Ceddi Journal of Information Systems and Technology (JST)
Vol. 2 No. 2 December (2023)
https://doi.org/10.56134/jst.v2i2.47

ISSN:2829-808X(print)
ISSN:2829-6575(online)

Application of the Adaptive Boosting Method to Increase the Accuracy of Classification of Type Two Diabetes Mellitus Patients Using the Decision Tree Algorithm

Hao Chieh Chiua[1]*, Robbi Rahim[2], Mahmud Mustapa[3], Kamaruddin[4], Akbar Hendra[5], Asnimar[6], Omita Abigail[7]

[1]*Department of Computer Science & Information Engineering, National Taiwan University, Taiwan*
[2]*Department of Management, Sekolah Tinggi ilmu Manajemen Sukma, Indonesia*
[3]*Electronics engineering education, Universitas Negeri Makassar, Indonesia*
[4-7]*Department of Informatics, Universitas Teknologi Akba Makassar, Makassar, Indonesia*

Abstract

One of the data mining processes that is often used in machine learning is the data classification process. A decision tree is a classification algorithm that has the advantage of being easy to visualize because of its simple structure. However, the decision tree algorithm is quite susceptible to incorrect classification calculations due to the presence of noise in the data or imbalance in the data, which can reduce the overall level of accuracy. Therefore, the decision tree algorithm should be combined with other methods that can increase the accuracy of classification performance. Machine Learning is used through an artificial intelligence approach to solve problems or carry out optimization. Adaptive Boosting is used to optimize classification calculations. This study aims to examine the performance of Adaptive Boosting in the process of classifying second-degree diabetes mellitus patients using the Decision Tree algorithm. Diabetes mellitus is known as a chronic condition of the human body, the cause of which is an increase in the body's blood sugar levels because the body is unable to produce insulin or is unable to utilize insulin effectively, which is usually referred to as hyperglycemia.. By using a 60:40 data split, the Decision Tree algorithm produces an accuracy value of 95.71%, while the Adaptive Boosting-based Decision Tree results reach a value of 98.99%.

Keywords: Adaptive Boosting; Decision Tree; Data Mining; Classification; Diabetes Mellitus.

Received: 15 June 2023 Revised: 10 November 2023 Accepted: 6 December 2023

Introduction

Diabetes mellitus is known as a chronic condition of the human body, the cause of which is an increase in the body's blood sugar levels because the body is unable to produce insulin or is unable to utilize insulin effectively, which is usually referred to as hyperglycemia.(Suryati & Kep, 2021). Symptoms of severe hyperglycemia can include polydipsia, polyuria, polyphagia, unexplained weight loss, fatigue, decreased performance, impaired vision, and susceptibility to infection with ketoacidosis or non-ketoacidosis (Muhlisin, 2019). The classification of diabetes mellitus is divided into two levels, namely grade one and grade two diabetes mellitus(Prabowo et al., 2021). First-degree diabetes mellitus is a condition where there is an abnormality in the production of insulin, the way insulin works, or an abnormality in both(Fatimah, 2015). Second-degree diabetes mellitus is also often referred to as lifestyle diabetes(Sibarani, 2023).

In 2021, the International Diabetes Federation (IDF) researched that there are around 537 million people of average adult age worldwide (age range 20-79 years) who suffer from diabetes, which is equivalent to 1 in 10 people(Suprayitna et al., 2023). Indonesia is ranked fifth because the number of sufferers is around 19.47 million people(Sudargo et al., 2018). IDF also notes that 4 out of 5 people with diabetes (or around 81%) are residents of low to middle-income countries(Sudargo et al., 2018). Therefore, the IDF estimates that there are still around 44% of adults who suffer from diabetes but have not been diagnosed. Grade 2 diabetes mellitus is one of the most common types of diabetes, with around 90% of all diabetes mellitus cases falling into this category (Horizon, 2021).

*Corresponding author.
E-mail address: haochiehchiua@gmail.com (Hao Chiehc Hiua)

As time progresses, Machine Learning (ML) is used through an artificial intelligence (AI) approach to solve problems or carry out optimization(Ula et al., 2021). One of the algorithms that performs optimization is the Adaboost algorithm. Adaboost is used to boost accuracy results to produce better accuracy values. Accuracy by utilizing the Adaboost algorithm is used to evaluate performance in machine learning(Byna & Basit, 2020),(Kong et al., 2020). The definition of accuracy is the degree to which predicted values correlate with actual values(Argina, 2020). In searching for information or knowledge to be more accurate from data, there is a computational process that generally needs to use data mining in the form of predictions, descriptions, estimates, clustering, classification, and associations. (Urva et al., 2023). By the needs and characteristics of a case, to determine the latest knowledge pattern of existing data. Classification is one of the data mining processes most often implemented in processing data(Tangkelayuk, 2022). Classification is a process where models or functions are used to describe and separate certain classes of data(Kusuma et al., 2023). The purpose of this classification is to predict the class of objects that do not yet have a known class label(Lorena Br Ginting et al., 2015). In completing a computation using classification techniques, of course, various algorithms can be used, such as the Naïve Bayes algorithm, K-Nearst Neighbor (KNN), and Support Vector Machine (SVM), but the algorithm that is quite popular in handling data classification cases is Decision Tree (Permana et al., nd).

The classification method using the decision tree algorithm has various advantages that can be highlighted(Fajri, 2015). One of them is its ability to present decision tree visualizations that are easy to interpret. Apart from that, decision trees are also able to provide a high level of accuracy and are efficient in classifying attributes, both discrete and numerical data.(Pirmansyah & Wahyudi, 2023). Decision trees can overcome the problem of overfitting on continuous attributes, choose appropriate attribute selection, and produce efficient calculations(Junita & Bachtiar, 2019).

