

# Analisis Pola Konsumsi Energi Listrik pada Gardu Induk Kelapa Gading dan Gardu Induk Sunter Menggunakan Metode *Naïve Bayes*

Putri Angraeni<sup>1</sup>, Putri Angraeni Badar<sup>2</sup>

Program Studi Teknik Komputer, Universitas Negeri Makassar  
Jl. Mallengkeri Raya, Parangtambung, Kec. Tamalate, Kota Makassar, Sulawesi Selatan 90224

## Abstrak

Peningkatan permintaan energi listrik di Indonesia menuntut ketersediaan dan distribusi energi listrik yang optimal dan efisien. Untuk mengatasi tantangan ini, diperlukan upaya untuk mengelola dan mengoptimalkan sistem distribusi energi listrik secara cerdas dan efektif. Gardu induk merupakan bagian penting dari sistem distribusi energi listrik. Gardu induk berfungsi untuk menurunkan tegangan listrik dari pembangkit atau transmisi menjadi tegangan yang sesuai untuk didistribusikan ke pelanggan melalui jaringan distribusi. Oleh karena itu, analisis pola konsumsi energi pada gardu induk menjadi kritis untuk perencanaan dan peningkatan sistem distribusi. Area gardu induk yang digunakan dalam penelitian ini adalah gardu induk Kelapa Gading dan gardu induk Sunter di Jakarta Utara. Metode yang digunakan untuk menganalisis pola konsumsi energi listrik adalah metode *Naïve Bayes* dengan menggunakan pemrograman python. Hasil penelitian menunjukkan bahwa hasil prediksi sesuai dengan label sebenarnya, dengan tingkat akurasi yang dihasilkan untuk data latih sebesar 96.428%. Sementara akurasi yang dihasilkan untuk data uji sebesar 98.214%. Dengan demikian, metode *Naïve Bayes* terbukti efektif dalam memprediksi pola konsumsi energi listrik pada kedua gardu induk. Informasi yang dihasilkan diharapkan dapat menjadi dasar bagi PLN, pemerintah, dan masyarakat dalam mengoptimalkan pengelolaan energi listrik untuk memenuhi kebutuhan yang terus meningkat.

**Kata kunci:** Gardu Induk, Listrik, *Naïve Bayes*, Python, Prediksi

## Abstract

*The increasing demand for electrical energy in Indonesia demands optimal and efficient availability and distribution of electrical energy. To overcome this challenge, efforts are needed to manage and optimize the electrical energy distribution system intelligently and effectively. Substations are an important part of the electrical energy distribution system. The substation serves to reduce the voltage of electricity from the generator or transmission to a voltage suitable for distribution to customers through the distribution network. Therefore, analyzing energy consumption patterns at substations is critical for distribution system planning and improvement. The substation areas used in this study are Kelapa Gading substation and Sunter substation in North Jakarta. The method used to analyze electrical energy consumption patterns is the Naïve Bayes method using python programming. The results showed that the prediction results were in accordance with the actual label, with an accuracy rate generated for training data of 96.428%. While the accuracy generated for testing data is 98.214%. Thus, the Naïve Bayes method proved effective in predicting the pattern of electrical energy consumption at both substations. The resulting information is expected to be the basis for PLN, the government, and the community in optimizing the management of electrical energy to meet increasing needs.*

**Keywords:** Substation, Electricity, *Naïve Bayes*, Python, Prediction

## I. PENDAHULUAN

Peningkatan kecepatan perkembangan teknologi di berbagai sektor, menyebabkan dampak berupa kenaikan permintaan akan energi listrik. Energi listrik merupakan kebutuhan pokok yang penting dalam kehidupan manusia. Energi listrik selain digunakan untuk keperluan pencahayaan dan kebutuhan rumah tangga semacam televisi, kulkas, kipas angin, dan setrika, energi listrik juga menjadi penunjang bagi sektor pendidikan, industri, ekonomi, dan umum. Hampir semua aktivitas manusia membutuhkan energi listrik. PT. PLN (Perusahaan Listrik Negara) dibentuk oleh pemerintah untuk mengelola dan mendistribusikan energi listrik [1]. Perusahaan Listrik Negara (PLN) adalah badan usaha milik negara yang bertanggung jawab atas penyediaan energi listrik di Indonesia. PLN menjalankan kegiatan usahanya mulai dari pembangkitan, transmisi, distribusi, hingga penjualan energi listrik kepada konsumen [2].

