Comparison of the Performance Results of C4.5 and Random Forest Algorithm in Data Mining to Predict Childbirth Process

Muhasshanah[1]*, Mohammad Tohir[2], Dewi Andariya Ningsih[3], Neny Yuli Susanti[4], Astik Umiyah[5], and Lia Fitria[6]
[1]Program Studi Teknologi Informasi, Universitas Ibrahimy
Situbondo 68374, Indonesia
[2]Program Studi Tadris Matematika, Universitas Ibrahimy
Situbondo 68374, Indonesia
[3−5]Program Studi Kebidanan, Universitas Ibrahimy
Situbondo 68374, Indonesia
[6]Program Studi Pendidikan Profesi Bidan, Universitas Ibrahimy
Situbondo 68374, Indonesia
Email: [1]muhasshanah@ibrahimy.ac.id, [2]matematohir@ibrahimy.ac.id, [3]dewiandariya01@gmail.com, [4]nenyyulisusanti@gmail.com, [5]astik.umiyah86@gmail.com, [6]leeafitria@gmail.com

***Abstract*—Technology advancements in the world of information have made it easier for many people to process data. Data mining is a process of mining more valuable information from large data sets. The research aims to determine the difference between the C.45 and random forest algorithms in data mining to predict the childbirth process of pregnant women. It compares the accuracy of the performance results of the C4.5 and random forest algorithms to predict the delivery process for pregnant women. Then, experimental research is conducted to classify the childbirth process in Situbondo, Indonesia, by applying the C.45 and the random forest algorithm in the data mining. The decision tree J48 algorithm is used for the C4.5 algorithm in the research. Both algorithms are compared for their error classification and accuracy level. The research uses 1,000 data for training and 200 data for testing. The results show the accuracy of implementing the C4.5 and random forest algorithms with data mining using 10-fold cross-validation, generating 96% and 95% as correctly classified data. Then, the Relative Absolute Error for both algorithms has the same result. It is 15%. The C4.5 algorithm has a better result than the random forest algorithm by comparing the performance results. Further research can add more data to improve the accuracy of the analysis results by using another algorithm.**

***Index Terms*—C4.5 Algorithm, Random Forest Algorithm, Data Mining, Childbirth Process**

Received: Feb. 19, 2022; received in revised form: July 15, 2022; accepted: July 18, 2022; available online: March 17, 2023.
*Corresponding Author

I. Introduction

TECHNOLOGY advancements in the world of information have made it easier for many people to process data. The more data exist, the more day it will take to process it. So, data processing is not optimal. There will be mountains of data resulting in little information being generated. Data mining is a process of collecting necessary information from big data [1]. Important information is carried out using several methods, including statistical methods, mathematic-section, and artificial intelligence technology. Data mining is described more explicitly as tools and programs that employ statistical data analysis and filters to save as much data as possible. Data mining is a set of methods that examine new values in the form of previously unknown information from a database by conducting the extraction process and discovering important patterns in existing data [2]. Machine learning classification methods have long been utilized in data mining and various other fields of computer science. The process of establishing a model or function that identifies and separates data classes or concepts is referred to as classification. The most often used techniques for data categorization include Naïve Bayes, decision tree, linear regression, k Nearest Neighbors (k-NN), neural network, support vector machine, and logistic regression [3–5].

The C4.5, random forest algorithms, and gradient boosting are examples of decision tree algorithms that have been developed and frequently utilized in dealing with various classification and prediction scenarios [6]. A decision tree is a well-known and strong classification and prediction approach. The decision tree approach converts massive volumes of data into a decision tree that contains the rules. A decision tree is a structure that uses a set of decision rules to split huge data collection into smaller record sets. The members of the result set become more similar to one another with each division series. The members of the result set become more similar to one another with each division series [7–9].

The C4.5 method is a decision tree formation technique that calculates the gain value, with the highest gain serving as the first node or root node [10]. C4.5 decision tree is the first fundamentally supervised machine learning classification algorithm widely applied and consistently achieves excellent predictive performance [11]. Meanwhile, random forest is an easy-to-use and versatile machine learning algorithm with amazing results most of the time, even without a hyperparameter setting [12]. It is also one of the most commonly utilized algorithms due to its simplicity and diversity, as it can be applied to classification and regression problems [13]. One of the essential characteristics of the random forest algorithm is that it can handle datasets with both continuous and categorical variables in regression and classification [14]. It outperforms other algorithms in categorization tasks [15].

