Indonesian Journal of Electrical Engineering and Computer Science
Vol. 18, No. 1, April 2020, pp. 209~217
ISSN: 2502-4752, DOI: 10.11591/ijeecs.v18.i1.pp209-217

Developed third iterative dichotomizer based on feature decisive values for educational data mining

Saja Taha Ahmed[1], Rafah Al-Hamdani[2], Muayad Sadik Croock[3]
[1,2]The Informatics Institute for Postgraduate Studies, Iraqi Commission for Computers & Informatics (IIPS-ICCI)
[3]Computer Engineering Department, University of Technology, Iraq

Article Info

Article history:

Received Jun 11, 2019
Revised Sep 12, 2019
Accepted Oct 3, 2019

Keywords:

Accuracy
Decision tree
Decisive values
EDM
ID3
Prediction

ABSTRACT

Recently, the decision trees have been adopted among the preeminent utilized classification models. They acquire their fame from their efficiency in predictive analytics, easy to interpret and implicitly perform feature selection. This latter perspective is one of essential significance in Educational Data Mining (EDM), in which selecting the most relevant features has a major impact on classification accuracy enhancement. The main contribution is to build a new multi-objective decision tree, which can be used for feature selection and classification. The proposed Decisive Decision Tree (DDT) is introduced and constructed based on a decisive feature value as a feature weight related to the target class label. The traditional Iterative Dichotomizer 3 (ID3) algorithm and the proposed DDT are compared using three datasets in terms of some ID3 issues, including logarithmic calculation complexity and multi-values features selection. The results indicated that the proposed DDT outperforms the ID3 in the developing time. The accuracy of the classification is improved on the basis of 10-fold cross-validation for all datasets with the highest accuracy achieved by the proposed method is 92% for the student.por dataset and holdout validation for two datasets, i.e. Iraqi and Student-Math. The experiment also shows that the proposed DDT tends to select attributes that are important rather than multi-value.

Corresponding Author:

Saja Taha Ahmed,
The Informatics Institute for Postgraduate Studies,
Iraqi Commission for Computers & Informatics, Baghdad, Iraq.
Email: sajataha@ymail.com ; sajataha2@yahoo.com

1. INTRODUCTION

Educational Data Mining (EDM) is employed to extract the relevant information from the extensive and complex educational datasets and it is valuable for data analysis and predictions [1]. The prediction is commonly applied using EDM that considers the following techniques: classification, clustering, association rule mining, etc. Classification is the most popular EDM methodology used for student performance prediction. There are numerous classification methods that can be categorized such as decision tree, neural network, k Nearest neighbor, etc. These techniques are typically accustomed to building the classification model, which predicts the future trend based on the previous pattern [2-3].

The decision tree is a foremost widespread methodology for data classification, which incorporates numerous types, such as Third Iterative Dichotomizer (ID3) that selected optimal attribute using information gain [4]. Different decision tree methods are developed from the ID3 method, such as C4.5 based on gain ratio [5], as well as Classification and Regression Tree (CART) used Gini index [6].

In general, the decision tree assists educational institutions and universities in decision making in order to provide a student with the necessary assistance in the learning process. It is so popular because complex data can be presented in a visual representation with all possible outcomes and produce classification rules that are easy to interpret than other classification methods. The most relevant subset features for a decision automatically emerge through the process of developing the tree, the top nodes of the tree are the most essential, since they are deciding the subsequent decisions to be made. In addition, the tree demonstrates the order decisions must be made and eliminates ambiguity related to how each item influences the others [7]. Nevertheless, ID3 specifically, has some burdens, such as:

a) It is time-consuming due to information entropy calculation which is based on logarithmic algorithms [8-9] since the computation speed of the logarithmic expression is slower than four arithmetic operations that only include adding, subtract, multiply and divide [10].
b) It uses information gain as attribute selection criteria that pick the multi-values attribute, and the number of attribute values cannot be used to measure the attribute significance. This major shortcoming influences the accuracy of the decision tree [11].
c) The Decision tree can have overfitting, a phenomenon in which a model becomes more complex. When it is excessively dependent on irrelevant attributes of the training data, the result is that it works well on the training data but is relatively poorly predictive on unseen instances [12].

