International Journal of Technical Innovation in Modern Engineering & Science (IJTIMES)
Impact Factor: 5.22 (SJIF-2017), e-ISSN: 2455-2585
Volume 5, Issue 05, May-2019

An approach to identify the strength of the decision tree algorithm by comparing various classifiers for customer relationship improvising

[1]Patel Alex, [2]Akash Saxena

[1] *Research Scholar, Rai University, Ahmedabad,* [2] *PhD Guide, Rai University, QA Engineer, Yosa Technology, Surat, Gujarat, India*
[2]*PhD Guide, Rai University, Ahmedabad, Gujarat, India*

Abstract: ***Objectives: To Compare classification algorithm and to identify the Strength of Decision Tree algorithm. Methods/Statistical analysis: Current study is to compare various classification algorithms theoretically. There are various sources of such data that majorly help us to study the user's pattern or access pattern. One of the important ways is Data Mining. Data Mining is an effective tool to derive useful data from unstructured and unmanaged data. To compare classification algorithms, it requires to compare estimated accuracy between the algorithms as we want the best algorithm for the dataset. By study and referring various papers various methods are compared.***
Findings: There are various methods comparing algorithms, but before we jump into that it requires to make the theoretical concept clear. We may not know best algorithm or best parameters to use for algorithms because we never knew which algorithm is going to give the best result for our dataset. Before one applies an algorithm to the dataset it requires to clear various terms, to help and decide which algorithm to use for specific problems. In this study, the result has been derived by comparison of various algorithms and its methods. Algorithms are compared on the basis of Data size, Accuracy, Speed of working, Type, Training time duration and its characteristics.
Application/Improvements: Purpose of comparison is not to prove any model high or low, but to search the accurate model as per problem. This comparison is theoretical so for the improvement it requires to work with data to compare the results.

Keywords: Decision Tree, Data mining, Comparison, Customers.

I. Introduction

Decision tree starts with a question which probably has two answers, where the first answer becomes the root node and second answer will be the leaf or nodes. Each answer will divert to another question or answer and this way data are classified or managed to get the proper identification, later these data can be arranged in proper structure or simply data can be categorized into pre-defined format. In Decision tree nodes are defined further on the bases of answers. It follows simple tree structure which can be diversified on previous nodes answers and on later stage pruning can be helpful to reduce the complexity from a final classifier for predicting accuracy in a much better way. When a tree is growing sometimes it requires removing noisy and erroneous data [1]. The final outcomes of decision tree depend on the attributes, precisely defined attributes are likely to result in accurately classified data. There is a relation between attributes and tree class, need to identify it and design for more effectiveness.

Decision tree can handle large amount of data with good accuracy. ID3, C4.5, CART are decision tree algorithms where C4.5 is the successor of ID3 algorithm and ID3 and CART (Classification and Regression Tree) were developed independently in the same time period. The decision tree is constructed in top to bottom manner. The decision tree is very easy to convert into classification rules, as each attribute is tested and outcomes are decided and this process can be derived from other attributes till proper outcomes are derived.

Whenever data is noisy or having missing values CART classification tree or regression tree is used for accuracy. It uses CHAID algorithm for handling the missing values. The input for the same is a random sample as it does not require data pre-processing. It is capable to select some random relevant attribute itself.

II. Proposed Work

The proposed is to state that Decision Tree classification is better than other classification for improvement of customer's satisfaction. There are a number of dimensions to choose the correct classification algorithm like if the problem is linearly separable or not? What is the size of data set? What are the requirements of the system? Each of classification methods behaves different as per data set size and real world situation where these are implemented. There is no single algorithm which fits best in every situation as every algorithm has its own advantages and disadvantages in its own situation [2]. One method can perform better under certain circumstances and similarly other methods can have different need, but classification methods can determine how useful information is derived using the proper selection as per the situation, behavior of the system, input data and size of the dataset. The classifier is chosen with the help of operational constraint

