ISSN: 04532198
Volume 62, Issue 08, September, 2020

Attribute Related Methods for Improvement of ID3 Algorithm in Classification of Data: A Review

Nur Farahaina Idris[1], Mohd Arfian Ismail[2]

Faculty of Computing, College of Computing and Applied Sciences, Universiti Malaysia Pahang, Malaysia[1,2.]

Abstract— Decision tree is an important method in data mining to solve the classification problems. There are several learning algorithms to implement the decision tree but the most commonly-used is ID3 algorithm. Nevertheless, there are some limitations in ID3 algorithm that can affect the performance in the classification of data. The use of information gain in the ID3 algorithm as the attribute selection criteria is not to assess the relationship between classification and the dataset's attributes. The objective of the study being conducted is to implement the attribute related methods to solve the shortcomings of the ID3 algorithm like the tendency to select attributes with many values and also improve the performance of ID3 algorithm. The techniques of attribute related methods studied in this paper were mutual information, association function and attribute weighted. All the techniques assist the decision tree to find the most optimal attributes in each generation of the tree. Results of the reviewed techniques show that attribute selection methods capable to resolve the limitations in ID3 algorithm and increase the performance of the method. All of the reviewed techniques have their advantages and disadvantages and useful to solve the classification problems. Implementation of the techniques with ID3 algorithm is being discussed thoroughly.

Keywords— ID3 Algorithm, Improvement, Attribute, Classification, Decision tree.

1. Introduction

Data mining is a significant concept established by computer scientists to lead a secure and relevant classification and deduction of data using important information gathers from various informative groups [1]. It can be defined as a process to gain information from data that has an adequate quality to generate decisions itself in the subsequent processes using algorithms, database technologies and artificial intelligence [1]. Data mining, in general, comprise of 5 vital components [2]. Firstly, the extraction and transformation of data into the data system. Secondly, storing and managing all the collected data in the multi-dimensional database system. Then, allow the information technology professionals to access data. Fourthly, the analysis of the data made using the software applications. Lastly, presenting the respective data in a more interpretable manner for example using graphs or in tables formats. In current times, data mining had been implemented in many areas of expertise likes healthcare, finance, education and real estate [3]. The application of data mining in various areas of expertise including the industries is to reduce expenditures, improve the existing research, boosting the sales and many more [4]. There are many effective and important techniques in data mining and one of them is classification [5]. Classification can be described as an action to assign a collection of data to the corresponding targeted classes [6]. An objective for the classification of data is to precisely forecasting the respective class or category in each case of the dataset. For the implementation of data classification, the applied training set would have the attributes that are already associated with stated class labels [4]. After that, the selected classification algorithm would learn from the training set and started to build the model. The principle of a classification process in Figure 1 based on [7] is the vital process to perform the classification of data. Every attribute in dataset utilize by the classification algorithms have the constant distinctive set of features whether the features are in continuous, categorical or binary forms [3]. The learning classification methods can be classified under three prominent category which are supervised learning, unsupervised learning and reinforced learning [2]. The focus of this study is leading more to supervised learning. If the following attributes are included with labels and insinuated with veracious output, thus the learning is known as supervised learning [8]. The supervised learning consists of classification and regression task [9]. The more commonly used supervised learning methods for data classification in data mining are decision tree, artificial neural network and support vector machine [7], [9]. In this paper, the data classification method that would be study is decision tree because decision tree is among the most effective techniques in the classification that has a broad prospect of application and great theoretical value [10]. Decision tree (DT) is an important method in data mining to solve the classification problems. It is an effective and efficient method as it is more interpretable, scalable and also can be expressed in both graphical and text forms [11]. The most well-known learning algorithms for DT are ID3, C4.5 and CART algorithm. The learning algorithm reviewed in this paper is ID3 algorithm as it is the most primarily use DT learning algorithm for now [12]–[14].