Over the past few years, a number of researchers have presented many related works for the use and/or suggestion of an enhancement in decision tree methods of various classification problems, below are some of the related works in this field.

ID3 has some exist disadvantages such as tending to select attributes biasing towards multi-values. The logarithmic expression has a high complexity computation and large-scale size. The authors of [13] proposed an improved ID3 algorithm that combines the simplified information entropy based on different weights with coordination degree in rough set theory. The traditional ID3 and the improved one are compared by exploiting three datasets, the experimental results showed that the proposed algorithm outperformed in the running time and tree size, but not in classification accuracy for small datasets.

The ID3 uses information gain tend to select the attribute with more values but it cannot measure the attribute importance via the number of attribute values. Therefore, the authors of [14] proposed a new method that selected the splitting attribute based on the utilization of conditional probability calculation of close contact between the attributes and the decision attributes. It joined with information gain to get higher predictive accuracy and less number of leaves without taking into consideration the running time. In perspective of the above issue, the authors of [15] suggested normalized association function combined with gain for each attribute to decide splitting decision, this can enhance accuracy but increase time complexity for proposed decision tree.

This paper aims to create a classification model particularly a decision tree algorithm that can effectively characterize students into one of two classes (Pass or Fail) by predicting the future grades of the students in their final examinations. The proposed algorithm aims to identify significant factors influencing student achievement and addresses the mentioned ID3 problems. A new methodology is utilized to build the proposed Decisive Decision Tree (DDT) based on the fact that the evaluation must consider the combination between the relevancy degree of each feature and the degree of classification accuracy enforcement. Therefore, the features relevancy degrees and the existing cross coupling are evaluated when they are combined together based on feature decisive (weighting) values. The proposed mechanism is examined by three datasets, namely, Iraqi dataset and UCI student performance dataset that includes mathematics, and Portuguese language courses datasets. The experimental results show that the proposed DDT obtains better performance than traditional ID3, in terms of, classification accuracy, running time and optimum multi-value feature selection.

2. RESEARCH METHOD

This study will include two phases as a part of methodology, as follow:

2.1. Dataset Collection

As mention earlier, this study incorporates three datasets. The first dataset is called Iraqi dataset which is uploaded at [16] and used for EDM preprocessing and Neural Network classification by [17]. It is collected during the second semester of 2018 by applying (or submitting) questionnaire in three Iraqi secondary schools for the applicable and biological branches of the final stage. The questionnaire initially contains 56 questions in three A4 sheets and 250 students (samples) respond to the questionnaire. Later, 130 samples are discarded due to lack of information, as pre-processing is used to obtain students ' most complete information. This study considers 120 instances with 55 features for experimental purposes after removing inconsistencies and incompleteness in the dataset. The attributes are divided into five main categories: Demographic, Economic, Education, Time and Marks. Furthermore, new features such as holidays and worrying effects are introduced. Also, the relationships between parents and schools and the student's use of books and references are considered.

The second used dataset in this study is (Student Alcohol Consumption Data Set), obtained from UCI Portugal [17-18]. This data set was collected during the 2005-2006 year from two public schools depending on two sources: school reports for the three-period grades and number of school absences, and questionnaires. The dataset consists of two datasets: student-mat.csv (Math), which holds 395 instances of Math course) and student-por.csv (Por), which holds 659 instances of Portuguese language course. Both of these datasets, consisting of 32 attributes.

2.2. The Proposed Methodology

A new criterion to build a decision tree for student performance prediction is presented. The Decisive Feature (Weight) value was calculated for both the training and the test set depending on the relative probability of the existing features occurring with respect to the target class.

The first stage is DDT building, in which the proposed system introduces the idea of obtaining each attribute in training set an importance via testing its significant degree with target class using the feature weight value calculated for each of the attributes, initially (1) [19-20] is used to compute a significant degree for target class:

$$D_t = \frac{\text{Ftsuccess}-\text{Ftfail}}{\text{Ftsuccess}+\text{Ftfail}} \tag{1}$$

Where; **t** is a target class.
D$_t$ is the Decisive value of the target.
Ftsuccess is the frequency of occurrence of success class.
Ftfail is the frequency of occurrence of fail class.