Indonesia sebagai negara berkembang membutuhkan energi listrik yang terus meningkat. Hal ini disebabkan oleh pertumbuhan ekonomi dan penduduk yang semakin pesat. Selain itu, perkembangan penduduk juga turut mempengaruhi perkembangan energi listrik, yaitu dalam hal jumlah pelanggan yang akan dialiri listrik. Menurut laporan Kementerian

ESDM, konsumsi listrik per kapita Indonesia pada tahun 2022 mencapai 1.173 kWh/kapita, naik sekitar 4% dari tahun sebelumnya. Kenaikan konsumsi listrik yang tinggi ini menuntut ketersediaan dan distribusi energi listrik yang optimal dan efisien. Meskipun ketersediaan dan distribusi energi listrik di Indonesia telah mengalami peningkatan, masih terdapat beberapa tantangan yang perlu dihadapi. Tantangan-tantangan tersebut meliputi kapasitas pembangkit yang terbatas, kerugian teknis dan non-teknis, gangguan pasokan, dan rendahnya efisiensi sistem. Untuk mengatasi tantangan-tantangan tersebut, diperlukan upaya untuk mengelola dan mengoptimalkan sistem distribusi energi listrik secara cerdas dan efektif. Kebutuhan energi listrik yang terus meningkat harus diantisipasi sejak dini. Hal ini agar ketersediaan energi listrik dapat terpenuhi dalam jumlah yang cukup dan harga yang terjangkau [3].

Gardu induk adalah bagian penting dari sistem distribusi energi listrik. Gardu induk berfungsi untuk menurunkan tegangan listrik dari pembangkit atau transmisi menjadi tegangan yang sesuai untuk didistribusikan ke pelanggan melalui jaringan distribusi. Gardu induk juga berperan penting dalam menjaga kualitas dan kelancaran pasokan listrik, serta mengelola beban dan kehilangan energi dalam jaringan. Oleh karena itu, analisis pola konsumsi energi pada gardu induk sangat penting untuk dilakukan. Analisis ini dapat memberikan informasi tentang karakteristik konsumsi energi listrik, seperti waktu puncak konsumsi, jenis beban yang dominan, dan faktor-faktor yang mempengaruhi konsumsi energi listrik. Informasi ini dapat digunakan untuk membantu merencanakan, memperbaiki, dan meningkatkan sistem distribusi energi listrik [4].

Gardu induk Kelapa Gading dan gardu induk Sunter merupakan dua gardu induk yang memasok energi listrik ke wilayah Jakarta Utara. Wilayah ini merupakan salah satu wilayah dengan konsumsi energi listrik tertinggi di Indonesia. Kedua gardu induk tersebut memiliki kapasitas terpasang masing-masing sebesar 150 MVA dan 100 MVA, serta menyuplai energi listrik ke sekitar 300.000 pelanggan. Pola konsumsi energi listrik pada kedua gardu induk tersebut dipengaruhi oleh berbagai faktor, seperti jenis pelanggan, waktu, musim, cuaca, dan aktivitas sosial [5].

Untuk menganalisis pola konsumsi energi listrik pada kedua gardu induk tersebut, penulis menggunakan metode *Naïve Bayes*. *Naïve Bayes* adalah metode klasifikasi probabilistik yang sederhana dan efisien. Metode ini menggunakan Teorema *Bayes* untuk memprediksi peluang suatu objek berada di dalam suatu kelas. *Naïve Bayes* mengasumsikan bahwa nilai dari setiap fitur atau variabel independen bersifat saling bebas atau tidak berkorelasi. Artinya, nilai satu fitur tidak mempengaruhi nilai fitur lainnya. Asumsi ini dapat menyederhanakan perhitungan probabilitas, sehingga metode *Naïve Bayes* menjadi lebih mudah diimplementasikan dan cepat dalam komputasi [6].

Beberapa penelitian sebelumnya telah dilakukan untuk klasifikasi menggunakan metode *Naïve Bayes*. Salah satunya yaitu penelitian memprediksi penggunaan listrik rumah tangga menggunakan metode *Naïve Bayes* dengan tingkat akurasi yang cukup tinggi, yaitu 78,3333% [7]. Selain itu, pada penelitian analisis sentimen perusahaan listrik negara cabang Ambon menggunakan metode *Support Vector Machine* dan *Naïve Bayes Classifier* menunjukkan hasil 67,20% [8]. Penelitian lainnya klasifikasi kestabilan jaringan listrik dengan SVM Dan *Naïve Bayes* dengan nilai akurasi sebesar 97,64% [9]. Lalu, penelitian analisis sentimen konsumen toko *online* JD.ID. menggunakan algoritma *Naïve Bayes Classifier* hasil analisa menunjukkan akurasi 96,44% [10].