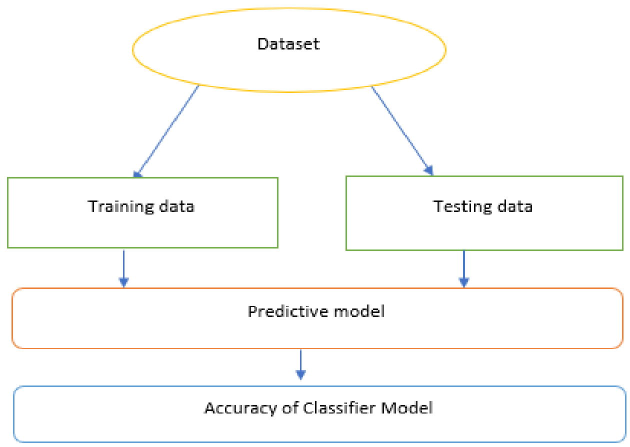

Figure 1. Principle of classification process.

2. Research Method

2.1 *ID3 Algorithm*

There are several learning algorithms to implement DT but the most commonly use is ID3 algorithm [15]. ID3 algorithm is a simple learning algorithm for DT that was firstly proposed by Quinlan in 1986 [16], [17]. The general concept of ID3 algorithm is that it works based on recursive partitioning which the collected data would be split to become subsets and the following subsets would be the partitions that represent the decision tree [11], [18]. The implementation of ID3 algorithm required the computation of Shannon’s information entropy and information gain as the attribute selection criteria [18]. The calculation of information entropy can be expressed in Equation 1 where the S represents the sets that separated to multiple classes while p_i stands for the proportion of the data belongs to the class. The computation of information gain can be conveyed in Equation 2 where A represents the current attribute, S stand for the sets that separated in to multiple classes, n is the potential values for the respective attribute and S_j be the subsets of S which holds the same credits as attribute A.

$$Entr(S) = -\sum_{i=1}^{k} p_i \log_2(p_i) \tag{1}$$

$$Gain(A,S) = Entr(S) - \sum_{j=1}^{n} \frac{|S_j|}{|S|} Entr(S_j) \tag{2}$$

The fundamental aspect of ID3 algorithm is the selection of the attributes from root to nodes of the tree by applying the highest value of information gain as the attribute selection criteria in each step until certain conditions achieve [16]. The action of branching the tree would occur recursively until all the attributes being classified or has the same category. ID3 algorithm form from the collected datasets that have more than one class attribute [18]. The characteristic of the algorithm is it will establish rules for the class prediction of the case in the dataset and at the same time, pinpoint the attributes of the respective class which are distinguished

ISSN: 04532198
Volume 62, Issue 08, September, 2020

from the other classes [15]. The significant aspect of the algorithm is it would reduce the tree size using the quality measure and logical reasoning.

In a study performed by [19], they apply Connect-4 dataset from UCI repository that has over 67557 instances to identify the performance of the ID3 algorithm. The used datasets randomly split into half to become a training and testing sets with the count of data in both sets are balanced by the count of classes. The results have shown that the average accuracy of classic ID3 algorithm is over 72.95% and the average computation time of training data is 3.0903 seconds after being conduct for 1000 times. Based on [10], shows that ID3 algorithm performs better compared to other methods like Bayes classifier and K-means algorithm in terms of proportion in classification errors and number of classification errors when applied with Fisher Iris dataset. Then, the research performed by [20], conveys that ID3 algorithm can achieve the accuracy of 99.66% when applying with Chess dataset and performs even better than ID3-A* search when applying with letter dataset with an accuracy of 78.49% in comparison to 77.64%. The performance of ID3 algorithm in classification better compare to other methods as it produces more accurate result with the shorter time taken [3]. Nevertheless, there are some limitations in ID3 algorithm that can affect the performance in the classification of data. The limitation that will be highlighted in this study is the likeliness of the algorithm to select attributes based on higher number of occurrences rather than the significance of the attributes to the class attribute which cause not necessarily best attribute chosen [15]. The use of information gain in ID3 algorithm as the attribute selection criteria also not assess the relationship between classification and the dataset's attributes [21]. Hence, the study is being conducted to focus on the attribute related methods for the improvement of ID3 algorithm.