The decisive values of the attributes are considered as leading indicators for feature weighting and significance analysis for the student's success/failure prediction task. The Decisive value (D) is within [1, -1] range. If the value is approximately 1, it implies that most of the feature is done with a successful student class. If the value is approximately to -1, it implies that the feature generally happens with a failure student class. While the value is near to 0, it implies that the feature in the success class is almost equivalent to failure class.

The Cumulative Decisive value (CD) is computed using (2) by multiplying the D value of each attribute’s category with its frequency. This takes into account the volume of the frequent occurrence of values that construct a specific attribute in relation to the target class.

$$\text{CD(i)} = \sum_{j=1}^{N} \left(\text{D(ij)} * \frac{\mathit{Frequancy\ of\ Occurance\ of\ Value\ j}}{\mathit{Total\ Number\ of\ Values\ within\ Attribute\ i}}\right) \tag{2}$$

Where; **i** is a specific attribute.
j is a value within attribute **i**.
N is the number of values (categories) within attribute **i**.
D(ij) is the Decisive value of specific category j within attribute i, the (1) of the target becomes (3) for attribute categories, with the description of the following parameters:

$$D(ij) = \frac{\text{Fisuccess(ij)}-\text{Fifail(ij)}}{\text{Fisuccess(ij)}+\text{Fifail(ij)}} \tag{3}$$

Fisuccess (ij) is the frequency of occurrence of value j of attribute i in success class. Fifail (ij) is the frequency of occurrence of value j of attribute i in a fail class.

Finally, the best attribute is selected using Gain by subtracting CD for each attribute from the target Dt using (4). The highest attribute gain is recommended to be the best attribute placed at the root for further splitting. The proposed DDT is continued in this way by testing every property with others until pure target class (all success or failure) is reached or no further splitting is found. In the latter case, when there is no combination of the values of attributes along the current path. The proposed DDT takes into consideration D(ij) for a specific category (current value) in the original training set, which has no combination within this path. Then DDT decides whether the leaf node will succeed or fail, if D(ij) value predominantly closes 1, at that point, the decision will succeed, otherwise, the decision will fail, this has a major impact on the tree classification accuracy enhancement. In contrast to traditional DT, which depends on the majority of the target class label when there is no combination of values (i.e. samples(value) is empty) and ignores the weight of current category on the classification. The important steps for building the proposed DDT, are illustrated in Algorithm (1).

$$Gain(i) = D_t - CD(i) \qquad (4)$$

Algorithm (1) Decisive Decision Tree Building

Input: Samples is a data table [#students, #attributes], target attribute,
array of attributes [#attributes].
Output: Decision Tree.
Algorithm Steps
If all sample positive, Return True.
If all sample negative, Return False.
If attributes are empty, Return the most distinct attribute as root.
Calculate Decisive Degree using (1), for target attribute:
For each attribute i in attributes
For each value j in attribute i
Calculate Decisive Degree D(ij) using (3), for each value
j of attribute i.
Calculate Cumulative Decisive Degree using (2), for
attribute i:
Calculate the difference between CD attribute and D target using (4)
Create a Root node for an attribute with the highest difference as a good
discriminating feature.
If (best attributes not best list), then add it to best attribute list.
For each value in the best attribute.
Begin
Select samples row when best attributes equal to value.
If samples (value) empty, then Begin
Select all samples with the value from the dataset.
Determining target class via D(ij) value.
Add leaf node with target class to Root.
End
Else Begin
Create child node using DDT (samples(value), target attribute,
attributes-best attribute).
Add child node to Root
End
End
Return Root

In the second stage, when a DDT is generated, the target class prediction for a new student in the test set is determined and the classification rules can be extracted using the DDT search clarified in the algorithm (2). Each new student information enters as a matrix of two tuples, tuple 0 contains the name of the attributes, and tuple 1 contains values corresponding to the attributes. DDT search mainly depends on matching student information at each node and tracing the path from the root to the target class at a leaf node.

