

Research Challenges and Comparative Study of Various Classification Technique Using Data Mining

Harshang G Patel¹, Prof. Ketan Sarvakar²

¹M. Tech Student, U.V Patel College of Engineering, Ganpat University, Kherva, Mehsana, India

²Assistant Professor, U.V Patel College of Engineering, Ganpat University, Kherva, Mehsana, India

Abstract: The Data Mining refers to extracting or mining knowledge, information from the large amount of Data. The main purpose of data mining is data analysis. In Data Mining various techniques that used are Association Rule Mining, Sequential Pattern Mining, Clustering, and Classification. Classification is a data mining technique used to predict the class label or membership data. In this paper, we present the basic classification techniques. Several major kinds of technique, including Decision trees (DTs), Naive Bayes, K-nearest neighbor (K-NN), Artificial Neural Networks (ANN). The main goal of this survey is to provide Comparative review of various classification techniques in data mining.

Keywords: - Data Mining, Classification, K-nearest Neighbor (KNN), Decision Tree, Naive Bayes, Artificial Neural Network (ANN)

I. INTRODUCTION

Data Mining [1] is the process of extracting previously unknown information from a large datasets. Data mining is also known as knowledge mining from data, knowledge extraction, data/pattern analysis, data archaeology, and data dredging. Remember that the mining of gold from rocks or sand is referred to as gold mining rather than rock or sand mining. [1] Thus, data mining should have been more appropriately named “knowledge mining from data,” which is unfortunately somewhat long. Data mining is considered one of the most important frontiers in database systems and one of the most promising interdisciplinary developments in the information industry.

Today, data mining is being used by several Industries, Medical, Education, Finance and Banking, etc... It is a clever technique that can be applied to extract useful patterns. These tools can include statistical models, machine learning methods, and mathematical algorithm. Accordingly, data mining consists of more than a collection and managing data, it also contains an analysis and prediction. The Knowledge Discovery in Databases (KDD) process includes: [1]

- Selection of data
- Preprocess on data (remove noise form data)
- Transformation of data (sort or group)
- Data mining (identify a pattern)
- Evaluation (display o/p pattern)

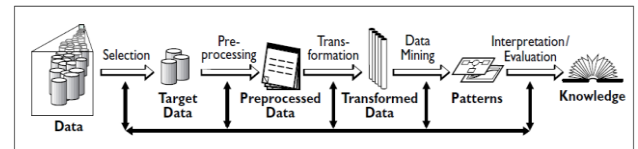


Figure1: KDD Process [1]

Data Mining Techniques which carry out the assigning of objects into related classes are called classifiers. According to the definition, Classification predicts the class label, Classification Techniques include two main phases in the first phase, they try to find a model for the class attribute as a function of other variables of the datasets, and in the second phase, they apply previously designed model for the new and unseen datasets for determining the related class of each record. [1]

In this paper, we discuss Comparative Study of Various Classification Technique Using Data Mining and research challenges. The paper is organized as follows. In Section I, we give the basic concept of data mining and Classification. In Section II, we describe a classification in data mining. Section III provides some of the research challenges in this field. Section IV contains different techniques with their limitations. A tabular comparison of different techniques of classification given by different authors is shown in section V. And finally we conclude in Section VI.

II. CLASSIFICATION DATA MINING

Classification is the process of finding a model (or function) that describes and distinguishes data classes or concepts, for the purpose of predicting the class of objects whose class label is unknown. Apply the model to previously unseen records to predict their class. [1]

The derived model is based on the analysis of a set of training data E.g. Classify Countries based on climate, Teachers classify students' grades as A, B, C, D, or F, Speech Recognition and Pattern Recognition. There are various data mining techniques are preprocessing, association, classification, pattern recognition and clustering [1]. Classification [21] is one of the most useful techniques in data mining to build classification models from an input data set. The used classification techniques to commonly build models that are used to predict future data trends [22] [23].