Dengan demikian, dilakukan penelitian “Analisis Pola Konsumsi Energi Listrik pada Gardu Induk Kelapa Gading dan Gardu Induk Sunter Menggunakan Metode *Naïve Bayes*” dengan bahasa pemrograman python. Penelitian ini bertujuan untuk mengidentifikasi pola konsumsi energi listrik pada gardu induk Kelapa Gading dan gardu induk Sunter, serta memberikan rekomendasi untuk meningkatkan efisiensi dan kualitas sistem distribusi energi listrik. Penelitian ini diharapkan dapat memberikan memberikan informasi yang berguna bagi PLN sebagai penyedia layanan energi listrik, serta bagi pemerintah dan masyarakat.

## II. TINJAUAN PUSTAKA

### 1.2 Listrik

Listrik adalah bentuk energi yang dihasilkan oleh pergerakan partikel bermuatan listrik, yaitu elektron. Elektron dapat bergerak dari satu atom ke atom lainnya melalui suatu penghantar. Listrik dapat dihasilkan dari berbagai sumber energi, seperti tenaga air, tenaga surya, dan tenaga nuklir. Listrik merupakan sumber energi yang sangat penting bagi kehidupan sehari-hari. Hampir semua perangkat yang digunakan dalam kehidupan sehari-hari, seperti lampu, televisi, komputer, dan mesin industri, membutuhkan listrik untuk beroperasi. Listrik dapat digunakan untuk berbagai keperluan, seperti penerangan, pemanas, pendingin, penggerak mesin, dan komunikasi [11].

## 2.2 KWH

KWH atau *kilowatt-hour* adalah satuan yang digunakan untuk mengukur jumlah energi listrik yang dikonsumsi. Satu KWH setara dengan energi yang dikonsumsi oleh perangkat dengan daya 1 kilowatt selama 1 jam [12].

## 2.3 Naïve Bayes Classifier

*Naïve Bayes Classifier* adalah model klasifikasi sederhana yang menggunakan Teorema Bayes. Model ini mengasumsikan bahwa variabel-variabel yang digunakan untuk klasifikasi saling independen. *Naïve Bayes* dapat digunakan untuk mengklasifikasikan suatu variabel tertentu dengan menggunakan metode probabilitas dan statistik.

Prediksi Bayes adalah metode untuk memprediksi probabilitas suatu peristiwa berdasarkan informasi yang tersedia. Metode ini didasarkan pada Teorema Bayes, yang menyatakan bahwa probabilitas suatu peristiwa dapat dihitung dengan mengalikan probabilitas a priori peristiwa tersebut dengan probabilitas posteriori peristiwa tersebut [13].

## 2.4 Persamaan Metode Naïve Bayes

Persamaan Teorema Bayes, yaitu [14][15]:

$$P(H|X) = \frac{P(X|H) \cdot P(H)}{P(X)} \quad (1)$$

Keterangan:

X adalah data yang belum diketahui kelasnya.

H adalah hipotesis bahwa data tersebut termasuk dalam kelas tertentu.

$P(H|X)$  adalah probabilitas hipotesis H berdasarkan kondisi X, atau disebut juga posteriori probabilitas.

$P(H)$  adalah probabilitas hipotesis H secara umum, atau disebut juga prior probabilitas.

$P(X|H)$  adalah probabilitas X berdasarkan kondisi pada hipotesis H.

$P(X)$  adalah probabilitas X secara umum.

