International Journal of Advances in
Soft Computing and Intelligent Systems (IJASCIS)
2023, Vol 02, Issue 01, 40-47

Science Transactions © 2023

Original Paper

A COMPARATIVE STUDY OF RULE BASED CLASSIFIER AND DECISION TREE IN MACHINE LEARNING

Onima Tigga[1], Jaya Pal[2], Debjani Mustafi[3]
Department of Comp. Sc. & Engg, Birla Institute of Technology, Mesra, Ranchi
[1]otirkey@bitmesra.ac.in, [2]jayapal@bitmesra.ac.in, [3]debjani.mustafi@bitmesra.ac.in

ABSTRACT

***In** Machine Learning, Decision Tree is an effective method of Classification. In this paper, the two methods Rule based classifier and Decision tree have been used to generate rules and compare the accuracies and precisions. As a result, it is found that decision tree classifier gives the good accuracy. The metrics entropy and information gain are used for finding the best splitting attribute for inducing tree of Iris dataset. Also, confusion matrix is used for finding accuracy of both the methods. Our experimental results show that Decision tree has excellent performance in terms of accuracy and error rates. Also, we find that confidence interval for accuracy of Decision Tree becomes tighter than Rule based classifier on increasing the number of data.*

KEYWORDS

Decision Tree, Rule based Classifier, Entropy, Information Gain.

1. INTRODUCTION

In machine learning, the analysis of classifier is a type of supervised learning as the class label of each training tuple is known. There are various classification tasks such as decision trees, rule-based classification, Bayesian classification etc. Classification is the most popular machine learning technique. The applications of this task include fraud detection, medical diagnosis, loan approval etc. it maps data into predefined classes. [1]. The model based on classification can be represented in various ways such as classification (IF- THEN) rules, neural networks, naïve Bayes ,decision trees classification, support vector machines etc. Unlike the classification model, predicts both numeric data values and class labels. It can also identify distribution trends on the basis of available data. In this paper, we use Decision Tree and Rule based classifier for classifying and predicting Iris data set. As we know, decision trees are easy to understand and have great accuracy. However, when they become very large in size, they possess some problems in interpreting the results. Thus, in such cases, a rule-based classifier is build by retrieving IF-THEN rules from derived trees. For doing this, each path creates one rule which starts from root and ends at leaf node. The IF part of the rule is formed by performing AND operation to each splitting criteria along a given path whereas THEN part would be the leaf node which holds the class prediction [2] .Decision Tree is a derived model for uniting tests in an efficient and cohesive manner where in a numeric characteristic value is compared with a threshold value in every test condition [3]. In classification problems, Decision Tree is useful. In this approach, a tree is built for modelling the classification process. The Decision Tree approach to classification divides the search space into rectangular regions. A tuple is classified on the basis of region into which it falls [5].

Ross Quinlan (1983) developed ID3, a learning algorithm for decision tree. Constructing the decision tree by using a top to down and greedy algorithm through the given sets for testing every attribute at each node in the tree is the basic idea of ID3[12]. For measuring the amount of uncertainty, the entropy is used in a set of a data. The value of Entropy always placed between 0 and 1[7].

Consider probabilities P_1, P_2, P_m where Σ P = 1, The Entropy is defined as

$$H(P_1, P_2 \ldots P_m) = \Sigma P_k \log(1/P_k) \tag{1}$$

Where k=1 to m

And the Gain of a particular split is defined as :

For k=1 to m, $$\text{Gain}(D, S) = H(D) - \Sigma P(D_k) H(D_k) \tag{2}$$

This paper is arranged as follows: Section 2 explains the related work for Rule based classifier and Decision Tree. Section 3 describes the evaluation criteria for accuracy of the Decision Tree and Rule based classifier for Iris dataset and also shows the confidence interval for accuracy of both the methods. Section 4 gives the experimental results and our paper concludes in Section 5.