2.2 *Attribute Related Methods*

There are several attribute related methods that can be implemented into the algorithm in order to attain improvement in terms of accuracy, computation time and depth of the tree. The attribute related method is being implemented to compute the significance of every attributes data in the dataset [15], [21]. The execution of these methods would assist to lessen the limitation of ID3 algorithm of having only information gain as the attribute selection criteria [19]. Thus, the dependency on attributes with more values would be reduced [13]. Aside from that, the use of the attribute related methods also will compute the relation and co-dependency between the attributes and the classification [14], [21]. The attribute related methods are proposed by several researchers based on mathematical theory concept like the probability theory, information theory, functions and many more [22]. There are three attribute related methods that will be explored in this study which are mutual information, association function and attribute weighted. The features and characteristics of the methods will be analysed to determine the pros and cons of each method.

2.2.1 *Mutual Information (MI)*

Mutual information (MI) can be defined as a value that assesses the relationship between the two random variables in a sample dataset at the same time [23]. MI utilization as the attribute selection criteria is inspired based on the information theory [24]. The special characteristic of information theory is it capable to select the transection feature subsets from the collected dataset and can straight away extricate the important required data with an exception of rare conditions [25]. MI is being proposed for DT as most of the learning algorithms including ID3 algorithm only take regards in a correlation between the conditional attributes and decision attributes like the information entropy. Nonetheless, this can cause a problem like repetition in selecting same precursory attributes and generates complex DT structures with subpar accuracy results [24]. The main purpose of implementing MI is to compute the mutual inclusion relation for all the attributes so the learning algorithms can select the feature subsets with the least redundancy. As the MI be applied to depict the relationship between the random variables, the relationship between the attributes and classes would be able to determine using this method. For example, an attribute of the dataset is equivalently distributed in all classes, thus the MI value for the class will be zero signify that the respective attribute has a poor relationship with the class [26]. The formula of mutual information is shown in Equation 3 where $p(z_i)$ indicates that the probability of attribute z is *i*, while x represents the category in the dataset, and $p(x_j)$ shows the probability that the category is *j*. The formula of average mutual information is expressed in Equation 4 where $p(z_i, x_j)$ indicating the joint probability of attribute z and class x when z is *i* and x is *j*. MI has large randomness so the computation on the distribution between two associate random variables is required. Thus, the computation of average mutual information is significance for decision tree implementation.

$$M(z,x) = \log \frac{p(z_i, x_j)}{p(z_i)p(x_j)} \tag{3}$$

$$M(z,x) = \sum_{i,j} p(z_i, x_j) \log \frac{p(z_i, x_j)}{p(z_i)p(x_j)} \tag{4}$$

As the value of mutual information higher, the association between the respective attribute and the class is more prominent. The deficiency of MI is it gravitated to select multi-valued attributes thus some modification had been applied to reformulate MI. Below show the formula of reformulated MI where *w* represents the number of attributes. All the formulas listed for MI computation is based on [26]. The replacement of information gain with mutual information as the attribute selecting criteria improves the accuracy results and is more productive in terms of performance in comparison to classic ID3 algorithm [27]. The study performs by [26], also shows that MIDT can improve the classification accuracy over 8.3% in comparison to classic ID3 algorithm when implementing with Contact lenses dataset meanwhile with Solar-flare-2 dataset, MIDT can achieve the accuracy of 99.1557% in comparison to classic ID3 algorithm that achieves the accuracy of 98.6867%. Based on the study by [28], the joint mutual information (MI) can achieve the classification accuracy of 88.97% using 8 features with Parkinson dataset which higher compare to using information gain with only 87%. MI also can achieve better accuracy compared to normalized joint mutual information maximization (NJMIM) with staggering 82.92% using 4 features when applied with credit approval dataset. Both of these datasets are from UCI repository. The study conducted by [27], expressed that MI can generate better accuracy and performance compared to classic ID3 algorithm.