Untuk menjelaskan metode *Naïve Bayes*, perlu dipahami bahwa proses klasifikasi memerlukan sejumlah fitur untuk menentukan kelas apa yang cocok bagi sampel yang dianalisis tersebut. Oleh karena itu, metode *Naïve Bayes* di atas disesuaikan [16]:

$$P(C|F_1 \dots F_n) = \frac{P(C)P(F_1 \dots F_n|C)}{P(F_1 \dots F_n)} \quad (2)$$

Rumus tersebut menjelaskan bahwa peluang masuknya sampel karakteristik tertentu dalam kelas C (Posterior) dapat dihitung dengan mengalikan peluang munculnya kelas C (prior) dengan peluang kemunculan karakteristik-karakteristik sampel pada kelas C (likelihood), kemudian membaginya dengan peluang kemunculan karakteristik-karakteristik sampel secara global (evidence). Variabel C merepresentasikan kelas, sedangkan variabel  $F_1 \dots F_n$  merepresentasikan karakteristik petunjuk yang dibutuhkan untuk melakukan klasifikasi. Rumus tersebut dapat ditulis secara sederhana sebagai berikut [17]:

$$\text{Posterior} = \frac{\text{prior} \times \text{likelihood}}{\text{evidence}} \quad (3)$$

Nilai evidence selalu sama untuk setiap kelas pada satu sampel. Nilai dari posterior tersebut nantinya akan dibandingkan dengan nilai-nilai posterior kelas lainnya untuk menentukan ke kelas apa suatu sampel akan diklasifikasikan. Rumus Bayes dapat dijabarkan lebih lanjut dengan menggunakan aturan perkalian sebagai berikut [18]:

$$\begin{aligned} P(C|F_1, \dots, F_n) &= P(C)P(F_1, \dots, F_n|C) \\ &= P(C)P(F_1|C)P(F_2, \dots, F_n|C, F_1) \\ &= P(C)P(F_1|C)P(F_2|C, F_1)P(F_3, \dots, F_n|C, F_1, F_2) \\ &= P(C)P(F_1|C)P(F_2|C, F_1)P(F_3|C, F_1, F_2)P(F_4, \dots, F_n|C, F_1, F_2, F_3) \\ &= P(C)P(F_1|C)P(F_2|C, F_1)P(F_3|C, F_1, F_2) \dots P(F_n|C, F_1, F_2, F_3, \dots, F_{n-1}) \end{aligned} \quad (4)$$

Dari hasil penjabaran rumus Bayes, terlihat bahwa semakin banyak faktor yang mempengaruhi nilai probabilitas, maka semakin kompleks perhitungannya. Hal ini hampir mustahil untuk dianalisis satu per satu. Oleh karena itu,

digunakan asumsi independensi yang sangat tinggi (naif), yaitu masing-masing petunjuk ( $F_1, F_2, \dots, F_n$ ) saling bebas satu sama lain. Dengan asumsi tersebut, maka berlaku suatu kesamaan sebagai berikut [18]:

$$P(F_i|F_j) = \frac{P(F_i \cap F_j)}{P(F_j)} = \frac{P(F_i)P(F_j)}{P(F_j)} = P(F_i) \quad (5)$$

Untuk  $i \neq j$ , sehingga

$$P(F_i|C, F_j) = P(F_i|C) \quad (6)$$

Persamaan di atas merupakan model dari teorema *Naïve Bayes* yang digunakan dalam proses klasifikasi. Untuk klasifikasi dengan data kontinyu, digunakan rumus *Densitas Gauss* [18].

$$P(X_i = x_i | Y = y_j) = \frac{1}{\sqrt{2\pi\sigma_{ij}}} e^{-\frac{(x_i - \mu_{ij})^2}{2\sigma_{ij}^2}} \quad (7)$$

Keterangan:

$P$  : Peluang terjadinya suatu peristiwa

$X_i$  : Atribut ke- $i$ , yaitu salah satu fitur atau karakteristik yang digunakan untuk mengklasifikasikan data

$x_i$  : Nilai atribut ke- $i$

$Y$  : Kelas yang dicari, yaitu kategori data yang ingin diklasifikasikan

$y_i$  : Sub kelas  $Y$  yang dicari, yaitu kategori data yang lebih spesifik dari  $Y$

$\mu$  : Rata-rata dari seluruh atribut

$\sigma$  : Varian dari seluruh atribut

## 2.5 Confusion Matrix

*Confusion Matrix* memiliki perhitungan seperti akurasi, presisi, dan *recall*. Akurasi adalah persentase data yang diklasifikasikan dengan benar. Presisi adalah metrik untuk mengukur seberapa akurat sistem dalam menghasilkan data yang relevan. *Recall* adalah metrik untuk mengukur seberapa lengkap sistem dalam menghasilkan data yang relevan [19]. Rumus untuk menghitung nilai akurasi [20]:

$$\text{Akurasi} = \frac{TP + TN}{TP + TN + FN + FP}$$

Rumus untuk menghitung nilai presisi [20]:

$$\text{Precision} = \frac{TP}{TP + FP}$$

Rumus untuk mendapatkan *recall* [20]:

$$\text{Recall} = \frac{TP}{TP + FN}$$

Keterangan:

TP (True Positive): prediksi yang benar bahwa suatu data termasuk dalam kelas positif dan kelas sebenarnya juga positif.  
TN (True Negative) adalah prediksi yang benar bahwa suatu data termasuk dalam kelas negatif dan kelas sebenarnya juga negatif.

