

Iterative Dichotomiser Three (ID3) Algorithm For Classification Community of Productive and Non-Productive

Ida¹, Muhammad Faisal²

¹Computer Science, Information Systems , Professional College of Informatics and Computer Management

^{1,2}Jl. A. P. Pettarani No. 27 Macassar, Indonesia

ABSTRACT

Article:

Accepted: April 16, 2022

Revised: May 13, 2022

Issued: June 15, 2022

© 2023 The Author(s).



This is an open-access article under the [CC BY-SA](https://creativecommons.org/licenses/by-sa/4.0/) license

*Alamat Korespondensi:

idamulyadi@stmikprofesiodnal.ac.id

One way to tackle poverty is to provide information about productive and non-productive communities in each rural. This is very beneficial for the government, especially in each rural regarding the classification of community data. This research aims to classify productive and non-productive people so that the government can prioritize assistance for people deemed necessary to be more creative in fulfilling their family's economy. The research method used is the Iterative Dichotomiser Three (ID3) algorithm to build a decision tree. The process in the decision tree is changing the shape of the data (table) into a tree (tree) and generating rules based on the highest Entropy and Gain values. The study's conclusion shows that this algorithm can be processed in a shorter time, with shorter decision rules and higher prediction accuracy, by displaying the highest gain value. The parameters used to consist of education, age, income, and employment status, which results in the following rule if higher education and high income, then the result is a productive society, whereas if high school education and low income, then the result is a non-productive society.

Keywords: *web, asset management, codeigniter, bootstrap*

1. INTRODUCTION

Science and technology are developing rapidly, so that it becomes a motivator for scientists and the public to increase creativity and innovation in helping users, especially the government, in recording community productivity by classifying productive and non-productive communities so that the government can prioritize non-productive communities to get assistance from the government and this is one way to overcome poverty. The application of research methods and technology in the decision-making system makes it easy for the government to process community data more effectively and efficiently, in this case the government (Pisang Utara kelurahan).

Pisang Utara Village, Ujung Pandang District, Makassar City, covers an area of 199.26 km² and has 25,866 people, 10,358 family heads, and 3.23 average family members. And the area's self-sufficiency status. The government of the northern banana sub-district has implemented poverty alleviation programs by organizing various types of social services, collecting data on street children, and improving the socio-economic welfare of the poor. but have not carried out data collection on productive and non-productive people. For this reason, researchers will provide solutions in collecting community data by using the classification method. non-productive community information will be a priority for the government to focus on providing and creating creativity in developing the community's economy. The result of classification information is one of the government's ways to alleviate poverty.

In order to map productive and non-productive communities in the context of equal distribution of work for the community, the government of the northern banana sub-district requires reference data[1][2], A source of information that has been transformed from raw to valid data is referred to as reference data. However, the lack of technological support in the community's classification process regarding productive or non-productive status means achieving government goals, inequitable welfare distribution, and providing employment opportunities is extremely slow. A productive society is defined as a group of people who have a pattern of behavior that leads to the production of something valuable.

Non-productive people have uncertain incomes. Iterative Dichotomiser Three(ID3)[3] is a decision tree classification algorithm with information acquisition as one of its attributes. There are numerous advantages[4] to classifying a large amount of data using the ID3 algorithm, which has intuitive solid properties and is simple to parse.

Data mining is the process of exploring large amounts of data to find patterns that can be used to make decisions. Classification is one technique for decision-making. Data classification is a type of data analysis used to extract models that describe key data classes. [5].