The previous research results show that the findings of the comparative analysis are the best alternative algorithm choice in airline customer satisfaction classifications. In this comparison, the random forest algorithm outperforms the C4.5 methods [16]. However, another previous research mentions that the Naïve Bayes algorithm has higher accuracy compared with the random forest algorithm and C4.5. There is a visible difference in accuracy between the Naïve Bayes with a random forest of 2.84%. The difference between the Naïve Bayes with C4.5 is 3.53% [17].

In another example, the algorithm with the best performance for classification is the random forest algorithm with the condition that it uses shuffle sampling, and the majority of linear sampling produces poor performance. Meanwhile, shuffle sampling performs very well for tree-based algorithms [18]. Next, another previous research result shows that the prediction of the resilient backpropagation algorithm is 100%. However, the C4.5 and random forest algorithms have 97.6% and 98.4% accuracy for evaluating seismic soil liquefaction potential, respectively [19]. The performance test results of the three algorithms (C4.5, random tree, and random forest) by another research suggest that random forest with pruning and pre-pruning is the best for an accuracy value of 74.63% and Area Under Curve (AUC) value of 0.743 [20].

The research focuses on implementing the C4.5 and random forest algorithms in data mining to predict the birth process of pregnant women before delivery. Implementing these two algorithms is essential because improving maternal health and reducing maternal and neonatal mortality are the primary purposes of expanding the number of facilities. Therefore, childbirth can take place safely, and emergency obstetric services are available in line with the principle that every pregnant woman is at risk of life-threatening complications [21]. As a result, the earlier the problem is predicted and handled, the lower the likelihood of an emergency. As a result, the quality of service provided throughout pregnancy, delivery, and puerperium is the main principle for reducing maternal mortality. In addition to an effective referral system, emergency obstetric services greatly determine maternal mortality [22]. A solid health system will make achieving health development goals or targets easier, specifically minimizing maternal mortality during childbirth. The delivery process can be done by two methods: normal birth and cesarean section. Thus, pregnant women need to know various things about childbirth to anticipate things that may happen and ensure their condition is good. Even though they already have an estimated date of birth, only a few pregnant women give birth on the day of their estimated childbirth. Therefore, pregnant women must be aware of the indications of the childbirth process, which can arrive anywhere from three weeks before the estimated date until two weeks after the birth.

Based on the description, the researchers are interested in comparing the C4.5 and random forest algorithms to see which provides the best accuracy in assisting health workers (midwives) in making decisions about the delivery process. The research aims to determine the effectiveness of applying the C4.5 algorithm with the random forest algorithm in data mining to predict the birth process of pregnant women before delivery. The research compares both algorithms for error classification and accuracy level. For the novelty of the research, there have been no previous studies comparing the accuracy of the C4.5 and random forest algorithms in predicting the childbirth process.

II. Research Method

A. Research Design

Data mining is a multidisciplinary scientific discipline that encompasses database technology, machine learning, statistics, pattern recognition, information retrieval, artificial neural networks, knowledge-based systems, artificial intelligence, high-performance

TABLE I
DISCRETIZATION AND ATTRIBUTE DESCRIPTION.

Attribute	Description	Unit	Discretization and Description
Birth Canal (BC)	The birth canal used is whether there is something that blocks it or not	-	Normal (if nothing covers the birth canal); Abnormal (placenta previa, tumor, and others)
Blood Pressure (BP)	Blood pressure of pregnant women	Mmhg	Hypotension (<110); Normal (110–130); Hypertension (>130)
Estimated Baby Weight (EBW)	Baby weight	Gram	Small (<2500); Normal (2500–4000); Large (>4000)
Fetus Position (FP)	The position of the fetus in the uterus is normal, breech, or transverse	-	Normal (head is on bottom); Transverse (head on the right/on the left); Breech (head is on top)
Heart Rate (HR)	Baby’s heart rate	Times/minute	Normal (120–160); Abnormal (>160 or <120)
Lab Examination (LAB)	Hemoglobin (Hb) examination results, reduction, and albumin	-	Normal (Hb negative, reduction, albumin); Abnormal (positive)
Mother’s Disease (MD)	Diseases suffered by the mother (heart, syphilis, HIV/AIDS, eclampsia, gonorrhea)	-	No; Yes (heart, syphilis, HIV/AIDS, eclampsia, gonorrhea)
C-section History (CS)	It is whether pregnant women ever have c-section surgery or not	-	Ever (mother ever has had childbirth process with c-section before); Never (mother has never had childbirth process with c-section before)
Pelvis Size (PS)	Pelvic size for pregnant women	Cm	Narrow (≤145); Normal (>145) (KIA-SPR)
Gestational Age (GA)	Gestational age of pregnant women	Weeks	Premature (<37); Aterm (38–40); Post Date (>41)

computing, and data visualization [23]. The research method is fundamental experimental research with the decision tree J48 method. It leads to the impact resulting from the experiment on applying the C4.5 and random forest algorithms in data mining to classify the birth process.