However, the decision tree algorithm is quite vulnerable to incorrect classification calculations due to the presence of noise in the data or imbalance in the data, which can reduce the overall level of accuracy.(Bisri & Wahono, 2015). Therefore, the decision tree algorithm should be combined with other methods that can increase the accuracy of classification performance(Iftikar & Sibaroni, 2022). One algorithm that can improve classifier performance is AdaBoost (Adaptive Boosting)(Pradana, 2018). This method assigns different weights to the training data distribution in each iteration. At each iteration, AdaBoost adds weight to incorrect classifications and reduces weight to correct classifications, effectively changing the distribution of the training data.(Nusrhendratno, 2022).

Several previous studies, such as research(Pebrianti et al., 2022)with the title "Implementation of the Adaboost Method to Optimize Diabetes Classification with the Naïve Bayes Algorithm" have implemented the Naïve Bayes algorithm to optimize the accuracy of data classification and combined it with the AdaBoost method. This research concluded that the Naïve Bayes Algorithm produced an accuracy of 0.7608. Meanwhile, the Naïve Bayes results boosted using the AdaBoost algorithm were 0.7694.

In addition, research conducted by(Domas & Rakhmadi, 2022)with the title "Improving Decision Tree Performance with AdaBoost for Classifying Lack of Transparency of Anti-Corruption Information" concludes that the application of the AdaBoost method in classifying class predictions is feasible to increase the level of better performance. Experimental results show that the accuracy of the Decision Tree is 69.5%, while the average calculation results from the Decision Tree using the AdaBoost method achieve greater accuracy, around 71.16%. In this research, researchers will combine the AdaBoost method using the Decision Tree algorithm on a dataset of type 2 diabetes mellitus patients provided on the kaggle.com site. Classification calculations using the RapidMiner tool.

Method

The research method for implementing the Adaptive Boosting Method in Increasing the Accuracy of Classification of Type Two Diabetes Mellitus Patients Using the Decision Tree Algorithm is:

1. Research stage
   The following are the stages carried out by the author in analyzing the application of the Adaptive Boosting Method in increasing the accuracy of classification of type two diabetes mellitus patients using the Decision Tree Algorithm(Aziz et al., 2023)are as follows:

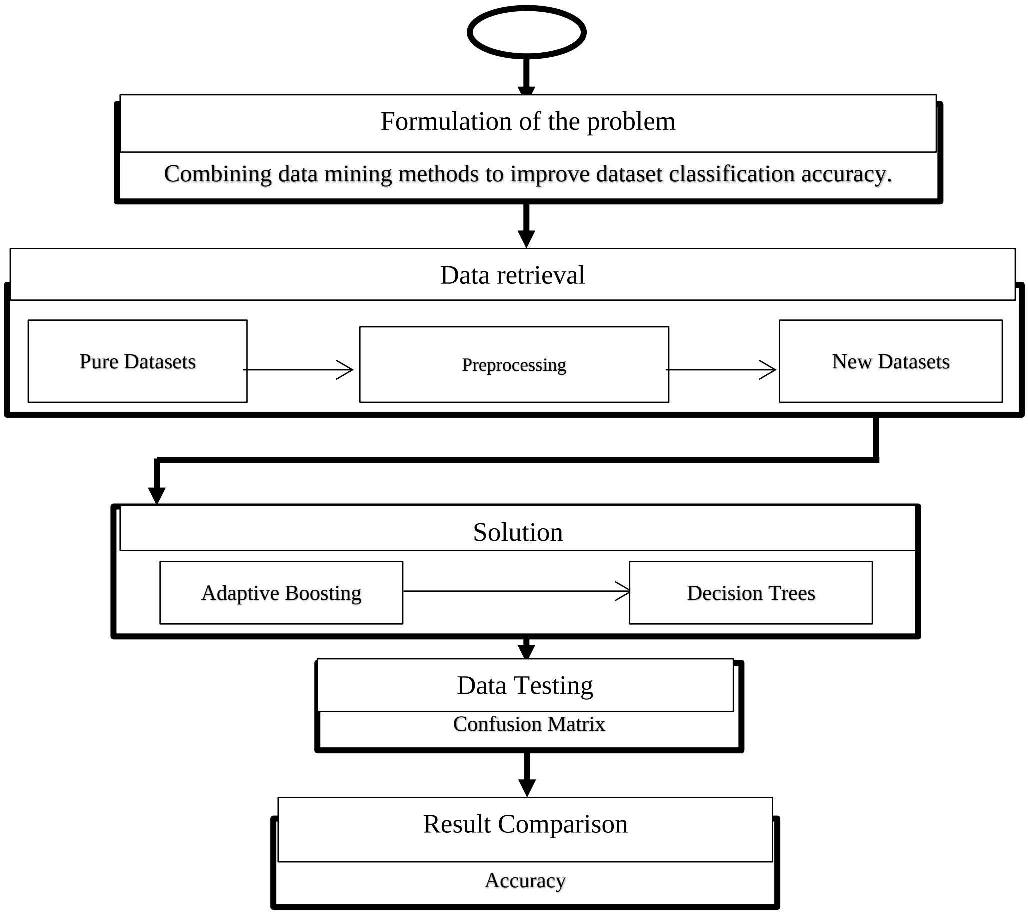


Figure 1. Research Flow

The desired result of this research is to compare the accuracy results of the dataset classification process, using the Decision Tree algorithm with the Decision Tree algorithm based on Adaptive Boosting.