and once it is configured based on learning, the accuracy of the classifier is the criteria for characterizing the performance of the classifier and estimating the same rate considering the generalized factor. The classifiers are categorized on the basis of learning and widespread base which depends on classification problem that have been studied. The dependencies of classifiers are studied experimentally or theoretically to find results and relations between the classifiers [3].
For Customer care improvement, Decision tree is the best option as it is a simple structure. And guides to manage every possible solution for a problem. It breaks down the data into a smaller size of useful information. Decision tree takes us step by step to find a solution related to our decisions. Every solution may have multiple outcomes having its own branches and leading to a solution. Decision Tree can be Multivariate or Univariate but is a useful classification when the results are interpreted by humans, especially in case where the data are to be collected from e-commerce sites; and study shows that multivariate approach is performing better than the univariate approach while data are large in size [4]. As we all know that today it is word of information and data and huge amount data sets are generated on social media, e-commerce website, search engines etc., Hence we need to have a system that utilizes the same information to improve customer's satisfaction and help the organizations provide better services, fulfil their need in the form they need by providing a personalize systems. Seek. The decision tree algorithm is efficient mining method to mine and extract data from Web stream using pattern and helpful to find different information related to log file and using Decision tree author increase the accuracy of generating non redundant association rule for both nominal and numerical data with less time complexity and memory space [5].
Customer behavior and interests cannot be the constant and dataset can be just storage of large customer data samples, but to use it effectively Decision Tree can be the key as a modification of the decision tree is an advantage. Decision Tree and C4.5 shows accurate and effective performance with large size of database with simulated data and generated database for AUC where AUC is an abbreviation of the area under the curve [3]. As we are dealing with customer data which is considered as always having increasing size, we should focus on the algorithm which gives a consistence performance with any size of data. Similarly, Decision Tree branches which are not part of the current trend can be removed by the pruning process. Change is the only constant and it similarly applied to customer data also so upcoming new changes in the customer behavior can be added like a new branch to the tree and vice a versa, old can be removed. Another important factor of the Decision Tree is with Time it keeps on growing with increasing size of the customer dataset, but as classification process this can be viewed as large information chunk.

Customer behavior data can be the key to the company's success and with the help of Decision Tree it can be improved and managed well with accuracy and speed. Decision trees can be split or with large dataset and many decision tree can be used as random forest. Which can be viewed as assemble of multiple data trees. Decision Trees are likely to be a better fit for problems where you have a small number of features or may be a large number of features train examples like customer feedback data. The organization can't identify the need of each customer, so data mining with the technique of decision tree can enables you to handle and customize the needs of customers like suggestions, targets, bunches, discounts and many more things. The choice of the best and suitable classifier is already guided by operational constraints, but beyond this, and after that the classifier is configured through a learning basis, the rate of generalization of the accuracy of the classifier which is the criterion characterizing its performance. This rate, usually unknown, is estimated using a generalized basis. This estimate, therefore, depends on the classification problem as well as dataset, the classifier uses the learning base and widespread basis in the real world. These dependencies are studied theoretically and experimentally over a dozen different classifiers [6].

III. CLASSIFICATION METHODS

2.1 Classification Techniques

Decision Tree is a supervised method for classification and Regression. Classification and prediction methods can be dependent on many criteria like Accuracy of the prediction, the scalability to handle large amount of data, Robustness to make a correct prediction, and speed of working. Many classification techniques are available like Neural Networks, Support Vector Machines (SVM), K-Means (KNN) and Decision Tree. All techniques having advantage at one point and lacking something at the other end. In this research, we are going to study the approach and working on Decision Tree compared to other techniques and look into some methodologies of the decision tree. The decision tree is easy and convenient to use for estimating the different probable outcomes and classifies those records that contains unknown values. Trees are even able to adapt to the probabilities of the variables that are to be tested [1].
Major challenges with web usage mining are: First pre-processing the data, that needs data filtration and making the data available in usable and correct format. As we have to find exactly how the web site is being used to find some interesting patterns. Another issue is again filtered the data set to present the rules and patterns in order to understand the data mining of computer instructions which led to some activities during navigation and website surfing [1]. Decision Tree starts with the root node and each internal node is a condition or simply a question, each question has answers associated to and throughout this leaf node it has some predictions or solutions of the problem. Decision Tree breaks the data and keep it developing in the form of branches and nodes. Just as a normal tree grows from one shoot to two branches and then into many branches. Branches are attribute values and leaf nodes are the class values to which values are separated. Decision trees can be constructed fast compared to other classification methods and having better accuracy [7].
Classification Methods: Classification methods in data mining can be classified into three basic types: Supervised, unsupervised and reinforced [8]. Decision Tree is a supervised learning method and can be used for both classification and

regression. Decision trees are considered to be on the most effective method for supervised learning. It aims at the partition of datasets into groups of homogeneous possible variables to be predicted [1]. Unsupervised learning learns from data which are not being labelled or classified while supervised learning is where data with a label and classification. In supervised learning an algorithm is used to map input and output variables while in unsupervised learning we have only input variables.
Classification process is a meaningful way to derive the useful information from the unmanaged and noisy dataset. Decision Tree, Naïve Bayes, Support Vector Machine, Neural network and each has its working method and specialty depends on the type and size of data. Naive Bayes first scans data and due to this large data is required for better accuracy. Neural network works like a black box and initially it used to improve itself using the data and it continues for much cycle. A decision tree is aware of the classification function that fulfills the operations of concluding the value of dependent attribute over independent attributes.