B. Identification and Selection Attribute

The data are classified using the available C4.5 (J48) and Random Forest algorithms in the Weka software. The research relies on information gathered from Private Practice Midwives (PMB) in Situbondo. After selecting all data used in the research, the attributes are obtained as follows: birth canal, blood pressure, estimated baby weight, fetal location, heart rate, laboratory examination, maternal disease, history of section caesarian (c-section), pelvic size, gestational age, and birth process. Then, the management of prenatal care in the third trimester in *Buku Saku Pelayanan Kesehatan Ibu di Fasilitas Kesehatan Dasar dan Rujukan*, published by the Ministry of Health of the Republic of Indonesia, is used to determine these attributes [24]. The research uses 1,000 data for training and 200 data for testing.

C. Attribute Discretization

Attribute discretization of the birth canal, heart rate, and laboratory examination is divided into normal and abnormal. Meanwhile, blood pressure is divided into hypotension, normal, and hypertension. Estimated baby weight is divided into small, normal, and large (Johnson Tousac). Then, the size of the pelvis is divided into narrow and normal. Gestational age is divided into premature, aterm (normal), and post-date. Next, the fetus’s position is divided into normal, transverse, and breech. The discretization for each attribute is based on the description in *Buku Rustam Mochtar Sinopsis Obstetri* [25]. The details are shown in Table I.

D. Decision Tree Preparation Phase Using J48

The C4.5 algorithm, which generates a decision tree, is implemented in decision tree J48. In data mining, a decision tree is one of the categorization techniques. A classification algorithm is a learning method that constructs a model from pre-classified samples using inductive learning. The value of each property determines the data item. Classification is a mapping of a set of characteristic-section of a certain class. The decision tree categorizes the provided data based on the attribute value [26–28]. Then, the dataset with the choice attributes is classified using the decision tree J48.

E. Evaluation of the J48 Decision Tree Classifier Using K-Fold Cross-Validation

In k-fold cross-validation, the test data are randomly divided into k mutually exclusive subsets or “folds” of $D1, D2, \ldots$, and Dk, each of which has roughly the same size. The K sessions of training and testing have been completed. In the $i-th$ iteration, the Di partition is used as test data, while the remaining partitions are mixed to train the model. In the first iteration, the subsets $D2, \ldots,$ and Dk are used as training data to construct the first model, which is tested on $D1$. Then, the second iteration is trained on the subsets $D1, D3, \ldots Dk$ and tested on $D2$ and so on.

F. Completion of the C4.5 Algorithm Phase

The C4.5 algorithm is a modification of the ID3 algorithm that employs information entropy, continuous and discrete characteristics, categorical and numeric attributes, and missing values [29]. The phase of testing the C4.5 algorithm is carried out using the following steps:

1) select the root from one of the attributes
2) branch each value
3) on the branch, divide the case
4) until each case has the same class on the branch, repeat the process

The root of the attribute is chosen based on the attribute with the highest gain value. It uses Eq. (1). It shows S as the case group, A as an attribute, n as the number of parts of attribute A, $|Si|$ as the number of cases in part i, and $|S|$ as the number of cases in S. Then, the research calculates the entropy value with Eq. (2). It has n as the number of partitions S and pi as the ratio of Si to S.

$$\text{Gain}(S, A) = \text{Entropy}(S) - \sum_{i=1}^{n} \frac{|S_i|}{S} * \text{Entropy}(S_i), \quad (1)$$

$$\text{Entropy}(S) = \sum_{i=1}^{n} -pi * \log_2 pi. \quad (2)$$

G. Completion of the Random Forest Algorithm Phase

Random Forest maps the class's attributes so that it can be used to find classifications for data that have not yet appeared. It is named a random forest because it is a descendant of the ID3 approach to constructing decision trees [30]. Following are the stages of testing the performance of the random forest algorithm.

1) Pay attention to the labels on the data. A leaf will be formed with the overall data label value if they are all the same.
2) Calculate the value of information using all existing data, with Eq. (3). The equation is the probability of a tuple in D being a class with the assumption that the entropy of D is the average of the information needed to identify tuples in D. If the value of A is discrete, the D data will be separated by a number of A data values so that the value of each branch will be pure and similar. After the first branch, the number of possible branches is measured by Eq. (4).
3) Calculate the value of information with Eq. (5).