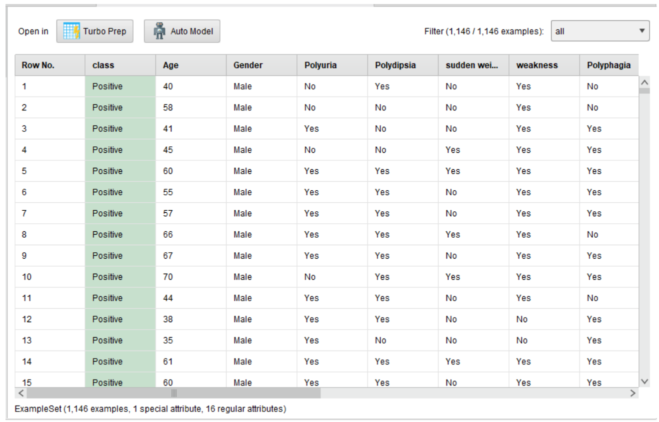
2. Data preprocessing
The dataset used in this research is the Early Stage Diabetes Risk Prediction Dataset which contains data on diabetes patients diagnosed with type two diabetes. (Hana, 2020). The amount of data used is 1,146 data records which are public data taken from the kaggle.com site with the following dataset information, attention to Table 1:

Table 1. Dataset Information

Data Characteristics	:	Multivariate	Amount of data	:	1,146
Attribute Type	:	Nominal	Missing Values	:	There is
Year of Donation	:	2020	Information	:	Cleaning
Number of Attributes	:	17	Number of Labels	:	1

The dataset before cleaning was 1,146, then after using the examples filter feature the remaining dataset was 990, because 156 data had missing values which had been deleted before processing. Then the 990 dataset is a new dataset and is ready for classification processing.

3. Algorithm Testing
This research uses the Rapid Miner tool to test the algorithm, and the dataset is in CSV format which will be imported into Rapid Miner using Read CSV.



Open in Turbo Prep Auto Model Filter (1,146 / 1,146 examples): all

Row No.	class	Age	Gender	Polyuria	Polydipsia	sudden wei...	weakness	Polyphagia
1	Positive	40	Male	No	Yes	No	Yes	No
2	Positive	58	Male	No	No	No	Yes	No
3	Positive	41	Male	Yes	No	No	Yes	Yes
4	Positive	45	Male	No	No	Yes	Yes	Yes
5	Positive	60	Male	Yes	Yes	Yes	Yes	Yes
6	Positive	55	Male	Yes	Yes	No	Yes	Yes
7	Positive	57	Male	Yes	Yes	No	Yes	Yes
8	Positive	66	Male	Yes	Yes	Yes	Yes	No
9	Positive	67	Male	Yes	Yes	No	Yes	Yes
10	Positive	70	Male	No	Yes	Yes	Yes	Yes
11	Positive	44	Male	Yes	Yes	No	Yes	No
12	Positive	38	Male	Yes	Yes	No	No	Yes
13	Positive	35	Male	Yes	No	No	No	Yes
14	Positive	61	Male	Yes	Yes	Yes	Yes	Yes
15	Positive	60	Male	Yes	Yes	No	Yes	Yes

ExampleSet (1,146 examples, 1 special attribute, 16 regular attributes)

Figure 2. Display of Dataset Imported to Rapid Miner

The initial dataset was 1,146 records. To carry out the data cleaning process, researchers use the Filter Examples feature, which is as follows, attention to Figure 3:

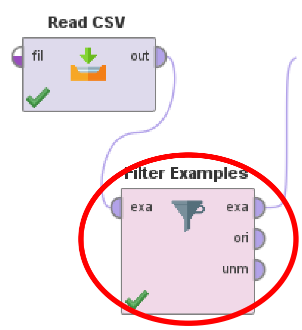


Figure 3. Data Cleaning Process

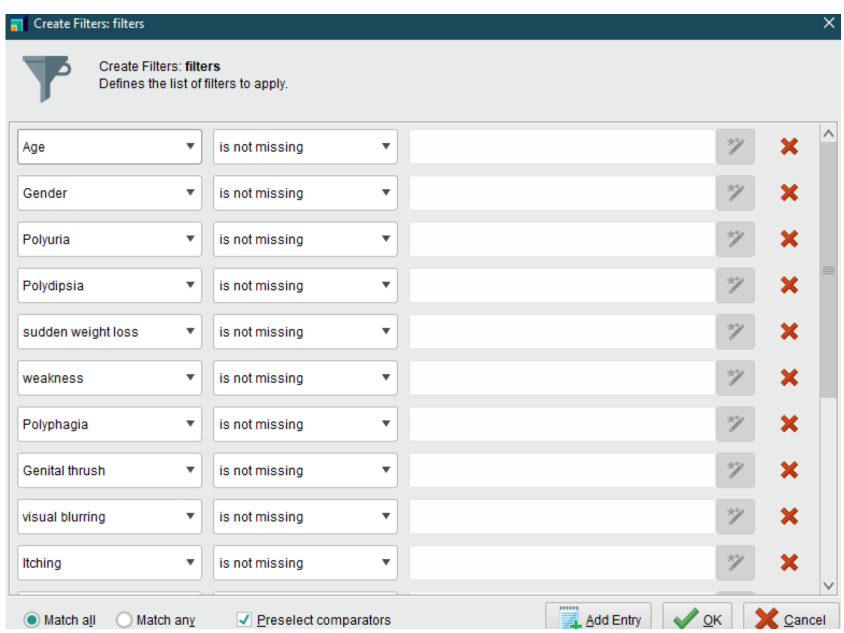


Figure 4. Examples Filter Settings

Figure 4. is the desired data filter setting. This data filter process aims to delete data records that have missing values. The following is an overview of the data-cleaning process:

Open in Turbo Prep Auto Model Filter (1,146 / 1,146 examples): all

Row No.	class	Age	Gender	Polyuria	Polydipsia	sudden wei...	weakness	Polyphagia
1	Positive	40	Male	No	Yes	No	Yes	No
2	Positive	58	Male	No	No	No	Yes	No
3	Positive	41	Male	Yes	No	No	Yes	Yes
4	Positive	45	Male	No	No	Yes	Yes	Yes
5	Positive	60	Male	Yes	Yes	Yes	Yes	Yes
6	Positive	55	Male	Yes	Yes	No	Yes	Yes
7	Positive	57	Male	Yes	Yes	No	Yes	Yes
8	Positive	66	Male	Yes	Yes	Yes	Yes	No
9	Positive	67	Male	Yes	Yes	No	Yes	Yes
10	Positive	70	Male	No	Yes	Yes	Yes	Yes
11	Positive	44	Male	Yes	Yes	No	Yes	No
12	Positive	38	Male	Yes	Yes	No	No	Yes
13	Positive	35	Male	Yes	No	No	No	Yes
14	Positive	61	Male	Yes	Yes	Yes	Yes	Yes
15	Positive	60	Male	Yes	Yes	No	Yes	Yes

ExampleSet (1,146 examples, 1 special attribute, 16 regular attributes)

Figure 5. Dataset View Before Preprocessing

Figure 5 displays the number of datasets before cleaning, namely 1146 data records, then after cleaning processing as in Figure 3.5, the remaining data results are 990, here's what it looks like:

Open in Turbo Prep Auto Model Filter (990 / 990 examples): all

Row No.	class	Age	Gender	Polyuria	Polydip...	sudden ...	weakne...	Polypha...	Genital t...	visual bl...	Itching
1	Positive	40	Male	No	Yes	No	Yes	No	No	No	Yes
2	Positive	58	Male	No	No	No	Yes	No	No	Yes	No
3	Positive	41	Male	Yes	No	No	Yes	Yes	No	No	Yes
4	Positive	45	Male	No	No	Yes	Yes	Yes	Yes	No	Yes
5	Positive	60	Male	Yes	Yes	Yes	Yes	Yes	No	Yes	Yes
6	Positive	55	Male	Yes	Yes	No	Yes	Yes	No	Yes	Yes
7	Positive	57	Male	Yes	Yes	No	Yes	Yes	Yes	No	No
8	Positive	66	Male	Yes	Yes	Yes	Yes	No	No	Yes	Yes
9	Positive	67	Male	Yes	Yes	No	Yes	Yes	Yes	No	Yes
10	Positive	70	Male	No	Yes	Yes	Yes	Yes	No	Yes	Yes
11	Positive	44	Male	Yes	Yes	No	Yes	No	Yes	No	No
12	Positive	38	Male	Yes	Yes	No	No	Yes	Yes	No	Yes
13	Positive	35	Male	Yes	No	No	No	Yes	Yes	No	No
14	Positive	61	Male	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
15	Positive	60	Male	Yes	Yes	No	Yes	Yes	No	Yes	Yes

Figure 6. Dataset appearance after preprocessing

Then the next process is dividing the data into training data and test data. Researchers divided the data on a scale of 60: 40, using the Split Data feature.

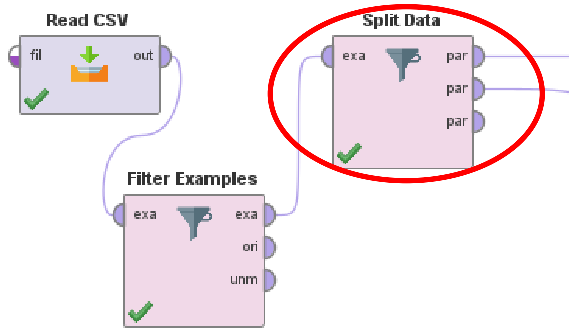


Figure 7. Division of Training Data and Test Data

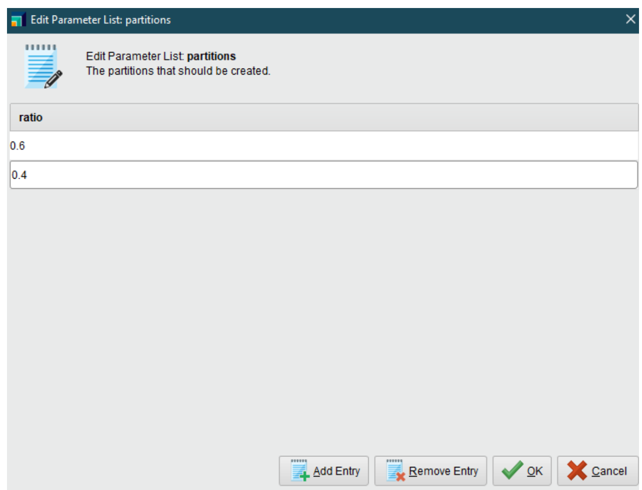
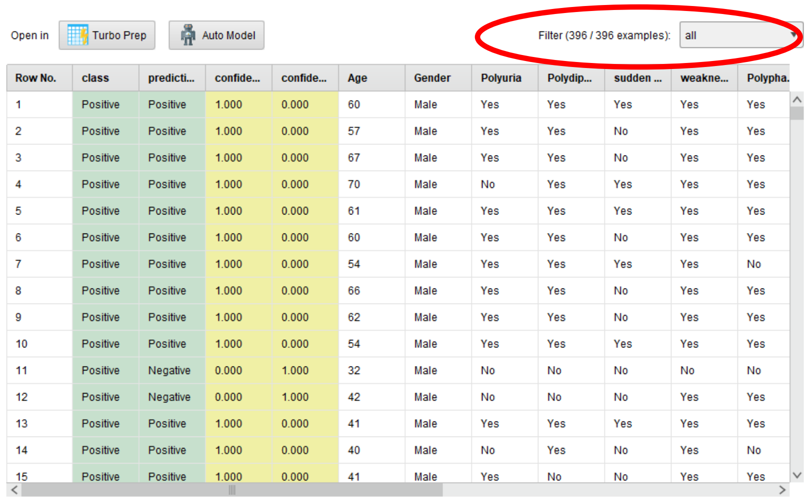


Figure 8. Data Ratio Division Settings

Figure 8 is the process of setting the ratio of training data and test data. The training data is 60% of the dataset that has been cleaned, while the test data is 40% of the dataset that has been cleaned. Therefore, the number of test data is 396 data, and the rest is training data.