The decision tree is a flowchart tree structure commonly used to obtain information for decision-making purposes. A rectangle represents each inner node, and an oval represents the leaf node. All inner nodes have two or more child nodes. All interior nodes contain a split, which tests the value of the attribute expression. The arcs from the inner vertex to its children are labeled with different test results. Each leaf node has a class label associated with it. The Decision Tree begins at the root node, where the user performs the activity. According to the Decision Tree learning algorithm, the user recursively divides each node from these nodes. As a result, a Decision Tree is formed, with each branch representing a possible decision scenario and its outcome. The most widely used DT learning algorithms are Hunt's ID3, CART, CHAID, and C4.5[6]. CART (Classification and Regression Tree) is a method or algorithm of one of the data exploration techniques, namely the decision tree technique. CART was developed to perform classification analysis on nominal, ordinal, and continuous response variables. CART produces a classification tree if the response variables are categorical and produces a regression tree if the response variables are continuous. The smallest error rate value in the resulting classification tree will tend to make this tree used to estimate the response. The principle of this classification method is to sort all observations into two groups of observations and re-sort these groups of observations into the next two groups of observations, so that the minimum number of observations is obtained in each subsequent group of observations[1].

The CHAID method is a nonparametric exploratory technique for analyzing large data sets, which divides the data into subsets that are mutually exclusive (mutually exclusive) that can describe response variables. If the larger value ($\alpha = 0.05$) is selected in the data classification analysis using the CHAID method on a classification tree for one response variable, the more significant estimators produced will be, and vice versa. The CHAID effect can see the more significant factors clearly and detect any interactions at each division in the classification tree. Furthermore, this method can be used in analyzing categorical survey data in very large numbers with various reliable comparisons of significant levels[2].

The differences in ID3, C4.5, CHART and CHAID are:

- ID3 : select the attribute that has the greatest information gain value
- C4.5 : attribute selection using gain ratio
- CHART : uses the smallest error rate value for the response
- CHAID: the greater the value ($\alpha = 0.05$) selected in classifying data for one response variable, the more estimating factors are produced, and vice versa.

The similarities in ID3, C4.5, CHART and CHAID are that they both make decision trees in classifying data.

ID3 characteristics in building decision tree is top-down and divide-and-conquer. Top-down means decision tree is built from the root node to the leaves, while divide-and-conquer means that the training data is recursively partitioned into smaller parts during tree construction[3].

Previous research, Improvement of ID3 Algorithm Based on Simplified Information Entropy and Coordination Degree Using ID3 (Iterative Dichotomiser 3), combines information entropy based on different weights with degrees of coordination[3].

Following research employs data mining methods to create predictive quality models that deal with complex variable relationships in multi-stage manufacturing systems. The Cascade Quality Prediction Method (CQPM) employs a variety of Principal Component Analysis and Iterative Dichotomous 3 Algorithms. Improved prediction results in positive and negative class predictions[4].

The following research proposes the fuzzy-ID3 (FID3) algorithm and the fuzzy decision tree as

a classification method in breast cancer detection to overcome the limitations of existing methods. The FID3 algorithm combines algorithms with fuzzy systems and decision tree techniques. Before generating the fuzzy rule base, ID3 as a decision tree generates a predefined fuzzy database. The fuzzified dataset is used by the FID3 algorithm, which is a fuzzy version of the ID3 algorithm. This study discovered that combining the FID3 algorithm with the FUZZYDBD method was reliable, solid, and performed well in breast cancer classification[7].

The following study examines the performance of the traditional ID3 algorithm in online data migration of sports competition action; the results show that the performance of the k-nearest neighbor-based ID3 optimization algorithm is significantly improved and can also solve the overfitting problem that exists in the traditional ID3 algorithm. Furthermore, provide more precise methods and high-performance online data migration models for sports competition action data mining[8].

The following research describes the Iterative Dichotomiser (ID3) algorithm in classification techniques, which constructs a decision tree from the data set. The results demonstrate the ID3 algorithm for building a decision tree (Decision Tree). The algorithm's implementation describes the Teaching Assistant's performance during the summer and the regular semester[6].

The following research uses data classification, a type of data analysis used to extract models that describe essential data classes. Decision Tree constructs a tree structure by gradually breaking down the data set into smaller subsets, as implemented by the ID3 and C4.5 algorithms to construct a decision tree using the "entropy" and "information gain" measures are critical components in the construction of the classifier model[5].