The DataSet can be divided into 2 parts:-

- **Training Set:** -Used to Build the Model.

- **Test Set:** - Used to determine the accuracy of the Model.

Model Construction: describing a set of predetermined classes [4]

- Each tuple/sample is assumed to belong to a predefined class, as determined by the class label attribute
- The set of tuples used for model construction: training set.
- The model is represented as classification rules, decision trees, or mathematical formulae.
- The accuracy rate is the percentage of test set samples that are correctly classified by the model.
- The test set is independent of training set, otherwise over-fitting will occur.

There are different Techniques [1] [2] used for data classification such as Decision Trees (DT), Naive Bayes (NB), K-Nearest Neighbor (K-NN), Artificial Neural Networks (ANN). The comparison of the classifiers and using the most predictive classifier is very important. Each of the classification methods show various effectiveness and accuracy based on the kind of dataset [4]. Classification and association are the popular techniques used to predict user interest and relationship between those data items which has been used by others. [2]

In this paper, we have worked with Various data mining, classification algorithm/technique, these algorithms have been applied to different data set to find out the efficiency of the algorithm and improve the performance by applying data preprocessing techniques and feature selection and also prediction of new class labels.

III. RESEARCH CHALLENGES

Today's, Data Mining used in Many Application. So, There are various areas where data mining, classification used, but the main research challenges are described below:-

1) *Data Cleaning*: - Preprocess data in order to remove or reduce the noise (by applying smoothing techniques) and handle those missing values. (i.e. By replacing a missing value with the most commonly occurring value for that attribute, or with most probable value based on statistics) although most classification algorithms have some mechanisms for handling noise or missing data, this Step can help reduce confusion during learning.

2) *Relevance analysis (feature selection)*: - Remove the insignificant or unnecessary attributes. Many of the attributes in the data may be irrelevant to the classification or prediction task. For example, data recording the day of the week on which a bank loan application was filed is unlikely to be relevant to the success of the application. Furthermore, other attributes may be unnecessary. Hence, relevance analysis May be performed on the data with the aim of removing any inappropriate or unnecessary attributes from the learning process. The time spent on relevance analysis, when added to the time spent on learning from the

resulting "reduced" feature subset, should be less than the time that would have been spent on learning from the original set of features. Hence, such Analysis can help to progress, classification efficiency and scalability.

3) *Data transformation*: -It Generalize and/or normalize the data.

Numerical attribute income \Rightarrow categorical {Low, medium, high}

Normalize all numerical attributes to [0,1] The data can be generalized to higher-level concepts. This is particularly useful for continuous valued attributes. For example, numeric values for the attribute income may be generalized to discrete ranges such as low, medium, and high. Similarly, nominal-valued attributes, like a street, can be generalized to higher-level concepts, like city, Since Generalization abbreviate the original training data, minor input/output operations may Be involved during learning. Also, Large Database, Data Scalability Overfitting, Automation, accuracy, robustness, interpretability.

IV. CLASSIFICATION DATA MINING TECHNIQUES

In this section, we focus on the various Classification Techniques which are used for the performance evaluation, below are listed.

1) *Decision Trees (DTs)*: -The Decision Trees [3] are a non-parametric supervised learning method used for classification and regression. Usually it follows a basic greedy approach algorithm. The goal is to create a model that predicts the value of a target variable by learning simple decision rules involved from the data features. A decision tree can easily modify to a set of classification rules. Some of the most well-known decision tree algorithms are C4.5 (successor of ID3), ID3 [24] (Iterative Dichotomiser 3) CART (Classification And Regression Tree), CHAID (CHi-squared Automatic Interaction Detector), MARS (Multivariate Adaptive Regression Splines) to extend decision trees to better handle numerical data, Hunt's Algorithm.

Some of the strength of DTs are they are computationally inexpensive, easy to use and implement and simple. It also provides objective analysis to decision making, allows flexibility and effective for decision making, quite fast at classifying unknown records.

