NEW DT + RF en inggris.

by _ _

Submission date: 01-Jul-2024 10:14AM (UTC+0530)
Submission ID: 2407070741
File name: NEW_DT_RF_en_inggris.pdf (499.64K)
Word count: 3908
Character count: 19940

APPLICATION OF DECISION TREE METHOD IN DIABETES MELLITUS DISEASE CLASSIFICATION

Nofa Auliyatul Maulidiyyah*[1] , Trimono[2] , Aviolla Terza Damaliana[3]

[1]National Development University "Veteran" East Java
[2]National Development University "Veteran" East Java
[3]National Development University "Veteran" East Java
Email: [1]20083010029@student.upnjatim.ac.id,[2]trimono.stat@upnjatim.ac.id, [3]aviolla.terza.sada@upnjatim.ac.id

(Manuscript submitted: dd mmm yyyy, accepted for publication: dd mmm yyyy)

Abstract

Diabetes mellitus is a deadly disease caused by the failure of the pancreas to produce enough insulin. Indonesia is ranked fifth in the world with the number of people with diabetes in 2021 around 19.47 million and will continue to increase. This research uses a machine learning approach to compare the Decision Tree and Random Forest methods in classifying type 1 and type 2 diabetes mellitus. The goal is to improve the accuracy and reliability of classification models to support the diagnosis and management of this disease. Calculation of prediction results is using a confusion matrix. The test results show that the Decision Tree model outperforms Random Forest with an accuracy of 93% compared to 91%. The Decision Tree model also has a lower number of prediction errors, namely 7 errors, compared to Random Forest which has 9 errors. In addition, the most influential variables in the classification results are also different between the two models. The conclusion is that Decision Tree can be a better choice to improve accuracy and reliability in DM diagnosis and control, by adjusting patient treatment based on more precise predictions.

Keywords: diabetes mellitus, machine learning, decision tree, random forest, confusion matrix

1. INTRODUCTION

Diabetes mellitus is a chronic metabolic disorder when the patient's body does not produce enough insulin or the patient's body cannot utilize insulin properly, causing excessive blood sugar in the body, which often causes complications in the body's organs [1]. Types of diabetes mellitus are classified into type 1, type 2, and gestational based on elevated blood sugar levels caused by autoimmune and lifestyle factors. However, type 1 and type 2 are the most common ones [2].

In Indonesia, there were 19.47 million diabetic patients in 2021, showing an annual increase in the number of people with diabetes. Data collected by the International Diabetes Federation (IDF) shows that the number of people with diabetes in Indonesia is expected to increase to 23.32 by 2030 [3]. Recent research shows that 40% of type 1 diabetes cases in adults over 30 years old are often misdiagnosed as type 2 diabetes. One of the main problems is the assumption that adults are more likely to have type 2 diabetes, and some adults with type 1 diabetes may not require insulin at diagnosis, leading to symptoms similar to type 2 diabetes. This misdiagnosis results in inappropriate treatment, negatively impacting the patient's quality of life and survival [4].

Prevention of diabetes mellitus is needed with medical personnel and mathematical quantitative models, to prevent DM with early detection and encourage a healthy lifestyle [5]. The purpose of this model approach is to use machine learning. Machine learning is a type of Artifical Intelligence (AI) that uses data and algorithms to mimic the way humans learn. The goal is to improve the accuracy and precision of predictions. One of the common techniques used in machine learning is classification.

Classification is the process of creating functions or models that describe classes of data or concepts to predict the class of objects whose labels have not yet been obtained [1]. In this study, classification techniques are used to predict patients suffering from type 1 and type 2 diabetes. The classification process can be done with several algorithms, namely Decision Tree and Random Forest. Decision Tree is a prediction model for a decision using a tree or hierarchy structure and looking for a problem solution by making criteria as interconnected nodes that form like a tree structure. Each tree has branches, these branches represent all the properties that must be fulfilled to grow to the next branch until the branches end up in the leaves [6]. Random Forest is a development of the Classification and Regression Trees (CART) method that uses bootstrap aggregating (bagging) techniques and random feature selection to form a collection of decision trees to classify data [7]. This method is effective because it can overcome overfitting and work quickly on large datasets, but tends to overfitting on unbalanced datasets [8].