Algorithm (2) DDT Search
Input: Root, new student information as string test [2, #attributes]//row 0: name of an attribute, row 1 values of each attribute Output: Path for a new student in the test set. Algorithm Steps Step1: Define index as -1 and tag as False. For each attribute i in the test set If test [0, i] equal to Root. Attribute Begin Set index to i; Break. End Set Path to Root.Attribute + test [1, index] If Root.Attribute. Values not equal to Null Begin For each value j in attribute If test [1, index] equal to Root.Attribute. Values[j] Begin Set Val to j Set Tag to True; Break; End If Tag equal to True Begin

```
Define Child_Node as TreeNode
Set Child_Node to Root.Child(Root.Attribute. Values[Val])
Set Root to Child_Node
Goto Step 1
End
End
Else Goto Step 2
Step 2: Return Path
```

3. RESULTS AND ANALYSIS

The experiments and the application system in this study are developed based on visual studio C# 2015. The model validation empowers locating the best features of the model while also shielding it from getting the chance to be over fitted. The proposed DDT model is assessed utilizing two of the most popular evaluation criteria 10-fold cross-validation and hold out methods. In 10-fold cross-validation [21], all the dataset has been divided into 10 subsets of approximately equal size. This is an iterative procedure, each time 9 subsets acts as a training data and one set is used as a testing data. In the holdout method [22], the data set is separated into two sets of training data is 70% of the entire dataset and testing data is 30%, represents the remaining dataset.

Since the decision tree needs the data to be in the categorical formulation, the grade features must have discrete values to obtain better results. The discretization mechanism has been exploited to convert the grade values from numerical values to nominal ones. Specific classes are defined, which represent classes label for student performance prediction, which can be either “Pass” or “Fail”. In UCI dataset, there are three average G1, G2 and G3 have ranged from 0 to 20. Thus, if the student has average equal or higher than 10, it should be defined within the “Pass” label, otherwise should be defined as “Fail” student. In Iraqi dataset, grade scores are within range 0-100, if the student has average equal or higher than 50, it should be defined within “Pass” label, otherwise is classified as “Fail” student.

A small training data set is examined to illustrate the difference between the structure of ID3 and DDT algorithms. Table 1 shows the dataset used in research work [14].

Table 1. The Dataset

ID	Chinese	Mathematics	English	Physics	Summary	Target Class
1	general	good	bad	general	qualified	Q
2	general	good	good	good	qualified	Q
3	good	general	general	good	qualified	Q
4	optimal	general	good	good	qualified	Q
5	general	general	general	general	qualified	Q
6	good	bad	general	bad	unqualified	U
7	optimal	bad	bad	general	unqualified	U
8	good	optimal	optimal	optimal	qualified	Q
9	general	general	optimal	good	qualified	Q
10	optimal	bad	general	general	qualified	Q
11	bad	good	good	bad	unqualified	U
12	good	general	good	good	qualified	Q
13	general	bad	good	general	qualified	Q
14	general	general	optimal	good	qualified	Q
15	good	bad	good	general	qualified	Q
16	optimal	general	optimal	good	qualified	Q
17	optimal	optimal	optimal	optimal	qualified	Q
18	good	bad	good	general	qualified	Q
19	good	general	bad	optimal	qualified	Q
20	general	general	general	general	qualified	Q

ID3 favors the selection of attribute that has a larger number of values (i.e. categories) because the attribute with more values has high information gain than others. Figure 1 shows the ID3 feature selection, which chooses the ID feature with 20 values as the root node for the decision tree.

The proposed DDT selects English attribute with four categories (bad, general, good, optimal) to be the root node of the decision tree and exclude ID as it has no predictive power of classification which explained in Figure 2. Since the proposed DDT tends to select the attribute that has high weight value regarding target labels, in the case of Table.1 there are two target labels qualified and unqualified.

