Jurnal ELTIKOM :
Jurnal Teknik Elektro, Teknologi Informasi dan Komputer

Vol. 9, No. 2, December 2025, page. 107-114
ISSN 2598-3245 (Print), ISSN 2598-3288 (Online)
DOI: http://doi.org/10.31961/eltikom.v9i2.1461
Available online at http://eltikom.poliban.ac.id

CLASSIFICATION OF DIABETES MELLITUS USING DECISION TREES

Rinci Kembang Hapsari[1*], Abdullah Harits Salim[2], Leonardo Fahsi Oktavian[1], Aldy Ramadhan Fitra[1]
[1)] Faculty of Electrical and Information Technology, Institut Teknologi Adhi Tama Surabaya, Surabaya, Indonesia
[2)] Faculty of Science and Data Analytics, Institut Teknologi Sepuluh November, Surabaya, Indonesia
e-mail: rincikembang@itats.ac.id, sharits0022@gmail.com, leonardofahdioktavian@gmail.com, aldyramadhanf112@gmail.com

Received: 3 February 2025 – Revised: 13 June 2025 – Accepted: 1 July 2025

ABSTRACT

Diabetes Mellitus is a global health concern, with its prevalence and incidence rising sharply worldwide, including in Indonesia. Several factors contribute to the onset of diabetes mellitus, such as heredity, age, weight, and blood pressure. Managing blood sugar levels, maintaining a balanced diet, exercising regularly, and undergoing early screening when necessary are among the key measures to prevent and control this disease. Early diagnosis is essential to reduce both the number of cases and the associated risks. This study aims to detect diabetes mellitus using classification techniques. The method involves several subprocesses within the classification procedure. The first stage, data preprocessing, includes feature selection and data cleaning. The resulting preprocessed data are then used in the classification stage, specifically the learning subprocess, to generate a decision tree model. Model construction employs pruning, followed by training and performance evaluation. The study utilizes a diabetes dataset obtained from kaggle.com, consisting of 768 records. The dataset includes attributes such as Pregnancies, Glucose, Blood Pressure, Skin Thickness, Insulin, Body Mass Index (BMI), Diabetes Pedigree Function, Age, and the label Outcome. Testing was conducted using decision trees with maximum depths of 3, 5, 7, 10, and 15. The results show that the highest accuracy (88.56%) occurred at a maximum depth of 5, while the highest recall (100%) was achieved at a depth of 3. The highest precision (47.37%) and specificity (95.85%) were also obtained at a depth of 3.

Keywords*: classification, decision tree, diabetes mellitus, pre-processing.*

I. INTRODUCTION

DIABETES Mellitus is a genetically and clinically heterogeneous metabolic disorder characterized by impaired carbohydrate tolerance. It has become a major global health problem, with its prevalence and incidence rates rising sharply worldwide, including in Indonesia. According to the World Health Organization (WHO) [1], type 2 diabetes mellitus is the most common form of the disease, and its incidence is increasing rapidly, particularly in low-income countries. The WHO also identifies type 2 diabetes mellitus as a growing global health threat.

The International Diabetes Mellitus Federation (IDF) estimated that in 2019, approximately 463 million people aged 20–79 years worldwide were living with diabetes mellitus, representing a prevalence rate of 9.3% within that age group. Based on gender, the IDF reported that the prevalence in 2019 reached 9% among women and 9.65% among men. The prevalence is expected to rise with age, reaching 19.9%, or about 111.2 million individuals aged 65–79 years. Furthermore, the number of people with diabetes mellitus is projected to increase to 578 million by 2030 and 700 million by 2045 [2]. The rising prevalence of diabetes mellitus in Indonesia must be addressed through preventive efforts.

Early detection of diabetes is essential to reduce the risk of complications in the future. Advances in science and information technology offer new approaches for diabetes prevention. One such approach is data mining, which can serve as a useful reference. Data mining techniques such as classification [4]

Figure 1. Block diagram of steps in the research

can be applied for this purpose. Classification is a supervised learning method used to assign new data to predefined categories [5].

Many classification algorithms have been developed to detect diabetes. One example is the k-nearest neighbors (k-NN) algorithm, a “lazy” learning method that has been applied to classify diabetes cases. In one study, tests were conducted using k values of 3 and 5. The highest accuracy, 39%, was achieved when k=3. The precision value for both k=3 and k=5 was 65%, while the highest recall, 36%, occurred at k=3. The F-measure also reached its highest value of 46% at k=3 [6].

Another study employed the Random Forest algorithm to detect diabetes mellitus. The research used a dataset from *kaggle.com* titled *Early Stage Diabetes Mellitus Risk Prediction*, consisting of 520 records and 17 attributes. The results showed a precision of 100%, a recall of 84.38%, and an accuracy of 90.38% [7].

