

## Komparasi Algoritme *Machine Learning* Untuk Prediksi Pemeliharaan Preventif *Precision Air Conditioning* di *Data Center*

Kahfi Heryandi Suradiradja<sup>1\*)</sup>; Dani Ramdani<sup>1</sup>; Karno Nano<sup>1</sup>

1. Universitas Pamulang, Jl. Suryakencana No.1, Pamulang Bar., Kec. Pamulang, Kota Tangerang Selatan, Banten 15417 Indonesia

<sup>\*)</sup>Email: dosen01514@unpam.ac.id

Received: 29 Juni 2024 | Accepted: 02 Januari 2025 | Published: 10 Januari 2025

### ABSTRACT

*Data Center is a primary facility for implementing an information technology service that applies the latest technology trends and must ensure the availability of services that can always be accessed, so supporting solutions are needed as part of the device maintenance strategy and operational monitoring process to maintain the availability of data center services. This research compares several machine learning classification algorithms to predict the condition of Precision Air Conditioning devices in data center operations. The dataset for this research is Precision Air Conditioning daily log data at the National Research and Innovation Agency. This research aims to identify several machine learning classification algorithms: Decision Tree, Random Forest, Artificial Neural Network Multi-Layer Perceptron, Naïve Bayes, and Support Vector Machine. The stages of this research method are analysis of the understanding of the maintenance status classification problem, data collection of Precision Air Conditioning log data, then continued with several machine learning algorithm modeling and evaluation to obtain an algorithm model with good accuracy results. The measurement results from the assessment of several machine learning methods in this research resulted in the C4.5 decision tree model having the best accuracy level, namely 98.75 percent. The roles generated from this model can be used to predict the condition of devices in the data center as a preventive.*

**Keywords:** Algorithm, Data Center, Decision Tree, Machine Learning, Precision Air Conditioning

### ABSTRAK

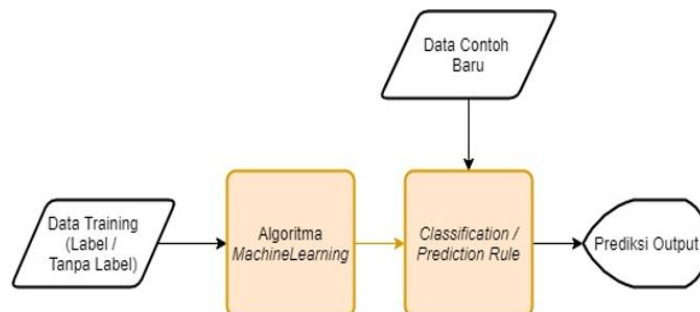
*Data Center atau pusat data merupakan fasilitas dasar untuk penerapan suatu layanan teknologi informasi yang menerapkan tren teknologi terkini harus terjamin ketersediaan layanannya selalu dapat diakses, sehingga dibutuhkan solusi pendukung sebagai bagian dari strategi pemeliharaan perangkat dan proses pemantauan operasional dalam rangka mempertahankan ketersediaan layanan data center. Pada penelitian ini membandingkan beberapa algoritme klasifikasi machine learning untuk memprediksi kondisi perangkat Precision Air Conditioning pada operasional data center. Dataset untuk penelitian ini adalah data log harian Precision Air Conditioning di Badan Riset dan Inovasi Nasional (BRIN). Tujuan dari penelitian ini, mengidentifikasi beberapa algoritme klasifikasi machine learning yakni Decision Tree, Random Forest, Artificial Neural Network Multi Layer Perceptron, Naïve Bayes dan Support Vector Machine. Tahapan dari metode penelitian ini yakni analisis pemahaman terhadap masalah pengklasifikasian status pemeliharaan, pengambilan data data log Precision Air Conditioning, kemudian dilanjutkan proses beberapa pemodelan algoritme – algoritme machine learning dan evaluasi untuk mendapatkan model algoritme dengan hasil akurasi yang baik. Hasil pengukuran pada evaluasi dari beberapa machine learning pada penelitian ini menghasilkan model decision tree C4.5 memiliki tingkat akurasi terbaik yakni 98,75 persen. Role yang dihasilkan dapat digunakan untuk memprediksi kondisi perangkat di data center sebagai tindakan preventif.*