Subsequent studies employing traditional decision trees for fault diagnosis frequently employ the ID3 construct algorithm. There are two parts to this study. The first section proposes a CV-DT based on the improved ID3 algorithm and the cluster validity index. This method selects the separation attribute with the highest classification credibility and improves diagnostic accuracy. The second step is to use a FR-DT created by the improved ID3 algorithm while accounting for the error rate. This algorithm considers each

attribute's partitioning ability and the priority of error isolation with a higher error rate[9].

An application has been developed that takes academic data from the university and generates a classification model using three different algorithms: artificial neural network, ID3, and C4.5. The performance of these models is compared to determine which model produces the best results and will be used to classify students. C4.5 and ID3 decision tree algorithms provide more accurate measurements of artificial neural networks. The C4.5 algorithm-generated tree has the best performance metrics, with precision, accuracy, and sensitivity of 0.83, 0.87, and 0.90, respectively[10].

This research aims to develop a decision support system for determining the performance status of productive and non-productive communities to optimize good results and avoid mistakes that occur during decision making in the northern banana village, Ujung Pandang sub-district, South Sulawesi province. The researcher honestly states that the research titled Application of the Iterative Dichotomiser Three (Id3) Algorithm in the Classification of Productive and Non-Productive Societies is entirely his or her idea.

Limitations of research data in terms of sample size, data quality and data generalization are as follows:

- The sample size used is the highest entropy and gain value
- The quality of the data used is age, income, employment status and education
- generalization of the research results, namely providing information based on the process of calculating data using the ID3 algorithm is a productive and non-productive society

2. METHODS

2.1. The implementation of this research uses the following methods:

a. Data collection

This method uses three methods: literature study, Interview, and Observation.

b. Literature or literature study

This method is carried out by studying references in scientific journals, the internet, and discussions.

c. Interview

Data collection was carried out by a question and answer process with related

parties, in this case, the Pisang Utara ward office. This method is intended to obtain information through staff and leaders regarding population data.

d. Observation

This method is carried out by direct Observation of population data processing consisting of education, income, age, and community performance status.

The following information was gathered in December 2021. Table 1 depicts population data for the northern banana ward community.

Table 1. Population Data for December 2021

Pen dud uk	Umur	Pengha silan	Status Pekerjaan	Pend idika n	Krite ria
1	<=30	Tinggi	Wirausaha	Sarja na	Produ ktif
2	<=30	Tinggi	Pegawai	Sarja na	Produ ktif
3	31..40	Tinggi	Wirausaha	Sarja na	Produ ktif
4	>40	Sedang	Pegawai	SMA	Produ ktif
5	>40	Rendah	Buruh	SMA	Non Produ ktif
6	>40	Tinggi	Pegawai	Sarja na	Produ ktif
7	31..40	Rendah	Buruh	SMA	Non Produ ktif
8	<=30	Sedang	Wirausaha	SMA	Produ ktif
9	<=30	Rendah	Buruh	SMA	Non Produ ktif
10	>40	Tinggi	Wirausaha	SMA	Produ ktif
11	<=30	Sedang	Pegawai	SMA	Produ ktif
12	31..40	Sedang	Wirausaha	SMA	Produ ktif
13	31..40	Tinggi	Wirausaha	Sarja na	Produ ktif
14	>40	Sedang	Pegawai	SMA	Produ ktif

2.2. Analysis Steps

The steps taken in this study by applying the ID3 algorithm are as follows:

- Setting up the dataset
- Calculating the value of entropy
- Calculating the value of Gain
- Create a branch node of the highest Gain value
- Repeat steps (2) to (4) until all nodes are partitioned.

The formula for counting Entropy:

$$Entropy(S) = \sum_{j=1}^k -P_j \log_2 P_j \quad (1)$$

description:

S = dataset

k = partition S

P_j = probability.