2. RELATED WORK

Patel B. R., Rana K. K [6] discussed classification task using decision tree. In this paper, finds that suitable property for each node of a derived decision tree divides the sample set included in current node, information theory technique is used which is going to make the composite degree of different types for all generated samples subsets reduced to a minimum. Therefore, using such an information theory technique, the required dividing number of objects classification can be minimized. Jijo B. T., Abdulazeez A. M. [3], discussed how we can measure a dataset's impurity and randomness about entropy and information gain. The entropy value is always placed between 0 and 1. When its value equals 0, it is better but when it equals 1, it is worse. i.e. it's always better when its value is closer to 0. SONG Y. V., Ying L. U. [7], discussed limitations of the decision tree. The main disadvantage is that it can be subject to over fitting and under fitting, when using a small dataset. This problem can limit the generalization ability and robustness of the resultant models. Patel H. H., Prajapati P. [4], applied ID3, CART decision tree algorithm and also C4.5 their performance on the basis of accuracy, time and precision. In this paper, they found CART to be very precise and most accurate among the others. CART takes a lot of time and is the slowest among them. Biao Qin, Y.Xia, S. Prabhakar, Y.TU [9], explained a new rule based algorithm. To classify and predict uncertain datasets, and they also proposed new ways to derive ideal rules out of highly undetermined data, cut down, boosting rules and class prediction for undetermined data. Dutta R. P., Saha S. [10], tried to facilitate for choosing an appropriate classification algorithm without the need for trial-and-error testing of the wide array of available algorithms. They experimented to produce actual evidence of the superiority of one algorithm over another based on different datasets.

3. EVALUATION CRITERIA

3.1 Confusion Matrix

Confusion matrix is used to find the accuracy and error rate of the model as follows:

$$Accuracy = \frac{Number\ of\ correct\ Predictions}{Total\ Number\ of\ Predictions} \tag{3}$$

$$Error\ Rate = \frac{Number\ of\ wrong\ Predictions}{Total\ Number\ of\ Predictions} \quad (4)$$

3.2 Binomial Confidence Interval for Accuracy

A test set has M tuples, the model predicts correct tuples is Y and s is the accuracy. A binomial-experiment is assumed, if the prediction task is taken, then Y is considered as a binomial-distribution with mean Ms and variance Ms (1-s). The accuracy, accuracy = Y/M, has a binomial-distribution with mean s and variance s (1-s)/M. When M is large, the binomial-distribution is treated as a normal distribution [8]. The confidence interval for accuracy based on normal-distribution is as follows:

$$P\left(-Z_{\frac{\alpha}{2}} \leq \frac{accuracy - s}{\sqrt{s(s-1)/M}} \leq Z_{1-\frac{\alpha}{2}}\right) = 1 - \alpha \quad (5)$$

where $Z_{\alpha/2}$ = upper bound and $Z_{1-\alpha/2}$ = lower bound , when confidence level is (1- α) obtained from a standard normal-distribution [8]. It shows $Z_{\alpha/2} = Z_{1-\alpha/2}$ if Z=0, since standard normal-distribution is symmetric.

The confidence interval for s is obtained after rearranging this inequality is as follows:

$$\frac{2 * M * accuracy + Z^2_{\alpha/2} \pm Z_{\alpha/2}\sqrt{Z^2_{\alpha/2} + 4 * M * accuracy - 4 * M * accuracy^2)}}{2(M + Z^2_{\alpha/2}\ )} \quad (6)$$

4. EXPERIMENTAL RESULTS

4.1 Rule Based Classifier

Using 1R Algorithm in Rule based classifier, IF-THEN rules on IRIS dataset are depicted in Table 1.

Table 1. Rules using 1R

Option	Attribute	Rules	Errors
1.	S-Length	(0, 4]→ Setosa	0/0
		(4, 5] → Setosa	1/4
		(5, 6] → Versicolor	3/4
		(6,7] →Versicolor	3/6
		(7,8] →Virginica	0/1
		(8, ∞] → Virginica	0/0

2.	S-Width	(0,2.5] →Versicolor (2.5, 2.8] →Versicolor (2.8, 3.1] →Virginica (3.1, 3.4] →Virginica (3.4, 3.7] →Setosa (3.7, 4.0] →Setosa	0/2 1/1 1/4 4/5 0/2 0/1
3.	P-Length	(0,1.8] →Setosa (1.8, 2.6] →Setosa (2.6, 3.4] →Versicolor (3.4,4.2] →Versicolor (4.2, 5.0] →Versicolor (5.0, 5.8] →Virginica (5.8, ∞]→Virginica	0/5 0/0 0/0 0/2 0/3 0/2 0/3
4.	P-Width	(0,0.5] →Setosa (0.5,1] →Setosa (1,1.5] →Versicolor (1.5,2.0] →Versicolor (2.0,2.5] →Virginica (2.5,∞] →Virginica	0/5 0/1 0/3 2/3 0/3 0/0

Where S-length=Sepal Length, S-width=Sepal Width, P-Length=Petal Length and P-Width = Petal Length.

Thus S-length attribute is chosen to first generate rules with minimum error. Probability of putting a tuple in the Setosa class on the given Attribute-value pair is shown in Table 2.

