PRINT ISSN 1119-8362
Electronic ISSN 2659-1499

Full-text Available Online at
https://www.ajol.info/index.php/jasem
https://www.bioline.org.br/ja

J. Appl. Sci. Environ. Manage.
Vol. 29 (3) 799-805 Mar. 2025

Benchmarking Assessment of Supervised Machine Learning Algorithms of K-Nearest Neighbor, Random Forest, Decision Tree and Its Variants Based On Efficiency and Performance Metrics

***[1]ABDULLAHI, YY; [2]NUR, AS; [3]SALE, A**

**[1]North Eastern University Gombe, Nigeria*
[2]Nile University of Nigeria, Abuja
[3]Binyaminu Usman Polytechnic Hadejia, Jigawa, Nigeria

**Corresponding Author Email: yahaya.yusuf@neu.edu.ng*
**ORCID: https://orcid.org/0009-0000-7350-1269*
**Tel: +2348035987089; +2349065952760*

Co-Author Email: Abdulsalam.nur@nileuniversity.edu.ng; Auwalus6@bupoly.edu.ng

ABSTRACT: Machine learning provides more verbose algorithms capable of accurately predicting, classifying groups as needed. Consequently, the objective of this paper is to assess the benchmarking of Supervised Machine Learning Algorithms of K-Nearest Neighbor, Random Forest, Decision Tree and it variants (ID3, C4.5, C5.0 and CART) based on efficiency and performance metrics using python programming after downloading dataset from Kaggle repository. Dataset to the aforementioned models reveals that, the C4.5 variant of decision tree had the highest prediction accuracy, CART and KNN had the minimal learning and prediction time. If accuracy is the based preference, C4.5 variant of decision tree should be recognized, but when the chief concern is nominal time for training and prediction, then CART and KNN standout.

DOI: https://dx.doi.org/10.4314/jasem.v29i3.15

Cite this Article as: ABDULLAHI, Y. Y; NUR, A. S; SALE, A (2025). Benchmarking Assessment of Supervised Machine Learning Algorithms of K-Nearest Neighbor, Random Forest, Decision Tree and Its Variants Based On Efficiency and Performance Metrics. *J. Appl. Sci. Environ. Manage.* 29 (3) 799-805

Dates: Received: 31 January 2025; Revised: 28 February 2025; Accepted: 16 March 2025; Published: 31 March 2025

Keywords: Decision tree; supervised; CART; confusion matrix; precision; recall; accuracy; learning.

Machine learning is a subset of artificial intelligence that enables systems to learn from data to improve their performance overtime without explicitly/strictly specifying it. It has to do with algorithms and development of statistical models that give computers ability to correctly predict data from a given dataset (Müller and Guido, 2017). Machine learning is a computer program and said to learn from experience E with respect to some class of tasks T and performance measure P, if its performance at tasks in T, as measured by P, improves with experience E (Dutt *et al.*, 2019). Machine can be considered to learn if it is able to gather experience by doing a certain task and improve its performance in doing the similar tasks in the future. In machine learning, data is split into training data and test data. The first split of data, i.e. the initial reserve of data you use to develop your model, provides the training data. After you have successfully developed a model based on the training data and are satisfied with its accuracy, you can then test the model on the remaining data, known as the test data (Theobald, 2017). Decision tree and k-Nearest Neighbor algorithms are adopted by many machine learning practitioners (Dutt *et al.*, 2019). Supervised machine learning is one of the most commonly used and successful types of machine learning (Muller and Guido, 2017). supervised learning is used whenever we want to

