Improved Decision Tree Accuracy (C4.5) with Attribute Reduction Using Forward Selection in Data Classification

Raden Mas Rizky Yohannes Cristanto[1], Elviawaty Muisa Zamzami[2], Fahmi[3]

{ramariyocris@gmail.com[1], elvi_zamzami@usu.ac.id[2], fahmimn@usu.ac.id[3]}

Faculty of Computer Science and Information Technology, University of Sumatera Utara, Medan, Indonesia[1,2]
Faculty of Engineering, University of Sumatera Utara, Medan, Indonesia[3]

Abstract. The main process in the formation of Decision Tree C4.5 is the separation of attributes. However, the attribute separation procedure in C4.5 still cannot optimize prediction accuracy in decision tree formation because unwanted features can lead to noisy data and less relevant features, which in turn can result in very large decision tree sizes (overfitting). As a result, the data becomes unbalanced and the classification accuracy of the Decision Tree C4.5 model becomes lower. To improve the accuracy of the classification process, attribute reduction is performed as a technique to simplify less relevant attributes. Therefore, forward selection is proposed as an attribute reduction method to produce mutually uncorrelated features, which are then used in Decision Tree C4.5 for classification. This study used datasets from the UCI Machine Learning Repository and Kaggle.com namely Diabetic Retinopathy Debrecen and South German Credit. Debrecen's Diabetic Retinopathy consists of 1,151 data records with 20 attributes, while South German Credit consists of 1000 data records with 20 attributes. Evaluation of classification performance is carried out based on the calculation of the Confusion Matrix. The test results showed that the proposed method was able to increase classification accuracy by 7.68%. Therefore, forward selection is considered an effective technique in reducing attributes and improving classification accuracy in Decision Tree C4.5

Keywords: Classification, Decision Tree C4.5, Attribute Reduction, Forward Selection.

1 Introduction

Decision trees are one of the most popular machine learning algorithms, dividing data repeatedly to form classes or groups [1]. Decision tree as a classification method is very effective [2], where classification tasks are modeling with a set of hierarchical decisions on feature variables in the form of a tree [3]. Classification algorithms in the decision tree include ID3, C4.5, and CART [4]. The research in this paper uses a C4.5 decision tree. In C4.5 the decision tree uses the concept of entropy of classification information, using the separation criterion Improved Iterative Dichotomi 3 (ID3) called Gain Ratio [5]. In the research of Hasdyna, et al [6] used Gain Ratio in reducing attributes to improve the performance of the

ABEC 2023, September 21, Bengkalis, Indonesia
Copyright © 2024 EAI
DOI 10.4108/eai.21-9-2023.2342901

K-Nearest Neighbor (KNN) algorithm. In the C4.5 method using Gain Ratio (GR), where the attribute with the highest gain is chosen as root.

The decision tree classification method can go wrong if it is overfitting or the data is too noisy. Unnecessary nodes generate noisy data and attributes with low correlation. This leads to overfitting in the decision tree. Overfitting makes the classifier decrease in accuracy due to failure to properly generalize unseen instances [7]. For this, it is necessary to pruning [2]. Pruning is the process of cutting or removing unwanted nodes and branches, overfitting the decision tree [8].

There is research to eliminate variables (attributes) irrelevant to partial least squares regression models using forward selection [9]. Selection of attribute variables to produce a simple, robust and easily interpretable model against the selected data set. On a study [10], Selection or selection of attributes applied to the classification of heart disease. The researcher used K-Nearest Neighbor and the Forward Selection attribute selection method, resulting in a precision value of 78.66%. Research results by [10] obtained an increase compared to the precision value without Forward Selection of 73.44%. With the results of this study, the K-Nearest Neighbor algorithm using Forward Selection can increase the accuracy value.

The method for attribute reduction Forward Selection is a stepwise regression method that starts by adding variables one by one based on which variables are most statistically relevant and which will eliminate extraneous or irrelevant variables one by one statistically [11]. Each process considers statistical consequences using criteria determined from the standard estimation of coefficients. Previously, the data normalization process was carried out using the min-max method to avoid large value weights that could complicate the computational process in the test program. Min-Max normalization is the simplest method based on rescaling the range of feature values to a scale of [0,1] or [-1,1] [12].

2 Research Methods

2.1 Stages of Research

The following contains the stages of research (Figure 1) along with their explanations.