$$\text{info}(D) = -\sum_{i=1}^{m} pi \log^{2(pi)}, \quad (3)$$

$$\text{info}A(D) = \sum_{j}^{\nu} \frac{Dj}{D} * \text{info}A(Dj), \quad (4)$$

$$\text{Gain}(A) = \text{info}(D) - \text{info}A(D). \quad (5)$$

4) For each attribute, pay attention to the attribute's data content. It shows $\frac{Dj}{D}$ as the weight of partition j and info$A(D)$ as the information needed to classify the tuple of D in partition A. The smaller the result of Eq. (4) is, the better the resulting partition is going to be. The value of an attribute determines the importance of that attribute in the preparation of a decision tree. If the attribute is continuous, the split point will be searched by sorting all data according to the attribute from small to large. Then it sees the average between one data and the data afterward. The information value will be calculated according to the split point candidates one by one, and the smallest split point value will be selected. The gain value for each attribute will be calculated by Eq. (5). The highest gain will be used as a branch in the decision tree.
5) After the decision tree branch is formed, the calculation is repeated from steps 1 to 4. However, if the branch has reached the maximum allowed branch, a leaf will be formed with the majority value of the data value.

H. Weighted Mean Recall and Weighted Mean Precision

The weighted mean recall is the number of true positives over the number of true positives plus the number of false negatives used with weighted data [31]. Meanwhile, weighted mean precision is the number of true positives over the number of true positives plus the number of false positives [32].

III. Results and Discussion

A. The C4.5 Algorithm Model Testing

After processing and testing using the decision tree J48, the information is compiled in the form of a tree. It can be seen that the pelvis size is the root of the tree. If the pelvis size is narrow, the classification results indicate the birth process by c-section. However, if the pelvis size is normal and the heart rate is normal, the classification results indicate normal childbirth. Then, if the pelvis size is normal and the heart rate

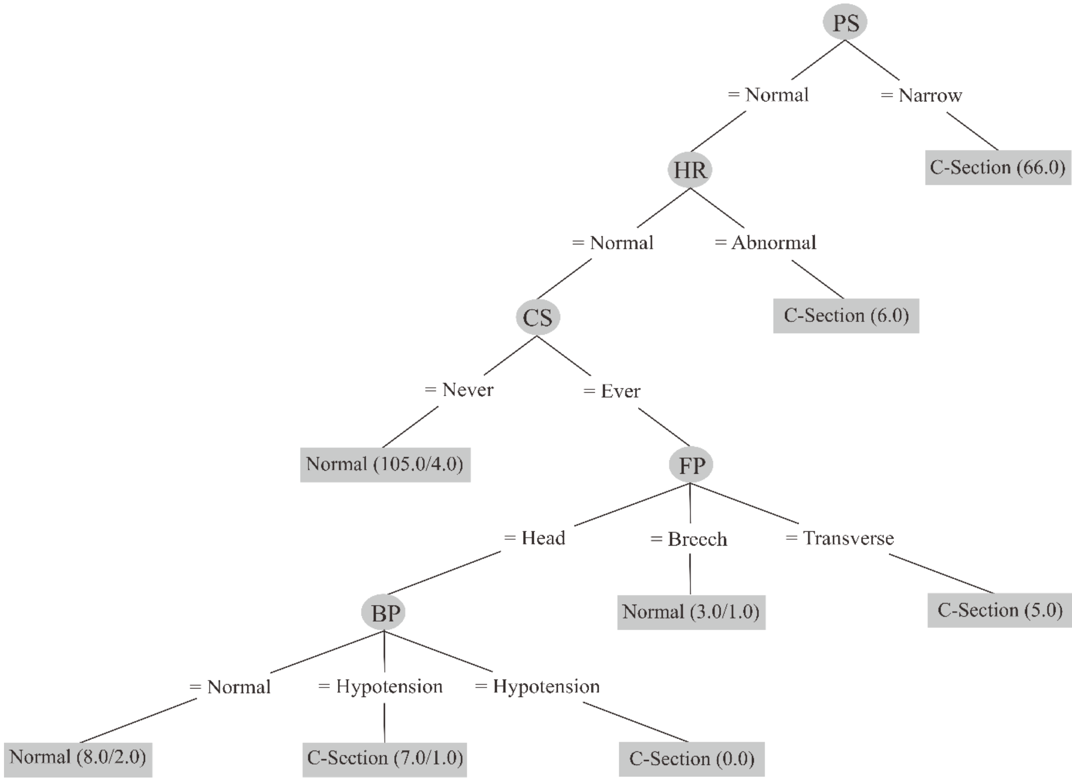

Fig. 1. Decision tree of C4.5 algorithm.

is abnormal, the classification results show a c-section childbirth process.

