ANALYSIS ON VARIOUS CLASSIFICATION MECHANISMS

Nomula Madhavi
Assoc. Prof.
Department of CSE
Siddhartha Institute of Technology and Sciences
Narapally, Hyderabad, Telangana, India

Chada Suresh
Asst. Professor
Department of CSE
Siddhartha Institute of Technology and Sciences
Narapally, Hyderabad, Telangana, India

Abstract- **In this study, different data mining classification approaches are examined using diverse datasets from UCI machine learning. Every method's precision and execution time are noticeable. The term "data mining" refers to the process of extracting knowledge from enormous amounts of data. Classification is an important data mining technique that has a wide range of applications. It organizes many types of information. Classification is used in almost every aspect of our lives. Everything in a group of data is organized into one of a preset set of classes or bunches using classification. This research was carried out on the Labor dataset using classification methods such as J48, Multilayer Perceptron, and Bayes Net. The performance of each method is evaluated using experimental results.**

Key Terms: Multilayer Perceptron, Bayes Net, J48, Data mining and classification

1. INTRODUCTION

Data mining is the process of identifying undiscovered, valid patterns and relationships in large data sets using a variety of sophisticated data analysis methods. Machine learning methods, statistical models, and mathematical algorithms are all examples of these tools. Data mining entails not only data gathering and management, but also analysis and prediction. The classification technique in data mining is becoming more popular since it can analyses a wider range of data than regression. Machine Learning (ML) has a variety of applications, the most important of which being data mining. Data mining, also known as Knowledge Discovery in Databases (KDD), is the process of extracting nontrivial, potentially useful, and previously unknown information from databases [1]. Association, classification, pattern recognition, and clustering are some of the data mining techniques [2]. Classification and association are common strategies for predicting user interest and the relationship between data items used by users [3, 4].

The labor dataset used in this paper taken from the UCI machine learning library. It contains a company's labor details in the form of 17 attributes, including duration, first-year wage increase, second-year wage increase, third-year wage increase, cost of living adjustments, working hours, pension, standby pay, shift differential, education allowance, statutory holidays, vacation, long-term disability assistance, contribution to dental plan, bereavement assistance, contribution to health plan, and class. To analyze the performance of this labor dataset, classification techniques such as J48 Decision Tree, Bayesian Network, and Multilayer Perceptron are used.

2. CLASSIFICATION TECHNIQUES

2.1. J48

It's also known as "advanced C4.5 execution" or "advanced C4.5 form." [5]. The Choice tree is the yield given by J48. A choice tree is a tree structure with a variety of nodes, such as a root node, a moderate node, and a leaf node. [6]. As the name implies, every node in the tree contains a choice, and that decision determines our outcome. The information space of an information set is divided into basically unconnected areas by a choice tree, with each range having a mark, a value, or an action to represent or expand its information emphasis. Part paradigm is used as part of a choice tree to choose which credit is best for parting that partition tree of the preparation information that reaches a specific node. [7].

Flow Chart:

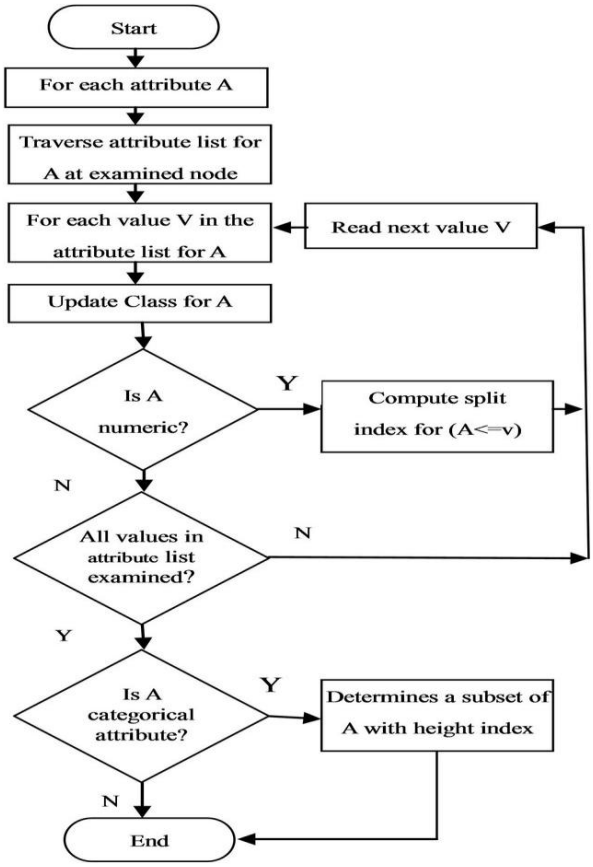

Figure 1: Flow chart of J48

Merits:

1. Decision trees are simple and straightforward
2. Decision trees are simple to convert to a set of production rules
3. Decision trees can classify both categorical and numerical data, but the output attribute must be categorical
4. There are no a priori assumptions about the data's nature.