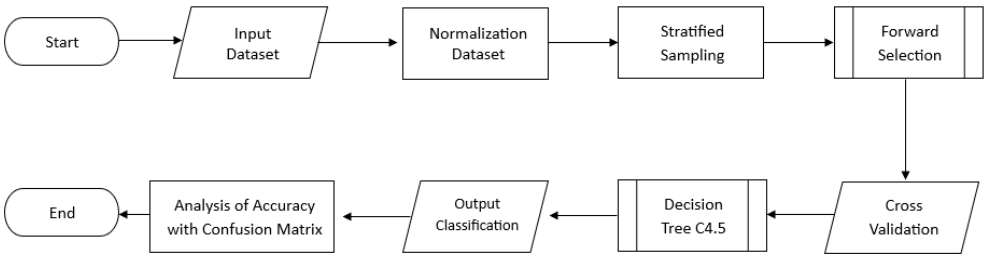

Fig. 1. Stages of research conducted by researchers.

Dataset

The data used in this study consisted of two (2) datasets. The first dataset was Debrecen's Diabetic Retinopathy (http://archive.ics.uci.edu/ml). This data is predictive data for medical testing tested on patients suspected of being affected by Diabetic Retinopathy. The number of attributes is 20 attributes, with the number of data records is 1151 records and consists of 2 attribute classes. The second dataset used is the dataset obtained from Kaggle.com, namely South Germany Credit which is a credit application dataset. The number of data records in the data set is 1000 data records with the number of data attributes, namely 20 attributes and 1 output attribute with 2 attribute classes. The following data on Cervical Cancer and South German Credit used are listed in Table 1.

Table 1. Dataset used

Dataset	Dataset Type	Number of Attributes	Number of Data Records
Diabetic Retinopathy Debren	Multivariate	20	1.151
South German Credit	Multivariate	20	1.000

Data Normalization

Data normalization aims to remove any more invalid data before proceeding to the next step [13]. Normalize data using the Min-Max method with the following formula [14]:

$$\frac{(Data-Min)*(NewMax-NewMin)}{(Max-Min)} + NewMin \quad (1)$$

Decision Tree C4.5

Decision Tree C4.5 to form a decision tree which is a very powerful classification and prediction method [15]. The decision tree method transforms very large facts into decision trees that represent rules that can be easily understood. The stages in Decision Tree C4.5, namely [16]:

a. Calculates the Entropy value of each attribute:

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

b. Calculate the value of Information Gain on each attribute:

$$InfoGain\ (S, A) = Entropy\ (S) - \sum_{i=1}^{n} \frac{|S_i|}{|S|} \times Entropy(S_i) \quad (3)$$

c. Calculate the Split Information value for each attribute:

$$SplitInfo_A(D) = -\sum_{j=1}^{v} \frac{|D_j|}{|D|} \times log_2\left(\frac{D_j}{D}\right) \quad (4)$$

d. Calculates the Gain Ratio value for each attribute:

$$GainRatio\ (A) = \frac{InfoGain\ (A)}{SplitInfo\ (A)} \quad (5)$$

e. The attribute has the highest Gain Ratio selected to be a measure (splitting attribute) and attribute that has a Gain Ratio value that is lower than root (root) selected to branch,

f. Calculate the value of Gain Ratio each attribute with exclude attribute selected to be root in previous stage,
g. The attribute that has the highest Gain Ratio is chosen to be branches. Repeat steps 4 and 5 until the resulting value is Gain = 0 for all remaining attributes.

Forward Selection

One of the many attribute reduction processes that involves an empty set of attributes that need to be changed is Forward Selection [17]. Then, each attribute is evaluated individually, and the best attribute is selected with the highest possible amplification. Then, proceed to the next iteration of testing continuously and stop until the tested attribute does not have a significant impact on accuracy [18]. Forward Selection is formulated as follows [19]:

a. Determining the initial model.

$$\hat{y} = b_0 \tag{6}$$

Input variable response with each predictor variable, e.g. *X1, X2*, ... Xn is related to *y*. Suppose *X1* so that form a model:

$$\hat{y} = b_0 + b_1 X_1 \tag{7}$$

b. Test F against the first selected variable provided that if Fcalculate < Ftable then the selected variable is deleted and the process stops. If Fcalculate > Ftable, then the selected variable has a real influence on the variable related to y so that it deserves to be taken into account in the model.

$$\hat{y} = b_0 + b_1 X_1 + b_2 X_2 \tag{8}$$

Test F, f Fcalculate < Ftabel, then the process is stopped and the best model is the previous model. However, if Fhtung > Ftable, the free variable deserves to be included in the model. The process will end if there are no more variables left that can be inserted into the model.