$$M(z,x) = \frac{1}{w} \sum_{i,j} p(z_i, x_j) \log \frac{p(z_i, x_j)}{p(z_i)p(x_j)} \tag{5}$$

2.2.2 *Association Function (AF)*

Association function (AF) derived from Correlation Function that highlight the relationship of the random variables [15]. One of the advantages of AF implementation is to overcome the limitation of classic ID3 algorithm that tends to select attributes with many values and at the same time depict the relationship between the features and the respective class attributes [29]. In order to implement AF, we need to compute the AF for each attribute then utilized the values to calculate the normalized gain. The normalized gain value would combine with the old information gain to produce new gain that would be used for attribute selection criteria [30]. There are several attributes in the dataset *D* and let *C* be any class attribute of dataset *D*. Thus, the relation degree function between *S* which is one of the attributes and *C* can be seen in Equation 6. In Equation 6, the representation is listed where n is the number of types of attributes in *S*. Both x_{i1} and x_{i2} depict as two types of cases where the attribute *S* of dataset *D* represents the ith value while category *C* represents as jth value (j=1,2) [14]. The computation of AF is performed on every attribute on dataset *D*. Then, we required to compute the normalized relation degree function value which also stands for normalized gain. The equation of normalized gain for every attribute is in Equation 7 where *m* is the total count of attributes in the dataset.

$$AF(S) = \frac{\sum_{i=1}^{n} |x_{i1} - x_{i2}|}{n} \tag{6}$$

ISSN: 04532198
Volume 62, Issue 08, September, 2020

$$V(k) = \frac{AF(k)}{AF(1) + AF(2) + ... + AF(m)} \tag{7}$$

After that, the combination of the initial (old) information gain with a normalized gain is depicted in Equation 8 where *Gain'(A)* represent as the new gain. The old gain multiplies with normalized gain to get new gain which would be implemented as a new attribute selection criterion for ID3 algorithm. Similar to the classic ID3 algorithm, the largest value would be selected as the node of the tree. This method produces new attribute splitting criteria for ID3 algorithm [15].

$$Gain'(A) = (I(s_1, s_2, ..., s_m) - E(A)) * V(A) \tag{8}$$

The research conducted by [14] shows that AF managed to solve the limitations of classic ID3 algorithm which frequently select the attributes that have a greater number of types as the generated decision tree selected age which has 3 attribute values as the root compared to classic ID3 algorithm that selected colour-cloth that has 4 attribute values when applied with Customer dataset. Based on [29], the application of AF produces more reasonable tree as the DT selected status that has 2 attribute values as the root instead of classic ID3 algorithm that selected age that has 3 attribute values as the root when implementing with Mutual Fund Application dataset. The application of AF decreases the importance of attributes with many values and increase the importance of attributes with fewer values to resolve the shortcomings of the greedy algorithm and use the relation degree values to reflect the importance of the attributes [14]. Then, the study performs by [31], that apply AF also shows that the improved DT has higher accuracy compare to classic ID3 algorithm with 97.5% in comparison to 92.5% when using Student Dropout dataset. Aside from that AF also generate more productive rules for the decision tree as it produces more optimal DT [29].

2.2.3 *Attribute Weighted*

The improved ID3 algorithm that implements attribute weighted emphasizing the attributes that have a higher degree of importance and reduce the shortcoming of the ID3 algorithm that tends to select attributes with more values as the node. The new attribute selection criteria introducing the importance of attribute were $q(0 < q < 1)$ [16]. The size or value regulated by the policymakers themselves based on their experience, knowledge or the situation in the field [32]. For a decision tree, the size refers to the rules and selection factors. Both of them give impact to the decision tree as the value change for every node of the tree development. The values of $q(A)$ can be determined in Equation 9 which p_i is the probability of attribute A belong to class C_I, leading to $0 \le q \le 1$. The formula of ID3 algorithm entropy being alter based on the attribute weighted method as shown in equation 10. The *AIE* in Equation 10 stands for attribute important entropy. Then, the information gain of this method depicted as in Equation 11.