Classification techniques using tree structures have also been widely developed in the healthcare sector. These include decision tree algorithms, C4.5, random forest, and Chi-Square Automatic Interaction Detector (CHAID). Research on disease diagnosis and prediction using such classification techniques has increased significantly. For instance, one study applied a decision tree algorithm to classify liver disease using a liver disease dataset. The dataset was divided using the holdout validation method with a test size of 0.2 and k-fold cross-validation with a test size of 0.6. Data normalization was performed using either the min-max or standard scaler method. The results showed that testing with k-fold cross-validation and min-max normalization produced an accuracy of 0.7, while k-fold cross-validation with a standard scaler achieved an accuracy of 0.7333. Using holdout validation, the min-max method yielded an accuracy of 0.5342, whereas the standard scaler achieved 0.6027 [8].

In another study on predicting Alzheimer’s disease (AD) in elderly patients, researchers used the decision tree algorithm with a dataset of 150 subjects. The decision tree was pruned, and model validation was conducted using 10-fold cross-validation. Performance was evaluated based on accuracy, precision, and receiver operating characteristic (ROC) curve values. The results showed an accuracy of 88.7%, precision of 86.7%, and an ROC value of 91.8% [9].

The Decision Tree algorithm has also been used to examine correlations between age and disease. One study aimed to determine which diseases most affected each age group. The dataset included ten diseases: cholera (28,590 cases), diarrhea (28,550), Ebola (28,179), malaria (28,535), Marburg virus (28,430), measles (28,471), meningitis (28,362), rubella (28,539), dengue virus (28,401), and yellow fever (28,428). The results showed that disease outbreaks were most common among children aged 10 years and below [10].

The present study addresses the problem of detecting diabetes mellitus (DM) using classification techniques and investigates whether decision tree depth significantly affects classification accuracy in the early detection of diabetes. Previous studies have demonstrated that the Decision Tree algorithm is widely applied in the medical field and has shown strong performance. Therefore, this study aims to classify DM using the decision tree algorithm to support early identification of the disease.

II. RESEARCH METHOD

Data mining is the process of using statistics, mathematics, artificial intelligence, and machine learning to extract and summarize information for practical use [11]. It is defined as the process of discovering patterns in data. Based on its function, data mining can be categorized into description, estimation, prediction, classification, clustering, and association [12]. Data mining algorithms help explore data and facilitate information discovery [13]. However, as the data size increases, processing time also becomes longer. This study consists of three stages: data preprocessing, classification using the Decision Tree algorithm, and performance measurement of the system. The process flow is illustrated in Figure 1.

A. Dataset

This research uses *Indian Liver Patient Dataset* obtained from https://www.kaggle.com/datasets/akshaydattatraykhare/diabetes-dataset. The dataset contains data collected from patients in northeastern Andhra Pradesh, India, consisting of nine attributes and 768 records, with one attribute serving as the label. The attributes include Pregnancies, Glucose, Blood Pressure, Skin Thickness, Insulin, Body Mass Index (BMI), Diabetes Pedigree Function, Age, and the label Outcome.

B. Preprocessing

The dataset obtained must first be processed to ensure the accuracy of the information. In this study, two preprocessing steps were carried out: feature selection and data cleaning.

The purpose of feature selection is to reduce the number of features involved in determining the target class value. Features that are irrelevant or redundant are excluded. The main goal of feature selection is to choose the most relevant features from a dataset to improve the predictive model’s learning performance by increasing model accuracy, reducing computational cost or processing time, and enhancing model interpretability [14].

The next step is data cleaning, which is essential because the dataset may contain incomplete, irrelevant, or inaccurate data. Preprocessing also helps eliminate potential issues such as missing values [15]. In this study, data cleaning was performed by removing unnecessary data.

C. Decision Tree

Decision trees are one of the classification methods used in data mining. Classification is the process of developing a mathematical model to differentiate, identify, and explain data classes in order to predict unknown classes in a dataset [16]. It is a supervised learning method used to categorize information into predefined groups, enabling the creation of models that can classify large data populations [17].

A decision tree is a type of classification algorithm structured like a tree, consisting of roots, branches, and leaves [16]. The root node represents the attribute being tested, each branch represents a possible outcome, and each leaf node represents a class label [18]. The dataset's features correspond to the root node, the decision rules are represented by branch nodes, and each leaf node produces an output result [19].

Decision trees are constructed by organizing classifications in a hierarchical tree structure. The tree is divided into smaller subsets while simultaneously generating decision and leaf nodes [20]. Decision trees are widely used to produce reliable predictions. This method transforms large amounts of data into a structured decision tree that represents clear and interpretable rules [21], [22].