Stratified Sampling

Stratified sampling is a technique of sampling by tracking that takes into account the levels (strata) in a population [20]. The dataset tested is first broken down into separate layers and then sampled by tracking based on the layer that has been created. The stages of stratified sampling are as follows [21]:

a. First divide the population N into subpopulations consisting of elements *N1, N2, N3, ..., NL*.
b. Thus, pepper cannot overlap between subpopulations, so $N1 + N2 + N3 + \cdots + NL = N$.
c. Finally, samples are taken by tracking each subpopulation by distributing the samples proportionally.

Before sampling, it is important to determine the sample size. The sample taken should reflect the overall situation. There are several ways to determine the sample size. One of Slovin's theories that is most widely applied and used is described by the following formula:

$$n = \frac{N}{1+Ne^2} \tag{9}$$

Information:
n = Large sample
N = Size of population

e = Precision value

2.2 Data Classification Performance Measurement

Cross Validation

Cross Validation is the process of converting a set of data into a set of subsets with the same size. A subset of each used for testing and training data. As a result, each data has the same opportunity for training and testing data. Cross Validation is used to correct the wrong data testing results. 10-Fold Cross-Validation will reduce testing by 10 times, and the result will be a percentage of 10 tests [22].

Confusion Matrix

In measuring the results of classification performance in line research carried out by testing Confusion Matrix to obtain the results of Accuracy, Precision, Recall and F1-Score and also to analyses the quality of the classifier in recognizing different classes [23]. Table Confusion Matrix can be seen in Table 2.

Table 2. Table of confusion matrix

Actual Class	Assigned Class	
	Positive	Negative
Positive	True Positive	False Negative
Negative	False Positive	True Negative

The True Positive and True Negative are actual conditions where the predicted results correspond to the actual conditions that occur. While False Positive and False Negative are conditions in which the results of prediction do not correspond to the actual conditions. Then to calculate the value Accuracy, precision, recall and f-1 score can be calculated using the formula:

$$Accuracy = \frac{TP+FN}{TP+TN+FP+FN} \tag{10}$$

$$Recall = \frac{TP}{TP+FN} \tag{11}$$

$$Precision = \frac{TP+TN}{TP+FP} \tag{12}$$

$$F1 - Score = 2 \times \frac{Recall \times Precision}{Recall+Precision} \tag{13}$$

3. Result and Discussion

3.1 Initial Data Processing

Based on the collected data, there is no need for data integration, because the storage place used only contains data from one specific storage place, so no further data integration is needed. However, the data normalization process must still be carried out. This is due to the possibility of inaccurate data or missing numbers in the data that has been collected. The results of data normalization can be seen in table 3 and table 4 below:

Table 3. Diabetic retinopathy debrecen normalization results

No.	X1	X2	X3	X4	.....	X19	*Class*
1	0.0	0.205	1.0	0.2	.....	1.0	Class 0
2	0.0	0.073	1.0	0.0	.....	1.0	Class 0
3	0.333	0.117	0.5	0.9	.....	1.0	Class 1
4	0.0	0.117	1.0	0.0	.....	0.0	Class 0
5	0.0	0.117	1.0	0.0	.....	0.0	Class 1
6	0.0	0.088	1.0	0.0	.....	0.0	Class 1
7	0.0	0.058	1.0	0.0	.....	0.0	Class 1
8	0.0	0.029	1.0	0.0	.....	0.0	Class 0
9	1.0	0.205	1.0	0.3	.....	1.0	Class 1
10	0.333	0.294	0.5	0.3	.....	1.0	Class 1
11	0.0	0.102	1.0	0.0	.....	1.0	Class 0
12	0.0	0.382	1.0	0.1	.....	1.0	Class 0
13	0.0	0.029	1.0	0.3	.....	1.0	Class 1
14	0.333	0.647	0.75	1.0	.....	1.0	Class 1
15	0.0	0.205	0.5	0.3	.....	1.0	Class 0
16	0.0	0.029	0.5	0.3	.....	1.0	Class 0
⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮
⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮
⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮
⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮
⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮
⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮
1150	1	1	39	36	.....	1	Class 1
1151	1	1	7	7	.....	0	Class 0

