Predictive Maintenance Air Conditioner Using Machine Learning

Ranggi Tino Fambudi[1] *, Sani Muhamad Isa*[2]
[1]*Computer Science Department, Binus Graduate Program–Master of Computer Science, Bina Nusantara University, Jakarta, Indonesia*
[2]*Computer Science Department, Lecturer, Bina Nusantara University, Jakarta, Indonesia*
E-mail: [1]*ranggi.fambudi@binus.ac.id,* [2]*sisa@binus.edu*

Abstract

Predictive maintenance will take care of the machine's needs in terms of power loss from damage that lowers performance, operational costs from severe damage, business interruptions from damage that renders the machine unusable, and much more. Almost every home has an air conditioner, the machine that requires constant maintenance of temperature and humidity, especially in offices with servers or control rooms. Preventive and predictive maintenance is necessary to identify the necessary steps for technicians to take when handling an AC before the damage worsens. In this research we implemented and proposed an Air Conditioner detection system using machine learning with three methods, namely K-Nearest Neighbor, Decision Tree, and Random Forest. In order to understand the actual conditions of each AC, we use data sheets that we gathered through surveys with engineering teams at multiple hotels as well as technical teams that handle servers and control rooms. There are 20 features in the gathered data set; however, since only 14 of the features affect the value, extraneous data will be removed. Then the data was divided into two groups, namely 23 AC Failures yes, which means the AC condition is not normal and 110 AC Failures No, which means the AC condition is not damaged. Using the stratified random sample method, 25% of the data will be oversampled. In this study, Kbest and backward elimination were employed for feature selection. The SMOTE approach was then applied for oversampling due to the unbalanced groups. With accuracy values of 91.18%, precision 91.18%, recall 90.90%, and f1-score 90.92%, the Random Forest model with the suggested model outperformed the Decision Tree and KNN models, according to the experimental findings.

Keywords: *We would like to encourage you to list your keywords in this section*

1. Introduction

Research related to predictive maintenance of air conditioners still belongs to few and mostly uses data sets themselves, or in the sense of unshared data sets. As shown in Table 2.1, the Unified Modeling Language with the Agile XP modeling method can produce a similarity accuracy of 72% [1]. Then further studies with the simulation results show that MLP has an accuracy of up to 99.4% and SVM has an accuracy of 97% [2]. In other studies that have different research objects, the random forest method has relatively the same results as the SVM method; the study mentioned an estimated precision of 96.72% [3]. This makes it possible that the results from Random Forest can be used as a reference for researchers' research, i.e., on predictive maintenance of air conditioners.

Later in the study [5], found that KNN with feature selection backward elimination showed an optimum improvement in RMSE and AE. This made the researchers choose KNN and feature selection backward elimination in predictive maintenance air conditioners. Then research related to the decision tree also has results that are categorized well to reach an accuracy of 96.05%, which makes the researchers think of adding the decision tree algorithm to the comparison. With the data that researchers take independently, researchers feel they need to add data to better predict results, but researchers do not want to corrupt the percentage of the existing data, so researchers use

ISSN: 2720-992X

Copyright © KESATRIA: Jurnal Penerapan Sistem Informasi (Komputer & Manajemen)

stratified random sampling propotional so that the data obtained has the same percentages as the original data.

Because this research has different datasets than other research data sets. So the results of this study evaluation are comparisons between KNN, Decision Tree, Random Forest, and SVM methods in predicting air conditioners. The evaluation will focus on performance metrics such as accuracy, precision, recall, and f1-score. By comparing the results of the method, this study will identify whether the proposed approach is able to provide more accurate predictions than the SVM method [2]. Of course, the results obtained by the SVM method are not the same as in previous studies because they have different datasets. For that, the researchers also added SVM methods to compare them.