Major Weakness of DTs is that the whole process depends on the accuracy of the input data used and also requires close together data to determine the accuracy of the output, error-prone with too many classes.

Decision trees used in data mining are of two main types:

- *Classification tree*: - analysis is used to predict outcome of the class to which the data belongs.
- *Regression tree*: - analysis is predicted when the outcome is considered as a real number. (e.g.. The price of a House, or a patient's length of stay in a hospital).

2) *K-NEAREST NEIGHBOR*: -The K-NEAREST NEIGHBOR algorithm (K-NN) [27] is a non-parametric method used for classification and regression. K-NN is a

type of instance-learning, or lazy learning, where the function is only approximated locally and all computation is deferred until classification. The K-NN algorithm is among the simplest of all machine learning algorithms and in which the training data set is stored, so that a classification of a new unclassified record may be found simply by comparing it to the most similar records in the training set. (E.g., Distance functions).

The algorithm can be summarized as:

- A positive integer k is specified, along with a new sample
- We select the k entries in our database which are closest to the new sample
- We find the most common classification of these entries
- This is the classification we give to the new sample

The K-nearest-neighbor (KNN) algorithm measures the distance between a query scenario and a set of scenarios in the data set. We can compute the distance between the two scenarios using some distance function $d(x, y)$, where x, y , are scenarios composed of N features, such that $x = \{x_1, \dots, x_n\}$, $y = \{y_1, \dots, y_n\}$, Two distance functions are discussed below summary:

-Distance Function:-

Euclidean $\sqrt{\sum_{i=1}^k (x_i - y_i)^2}$

Manhattan $\sum_{i=1}^k |x_i - y_i|$

Manhattan $(\sum_{i=1}^k (|x_i - y_i|)^q)^{1/q}$

Because the distance between the two scenarios is dependant of the intervals, it is recommended that resulting distances be scaled such that the arithmetic mean across the dataset is 0 and the standard deviation 1.

Some of strength is, it performs better with missing data, easy to implement and debug, Provides more accurate results, noise reduction techniques are used that improve the accuracy.

Some of its weaknesses are, its memory dependency, time consuming, computational complexity and also its reliance on k-value, slow technique since all process is done during the run time, low efficiency- prevents a lazy learning method it in many such applications such as dynamic web mining for a large repository. [4] [5]

3) *Naive Bayes*: - The Naive Bayes (NB) [7] classifier is based on Bayes' theorem with independence presumptions between the predictors. A Naive Bayesian model is easy to build, with no complicated iterative parameter estimation which makes it particularly useful for very large datasets. It is particularly suited when the dimensionality of the inputs is high. The Bayesian classification is used as a probabilistic learning method (Naive Bayes text classification). Naive Bayes classifiers are among the most successful known algorithms for Learning to classify text documents.

Bayesian networks are graphical models, which unlikely naive Bayesian classifier, it allows the

representation of dependencies among subsets of attributes [25]. Bayesian belief networks can also be used for classification. A simplified assumption: attributes are conditionally independent:

$$P(c|x) = \frac{P(x|c)P(c)}{P(x)}$$

- $P(c|x)$ is the posterior probability of class (target) given predictor (attribute).
- $P(c)$ is the prior probability of class.
- $P(x|c)$ is the likelihood which is the probability of predictor given class.
- $P(x)$ is the prior probability of predicting.

The Naive Bayes classifier provides a simple approach, with clear meaning, representation and learning probabilistic knowledge. It is also known as naive because is to depend on two important simplifying assumes that the predictive attributes are conditionally independent given the class, and it assumes that no hidden or unrealized attributes determine the prediction process. Some of strength are, Easy to implement & understand, Good result obtains most of the classes, It provides high accuracy and speed on large database, It can also handle real and discrete values. Some of its weaknesses are Class consider as an independent, so loss of accuracy.