Table 4. South german credit normalization results

No.	X1	X2	X3	X4	.....	X19	*Class*
1	0.0	0.205	1.0	0.2	.....	1.0	Class 0
2	0.0	0.073	1.0	0.0	.....	1.0	Class 0
3	0.333	0.117	0.5	0.9	.....	1.0	Class 1
4	0.0	0.117	1.0	0.0	.....	0.0	Class 0
5	0.0	0.117	1.0	0.0	.....	0.0	Class 1
6	0.0	0.088	1.0	0.0	.....	0.0	Class 1
7	0.0	0.058	1.0	0.0	.....	0.0	Class 1
8	0.0	0.029	1.0	0.0	.....	0.0	Class 0
9	1.0	0.205	1.0	0.3	.....	1.0	Class 1
10	0.333	0.294	0.5	0.3	.....	1.0	Class 1
11	0.0	0.102	1.0	0.0	.....	1.0	Class 0
12	0.0	0.382	1.0	0.1	.....	1.0	Class 0
13	0.0	0.029	1.0	0.3	.....	1.0	Class 1
14	0.333	0.647	0.75	1.0	.....	1.0	Class 1

15	0.0	0.205	0.5	0.3	.....	1.0	Class 0
16	0.0	0.029	0.5	0.3	.....	1.0	Class 0
⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮
⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮
⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮
⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮
⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮
999	1	1	39	36	.....	1	Class 1
1000	1	1	7	7	.....	0	Class 0

The data that has been obtained must be processed first before creating a model. Initialization of data is done using stratified sampling. Data sampling is carried out anonymously from the population by considering the ratio of data distribution to create a new data collection with dimension sampling determined using the slovin formula and 90% significance limit symbol with formula as follows:

$$Sample = \frac{583}{1 + 583 * 0.1^2} = 85$$

The sample size in the Diabetic Retinopathy Debrecen Dataset was determined using the slovin formula with the following calculations:

$$Sample = \frac{1.151}{1 + 1.151 * 0.1^2} = 92$$

From 92 samples taken in the Diabetic Retinopathy Debrecen Dataset, allocate proportionally from each attribute class in the Diabetic Retinopathy Debrecen Dataset so that the samples taken reflect the population. l Calculation of sample allocation as follows:

$$Sample\ (Class\ 0) = \frac{540}{1.151} \times 92 = 43$$

$$Sample\ (Class\ 1) = \frac{611}{1.151} \times 92 = 49$$

The sample size in the South German Credit Dataset is determined using the slovin formula with the following calculation:

$$Sample = \frac{1.000}{1 + 1.000 * 0.1^2} = 91$$

From the 92 samples taken in the South German Credit Dataset, allocate proportionally from each attribute class in the South German Credit Dataset so that the sample taken reflects the population. Calculation of sample allocation as follows:

$$Sample\ (Good) = \frac{700}{1.000} \times 91 = 64$$

$$Sample\ (Bad) = \frac{300}{1.000} \times 91 = 27$$

3.2 Model Experimentation and Testing

After getting a new dataset with a small size (sampling), then test by selecting the best attributes and removing those that have no effect. To select the best attribute with Forward Selection. Test each attribute individually by building a model, then test the model to see how accurate the results are. Select attributes with the highest precision. If the tested properties do not significantly improve accuracy, the process continues and stops. Experiments and tests on this study using RapidMiner Studio.

Fig. 2 contains an assessment of the attributes of each dataset tested which is then shown in Table 5.