2. Research Methodology

Predictive maintenance is a topic that has attracted the attention of researchers. With predictive maintenance, it will really help the team determine what actions must be taken. What is needed in predictive maintenance is high accuracy so that the results received are the correct results according to the actual results. Researchers intend to use machine learning with the Decision Tree, K-Nearest Neighbor, and Random Forest methods for the predictive maintenance process. It is hoped that with this method, one of them will get better results in terms of accuracy, recall, precision, and F1 score. For better predictions, researchers conducted experiments using the K-Best selection and Backward Regression selection features.

In this research, before modeling was carried out, the researcher carried out stratified random sampling (proportionate) to increase the number of sample data without changing the percentage of each class so as not to change the results achieved. Then feature selection is carried out using the K-Best and Backward regression models to find the best K value. After the K value is found, the researcher will conduct the research twice, the first using SMOTE and the second without using SMOTE. With this method in the research "Predictive Maintenance of Air Conditioners Using Machine Learning" it is hoped that we can create a better predictive maintenance system for air conditioners. By using these three methods, the best combination of feature selection and machine learning models will be obtained.

2.1. Literatures Review

Research related to predictive maintenance of air conditioners is still relatively small, and most of it uses its own datasets, or in the sense of datasets that are not shared. As seen in Table 2.1. that the Unified Modeling Language with the Agile XP system development method can produce a similarity accuracy of 72% [1], then subsequent research with simulation results shows that MLP has an accuracy of up to 99.4% and SVM up to 97% [2]. In other research that has a different research object, the Random Forest method has relatively the same results as the SVM method, in the research it is stated that the accuracy is 96.72% [3]. This makes it possible that the results from Random Forest can be used as a reference for research by researchers, namely on predictive maintenance of air Conditioners.

Then, in research [5], it was found that KNN with backward elimination feature selection showed quite optimal increases in RMSE and AE. This is what made researchers choose KNN and backward elimination feature selection in predictive air conditioner maintenance. Then research related to decision trees also had results that were categorized as good, reaching an accuracy of 96.05%. This made researchers think about adding a decision tree algorithm for comparison. With the data that the researcher took independently, the researcher felt that he had to add data to make the prediction results better, but the researcher did not want to damage the percentage of existing data, so the researcher used proportional stratified random sampling so that the data obtained had the same percentage as the original data.

Because this research has a different dataset from other research datasets. So the results of this research evaluation are a comparison between the KNN, Decision Tree, Random Forest, and SVM methods for predicting air conditioners. This evaluation will focus on performance metrics such as accuracy, precision, recall, and f1-score. By comparing the results of these methods, this research will identify whether the proposed approach is able to provide more accurate prediction results than the SVM method [2]. Of course, the results obtained by the SVM method are not the same as in previous research because it has a different dataset. For this reason, researchers also added the SVM method to compare them.

Before starting the research, the researcher did the most important thing, namely identifying the problem based on literature studies and also case studies in the field. By focusing on predictive maintenance on Air Conditioners using machine learning with three methods, namely K-Nearest Neighbor, Decision Tree and Random Forest, this research will be able to contribute to the planogram evaluation process which can be used by engineering teams to determine in-depth maintenance.

This research was conducted to be able to provide a model that is able to predict abnormal air conditioner conditions with good accuracy. The approach taken is through research [2]. The literature study that the researchers have carried out has made them decide to use the machine learning classification models Random Forest, Decision Tree, and KNN with the K-Best and Backward Elimination feature selection techniques. The raw data will be subjected to stratified oversampling to get larger data but with the same percentage. After that, it continues with model classification and will continue with the evaluation process with the confusion matrix: accuracy, precision, recall, and F1-score.

To achieve the research objectives, this research is divided into several research stages, namely literature study, problem identification, determining research topics, dataset collection, data pre-processing, feature selection, classification, evaluation, and report preparation.

2.2. Questionnaire

Questionnaire To obtain data related to research, researchers conducted interviews and questionnaires with a team of engineers who work in hotels and also selected companies engineers such as Swiss Bel Resort Belitung, Harper Malioboro, Semarang Container Terminal, and so on. Then the questions will become attributes that will later become data in this research.