4) *Artificial Neural Networks*: - Artificial Neural Networks (ANN) [6] [8] are one of the familiar classification methods in data mining. To Employ Neural Network based classifiers, Multi Layer Perceptron (MLP) and Radial Base Function (RBF) were used in this paper. MLP is a feed forward network that makes a model to show input data to output data. Hidden layer in MLP can allow various layers between input and output. The structure of MLP is shown in Fig.1

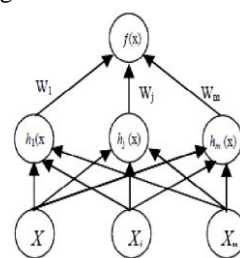


Fig.1 Radial schema Base network

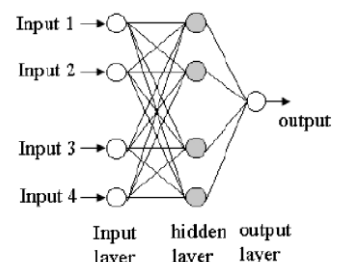


Fig.2 Multi Layer Perceptron

RBF is another type on ANN. The input of NN in RBF is linear and the output is nonlinear. The output of this type of ANN is taken from a weighted sum of hidden layer's output. The RBF networks are Divided into feed-forward layer. Fig2 illustrating the structure of this network (The figure is adapted from [5] [8]) Artificial Neural Network (ANN) is a computational model based on biological neural network. ANN also called Neural Network [26].

Some of strength are, Data Driven, Self-adaptive, high accuracy, Fault & Noise Tolerance.

Some of its weaknesses are lack of transparency, Define classification rules very difficult.

I. TABLE I: TABULAR COMPARISON OF DIFFERENT TECHNIQUES

Sr. No	Authors	Pub. Yr	Technique used for classification	Approach	Result
1	Kanu Patel, Jay Vala, Jaymit Pandya	2014	Decision Trees (DT), Rule Based Methods, Logistic Regression (LogR), Linear Regression (LR), Naive Bayes (NB), Support Vector Machine (SVM), K-Nearest Neighbor (K-NN), Artificial Neural Networks, Linear Classifier (LC)	A Concept of various classifications is used to predict the class label, The comparison of the classifiers and using the most predictive classifier is very important. Each of the classification methods show different efficiency and accuracy based on the kind of dataset & different techniques handle incomplete data, offer a natural interpretation to instructors and be accurate working with small samples.	In this paper among all techniques multilayer perceptron is more accurate and efficient in all parameters like TP-rate, FP-rate, Precision, Recall and ROC area compared to the others classification technique.
2	Rashedur M. Rahman, Farhana Afroz	2013	Bayes network, Multilayer Perceptron, Decision Tree, J48graft, J4.5	The Technique that has the potential to significantly improve the common or conventional methods will be suggested for use in large scale data, Bioinformatics or other general applications.	In this paper among all techniques the highest accuracy belongs to J48graft and lowest accuracy that belongs to FLR.J48graft is the best, second best is the Bayes Net and MLP & JRip is moderate but FLR is arguable.
3	Nikhil N. Salvithal, Dr. R. B. Kulkarni	2013	Decision Tree Tech, Naïve Bayes, Lazy IBK	This Approach works basically performance analysis, factor like classification technique, various nature of dataset, type of class and various tools.	In this paper among all techniques J48 performance and accuracy is good, take Less time to build the model compared to the others classification technique.
4	V. Vaithyanathan, K. Rajeswari, Kapil Tajane, Rahul Pitale	2013	J48, Naive Bayes updatable, Multilayer perceptron	They tried to compare performance of various classification techniques for different datasets using Weka tools and Measure the Performance.	In this paper among all techniques the efficiency of Naive Bayes Updatable gives the best result compare to others classification technique
5	K. Wisaeng	2013	Decision Tree, J48, -Graft, LAD tree, NavieBayes	They tried to compare bank dataset with different classification technique and weak tools to measure the performance, which one give high accuracy, Sensitivity or not.	In this paper among all techniques support vector machines achieves highest sensitivity, specificity and accuracy. By other hand, the worst classification was performed by radial basis function network.
6	Iain Brown, University of Southampton	2012	Logistic Regression (LOG), Linear and Quadratic Discriminant analysis (LDA, QDA), Decision Trees (C4.5), Neural Networks (NN), Nearest-Neighbor Classifiers	This study aims to compare the performance of a wide range of classification techniques within a credit scoring context, thereby assessing	In this paper among all techniques the credit scoring techniques, linear discriminant analysis (LDA) and logistic regression (LOG), give