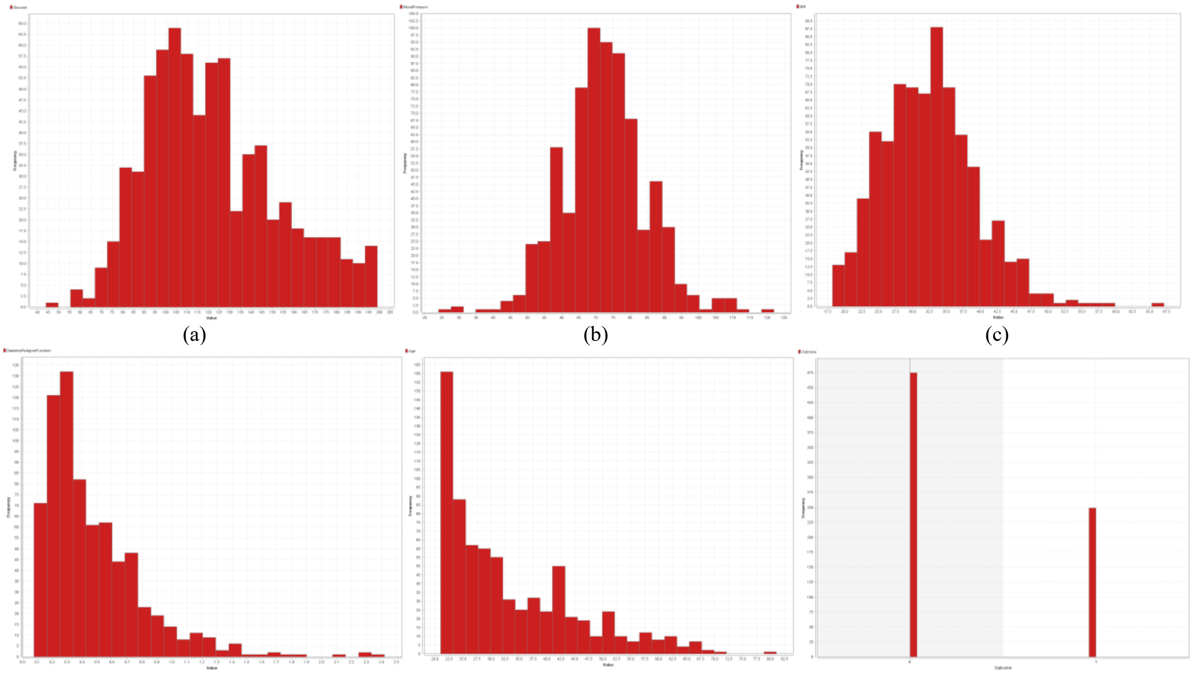
The Decision Tree algorithm divides data into subsets based on specific features, forming a structure where internal nodes represent features and edges denote possible outcomes or values. Its main objective is to improve the organization or purity of these subsets by minimizing impurity measures such as Gini impurity or entropy.

These measures assess the degree of disorder or unpredictability in a dataset. The algorithm determines the optimal feature and value combination to split the data at each node based on the reduction in impurity achieved. This process continues iteratively until a stopping criterion is met, such as reaching a maximum tree depth or having a minimum number of samples in a leaf node [23].

Pruning is performed on the decision tree to prevent overfitting. Based on a model constructed from training data, the dataset is recursively traversed to each leaf node. At this stage, the tree may become overly complex and fit the training data too closely, resulting in reduced accuracy on unseen data [18]. The mean squared error (MSE) of the dataset is then calculated. If the MSE decreases after pruning, the node is removed; otherwise, it is retained [24].

D. System Performance Measurement

Testing in this study was conducted using calculations derived from the confusion matrix, which evaluates the performance of the classification model [24]. The confusion matrix is used to calculate accuracy, recall, precision, and specificity values [25],[26],[27] and provides insight into the types of errors made by the classification model [28]. It consists of four possible outcomes [29],[28],[30]: False Positive (FP): negative data incorrectly predicted as positive, False Negative (FN): positive data incorrectly predicted as negative, True Positive (TP): positive data correctly predicted as positive, and True Negative (TN): negative data correctly predicted as negative.



(a) (b) (c)

(d) (e) (f)

Figure 2. (a) Histogram of the distribution of Glucose values; (b) Histogram of the distribution of Blood Pressure values; (c) Histogram of the distribution of BMI values; (d) Histogram of the distribution of Diabetes Pedigree Function values; (e) Histogram of the distribution of Age values; (f) Histogram of the distribution of Outcome values.