**Corresponding Author Email: yahaya.yusuf@neu.edu.ng*
**ORCID: https://orcid.org/0009-0000-7350-1269*
**Tel: +2348035987089; +2349065952760*

predict a certain outcome from a given input, and we have examples of input/output pairs. We build a machine learning model from these input/output pairs, which comprise our training set. Our goal is to make accurate predictions for new, never-before-seen data (Müller and Guido, 2017). As far as tabulated data is concern, decision tree as a supervised machine learning algorithm standout because of its accuracy in prediction (Bechler-Speicher *et al.,* 2024). However, the authors claim that when it comes to graph structured data, decision is inefficient; they proposed TREE-G, a relatively new splitting function that is specialized for graph data. Most of the time, information is stored in a tabular form which is suitable for prediction and classification, the claim (graph structured data) is not always supported by most organizations. Izza *et al.,* (2020) claim that even though decision tree is simple to understand, sometimes it might look uninterpretable especially when the data path of the decision tree is arbitrarily larger than a PI-Explanation, the suggest a new model for computing PI-Explanation of a decision tree. They stated that recent work suggested that interoperability should match with how shallow a decision tree is, but their work also investigate the limits of interoperability of a decision tree. Wan, *et al*., (2021) says that earlier research focus on either prioritizing interoperability over accuracy while others prioritize accuracy over interoperability, for this dilemma, they suggest a new strategy called NBDTs (Neural Backed Decision Tree) what the NBDT does is to replace last network layer with a decision tree, the claim here is that, their proposed model will prioritize accuracy as well as interoperability at the same time. Priorities are mostly based on necessities, so if inclination is to be towards accuracy, accuracy is considered and if it should be toward interoperability, it will be the central focus. Sometimes the more priorities you have, the more resource/energy your model will consume.

Nandini, (2024), presented a comparative study between different machine learning algorithms including SVM, KNN, Decision Tree, Logistic Regression, XGBoost and Gradient Boosting to find the algorithm with the highest level of accuracy and proposed it for early detection of cardiovascular diseases. However, the result of their experiment shows the effectiveness of ensemble methods in providing reliable prediction, but in their paper, there was no presentation of correct and incorrect predictions from their experiment like confusion matrix. Moritz, (2020), compares five different machine learning algorithms in other propose more robust one for efficient document classification, in his documentation, he stated uprightly that neural network achieve high accuracy but takes the highest computational time than the rest. They compare SVM, Suryankanthi, (2020), carryout an incredibly critical comparative study to analyse the performance of GINI index and information gain, he concluded that, both of them give same accuracy when applicable to classification (splitting indeces) model. i.e. there is no difference as to whether GINI index or information gain is prioritized to be implemented for a classification model as both produces same accuracy level. They used a supervised learning algorithm as a decision tree classifier for creating the decision tree, specifically classification and regression tree (CART) model. Based on the statistics and inferences by this author both the two can be used to measure the quality of a split at a particular node. However, combining multiple decision tree, i.e. Random forest can reliably provide accuracy and also eliminate overfitting. Anuja *et al.,* (2013) conducted a comparative study among decision tree classification algorithm including ID3, CERT, C4.5, SPRINT, SLIQ and then apply few out of these to student's record for predicting their performances. They said that even though the size of the dataset is a bit smaller and that the same student record has to reside permanently in the database (the training dataset and the testing dataset), but nevertheless useful prediction can be made out of it. Base on the result of their comparison between SPRINT and SLIQ, the latter produces more accurate results. However, this study comes with a lot of short comings, like same dataset that was used during training has to be maintain for accurate prediction of the student performance. Secondly if size of the data set changes, this model might not be efficient for accurate prediction.