Demerits:

1. The decision lines are rectilinear.
2. Minor differences in the data can result in extremely different-looking trees.
3. A sub-tree can be copied multiple times.
4. Error-prone due to a large number of classes
5. It's not very accurate at guessing the value of a continuous class attribute's value.

2.2.Multi-Layer Perceptron

Multi-Layer Perceptron can be described as a combination of a neural network and artificial intelligence. A multilayer perceptron (MLP) is a feed forward neural system with one or more layers between the input and output layers. [8]. There are three levels in total: the information layer, the concealed layer, and the yield layer. There could be more than one hidden layer. Every neuron (node) in each layer is linked to every neuron (node) in the layers above and below it. The preparation or testing vectors are linked to the info layer, making it easier for the covered up and yield layers to handle them.

Back Propagation Algorithm:Back propagation, also known as the generalized delta rule, is a type of supervised learning for multi-layer nets. Error data from the output layer is transmitted back to earlier levels, allowing these layers' incoming weights to be modified. In today's neural network applications, it's most commonly employed as a training algorithm. The back propagation method is a complex mathematical tool, yet iterative procedures are used to execute the training equations, making it simple to apply on a computer. [9].

A pair of patterns named input pattern and desired pattern are used during the network's training session. The disparity between the actual and target outputs generates an error signal at the output layer. The values of the weights of the neurons in each layer determine the error signal. This error is minimized, and new weight values are obtained as a result of this process. The learning rate is a factor that affects the speed and precision of the learning process, which is the process of updating the weights.

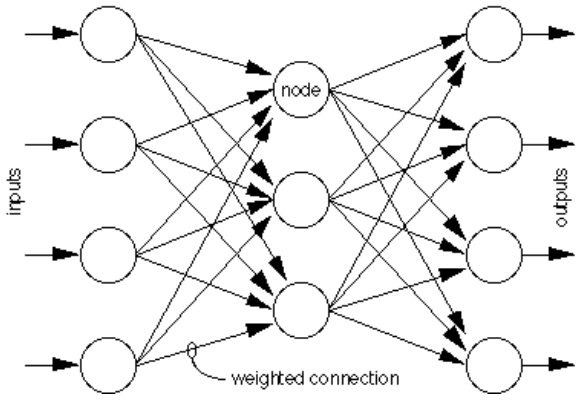

Figure 2: Multilayer perceptron structure

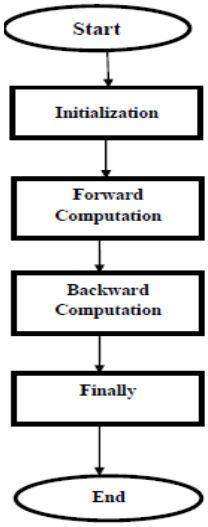

Figure 3: Flow of Multi-Layer Perceptron

Merits:

1. Tolerance for Errors Simple to Recognize
2. In the Main Approach, speech synthesis is used.
3. Monetary applications
4. Recognizing Patterns

Demerits

1. Largely unaffected by outliers (wild points).
2. The Main Drawback is the Slow Process.
3. The model must be stored in a little amount of memory.

2.3 Bayes Net classifier

The Bayes Net classifier is based on Bayes' theorem. In the Bayes Net classifier, the restricted likelihood for each node is determined first, and then a Bayesian System is formed. Whatever the case may be, a Bayesian System is a coordinated non-cyclic chart. The assumption in Bayes Net is that all qualities are obvious and that there are no missing values or esteems that have supplanted everything. In Bayes Net, Hereditary Calculation and K2 are two different types of computations that are used to determine restricted likelihood. The yield of can be imagined as far as the chart in Bayes Net[10].