			(k-NN10, k-NN100) and Gradient Boosting algorithm.	to what extent they are affected by increasing class imbalance.	good results reasonably the good result compare to others classification technique.
7	Peiman Mamani Barnaghi, Vahid Alizadeh Sahzabi Azuraliza Abu Bakar	2012	ANN (MLP, RBF), DT (J48, LMT), Bayes (Nave Bayes, Bayesian network) and rough set	An approach of the blood test dataset and different Classification Techniques to learn from the test data set and develop a system that is able to identify the existing of a liver disorder by processing the blood test data. And measured the accuracy of the Classifications Techniques.	In this paper among all techniques J48, MLP and RBF with have higher accuracy compared to other methods, Compared to Bayesian and Rough Sets, Neural Networks classifier methods obtain a good result. MLP obtains higher Results than RFB and also J48 shows good results, but Rough Sets did not perform well to classify the experimental dataset compare to other methods. MLP shows that can it provide better results with larger training set.
8	K.R. Lakshmi , M.Veera rishna , S.Prem Kumar	2013	SVM, LDA, C4.5, k-NN, BLR, MLR, PLS-LDA, k-means, EMC(Entropy based MeanClustering) and Apriori algorithm	Analyzing system used to transform original dataset to predict the actual Disease, measure the accuracy, Performance, computational time.	In this paper among all techniques the PLS-DA is the best one among ten Classifier, so the Accuracy of PLS-DA is high compared to others results. PLS-DA algorithm plays a vital role in data mining techniques, The performance of PLS-DA shows the high level compare with other classifiers.
9	Sonam Narwal, Mr. Kamaldeep Mintwal	2013	Bayesian Networks, J48, C4.5, ID3, K-means	Approach is to provide a detailed introduction of weka clustering algorithms. It provides the past project data for analysis. With the help of dataset, implementation showing the working of various algorithms used in weka.	In this paper among all techniques we found that k-means clustering algorithm is the simplest algorithm as compared to other algorithms. In clustering J48 shows the best performance considering both accuracy and speed.
10	Anshul Goyal and Rajni Mehta	2012	Naive Bayes, J48	Performance of dataset measurement through easily identify the which classification techniques give better accuracy, better performance	In this paper among all J48 gives more classification accuracy for class, gender in bank dataset having two values Male and Female. The result in the study on these datasets also shows that the efficiency and Accuracy of j48 and Naive Bayes is good.
11	S. Aruna, Dr S.P.	2011	Naive Bayes, Support vector	A supervised learning	In this paper among all

	Rajagopalan L.V. Nandakishore		machines, Radial basis neural networks, Decision trees J48 and simple CART	algorithm is used for data set and find out the best classifier with respect to disease detection, real data sets also compared the performance criterion of supervised techniques.	techniques the results are compared and found that SVM RBF Kernel is excellent in performance than the other classifiers with respect to accuracy, sensitivity, specificity and precision for both binary and multiclass datasets. SVM RBF Kernel always outperforms other classifiers for all datasets.
12	Diego Garcia-Saiz	2011	Owner, J48, Naive Bayes, BayesNet TAN, NNge.	The performance and interpretation level of the output Of the different classification techniques applied to educational datasets and propose a meta-algorithm to preprocess the datasets and improve the accuracy of the model and to determine which one is more suitable for wrapping in our EIWM tool.	In this paper among all techniques the accuracy depends on the sample size and the type of attributes. When the sample size is very small (less than 100 instances) and contains numeric attributes, Naive Bayes performs appropriately, on the other hand, when the dataset is bigger, BayesNet TAN is a better alternative. J48 is suitable for datasets with more instances and/or with the presence of nominal attributes with missing data, although in this Last context Naive Bayes is the best, but less interpretable.