2.2 Tree Construction

There are many decision tree algorithms like ID3, C4.5, C5.0, CART and these have merits of strong learning ability, to deal with noisy data, using less memory and with simple flow charts like structure it can provide accurate results. DTA is the classification technique that divides the elements starting from the root node to a leaf node. Decision tree algorithms are based on classification using attribute values for taking decisions. Decision tree represents data in a fashion which is very easily interpreted by users [9]. It is majorly used to make decisions based upon the tree formed that helps us to choose a path of activities which is next to happen or some parallel decision that can be taken. Something like plan A and Plan B or Plan N based upon the success of a pattern formed and decisions made [10]. The proposed model of the decision tree is used to form a structure from the taken data set and can be used as follows: To build tree first pick an attribute from given data and divide the dataset into parts on the basis of attributes and repeat the process until leaf is created. In case of large dataset sometime it also needs to remove branches which are not useful for classification through pruning, it could be post pruning or pre pruning process.

An attribute selection measure is splitting criteria that separate the given data by partitioning. Partitioning is a process where we divide datasets into branches kind of structure that can further split or bisect into many sub nodes or as needed. Tree node is created for that partition and like this tree grows with each criteria and data are partitioned accordingly.
There are many popular attribute selection measures like information gain, Gini index. ID3 uses information gain as its attribute selection measure, attribute with highest information gain is used to split attribute for the node. It helps to select the most informative feature. It has Key Entropy, it is the measurement of impurity or randomness in the database, simply high entropy with random database and vice versa. Gini Index is used in CART algorithm and it provides the smallest split value to choose the split node. Gini Coefficient of zero means equality, where the Gini coefficient of one express maximal inequality. Gini index can minimize the misclassification. Decision tree has one disadvantage of overfitting and this overfitting of branches can be solved by pruning methods where unwanted resulted branches can be removed using this technique. There are two types of pruning methods, Pre-Pruning and Post Pruning. While developing decision tree especially with large dataset, there are chances to add noise and these noise can be removed by tree pruning methods. Noise can create unwanted branches and we can manage to remove those unwanted branches for better results.

2.3 Neural Network

When it is about the small dataset Neural networks are not good with it, while in case of Decision Tree data set size does not a matter. On most of the example decision tree are able to handle with any amount of data except sometime the structure size gets complex. Neural network requires a lot of hard work on choosing the correct algorithms to solve the problem. Decision Tree use Entropy, Gini index, Information Gain measures to select the proper attributes. The neural network is a lengthy process to evaluate and test; while Decision Trees are based on divide and conquer rule and have simple structures of flowchart, so are easy to train. Decision tree does not require a lot of parameter managing criteria. Both Neural Networks and Decision Tree can handle nonlinear relationship and interaction between data. Neural Network works like a Black Box that is like a mobile phone, which delivers the result, but the user is not aware what is happening inside. So it is difficult to be aware how decisions are made and on what basis it has processed. Decision Tree provides open decision flows where decisions are made on selected attributes. As explained in Figure 1. Hidden layer behaves like a Black Box that gets the input and delivers the result without explanation of methods.

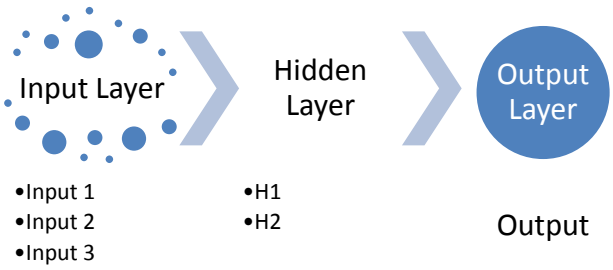

Figure 1. Understanding the working of Neural Network

2.4 Support *Vector*

Decision Tree is one of the good classifiers while SVMs are seen as a complex method and does not perform like Decision Tree. Major classifiers like SVM, Neural Network, KNN, Logistic Regression, Naïve Bayes, and Random Forests are tested on eleven different datasets and compared and found Boosted Decision tree and Random Forests are performing better [11].