$$Accuracy = \frac{TF + TN}{TP + FN + FP + TN} \times 100\% \tag{1}$$

$$Sensitivity = \frac{TP}{(TP + FN)} \tag{2}$$

$$Precision = \frac{TP}{(TP + FP)} \tag{3}$$

$$Specificity = \frac{TN}{TN + FP} \tag{4}$$

The accuracy of the algorithm indicates how precisely the model’s predictions match the actual data. The closer the accuracy is to 1 or 100%, the better the model’s performance [31]. Accuracy is calculated using (1). The sensitivity or recall of an algorithm indicates how accurately the system identifies the positive class. In predictive analysis, sensitivity measures how many actual positive cases are correctly classified as positive. It represents the ratio of true positive predictions to the total number of actual positive cases, reflecting the model’s ability to detect positive instances. Sensitivity is calculated using (2). Precision measures how accurately an algorithm predicts positive results. It represents the proportion of true positive predictions among all predicted positive outcomes [32]. Precision indicates the system’s ability to produce relevant data and is computed as the ratio of true positive data to all data predicted as positive. Precision is calculated using (3). Specificity reflects how accurately the system identifies the negative class. It measures the proportion of correctly classified negative cases compared with the total number of actual negative cases, helping to evaluate the model’s ability to distinguish negative outcomes [33]. Specificity is calculated using (4). Sensitivity and specificity are key metrics for evaluating the detailed performance of a classification model [29].

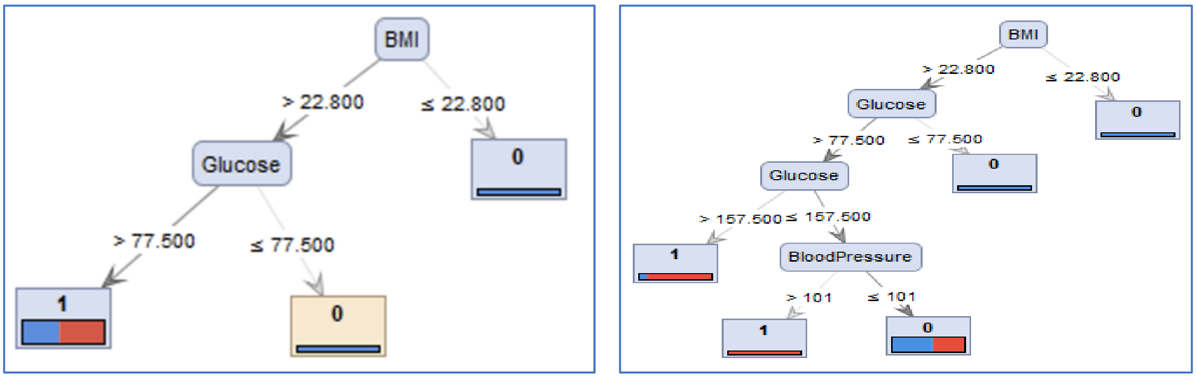
III. RESULTS AND DISCUSSION

The dataset in this study consists of 768 records. In the feature selection subprocess, six attributes were selected as relevant features: Glucose, Blood Pressure, Body Mass Index (BMI), Diabetes Pedigree Function, Age, and Outcome. In the subsequent data cleaning subprocess, incomplete data were removed, resulting in a final dataset of 724 records. The preprocessed dataset was then divided into two parts: training data and testing data. The training data consisted of 464 records, comprising 234 positive diabetes cases and 234 negative cases, while the testing data consisted of 260 records.

The distribution of glucose values is shown in Figure 2(a), ranging from 57 to 194, with a mode of 99. The distribution of blood pressure values, shown in Figure 2(b), ranges from 24 to 122, with a mode of 70. Figure 2(c) illustrates the distribution of BMI values, which range from 18.4 to 52.9, with a mode of 27.8.

Figure 2(d) shows the distribution of Diabetes Pedigree Function values, ranging from 0.085 to 2.42. The bars represent the frequency of each value. Figure 2(e) presents the distribution of age values, ranging from 21 to 70 years, with bars indicating the frequency of each age group. Figure 2(f) displays the distribution of the Outcome variable, which contains two values: 0 and 1, where 0 indicates non-diabetic and 1 indicates diabetic.

After completing the preprocessing stage, the next step was classification using the decision tree algorithm. The classification process consisted of two stages: training and testing. During training, the parameters used were *confidence* = 0.25, *min_gain* = 0.1, and *min_leaf_size* = 2. The learning process with a maximum depth of 3 produced the tree model shown in Figure 3(a), while Figure 3(b) presents the decision tree model with a maximum depth of 5.



(a) (b)

Figure 3. (a) Decision tree model with maximum depth = 3; (b) Decision tree model with maximum depth = 5.