Moreover, if there are normal heart rate, a history of the previous c-section, and the fetus's transverse position, the classification results indicate the process of childbirth by c-section. Next, suppose there are normal pelvis size, normal heart rate, a history of c-section, normal fetal position, and hypertensive or hypotensive blood pressure. In that case, the classification results are the process of childbirth by c-section. All information can be obtained by reading all the branches of the tree. In Fig. 1, a decision tree chart of the C4.5 algorithm is presented using decision tree J48.

Figure 1 shows that the calculation of entropy and gain value generated by the C4.5 algorithm. It has the highest gain value on the pelvis size attribute, which becomes the root of the decision tree. If the pelvis size is "narrow", the childbirth process only generates "c-sections" decision of 66 data. However, if the pelvis size is "normal", the heart rate is "normal", and the history of c-section is "never", it will generate 105 data of childbirth process "normal" and 4 data of "c-section". Then entropy and gain value calculations are continued to determine other branches.

Table II shows that the number of testing data for mothers who give normal birth is 110. Then, 109 (true-positive) are correctly identified as giving normal birth. In contrast, 1 (false-positive) is incorrectly identified by the decision tree J48 classifier that the actual condition is that the mother gives birth with the c-section childbirth process.

Meanwhile, in testing pregnant women who use c-sections, 83 people are correctly identified as doing c-sections (true-negative). In comparison, 7 people (false-negative) are incorrectly identified as not doing c-sections. Thus, it can be calculated that the accuracy of the decision tree J48 classifier reaches 96%, with 192 correctly classified data. The amount of error that causes a decrease in accuracy occurs in false-negative conditions.

TABLE II
CONFUSION MATRIX FROM C4.5 DECISION TREE TESTING WITH 10-FOLD CROSS-VALIDATION.

Classification	Childbirth Process Identified		Amount of Data
	Normal	C-section	
Normal	109	1	Normal
C-Section	7	83	C-Section

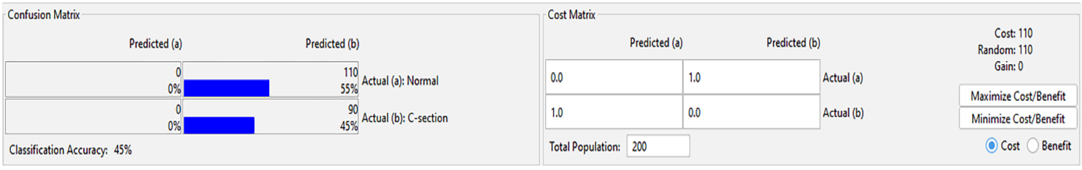

Fig. 2. Confusion matrix for normal childbirth process.

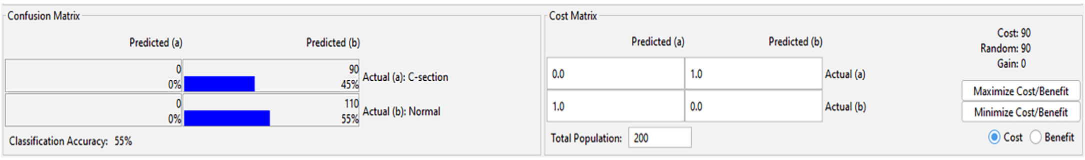

Fig. 3. Confusion matrix for c-section childbirth process.

TABLE III
CONFUSION MATRIX FROM RANDOM FOREST ALGORITHM WITH DECISION TREE TESTING WITH 10-FOLD CROSS-VALIDATION.

Classification	Childbirth Process Identified		Amount of Data
	Normal	C-section	
Normal	107	3	Normal
C-section	7	83	C-section

B. The Random Forest Algorithm Model Testing

Figures 2 and 3 are the result of the confusion matrix by testing all data using the random forest algorithm. In a normal childbirth process, the classification accuracy is 55%. Meanwhile, the classification accuracy for the c-section childbirth process is 45%.

The number of testing data for mothers who give normal birth is 110. Around 107 data (true-positive) are correctly identified as giving normal birth. In contrast, 3 data (false-positive/FP) are incorrectly identified by random forest that the actual condition is a c-section childbirth process. Meanwhile, in testing pregnant women who use c-sections, 83 people are correctly identified as doing c-sections (true-negative). In comparison, 7 people (false-negative) are incorrectly identified as not doing c-sections. As a result, it can be calculated that the accuracy of the random forest with decision tree classifier reaches 95%, with 190 correctly classified data. The amount of error that causes a decrease in accuracy also occurs in false-negative conditions. Table III shows the confusion matrix from random forest algorithm with decision tree testing with 10-fold cross-validation.

C. The Comparison of C4.5 and Random Forest Algorithms for Data Mining

The results of testing the C4.5 Algorithm and Random Forest for data mining in predicting the childbirth process can be compared, as shown in Table IV. All the algorithm trials in Table IV show that a good algorithm for the characteristics of the classification data in the research is the C4.5 algorithm using shuffle sampling (gain ratio). The weighted mean recall and precision for both algorithms are obtained from decision tree testing with 10-fold cross-validation using Weka software.