FP (False Positive) adalah prediksi yang salah bahwa suatu data termasuk dalam kelas positif padahal kelas sebenarnya negatif.

FN (False Negative) adalah prediksi yang salah bahwa suatu data termasuk dalam kelas negatif padahal kelas sebenarnya positif.

## 2.6 Pseudocode

Adapun *pseudocode* algoritma *Naïve Bayes* [21]:

```

Begin

// Memisahkan data menjadi data training dan data test (split data):
data_training, data_test = Split_Data(data, persentase_split)

//Inisialisasi variabel:

```

```

classes = {tinggi, rendah, normal}
probabilitas_kelas = Dictionary()
probabilitas_fitur_diberi_kelas = Dictionary()
hasil_klasifikasi = List() // Untuk menyimpan hasil prediksi kelas data test
label_asli = List() // Untuk menyimpan label kelas asli data test

// Hitung jumlah total data training (N) dan jumlah data dalam setiap kelas:
for class in classes do
    count_class = Hitung jumlah data training dengan kelas 'class'
    probabilitas_kelas[class] = count_class / N
end

// Untuk setiap fitur pada dataset, hitung probabilitas kemunculan fitur pada setiap kelas:
for class in classes do
    for setiap fitur pada dataset do
        count_fitur_diberi_kelas = Hitung jumlah data training dengan fitur dan kelas 'class'
        probabilitas_fitur_diberi_kelas[(fitur, class)] = count_fitur_diberi_kelas / count_class
    end
end

// Untuk setiap data dalam data test do
// Hitung probabilitas posterior untuk setiap kelas:
for class in classes do
    probabilitas_posterior[class] = probabilitas_kelas[class]

    for setiap fitur pada data test do
        probabilitas_posterior[class] *= probabilitas_fitur_diberi_kelas[(fitur, class)]
    end
end

Tentukan kelas dengan probabilitas posterior tertinggi sebagai prediksi kelas data.
End

Simpan hasil prediksi kelas dalam hasil_klasifikasi
Simpan label kelas asli dalam label_asli
end

//Evaluasi hasil klasifikasi:
hitung_akurasi = Hitung akurasi(hasil_klasifikasi, label_asli)
hitung_presisi = Hitung presisi(hasil_klasifikasi, label_asli)
hitung_recall = Hitung recall(hasil_klasifikasi, label_asli)
hitung_f1 = Hitung F1-score(hasil_klasifikasi, label_asli)

//Output hasil evaluasi:
Tampilkan "Akurasi:", hitung_akurasi
Tampilkan "Presisi:", hitung_presisi
Tampilkan "Recall:", hitung_recall
Tampilkan "F1-score:", hitung_f1
End

```

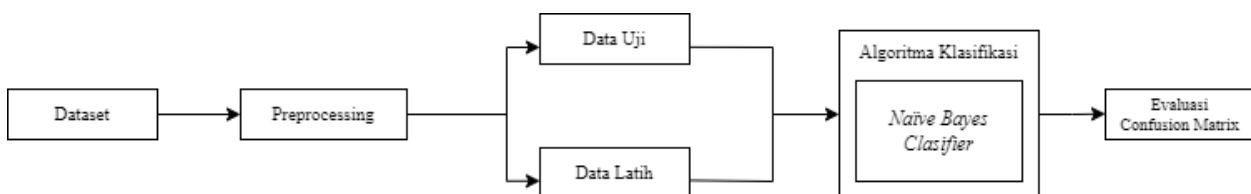
Adapun penjelasan lebih rinci untuk *pseudocode* yang mencakup pembagian data, pelatihan model *Naïve Bayes*, dan evaluasi hasil klasifikasi adalah sebagai berikut:

- 1) Memisahkan data menjadi data *training* dan data *test* (split data):
  - a. Data dibagi menjadi dua bagian, yaitu data\_training dan data\_test. Pembagian data ini dapat dilakukan sebelum memulai algoritma *Naive Bayes*, dan biasanya dilakukan untuk menguji kinerja model pada data yang tidak digunakan selama pelatihan. Data\_test digunakan untuk mengevaluasi model.
- 2) Inisialisasi variabel:
  - a. *classes*: Kumpulan kelas yang mungkin, dalam kasus ini {tinggi, rendah, normal}.
  - b. *probabilitas\_kelas*: Dictionary untuk menyimpan probabilitas prior dari setiap kelas.
  - c. *probabilitas\_fitur\_diberi\_kelas*: *Dictionary* untuk menyimpan probabilitas kemunculan fitur dalam konteks setiap kelas.