Apart from questionnaires, researchers carry out literature studies by collecting, reading, and understanding theoretical references originating from theoretical books, electronic books (ebooks), research journals, and other authentic library sources related to research.

2.3. Dataset

The dataset consists of 20 features and is divided into two classes. Class 0 is a normal real air conditioner condition, while class 1 is an abnormal air conditioner condition. The categories of each air conditioner condition with the original dataset features are shown in Table 1 below

Table 1. Attribute Table used

Attribut Name	Description
Timestamp	The timestamp of the data was captured
Email Address	Email from the filler
Nama Pengisi	Name of the filler
Jabatan / Posisi Pekerjaan	Job title of the filler
Instansi / Unit kerja saat ini	Job location of the filler
Suhu dalam ruangan?	Indoor temperature
Suhu remote atau Panel Control AC?	Remote Setting Temperature

Attribut Name	**Description**
Kelembaban ruangan?	Percentase of Humidity indoor
Jenis AC?	AC type
Berapa PK?	PK Of the AC
Rentang berapa minggu pengecekan komponen indoor?	How long does it take for indoor AC to be maintained?
Rentang berapa minggu dalam pengecekan komponen outdoor?	How long does it take for outdoor AC to be maintained?
Berapa Voltase pada outdoor ac?	How many Voltage for Outdoor AC?
Berapa Ampere pada outdoor ac?	How many Ampere for Outdoor AC?
Outdoor Mati	Is Outdoor AC died?
Pipa Kering	Is the pipe of outdoor dry
Kipas Indoor Bising	Is the indoor fan noisy?
Tetesan Air	Are there any water droplets?
Indikasi	what are the indications?
AC Failure	Is AC Failure?

3. Results and Discussion

In this research, the dataset used is real data taken from several companies and hotels. The dataset that has been collected is identified first, and then the data quality is checked. The data will be cleaned to remove noise, inconsistent data, detect and delete outlier data.

Rename features are used to make features shorter and easier to understand from data that was originally in the form of question data. The answers or contents of features containing the words yes and no will be transformed into binary so that they can be processed further. Next, changes are made to all features that have numeric data and make them decimal.

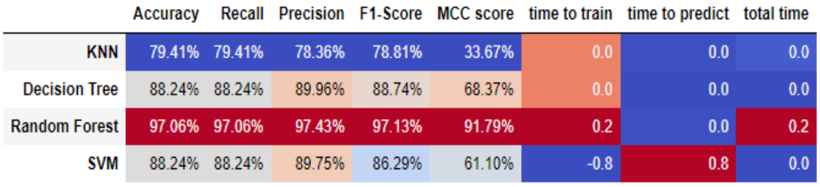
After all the numeric data has been converted to a float, we delete data that does not affect the results, for example, name, email, and so on. The next step is to take a percentage of the original data of 20% to carry out stratified random sampling (proportionate).

The next step is to group the data to find out whether the data is imbalanced or balanced, which will then be tested by making it balanced and without making it balanced. Then pipe to pass the resulting DataFrame from the previous step into the next operation in the pipe series. The pipe function is used to apply transformations or sequential operations to an object. Pipes make it possible to apply a series of operations to the object more easily and in a structured manner.

3.1. Model Classification

In this research, the classification process begins by preparing or dividing the data into training data and test data. The proportion of dataset division is 80% training data and 20% test data. Then we applied the KNN, decision tree, and random forest methods with K-best and backward elimination feature selection, as well as the imbalanced oversampling Smote technique.

The first experiment for classification used the Random Forest, K-Nearest Neighbors, Decision Tree, and Support Vector Machine algorithm models to compare the results without feature selection. Performance evaluation is carried out by comparing the accuracy, precision, recall, and f1-score results of the three algorithms. The results of the evaluation can be seen in Figure 1 below.