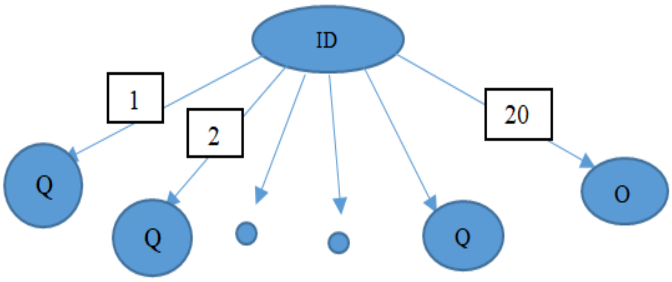

Figure 1. ID3 Decision tree construction

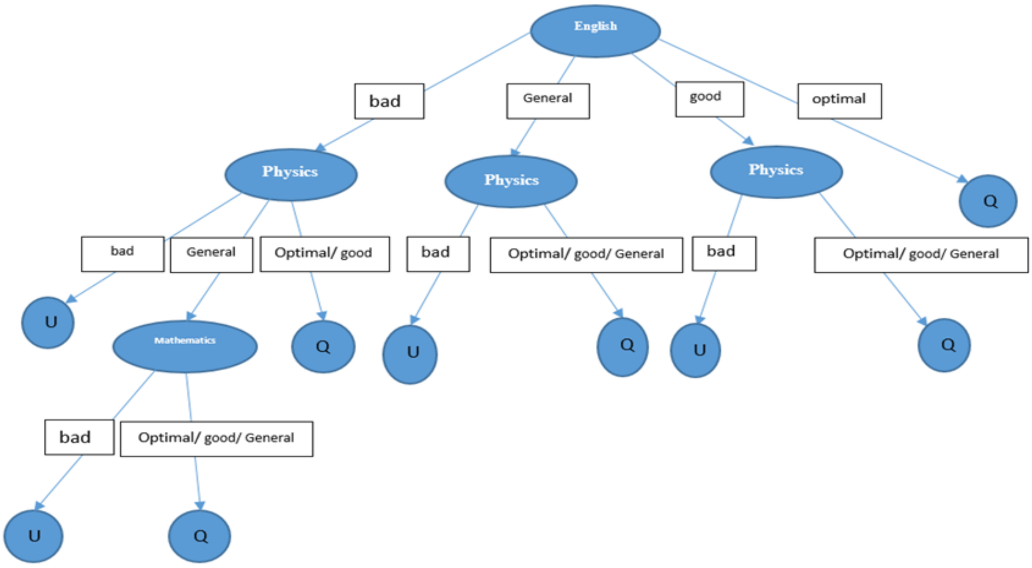

Figure 2. DDT Decision tree construction

The evaluation on the basis of Accuracy (ACC) value is executed. Accuracy measures the degree to which the instances correctly classified by machine learning algorithm and can be computed using a confusion matrix with (5) as follows [23]:

$$ACC = \frac{\sum True\ Positive + \sum True\ Negative}{\sum Total\ Population} \tag{5}$$

Holdout cross-validation for three datasets Iraqi dataset, Por, and Math depend on confusion matrix that can be illustrated in Tables of 2, 3, and 4. It can be shown that the achieved accuracies of the predicted classes are 88.88, 61.5, and 74.7, respectively.

Table 2. Confusion Matrix of Iraqi Dataset

Total Population=36		Actual Calss	
Acc=88.88		SUCCESS	FAIL
Prediction	SUCCESS	TP=32	FP=4
Class	FAIL	FN=0	TN=0

Table 3. Confusion Matrix of Por Dataset

Total Population=195		Actual Calss	
Acc= 61.5		SUCCESS	FAIL
Prediction	SUCCESS	TP=105	FP=70
Class	FAIL	FN=5	TN=15

Table 4. Confusion Matrix of Math Dataset

Total Population=119		Actual Calss	
Acc=74.7		SUCCESS	FAIL
Prediction	SUCCESS	TP=82	FP=25
Class	FAIL	FN=5	TN=7

Holdout cross-validation may waste datasets and produce a high error rate. Since the aim is generalizing proposed model well without overfitting, therefore 10-fold cross-validation is used to ensure all observations are used for both training and testing. Each observation is used for testing exactly once.

At the point when the tree is built based on specific features and gives better exactness then the tree can be utilized for feature selection and can consider these features as the best parameters with high predictive power. The best parameters can be determined from datasets using the proposed DDT with the highest accuracy. The perfect accuracies of Iraqi, Por and Math are achieved at iterations 10, 6 and 8, respectively. Table 5 shows 10 iterations and the overall accuracy using 10-fold cross-validation and holdout of the proposed DDT for three datasets.