$$0 \le q = q(A) \le \min(p_{1,}, p_2 ... p_m) \tag{9}$$

$$AIE(A) = \sum_{i=1}^{N} \left(\frac{S1i + S2i + ...Sni}{S} + q(A)\right) I(S1i + S2i + ...Sni) \tag{10}$$

$$Gain'(A) = I(S1i + S2i + ...Sni) - AIE(A) \tag{11}$$

Thus, the attribute selection criteria for this method use Gain'(A) like in equation 11. The attributes with the highest value of Gain'(A) will be chosen as the splitting attribute of the tree nodes. The analysis on the results of experiment conduct by [16], [32], stated that the improved ID3 algorithm has better classification accuracy and has more reasonable and effective classification rules. In the classic ID3 algorithm, all the attribute that have a higher number of attribute values listed first whereas, in improved ID3 algorithm, the attributes with more values will weaken their importance in feature selection [14], [16], [21]. Based on the study performs [16], that apply Affecting Summer Weather Comfort dataset shows that the improved DT selected humidity that has 2 attribute values as root compare to classic ID3 algorithm that selected dressing index that has 3 attribute values. Then, the study made by [32], which utilized a real-life dataset of software firm shows that the improved ID3 algorithm selected productivity that has 2 attribute values as the root compared to classic ID3 algorithm that selected technical skills that have 3 attribute values as the root node. Thus, the drawback of ID3 algorithm that inclined to choose any attributes with more values is solved. The research conducted by [21], shows that the classification accuracy of improved ID3 algorithm higher compares to classic ID3 algorithm when implement with Nursery dataset with an accuracy of 96% in comparison to 89 % and the number of leaf nodes also lesser with 359 nodes in comparison to 372 nodes. While when they using the Adult dataset, the accuracy of the improved ID3 algorithm is 82 % with 375 nodes compare to classic ID3 algorithm which only 74% with 386 nodes. Both of the datasets are from UCI. The proposed improved ID3 algorithm also generates results that comply with the policymakers as they can regulate the values themselves [32].

3. Comparative Analysis Between Attribute Related Methods

All the studied attribute related methods have their pros and cons. There are many advantages and disadvantages of implementing the improved ID3 algorithm whether in terms of classification accuracy, computation speed, time complexity and the interpretability of the model based on the proposed methods. Each of the studied attribute related methods have different characteristics among one another. Table 1 shows the comparative analysis between the attribute related methods.

Table 1. Comparative analysis between the attribute related methods.

Method	Advantages	Disadvantages
Mutual Information	-Select the feature subsets with the least redundancy [26], [28]. -Increase classification accuracy [26]–[28].	-Not suitable for continuous-valued attributes [14], [26]. -More complex to understand compares to information gain [28].
Association Function	-Generate more productive rules [29]. -Overcomes shortcoming of tends to select attributes with many values [14], [15], [29]. -Increase classification accuracy [33].	-Increase the computational complexity and runtime [29].
Attribute Weighted	-Overcomes shortcoming of tends to select attributes with many values [16]. -Policy makers can regulate the values themselves [32]. -Increase classification accuracy [21]. -Lesser number of leaf node [21].	-Requires expertise and vast knowledge to regulate the values and weightage [32].

4. Conclusion

This paper presents the study of improved ID3 algorithm using the attribute selection methods. The decision tree learning reviewed in this paper is ID3 algorithm because it was widely used to solve the classification problems. This paper discussed three techniques in attribute selected methods which are MI, AF and attribute weighted in details which covered on the formulas, characteristics and mechanisms of the methods. Based on the comparative analysis of the attribute selection methods that had been conducted, it is intricate to select and determine the best techniques as all the methods have their own and advantage and disadvantage. MI and attribute weighted can increase the classification accuracy. Nonetheless, MI is more complex to understand and unsuitable for continuous data while attribute weighted required experience and vast knowledge regarding the method and regulation. Then, AF is overcoming the shortcoming of the ID3 algorithm, can increase classification accuracy and generate more productive rules but increase the computational complexity and runtime. The selection of the technique is based on the problems and features selected because no method works best on every single problem [34].