Open in Turbo Prep Auto Model Filter (396 / 396 examples): all

Row No.	class	predicti...	confide...	confide...	Age	Gender	Polyuria	Polydip...	sudden ...	weakne...	Polypha.
1	Positive	Positive	1.000	0.000	60	Male	Yes	Yes	Yes	Yes	Yes
2	Positive	Positive	1.000	0.000	57	Male	Yes	Yes	No	Yes	Yes
3	Positive	Positive	1.000	0.000	67	Male	Yes	Yes	No	Yes	Yes
4	Positive	Positive	1.000	0.000	70	Male	No	Yes	Yes	Yes	Yes
5	Positive	Positive	1.000	0.000	61	Male	Yes	Yes	Yes	Yes	Yes
6	Positive	Positive	1.000	0.000	60	Male	Yes	Yes	No	Yes	Yes
7	Positive	Positive	1.000	0.000	54	Male	Yes	Yes	Yes	Yes	No
8	Positive	Positive	1.000	0.000	66	Male	Yes	Yes	No	Yes	Yes
9	Positive	Positive	1.000	0.000	62	Male	Yes	Yes	No	Yes	Yes
10	Positive	Positive	1.000	0.000	54	Male	Yes	Yes	Yes	Yes	Yes
11	Positive	Negative	0.000	1.000	32	Male	No	No	No	No	No
12	Positive	Negative	0.000	1.000	42	Male	No	No	No	Yes	Yes
13	Positive	Positive	1.000	0.000	41	Male	Yes	Yes	Yes	Yes	Yes
14	Positive	Positive	1.000	0.000	40	Male	No	Yes	No	Yes	No
15	Positive	Positive	1.000	0.000	41	Male	Yes	No	No	Yes	Yes

ExampleSet (396 examples, 4 special attributes, 16 regular attributes)

Figure 9. Test Data Display

After implementing data cleaning and dividing the ratio of training data and test data, the next process is to use the required algorithm(Iskandar, 2023). In this research, we will compare the Decision Tree algorithm with the Decision Tree algorithm based on the Adaptive Boosting method(Etika et al., 2023), which later this process will produce different performance results on the data set used. The following displays the process of using the algorithm, figure 11:

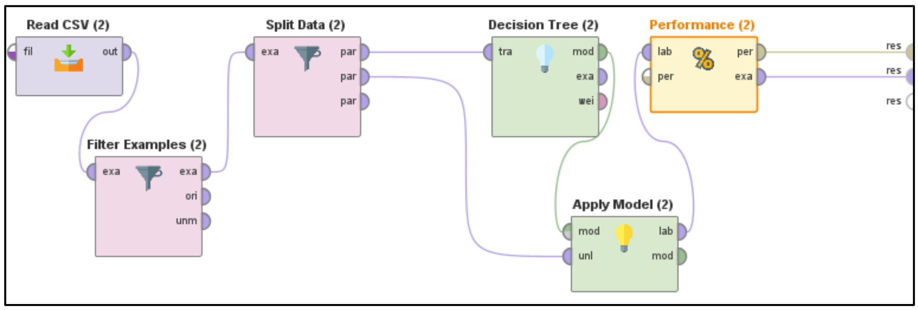


Figure 10. Decision Tree Algorithm Testing Process

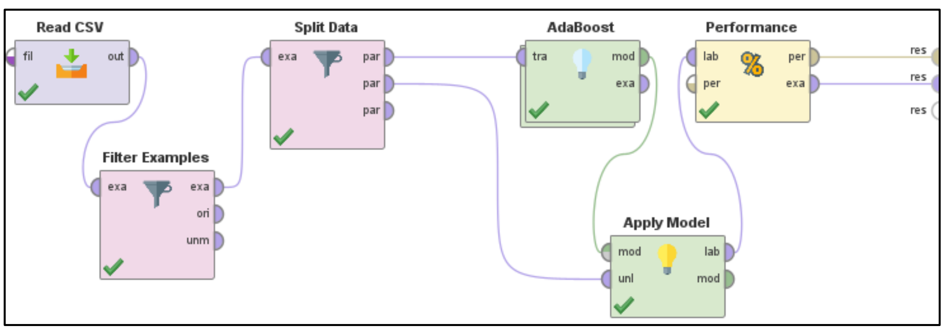


Figure 11. Decision Tree + AdaBoost Algorithm Testing Process

OnAdaBoost feature added Decision Tree Algorithm in the process, here's what it looks like figure 12:

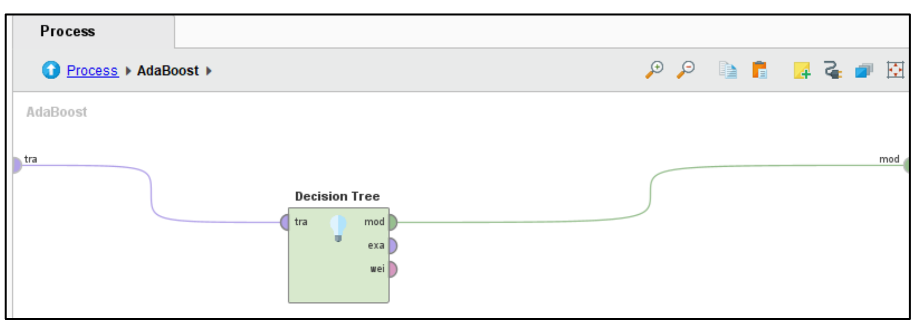


Figure 12. Decision Tree Algorithm in the AdaBoost Method

4. Research Supplies
   a) Software
      1) Windows 10 Home 64-bit Operating System.
      2) RapidMiner Studio Version 10.1.
      3) Microsoft Office Word & Excel 2013.
   b) Hardware
      1) HP Laptop, AMD A9-9425, RADEON R5
      2) 4GB RAM.
      3) 1000GB hard disk

