificial knowledge, insights and example acknowledgment for investigation of vast volumes of data. Classification techniques are normally solid for

displaying correspondences. Every one of these techniques can be utilized as a part of different circumstances as required where one has a tendency to be utilized while the other may not and the other way around. These classification calculations can be executed on various kinds of data sets like offer market data, data of patients, monetary data, and so on. Consequently, these classification techniques indicate how a data can be resolved and gathered when another arrangement of data is accessible. Every strategy has got its own component and confinements as given in the paper. In view of the Conditions, comparing execution and highlight every one as required can be chosen.

V. REFERENCES

[1]. Nike, Orient. J. Comp. Sci. & Technol., Vol. 8(1), 13-19 (2015), ORIENT JOURNAL COMPUTER SCIENCE AND TECHNOLOGY

[2]. J. Han and M. Camber, "Data Mining Concepts and Techniques", Elsevier, 2011.

[3]. V. Vapnik and C. Cortes, "Support Vector Network," Machine Learning, 20; 273-297, (1995).

[4]. C. J. Burges, "A Tutorial on Support Vector Machines for Pattern Recognition," Data Mining and Knowledge Discovery, 2; (1998).

[5]. H. Bhavsar, A. Ganatra, "A Comparative Study of Training Algorithms for Supervised Machine Learning", International Journal of Soft Computing and Engineering (IJSCE) ISSN: 2231 -2307, 2(4); (2012)

[6]. G. Wang, "A Survey on Training Algorithms for Support Vector Machine Classifiers", Fourth International Conference on Networked Computing and Advanced Information Management, 2008, IEEE.

[7]. G Madzarov, D. Gjorgievikj and I. Chorbev, " A Multi-class SVM Classifier Utilizing Binary Decision Tree", Informatica, pp. 233-241 (2009).

[8]. M. Aly, "Survey on Multiclass Classification Methods", November (2005).

[9]. V. Vapnik, "Statistical Learning Theory", Wiley, New York, (1998).

[10]. T.Joachims, "Making large-scale support vector machine learning practical", In Advances in Kernel Methods: Support Vector Machines, (1999).

[11]. J.Platt, "Fast training of SVMs using sequential minimal optimization", In B. Sch¨olkopf, C.Burges and A.Smola (ed.), Advances in Kernel Methods: Support Vector Learning, MIT Press, Cambridge, MA, 1999, 185-208.

[12]. D. Michie, D.J. Spiegelhalter, C.C. Taylor "Machine Learning, Neural and Statistical Classification", February 17, (1994).

[13]. Delveen Luqman Abd Al.Nabi, Shereen Shukri Ahmed, "Survey on Classification Algorithms for Data Mining: (Comparison and Evaluation)" (ISSN 2222-2863)4(8); (2013)

[14]. Riaan Smit" An Overview of Support Vector Machines, 30 March 2011.

[15]. Nikita Jain, Vishal Srivastava, "DATA MINING TECHNIQUES: A SURVEY PAPER"-IJRET - Nov 2013

[16]. Qiang Yang and Xindongwu, "10 CHALLENGING PROBLEMS IN DATA MINING RESEARCH", International Journal of Information Technology & Decision MakingVol. 5, No. 4 (2006) 597-604

[17]. R. Rojas: "The Backpropagation Algorithm", Neural Networks, Springer-Verlag, Berlin, 1996

[18]. Chih-Wei Hsu; Chih-Jen Lin, "A comparison of methods for multiclass support vector machines", Neural Networks, IEEE Transactions on (Volume: 13, Issue: 2) on Mar 2002

[19]. Bing Liu Wynne Hsu Yiming Ma; "Integrating Classification and Association Rule Mining", American Association for Artificial Intelligence -1998

[20]. Pernkopf, F.; Wohlmayr, M.; Tschiatschek, S.; "Maximum Margin Bayesian Network Classifiers"; Pattern Analysis and Machine

Intelligence, IEEE Transactions on (Volume: 34, Issue: 3)March 2012

[21]. [20]Barros, R.C.; de Carvalho, A.C.P.L.F.; Basgalupp, M.R.; Quiles, M.G.;"A clustering-based decision tree induction algorithm", Intelligent Systems Design and Applications (ISDA), 2011 11th International Conference on Nov 2011

[22]. P. Domingos and G. Hulten, "Mining high-speed data streams", Proc. 6th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, pp. 71-80, 2000.

[23]. V. Garcia, E. Debreuve, and M. Barlaud, "Fast k nearest neighbor search using GPU," in Proc. of 2008 IEEE Computer Society Conference on Computer Vision and Pattern Recognition Workshops, Anchorage, Alaska, USA, June 23-28, 2008, pp. 1-6

[24]. Masud, M.M., Jing Gao; Khan, L.; Jiawei Han; Thuraisingham, Bhavani, "A Practical Approach to Classify Evolving Data Streams: Training with Limited Amount of Labeled Data ", Data Mining, 2008. ICDM '08. Eighth IEEE International Conference on 15-19 Dec. 2008

[25]. A. Bifet and R. Kirkby, Data Stream Mining a Practical Approach, University of WAIKATO, Technical Report, 2009.

[26]. James Dougherty, Ron Kohavi, and MehranSahami.Supervised and unsupervised discretization of continuous features. In International Conference on Machine Learning, pages 194-202, 1995.

[27]. David Hand, HeikkiMannila and Padhraic Smyth, Principles of Data Mining, MIT Press, 2001

[28]. Their Nu Phyu, Survey of Classification Techniques in DataMining, Proceedings of the International MultiConference of Engineers and Computer Scientists 2009

[29]. E.W.T. Ngai a,*, Li Xiu b, D.C.K. Chau a, Application of data mining techniques in customer relationship management: A literature review and classification, Expert Systems with Applications (2009)

[30]. R. Agrawal, T. Imielinski, and A. Swami. Database mining: A performance perspective. IEEE Trans. on Knowledge and Data Engineering, 5(6), Dec. 1993.

[31]. E. Manolakos and I. Stamoulis, "IP-cores design for the kNN classifier," in Proc. of IEEE International Symposium on Circuits and Systems, Paris, France, May 30- June 2, 2010, pp. 4133 - 4136.