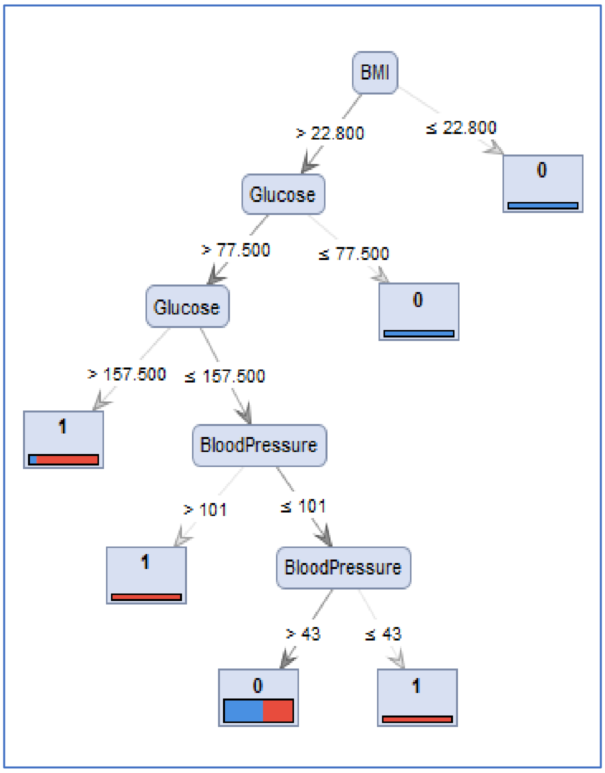


Figure 4. Decision tree model with maximum depth = 7.

The learning process with a maximum depth of 7 produced the tree model shown in Figure 4, and the model with a maximum depth of 10 is presented in Figure 5. The resulting decision tree models were used as test patterns for the testing dataset. The predicted values were determined by matching the decision rules with the dataset, and the outcomes were compared with the actual values to generate the confusion matrix. Based on the confusion matrix results, the accuracy, precision, recall, and specificity values were calculated, as shown in Table 1.

Based on Table 1, the lowest accuracy value was 23.99% at a maximum depth of 3, while the highest accuracy was 88.56% at a maximum depth of 5. The minimum recall value was 30% at maximum depths of 5, 7, and 10, and the highest recall was 100% when the maximum depth was 3. The lowest precision value was 12.71% at a depth of 3, and the highest was 47.37% at a depth of 5. The lowest specificity value was 14.52% at a depth of 3, while the highest was 95.85% at a depth of 5.

Based on the test results, a higher recall value indicates that the system is more capable of identifying positive classes that are truly positive, thereby improving prediction accuracy. However, the results of this study require further validation due to the unbalanced proportion of test data. The dataset contained 11% positive diabetes cases and 89% negative cases. To enhance system performance, future research should employ balanced training and testing datasets.

The results also show that when the maximum depth value is too small, the model becomes overly simple and less adaptable to the data's complexity. This leads to low accuracy because the model cannot capture the intricate relationships between features and the target variable. A small maximum depth may cause underfitting, where the model fails to recognize underlying patterns in the training data, resulting in poor performance in both training and testing phases. A simple model contributes to underfitting, while a highly complex model increases the risk of overfitting [29].