Results and Discussion

Results

The results of the research on the Application of the Adaptive Boosting Method in Increasing the Accuracy of Classification of Type Two Diabetes Mellitus Patients Using the Decision Tree Algorithm are in the form of the following test results:

1. Test 1 (Decision Tree)
The first test was carried out using the Decision Tree algorithm on the dataset. The following are the results of the accuracy value obtained from this test, which is 95.71%, for more details, see Figure 4.1.1 as follows:

accuracy: 95.71%

	true Positive	true Negative	class precision
pred. Positive	221	5	97.79%
pred. Negative	12	158	92.94%
class recall	94.85%	96.93%	

Figure 13. Test Accuracy Results 1

The following are 15 representative classification results that can be displayed from the total test data in the test

Row No.	class	prediction(cl...	confidence(...	confidence(...	Age	Gender	Polyuria	Polydipsia
1	Positive	Negative	0.077	0.923	40	Male	No	Yes
2	Positive	Negative	0.018	0.982	45	Male	No	No
3	Positive	Positive	1	0	70	Male	No	Yes
4	Positive	Positive	1	0	35	Male	Yes	No
5	Positive	Positive	1	0	66	Male	Yes	Yes
6	Positive	Positive	1	0	43	Male	Yes	Yes
7	Positive	Positive	1	0	58	Male	Yes	Yes
8	Positive	Positive	1	0	30	Female	Yes	No
9	Positive	Positive	1	0	35	Female	Yes	Yes
10	Positive	Positive	1	0	58	Male	No	No
11	Positive	Positive	1	0	55	Male	Yes	Yes
12	Positive	Negative	0.083	0.917	60	Female	Yes	No
13	Positive	Positive	1	0	45	Female	No	No
14	Positive	Negative	0.077	0.923	40	Male	No	Yes
15	Positive	Negative	0.018	0.982	45	Male	No	No

Figure 14. Test Classification Results 1

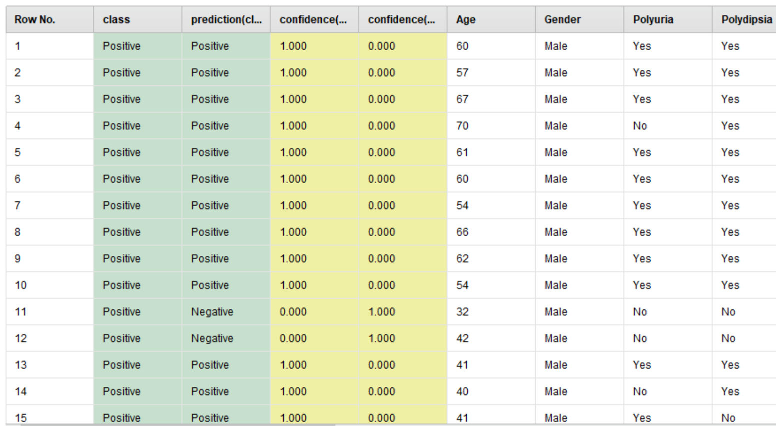
2. Test 2 (Decision Tree + AdaBoost)
The second test was carried out using the AdaBoost-based Decision Tree algorithm on the dataset. Following are the results of the accuracy value obtained from this test, which is 98.99%, for more details, see the following image:

accuracy: 98.99%

	true Positive	true Negative	class precision
pred. Positive	229	0	100.00%
pred. Negative	4	163	97.60%
class recall	98.28%	100.00%	

Figure 15. Accuracy Results of Test 2

The following are 15 representative classification results that can be displayed from the classification results of test 2.



Row No.	class	prediction(cl...	confidence(...	confidence(...	Age	Gender	Polyuria	Polydipsia
1	Positive	Positive	1.000	0.000	60	Male	Yes	Yes
2	Positive	Positive	1.000	0.000	57	Male	Yes	Yes
3	Positive	Positive	1.000	0.000	67	Male	Yes	Yes
4	Positive	Positive	1.000	0.000	70	Male	No	Yes
5	Positive	Positive	1.000	0.000	61	Male	Yes	Yes
6	Positive	Positive	1.000	0.000	60	Male	Yes	Yes
7	Positive	Positive	1.000	0.000	54	Male	Yes	Yes
8	Positive	Positive	1.000	0.000	66	Male	Yes	Yes
9	Positive	Positive	1.000	0.000	62	Male	Yes	Yes
10	Positive	Positive	1.000	0.000	54	Male	Yes	Yes
11	Positive	Negative	0.000	1.000	32	Male	No	No
12	Positive	Negative	0.000	1.000	42	Male	No	No
13	Positive	Positive	1.000	0.000	41	Male	Yes	Yes
14	Positive	Positive	1.000	0.000	40	Male	No	Yes
15	Positive	Positive	1.000	0.000	41	Male	Yes	No

Figure 16. Classification Results of Test 2

Discussion

The research that has been carried out uses two testing models, namely the first test uses a Decision Tree, then the second test uses a Decision Tree based on Adaptive Boosting (AdaBoost). Based on the accuracy results, it is stated as follows in Table 2:

Table 2. Accuracy Comparison

Datasets	Accuracy Test Results	
	Decision Trees	Decision Tree + AdaBoost
Early Stage Diabetes Risk Prediction	95.71 %	98.99 %
Difference	3.28 %	

The calculation above uses a division ratio of 60:40. The number of test data is 396 examples. With the results of manual calculations of accuracy test data for each method using the confusion matrix as follows, table 3:

1. Decision Trees

Table 3. Decision Tree Accuracy Testing Table

Confusion Matrix		Predicted Value	
		Positive	Negative
Actual Value	Positive	TP = 221	FP = 5
	Negative	FN = 12	TN = 158

Test the accuracy value using manual confusion matrix calculations in the Decision Tree algorithm:

$$Accuracy = \frac{221 + 158}{(221+5+12+158)}$$

$$= \frac{221+158}{396}$$

$$= \frac{379}{396} = 0{,}957$$

2. Decision Tree + Adaptive Boosting (AdaBoost)

Table 4. Decision Tree Accuracy Testing Table

Confusion Matrix		Predicted Value	
		Positive	Negative
Actual Value	Positive	TP = 229	FP = 0
	Negative	FN = 4	TN = 163

Test the accuracy value using manual confusion matrix calculations on the AdaBoost-based decision tree algorithm:

$$Accuracy = \frac{229+163}{(229+0+4+163)}$$

$$= \frac{229+168}{396}$$

$$= \frac{392}{396} = 0{,}989$$

A comparison of dataset classification using the Decision Tree algorithm and the AdaBoost-based Decision Tree Algorithm has indeed proven that adding Adaptive Boosting to the Decision Tree algorithm can increase the accuracy value. So, the aim of this analysis is so that researchers can use machine learning effectively in processing data. This has previously been researched by Lidia Pebrianti, et al. With the research title "Implementation of the Adaboost Method to Optimize Diabetes Classification with the Naïve Bayes Algorithm" the Naïve Bayes algorithm has been used to optimize the accuracy of data classification by applying the AdaBoost method. This research concluded that the Naïve Bayes Algorithm produced an accuracy of 0.7608. Meanwhile, the results of Naïve Bayes which were boosted using the AdaBoost algorithm were 0.7694. Apart from that, research conducted by Zico Karya Saputra Domas and Roby Rakhmadi with the title "Improving Decision Tree Performance with AdaBoost for Classifying Lack of Transparency in Anti-Corruption Information" concluded that the application of the AdaBoost method in classifying class predictions is feasible to increase the level of better performance. Experimental results show that the accuracy of the Decision Tree is 69.5%, while the average results of the Decision Tree using the AdaBoost method reach an accuracy of 71.16%. This is because the AdaBoost method adds weight to error classifications and reduces weight to correct classifications, thereby effectively changing the distribution of the training data.

Conclusions and Suggestions

Conclusions

Based on the research results above, the researcher concluded that:

1. several other factors have a big influence on increasing the accuracy of the Decision Tree algorithm without having to use adaboost, namely at the preprocessing stage or during the data cleaning process, where at the preprocessing stage accuracy is required in preparing the data, so that the data is completely clean of noise. , unbalanced data or outliers in the data, so that the classification process can run well and the results are more accurate.
2. Based on research that has been carried out, the level of effectiveness resulting from the Decision Tree method based on Adaptive Boosting (AdaBoost) on the Early Stage Diabetes Risk Prediction dataset, can increase accuracy by 3.28%, with a success rate of AdaBoost on the dataset reaching 98.99%, so that the Boosting method works quite effectively in increasing the accuracy of the Decision Tree algorithm

Suggestions

From the results of the research that has been carried out, researchers have several suggestions that can be used for further research as reference material, namely as follows:

1. Try to use other boosting, such as Bayesian Boosting for the Naïve Bayes algorithm or other algorithms.
2. Try to add 2-5 (two to 5) or more datasets when testing the model using the AdaBoost-based Decision Tree algorithm so that the test results can be known in detail.
3. View test results other than accuracy values, such as precision, recall, or AUC graphs.
4. Try using other tools such as Orange, Weka Studio, or Matlab to carry out the data mining process.

References

Argina, AM (2020). Application of the K-Nearest Neighbor Classification Method on a Dataset of Diabetes Patients. Indonesian Journal of Data and Science, 1(2), 29–33.

Aziz, MI, Fanani, AZ, & Affandy, A. (2023). Analysis of Ensemble Methods in Decision Tree Based Heart Disease Classification. Budidarma Media Informatics Journal, 7(1), 1–12.

Bisri, A., & Wahono, RS (2015). Application of Adaboost to resolve class imbalances in determining student graduation using the Decision Tree method. In Computer Science (Vol. 1, Issue 1). Computer Science. com.

Byna, A., & Basit, M. (2020). Application of the Adaboost Method to Optimize Stroke Prediction Using the Na{\"\i}ve Bayes Algorithm. Sisfokom Journal (Information and Computer Systems), 9(3), 407–411.

Horizon. (2021, September 9). Cost of Illness in Type 2 Diabetes Mellitus Patients - Unair News. Newssunair. https://news.unair.ac.id/2021/09/09/cost-of-illness-pada-patient-diabetes-mellitus-tipe-2/?lang=id

Domas, ZKS, & Rakhmadi, R. (2022). Improving Decision Tree Performance with AdaBoost for Classifying Lack of Transparency of Anti-Corruption Information. AppliedInformationSystemsandManagement (AISM), 5(2), 75–82.

Etika, K., Pandu, A., Syafar, F., Akbar, I., Arman, P., & Robbi, R. (2023). Application of forward selection strategy using C4. 5 algorithms to improve the accuracy of classification's data set. Application of Forward Selection Strategy Using C4. 5 Algorithms to Improve the Accuracy of Classification's Data Sets, 30(1), e14--e23.