D. Discussion

The C4.5 algorithm builds a decision tree that has a set of rules for generating predictions. On the other hand, the random forest method chooses observations and characteristics at random to build numerous decision trees before averaging the results. The C4.5 algorithm correctly classifies the data by 96%, while the random forest algorithm has 95%. The C4.5 algorithm is higher than the random forest algorithm because the C4.5 algorithm correctly classifies the data for the normal childbirth process in as much as 109 of the 110 data. In comparison, the random forest algorithm only classifies 107 of 110 data. The results for the c-section childbirth process are the same in both algorithms.

The research indicates that the correctly classified data accuracy of the C4.5 algorithm obtained better results than the random forest algorithm. However, the results contradict the results of previous research that also compares the C4.5 algorithm with the random forest algorithm [16, 18]. In that previous research, the random forest algorithm performs better than the C4.5 algorithms [16]. Then, another previous research shows that the algorithm with the best performance for

TABLE IV
THE RESULT COMPARISON OF C4.5 AND RANDOM FOREST ALGORITHMS.

Algorithm	Class	Accuracy (%)	Weighted Mean Recall (%)	Weighted Mean Precision (%)	Relative Absolute Error (%)
C4.5	Normal	96.00	99.10	94.00	15
	C-Section		92.20	98.80	
Random Forest	Normal	95.00	97.30	93.90	15
	C-Section		92.20	96.50	

classification is the random forest algorithm with the condition that it uses shuffle sampling [18]. The C4.5 algorithm based on particle swarm optimization can improve the accuracy of the C4.5 algorithm [33].

Meanwhile, the research results on the random forest algorithm have a lower accuracy of 1% compared to the C4.5 algorithm. It has the same results as previous research that the accuracy of the random forest to predict heart disease is above 90% [34]. Another previous research shows that after filtering data with the random forest algorithm, the accuracy value is 99.98%. It indicates an increase in the performance of the random forest algorithm on big data [35]. However, the results differ from previous research [36]. The random forest algorithm provides a higher classification accuracy than other methods. For Ikonos images in urban areas, the results show that the random forest algorithm has a classification accuracy of 10% higher than Support Vector Machine (SVM). In comparison, the Gentle AdaBoost (GAB) algorithm has the lowest classification accuracy (14% lower than random forest).

IV. CONCLUSION

The C4.5 and random forest algorithms can be applied to predict the childbirth process. As a result, it can assist midwives in determining whether the mother will give normal birth or surgically. The result is expected to reduce maternal and infant mortality due to decision-making mistakes. Each algorithm uses 10-fold cross-validation with 96% accuracy in correctly classified data for the C4.5 algorithm and 95% for the random forest algorithm. In the C4.5 algorithm, the true positive value is 0.991 for the class of normal childbirth and 0.922 for c-section childbirth. Meanwhile, the random forest algorithm obtains a true positive value of 0.973 for the class of normal childbirth and 0.922 for c-section childbirth. The C4.5 algorithm is higher for correctly classified data on the true positive value in the class of normal childbirth than random forest. Classification of more data will provide a more accurate rule in predicting the childbirth process.

The research is limited to only 200 data used as testing data. The research also only uses 10 attribute descriptions in predicting the childbirth process: birth canal, blood pressure, estimated baby weight, fetal location, heart rate, laboratory examination, maternal disease, history of cesarean section (c-section), pelvic size, gestational age, and birth process. Hence, more testing data and more complete attributes can be used for further research to get more accurate results.

ACKNOWLEDGEMENT

We would like to thank the support from the Faculty of Science and Technology and the Faculty of Health Science, Universitas Ibrahimy, Situbondo, Indonesia, in 2022.

REFERENCES

[1] J. Yang, Y. Li, Q. Liu, L. Li, A. Feng, T. Wang, S. Zheng, A. Xu, and J. Lyu, "Brief introduction of medical database and data mining technology in big data era," *Journal of Evidence-Based Medicine*, vol. 13, no. 1, pp. 57–69, 2020.

[2] S. Lv, H. Kim, B. Zheng, and H. Jin, "A review of data mining with big data towards its applications in the electronics industry," *Applied Sciences*, vol. 8, no. 4, pp. 1–34, 2018.

[3] S. Dutta and S. K. Bandyopadhyay, "Employee attrition prediction using neural network cross validation method," *International Journal of Commerce and Management Research*, vol. 6, no. 3, pp. 80–85, 2020.