Table 2. Probability of putting tuple in the classes

Rules	**Setosa**	**Versicolor**	**Virginica**
S-Length ≤ 4	0/0	0/0	0/0
4 < S-Length ≤ 5	¾	1/4	0/4
5 < S-Length ≤ 6	2/4	1/4	1/4
6 < S-Length ≤7	0/6	3/6	3/6
7 <S-Length ≤ 8	0/1	0/1	1/1

8 < S-Length	0/0	0/0	0/0

1R Rule finds that the attribute test 4< S-Length ≤ 5 and 5 < S-Length ≤ 6 best improves the accuracy of our current rule.

If 4 < S-Length ≤ 6 Then class = Setosa.

Every time we add an attribute test to a rule, finally rule should cover more of the Setosa tuple. We have obtained the Rules as follows:

Rule1: If 4 < S-Length ≤ 6 AND 2.8 < S-Width ≤ 4.0 AND P-Length ≤ 1.8 AND Petal_ width ≤ 0.5 Then class = Setosa

Rule2: If 4 < P-Length ≤ 7 AND 2.5 < S-Width ≤ 3.4 AND 2.6 P-Length ≤ 5.0 AND 0.5 P-Width ≤ 2.Then class = Versicolor

Rule3: If 5 < S-Length ≤ 8 AND 2.8 < S-Width ≤ 3.4 AND 0.5 < P-Length AND 1.5 <P-Width ≤ 2 Then class = Virginica

4.2 Decision Tree

Further, we have constructed decision tree for three classes of IRIS dataset as shown in Table 3.

Table 3. Conditions Chosen for classes

Attribute	**Setosa**	**Versicolor**	**Virginica**
S-Length	S-Length ≤ 5.4	5.4 < S-Length ≤ 6.4	6.5≤ S-Length < 6.9
S-Width	S-Width ≤ 2.7	2.7 < S-Width ≤ 3.5	3.6 ≤ P-Width < 3.9
P-Length	P-Length ≤ 2.4	2.4 < P-Length ≤ 4.4	4.5≤ P-Length < 5.4
P-Width	P-Width ≤ 0.7	0.7 < P-Width ≤ 1.7	1.8 ≤ P-Width < 2.2

The comparison between Decision tree & rule based classifier is shown in Table 4.

Table 4. Comparison between classifier models

S-Length	**S-width**	**P-length**	**P-width**	**Calculated Classes using Decision Tree**	**Calculated Classes using Rule based Classifier**	**Actual class**
5.0	3.3	1.4	0.2	Setosa	Setosa	Setosa
5.3	3.7	1.5	0.2	Setosa	Setosa	Setosa
4.6	3.2	1.4	0.2	Setosa	Setosa	Setosa
5.1	3.8	1.6	0.2	Setosa	Setosa	Setosa

4.8	3.0	1.4	0.3	Setosa	Setosa	Setosa
5.1	3.8	1.9	0.4	Setosa	∞	Setosa
5.0	3.5	1.6	0.6	Setosa	Setosa	Setosa
5.7	2.8	4.1	1.3	Versicolor	Versicolor	Versicolor
5.1	2.5	3.0	1.1	Versicolor	Versicolor	Versicolor
6.2	2.9	4.3	1.3	Versicolor	Versicolor	Versicolor
5.7	2.9	4.2	1.3	Versicolor	Versicolor	Versicolor
5.7	3.0	4.2	1.2	Versicolor	Versicolor	Versicolor
5.6	2.7	4.2	1.3	Versicolor	Versicolor	Versicolor
5.0	2.3	3.3	1.0	Versicolor	∞	Versicolor
6.0	2.2	5.0	1.5	Versicolor	∞	Virginica
6.2	3.4	5.4	2.3	Virginica	Virginica	Virginica
6.5	3.0	5.2	2.0	Virginica	Virginica	Virginica
6.3	2.5	5.0	1.9	Versicolor	Versicolor	Virginica
6.7	3.0	5.2	2.3	Virginica	Virginica	Virginica
6.4	2.7	5.3	1.9	Versicolor	Versicolor	Virginica

The Confusion Matrix for both the methods Rule based Classifier and Decision Tree are shown below:

Table 5. Confusion matrix for Decision Tree

classes	**Setosa**	**Versicolor**	**Virginica**
Setosa	7	0	0
Versicolor	0	7	0
Virginica	0	3	3

Table 6. Confusion matrix for Rule based Classifier

classes	**Setosa**	**Versicolor**	**Virginica**
Setosa	6	1	0
Versicolor	0	6	1
Virginica	0	3	3

The Accuracy and error rates are depicted in Table 7:

Table7. Measures Used For Methods

Measures Used	**Rule Based Classifier**	**Decision Tree**
Accuracy	0.75	0.85
Error Rate	0.333	0.15

The binomial distribution of Confidence Interval for Accuracy of both the models Rule based classifier and Decision Tree in Table 8 and 9 are as follows:

Table 8. Confidence Interval for Accuracy of Rule based classifier

N	**20**	**40**	**60**	**80**	**100**
Confidence Interval	0.8881 0.5333	0.9294 0.7093	0.9190 0.7388	0.8792 0.7180	0.8911 0.7445

Table 9. Confidence Interval for Accuracy of Decision Tree

N	**20**	**40**	**60**	**80**	**100**
Confidence Interval	0.9476 0.6396	0.9454 0.7389	0.9534 0.7985	0.9485 0.8149	0.9519 0.8377

It is found that as the number of data N increases, confidence interval for accuracy becomes tighter.