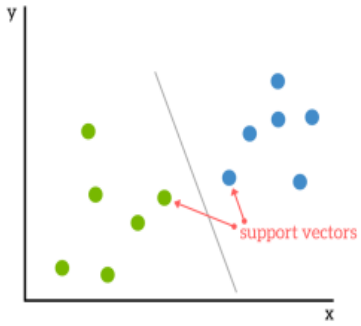

Figure 2. Support Vector

Support vectors are the data points near the hyperplane, the line separating two different data in Figure 2 and this is called hyperplane. Hyperplane segregates two class better. A margin of two data sets from the hyperplane helps to select the better hyperplane. SVM work better for text classification and in case problem is not linearly separable SVM can be used with non-linear kernel, but the problem here is it requires a lot of time for training, so it is not recommended with large number of training data set. In some case the inappropriate use of kernels, like a poor choice of the kernel or a linear kernel for non-linear problem is the problem, SVM expected to give proper result when the choice of the kernel is done correctly. It also faces the problem when the dataset has more noise. In some case it requires clear decision boundary to work well as sometime selecting right kernel may be challenged.

2.5 Logistic Regression

Logistic Regression is efficient in terms of time and memory and behave well when the problem is linearly separable. Logistic Regression can also be used in large dataset as it is efficient but on the other side, it gives you probability for each output. When a dataset is large in size and amount of data are missing, then it is a problem for the Logistic Regression as many attributes are missing in this case as it uses the entire data to derive the result. Logistic regression relies on the entire data and it is a global model influenced by the entire size of dataset. Decision Tree can automatically handle the interaction between variables, but in case of logistic regression it has to be manually. So in case dataset is having interaction between variables, then one has to perform it manually. It depends on the dataset and specific problem definition to use it or not. It generates the result based on rules along with metrics like Lift, Confidence and Support. Logistic regression Analysis is used to find the ratio of the probability of having the outcome by probability of not having Logistic regression works better for single decision boundary and high dimensional data, which may not necessarily be parallel to the axis whereas Decision trees can be applied for more than one underlying decision boundary, and high dimension data.
It still works best when the class labels may lie in hyper-rectangular regions. Logistic regression is simple and has low variance and hence it is less prone to over-fitting while Decision trees can be very complex, are more prone to over-fit as a result Pruning is done to remove or avoid the issue of over-fit.

2.6 Naïve Bayes

Naïve Bayes requires to build a classification manually. Unlike Decision Tree it does not have the attributes to pick and start classifier. Naïve Byes are simple algorithm to train and implement, but difficult to debug, especially when dataset does not contain all possibilities. It works on Bayesian theorem and it will answer the continuous classifier. Bayes theorem provides a way of calculating for posterior probability is given by:
P(c|x) from P(c) , P(x) and p(x|c) (2.1)
Where,
P(c|x) is the posterior probability of class (c, target) given predictor (x, attributes).
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 predictor.
Postrior = Prior x likelihood Evidence

Naive Bayes classifiers are used to handle continuous or categorical arbitrary number of independent variables. It can be presented as:
Set of variables, X = {x1, x2, x3..., xn},
Event Cj, from which we want to construct the posterior probability

C = {c1, c2, c3 ..., cn}, set of possible outcomes.
Or we can even say that X is a predictor and C is the set of levels in dependent variable.
Using Bayes' rule:

$$p(Cj|x1, x2, \ldots . xd)\alpha\ p(Cj|x1, x2, \ldots . xd)|Cj)\ p(Cj) \qquad (2.2)$$

Where p (Cj | x1, x2, x3..., xn) is the posterior probability of X's belonging to Cj. Nave Bayes requires a large number of datasets, but still it may perform with less accuracy [12].

I. *DISCUSSION*

3.1 Theoretical and Comparative Study of Classification Techniques considered for customer's satisfaction

Comparison of classification Techniques and their analysis is done to find a better way for providing data and result, so that customer's satisfaction can be improved based on the data and pattern available and results as shown in Table 1.