**Kata kunci:** Algoritme, Data Center, Decision Tree, Machine Learning, Precision Air Conditioning

## 1. PENDAHULUAN

Pusat data merupakan serangkaian produk dan solusi yang menyediakan sumber daya komputasi dan penyimpanan [1]. Salah satu perangkat yang ada di *data center* adalah PAC (*Precision Air Conditioning*) yakni sistem pendingin yang menjaga suhu (18 sampai dengan 24 derajat celcius) serta kelembaban (*Relative Humidity* 50% dan toleransi  $\pm 5\%$ ) secara konstan. Layanan sistem yang berada dalam *data center* merupakan fasilitas dasar penerapan suatu layanan teknologi informasi seperti teknologi kecerdasan artifisial, *big data* hingga sistem komputasi berkinerja tinggi harus senantiasa dapat beroperasi dengan baik, sehingga layanan fasilitas data (catu daya, pendingin dan jaringan) harus dipastikan selalu tersedia. Salah satu yang harus dilakukan untuk memastikan agar ketersediaan layanan *data center* tetap dapat dijaga adalah dengan penerapan pemeliharaan yang baik agar tidak terjadi kegagalan dalam memberikan layanan. Kegagalan yang terjadi dapat menjadi akar penyebab downtime pusat data [2]. Pemeliharaan memiliki dampak yang besar dalam bisnis perusahaan dapat mengakibatkan berhentinya proses produksi, sehingga perlu dilakukan pemilihan strategi dalam pemeliharaan [3].

Menurut [4] definisi *machine learning* yaitu pemrograman komputer sebagai optimalisasi kriteria kinerja menggunakan data masa lalu. Menurut pendapat dari [5] bahwa umumnya metodologi dari Machine Learning melibatkan proses pembelajaran yang tujuan untuk belajar dari "pengalaman" dalam melaksanakan *task*. Ditunjukkan pada Gambar 1 proses secara umum *machine learning*.



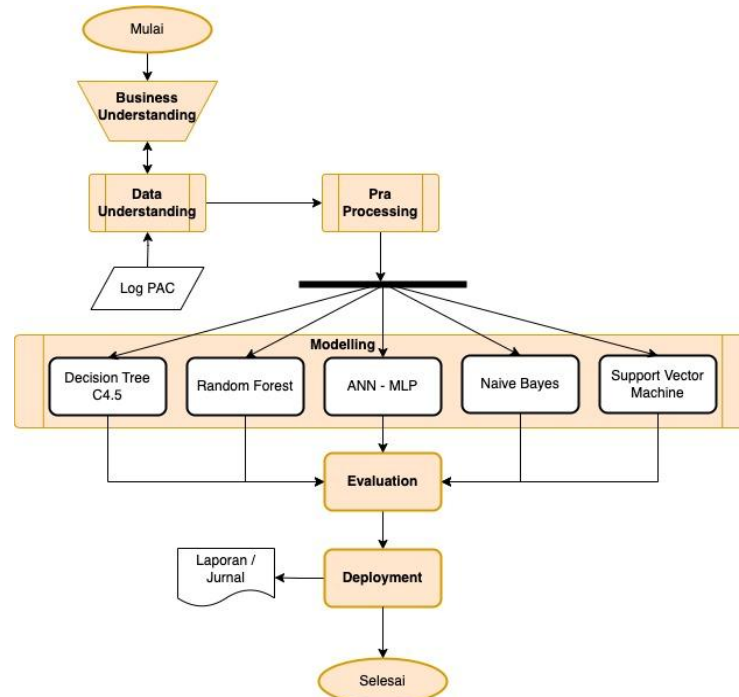
**Gambar 1.** Pendekatan machine learning secara umum (diolah dari [5])

Klasifikasi dalam *machine learning* yakni suatu proses untuk menemukan suatu fungsi atau model yang menggambarkan serta membedakan konsep atau kelas, bertujuan agar dapat menggunakan model dalam memprediksi sebuah kelas objek. Model tersebut berasal didasarkan pada analisis dari suatu set data pelatihan yakni, data objek yang kelas label diketahui [6].

Salah satu solusi terhadap permasalahan guna memastikan ketersediaan layanan *data center* tetap terjaga terutama pada pemeliharaan perangkat PAC yakni dibutuhkan sebuah model algoritme *machine learning* yang dapat membantu memberikan prediksi waktu pemeliharaan preventif PAC. Maka tujuan dari penelitian ini adalah membandingkan beberapa algoritme klasifikasi *machine learning* yakni *Decision Tree*, *Random Forest*, *Artificial Neural Network - Multi Layer Perceptron*, *Naïve Bayes* dan *Support Vector Machine* yang memiliki akurasi terbaik untuk memprediksi waktu pemeliharaan perangkat PAC di *data center*. Beberapa penelitian yang telah dilakukan [7], [8], [9] menggunakan machine learning yang terkait *air conditioning* namun berbeda atribut dataset yang digunakan.