5. Conclusion

On comparison, Decision Tree gives more accurate results than Rule based classifier. Also, it is found that the confidence interval for accuracies of both the methods becomes closer when number of data (N) increases. The confidence interval for accuracy of rule based classifier and decision tree are calculated and have found that the confidence interval for accuracy will become tighter when the number of data N increases,. We have found that when N = 20, the difference between the upper and the lower limit of the confidence interval for accuracy of ruled based classifier is 0.3548 while using decision tree, it is 0.3080. Again, when N = 100 the same difference for ruled based classifier is 0.1466 while using decision tree it is 0.1142.

References

[1] Dunham M. H., Data Mining, Pearson, (2006).

[2] Tigga O., Priyam A., An Application of Predictive Data Mining Techniques for Classifying Eco gene data, International Journal of Engineering Research & Technology, ISSN-0181,Vol.3 Issue 1 (2014).

[3] Jijo B. T. , Abdulazeez A. M. , Classification Based on Decision Tree Algorithm for Machine Learning, Journal of Applied science and Technology Trends, Vol.02, No. 01, pp. 20-28 (2021), ISSN: 2708-0757.

[4] Patel H. H., Prajapati P., Study and Analysis of Decision Tree Based classification Algorithms, International Journal of Computer Science and Engineering, Vol.-6, Issue-10, Oct, (2018).

[5] Devasena C. L., Sumanthi T., GomathiV. V. , Hemalatha M., Effectiveness Evaluation of Rule Based classifiers for the classification of Iris Data set, Bonfirong International Journal of Man Machine Interface, Vol. 1, special Issue, (2011).

[6] Patel B. R., Rana K. K., A Survey on Decision Tree Algorithm for classification, IJEDR, Volume 2, Issue 1 / ISSN: 2321-9939, (2014).

[7] SONG Y. V., Ying L. U., Decision tree methods: applications for classification and prediction, Shanghai Archives of Psychiatry, vol.27, No.2, (2015).

[8] Tan P. N., Steinbach M., Karpatne A., Kumar V., Introduction To Data Mining, Second Edition, Pearson, (2021).

[9] Xia B. Q. Y., Prabhakar S., Yichengtu, A Rule-based classification Algorithm for uncertain Data, IEEE International conference on Data Engineering, (2009).

[10] Dutta R.P., Saha S., Applying rule-based classification techniques to medical databases: an Empirical study, International Journal of Business intelligence and systems Engineering vol. 1, No. 1, (2016).

[11] Iris Data Set, http:// archive.ics.edu/ml/iris

[12] Lakshmi D. B., Arundathi S., Dr.Jagadeesh, Data Mining: A prediction for student's Performance Using Decision Tree ID3 Method.

Authors

Myself Mrs. Onima Tigga received MCA and M. Tech (Computer Science) degree from Birla Institute of Technology, Mesra, Ranchi. I am currently pursuing Ph. D. degree in Technology from Birla Institute of Technology, Mesra, Ranchi. I am associated with Birla Institute of Technology, Mesra, Ranchi as an Assistant Professor. My interests and research area are Data Mining, Machine Learning and Databases.

Dr. Jaya Pal received the M.Sc.(Maths) degree from Ranchi University Ranchi and M.C.A. degree from Birla Institute of Technology, Mesra, Ranchi She received the Ph.D. degree in Technology from Birla Institute of Technology, Mesra, Ranchi in 2017.Currently, she is an Assistant Professor at Birla Institute of Technology, Ranchi, India. Her research interests include Machine Learning, Fuzzy Logic & its Applications, Soft Computing, Software Quality Prediction and Data Mining.

Dr. Debjani Mustafi is affiliated to Birla Institute of Technology, Mesra, India. She is currently working as an Assistant Professor in the Department of Computer Science and Engineering. She has been actively associated with academics and research. She has authored and co-authored multiple peer reviewed journals and conferences. Her research interests include Machine Learning, Text mining, Data analysis and visualization, Evolutionary Computing.

.