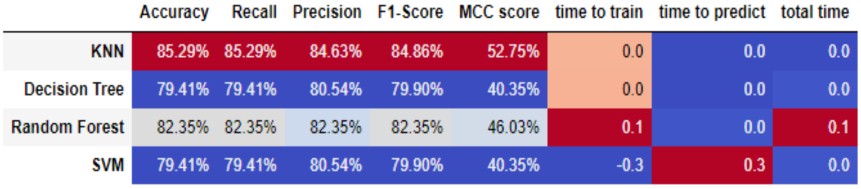


	Accuracy	Recall	Precision	F1-Score	MCC score	time to train	time to predict	total time
KNN	79.41%	79.41%	78.36%	78.81%	33.67%	0.0	0.0	0.0
Decision Tree	88.24%	88.24%	89.96%	88.74%	68.37%	0.0	0.0	0.0
Random Forest	97.06%	97.06%	97.43%	97.13%	91.79%	0.2	0.0	0.2
SVM	88.24%	88.24%	89.75%	86.29%	61.10%	-0.8	0.8	0.0

Figure 1. Without Elimination and Smote

In Figure 1, you can see that the best result is Random Forest with Accuracy 97.06%, Precision 97.43%, Recall 97.06%, F1-Score 97.13%, and time to train 0.2. beats the results obtained from KNN Accuracy 79.41%, Precision 78.36%, Recall 79.41%, F1-Score 78.81%, and time to train 0.0, and also Decision Tree Accuracy 88.24%, Precision 89.96%, Recall 88.24%, F1-Score 88.74%, and time to train 0.0. Even SVM Accuracy 88.24%, Precision 89.75%, Recall 88.24%, F1-Score 86.29%, and time to train -0.8, which was previously carried out in research, is still below that of Random Forest.

Then the second trial continued using the KNN, Decision Tree, Random Forest, and Support Vector Machine algorithms, but we used the imbalanced oversample SMOTE method to compare the results. The results of the evaluation can be seen in Figure 2 below.



	Accuracy	Recall	Precision	F1-Score	MCC score	time to train	time to predict	total time
KNN	85.29%	85.29%	84.63%	84.86%	52.75%	0.0	0.0	0.0
Decision Tree	79.41%	79.41%	80.54%	79.90%	40.35%	0.0	0.0	0.0
Random Forest	82.35%	82.35%	82.35%	82.35%	46.03%	0.1	0.0	0.1
SVM	79.41%	79.41%	80.54%	79.90%	40.35%	-0.3	0.3	0.0

Figure 2. Without Elimination but using Smote

Based on the results in Figure 2, it can be seen that Oversample Smote caused the results of the three proposed methods to decrease, both in terms of accuracy, precision, recall, and F1-score, but only one experienced an increase in all sessions, namely KNN, with the results being The highest are accuracy 85.29%, precision 84.63%, recall 85.29%, and F1-score 84.86%.

Then we tried to use the K-Best and Backward Elimination features to get better results, both using SMOTE oversampling and not using oversampling. Figure 4 and 6 will explain the evaluation results of using K-Best and backward elimination without using Smote oversampling technique.

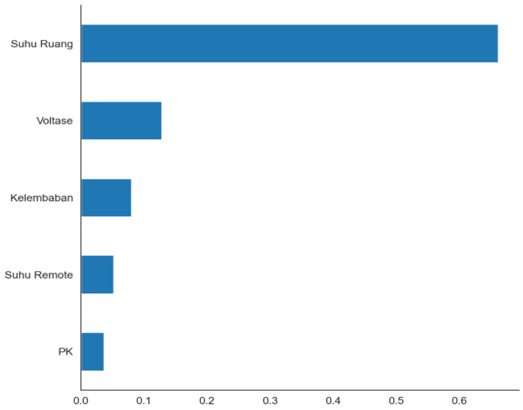
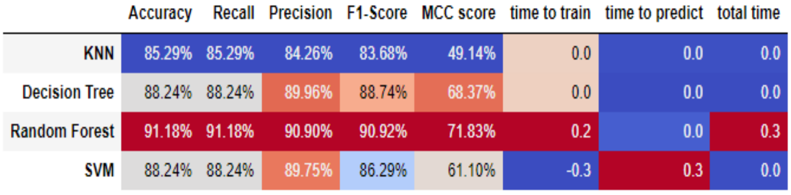