## 2. METODE PENELITIAN

Prosedur kerja dalam penelitian ini menggunakan metode standar CRISP-DM yang disesuaikan dengan kondisi penelitian. Menurut [10] CRISP-DM, proyek *data mining* memiliki siklus dengan enam fase atau tahapan, yakni *Business Understanding*, *Data Understanding*, *Data Preparation*, *Modelling*, *Evaluation* dan *Deployment*. Urutan fase adalah bersifat adaptif, artinya fase berikutnya sering kali bergantung pada hasil yang terkait dengan fase sebelumnya (Gambar 2).



**Gambar 2.** Metode penelitian prediksi *maintenance* PAC

Digambarkan setiap fase dalam bagan siklus alir penelitian ini dengan penjelasan setiap proses sebagai berikut:

*a. Business/Research Understanding*

Tahapan pemahaman penelitian guna mengembangkan model klasifikasi prediksi sebagai langkah preventif kapan sebuah PAC memasuki waktu pemeliharaan dengan strategi perumusan identifikasi masalah, batasan serta tujuan penelitian.

*b. Data Understanding*

Penelitian dilaksanakan di Badan Riset dan Inovasi Nasional (BRIN) dengan mengambil data Laporan Bulanan Pengelolaan dan Pengembangan Pusat data Badan Riset dan Inovasi Nasional. Hasil pengumpulan data selama minimal 5 tahun dari lapangan tersebut diatur penempatannya dalam *data training*, *data testing* dan *data validation*.

*c. Data Preparation*

Persiapan dataset untuk penelitian ini mencakup penyesuaian target *class* dengan cara memberikan label berdasarkan arahan dari bagian *management engineering* sebagai pakar. Proses *preprocessing* pembagian data dilakukan sebab tidak meratanya sebaran setiap kelas atau label.

*d. Modelling*

Tahap pemilihan dan penerapan model yang tepat dan pengaturan model untuk hasil yang optimal. Pada proses ini data yang telah dipersiapkan pada proses *preprocessing* menjadi

bahan input untuk pembentukan model, baik sebagai data training testing dan validation. Luaran dari proses ini adalah model algoritme machine learning. Secara siklus prosesnya dari *modeling* dilanjutkan ke *evaluation* kemudian apabila ukuran *metric* pengukuran masih perlu ditingkatkan akan diproses kembali proses model dengan *tuning parameter* ataupun rekayasa algoritme.

e. *Evaluation*

Tahapan ini yakni proses evaluasi terhadapluaran proses model. Metode evaluasi yang akan digunakan adalah *confusion matrix* serta satuan metrik pengujian lainnya seperti *recall*, *precision*, dan F-1 score.

f. *Deployment*

Tahapan ini sebagai pemanfaatan model yang telah dibuat sesuai dengan hasil akhir *evaluation* yang diharapkan, diimplementasikan pada sebuah *prototype* model dan laporan penelitian serta publikasi yang merupakan luaran dari fase deployment ini dengan capaian indikatornya adalah model prediksi pemeliharaan preventif perangkat PAC di *data center*.

### 3. HASIL DAN PEMBAHASAN

Tahapan data understanding yakni pengumpulan dataset dengan sumber dari *Management Engineering* BRIN terdapat 289 data log harian dari bulan Juni 2023 hingga bulan November 2023 memiliki 11 inti atribut atau field (Tabel 3).