Vasile and Ştefan-Gheorghe, (2014) Had done a research study, they compare two decision tree classification algorithms, namely C4.5 and C5.0 and said that their proposed method predicts more accurately compare to C5.0 even though it produces a very large tree compare to the C5.0. They improve their method by introducing data compression algorithm at dataset testing level. As the entire concept deals with extremely large amount of realistic information, compression algorithm can play a pivotal role in making the whole process more efficient. Rai *et al*., (2016) pointed out the essentiality of intrusion detection systems and also compare different decision tree classification algorithms, they proposed the C4.5 decision tree classification algorithm, and they build the tree using information gain. The general idea behind their study is using a decision tree classification algorithm to improve intrusion detection over a network. However, this C4.5 decision tree classification algorithm produces extremely large tree. They claim that these machine learning algorithms can improve detection of anomalies, suspicious action, unauthorize access and many other cybercrimes but have not stated clearly how the algorithm will be inculcated with the intrusion detection system to improve the whole process. Jijo and Abdulazeez, (2021) discusses different types of decision tree classification algorithms (like the CART, C4.5, CHAIT and QUEST), compare them and draw meaningful conclusions over them, this conclusion is that decision tree performed extremely better with an accuracy of 99.93% when the dataset used for training and testing is from a known repository. Aside these they've reviewed many current related articles like analysis of medical diseases, text classification, patterns, images and the rest. However, based on their review, they concluded that decision tree produces reliably accurate results. Decision tree, KNN, Naïve Bayes and artificial neural Network. Other decision tree variants can provide higher accuracy without overfitting or underfitting if the model is carefully developed.

Consequently, the objective of this paper is to assess the benchmarking of Supervised Machine Learning Algorithms of K-Nearest Neighbor, Random Forest, Decision Tree and its Variants (ID3, C4.5, C5.0 and CART) based on efficiency and performance metrics.

MATERIALS AND METHOD

We have implemented various machine learning algorithms including KNN, Random Forest, decision tree and its variants to measure the performance and as well efficiency metrics, for the performance, we consider: precision, accuracy, recall, F1-score and for efficiency we use training and prediction time of the various algorithms. Our models are main to predict whose loan should be approve and whose own should be rejected. We have use loan approval dataset available on Kaggle.com/repository, the dataset considered for training and testing contains exactly thirteen columns (13) and one thousand two hundred and eighty-one rows (1281). The dataset was split into training and test data, 20 percent training set and 80 percent test set. Python programming was used for the implementation.

RESULT AND DISCUSSION

Attribute Selection Measures: While implementing a Decision tree, the main issue that arises is that how to select the best attribute for the root node and for sub-nodes. So, to solve such problems there is a technique which is called as Attribute selection measure or ASM decision (Müller and Guido, 2017). By this measurement, we can easily select the best attribute for the nodes of the tree. There are two popular techniques for ASM, which are: (1) Information Gain (2) Gini Index

Information Gain: Information gain is the measurement of changes in entropy after the segmentation of a dataset based on an attribute. It calculates how much information a feature provides us about a class. According to the value of information gain, we split the node and build the decision tree (javatpoint, 2025).

(Gao, 2021) It is given mathematically in equation 1.

$$\text{Information gain} = -\sum_{i=1}^{k} P(c_i)\log_2(P(c_i)) \quad (1)$$

Evaluation metrics

Gini Index: Gini index is a measure of impurity or purity used while creating a decision tree in the CART (Classification and Regression Tree) algorithm. An attribute with the low Gini index should be preferred as compared to the high Gini index. It only creates binary splits, and the CART algorithm uses the Gini index to create binary splits (Javatpoint, 2025). It is given in equation 2:

$$\text{Gini} = 1 - \sum_{i=1}^{n} P_i^2 \quad (2)$$

K-Nearest Neighbors: The k-NN algorithm is arguably the simplest machine learning algorithm. Building the model consists only of storing the training dataset. To make a prediction for a new data point, the algorithm finds the closest data points in the training dataset—its "nearest neighbors (Muller and Guido, 2017). One of the strengths of *k*-NN is that the model is very easy to understand, and often gives reasonable performance without a lot of adjustments. To find the distance between any given points, it is evaluated using equation 3.

$$D = \sqrt{((x2 - x1)^2 + (y2 - y1)^2)} \quad (3)$$

This section narrates performance metrics and their statistics as well as details about them. We have considered two evaluation metrics namely performance (which focuses on accuracy, precision, recall and F1-score) and efficiency (which entails things like training time, prediction time and the rest)

Performance Metrics: Accuracy: measure the proportion of correct prediction (both positive and negative) out of the total number of predictions. Given as:

$$\text{Accuracy} = \frac{\text{True Positive} + \text{True Negative}}{\text{Total Samples}} \quad (4)$$