5. Acknowledgement

This study was supported by Fundamental Research Grant Scheme (FRGS) with Vot No. FRGS/1/2018/ICT02/UMP/02/2: RDU190113 from Ministry of Higher Education (MOHE) and managed by Universiti Malaysia Pahang.

6. References

[1] H. I. Bülbül and Ö. Ünsal, "Comparison of classification techniques used in machine learning as applied on vocational guidance data," in *Proceedings - 10th International Conference on Machine Learning and Applications, ICMLA 2011*, 2011, vol. 2, pp. 298–301.

[2] B. N. Patel, S. G. Prajapati, and K. I. Lakhtaria, "Efficient Classification of Data Using Decision Tree," *Bonfring International Journal of Data Mining*, vol. 2, no. 1, pp. 06–12, 2012.

[3] R. Chai and M. Wang, "A more efficient classification scheme for ID3," in *2010 2nd International Conference on Computer Engineering and Technology*, 2010, vol. 1, pp. 329–332.

[4] I. Gazalba, N. Gayatri, and I. Reza, "Comparative Analysis of K-Nearest Neighbor and Modified K-Nearest Neighbor Algorithm for Data Classification," pp. 294–298, 2020.

[5] D. T. Larose, *Discovering Knowledge in Data: An Introduction to Data Mining*. John Wiley & Sons, Inc, 2005.

[6] S. Taneja, B. Suri, S. Gupta, and H. Narwal, "A Fuzzy Logic Based Approach for Data Classification," *Data Engineering and Intelligent Computing, Advances in Intelligent Systems and Computing*, pp. 605–616, 2018.

[7] M. Sharma, G. Singh, and R. Singh, "Stark Assessment of Lifestyle Based Human Disorders Using Data Mining Based Learning Techniques," *IRBM*, vol. 38, no. 6, pp. 305–324, 2018.

[8] S. B. Kotsiantis, "Supervised Machine Learning : A Review of Classification Techniques," vol. 31, pp. 249–268, 2007.

[9] D. M. Abdulqader, A. M. Abdulazeez, and D. Q. Zeebaree, "Machine Learning Supervised Algorithms of Gene Selection : A Review," vol. 62, no. 03, pp. 233–244, 2020.

[10] H. Zhang and R. Zhou, "The Analysis and Optimization of Decision Tree Based on ID3 Algorithm," in *The 9th International Conference on Modelling, Identification and Control*, 2017, no. ICMIC, pp. 924–928.

[11] S. B. Begenova and T. V. Avdeenko, "Building of fuzzy decision trees using ID3 algorithm," *Journal of Physics: Conference Series*, vol. 1015, no. 2, 2018.

[12] Q. Liu and Y. Wang, "Improved ID3 algorithm using ontology in computer forensics," in *ICCASM 2010 - 2010 International Conference on Computer Application and System Modeling, Proceedings*, 2010, vol. 11, no. Iccasm, pp. 494–497.

[13] H. Luo, Y. Chen, and W. Zhang, "An improved ID3 algorithm based on attribute importance-weighted," in *2010 2nd International Workshop on Database Technology and Applications*, 2010.

[14] J. Chen, D. Luo, and F. Mu, "An Improved ID3 Decision Tree Algorithm," in *Proceedings of 2009 4th International Conference on Computer Science & Education*, 2009, pp. 127–130.

[15] S. Teli and P. Kanikar, "A Survey on Decision Tree Based Approaches in Data Mining," *International Journal of Advanced Research in Computer Science and Software Engineering*, vol. 5, no. 4, pp. 613–617, 2015

[16] Y. Liu and N. Xie, "Improved ID3 algorithm," in *2010 3rd International Conference on Computer Science and Information Technology*, 2010, no. 4, pp. 465–468.

[17] V. K. Nijhawan, M. Madan, and M. Dave, "The Analytical Comparison of ID3 and C4 . 5 using WEKA," *International Journal of Computer Applications*, vol. 167, 2017.