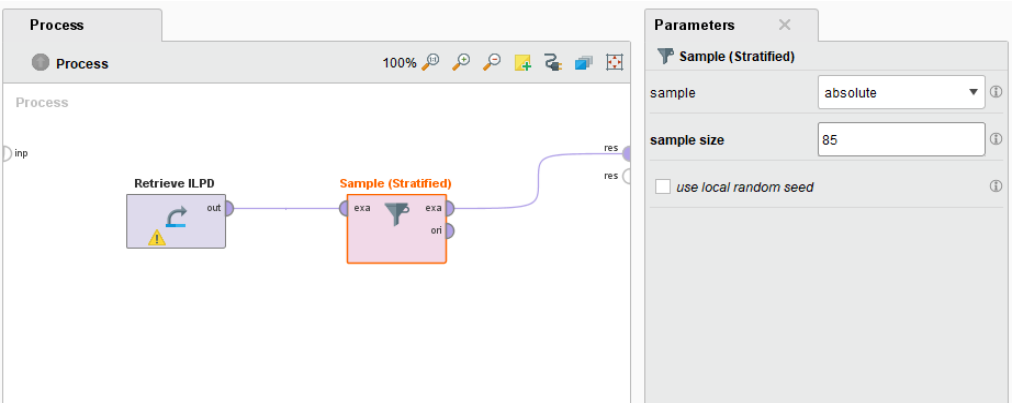

Fig. 2. Selection testing scheme forward selection feature

Table 5. Results of attribute reduction with forward selection on diabetic retinopathy debrecen dataset

No	Attribute	Weight
1	Attributes 0	0
2	Attributes 1	0
3	Attributes 2	0
4	Attributes 3	0
5	Attributes 4	0
6	Attributes 5	0
7	Attributes 6	1
8	Attributes 7	0
9	Attributes 8	0
10	Attributes 9	0
11	Attributes 10	1

12	Attributes 11	0
13	Attributes 12	0
14	Attributes 13	0
15	Attributes 14	0
16	Attributes 15	0
17	Attributes 16	0
18	Attributes 17	0
19	Attributes 18	0

Table 6. Results of attribute reduction with forward selection on diabetic retinopathy debrecen dataset

No	Attribute	Weight
1	Status	1
2	Duration	1
3	Credit History	1
4	Purpose	0
5	Amount	0
6	Savings	0
7	Employment Duration	0
8	Installment Rate	0
9	Personal Status Sex	0
10	Other Debtors	0
11	Present Residence	0
12	Property	0
13	Age	0
14	Other Installment Plants	0
15	Housing	0
16	Number Credits	0
17	Job	0
18	People Liable	0
19	Telephone	0
20	Foreign Worker	0

Based on Table 5 and Table 6, the attribute that has the weight 1 is the best attribute and selected for created model. While the attribute with l0 weight is the attribute does not have the influence of and will removed so that produces a (new) dataset whose number of attributes less. After testing is done, then enter the modeling stage by creating a decision tree as in Fig. 3:

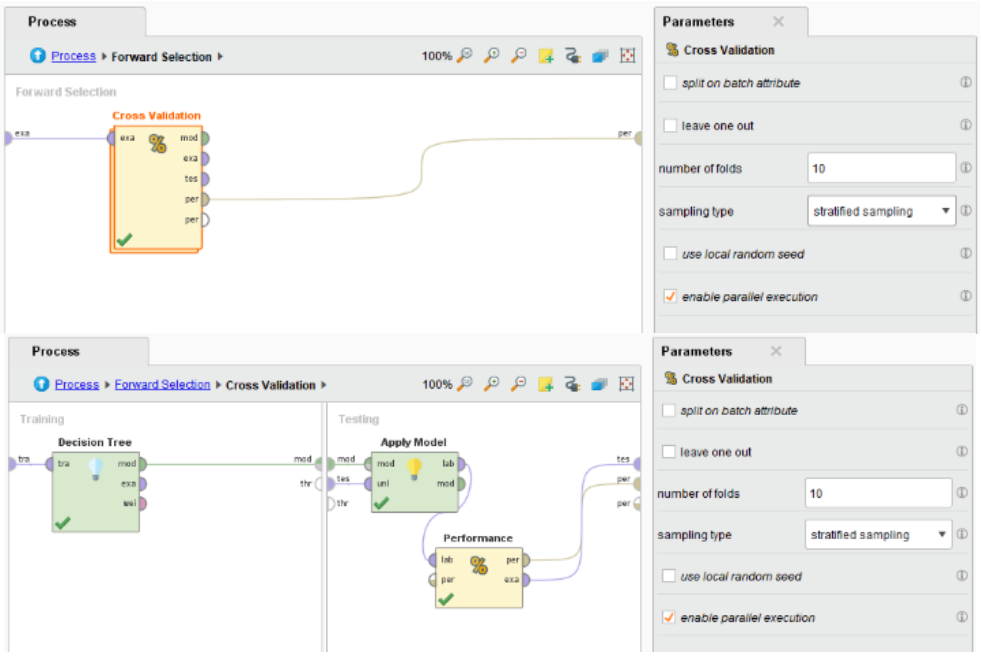

Fig. 3. Classification model test scheme

The process in Fig. 3 produces the pattern described in Fig. 4:

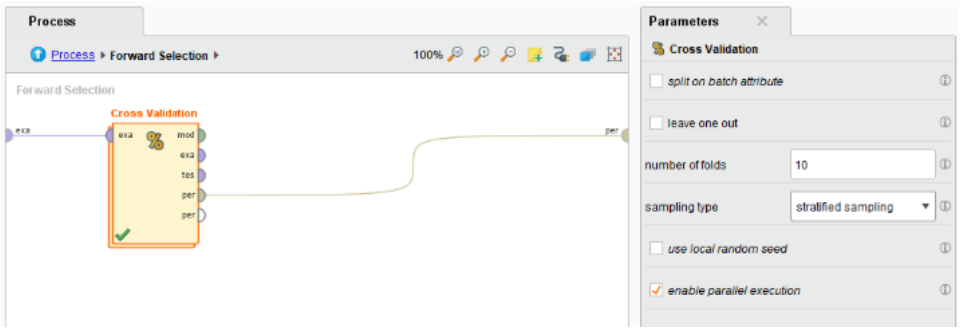

Fig. 4. Classification model accuracy testing scheme

A. Evaluation of Testing and Validation of Results

The test was carried out with the help of using RapidMiner Studio software based on Decision Tree C4.5 without reduction of Tree 4.5 attributes and Decision Tree C4.5 with attributes of Forward Selection reduction results. Then the classification evaluation is calculated based on the Confusion Matrix with the results can be seen in Table 7.

Table 7. Evaluation of classification results

Data Set	Accuracy Comparison (%)		Accuracy Difference (%)
	C4.5	C4.5 + Forward Selection	
Diabetic Retinopathy Debrecen	53.09	55.00	1.91
South German Credit	70.00	83.44	13.44
Average	61.54	69.22	7.68

Based on the results in Table 6, the accuracy of Decision Tree C4.5 + Stratified Sampling + Forward Selection was compared with Decision Tree C4.5 without optimization described in the following Fig. 5:

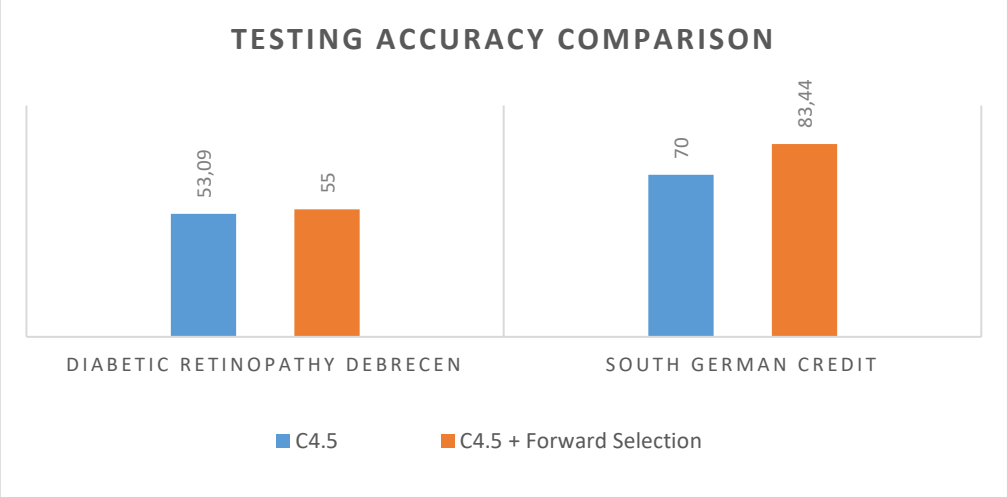

Fig. 5. Classification model test scheme

Based on 5 Figure Decision Tree C4.5 + Stratified Sampling + Forward Selection was able to increase accuracy by 7.68% when compared to Decision Tree C4.5 No optimization in data classification.