Fajri, M. (2015). Splitting Rule and Application of Bagging in Classification Trees. Faculty of Mathematics and Natural Sciences, Sepuluh Nopember Institute of Technology.

Fatimah, RN (2015). DIABETES MELLITUS TYPE 2. Majority Journal, 4(5). https://juke.kedokteran.unila.ac.id/index.php/majority/article/view/615

Hana, FM (2020). Classification of Diabetes Patients Using the C4 Decision Tree Algorithm. 5. SISKOM-KB Journal (Computer Systems and Artificial Intelligence), 4(1), 32–39.

Iftikar, MA, & Sibaroni, Y. (2022). Twitter Sentiment Analysis: Handling the Covid-19 Pandemic Using the Hybrid Method of Na{\"\i}ve Bayes, Decision Tree, and Support Vector Machine. EProceedings of Engineering, 9(3).

Iskandar, A. (2023). Introduction to Data Analysis With R Studio. Indonesian Digital Innovation Scholars Foundation.

Junita, V., & Bachtiar, F.A. (2019). Human Activity Classification using the C4 Decision Tree Algorithm. 5 and Information Gain for Feature Selection. Journal of Information Technology and Computer Science Development, 3(10), 9426–9433.

Kong, M., Zhang, H., Cao, X., Mao, X., & Lu, Z. (2020). Higher levels of neutrophil-to-lymphocyte are associated with severe COVID-19. Epidemiology \& Infection, 148.

Kusuma, IGNA, Pradipta, IM, Santosa, IMA, Dharmendra, IK, & others. (2023). Handling Data Imbalances in the Classification of Public Complaints. Journal of Information and Computer Technology, 9(5).

Lorena Br Ginting, S., Zarman, W., & Darmawan, A. (2015). Data Mining Technique for Predicting Student Study Period Using the K-Nearest Neighborhood Algorithm. KOMPUTIKA-UNIKOM Computer Systems Journal, 3(2).

Muhlisin, A. (2019, February 22). Hyperglycemia - Causes, Symptoms, & Treatment | HonestDocs. Honestdocs. https://www.honestdocs.id/hiperglycemia-gula-darah-tinggi

Nusrhendratno, SS (2022). Synthesis of Density Based Feature Selection (DBFS) and AdaBoots Features with XGBoost to Improve Prediction Model Performance. Proceedings of the National Science and Technology Seminar, 12(1), 305–313.

Pebrianti, L., Aulia, F., Nisa, H., & Saputra, K. (2022). Implementation of the Adaboost Method to Optimize Diabetes Classification with the Naive Bayes Algorithm. JUSTINDO (Indonesian Journal of Information Systems and Technology), 7(2), 122–127.

Permana, AA, Wahyuddin, S., Santoso, LW, Wibowo, GWN, & Wardhani, AK (nd). Ahmad Jurnaidi Wahidin Gusti Eka Yuliastuti Elisawati Rima Rizqi Wijayanti Abdurrasyid.

Pirmansyah, F., & Wahyudi, T. (2023). Implementation of Data Mining Using the C4 Algorithm. 5 To Predict Evaluation of Security Unit Members Case Study Pt. Yimm Pulogadung. Indonesian Journal: Information and Communication Management, 4(3), 1566–1580.

Prabowo, NA, Ardyanto, TD, Hanafi, M., Kuncorowati, NDA, Dyanneza, F., Apriningsih, H., & Indriani, AT (2021). Increasing Diabetes Diet Knowledge, Diabetes Self Management and Reducing Stress Levels During Dieting in Type 2 Diabetes Mellitus Patients at Sebelas Maret University Hospital. LPM Newsletter, 24(2), 285–296.

Pradana, E. (2018). Analysis of the Application of Adaptive Boosting (Adaboost) in Improving the Performance of the C4 Algorithm. 5. Pelita Bangsa College of Technology.

Sibarani, RP (2023, January 3). Recognizing the Symptoms of Type 2 Diabetes Mellitus that You Need to Know - EMC Healthcare - SAME. Emc. https://www.emc.id/id/care-plus/mengenal-symptomatic-diabetes-melitus-tipe-2-yang-perlu-diketahui

Sudargo, T., Freitag, H., Kusmayanti, NA, & Rosiyani, F. (2018). Diet and obesity. UGM press.

Suprayitna, M., Hajri, Z., Fatmawati, BR, Prihatin, K., & Nadrati, B. (2023). Early Detection of Diabetes Mellitus (DM) Through "DM Awareness." BERNAS: Journal of Community Service, 4(3), 2291–2296.

Suryati, NI, & Kep, M. (2021). Effective Nursing Exercise Book for Diabetes Mellitus Patients Based on Research Results. Deepublish.

Tangkelayuk, A. (2022). The Water Quality Classification Using KNN, Na{\"\i}ve Bayes, and Decision Tree Methods. JATISI (Journal of Informatics Engineering and Information Systems), 9(2), 1109–1119.

Ula, M., Ulva, AF, Mauliza, M., Sahputra, I., & Ridwan, R. (2021). Implementation of Machine Learning in Determining Nutritional Status using the Complete Linkage Agglomerative Hierarchical Clustering Method. Mantic Journal, 5(3), 1910–1914.

Urva, G., Albanna, I., Sungkar, MS, Gunawan, IMAO, Adhicandra, I., Ramadhan, S., Rahardian, RL, Handayanto, RT, Ariana, AAGB, Atika, PD, & others. (2023). APPLICATION OF DATA MINING IN VARIOUS FIELDS: Concepts, Methods, and Case Studies. PT. Sonpedia Publishing Indonesia.