[4] D. S. Abdelminaam, N. Neggaz, I. A. E. Gomaa, F. H. Ismail, and A. A. Elsawy, "ArabicDialects: An efficient framework for Arabic dialects opinion mining on Twitter using optimized deep neural networks," *IEEE Access*, vol. 9, pp. 97 079–97 099, 2021.

[5] S. Pouriyeh, S. Vahid, G. Sannino, G. De Pietro, H. Arabnia, and J. Gutierrez, "A comprehensive investigation and comparison of machine learning techniques in the domain of heart disease," in *2017 IEEE Symposium on Computers and Communications (ISCC)*. Heraklion, Greece: IEEE, July 3–6, 2017, pp. 204–207.

[6] Z. Xu and Z. Wang, "A risk prediction model for type 2 diabetes based on weighted feature selection of random forest and XGBoost ensemble

classifier," in *2019 Eleventh International Conference on Advanced Computational Intelligence (ICACI)*. Guilin, China: IEEE, June 7–9, 2019, pp. 278–283.

[7] T. T. Huynh-Cam, L. S. Chen, and H. Le, "Using decision trees and random forest algorithms to predict and determine factors contributing to first-year university students' learning performance," *Algorithms*, vol. 14, no. 11, pp. 1–17, 2021.

[8] S. Poudyal, M. Nagahi, M. Nagahisarchoghaei, and G. Ghanbari, "Machine learning techniques for determining students' academic performance: A sustainable development case for engineering education," in *2020 International Conference on Decision Aid Sciences and Application (DASA)*. Sakheer, Bahrain: IEEE, Nov. 8–9, 2020, pp. 920–924.

[9] A. Hamoud, A. S. Hashim, and W. A. Awadh, "Predicting student performance in higher education institutions using decision tree analysis," *International Journal of Interactive Multimedia and Artificial Intelligence*, vol. 5, pp. 26–31, 2018.

[10] R. Sudrajat, I. Irianingsih, and D. Krisnawan, "Analysis of data mining classification by comparison of C4.5 and ID algorithms," in *IOP Conference Series: Materials Science and Engineering*, vol. 166. IOP Publishing, 2017, pp. 1–8.

[11] R. Jothikumar and S. R. Balan, "C4.5 classification algorithm with back-track pruning for accurate prediction of heart disease," *Biomedical Research*, pp. S107–S111, 2016.

[12] M. T. Yazici, S. Basurra, and M. M. Gaber, "Edge machine learning: Enabling smart internet of things applications," *Big Data and Cognitive Computing*, vol. 2, no. 3, pp. 1–17, 2018.

[13] A. Priyam, G. R. Abhijeeta, A. Rathee, and S. Srivastava, "Comparative analysis of decision tree classification algorithms," *International Journal of Current Engineering and Technology*, vol. 3, no. 2, pp. 334–337, 2013.

[14] K. Kim, "A hybrid classification algorithm by subspace partitioning through semi-supervised decision tree," *Pattern Recognition*, vol. 60, pp. 157–163, 2016.

[15] P. Y. Pawar and S. H. Gawande, "A comparative study on different types of approaches to text categorization," *International Journal of Machine Learning and Computing*, vol. 2, no. 4, pp. 423–426, 2012.

[16] W. Baswardono, D. Kurniadi, A. Mulyani, and D. M. Arifin, "Comparative analysis of decision tree algorithms: Random forest and C4.5 for airlines customer satisfaction classification," *Journal of Physics: Conference Series*, vol. 1402, pp. 1–6, 2019.

[17] W. Gata, H. Basri, R. Hidayat, Y. E. Patras, B. Baharuddin, R. Fatmasari, S. Tohari, and N. K. Wardhani, "Algorithm implementations Naïve Bayes, random forest. C4.5 on online gaming for learning achievement predictions," in *2nd International Conference on Research of Educational Administration and Management (ICREAM 2018)*. Bandung, Indonesia: Atlantis Press, Oct. 18, 2019, pp. 1–9.

[18] E. Ismanto and M. Novalia, "Komparasi kinerja algoritma C4.5, random forest, dan gradient boosting untuk klasifikasi komoditas," *Techno.Com*, vol. 20, no. 3, pp. 400–410, 2021.

[19] R. B. Bhardwaj and S. R. Chaurasia, "Use of ANN, C4.5 and random forest algorithm in the evaluation of seismic soil liquefaction," *Journal of Soft Computing in Civil Engineering*, vol. 6, no. 2, pp. 92–106, 2022.