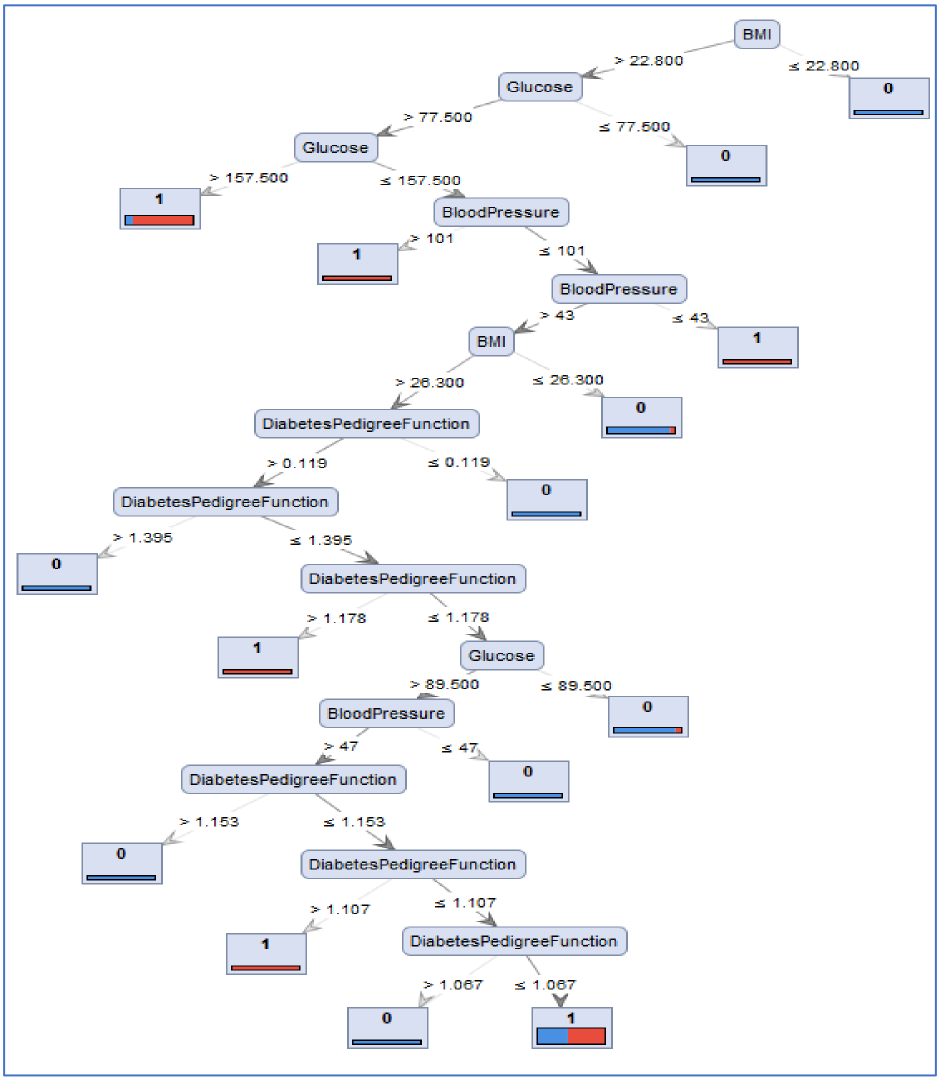


Figure 5. Decision tree model with maximum depth = 15.

TABLE 1
SYSTEM PERFORMANCE

Variable	Maximum Depth				
	3	5	7	10	15
Accuracy	23.99	88.56	87.45	87.08	49.08
Precision	12.71	47.37	40.91	39.13	17.47
Recall	100.00	30.00	30.00	30.00	96.67
Specificity	14.52	95.85	94.61	91.57	43.15

The results also reveal that when the minimum sample split value is too small, the tree tends to form branches for individual observations. This makes the model overly complex and sensitive to noise in the training data, reducing its ability to generalize patterns and lowering test accuracy. Conversely, if the maximum depth is too large, the model becomes overly complex and may capture noise in the data. A complex model leads to overfitting, reducing accuracy when applied to new data [18][34][35]. In this condition, the decision tree becomes excessively deep, with many irrelevant branches. Therefore, selecting an appropriate maximum depth is crucial to balance bias and variance in a decision tree [36].

When the minimum sample split value is too large, the tree stops dividing nodes even though useful information may still be extracted from the data. This results in an overly simple model that cannot capture data complexity, leading to underfitting.

IV. CONCLUSION

Based on the results of study using the Decision Tree algorithm, it can be concluded that after preprocessing through feature selection and data cleaning, the dataset used consisted of six attributes: Glucose, Blood Pressure, Body Mass Index (BMI), Age, Diabetes Pedigree Function, and Outcome, with a total of 724 records. The Decision Tree algorithm demonstrated optimal performance when the maximum depth was set to 5. The relatively low precision value was likely influenced by the imbalance in the test dataset.

REFERENCES

[1] WHO, "Cardiovascular diseases," *World Health Organization: Geneva, Switzerland*, 2017. https://www.who.int/news-room/fact-sheets/detail/cardiovascular-diseases-(cvds) (accessed Mar. 12, 2021).

[2] "Diabetes Mellitus," *Pusat Data Dan Informasi Kementrian Kesehatan Republik Indonesia*, 2020. https://pusdatin.kemkes.go.id/article/view/20111800001/diabetes-melitus.html

[3] I. Oktanisa and A. A. Supianto, "Comparison of Classification Techniques in Data Mining for Bank Direct Marketing," *J. Teknol. Inf. dan Ilmu Komput.*, vol. 5, no. 5, p. 567, 2018, doi: 10.25126/jtiik.201855958.

[4] V. C. Bavkar and A. A. Shind, "Machine learning algorithms for Diabetes prediction and neural network method for blood glucose measurement," *Indian J. Sci. Technol.*, vol. 14, no. 10, pp. 869–880, 2021, doi: https://doi.org/ 10.17485/IJST/v14i10.2187 ∗.

[5] E. S. Almutairi and M. F. Abbod, "Machine Learning Methods for Diabetes Prevalence Classification in Saudi Arabia," *Modelling*, vol. 2023, no. 4, pp. 37–55, 2023, doi: https://doi.org/10.3390/ modelling4010004.

[6] A. M. Argina, "Application of the K-Nearest Neighbor Classification Method on a Dataset of Diabetes Patients," *Indones. J. Data Sci.*, vol. 1, no. 2, pp. 29–33, 2020, doi: 10.33096/ijodas.v1i2.11.

[7] A. C. Mawarni, R. Rusdah, L. L. Hin, and D. Anubhakti, "Early Detection of Early Symptoms of Diabetes using the Random Forest Algorithm," *IDEALIS Indones. J. Inf. Syst.*, vol. 6, no. 2, pp. 165–171, 2023, doi: 10.36080/idealis.v6i2.3018.