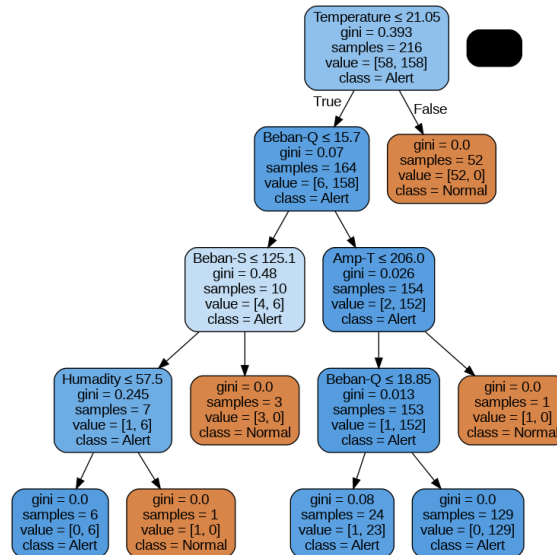
**Tabel 1.** Dataset log PAC

| VR  | VS  | VT  | AR  | AS  | AT  | BTA | BTQ | BTS | Temp | Hum | Warning |
|-----|-----|-----|-----|-----|-----|-----|-----|-----|------|-----|---------|
| 218 | 218 | 223 | 216 | 221 | 197 | 137 | 24  | 139 | 21   | 53  | N       |
| 218 | 218 | 223 | 216 | 221 | 197 | 137 | 24  | 139 | 20   | 53  | N       |
| 216 | 217 | 221 | 216 | 221 | 197 | 138 | 21  | 140 | 21   | 53  | N       |
| 218 | 219 | 223 | 223 | 227 | 203 | 141 | 23  | 143 | 21   | 55  | N       |
| 217 | 218 | 223 | 228 | 240 | 202 | 144 | 27  | 147 | 20   | 55  | N       |
| 217 | 218 | 223 | 228 | 240 | 202 | 144 | 27  | 147 | 20   | 58  | N       |
| 214 | 216 | 219 | 234 | 234 | 208 | 143 | 19  | 144 | 21   | 53  | Y       |
| ... | ... | ... | ... | ... | ... | ... | ... | ... | ...  | ... | ...     |

Selanjutnya pada tahapan *preparation* data dari 289 data, dibagi kedalam 216 buah *data training* dan 73 data untuk kebutuhan *testing data*. Pelaksanaan penelitian pada pemodelan menggunakan beberapa pilihan dari arsitektur algoritme yakni *Decision Tree C4.5*, *Random Forest*, *ANN Multi Layer Perceptron*, *Naïve Bayes* dan *Support Vector Machine*.

#### 3.1 *Decision Tree C4.5*

Model tree yang dihasilkan dari proses algoritme C4.5 dengan parameter *max\_depth*=4 dan *random\_state*=0 seperti ditunjukkan pada Gambar 3.



Gambar 3. Model *decision tree* C4.5 prediksi *maintenance PAC*

Rule dari model *decision tree* C4.5 *maintenance PAC* yang terbentuk dari Gambar 4 adalah sebagai berikut:

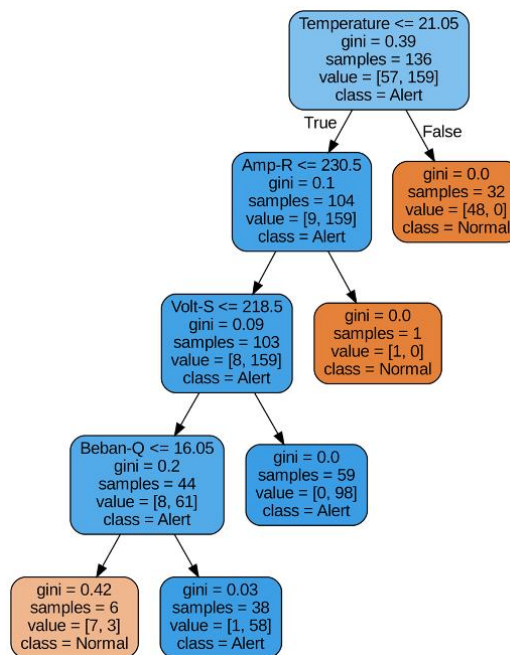
```

if Temperature <= 21.05:
    if Beban_Q <= 15.7:
        if Beban_S <= 125.1:
            if Humadity <= 57.5:
                return Normal
            if Humadity > 57.5:
                return Alert
        if Beban_S > 125.1:
            return Alert
    if Beban_Q > 15.7:
        if Amp_T <= 206:
            if Beban_Q <= 18.85:
                return Normal
            if Beban_Q > 18.85:
                return Normal
        if Amp_T > 206:
            return Alert
if Temperature > 21.05:
    return Alert
    
```

Evaluasi terhadap model *decision tree* tersebut menggunakan *cross validation fold* = 10, dimana *confusion matrix* dari hasil evaluasi tersebut ditunjukkan pada Gambar 4 untuk target *class* klasifikasi *alert* dan *normal*.