Table 5. DDT Holdout and 10-Fold Cross-Validation

DDT	Holdout	1	2	3	4	5	6	7	8	9	10	10Fold AVG
Iraq	88.88	58.3	58.3	91.6	83.3	91.6	91.6	91.6	91.6	83	91.6	83.3
Por	61.5	92	87.5	70.3	87.5	84.3	92	76.5	73.4	48.4	57.8	77
Math	74.7	69	71.9	61	58	61.5	64	64	87	69	66.6	67.2

Table 6 shows ID3 based on Holdout and 10- Fold Cross-Validation, from Tables 5 and 6, it can be inferred that the proposed DDT has a higher prediction accuracy than ID3 on the basis of holdout and average of 10-fold cross-validation for two reasons, the first DDT can select the feature based on its importance (weight) taking into account the target class, as opposed to traditional ID3, which chooses a feature of a high category that may not have a predictive classification power, secondly, when there is no combination between features (i.e. sample(value) is empty), the DDT depends on D(ij) for the current value to determine class of leaf nodes, while traditional ID3 decides on a leaf node based on the majority of the class of target attribute, ignoring the tendency of a current value towards a specific class.

Table 6. ID3 Holdout and 10-Fold Cross-Validation

ID3	Holdout	1	2	3	4	5	6	7	8	9	10	10Fold AVG
Iraq	83	59.3	78	59.3	77	86	91.6	91.6	90	66.6	91.6	79
Por	67	87	82.8	64	84	81	87.5	73.4	71.8	57.8	60.9	75
Math	62	53.8	64	74	51	58.9	61.5	64	66.66	69	58.9	62

In terms of running time, the proposed DDT surpass the traditional ID3, which has faster decision tree construction time than that of ID3. Figure 3, showing that the proposed DDT reduces the time complexity of the traditional ID3 for three datasets since the proposed DDT utilizes simple mathematical expressions incorporates subtraction, addition, and division. All these operations are less computational complexity than ascertaining entropy information that implies calculation of the logarithm algorithm in traditional ID3, which makes DDT useful for improving real-time capability such as online learning systems.

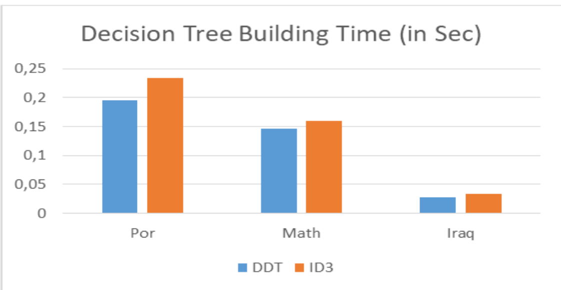

Figure 3. Decision Tree Construction Time for ID3 and DDT

Since the proposed DDT building algorithm selects features locally based on their weight (decisive value), and with relation to the feature selected in earlier stages, so that the features that occur in the DDT are complementary. Therefore, DDT gives a set of extremely important features that lead to a significant increase in the model's predictive accuracy. Table 7 shows the best DDT feature subset, which results in higher accuracy for three datasets. Once the best parameter combination has been discovered, a set of classification rules can be extracted from the proposed DDT. These rules help to classify students and foresee the final status of the students.

Table 7. DDT Best Feature Subset

Datasets	Accuracy	#Iteration	Features
Iraq	91.6	10	Higher Education Willing, sleep Hour, Father Alive, Attendance, Failure Year, Study Hour, Internet Usage, Parent Meeting, Worry Effect, Arrival Time, Holiday Effect, Transport.
Por	92	1	Fedu, higher, Fjob, absence, study time, health, famrel, walc, dalc, activities, free time, famsize. gaurdian
Math	87	8	Internet, freetime, famrel, failure, health, absence, walc,dalc, study time, romance, reason, health, medu, higher,paid, schoolsup, gout.

Table 8 shows a comparison of the proposed DDT with the research work of [24]. This research uses Por dataset from UCI to predict student performance based on eight features G2, G1, failures, higher, Medu, school, studytime, Fedu. In addition, a comparison of the proposed DDT with the research work of [25]. This research uses Math dataset from UCI to predict student performance based on 19 features including the class attribute: sex, famsize, address, Pstatus, Medu, Fedu, Mjob, Fjob, traveltime, studytime, schoolsup, higher, internet, romantic, freetime, Dalc, Walc, health, success. It is clear that the proposed DDT surpass all methods utilized in these researches for two UCI (Por and Math) datasets.