IV. CONCLUSION AND FUTURE WORK

Data mining techniques and algorithms such as classification, clustering, etc., helps to find the patterns to decide upon the future trends to expand the businesses. Data mining is a very large area that Combined techniques from different fields including machine learning, artificial intelligence, statistics and pattern recognition, for the analysis of large amounts of data. There are so many data mining algorithms integrated in these fields to perform different data analysis tasks. The role of classification is to generate more proper and accurate system results classification techniques are being used in various application areas, and there is no single classifier which can perform better all the time for alternative data.

Our survey in this paper focuses on the existing literature present in the field of Classification Techniques (DTs, KNN, NBs, ANN) and research challenges in Data Mining. From our analysis, we have found that there is no single technique that is consistent with all domains. All methods perform in a different way depending on the type of dataset as well as the type of application or domain. But still from our analysis, we can conclude that Classification Techniques and algorithms perform better than the other existing methods. So these classification techniques show

how a data can be resolved and grouped, when a new set of data is available. Each technique has got its own Strength and weakness as given in the paper. Based on the needed conditions each one as needed can be selected.

Our future work will focus on the implement various classification techniques using various data set using WEKA Tools and to improvement of Classification Technique efficiency, accuracy Also, we want to propose a hybrid classification techniques approach to improve the performance.

REFERENCES

- [Book]**
- [1] J. Han and M. Kamber, "Data Mining: Concepts and Techniques", 2nd Ed., The Morgan Kaufmann Series in Data Management Systems, Jim Gray, Series Editor 2006.
- [Reference]**
- [2] N. Abirami, T. Kamalakannan, Dr. A. Muthukumaravel" A Study on Analysis of Various Data Mining Classification Techniques on Health Care Data" IJETAE 2013
 - [3] V. Vaithyanathan, K. Rajeswari, Kapil Tajane, Rahul Pitale", Comparison of Different Classification Techniques Using Different Datasets" IJETAE 2013
 - [4] Syeda Farha Shazmeen, Mirza Mustafa Ali Baig"Performance Evaluation of Different data mining Classification Algorithm and Predictive analysis", IOSR-JCE2013
 - [5] M. Soundarya, R. Balakrishnan" Survey on Classification Techniques in Data mining", IJARCCCE2014