### 3.2 Random Forest

Model *tree* yang dihasilkan dari proses algoritme *Random Forest* dengan parameter *max\_depth*=4 dan *random\_state*=0 seperti ditunjukkan pada Gambar 5.



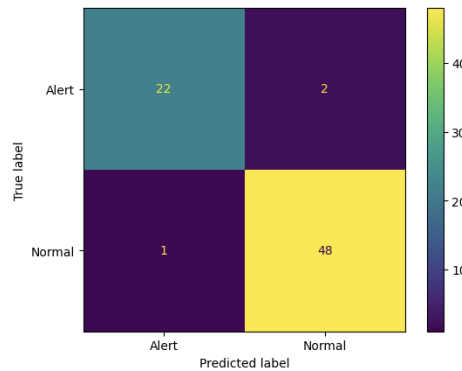
**Gambar 4.** Model *random forest* prediksi *maintenance PAC*

Rule dari model *random forest maintenance PAC* yang terbentuk dari Gambar 4 adalah sebagai berikut:

```

if Temperature <= 21.05:
    if Amp-R <= 230.5:
        if Volt-S <= 218.5:
            if Beban-Q <= 16.05:
                return Normal
            if Beban-Q > 16.05:
                return Alert
        if Volt-S > 218.5:
            return Alert
    if Amp-R > 230.5:
        return Normal
if Temperature > 21.05:
    return Normal
    
```

Evaluasi terhadap model *random forest* tersebut menggunakan *cross validation fold* = 10, dimana *confusion matrix* dari hasil evaluasi tersebut ditunjukkan pada Gambar 6 untuk target *class* klasifikasi *alert* dan *normal*.

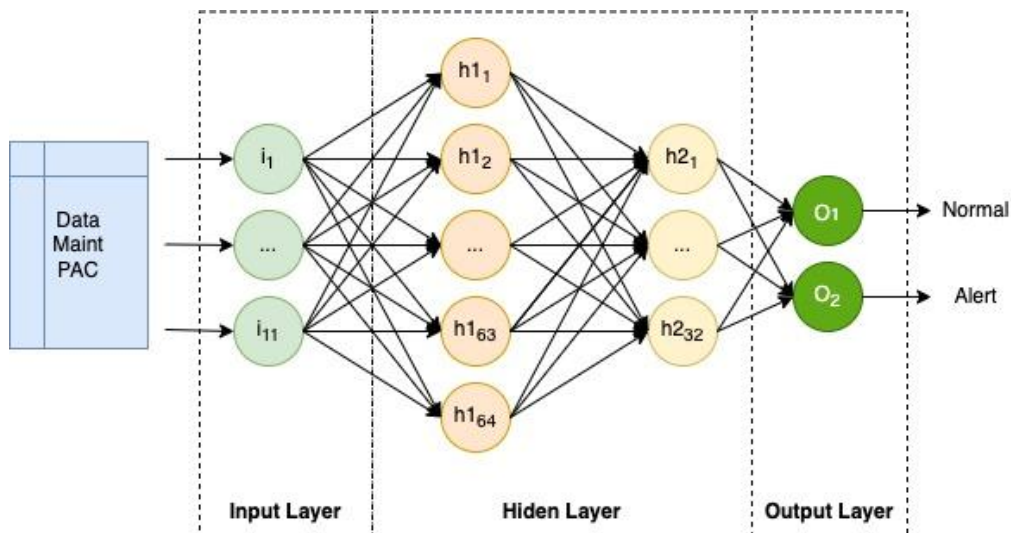


**Gambar 5.** Confusion matrix random forest maintenance PAC

Hasil pengujian pada tahapan evaluasi dari 79 data *testing*, terdapat 22 data target *class* untuk *alert* tepat diprediksi *alert* dan 48 data target *class* untuk normal, sedangkan data yang seharusnya dihasilkan target *class alert* tetapi hasilnya adalah normal sebanyak 2 buah data dan *target class* yang seharusnya normal tetapi dihasilkan *alert* sebanyak 1 buah data. Hasilkan akurasi dari evaluasi pengujian model *random forest* ini menghasilkan 0,975 atau 97,5 persen.