Precision: measure the proportion of true positive predictions out of all positive predictions made by the model.

$$\text{Precision} = \frac{\text{True Positive}}{\text{True Positive} + \text{False Positive}} \quad (5)$$

Recall (Sensitivity or true positive rate): measure the proportion of true positive prediction out of all actual positive cases.

$$\text{Recall} = \frac{\text{True Positives}}{\text{True Positives} + \text{False Negative}} \quad (6)$$

F1-score**:** measure the harmonic mean of precision and recall, providing a balance between the two

$$\text{F1} - \text{score} = 2 * \frac{\text{Precision} * \text{Recall}}{\text{Precision} + \text{Recall}} \quad (7)$$

Table 1: C4.5 Model Performance

Metrics	Accuracy	Precision	Recall	F1-score
0	98%	0.98	0.99	0.98
1		0.98	0.96	0.97

The C4.5 variant of decision tree had correctly predicted 1250 positive classes and had incorrectly predicted 31 negative classes with about 98% accuracy, in the 1281 test dataset considered. And it is shown in figure 1. The C5.0 also is a variant of a decision tree which achieved an accuracy of 96% as shown in table 2. The confusion matrix below shows the details of how the 96% of accuracy is attained from the given dataset, it has correctly predicted 1227 positive classes and had incorrectly predicted 54 negative classes, these is detailed in figure 2 using head map.

The CART is also a variant of decision tree algorithm and had achieve the following metrics. The confusion metrics below details how the 97% accuracy was achieved, the model correctly predicted 1248 positive classes and incorrectly predicted 33 negative classes from a total of 1281 rows of a loan approval dataset. It is shown in figure 3. The table 4 shows the performance metrics statistics of Random Forest model. The chart in figure 4 details the accuracy of Random Forest model achieving 97%

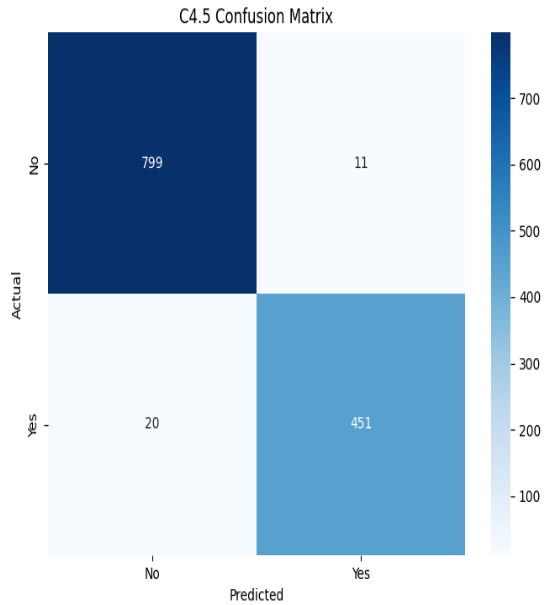

Fig 1: C44.5 Confusion Matrix

Table 2:C5.0 Model Performance

Metrics	Accuracy	Precision	Recall	F1-score
0	96%	0.97	0.97	0.97
1		0.94	0.94	0.94