Figure 3. Feature Elimination K-Best without Smote

In the experimental results with the K-Best Elimination technique, we carried out feature selection by identifying the five most important features to improve classification performance. After carrying out the elimination process, the five most important features are obtained, namely: room temperature, voltage, humidity, remote temperature, and PK. After that, we carried out the process method on the feature results, and the results can be seen in Figure 4.



	Accuracy	Recall	Precision	F1-Score	MCC score	time to train	time to predict	total time
KNN	85.29%	85.29%	84.26%	83.68%	49.14%	0.0	0.0	0.0
Decision Tree	88.24%	88.24%	89.96%	88.74%	68.37%	0.0	0.0	0.0
Random Forest	91.18%	91.18%	90.90%	90.92%	71.83%	0.2	0.0	0.3
SVM	88.24%	88.24%	89.75%	86.29%	61.10%	-0.3	0.3	0.0

Figure 4. Result of Using K-Best without Smote

Figure 4 explains that using K-Best makes the results from KNN better with an average increase of 6%, both in terms of accuracy, precision, recall, and F1-Score, but for Random Forest, the results are the opposite, decreasing from when not using the K-Best and Smote filters. Then, for Decision Tree and SVM, there is no decrease or increase in performance; only SVM has an effect on time to train and time to predict.

Next, the second filter elimination process uses backward regression. Based on the experimental results using the backward regression elimination technique, we carried out feature selection by identifying the five most important features to improve classification performance. After carrying out the elimination process, the five most important features are obtained, namely: room temperature, remote temperature, voltage, ampere, and indoor check. The feature selection results can be seen in Figure 5.

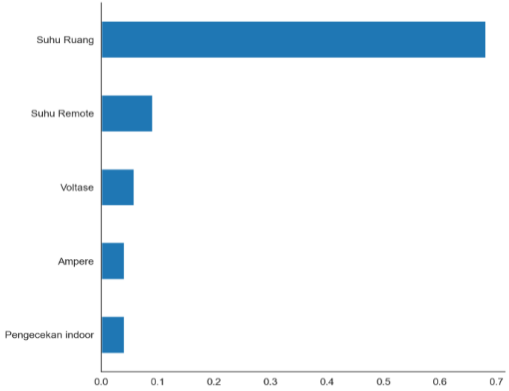
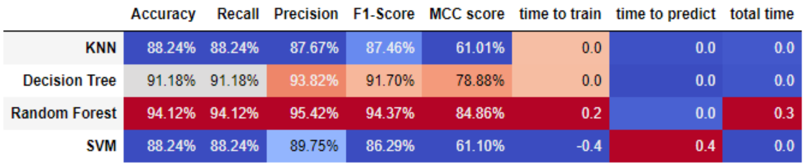


Figure 5. Feature Backward Regression without Smote

After that, we carried out the process method on the feature results, and the results can be seen in Figure 6.



	Accuracy	Recall	Precision	F1-Score	MCC score	time to train	time to predict	total time
KNN	88.24%	88.24%	87.67%	87.46%	61.01%	0.0	0.0	0.0
Decision Tree	91.18%	91.18%	93.82%	91.70%	78.88%	0.0	0.0	0.0
Random Forest	94.12%	94.12%	95.42%	94.37%	84.86%	0.2	0.0	0.3
SVM	88.24%	88.24%	89.75%	86.29%	61.10%	-0.4	0.4	0.0

Figure 6. Result of Backward Elimination without Smote

In Figure 6, it can be seen that the backward regression filter improves performance in KNN and Decision Tree both in terms of accuracy, precision, recall, and F1-score, but for Random Forest, the results are the opposite, which decreases compared to when not using the K-Best and Smote filters. Then SVM does not experience a decrease or increase in performance; only SVM has an effect on time to train and time to predict.