Research conducted by [9] shows that the Decision Tree method in classifying best-selling products (private data) obtained a result of 90% and an AUC value of 0.709, this value is included in Good Classification. [10] showed that Random Forest achieved 87.20% accuracy in the classification of infectious diseases.

This research compares Decision Tree and Random Forest in classifying diabetes mellitus. The purpose of this study is to determine the extent to which Decision Tree and Random Forest can improve the accuracy and reliability of classification models when assisting the diagnosis and control of diabetes mellitus.

2. RESEARCH METHODS

2.1 Decision Tree

A Decision Tree is a tree structure where each node represents the attribute being tested, each branch shows the result of the test, and each leaf node represents a specific class group. The highest level node of the decision tree is called the root, which is usually the attribute with the greatest influence on a particular class. In general, Decision Tree uses a top-down search strategy to find a solution. In the classification process, attribute values are tested by following the path from the root node to the leaf nodes, then the new class is determined based on the results [9]. Decision Tree is used to classify a data sample that has no known class into existing classes.

The stages in building a decision tree are as follows [1]:

1. Selecting an attribute as the root
2. Create a branch for each value of the attribute
3. Division of cases into each branch
4. Repeating this process on each branch until all cases in each branch have the same class to determine the attribute as the root, the attribute with the highest gain value of all existing attributes is selected. This gain value is calculated using the following equation:

$$\text{Gain}(S, A) = \text{Entropi}(S) - \sum_{i=1}^{n} - \left|\frac{S_i}{S}\right| \times \text{Entropi } S_i \quad (1)$$

Description:
S = states the set of cases
A = state attribute
n = states the number of partitions of attribute A
$|S_i|$ = states the number of cases in the i-th partition
$|S|$ = stated number of cases in s

Meanwhile, to produce the Entropy value with the following formula:

$$\text{Entropy}(S) = \sum_{i=1}^{k} - P_i \log_2 P_i \quad (2)$$

Description:
S = states the set of cases
k = states the number of partitions S
P_i = expresses the probability obtained from the total number of samples.

The following are some of the general characteristics of a Decision Tree [11]:

1. Decision Tree is a nonparametric approach to building classification models.
2. The techniques used to build the Decision Tree allow for rapid model building from a large training set.
3. Decision trees with small tree sizes are relatively easy to interpret.
4. Decision Tree provides an expressive overview in learning discrete value functions
5. Decision Tree is quite resistant to noise, especially for methods that can handle the problem of overfitting.

2.2 Random Forest

Breiman introduced Random Forest in 2001. Random Forest has two main objectives in solving problems, namely to perform classification and regression using several decision trees [12]. Random Forest is an ensemble method that improves classification accuracy by combining several classification methods. Random Forest consists of a set of decision trees, and the more decision trees used, the stronger the Random Forest algorithm. In each decision tree, the data starts from the root and moves to the leaf to determine the class or predicted value of the data [13]. The number of decision trees in Random Forest affects the accuracy of the overall random structure.

Random Forest works by building many decision trees and generating predictions in the form of class modes or averages of individual decision tree results. The Random Forest concept combines multiple random decision trees into a model because the number of decision trees affects the overall accuracy and stability of the model [14]. [15]said in the decision tree, the process starts by calculating entropy to assess the degree of impurity of the attribute, and then the information gain value is used to determine which attribute will be used to perform data splitting.

The following is how the Random Forest algorithm works including [10]:

1. Determines the number of trees (k) selected from the total features m, where k is less than m.
2. N random samples are taken from the dataset for each tree.

3. In each tree, p subsets of predictors are randomly picked, where m < p, p is the number of predictor variables.
4. The process in the second and third steps is repeated until k trees are reached.
5. The prediction result is obtained from the most votes from the classification results of the k tree.

The advantages of this Random Forest Algorithm are that it can produce relatively low errors, has good classification performance, and is suitable for large data.

2.3 Confusion Matrix

Confusion matrix table that displays four different combinations of predicted and actual values. Each row of the matrix shows the actual data classification and the predicted classification, or vice versa [16].