Table 8. Accuracy Comparison of Our Proposed DDT and other Methods for UCI Datasets

Dataset	Research Work	Method	Accuracy
Por	[24] (2019)	Naïve Bayes	73.18 %
		Decision Tree	76.27 %
		RandomTree	67.95 %
		REPTree	76.73%
		JRip	74.11 %
		OneR	76.73 %
		SimpleLogistic	73.65%
		ZeroR	30.97%
	Our Proposed Model	The Proposed DDT	92%
Math	[25] (2016)	PCF with k-medoids algorithm	65.82 %
		PCF with k-means algorithm	63.50%
	Our Proposed Model	The Proposed DDT	87%

4. CONCLUSION

This paper proposed an improved ID3 algorithm, which employs attribute weight between attributes and class labels for selection splitting attribute. Constructing the proposed DDT based on feature decisive value ensures that each time important rather than more attribute value is selected. This has a major impact on enhancing classification accuracy. It also has a faster constructing time than classical ID3 which implies time complexity of logarithm computation, as the proposed DDT depends only on calculation attribute frequency of occurrences, which overcomes the limitations of the ID3 algorithm. The proposed algorithm was tested over three datasets. These include Iraqi and two UCI datasets. The obtained results showed that the developed ID3 algorithm beat the traditional ID3 in terms of accuracy and consumed execution time.

REFERENCES

[1] M. Abdullahi, et al., "The Patterns of Accessing Learning Management System Among Students", *Indonesian Journal of Electrical Engineering and Computer Science (IJEECS)*, Vol. 13, No. 1, ISSN: 2502-4752, DOI: 10.11591/ijeecs. v13.i1. pp15-21, 2019.

[2] M. Gandhi and S.N. Singh, "*Predictions in Heart Disease Using Techniques of Data Mining*". In Proceedings of the International Conference on Futuristic Trends on Computational Analysis and Knowledge Management, pp. 520-525, 2015.

[3] Özdemir, U. Yavuz, F. Abdulhafidh Dael, "Performance Evaluation of Different Classification Techniques Using Different Datasets", *International Journal of Electrical and Computer Engineering (IJECE)*, Vol. 9, No. 5, ISSN: 2088-8708, DOI: 10.11591/ijece. v9i5. pp. 3584-3590, 2019.

[4] V.N. Phu, V.T.N. Tran, V.T.N. Chau, N.D. Dat, K.L.D. Duy, "A Decision Tree Using ID3 Algorithm for English Semantic Analysis". *Int. J. Speech Technol.*, 20, pp. 593-613, 2017.

[5] E. Budiman, Haviluddin, N. Dengan, A. H. Kridalaksana, M. Wati, Purnawansyah, "Performance of Decision Tree C4.5 Algorithm in Student Academic Evaluation", *Springer Nature Singapore Pte Ltd*, pp. 380-389, 2018.

[6] T. Roya, S. Sundaramb, P. Ranjana, R. Balasubramaniama, "*Prediction of Material Removal in Single Spark Micro-EDM Using Multiple Linear Regression and CART*", the 10th International Conference of Precision, Meso, Micro, Nano Engineering, ISBN: 978-93-80689-28-9, 2017.

[7] S. T. Ahmed, R. S. Al-hamdani, M. S. Croock, "Studying of Educational Data Mining Techniques", *International Journal of Advanced Research in Science, Engineering and Technology*, Vol. 5, Issue 5, 2018.

[8] J. Liang, Z. Shi, "The Information Entropy, Rough Entropy and Knowledge Granulation in Rough Set Theory", *Int. J. Uncertain. Fuzziness Knowl.-Based Syst.*, 12, pp. 37-46, 2008.

[9] T.P. Exarchos, M.G. Tsipouras, C.P. Exarchos, C. Papaloukas, D.I. Fotiadis, L.K. Michalis, "A Methodology for The Automated Creation of Fuzzy Expert Systems for Ischaemic and Arrhythmic Beat Classification Based On a Set of Rules Obtained by A Decision Tree", *Artif. Intell. Med.*, 40, pp.187-200, 2007.