And in the last two steps, we used the K-Best feature with Oversampling Smote and then used the Backward Regression feature with Oversampling Smote, and the feature results obtained were as explained in the image below.

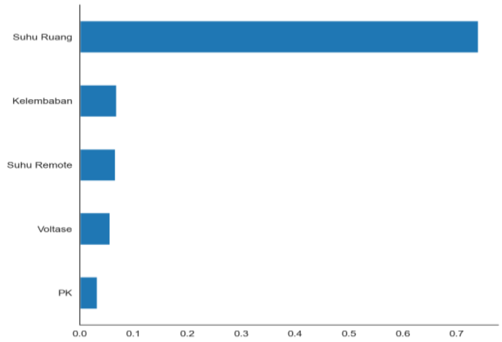
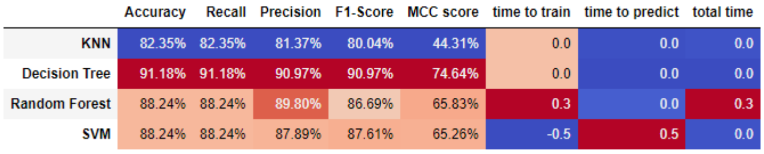


Figure 7. Feature Backward Elimination with Smote

By using backward regression and smote filters, the five best features are produced, namely room temperature, humidity, remote temperature, voltage, and PK. This, of course, changes the results of predictive maintenance.



	Accuracy	Recall	Precision	F1-Score	MCC score	time to train	time to predict	total time
KNN	82.35%	82.35%	81.37%	80.04%	44.31%	0.0	0.0	0.0
Decision Tree	91.18%	91.18%	90.97%	90.97%	74.64%	0.0	0.0	0.0
Random Forest	88.24%	88.24%	89.80%	86.69%	65.83%	0.3	0.0	0.3
SVM	88.24%	88.24%	87.89%	87.61%	65.26%	-0.5	0.5	0.0

Figure 8. Result of Backward Regression with Smote

Figure 8 shows that by using Oversampling Smote and Filter Backward Regression, the results of the Decision Tree method's performance increased, beating Random Forest, which experienced a decrease in performance in its predictive results. The decision tree is the most superior with the techniques applied in terms of accuracy, precision, recall, and F1-score, with an average increase of 3%. This is supported by the use of smote oversampling with feature selection results, as explained in Figure 9.

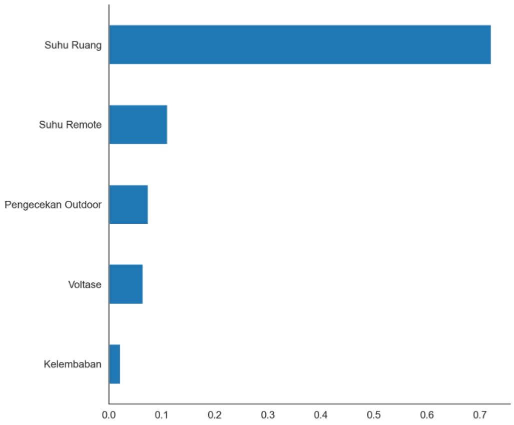
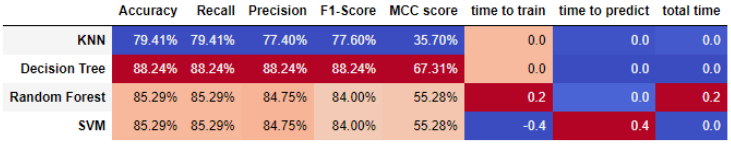


Figure 9. Feature with K-Best with Smote

By using the K-Best Filter and Smote Oversampling, the five best features are produced, namely room temperature, remote temperature, outdoor check, voltage, and humidity. This also changes the results of predictive maintenance.