[8] E. Patimah, V. B. Haekal, and D. Sandya Prasvita, "Classification of Liver Disease Using the Decision Tree Method," in *Seminar Nasional Mahasiswa Ilmu Komputer dan Aplikasinya (SENAMIKA) Jakarta-Indonesia*, Universitas Pembangunan Nasional Veteran Jakarta, 2021, pp. 655–659. [Online]. Available: https://conference.upnvj.ac.id/index.php/senamika/article/view/1388

[9] B. Gopi, C. Nalini, and A. Francesco, "Late-Life Alzheimer's Disease (AD) Detection Using Pruned Decision Trees," *Int. J. Brain Disord. Treat.*, vol. 6, no. 1, pp. 8–11, 2020, doi: 10.23937/2469-5866/1410033.

[10] A. Fuoco, A. Komeshak, C. Farida, C. White, and M. Mahmoud, "Decision Tree Classifier Based Model for Disease Prediction," in *Conference: The 2021 World Congress in Computer Science, Computer Engineering, & Applied Computing (CSCE)*, 2021. [Online]. Available: https://www.researchgate.net/publication/353654932_Decision_Tree_Classifier_Based_Model_for_Disease_Prediction

[11] A. Khalemsky, R. Gelbard, and Y. Stukalin, "Constructing a Course on Classification Methods for Undergraduate Non-STEM Students: Striving to Reach Knowledge Discovery Constructing a Course on Classification Methods for Undergraduate Non-STEM Students: Striving to Reach Knowledge Discovery," *J. Stat. Data Sci. Educ.*, vol. 33, no. 1, pp. 68–76, 2025, doi: 10.1080/26939169.2024.2320218.

[12] N. L. W. S. R. Ginantra *et al.*, *Data Mining and Application of Algorithms*, Pertama. Yayasan Kita Menulis, 2021.

[13] N. S´anchez-Maro˜no, A. Alonso-Betanzos, and M. Tombilla-Sanrom´an, "Intelligent Data Engineering and Automated Learning - IDEAL 2007," in *Conference: Intelligent Data Engineering and Automated Learning - IDEAL 2007*, 2007, pp. 790–799. doi: 10.1007/978-3-540-77226-2.

[14] H. Wang, S., Tang, J., & Liu, *Encyclopedia of Machine Learning and Data Science*. 2020. doi: 10.1007/978-1-4899-7502-7.

[15] J. N. B. S. Ringo, W. J. Mursalin, N. C. Nurfadilah, D. R. Ramadhan, and W. O. Z. Madjida, "Perbandingan Metode Klasifikasi Multiclass untuk Pemetaan Zona Risiko COVID-19 di Pulau Jawa," *J. Komput. dan Inform.*, vol. 9, no. 1, pp. 98–107, 2021, doi: 10.35508/jicon.v9i1.3602.

[16] R. K. Hapsari, E. Purwanti, W. Widyanto, R. Gunawan, F. Nurlaily, and A. H. Salim, "Optimization Based Random Forest Algorithm Modification for Detecting Monkeypox Disease," in *2023 Sixth International Conference on Vocational Education and Electrical*

Engineering (ICVEE), IEEE, 2023, pp. 340–346. doi: 10.1109/icvee59738.2023.10348223.

[17] A. Jain, D. Somwanshi, K. Joshi, and S. S. Bhatt, "A Review: Data Mining Classification Techniques," in *Proceedings of 3rd International Conference on Intelligent Engineering and Management, ICIEM 2022*, 2022, pp. 636–642. doi: 10.1109/ICIEM54221.2022.9853036.

[18] J. E. Black, J. K. Kueper, and T. S. Williamson, "An introduction to machine learning for classification and prediction," *Fam. Pract.*, vol. 40, no. October, pp. 200–204, 2023.

[19] B. T. Jijo and A. Abdulazeez, "Classification Based on Decision Tree Algorithm for Machine Learning," *J. Appl. Sci. Technol. Trends*, vol. 2, no. 01, pp. 20–28, 2021, doi: 10.38094/jastt20165.

[20] G. Nanfack, P. Temple, and B. Frénay, "Constraint Enforcement on Decision Trees: A Survey," *ACM Comput. Surv.*, vol. 54, no. 10, 2022, doi: 10.1145/3506734.

[21] P. Bhargav and K. Sashirekha, "A Machine Learning Method for Predicting Loan Approval by Comparing the Random Forest and Decision Tree Algorithms.," *J. Surv. Fish. Sci.*, vol. 10, no. 1S, pp. 1803–1813, 2023, [Online]. Available: https://sifisheriessciences.com/journal/index.php/journal/article/view/414

[22] M. S. A. Rahman, N. A. A. Jamaludin, Z. Zainol, and T. M. T. Sembok, "The Application of Decision Tree Classification Algorithm on Decision-Making for Upstream Business," *Int. J. Adv. Comput. Sci. Appl.*, vol. 14, no. 8, pp. 660–667, 2023, doi: 10.14569/IJACSA.2023.0140873.