- d. hasil\_klasifikasi: Daftar untuk menyimpan hasil prediksi kelas data\_test.
- e. label\_asli: Daftar untuk menyimpan label kelas asli data\_test.
- 3) Hitung jumlah total data *training* (N) dan jumlah data dalam setiap kelas:
  - a. Iterasi melalui setiap kelas (tinggi, rendah, normal).
  - b. Hitung jumlah data dalam setiap kelas, yaitu count\_class.
  - c. Hitung probabilitas prior untuk setiap kelas (probabilitas\_kelas[class]) sebagai rasio jumlah data dalam kelas tersebut terhadap jumlah total data training (N).
- 4) Untuk setiap fitur pada dataset, hitung probabilitas kemunculan fitur pada setiap kelas:
  - a. Iterasi melalui setiap kelas.
  - b. Iterasi melalui setiap fitur pada data\_training.
  - c. Hitung jumlah data *training* yang memiliki fitur tersebut dan berada dalam kelas tertentu (count\_fitur\_diberi\_kelas).
  - d. Hitung probabilitas kemunculan fitur dalam konteks setiap kelas (probabilitas\_fitur\_diberi\_kelas[(fitur, class)]) sebagai rasio count\_fitur\_diberi\_kelas terhadap count\_class.
- 5) Untuk setiap data dalam data *test* do:
  - a. Hitung probabilitas posterior untuk setiap kelas.
  - b. Iterasi melalui setiap kelas.
  - c. Inisialisasi probabilitas\_posterior[class] dengan probabilitas\_kelas[class].
  - d. Iterasi melalui setiap fitur pada data\_test.
  - e. Kalikan probabilitas\_posterior[class] dengan probabilitas\_fitur\_diberi\_kelas[(fitur, class)].
- 6) Tentukan kelas dengan probabilitas posterior tertinggi sebagai prediksi kelas data.
  - a. Pilih kelas dengan probabilitas\_posterior tertinggi sebagai kelas prediksi untuk data tersebut.
  - b. Simpan kelas prediksi dalam hasil\_klasifikasi.
- 7) Simpan label kelas asli dalam label\_asli:
  - a. Simpan label kelas asli dari data\_test yang sedang diolah dalam label\_asli.
- 8) Evaluasi hasil klasifikasi:
  - a. Hitung akurasi: Akurasi mengukur sejauh mana model memprediksi dengan benar, yaitu seberapa banyak data\_test yang telah diprediksi dengan benar.
  - b. Hitung presisi: Presisi mengukur sejauh mana prediksi positif model adalah benar, yaitu seberapa banyak dari kelas positif yang benar-benar positif.
  - c. Hitung *recall*: *Recall* mengukur sejauh mana model dapat mendeteksi semua kasus positif, yaitu seberapa banyak data positif yang diprediksi dengan benar.
  - d. Hitung F1-score: F1-score adalah ukuran gabungan yang memperhitungkan baik presisi maupun *recall*.
- 9) Output hasil evaluasi:
  - a. Tampilkan metrik evaluasi seperti akurasi, presisi, recall, dan F1-score untuk mengevaluasi kinerja model *Naïve Bayes* dalam memprediksi kelas konsumsi energi listrik pada data\_test.

### III. METODE PENELITIAN

Penelitian ini menggunakan metode naive bayes untuk menganalisis pola konsumsi energi listrik pada gardu induk kelapa gading dan gardu induk sunter. Diagram tahapan penelitian dapat dilihat pada gambar 1.