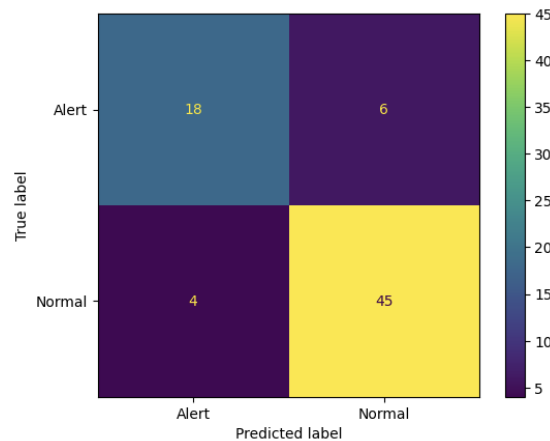
### 3.3. Artificial Neural Network - Multi Layer Perceptron

Model klasifikasi *maintenance PAC* dengan proses algoritme ANN - MLP menggunakan parameter *hidden\_layer\_sizes* = (64, 32), *random\_state* = 42, *max\_iter* = 1000 dan *activation* = *default* atau relu seperti ditunjukkan pada Gambar 7.



**Gambar 6.** Model ANN-MLP prediksi *maintenance PAC*

Evaluasi terhadap model ANN-MLP tersebut menggunakan *cross validation fold* = 10, dimana *confusion matrix* dari hasil evaluasi tersebut ditunjukkan pada Gambar 8 untuk *target class* klasifikasi *alert* dan normal.

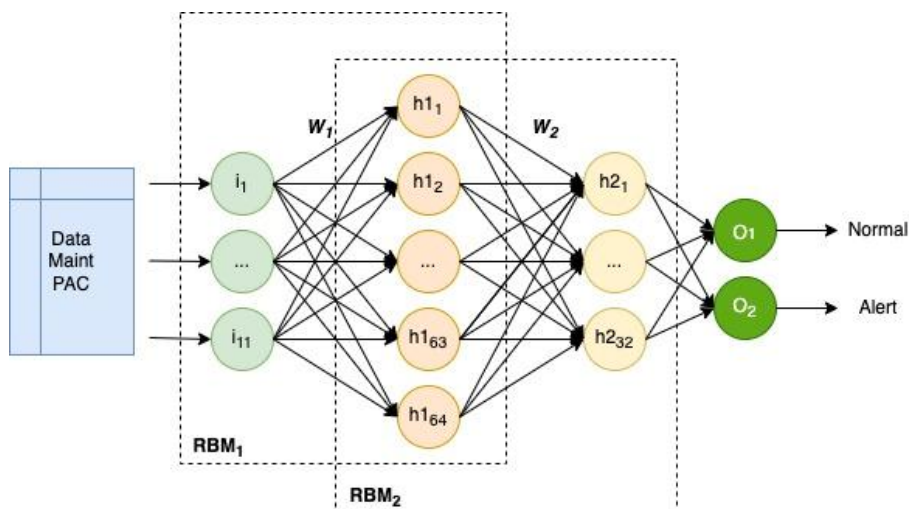


Gambar 7. Confusion matrix random forest maintenance PAC

Hasil pengujian pada tahapan evaluasi dari 79 data *testing*, terdapat 18 data target *class* untuk *alert* tepat diprediksi *alert* dan 45 data target *class* untuk normal, sedangkan data yang seharusnya dihasilkan target *class alert* tetapi hasilnya adalah normal sebanyak 6 buah data dan target *class* yang seharusnya normal tetapi dihasilkan *alert* sebanyak 4 buah data. Hasilkan akurasi dari evaluasi pengujian model ANN-MLP ini menghasilkan 0,9321428571428572 atau 93,21 persen.

### 3.4. Naïve Bayes

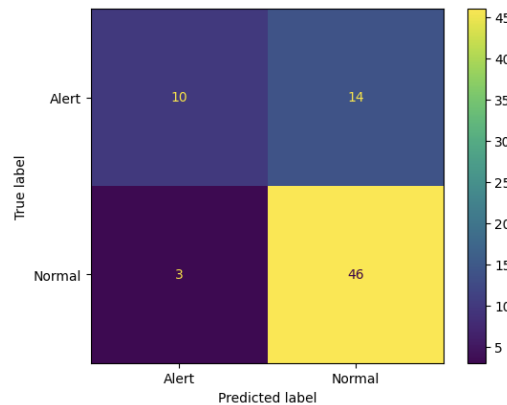
Model arsitektur klasifikasi *maintenance* PAC dengan proses algoritme *Naïve Bayes* atau juga dikenal dengan algoritme BernoulliRBM dengan default parameter yakni *learning\_rate* = 0.1, *n\_iter*=10 dan *batch\_size*=10 seperti ditunjukkan pada Gambar 9.