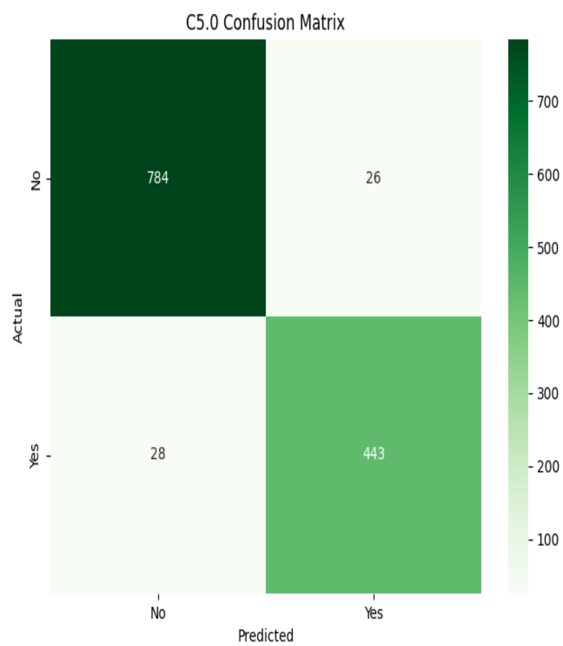

Fig 2 : C5.0 Confusion Matrix

Table 3: CART Model Performance

Metrics	Accuracy	Precision	Recall	F1-score
0 (rejected)	97%	0.97	0.99	0.98
1 (approve)		0.97	0.96	0.96

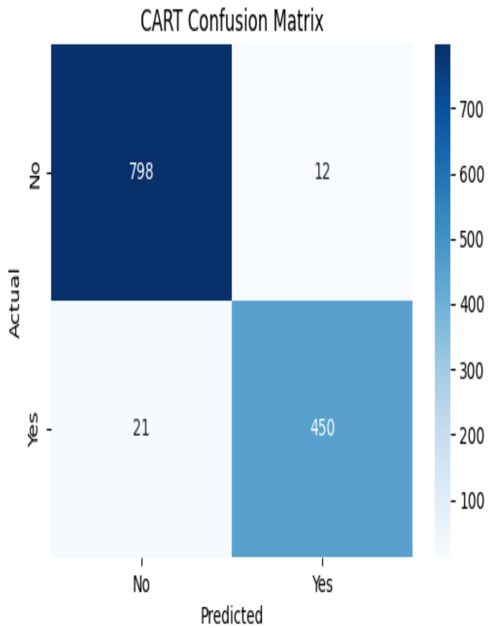

Fig. 3 : The CART Confusion Matrix

Table 4 : Random Forest Performance

Metrics	Accuracy	Precision	Recall	F1-score
0	97%	0.97	0.99	0.98
1		0.98	0.95	0.97

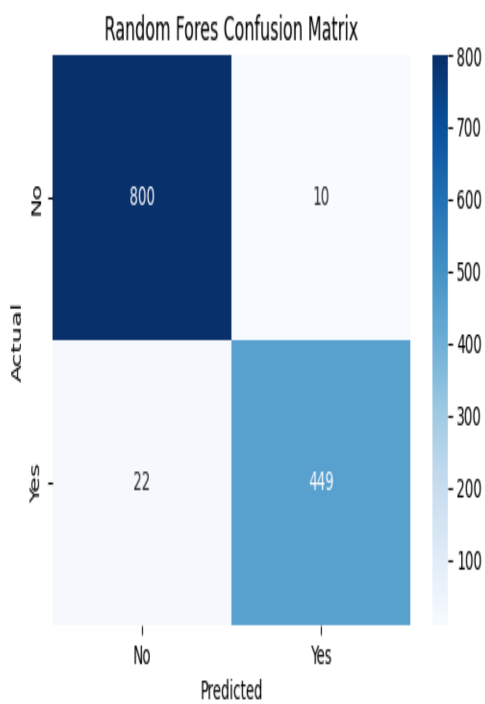

Fig. 4 : Random Forest Confusion Matrix

The model based on the K-Nearest Neighbor has the following values in the metrics considered.

Table 5 : K-Nearest Neighbor Model Performance

Metrics	Accuracy	Precision	Recall	F1-score
0	87%	0.91	0.88	0.90
1		0.81	0.86	0.83

The detail is graphically shown in fig. 5:

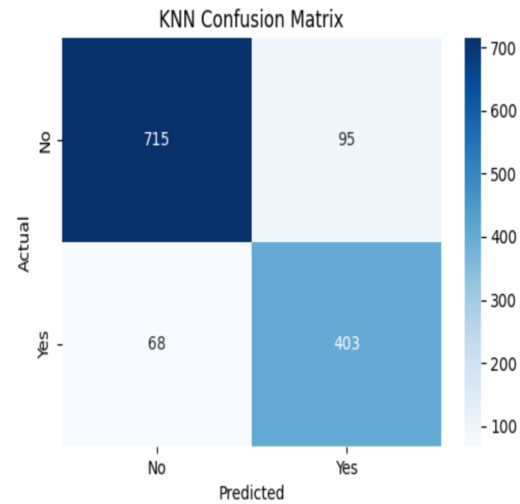

Fig. 5: K Nearest Neighbor Confusion Matrix

The last machine learning models we've considered for our experiment is the ID3 variant of a decision tree and had produced the values shown in table 6.

Table 6: ID3 Model Performance

Metrics	Accuracy	Precision	Recall	F1-score
0	63%	0.63	1.00	0.77
1		0.00	0.00	0.00

Confusion matrix: Also detail of *table 6* is depicted in fig. 6.

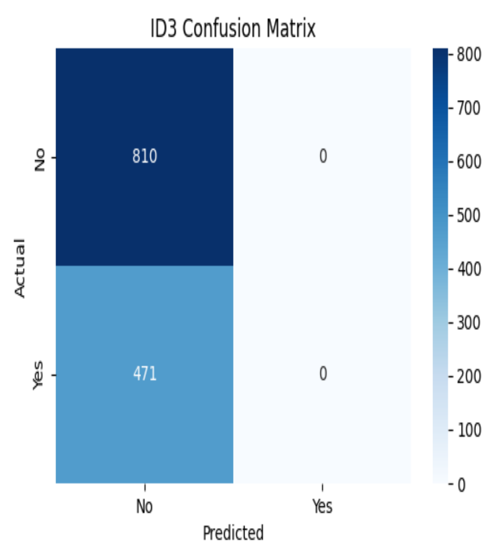

Fig. 6: ID3 Confusion Matrix

Comparing the accuracy: The accuracy determines how reliable model is, for the six (6) machine learning algorithms we've sampled, *table 7* presented the algorithm, metrics and their corresponding values.

Table 7 : Accuracy Comparison of the Models

S/N	Algorithms	Accuracy
1.	ID3	63%
2.	C4.5	98%
3.	C5.0	96%
4.	CART	97%
5.	RANDOM FOREST	97%
6.	KNN	87%

Figure 7 correlate the accuracy of the algorithms considered. Based on the result, C4.5 variant of decision tree achieved higher accuracy of 98% followed by CART also a variant decision tree and Random Forest, then C5.0, KNN and lastly ID3.

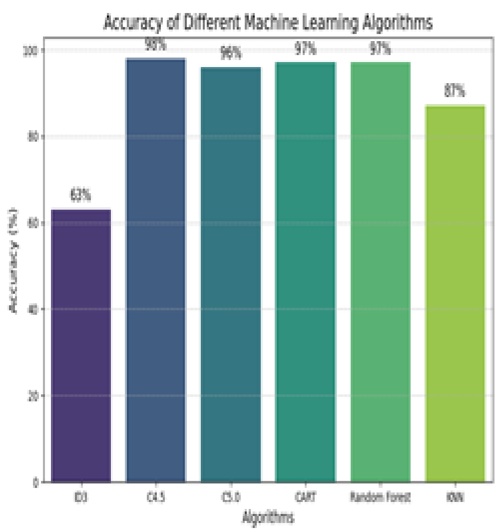

Fig. 7: Models Accuracy Comparison

Efficiency Metrics: This measure how long the model takes to learn and as well predict if new dataset is supply to it.

Table 8 : Training and Prediction time of the Models

S/N	Metrics	Training Time	Prediction Time
1.	ID3	1.0014s	0.5007s
2.	C4.5	1.2003s	0.6006s
3.	C5.0	0.8005s	0.4007s
4.	CART	0.1559s	0.0260s
5.	RANDOM FOREST	0.6376s	0.0260s
6.	KNN	0.1030s	0.2578s

Based on the result of our experiment on efficiency metrics, presented in *fig. 8*, CART and KNN takes less training and prediction time, they have similar prediction time, but KNN takes less time to learn (training time).