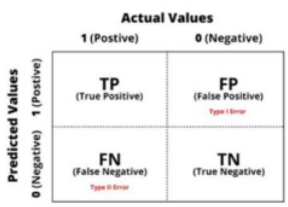

Figure 1. Confusion Matrix

Confusion matrix has four terms as classification results. True Positive (TP) is the number of correctly classified positive data, True Negative (TN) is the number of correctly classified negative data; False Positive (FP) is negative data but considered as positive data; and False Negative (FN) is positive data but considered as negative. There are five indicators: accuracy, precision, recall, and F1-Score.

$$accuracy = \frac{TP+TN}{TP+FP+FN+TN} \quad (3)$$

$$recall = \frac{TP}{TP+FN} \quad (4)$$

$$Precision = \frac{TP}{TP+FP} \quad (5)$$

$$f1 - score = 2.\frac{1}{\frac{1}{recall}+\frac{1}{precision}} \quad (6)$$

2.4 Description of Data and Research Steps

The data used in this study came from patients suffering from diabetes mellitus at Bungah Health Center, Bungah District, Gresik Regency. The data consisted of 513 with 10 variables. From this data, there are 341 patients with type 2 diabetes and 172 patients with type 1 diabetes.

Table 1. Research Variables

Variables	Data	Description
X_1 (Age)	1 : <20 Years 2 : 20-40 Years 3 : >40 Years	Information about the patient's age
X_2 (GDA)	Numerical	Random blood sugar
X_3 (GDP)	Numerical	Patient's blood sugar level during fasting
X_4 (Blood Sugar 2 HourPP)	Numerical	Postprandial or post-meal blood sugar levels
X_5 (Hba1C)	Numerical	Blood HbA1c levels (long-term indicator of blood sugar control)
X_6 (BMI)	Numerical	Body mass index of the patient
X_7 (Physical Activity)	0 : Lightweight 1 : Weight	Physical activity level or lifestyle of the patient
X_8 Systolic	Numerical	Blood pressure when the heart pumps blood
X_9 Diastolic	Numerical	Blood pressure during heart relaxation
Y	1 : Type 1 diabetes 2 : Type 2 diabetes	Diagnosis results based on doctors

Figure 2 shows the flow of the research methodology used to achieve the research analysis objectives. The process starts with data collection, pre-processing, the use of SMOTE to handle unbalanced data, and then classification using Random Forest.

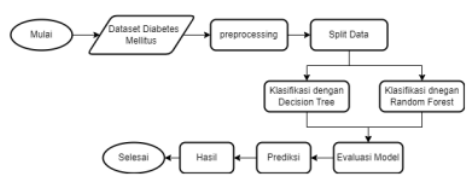

Figure 2. Research Flowchart

1. Diabetes Mellitus Dataset
   This study used secondary data from Bungah Community Health Center, Bungah District,

Gresik Regency. Data included medical records of patients with diabetes mellitus. The variables analyzed included age, random blood sugar, fasting blood sugar, 2-hour PP blood sugar, HbA1C, BMI, physical activity, systolic, and diastolic blood pressure.

2. Preprocessing
   Data preprocessing includes checking for blanks and duplicates, separating systolic and diastolic blood pressure, and encoding the diagnosis from text to numeric.
3. Split Data
   The data is divided into two parts, namely training and testing data. In this research, training data is used for the classification process using the Random Forest method. After that, testing data is used to evaluate the performance and final results of the model. In this study, the data is divided into 80% for training data and 20% for testing data.
4. Decision Tree and Random Forest Classification
5. Model evaluation
   At this stage, the performance of the Decision Tree and Random Forest algorithm classification models is evaluated. In this stage, it is seen based on the accuracy, precision, recall, and f1-score values.
6. Prediction
   At this stage, the trained model will produce a prediction stating whether the patient belongs to type 1 or type 2 diabetes.
7. Results
   At this stage, it determines how well the model can predict the diabetes status of the patient and to make decisions based on the prediction results.

3. RESULTS AND DISCUSSION

3.1 Diabetes Mellitus Dataset

This study uses medical record data of diabetes mellitus patients from one of the health centers in Gresik. The data were classified into two types, namely type 1 and type 2 diabetes. A total of 513 data were collected with 9 dependent variables and 1 independent variable. The data consisted of 341 patients with type 2 diabetes mellitus and 172 with type 1 diabetes mellitus, as shown in the figure 3.

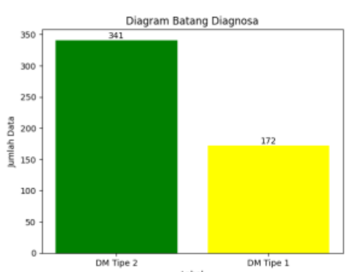

Figure 3. Patient Diagnosis Type Bar Chart

3.2 Preprocessing

In this step, checking for empty data, duplicates, and the encoding process is carried out by using the LabelEncoder function from the Scikit-Learn library to convert the diagnosis variable which was originally text into numeric, because Random Forest only accepts data in numeric form.