**Gambar 1.** Diagram tahapan penelitian

Diagram pada gambar 1 menggambarkan alur kerja penelitian menggunakan metode naive bayes. Penelitian ini dimulai dengan pengumpulan dataset yang disusun dalam bentuk tabel, selanjutnya melakukan preprocessing data, lalu menghitung data uji dan data latih, setelah itu dilakukan klasifikasi menggunakan algoritma naive bayes, dan yang terakhir melakukan evaluasi. Penjelasan rinci dapat diuraikan sebagai berikut:

### 1) Dataset

Dataset merupakan kumpulan data terorganisir yang akan dianalisis dan diuji, dimana setiap kolom mewakili variabel dan setiap baris mempresentasikan jumlah data. Dalam penelitian ini, data yang digunakan diperoleh dari salah satu operator gardu induk kelapa gading dan gardu induk sunter. Data ini berupa laporan KWH meter yang mencakup informasi seperti gardu induk, pemilik, nama, stan awal dan stan akhir per hari pada bulan agustus, dengan jumlah data sebanyak 140. Dapat dilihat pada tabel berikut.

**Table 1.** Atribut Data

No	Atribut	Keterangan
1	Gardu Induk	Berisi nama gardu induk yang terdiri dari GI Kelapa Gading dan GI Sunter
2	Pemilik	Berisi nama pemilik yang terdiri atas Disjaya dan PLN
3	Nama	Berisi nama-nama penyulang (feeder) seperti moneter, market, bisnis, inflasi, kuota, dll.
4	Stan Awal	Berisi nilai stan awal sebanyak 140 data masing-masing 70 data tiap gardu induk
5	Stan Akhir	Berisi nilai stan akhir sebanyak 140 data masing-masing 70 data tiap gardu induk
6	Selisih Stan Awal-Stan Akhir	Berisi nilai selisih dari 140 data stan awal-stan akhir

### 2) Preprocessing data

*Preprocessing* data merupakan proses untuk membersihkan, mengorganisir, dan mempersiapkan data agar sesuai dengan kebutuhan analisis sebelum melakukan tahap klasifikasi. Dalam penelitian ini, tahap preprocessing data meliputi: organisir data, dimana data dibagi menjadi 2 bagian yaitu GI Kelapa Gading dan GI Sunter. Selanjutnya menyeleksi fitur yaitu gardu induk, pemilik, dan nama, serta merata-ratakan stan awal – stan akhir. Hasil preprocessing data dapat dilihat pada tabel 2.

**Table 2.** Perbandingan selisih stan awal dan stan akhir gardu induk

KELAPA GADING									
NO	Gardu Induk	Pemilik	Nama	Selisih KWH					
				1	2	3	4	Dst.	31
1	GI Kelapa Gading	Disjaya	Sakura/H	6,21	5,94	6,39	6,29	...	10,43
2	GI Kelapa Gading	Disjaya	Sakura/L1	10,25	10,27	10,28	10,21	...	14,58
3	GI Kelapa Gading	Disjaya	Sakura/L2	17,64	15,9	21,85	21,49	...	44,18
4	GI Kelapa Gading	Disjaya	Adenium/H	8,1	8,24	8,4	8,43	...	8,42
5	GI Kelapa Gading	Disjaya	Adenium/L1	14,67	14,73	14,89	15,03	...	15,34
...	...	...	...	...	...	...	...	...	...
70	GI Kelapa Gading	PLN	Trafo2/H	79,65	71,91	74,96	76,53	...	54,66
SUNTER									
1	GI Sunter	Disjaya	ekonomi / H	14,621	15,07	15,036	15,577		14,66
2	GI Sunter	Disjaya	ekonomi / L1	22,342	23,42	23,614	24,337		22,78
3	GI Sunter	Disjaya	ekonomi / L2	32,591	39,58	35,751	28,064		29,25

4	GI Sunter	Disjaya	ekspor /H	7,497	7,42	7,544	7,83	7,62
5	GI Sunter	Disjaya	ekspor /L1	9,061	9,261	9,128	9,351	9,22
...	...	...	...	...	...	...	...	...
70	GI Sunter	PLN	TrafoPS2/L2	44,784	70,288	58,811	61,663	46,00

### 3) Baca Data *Training* (Data Latih)

Tahap berikutnya adalah membaca data *training* atau biasa disebut data latih. Data yang akan digunakan sebanyak 60% dari 140 data. Proses data *training* melibatkan perhitungan probabilitas kondisional pada setiap atribut dataset, berdasarkan kelas target. Proses ini akan dijalankan melalui implementasi algoritma naïve bayes menggunakan bahasa pemrograman python.