	Accuracy	Recall	Precision	F1-Score	MCC score	time to train	time to predict	total time
KNN	79.41%	79.41%	77.40%	77.60%	35.70%	0.0	0.0	0.0
Decision Tree	88.24%	88.24%	88.24%	88.24%	67.31%	0.0	0.0	0.0
Random Forest	85.29%	85.29%	84.75%	84.00%	55.28%	0.2	0.0	0.2
SVM	85.29%	85.29%	84.75%	84.00%	55.28%	-0.4	0.4	0.0

Figure 10. Result of K-Best with Smote

Similar to the backward regression filter, in Figure 10, the results of using the K-best feature and smote oversampling make the decision tree beat the others in terms of accuracy, precision, recall, and F1-score. However, the results obtained are better using the backward regression feature compared to K-Best combined with Smote. This is also supported by the use of feature selection techniques.

3.2. Data Evaluation

To evaluate the effectiveness of the method proposed in this research, a comparison was made with previous research. Because the cases that occurred were different and the datasets were also different, we compared the performance of the methods used in this research. So that the results of the methods and techniques recommended in previous research can be compared. The evaluation results in previous research recorded the level of accuracy of methods in predictive air conditioner maintenance, namely SVM, MLP, and deep learning. Of the three methods, SVM has the best results with 97%. However, through the approach proposed in this research, we succeeded in obtaining a method with a performance level that is better than the performance of SVM by increasing accuracy, precision, racall, and f1-score.

Our results show that the proposed model can outperform other previously researched methods. The highest level of accuracy that we achieved was 97.06%. This success was achieved by applying the Random Forest algorithm without utilizing feature selection techniques. In this research, there were 3 methods chosen, namely KNN, Random Forest, and Decision Tree. The reason we chose these three methods is because KNN is because these three methods have a history of good classification cases. We want to find the best method for predictive air conditioner maintenance. In this research, Random Forest was proven to get the best results so that it could beat SVM, which had been carried out in previous research. Random Forest algorithm that relies on a random subset of selected variables. In classification, predictions are made by taking the majority of the results from these trees (Radivilova et al., 2019). The classification process using Random Forest involves merging trees, where the more trees used, the better the resulting accuracy. In Table 2, you can see the comparison results with previous research which used the same method.

Table 2. Performance Results of the proposed model compared to other research models

Reference & Model	Accuracy	Precision	Recall	F1-Score
[6] Decision tree C4.5	96.05	99.8	96.18	NA
[3] Random Forest & SVM	96.72 96.72	100 100	0 0	NA
Proposed Model SVM	88.24	89.75	88.24	86.29
Proposed Model Decision Tree & Backward Regression	91.18	93.82	91.18	91.70
Proposed Model Random Forest	97.06	97.43	97.06	97.17
Proposed Model KNN	85.29	84.63	85.29	84.86

In Table 2, you can see a comparison of the results of previous research with the state-of-the art method produced in this research. The evaluation results show that the accuracy score is 97.06%, precision is 97.43%, recall is 97.06%, and F1 score is 91.70% obtained from the method proposed by the researcher. Improvements were also seen in the latest research, with an increase in accuracy of 8.82%, precision of 7.68%, recall of 8.82%, and F1 value of 10.88%.

In research [6], the Decision Tree C4.5 method was used in classification cases. The results were very good, reaching accuracy of 96.05%, precision of 99.8%, and recall of 96.18%. However, because the cases handled are different, we used the decision tree method for the cases in this research, namely predictive air conditioner maintenance. The results obtained were accuracy of 88.24%, precision of 89.96%, F1-score of 88.74%, and recall of 88.24%. Then we applied the backward regression technique, and the results obtained succeeded in increasing accuracy performance by 2.94%, precision by 3.86%, recall by 2.94%, and F1 value by 2.96%.