Table 2. Data after Preprocessing Stage

No.	Age	GDA	GDP	Blood Sugar 2 Hours PP	Systolic	Dias-tolik	HbA1c	BMI	Physical Activity	Diagnosis
1	64	93	102	220	175	96	8.5	33.6	1	1
2	49	108	115	180	150	70	7.0	26.6	1	0
3	56	186	126	300	174	96	9.0	23.3	1	1
4	57	255	137	250	110	60	8.0	28.1	1	1
5	42	292	148	190	104	63	7.5	43.1	1	1

After the encoding process is complete, in the Physical Activity column, patient data with physical activity (Yes) becomes 1, while patient data without Physical Activity (No) becomes 0. In the diagnosis column, type 1 diabetes miletus is represented with 0, and type 2 diabetes miletus with 1.

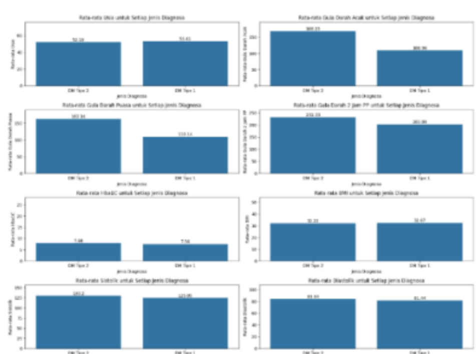

Figure 4. Bar Chart of Average Features per Diagnosis

The figure shows that the average of each feature for type 2 diabetes mellitus is higher than that of type 1. In addition to the bar chart, a heatmap of the correlation between features and targets can also show the features that most influence the diagnosis. The visualization of this heatmap can be seen in Figure 5.

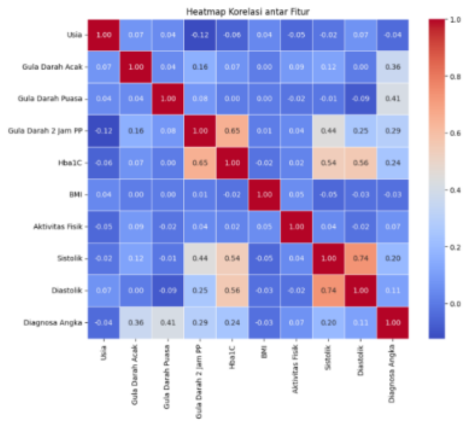

Figure 5. Correlation between Features and Diagnosis

The figure shows the correlation between features and diagnoses. The higher the correlation value, the greater the influence of the feature on the patient's diagnosis. The most influential features can be seen in Figure 6

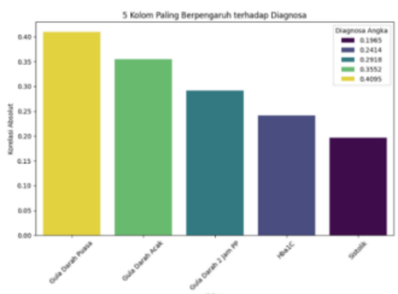

Figure 6. Features Most Influential to Diagnosis

The figure shows the features that have the most influence on patient diagnoses are fasting blood sugar, random blood sugar, 2-hour PP blood sugar, Hba1C, and Systolic.

3.3 Split Data

At this stage, the data is divided into training data and testing data with a ratio of 80% and 20%. Data division is done using the train_test_split function from the Scikit-Learn library. The results can be seen in Table 3.

Table 3. Split Data Desicion Tree an Random Forest

Type of Diabetes	Decision Tree Data		Random Forest Data	
	Train	Test	Train	Test
Type 1	137	35	137	35
Type 2	273	68	273	68

3.4 Decision Tree Classification

At this stage, training data is used to train the Decision Tree and Random Forest models, to compare their performance. At this stage, the Decision Tree model is built using the Scikit-Learn library with parameters 'max_depth' 5, 'min_samples_split' 4, and 'random_state' 42. 'max_depth' is used to control the maximum depth or complexity of the decision tree. After that 'min_samples_split' is used to set the minimum number of samples required at a node for a split to occur.