Gambar 8. Model naive bayes prediksi maintenance PAC

Evaluasi terhadap model *Naïve Bayes* tersebut menggunakan *cross validation fold* = 10, dimana *confusion matrix* dari hasil evaluasi tersebut ditunjukkan pada Gambar 10 untuk target *class* klasifikasi *alert* dan normal.



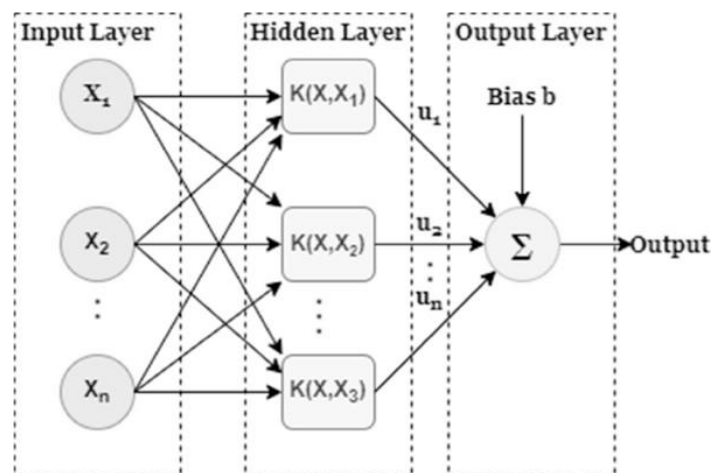


**Gambar 9.** Confusion matrix naive bayes maintenance PAC

Hasil pengujian pada tahapan evaluasi dari 79 data testing, terdapat 10 data target *class* untuk *alert* tepat diprediksi *alert* dan 46 data target *class* untuk normal, sedangkan data yang seharusnya dihasilkan target *class alert* tetapi hasilnya adalah normal sebanyak 14 buah data dan target *class* yang seharusnya normal tetapi dihasilkan *alert* sebanyak 3 buah data. Hasilkan akurasi dari evaluasi pengujian model *random forest* ini menghasilkan 0,9464285714285714 atau 94,64 persen.

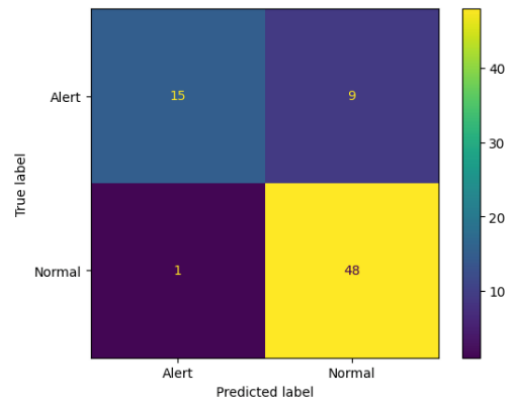
### 3.5. Support Vector Machine

Model arsitektur klasifikasi *maintenance PAC* dengan proses algoritme *Support Vector Machine* pada penelitian menggunakan default parameter yakni dengan  $C=1.0$ ,  $\text{kernel}='rbf'$  dan  $\text{degree}=3$  seperti ditunjukkan pada Gambar 11.



**Gambar 1.** Model SVM prediksi *maintenance PAC*

Evaluasi terhadap model SVM tersebut menggunakan *cross validation fold* = 10, dimana *confusion matrix* dari hasil evaluasi tersebut ditunjukkan pada Gambar 12 untuk target *class* klasifikasi *alert* dan normal.



**Gambar 2.** *Confusion matrix naive bayes maintenance PAC*

Hasil pengujian pada tahapan evaluasi dari 79 data testing, terdapat 15 data target *class* untuk *alert* tepat diprediksi *alert* dan 48 data target *class* untuk normal, sedangkan data yang seharusnya dihasilkan target *class alert* tetapi hasilnya adalah normal sebanyak 9 buah data dan target *class* yang seharusnya normal tetapi dihasilkan *alert* sebanyak 1 buah data. Hasilkan akurasi dari evaluasi pengujian model *random forest* ini menghasilkan 0,7285714285714284 atau 72,86 persen.