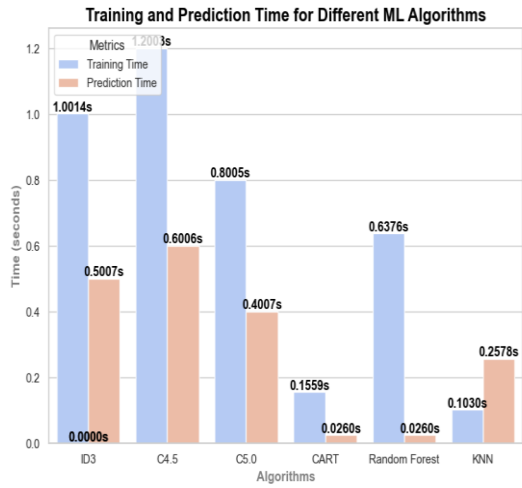

Fig. 8: Learning and Prediction Time in second

Conclusion: Considering accuracy as a major performance metrics, C4.5 algorithm had the highest accuracy, CART and KNN require limited training/learning time. And so, for this reason, we recommend C4.5 or multiples of it for better accuracy in classification, while K-Nearest Neighbor and should be consider for a limited/shorter learning time. However, other machine learning algorithms that have not been covered, might offer exceptional performance using metrics other than the one we use. For a future purpose, we are going to choose, three supervised machine learning, three unsupervised and three reinforcement machine learning algorithms for our experiments. C4.5 algorithm can ensure accuracy in classification problem if appropriately implemented.

Declaration of conflict of interest: The authors declare no conflict of interest.

Data Availability: Data are available upon request from the first author or the Kaggle repository.

REFERENCES

Anuja, P; Abhijeeta, RG; Anju, R; Saurabh, S. (2013). Comparative Analysis of Decision Tree Classification Algorithms. *Inter. J. Cur. Engineer. Technol.* 3 (2): 334-337

Bechler-Speicher, M; Globerson, A; Gilad-Bachrach, R (2024). TREE-G: Decision Trees Contesting Graph Neural Networks. *The Thirty-Eighth AAAI Conf. on Artif. Intel. (AAAI-24).* 11032-11042

Dutt, S; Chandramouli, S; Das, AK (2019). *Machine Learning.* Pearson. pp 33.

Gao, A. (2021). Decision Trees. pp 5.

Izza, Y; Ignatiev, A; Marques-Silva, J (2020). On Explaining Decision Trees. pp 23.

Javatpoint. (2025, January 17). machine-learning-decision-tree-classification-algorithm. Retrieved January 17, 2025, from https://www.javatpoint.com/machine-learning-decision-tree-classification-algorithm

Jijo, BT; Abdulazeez, AM (2021). Classification Based on Decision Tree Algorithm for Machine Learning. *Jour. of Appl. Sci. and Tech. Tre., Vol. II*, pp 20-28.

Nandini (2024). Comparative Study of Machine Learning Algorithms in Detecting Cardiovascular Diseases. pp 19.

Moritz, H (2020). A comparative study of machine learning algorithms for Document Classification. UPPSALA UNIVERSITET. pp 14.

Müller, AC; Guido, S (2017). Introduction to Machine Learning With Python. United States of America: O'Reilly Media. pp 116.

Rai, K; Devi, MS; Guleria, A (2016). Decision Tree Based Algorithm for Intrusion Detection. *Int. J. Adv. Net. App.,* (04), 2828-2834.

Suryakanthi. (2020). Evaluating the Impact of GINI Index and Information Gain on Classification using Decision Tree Classifier Algorithm. *Int. J Adv. Comp. Sci. App*. 11(2).

Theobald, O (2017). Machine Learning for Absolute beginners. pp 211.

Vasile, N; Ştefan-Gheorghe. (2014). Fast Decision Tree Algorithm. *Adv. in Elec. and Comp. Eng.* pp 1582-7445