3.5 Random Forest Classification

Random Forest uses the Scikit-Learn library with parameters 'n_estimators' of 100 and 'random_state' of 42. 'n_estimators' determines the number of decision trees in the model. The more 'n_estimators', the more complex the Random Forest model, improving its ability to learn complex patterns, but also extending the training and prediction time. Increasing 'n_estimators' generally improves the performance of the model in terms of generalization, but should be chosen carefully to avoid overfitting. 'random_state' controls randomization when building the tree, ensuring reproducible results, which is important for sharing data and comparing models.

3.6 Model Evaluation

After training the model with training data, its performance is tested using testing data using confusion matrix. The value of the confusion matrix is used to calculate accuracy, precision, recall, and f1-score. This performance calculation process uses the Scikit-Learn library.

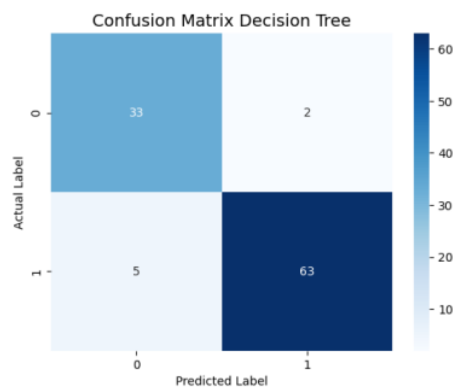

Figure 7. Confusion Matrix Desicion Tree

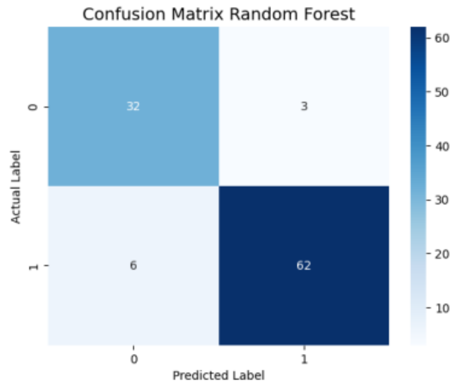

Figure 8. Confusion Matrix Random Forest

Based on the confusion matrix image above, the Decision Tree model shows better performance in predicting type 2 diabetes mellitus (1) compared to Random Forest. The Decision Tree model has the number of type 1 diabetes mellitus prediction errors (0) as many as 2 cases of False Positive (FP) and type 2 diabetes mellitus prediction errors (1) as many as 5 cases of False Negative (FN). Meanwhile, Random Forest has 3 FP cases, and 6 FN cases. Overall, Decision Tree has higher accuracy and lower number of prediction errors.

Based on the confusion matrix results, Table 4 below shows the calculation results of the performance generated by the model.

Table 4. Model Performance Comparison

Metrics	Decision Tree	Random Forest
F1-Score	93%	91%
Accuracy	93%	91%
Recall	93%	91%
Precision	93%	91%

The model trained with Decision Tree shows better performance with a 2% increase in performance value compared to Random Forest. Table 5 below is an example of data used in testing and the prediction results of each model.

Table 5. Model Prediction Results on Sample Data

No.	Age	GDA	GDP	Blood Sugar 2 Hours PP	HbA1c	BMI	Physical Activity	Systolic	Diastolic	Decision Tree	Random Forest	Actual
1	52	115	85	178	7.14	24	0	131	81	Type 2 DM	Type 2 DM	1
2	33	130	95	195	7.5	24.7	1	130	80	Type 2 DM	Type 2 DM	1
3	70	195	140	330	7	43.3	1	119	68	Type 1 DM	Type 1 DM	0
4	58	109	77	190	8.2	35.7	1	130	90	Type 1 DM	Type 2 DM	1
5	39	95	193	200	7	34.9	0	120	70	Type 2 DM	Type 2 DM	1

The models trained using Decision Tree and Random Forest are both good, with more than 90% accuracy. Both models can correctly classify diabetic patients based on medical records. The training results of the Decision Tree and Random Forest models have the weight of each feature, can be seen in the figure below.