The Formula for Gain :

$$Gain(A) = entropy(S) - \sum_{i=1}^k \frac{|S_i|}{|S|} \times entropy(S_i) \quad (2)$$

Description :

S = data training sample

A = Atribut

|S_i| = total sample V

|S| = total data sample

Entropy(S_i) = the sample - i

Node divides the high Gain

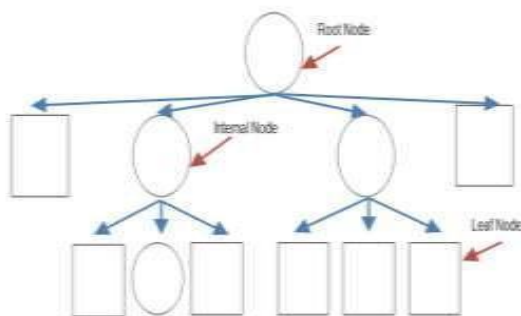


Figure 1. Node divide decision support.

Description:

- Root Node or root node: the node located at the very top of a tree is the Highest Gain Value
- Internal Node is a branching node that has one input and two outputs
- Leaf Node is an end node that only has one input and has no Outputs.

Reading the last rule

Conclude the existing decision tree to serve as a rule.

3. RESULTS AND DISCUSSION

The community classification system using the ID3 algorithm starts from the availability of datasets obtained from interviews and observations, then calculates attribute values, entropy values, and gain values. The following process makes decision tree branch nodes, then analyzes and produces the final rule reading.

3.1. Determine Attributes, Number of Cases, and Number of Predictions.

Determine the item and attribute values, the number of cases throughout the data, and the number of predictions for productive and non-productive people. The following is the display of the calculation of Figure 2 based on Figure 1 below.

Attribut		Jumlah Kasus	Prediksi		Rumus	
Item	Nilai	S	Produktif	Non Produktif	ENTROPY	GAIN
		14	11	3	0.749595257	
Umur	<=30	5	4	1	0.721928095	0.981389
	31.. 40	4	3	1	0.811278124	
	>40	5	4	1	0.721928095	
Penghasilan	Tinggi	6	6	0	0	0.749595
	Sedang	5	5	0	0	
	Rendah	3	0	3	0	
Status Pekerjaan	Wirasaha	6	6	0	0	0.749595
	Pegawai	4	4	0	0	
	Buruh	3	0	3	0	
Pendidikan	Sarjana	5	5	0	0	1.339928
	SMA	9	6	3	0.918295834	

Figure 2. Accounting for the Dataset

S = Obtained from the number of cases contained in the dataset
 Predictive values of productive and non-productive are obtained from the number of cases in the dataset.

3.2. Calculate the Entropy

$$Entropy(S) = \sum_{j=1}^k -P_j \log_2 P_j \quad (1)$$

$$Entropy(S) = (11/14 * \log_2(11/14)) + (3/14 * \log_2(3/14)) = 0.749595257$$

$$Entropy \text{ age } \leq 30 = (-4/5 * \log_2(4/5)) + (1/5 * \log_2(1/5)) = 0.721928095$$

$$Entropy \text{ age } \geq 31 \text{ and } \leq 40 = (-3/4 * \log_2(3/4)) + (-1/4 * \log_2(1/4)) = 0.811278124$$

$$Entropy \text{ age } > 40 = (-4/5 * \log_2(4/5)) + (-1/5 * \log_2(1/5)) = 0.721928095$$

The calculation for entropy continues until attribute the senior high school

3.3. Calculation Gain Value

$$Gain(A) = entropy(S) - \sum_{i=1}^k |S_i| \times entropy(S_i) \quad (2)$$

Gain value for attribute age <=30
 =0.749595257-
 (5/14*0.721928095)+(4/14*0.811278124)+(5/
 14*0.721928095)

Calculation gain value continues until Attribute The senior high school

The final result of the above calculation gets the highest gain value contained in the education attribute of 1.33993. Then the calculation process is continued based on the dataset centered on the educational attributes, the dataset displayed in Figure 3.