[20] W. Gata, G. Grand, R. Fatmasari, B. Baharuddin, Y. E. Patras, R. Hidayat, S. Tohari, and N. K. Wardhani, "Prediction of teachers' lateness factors coming to school using C4.5, random tree, random forest algorithm," in *2nd International Conference on Research of Educational Administration and Management (ICREAM 2018)*. Bandung, Indonesia: Atlantis Press, Oct. 18, 2019, pp. 161–166.

[21] A. Lalonde, K. Herschderfer, D. Pascali-Bonaro, C. Hanson, C. Fuchtner, and G. H. A. Visser, "The international childbirth initiative: 12 steps to safe and respectful MotherBaby-Family maternity care," *International Journal of Gynecology & Obstetrics*, vol. 146, no. 1, pp. 65–73, 2019.

[22] A. A. Daniels and A. Abuosi, "Improving emergency obstetric referral systems in low and middle income countries: A qualitative study in a tertiary health facility in Ghana," *BMC Health Services Research*, vol. 20, no. 1, pp. 1–10, 2020.

[23] R. Rahim, I. Zufria, N. Kurniasih, M. Y. Simargolang, A. Hasibuan, D. U. Sutiksno, R. F. Nanuru, J. N. Anamofa, A. S. Ahmar, and A. D. GS, "C4.5 classification data mining for inventory control," *International Journal of Engineering & Technology*, vol. 7, no. 2.3, pp. 68–72, 2018.

[24] E. M. Moegni and D. Ocviyanti, *Buku saku pelayanan kesehatan ibu di fasilitas kesehatan dasar dan rujukan*. Kementerian Kesehatan Republik Indonesia, 2013.

[25] A. Sofian, *Rustam Mochtar sinopsis obstetri*.

EGC, 2012.

[26] A. Craik, Y. He, and J. L. Contreras-Vidal, "Deep learning for electroencephalogram (EEG) classification tasks: A review," *Journal of Neural Engineering*, vol. 16, no. 3, pp. 1–28, 2019.

[27] W. Liu, J. Su, Z. Mao, P. Jin, Y. Huang, C. Dou, L. Zhou, and Y. Shang, "Research on text classification method of distribution network equipment fault based on deep learning," in *2020 7th IEEE International Conference on Cyber Security and Cloud Computing (CSCloud)/2020 6th IEEE International Conference on Edge Computing and Scalable Cloud (EdgeCom)*. New York, USA: IEEE, Aug. 1–3, 2020, pp. 11–16.

[28] W. Liu, B. Wang, and Z. Song, "Failure prediction of municipal water pipes using machine learning algorithms," *Water Resources Management*, vol. 36, no. 4, pp. 1271–1285, 2022.

[29] A. Rana and R. Pandey, "A review of popular decision tree algorithms in data mining," *Asian Journal of Multidimensional Research*, vol. 10, no. 10, pp. 230–237, 2021.

[30] M. Rakhra, P. Soniya, D. Tanwar, P. Singh, D. Bordoloi, P. Agarwal, S. Takkar, K. Jairath, and N. Verma, "Crop price prediction using random forest and decision tree regression:-A review," *Materials Today: Proceedings*, 2021.

[31] S. R. Hashemi, S. S. M. Salehi, D. Erdogmus, S. P. Prabhu, S. K. Warfield, and A. Gholipour, "Asymmetric loss functions and deep densely-connected networks for highly-imbalanced medical image segmentation: Application to multiple sclerosis lesion detection," *IEEE Access*, vol. 7, pp. 1721–1735, 2018.

[32] K. Sprute, V. Kramer, S. A. Koerber, M. Meneses, R. Fernandez, C. Soza-Ried, M. Eiber, W. A. Weber, I. Rauscher, K. Rahbar *et al.*, "Diagnostic accuracy of 18F-PSMA-1007 PET/CT imaging for lymph node staging of prostate carcinoma in primary and biochemical recurrence," *Journal of Nuclear Medicine*, vol. 62, no. 2, pp. 208–213, 2021.

[33] M. A. Muslim, S. H. Rukmana, E. Sugiharti, B. Prasetiyo, and S. Alimah, "Optimization of C4.5 algorithm-based particle swarm optimization for breast cancer diagnosis," *Journal of Physics: Conference Series*, vol. 983, pp. 1–7, 2018.

[34] M. Pal and S. Parija, "Prediction of heart diseases using random forest," *Journal of Physics: Conference Series*, vol. 1817, pp. 1–8, 2021.

[35] Y. A. Saadoon and R. H. Abdulamir, "Improved random forest algorithm performance for big data," *Journal of Physics: Conference Series*, vol. 1897, pp. 1–13, 2021.

[36] Ö. Akar and O. Güngör, "Classification of multispectral images using random forest algorithm," *Journal of Geodesy and Geoinformation*, vol. 1, no. 2, pp. 105–112, 2012.