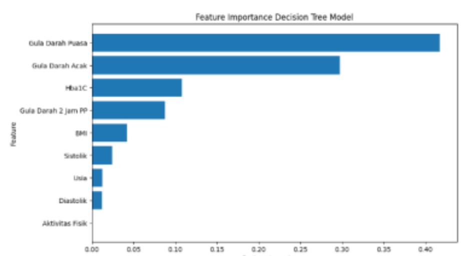

Figure 9. Feature Weights of Decision Tree Model

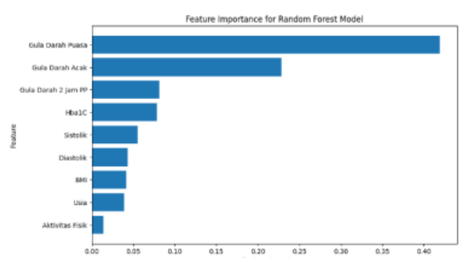

Figure 10. Feature Weights of Random Forest Model

3.7 Discussion

In testing the Decision Tree and Random Forest models, the model trained with Decision Tree showed higher accuracy than the Random Forest model. The performance value of the Decision Tree and Random

Forest models is 2% higher for each matrix, reaching 93% compared to 91%, as shown in Figure 11.

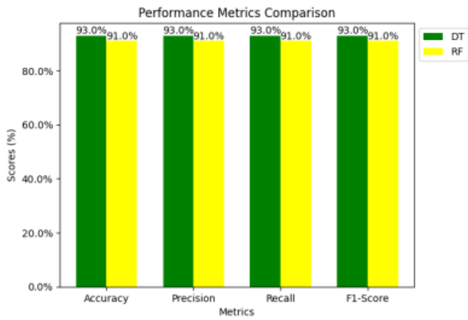

Figure 11. Comparison of Testing Performance Results

Data using the Decision Tree model produced better accuracy The model trained with Drsicion Tree had fewer errors with 7 total predictions, 2 for type 1 diabetes mellitus and 5 for type 2 diabetes mellitus. While the Random Forest model without SMOTE had 9 total prediction errors, 3 for type 1 diabetes mellitus and 6 for type 2. This shows that training the model with balanced data can improve prediction performance.

4. CONCLUSIONS

The model trained with data through Decision Tree performed better, with a total of 7 prediction errors (2 for type 1 and 5 for type 2) compared to the Random Forest model which had 9 prediction errors (3 for type 1 and 6 for type 2). The Decision Tree model achieved 2% higher accuracy, precision, recall, and f1-score, 93% compared to 91%. In addition, the variables that most influence the classification results of the Decision Tree and Random Forest models are different. In the Decision Tree results, the 5 most influential factors are GDP, GDA, Hba1c, Blood Sugar 2 Hours PP, AND BMI. While in the Random Forest model the 5 most influential factors are GDP, GDA, Blood Sugar 2 Hours PP, Hba1C, and Systolic.

5. REFERENCE

[1] F. M. Hana, "Klasifikasi Penderita Penyakit Diabetes Menggunakan Algoritma Decision Tree C4 . 5," *J. Sist. Komput. Kecerdasan Buatan*, vol. 4, no. 2, 2020.

[2] R. Marzel, "Terapi pada DM Tipe 1," *J. Penelit. Perawat Prof.*, vol. 3, no. 1, pp. 51–62, 2021, doi: 10.37287/jppp.v3i1.297.

[3] IDF, "Diabetes report 2000 — 2045," *Diabetes Atlas*, 2021. https://diabetesatlas.org/data/en/country/94/id.html

[4] S. P. Katongole, P. Akweongo, R. Anguyo, D. E. Kasozi, and A. Adomah-Afari, "Prevalence and Classification of Misdiagnosis Among Hospitalised Patients in Five General Hospitals of Central Uganda," *Clin. Audit*, vol. Volume 14, no. September, pp. 65–77, 2022, doi: 10.2147/ca.s370393.

[5] W. Nugraha and R. Sabaruddin, "Teknik Resampling untuk Mengatasi Ketidakseimbangan Kelas pada Klasifikasi Penyakit Diabetes Menggunakan C4.5, Random Forest, dan SVM," *Techno.Com*, vol. 20, no. 3, pp. 352–361, 2021, doi: 10.33633/tc.v20i3.4762.

[6] A. Tangkelayuk and E. Mailoa, "Klasifikasi Kualitas Air Menggunakan Metode KNN , Naïve Bayes Dan Decision Tree," vol. 9, no. 2, pp. 1109–1119, 2022.