[23] A. Arifuddin, G. S. Buana, R. A. Vinarti, and A. Djunaidy, "Performance Comparison of Decision Tree and Support Vector Machine Algorithms for Heart Failure Prediction," *Procedia Comput. Sci.*, vol. 234, pp. 628–636, 2024, doi: 10.1016/j.procs.2024.03.048.

[24] F. Hajjej, M. A. Alohali, M. Badr, and A. Rahman, "A Comparison of Decision Tree Algorithms in the Assessment of Biomedical Data," *Biomed Res. Int.*, vol. 2022, 2022, doi: 10.1155/2022/9449497.

[25] R. K. Hapsari, M. Miswanto, R. Rulaningtyas, and H. Suprajitno, "Identification of Diabetes Mellitus and High Cholesterol Based on Iris Image," *J. Hunan Univ. (Natural Sci.*, vol. 48, no. 10, pp. 151–160, 2021, [Online]. Available: http://jonuns.com/index.php/journal/article/view/776

[26] A. S. Paymode and V. B. Malode, "Transfer Learning for Multi-Crop Leaf Disease Image Classi fi cation using Convolutional Neural Network VGG," *Artif. Intell. Agric.*, vol. 6, pp. 23–33, 2022, doi: 10.1016/j.aiia.2021.12.002.

[27] D. Saputra, W. S. Dharmawan, M. Wahyudi, W. Irmayani, J. Sidauruk, and Martias, "Performance Comparison and Optimized Algorithm Classification," *J. Phys. Conf. Ser.*, vol. 1641, no. 1, 2020, doi: 10.1088/1742-6596/1641/1/012087.

[28] O. Y. Odufuwa, L. K. Tartibu, and K. Kusakana, "Artificial neural network modelling for predicting efficiency and emissions in mini-diesel engines: Key performance indicators and environmental impact analysis," *Fuel*, vol. 387, no. December, p. 134294, 2025, doi: 10.1016/j.fuel.2025.134294.

[29] M. Ahsan *et al.*, "Deep transfer learning approaches for Monkeypox disease diagnosis," *Expert Syst. Appl.*, vol. 216, no. December, p. 119483, 2023, doi: 10.1016/j.eswa.2022.119483.

[30] F. R. Razak, M. K. Biddinika, and H. Yuliasnyah, "Radial basis function model for obesity classification based on lifestyle and physical condition," *J. ELTIKOM J. Tek. Elektro, Teknol. Inf. dan Komput.*, vol. 8, no. 2, pp. 192–200, 2024, doi: http://doi.org/10.31961/eltikom.v8i2.1347.

[31] D. Muller, I. Soto-Rey, and F. Kramer, "Towards a Guideline for Evaluation Metrics in Medical Image Segmentation," *arXiv (Cornell Univ.*, pp. 1–7, 2022, doi: https://doi.org/10.48550/arxiv.2202.05273.

[32] R. K. Hapsari, M. Miswanto, R. Rulaningtyas, H. Suprajitno, and G. H. Seng, "Modified Gray-Level Haralick Texture Features for Early Detection of Diabetes Mellitus and High Cholesterol with Iris Image," *Int. J. Biomed. Imaging*, vol. 2022, 2022, doi: https://doi.org/10.1155/2022/5336373.

[33] G. M. Foody, "Challenges in the real world use of classification accuracy metrics: From recall and precision to the Matthews correlation coefficient," *PLoS One*, vol. 18, no. 10, pp. 1–19, 2023, doi: https://doi.org/10.1371/journal.pone.0291908.

[34] I. Ihoume, R. Tadili, N. Arbaoui, M. Benchrifa, A. Idrissi, and M. Daoudi, "Developing a multi-label tinyML machine learning model for an active and optimized greenhouse microclimate control from multivariate sensed data," *Artif. Intell. Agric.*, vol. 6, pp. 129–137, 2022, doi: 10.1016/j.aiia.2022.08.003.

[35] B. Pham, P. T. Le, T. Tai, Y. Hsu, Y. Li, and J.-C. Wang, "Electrocardiogram Heartbeat Classification for Arrhythmias and Myocardial Infarction," *Sensors*, vol. 23, pp. 1–18, 2023, doi: doi.org/10.3390/s23062993.

[36] J. Nemecek, T. Pevny, and J. Marecek, "Improving the Validity of Decision Trees as Explanations arXiv: 2306 . 06777v5 [cs . LG] 4 Jun 2024," *arXiv (Cornell Univ.*, vol. June, 2024, doi: https://doi.org/10.48550/arxiv.2306.06777.