Then, in research [3], Random Forest and SVM were carried out to solve classification problems as well. The results obtained from the Random Forest method were very good, namely accuracy of 96.72%, precision of 100%, and recall of 0%. The SVM method in this study produced the same figures. The results will be different if the training data is oversampled or undersampled. Then we used these two methods for predictive maintenance air conditioner research. The results of the Random Forest method without filters are accuracy 97.06%, precision 97.43%, and recall 97.06%. This resulted in an increase in accuracy of 0.34%, but precision decreased by 2.57%, and recall experienced a drastic increase of 97.06%.

Based on the results of this evaluation, the random forest model obtained better performance results compared to other methods in previous research. The contribution obtained from this research is to obtain the best method for increasing accuracy, precision, recall, and F1-Score in predictive air conditioner maintenance.

4. Conclusion

In this study, we conducted six experiments. The first approach proposed does not use feature selection techniques, resulting in the greatest accuracy of 97.06% using the Random Forest model algorithm. Then we continued with the second approach, namely using the smote oversampling technique, and obtained the greatest accuracy of 85.29% using the K-Nearest Neighbor model algorithm. Then the third experiment using K-Best Feature Selection without Smote oversampling produced the greatest accuracy by the Random Forest method, namely 91.18%. Then we continued with the fourth experiment using K-Best Feature Selection plus the Smote Oversampling technique, resulting in the highest accuracy of 88.24% obtained by the Decision Tree method. Then, in the fifth experiment, namely using the feature selection backward regression technique without using Smote, the highest accuracy was obtained by the random forest method with an accuracy of 94.12%. Then the sixth experiment, namely the final experiment, used the feature selection backward regression technique with smote oversampling, resulting in the highest accuracy of 91.18 obtained by the decision tree method.

Based on the comparison of methods presented in this research, The latest test results show the best methods for improvement in terms of accuracy, precision, recall, and F1-score. From the experimental results, it was found that the Random Forest model without using feature selection and smote oversampling, which had been analyzed, gave better results compared to other models and combinations of feature selection. The results obtained with accuracy were 97.06%, precision 97.43%, recall 97.06%, and F1 score 91.70%.

References:

[1] Rusmawan, Widyaningsih, T.W., “Identifikasi Kerusakan Air Conditioner Ruangan Dengan Metode Case Based Reasoning Berbasis Web”, Jurnal Sistem Komputer dan Kecerdasan Buatan, 2023

[2] Sulaiman, N.A., Abdullah, M.P., Abdullah, H., Zainudin, M.N.S , Yusop, A.M., “Fault detection for air conditioning system using machine learning”, International Journal of Artificial Intelligence (IJ-AI), 2020, https://doi.org/10.11591/ijai.v9.i1.pp109-116.

[3] Qadrini, L., Hikmah, Megasari, “Oversampling, Undersampling, Smote SVM dan Random Forest pada Klasifikasi Penerima Bidikmisi Sejawa Timur Tahun 2017”,

Journal of Computer System and Informatics (JoSYC), 2022, https://doi.org/10.47065/josyc.v3i4.2154.

[4] Rusmawan, Widyaningsih, T. W, “Identifikasi Kerusakan Air Conditioner Ruangan Dengan Metode Case Based Reasoning Berbasis Web”, Jurnal Sistem Komputer dan Kecerdasan Buatan, Volume VI Nomor 2 Maret 2023

[5] Nugroho, W, “Optimasi Metode K-Nearest Neighbours dengan Backward Elimination Menggunakan Dataset Software Effort Estimation”, Bianglala Informatika. Vol. 8 No. 2 – Tahun 2020

[6] Pambudi, R., Sriyanto , Firmansyah (2022). Klasifikasi Penyakit Stroke Menggunakan Algoritma Decision Tree C.45. Jurnal Teknika. https://doi.org/10.5281/zenodo.7535865