[10] J. Sneyers, T. Schrijvers, B. Demoen, "*The computational power and complexity of Constraint Handling Rules*", In Proceedings of the 2nd Workshop on Constraint Handling Rules, pp. 3-17, 2005.

[11] U.M. Fayyad, K.B. Irani, "*The Attribute Selection Problem in Decision Tree Generation*", In Proceedings of the National Conference on Artificial Intelligence, San Jose, CA, USA, pp. 104-110, 1992.

[12] M. Bramer, "Avoiding Overfitting of Decision Trees. In: Principles of Data Mining", book, *Springer*, DOI https://doi.org/10.1007/978-1-84628-766-4_8, ISBN 978-1-84628-766-4, PP. 119-134, 2007.

[13] L. Yi-bin, W. Ying-ying and R. Xue-wen, "Improvement of ID3 algorithm based on simplified information entropy and coordination degree", *Chinese Automation Congress (CAC), Jinan*, doi: 10.1109/CAC.2017.8243009, pp. 1526-1530, 2017.

[14] X. Liang, F. Qu, Y. Yang, H. Cai, "*An Improved ID3 Decision Tree Algorithm Based on Attribute Weighted*", International Conference on Civil, Materials and Environmental Sciences, DOI: 10.2991/cmes-15.2015.167, pp. 613-615, 2015.

[15] P. G. Ahire, S. Kolhe, K. Kirange, H. Karale, A. Bhole, "Implementation of Improved ID3 Algorithm to Obtain More Optimal Decision Tree", *International Journal of Engineering Research and Development*, Vol. 11, Issue 02, PP.44-47, 2015.

[16] S. Taha, "Iraqi Student Performance Prediction", *Mendeley Data*, v1, http://dx.doi.org/10.17632/smgx6s5pwr.1, DOI: 10.17632/smgx6s5pwr.1, 2018.

[17] Saja Taha Ahmed, Rafah Shihab al-Hamdani, Muayad Sadik Croock, "EDM Preprocessing and Hybrid Feature Selection for Improving Classification Accuracy", *Journal of Theoretical and Applied Information Technology*, Vol.96. No 1 ISSN: 1992-8645, 2019.

[18] P. Cortez and A. Silva.,"*Using Data Mining to Predict Secondary School Student Performance*", The 5th Future Business Technology Conference (FUBUTEC 2008), pp. 5-12, Porto, 2008.

[19] N. G. M. Jameel, L. E. George, "Detection Phishing Emails Using Features Decisive Values", *International Journal of Advanced Research in Computer Science and Software Engineering*, Vol.3, Issue.7, 2013.

[20] A. Abdullah, L. E. George, and Imad J. Mohammed," Email Phishing Detection System Using Neural Network", *Research Journal of Information Technology*, Vol.6, Issue.3, pp. 39-43, 2015.

[21] R. Kohavi, "*A study of cross-validation and bootstrap for accuracy estimation and model selection*", In Proceedings of International Joint Conference on AI (IJCA), Vol. 14, No. 2, P. 1137-1145, 1995.

[22] S. Raschka, "Model Evaluation, Model Selection, and Algorithm Selection in Machine Learning", arXiv preprint, arXiv:1811.12808v2, 2018.

[23] M. Z. H. Jesmeen, A. Hossen, J. Hossen, J. E. Raja, "AUTO-CDD: Automatic Cleaning Dirty Data Using Machine Learning Techniques", *TELKOMNIKA (Telecommunication, Computing, Electronics and Control)*, Vol.17, No.4, ISSN: 1693-6930, DOI: 10.12928/TELKOMNIKA.v17i4.12780, pp.2076-2086, 2019.

[24] Y. K. Salal, S. M. Abdullaev, M. Kumar, "Educational Data Mining: Student Performance Prediction in Academic", *Blue Eyes Intelligence Engineering & Sciences Publication, International Journal of Engineering and Advanced Technology (IJEAT)*, ISSN: 2249-8958, Vol.8, Issue-4C, 2019.

[25] N. U. Sati, "*Prediction of Students' Success in Mathematics by A Classification Technique Via Polyhedral Conic Functions*", International Conference on Research in Education and Science, the Eurasia Proceedings of Educational & Social Sciences (EPESS), ISSN: 2587-1730, Vol. 5, pp. 190-195, 2016.