4. Conclusion

Based on the test results show that the proposed method is able to improve the accuracy of classification on Decision Tree C4.5 by reducing attributes using Forward Selection. Therefore, forward selection is considered an effective technique in reducing attributes and improving classification accuracy in Decision Tree C4.5. The increase in accuracy obtained in Debrecen Diabetic Retinopathy was 1.91%, while the increase in accuracy obtained in South German Credit was 13.44%. The average result of increasing accuracy in all data sets was 7.68%. The average accuracy rate on Decision Tree C4.5 + Forward Selection is 69.22% and higher than the average accuracy value on Decision Tree C4.5 which obtained an average accuracy of

61.54%. The results obtained after attribute reduction in Decision Tree C4.5 with Forward Selection are much more accurate than Decision Tree C4.5 natively.

References

[1] M. M. Mijwil and R. A. Abttan, "Utilizing the Genetic Algorithm to Pruning the C4.5 Decision Tree Algorithm," *Asian J. Appl. Sci.*, vol. 9, no. 1, pp. 2321–0893, Feb. 2021, doi: 10.24203/AJAS.V9I1.6503.

[2] F. M. J. Mehedi Shamrat, S. Chakraborty, M. M. Billah, P. Das, J. N. Muna, and R. Ranjan, "A comprehensive study on pre-pruning and post-pruning methods of decision tree classification algorithm," *Proc. 5th Int. Conf. Trends Electron. Informatics, ICOEI 2021*, pp. 1339–1345, Jun. 2021, doi: 10.1109/ICOEI51242.2021.9452898.

[3] C. C. Aggarwal, *Data Mining*. Cham: Springer International Publishing, 2015. doi: 10.1007/978-3-319-14142-8.

[4] J. Jiang, X. Zhu, G. Han, M. Guizani, and L. Shu, "A Dynamic Trust Evaluation and Update Mechanism Based on C4.5 Decision Tree in Underwater Wireless Sensor Networks," *IEEE Trans. Veh. Technol.*, vol. 69, no. 8, pp. 9031–9040, Aug. 2020, doi: 10.1109/TVT.2020.2999566.

[5] R. Rusito and M. Firmansyah, "IMPLEMENTASI METODE DECISION TREE DAN ALGORITMA C4.5 UNTUK KLASIFIKASI DATA NASABAH BANK," *J. Ilm. Infokam*, vol. 12, no. 2, Oct. 2016, doi: 10.53845/INFOKAM.V12I2.103.

[6] N. Hasdyna, B. Sianipar, and E. M. Zamzami, "Improving The Performance of K-Nearest Neighbor Algorithm by Reducing The Attributes of Dataset Using Gain Ratio," *J. Phys. Conf. Ser.*, vol. 1566, no. 1, p. 012090, Jun. 2020, doi: 10.1088/1742-6596/1566/1/012090.

[7] A. Amro, M. Al-Akhras, K. El Hindi, M. Habib, and B. A. Shawar, "Instance Reduction for Avoiding Overfitting in Decision Trees," *J. Intell. Syst.*, vol. 30, no. 1, pp. 438–459, Jan. 2021, doi: 10.1515/JISYS-2020-0061/MACHINEREADABLECITATION/RIS.

[8] X. Zhou and D. Yan, "Model tree pruning," *Int. J. Mach. Learn. Cybern.*, vol. 10, no. 12, pp. 3431–3444, Dec. 2019, doi: 10.1007/S13042-019-00930-9/METRICS.

[9] D. C. Whitley, M. G. Ford, and D. J. Livingstone, "Unsupervised Forward Selection: A Method for Eliminating Redundant Variables," *J. Chem. Inf. Comput. Sci.*, vol. 40, no. 5, pp. 1160–1168, 2000, doi: 10.1021/CI000384C.

[10] J. Zeniarja, A. Ukhifahdhina, and A. Salam, "Diagnosis Of Heart Disease Using K-Nearest Neighbor Method Based On Forward Selection," *J. Appl. Intell. Syst.*, vol. 4, no. 2, pp. 39–47, Mar. 2020, doi: 10.33633/jais.v4i2.2749.

[11] G. Smith, "Step away from stepwise," *J. Big Data*, vol. 5, no. 1, pp. 1–12, Dec. 2018, doi: 10.1186/S40537-018-0143-6/FIGURES/1.