Flow Chart:

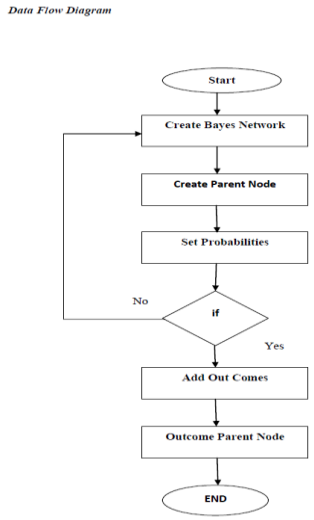

Figure 4: Flow of BayesNet classifier

Merits:

1. Bayesian learning methods interpolate all the way to pure engineering.
2. When working on the "think harder" phase of a solution, this is quite useful.
3. Bayesian learning entails defining a prior and integrating data, two tasks that appear to be uniformly beneficial. (See the section on intuitions)

Demerits:

1. Specifying a previous turns out to be quite difficult.
2. Assume we can precisely describe a prior over each and every air molecule in a room. Even yet, calculating a posterior can be challenging. This problem necessitates the use of computational approximation.
3. The Bayesian research program's "think harder" component is (in some ways) a "Bayesian employment" act.

4. RESULTS AND DISCUSSION

A comparison of classifiers for various datasets is made based on their accuracy and execution time. The number of instances successfully categorized is defined as accuracy. Table 4.1 shows that the Multilayer Perceptron worked well with the Labor dataset and Table 4.2 shows the 2 performance measures for various classification algorithm in Labor Dataset.

Algorithm	Correctly Classified (%)	In correctly Classified (%)	Time Execution (sec)
J48	73.68	26.13	0.02
Multilayer Perceptron	92.98	7.01	0.05
Bayes Net	87.71	12.28	0.02

Table: 4.1 Comparison of correctly and incorrectly classified for various Classification algorithms in Labor Dataset

Algorithm	Precision	Recall	F-Measure
J48	0.748	0.737	0.74
Multilayer perceptron	0.934	0.93	0.931
Bayes net	0.886	0.877	0.879

Table: 4.2 performance measures for various classification algorithm in Labor Dataset

5. CONCLUSION

As a result, we use several accuracy and time measurements to evaluate the performance of J48, Multilayer Perceptron, and Bayes Net algorithms in terms of classification accuracy. Each classification algorithm's accuracy has been measured, and the results are tabulated here. In the labor dataset, it is obvious that the Multilayer Perceptron algorithm outperforms the J48 and Bayes Net algorithms. Our next study will focus on changing the classification method in order to improve order proficiency in less time. To improve performance, a variety of characterization systems will be deployed.

6. REFERENCES

[1].Usama Fayyad, Gregory Piatetsky-Shapiro, Padhraic Smyth, and RamasamyUthurasamy, "Advances in Knowledge Discovery and Data Mining", AAAI Press/ The MIT Press, 1996.

[2]. Michael Berry and Gordon Linoff, "Data Mining Techniques (For Marketing, Sales, and Customer Support), John Wiley & Sons, 1997

[3]. K. Rajeswari, Dr. V. Vaithiyannathan,(2012) "Mining Association Rules Using Hash Table", International Journal of Computer Applications(9132-3320).

[4]. Rajeswari, Dr. V. Vaithiyannathan,(2011), " Heart Disease Diagnosis: An Efficient Decision Support System Based on Fuzzy Logic and Genetic Algorithm", International Journal of Decision Sciences, Risk and Management by Inderscience Publications. ISSN: 1753-7169.

[5]. J. Ross Quinlan, "C4.5: Programs for Machine Learning", Morgan Kaufmann Publishers, 1993

[6].Ozer, Patrick. "Data Mining Algorithms for Classification." (January 2008): 27. Web. 5 May 2013.[16].

[7]. Gholap, jay"Performance tuning of j48 algorithm for prediction of soil fertility." 5.web. 2 May 2013

[8]. Fine, T. L. 1999. Feedforward Neural NetworkMethodology, 3rd ed. New York: Springer-Verlag.

[9]. Rumelhart, David E.; Hinton, Geoffrey E.; Williams, Ronald J. (8 October 1986). "Learning representations by back-propagating errors".

[10]. Friedman, N., D Geiger, M Goldszmidt "Bayesian Networks Classifiers" *Machine Learning,* 29, 131–163 (1997).