PENDIDIKAN SMA					
Penduduk	UMUR	Penghasilan	Status Pekerjaan	PENDIDIKAN	Kriteria
1	>40	Sedang	Pegawai	SMA	Produktif
2	>40	Rendah	Buruh	SMA	Non Produktif
3	31..40	Rendah	Buruh	SMA	Non Produktif
4	<=30	Sedang	Wirausaha	SMA	Produktif
5	<=30	Rendah	Buruh	SMA	Non Produktif
6	>40	Tinggi	Wirausaha	SMA	Produktif
7	<=30	Sedang	Pegawai	SMA	Produktif
8	31..40	Sedang	Wirausaha	SMA	Produktif
9	>40	Sedang	Pegawai	SMA	Produktif

Figure 3. Dataset Senior High School

The dataset above is processed based on the highest gain value, namely the education parameter with high school values, and produces two decisions, namely six productive and three non-productive then the process of calculating the entropy value and gain value is based on Figure 3, the calculation process can be seen in Figure 4

Attribut		Jumlah Kasus	Prediksi		Rumus	
Item	Nilai	S	Produktif	Non Produktif	ENTROPY	GAIN
Pendidikan		9	6	3	0.918295834	
Umur	<=30	3	2	1	0.918295834	1.194988
	31 .. 40	2	1	1		1
	>40	4	3	1	0.811278124	
Penghasilan	Tinggi	1	1	0	0	0.918296
	Sedang	5	5	0	0	
	Rendah	3	0	3	0	
Status Pekerjaan	Wirausaha	3	3	0	0	0.918296
	Pegawai	3	3	0	0	
	Buruh	3	0	3	0	

Figure 4. Highest Gain Calculation Results

From the above calculation results, the highest Gain is age. The following process calculates the entropy and Gain

values based on the age attribute, and the dataset display for the age parameter <= 30 is shown in Figure 5.

Umur <=30					
Penduduk	UMUR	Penghasilan	Status Pekerjaan	PENDIDIKAN	Kriteria
1	<=30	Sedang	Wirausaha	SMA	Produktif
2	<=30	Rendah	Buruh	SMA	Non Produktif
3	<=30	Sedang	Pegawai	SMA	Produktif

Figure 5. Dataset attribute age <=30

Based on Figure 5, the age <= 30 dataset produces two decisions, namely two productive and one non-productive, as shown in Figure 6.

Attribut		Jumlah Kasus	Prediksi		Rumus	
Item	Nilai	S	Produktif	Non Produktif	ENTROPY	GAIN
Umur		3	2	1	0.918295834	
<=30						
Penghasilan	Sedang	2	2	0	0	0.918296
	Rendah	1	0	1	0	
Status Pekerjaan	Wirausaha	1	1	0	0	0.918296
	Pegawai	1	1	0	0	
	Buruh	1	0	1	0	

Figure 6. Result Calculation for Attribute Age <=30

The calculation results above do not get the highest gain value, so the process stops. The display of the dataset with the age attribute 31..40 is shown in Figure 7.

Umur 31..40					
Penduduk	UMUR	Penghasilan	Status Pekerjaan	PENDIDIKAN	Kriteria
1	31..40	Rendah	Buruh	SMA	Non Produktif
2	31..40	Sedang	Wirausaha	SMA	Produktif

Figure 7 Dataset The age >=31 and <=40

Based on the dataset age 31 until 40, two decisions were obtained, namely one productive and one non-productive, then the calculations were carried out as shown in Figure 8.

Attribut		Jumlah Kasus	Prediksi		Rumus	
Item	Nilai	S	Produktif	Non Produktif	ENTROPY	GAIN
Umur		2	1	1	1	
31..40						
Penghasilan	Sedang	1	1	0	0	1
	Rendah	1	0	1	0	
Status Pekerjaan	Wirausaha	1	1	0	0	1
	Buruh	1	0	1	0	

Figure 8. Result of Calculation Attribute Age

≥ 31 and ≤ 40

The calculation results above do not get the highest gain value, so the process stops. Furthermore, the dataset age > 40 is shown in Figure 9 with two decisions, namely three productive and one non-productive.

Umur >40					
Penduduk	UMUR	Penghasilan	Status Pekerjaan	PENDIDIKAN	Kriteria
1	>40	Sedang	Pegawai	SMA	Produktif
2	>40	Rendah	Buruh	SMA	Non Produktif
3	>40	Tinggi	Wirausaha	SMA	Produktif
4	>40	Sedang	Pegawai	SMA	Produktif

Figure 9. Dataset Age >40

Based on the dataset above, it produces two decisions, namely three productive and one non-productive then the calculation process is shown in Figure 10.