[12] D. Borkin, A. Némethová, G. Michaľčonok, and K. Maiorov, "Impact of Data Normalization on Classification Model Accuracy," *Res. Pap. Fac. Mater. Sci. Technol. Slovak Univ. Technol.*, vol. 27, no. 45, pp. 79–84, Sep. 2019, doi: 10.2478/RPUT-2019-0029.

[13] A. I. Lubis, U. Erdiansyah, and R. Siregar, "Comparison of Accuracy in Naïve Bayes and Random Forests in Classification of Liver Disease," *CESS (Journal Comput. Eng. Syst. Sci.*, vol. 7, no. 1, pp. 81–89, Dec. 2021, doi: 10.24114/CESS.V7I1.28888.

[14] U. Erdiansyah, A. I. Lubis, and K. Erwansyah, "Komparasi Metode K-Nearest Neighbor dan Random Forest Dalam Prediksi Akurasi Klasifikasi Pengobatan Penyakit Kutil," *J. MEDIA Inform. BUDIDARMA*, vol. 6, no. 1, pp. 208–214, Jan. 2022, doi: 10.30865/MIB.V6I1.3373.

[15] E. Patimah *et al.*, "Klasifikasi Penyakit Liver dengan Menggunakan Decision Tree," *Pros. Semin. Nas. Mhs. Bid. Ilmu Komput. dan Apl.*, vol. 2, no. 1, pp. 655–659, Jul. 2021, Accessed: Mar.

27, 2023. [Online]. Available: https://conference.upnvj.ac.id/index.php/senamika/article/view/1388

[16] P. Handayani *et al.*, "Prediksi Penyakit Liver Dengan Menggunakan Metode Decision Tree dan Neural Network," *CESS (Journal Comput. Eng. Syst. Sci.*, vol. 4, no. 1, pp. 55–59, Feb. 2019, doi: 10.24114/CESS.V4I1.11528.

[17] T. B. Sasongko and O. Arifin, "Implementasi Metode Forward Selection pada Algoritma Support Vector Machine (SVM) dan Naive Bayes Classifier Kernel Density (Studi Kasus Klasifikasi Jalur Minat SMA)," *J. Teknol. Inf. dan Ilmu Komput.*, vol. 6, no. 4, pp. 383–388, Jul. 2019, doi: 10.1016/j.ijar.2008.08.008.

[18] M. Reif and F. Shafait, "Efficient feature size reduction via predictive forward selection," *Pattern Recognit.*, vol. 47, no. 4, pp. 1664–1673, Apr. 2014, doi: 10.1016/J.PATCOG.2013.10.009.

[19] S. Luo and S. Ghosal, "Forward selection and estimation in high dimensional single index models," *Stat. Methodol.*, vol. 33, pp. 172–179, Dec. 2016, doi: 10.1016/J.STAMET.2016.09.002.

[20] S. F. Ulya, Y. Sukestiyarno, and P. Hendikawati, "Analisis Prediksi Quick Count Dengan Metode Stratified Random Sampling Dan Estimasi Confidence Interval Menggunakan Metode Maksimum Likelihood," *Unnes J. Math.*, vol. 7, no. 1, pp. 108–119, 2018.

[21] I. Ubaedi and Y. M. Djaksana, "OPTIMASI ALGORITMA C4.5 MENGGUNAKAN METODE FORWARD SELECTION DAN STRATIFIED SAMPLING UNTUK PREDIKSI KELAYAKAN KREDIT," *JSiI (Jurnal Sist. Informasi)*, vol. 9, no. 1, pp. 17–26, Mar. 2022, doi: 10.30656/JSII.V9I1.3505.

[22] U. Erdiansyah, A. I. Lubis, and G. Syahputra, "Klasifikasi Penyakit Diabetic Retinopathy Menggunakan Multilayer Perceptron," *J. Artif. Intell. Softw. Eng.*, vol. 2, no. 1, May 2022, doi: 10.30811/JAISE.V2I1.3084.

[23] A. Yuliana and D. B. Pratomo, "ALGORITMA DECISION TREE (C4.5) UNTUK MEMPREDIKSI KEPUASAN MAHASISWA TERHADAP KINERJA DOSEN POLITEKNIK TEDC BANDUNG," *Pros. SEMNAS INOTEK (Seminar Nas. Inov. Teknol.*, vol. 1, no. 1, pp. 377–384, 2017, doi: 10.29407/INOTEK.V1I1.429.