- [6] Ms. Aparna Raj, Mrs. Bincy G, Mrs. T. Mathu "Survey on Common Data Mining Classification Techniques", International Journal of Wisdom Based Computing 2012
- [7] M.Soundarya1, R.Balakrishnan. "Survey on Classification Techniques in Data Mining", International Journal of Wisdom Based Computing 2012
- [8] Peiman Mamani Barnaghi, Vahid Alizadeh Sahzabi, Azuraliza Abu Bakar "A Comparative Study for Various Methods of Classification", ICICN 2012
- [9] G. H. John and P. Langley, "Estimating Continuous Distributions in Bayesian Classifier" Proceedings of the 11th Conference on Uncertainty in Artificial Intelligence, San Francisco, 1995, pp. 338-345.
- [10] Kanu Patel, Jay Vala, Jaymit Pandya "Comparison of various classification algorithms on iris datasets using WEKA", IJAERD 2014
- [11] Rashedur M. Rahman, Farhana Afroz "Comparison of Various Classification Techniques Using Different Data Mining Tools for Diabetes Diagnosis", scores 2013
- [12] Nikhil N. Salvithal, Dr. R. B. Kulkarni "Evaluating Performance of Data Mining Classification Algorithm in Weka" IJAEM 2013
- [13] Vaithiyathan, K. Rajeswari, Kapil Tajane, Rahul Pitale "COMPARISON OF DIFFERENT CLASSIFICATION TECHNIQUES USING DIFFERENT DATASETS", IJAET 2013
- [14] K. Wisaeng "A Comparison of Different Classification Techniques for Bank Direct Marketing", IJSCE 2013
- [15] Iain Brown, University of Southampton, Southampton "An experimental comparison of classification techniques for imbalanced credit scoring data sets using SAS® Enterprise Miner", SAS GLOBAL FORUM 2012
- [16] Peiman Mamani Barnaghi, Vahid Alizadeh Sahzabi and Azuraliza Abu Bakar "A Comparative Study for Various Methods of Classification", IPCSIT 2012
- [17] K.R. Lakshmi, M. Veera rishna, S. Prem Kumar. "Performance Comparison Of Data Mining Techniques for Predicting Of Heart Disease Survivability", ISSN 2250-3153 2013
- [18] Sonam Narwal, Mr. Kamaldeep Mintwal, "Comparison the Various Clustering and Classification Algorithms of WEKA Tools" IPCSIT 2012
- [19] Anshul Goyal and Rajni Mehta. "Performance Comparison of Naïve Bayes and J48 Classification Algorithms" ISSN 0973-4562 201
- [20] S. Aruna1, Dr S.P. Rajagopalan, L.V. Nandakishore. "AN EMPIRICAL COMPARISON OF SUPERVISED LEARNING ALGORITHMS IN DISEASE DETECTION", IJITCS-2011
- [21] Diego Garcia-Saiz, Marta Zorrilla. "Comparing classification methods for predicting distance students' performance" JMLR: Workshop and Conference Proceedings 17 (2011)

[Book]

- [22] Mikko Järvinen, "Femtocell Deployment in 3rd Generation Networks", Master's Thesis, HELSINKI UNIVERSITY OF TECHNOLOGY, 2009
- [23] Vikram Chandrasekhar and Jeffrey G. Andrews, The University of Texas at Austin, Alan Gatherer, Texas Instrument, "Femtocell Networks: A Survey", IEEE Communications Magazine, 2008
- [24] Khaled Elleithy and Varun Rao, "Femto Cells: Current Status and Future Directions", International Journal of Next-Generation Networks (IJNGN) Vol.3, No.1, March 2011
- [25] J.S R Jang (1993). ANFIS Adaptive Network Based Fuzzy inference System. IEEE transaction on Systems, Man and Cybernetics. Vol. 23, no3, pp 665-685

[Reference]

- [26] Yu-Shan Liang, Wei-Ho Chung, Member, IEEE, Guo-Kai Ni, Ing-Yi Chen, Hongke Zhang, and Sy-Yen Kuo, Fellow, "Resource Allocation with Interference Avoidance in OFDMA Femtocell Networks", IEEE TRANSACTIONS ON VEHICULAR TECHNOLOGY, VOL. 61, NO. 5, JUNE 2012

[Book]

- [27] Artificial Neural Networks Ajith Abraham Oklahoma State University, Stillwater, OK, USA.
- [28] D. Hand, H. Mannila, P. Smyth.: Principles Of Data Mining. The MIT Press. (2001).

BIOGRAPHY



Harshang G Patel received the B.E. Degree in Information Technology from the Dharamsinh Desai University, Gujarat, in 2013, he is in the final year of M.TECH pursuing in Information Technology from the Ganpat University, Kherva, Mehsana, Gujarat, India. His research areas include data mining and its Classification issues.



KETAN J SARVAKAR received B.E. Degree and M. Tech degree. He currently Works as Assistant Professor in Dept., Information Technology in the Ganpat University, Kherva, Mehsana, Mahesana, Gujarat. His research area is data mining and wireless sensor .