### 3.5. Hasil Evaluasi

Hasil evaluasi dari perbandingan lima model algoritme *machine learning* dengan Metode evaluasi dengan model pengukuran *confusion matrix* [11] ditunjukkan pada Tabel 4 yang memiliki akurasi tertinggi untuk prediksi *maintenance PAC* adalah model algoritme *Decision Tree* dengan tingkat akurasi 98,75 persen.

**Tabel 2.** Perbandingan algoritme machine learning prediksi *maintenance PAC*

| Model Algoritme                                         | Akurasi (Persen) |
|---------------------------------------------------------|------------------|
| <i>Decision Tree</i>                                    | 98,75            |
| <i>Random Forest</i>                                    | 97,5             |
| <i>Artificial Neural Network Multi Layer Perceptron</i> | 93,21            |
| <i>Naïve Bayes</i>                                      | 94,64            |
| <i>Support Vector Machine</i>                           | 72,86            |

## 4. KESIMPULAN DAN SARAN

Kesimpulan hasil penelitian dan pengukuran evaluasi dari penelitian ini yakni penggunaan model algoritme machine learning menggunakan *decision tree* C4.5 dapat digunakan dalam memprediksi *maintenance PAC*, dimana rule yang dihasilkan dari model ini dapat digunakan oleh pengembang sistem sebagai output rekomendasi dari hasil prediksi *maintenance* perangkat. Saran terhadap hasil penelitian ini dapat dilanjutkan pada implementasi pengembangan system *alert maintenance PAC* secara otomatis dari sistem PAC terintegrasi ke *alert* sistem yang dapat dikembangkan.

## UCAPAN TERIMA KASIH

Ucapan terima kasih kepada Badan Riset Inovasi Nasional yang telah mendukung dalam penggunaan dataset penelitian ini.

## DAFTAR PUSTAKA

- [1] D. V. Dumsky and E. A. Isaev, "Data centers for physical research," in *Physics Procedia*, Elsevier B.V., 2015, pp. 298–302. doi: 10.1016/j.phpro.2015.08.330.
- [2] M. Wiboonrat, "Condition-Based Maintenance for Data Center Operations Management." [Online]. Available: [www.intechopen.com](http://www.intechopen.com)
- [3] I. Errandonea, S. Beltrán, and S. Arrizabalaga, "Digital Twin for maintenance: A literature review," *Computers in Industry*, vol. 123. Elsevier B.V., Dec. 01, 2020. doi: 10.1016/j.compind.2020.103316.
- [4] E. Alpaydm, *Introduction to Machine Learning Second Edition*. The MIT Press, 2010. [Online]. Available: <http://www.ncbi.nlm.nih.gov/pubmed/24272434>
- [5] K. G. Liakos, P. Busato, D. Moshou, S. Pearson, and D. Bochtis, "Machine learning in agriculture: A review," *Sensors (Switzerland)*, vol. 18, no. 8, pp. 1–29, 2018, doi: 10.3390/s18082674.
- [6] Han, Jiawei, Kamber, and Micheline, *Data Mining: Concepts and Techniques Second Edition*, 2nd ed. San Francisco: Morgan Kaufmann, 2007.
- [7] M. Manivannan, B. Najafi, and F. Rinaldi, "Machine learning-based short-term prediction of air-conditioning load through smart meter analytics," *Energies (Basel)*, vol. 10, no. 11, Nov. 2017, doi: 10.3390/en10111905.
- [8] M. Sari, M. A. Berawi, S. P. Larasati, S. I. Susilowati, B. Susantono, and R. Woodhead, "Developing Machine Learning Model to Predict HVAC System of Healthy Building: A Case Study in Indonesia," *International Journal of Technology*, vol. 14, no. 7, pp. 1438–1448, 2023, doi: 10.14716/ijtech.v14i7.6682.
- [9] N. A. Sulaiman, M. P. Abdullah, H. Abdullah, M. N. S. Zainudin, and A. M. Yusop, "Fault detection for air conditioning system using machine learning," *IAES International Journal of Artificial Intelligence*, vol. 9, no. 1, pp. 109–116, 2020, doi: 10.11591/ijai.v9.i1.pp109-116.
- [10] D. T. Larose, "Discovering Knowledge in Data: An Introduction to Data Mining," 2005.
- [11] D. Xhemali, C. J. Hinde, and R. G. Stone, "Naïve Bayes vs. Decision Trees vs. Neural Networks in the Classification of Training Web Pages."