Attribut		Jumlah Kasus	Prediksi		Rumus	
Item	Nilai	S	Produktif	Non Produktif	ENTROPY	GAIN
Umur	>40	4	3	1	0.811278124	
Penghasilan	Tinggi	1	1	0	0	0.811278
	Sedang	2	2	0	0	
	Rendah	1	0	1	0	
Status Pekerjaan	Wirausaha	1	1	0	0	0.811278
	Buruh	1	0	1	0	
	Pegawai	2	2	0	0	

Figure 10. result calculation for attribute the age ≥ 40 .

The calculation results above do not get the highest gain value, so the stop process.

3.4. Node divides the Highest Gain value

Based on the gain value calculation above, the branch nodes can be described in Figure 11 below.

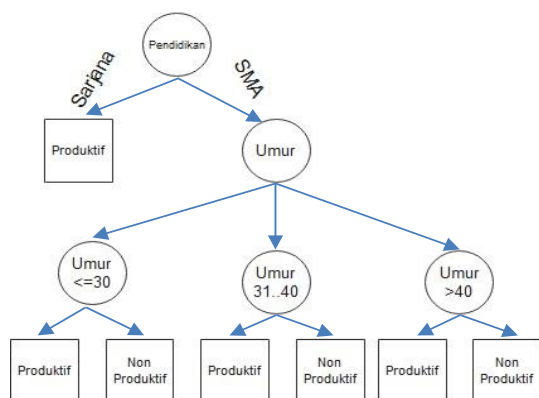


Figure 11. Node divides Gain value.

From the results of the calculation of the dataset in Figure 1, it is obtained that the highest gain value is obtained at the attribute of undergraduate education with one decision, namely productive, and high school education with the results of 2 decisions, namely six productive and three non-productive, because there are two decisions, so it is necessary to calculate entropy and gain on attributes of high school education so that the calculation results produce the highest gain value on the age attribute, each category produces two decisions, namely productive and non-productive. So that the following process is recalculated, however, the gain value produces the same value so that the process is terminated or completed.

3.5. Final Reading of Rules

The rule obtained from the steps mentioned above is:

1. If the population is aged ≤ 30 s/d >40 and has medium and high income and employment status as entrepreneur and employee and has a bachelor's degree, the result is productive.
2. If age ≤ 30 s/d >40 and medium and high income and employment status of the employee and self-employed and high school education, the result is productive.
3. If age ≤ 30 to >40 and low income and employment status as laborer and high school education, the result is non-productive.
4. If age ≤ 30 and moderate-income and employment status of employees and entrepreneurs and high school education, the result is productive.
5. If age ≤ 30 and medium and low income and employment status as laborer and high school education, the result is non-productive.
6. If the age is between 31..40, and medium-income and employment status is entrepreneur and high school education, the result is productive.
7. If the age is between 31..40 and, the income is low, and the employment status is labor and high school education, then the result is non-productive.
8. If age >40 and medium and high income and employment status of employees and entrepreneurs and high school education, the result is productive.

9. If age >40 and low income and employment status as laborer and high school education, then the result is non-productive.

The application of the ID3 algorithm in building decision trees using the attributes "entropy" and "information gathering" which are the basic components of the classification model really helps the government in generating information to find out non-productive people, so that the government can focus on providing assistance and teaching to be more creative in develop the community's economy, so automatically the results of the information from the implementation of the ID3 algorithm will become the focus of the government in alleviating the poor. In this study the authors created a system that can apply the calculation of the ID3 algorithm, so that when the admin inputs data, information from the application of the ID3 algorithm will automatically display information on productive or non-productive people. from the results of this information, the government can collect data on non-productive people to then provide follow-up in developing the community's economy by routinely conducting training in creating an economic creativity.