[18] S. Wu, L. Wu, Y. Long, and X. D. Gao, "Improved Classification Algorithm by Minsup and Minconf Based on ID3," in *2006 International Conference on Management Science and Engineering*, 2006, pp. 135–139.

[19] S. Kraidech and K. Jearanaitanakij, "Improving ID3 Algorithm by Combining Values from Equally Important Attributes," vol. 6, pp. 17–20, 2017.

[20] N. Kaewrod and K. Jearanaitanakij, "Improving ID3 Algorithm by Using A∗ Search," in *ICSEC 2017 - 21st International Computer Science and Engineering Conference 2017, Proceeding*, 2018, vol. 6, pp. 132–135.

[21] L. Xian, Q. Fuheng, Y. Yong, and C. Hua, "An Improved ID3 Decision Tree Algorithm Based on Attribute Weighted," in *International Conference on Civil, Materials and Environmental Sciences (CMES 2015)*, 2015, no. Cmes, pp. 613–615.

[22] J. H. Liu and N. Li, "Optimized ID3 algorithm based on attribute importance and convex function," in *ITME 2011 - Proceedings: 2011 IEEE International Symposium on IT in Medicine and Education*, 2011, vol. 2, pp. 136–139.

[23] E. G. Learned-Miller, "Entropy and Mutual Information," *University of Massachusetts, Amherst, Department of Computer Science*, pp. 1–4, 2013.

[24] H. Li, X. Wang, and Y. Li, "Using mutual information for selecting continuous-valued .attribute in decision tree learning .," no. November, pp. 2–5, 2003.

[25] C. Ding and H. Peng, "Minimum redundancy feature selection from microarray gene expression data," *Journal of Bioinformatics and Computational Biology*, vol. 3, no. 2, pp. 185–205, 2005.

[26] L. Fang, H. Jiang, and S. Cui, "An Improved Decision Tree Algorithm Based on Mutual Information," in *2017 13th International Conference on Natural Computation, Fuzzy Systems and Knowledge Discovery (ICNC-FSKD)*, 2017, pp. 1615–1620.

[27] I. D. Mienye, Y. Sun, and Z. Wang, "Prediction performance of improved decision tree-based algorithms: a review," in *2nd International Conference on Sustainable Materials Processing and Manufacturing*, 2019, vol. 35, pp. 698–703.

[28] M. Bennasar, Y. Hicks, and R. Setchi, "Feature selection using Joint Mutual Information Maximisation," *Expert Systems with Applications*, vol. 42, no. 22, pp. 8520–8532, 2015.

[29] P. A. M. Bhadgale and M. S. Natu, "Implementation of Improved ID3 Algorithm Based on Association Function," vol. 114, no. 10, pp. 1–9, 2017.

[30] P. P. G. Ahire, S. S. Kolhe, K. Kirange, H. Karale, and A. Bhole, "Implementation of Improved ID3 Algorithm to Obtain more Optimal Decision Tree .," vol. 11, no. 02, pp. 44–47, 2015.

[31] S. Sivakumar, S. Venkataraman, and R. Selvaraj, "Predictive Modeling of Student Dropout Indicators in Educational Data Mining using Improved Decision Tree," vol. 9, no. January, 2016.

[32] X. J. Suganya and R. Balasubramanian, "Enhanced ID3 algorithm based on the weightage of the Attribute," in *International Journal of Innovatice Research in Advanced Engineering*, 2016, vol. 3, no. 03, pp. 12–18.

[33] X. J. Chen, Z. G. Zhang, and Y. Tong, "An improved ID3 decision tree algorithm," *Advanced Materials Research*, vol. 962–965, pp. 2842–2847, 2014.

[34] J. A. Jupin, T. Sutikno, M. A. Ismail, M. S. Mohamad, and S. Kasim, “Review of the machine learning methods in the classification of phishing attack,” vol. 8, no. 4, pp. 1545–1555, 2019.