[7] K. Siti, "KLASIFIKASI PENYAKIT DIABETES MENGGUNAKAN METODE DECISION TREE DAN RANDOM FOREST," *repository.unsri.ac.id*, no. 8.5.2017, pp. 2003–2005, 2022.

[8] E. Rosta *et al.*, "Klasifikasi Data Kesehatan Mental di Industri Teknologi Menggunakan Algoritma Random Forest," vol. 1, no. 3, pp. 237–253, 2023.

[9] A. Husna Nasrullah, "IMPLEMENTASI ALGORITMA DECISION TREE UNTUK KLASIFIKASI PRODUK LARIS," vol. 7, no. 2, pp. 45–51, 2021.

[10] A. Prabowo, S. Wardani, R. Wijaya Dewantoro, W. Wesly, and Leonardo, "Komparasi Tingkat Akurasi Random Forest dan Decision Tree C4 . 5 Pada Klasifikasi Data Penyakit Infertilitas," vol. 4, no. 1, pp. 218–224, 2023, doi: 10.30865/klik.v4i1.1115.

[11] L. Qadrini, A. Seppewali, and A. Aina, "DECISION TREE DAN ADABOOST PADA KLASIFIKASI PENERIMA PROGRAM BANTUAN SOSIAL," *J. Inov. Penelit.*, vol. 2, no. 7, 2021.

[12] M. L. Suliztia, "PENERAPAN ANALISIS RANDOM FOREST PADA PROTOTYPE SISTEM PREDIKSI HARGA KAMERA BEKAS MENGGUNAKAN FLASK," *dspace.uii.ac.id*, 2020.

[13] M. Aqsha and N. Sunusi, "PERFORMA KLASIFIKASI DATA TIDAK SEIMBANG DENGAN PENDEKATAN MACHINE LEARNING (STUDI KASUS : DIABETES INDIAN PIMA)," vol. 12, no. 2, pp. 176–193, 2023.

[14] F. Mu'alim and R. Hidayati, "Implementasi Metode Random Forest Untuk Penjurusan," vol. 14, no. 1, pp. 116–125, 2022.

[15] U. Azmi, "Pendeteksian Aroma Ganja Kering Menggunakan Algoritma Random Forest," *JITSI J. Ilm. Teknol. Sist. Inf.*, vol. 4, no. 1, pp. 28–33, 2023, [Online]. Available: https://jurnal-itsi.org/index.php/jitsi/article/view/104%0A https://jurnal-

itsi.org/index.php/jitsi/article/download/104/82
[16] A. Nurwalikadani, "Implementasi Algoritme Smote Dan Klasifikasi Random Forest Pada Imbalanced Data Metilasi Sequence Protein Lisin," 2022, [Online]. Available: http://digilib.unila.ac.id/67956/%0Ahttp://digilib.unila.ac.id/67956/3/SKRIPSI FULL TANPA PEMBAHASAN.pdf

NEW DT + RF en inggris.

ORIGINALITY REPORT

22%	17%	16%	8%
SIMILARITY INDEX	INTERNET SOURCES	PUBLICATIONS	STUDENT PAPERS

PRIMARY SOURCES

1	jeeemi.org Internet Source	2%
2	"38th EASD Annual Meeting of the European Association for the Study of Diabetes", Diabetologia, 2016 Publication	1%
3	Submitted to University of Malta Student Paper	1%
4	docplayer.net Internet Source	1%
5	www.ncbi.nlm.nih.gov Internet Source	1%
6	ejournals.umn.ac.id Internet Source	1%
7	Submitted to Universiti Teknikal Malaysia Melaka Student Paper	1%
8	Purushottam R. Patil. "The Application of Data Mining for Direct Marketing", 2009 Second	1%

International Conference on Emerging Trends in Engineering & Technology, 12/2009
Publication

9	Jepri Banjarnahor, Friska Sinaga, Dedi Setiadi Sitorus, Wahyu Adventus Andreas Sitanggang, Mardi Turnip. "Application of Decision Tree Method in ECG Signal Classification For Heart Disorder Detection", sinkron, 2024 Publication	1 %
10	Submitted to University of Bristol Student Paper	1 %
11	Luthfia Rahman, Noor Akhmad Setiawan, Adhistya Erna Permanasari. "Feature selection methods in improving accuracy of classifying students' academic performance", 2017 2nd International conferences on Information Technology, Information Systems and Electrical Engineering (ICITISEE), 2017 Publication	1 %
12	epos.myesr.org Internet Source	1 %
13	Tessy Badriyah, Ina Ratudduja, Intan P. Desy, Iwan Syarif. "Assessing Risk Prediction of Cervical Cancer in Mobile Personal Health Records (mPHR)", 2018 International Conference on Applied Information Technology and Innovation (ICAITI), 2018	1 %