Table 1. Comparison of various Classification Techniques

Variable	Decision Tree	SVM	Neural Network	Naïve Bayes
Data	Large size of data and better fits into model, Requires more data to fit all possibilities	SVM work better for text classification, face the problem when the dataset has more noise	Good at handling Noise but it requires to select correct topology	Works well with low amount of training data, Only work with numeric data
Accuracy	More accurate results even with large dataset	High Accuracy	Good accuracy with a small network	Predict accurate results
Speed	Fastest to relate the attributes	Slower as compared to a decision tree	Normally slower compared to other algorithms	Simple and quick to evaluate
Type	Non-Probabilistic, Supervised Learning	Deterministic Algorithm	Non-Deterministic Algorithm, Supervised Learning	Probabilistic, Supervised Learning
Training	Easy to learn as well implement	Hard to learn in batch mode	Easy to learn in incremental fashion, but slow to train	It learns and reprogram
Property	Decision trees split the input space into hyper-rectangles according to the target.	SVM uses the kernel trick to turn a linearly non separable problem into a linearly separable one	Neural Networks learn from the iteration using data.	Naïve Bayes algorithm depends on Bayesian Theorem capable of calculating most possible outputs

II. CONCLUSION AND FUTURE WORK REWRITE

The key to a successful organization is their customer's satisfaction, the more they are satisfied more revenue can be generated and profit can be managed. This can be done by analyzing the customer's data so that a better and newer of providing services can be presented. In this paper, we have tried to present a detailed study of various classification techniques which are being used and can be used in a better way to improve the customer's satisfaction. Another aspect we need to consider is the increasing size of databases and we need to find the relevant dataset and their applications to fulfil our needs.

The data retrieved from the internet are of great use as they are collected directly or indirectly from the customers in the form of feedback, comments, registration, buying pattern etc. The decision tree is identified and studied as mean to find all possible outcomes that will be useful to improve the service. It helps us to identify the nodes and attributes that are more likely to support our decision and their dependency on each other. Also a major advantage of the decision tree is that, if there is no information available it uses a probabilistic approach to fill the gap or missing values based on some similarity and predictive models. Decision tree follows the same pattern like human to make decisions and implement in a more versatile manner. In future scope there is a lot to do like a framework for the same purpose and to derive the latest results from the big and recent data to provide more customized services. Using tools and sample dataset we can compare actual outcomes and can propose a more detailed work of method comparison for various classification techniques. The actual data can bring the real picture of all classification techniques and comparison can be clearer.

BIBLIOGRAPHY

[1] A. M. E. E. Badr HSSINA, "A comparative study of decision tree ID3 and C4.5," *(IJACSA) International Journal of Advanced Computer Science and Applications,*, p. 7.

[2] R. Geetika, "HIERARCHICAL CLUSTERING TECHNIQUES IN DATA MINING: A COMPARATIVE STUDY," *International Journal of Science Research and Technology*, p. 6, 25th March 2016.

[3] A. R. M. Reza Entezari-Maleki, "Comparison of Classification Methods Based on the Type of Attributes and Sample," *Journal of Convergence Information Technology ·*, p. 10, 2009.

[4] G. S. ,. R. B. ,. M. MiningDr. Neeraj Bhargava, "Decision Tree Analysis on J48 Algorithm for Data Mining," *International Journal of Advanced Research in Computer Science and Software Engineering*, p. 6, June 2013.

[5] S. S. L. Sheetal Sharma, "Development of Decision Tree Algorithm for mining Web Data Stream," *Internationl Journal of Computer Applications(0975-8887)*, p. 10, 2016.

[6] M. E. A. B. a. M. A. C. Nesma Settouti, "Statistical Comparisons of the Top 10 Algorithms in Data Mining for Classification Task," *International Journal of Interactive Multimedia and Artificial Intelligence*, , p. 6.

[7] S. K. Himani Sharma, "A Survey on Decision Tree Algorithms of Classification in Data Mining," *International Journal of Science and Research (IJSR)*, vol. Volume 5 , p. Issue 4, April 2016.

[8] S. G. P. a. D. K. I. L. Bhaskar N. Patel, "Efficient Classification of Data Using Decision Tree," *Bonfring International Journal of Data Mining,*, p. 7, March 2012.

[9] A. P. R. B. Pooja Sharma, "Implementation of Decision Tree Algorithm to Analysis the Performance," *International Journal of Advanced Research in Computer and Communication Engineering*, p. 4, December 2012.

[10] R. K. K. R. Suneetha, "Classification of Web Log Data to Identify Interested Users Using Decision Tree," *Classification of Web Log Data to Identify Interested Users Using Decision Tree*, p. 6.

[11] A. N.-M. Rich Caruana, "An Empirical Comparison of Supervised Learning Algorithms," *Appearing in Proceedings of the 23 rd International Con,Department of Computer Science, Cornell University, Ithaca, NY 14853 USA*, p. 8, 2006..

[12] H. P. C. Sayali D. Jadhav, "Comparative Study of K-NN, Naive Bayes and Decision Tree Classification Techniques," *International Journal of Science and Research (IJSR)*, p. 4, 2014.