CONCLUSION

The application of the ID3 algorithm in poverty alleviation is to generate information on productive and non-productive people by classifying data, based on the dataset (initial data) this study can perform data calculations based on the amount of available data, the number of predicted data on productive people and the number of non-productive people then calculate the value entropy and gain value, after getting the highest gain value from the initial data process, then attribute 1 with the highest gain value will be used as the initial value calculation like the first process, after getting attribute 2 with the highest gain value, then it will proceed to all the attributes until the last attribute, then from the results of the gain calculation it can produce information in the form of rules.

Based on the description above, the researchers applied the Iterative Dichotomiser Three (ID3) algorithm in classifying productive and non-productive communities. By calculating the value of entropy and Gain based on Microsoft Excel. The results of applying this algorithm can be processed in a shorter time, with shorter decision rules and higher

prediction accuracy by displaying the highest gain value. The parameters used to consist of education, age, income, and employment status, which results in the following rule if higher education and high income, then the result is a productive society, whereas if high school education and low income, then the result is a non-productive society.

In the future the author will implement the ID3 algorithm into the system which will make it easy for public or admin data processors. Inputting community data into the system will immediately display information that the community's input is in the productive or non-productive category.

REFERENCES

- [1] R. Aditya Sari, "Stimulus Kelompok Orangtua Siswa SD Ar Rasyid dalam Menciptakan Masyarakat Produktif Ekonomi," *Din. J. Pengabd. Kpd. Masy.*, vol. 5, no. 3, pp. 715–723, 2021, doi: 10.31849/dinamisia.v5i3.4782.
- [2] S. D. Utami, H. Hunaepi, S. N. Primawati, A. Imran, and S. R. Fajri, "Pemberdayaan Kelompok Masyarakat Non Produktif Melalui Budidaya Jamur Tiram Di Desa Darek," *BAKTIMAS J. Pengabd. pada Masy.*, vol. 2, no. 1, pp. 15–21, 2020, doi: 10.32672/btm.v2i1.2100.
- [3] Y. Wang, Y. Li, Y. Song, X. Rong, and S.Zhang, "Improvement of ID3 algorithm based on simplified information entropy and coordination degree," *Algorithms*, vol. 10, no. 4, pp. 1–18, 2017, doi: 10.3390/a10040124.
- [4] F. Arif, N. Suryana, and B. Hussin, "Cascade Quality Prediction Method Using Multiple PCA+ID3 for Multi-Stage Manufacturing System," *IERI Procedia*, vol. 4, pp.201–207, 2013, doi:10.1016/j.ieri.2013.11.029.
- [5] B. N. Theogene Uwizeyimana et al., "The Multi Sectorial Approach to COVID-19 Pandemic in Limited-Resource Settings: Discussing Rwandan Experience," *Jphi*, vol.3, no.4,p.14, 2021, doi: 10.14302/issn.2641.
- [6] K. Devasenapathy and S. Duraisamy, "Evaluating the Performance of Teaching Assistant Using Decision Tree ID3 Algorithm," *Int. J. Comput. Appl.*, vol.

- 164, no.7, pp.23–27, 2017, doi: 10.5120/ijca2017913658.
- [7] N. F. Idris and M. A. Ismail, "Breast cancer disease classification using fuzzy-ID3 algorithm with FUZZYDBD method:Automatic fuzzy database definition," *PeerJ Comput. Sci.*, vol. 7, pp. 1–22, 2021, doi: 10.7717/PEERJ-CS.427.
- [8] L. Ju, L. Huang, and S. Tsai, "Online Data Migration Model and ID3 Algorithm in Sports Competition Action Data Mining Application," vol. 2021, 2021.
- [9] J. Shi, Q. He, and Z. Wang, "GMM clustering-based decision trees considering fault rate and cluster validity for analog circuit fault diagnosis," *IEEE Access*, vol.7, pp.140637–140650, 2019, doi: 10.1109/ACCESS.2019.2943380.
- [10] N. Bedregal-alpaca, V. Cornejo-aparicio, J. Zárate-valderrama, and P. Yanquechuro, "Classification models for determining types of academic risk and predicting dropout in university students 2020 *International Journal of Advanced Computer Science and Applications* pdf," vol.11, no.1, pp. 266–272, 2020.