Publication

14	Submitted to Universitas Khairun Student Paper	1 %
15	Submitted to University College London Student Paper	1 %
16	repository-tnmgrmu.ac.in Internet Source	1 %
17	Submitted to Universiti Malaysia Kelantan Student Paper	<1 %
18	Submitted to Malaviya National Institute of Technology Student Paper	<1 %
19	hdl.handle.net Internet Source	<1 %
20	Submitted to Trinity College Dublin Student Paper	<1 %
21	studentsrepo.um.edu.my Internet Source	<1 %
22	A Andriani, S Hartati. "Prognosis of Diabetes Mellitus with Transfer Learning-Based Naïve Bayes Method", Journal of Physics: Conference Series, 2021 Publication	<1 %
23	www.adlittle.com Internet Source	<1 %

24	Agus Setiyono, Windu Gata. "Emotional Programmer's Behavior to Respond Problems by Using a Decision Tree", 2019 International Conference on Information and Communications Technology (ICOIACT), 2019 Publication	<1 %
25	peerj.com Internet Source	<1 %
26	www.jmir.org Internet Source	<1 %
27	www.jstage.jst.go.jp Internet Source	<1 %
28	mro.massey.ac.nz Internet Source	<1 %
29	Badat, Na'eema. "Keratoconjunctivitis Sicca and Its Correlation with Diabetes Mellitus", University of Johannesburg (South Africa), 2021 Publication	<1 %
30	Eswari, Potharlanka P. B.. "Cardiac Manifestations of Diabetes Mellitus", Dr. NTR University of Health Sciences (India), 2021 Publication	<1 %
31	digilib.unila.ac.id Internet Source	<1 %

jurnal.fikom.umi.ac.id

32	Internet Source	<1%
33	www.coursehero.com Internet Source	<1%
34	www.scilit.net Internet Source	<1%
35	www.warf.org Internet Source	<1%
36	Oktariani Nurul Pratiwi, Yenie Syukriyah. "Question Classification for e-Learning Using Machine Learning Approach", 2019 International Conference on ICT for Smart Society (ICISS), 2019 Publication	<1%
37	archive.umsida.ac.id Internet Source	<1%
38	ecampus.pelitabangsa.ac.id Internet Source	<1%
39	ejournals.itda.ac.id Internet Source	<1%
40	journal.universitasbumigora.ac.id Internet Source	<1%
41	jurnal.atmaluhur.ac.id Internet Source	<1%
	jurnal.iaii.or.id	

42	Internet Source	<1 %
43	repository.unsri.ac.id Internet Source	<1 %
44	www.mdpi.com Internet Source	<1 %
45	Nurul Ria Amelia, Yaddarabullah, Budi Arifitama, Yessi Jusman. "Classification of toddler nutrition status using data mining with algorithm C4.5 (Case study: Integrated service centers in Ciasihan Village)", AIP Publishing, 2022 Publication	<1 %
46	Guzzardo, Frank. "Spatial Conflict Prediction with Machine Learning Conflict Vulnerability in the Sahel Region", Universidade NOVA de Lisboa (Portugal), 2024 Publication	<1 %
47	Olivares, Roberto Jose Luna. "Palm Tree Image Classification : a Convolutional and Machine Learning Approach", Universidade NOVA de Lisboa (Portugal), 2024 Publication	<1 %
48	Rizky Embun Arifah, Tituk Diah Widajantie. "THE The Effect Of Perceived Ease Of Use, Perceived Usefulness, Perceived Security, And Perceived Credibility Mobile Banking As A	<1 %

Digital Payment On The Financial Management Of Undergraduate Accounting Students At State Universities In Surabaya", Journal of Economic, Bussines and Accounting (COSTING), 2024

Publication

Exclude quotes	Off	Exclude matches	Off
Exclude bibliography	On		

NEW DT + RF en inggris.

PAGE 1

PAGE 2

PAGE 3

PAGE 4

PAGE 5

PAGE 6

PAGE 7

PAGE 8