### 4) Data Uji

Tahap berikutnya adalah membaca data uji. Data yang akan digunakan sebanyak 40% dari 140 data. Data tersebut akan dimasukkan kedalam model yang nantinya akan melakukan klasifikasi data baru berdasarkan probabilitas kondisional yang telah dihitung.

### 5) Klasifikasi *Naïve Bayes*

Tahap selanjutnya yaitu klasifikasi menggunakan algoritma naïve bayes. Model klasifikasi ini dapat dilihat pada tabel 3. Hasil klasifikasi akan digunakan untuk mengevaluasi kinerja model naïve bayes.

**Table 3.** Model Klasifikasi KWH Meter

KWH	Level
0-20	Rendah
21-70	Normal
>70	Tinggi

Klasifikasi pada penelitian ini menggunakan 3 kondisi yaitu penggunaan listrik rendah, normal dan tinggi, berdasarkan perhitungan rata-rata dari data selisih KWH per hari pada bulan agustus. Klasifikasi tersebut dapat dilihat pada tabel 4.

**Table 4.** Klasifikasi Laporan KWH Meter

KELAPA GADING								
NO	Gardu Induk	Pemilik	Nama	Selisih KWH				
				1	Dst.	31	Rata-Rata	Peggunaan
1	GI Kelapa Gading	Disjaya	Sakura/H	6,21	...	8.10	6,57	Rendah
2	GI Kelapa Gading	Disjaya	Sakura/L1	10,25	...	14.26	14,09	Rendah
3	GI Kelapa Gading	Disjaya	Sakura/L2	17,64	...	30.95	26,37	Normal
4	GI Kelapa Gading	Disjaya	Adenium/H	8,1	...	7.86	8,63	Rendah
5	GI Kelapa Gading	Disjaya	Adenium/L1	14,67	...	13.61	15,31	Rendah
...	...	...	...	...	...	...	...	...
42	Gi Kelapa Gading	Disjaya	Aster / L2	10.93	...	19.26	13.818	Rendah
SUNTER								
1	GI Sunter	Disjaya	ekonomi / H	14,621		14,66	14.659	Rendah
2	GI Sunter	Disjaya	ekonomi /L1	22,342		22,78	23.019	Normal
3	GI Sunter	Disjaya	ekonomi /L2	32,591		29.25	35.850	Normal
4	GI Sunter	Disjaya	ekspor /H	7,497		7.62	7.549	Rendah
5	GI Sunter	Disjaya	ekspor /L1	9,061		9.22	9.184	Rendah
...	...	...	...	...		..	...	...
42	Gi Sunter	Disjaya	Niaga / L2	29.474	...	27.48	34.959	Normal



## 6) Evaluasi

Pada tahap ini, performa model naïve bayes akan dievaluasi menggunakan metrik yang relevan seperti akurasi, presisi, recall atau F1-score. Proses evaluasi dilakukan dengan membandingkan hasil klasifikasi model dengan label sebenarnya pada data uji. Evaluasi ini bertujuan untuk menilai sejauh mana kemampuan naïve bayes yang dijalankan menggunakan python [22].

## IV. HASIL DAN PEMBAHASAN

Analisis pola konsumsi energi listrik pada gardu induk kelapa gading dan gardu induk sunter menggunakan metode *Naïve Bayes* di implementasikan menggunakan jupyter dengan bahasa pemrograman python. Dalam melakukan prediksi, mesin belum dapat memahami data target yang bersifat kategorikal, seperti rendah, normal, tinggi. Mesin hanya dapat memahami data dalam bentuk angka. Oleh karena itu, data diubah menjadi:

**Table 5.** Variabel data target

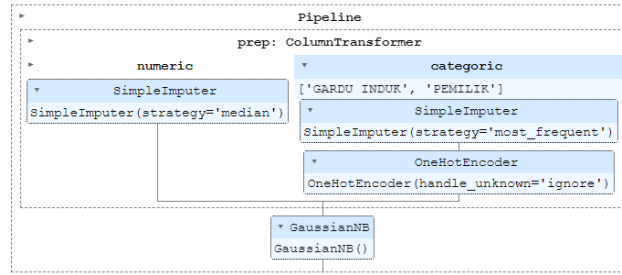
Variabel data target	Pengkodean Integer
Rendah	0
Normal	1
Tinggi	2

KELAPA GADING								
NO	Gardu Induk	Pemilik	Nama	Selisih KWH				
				1	Dst.	31	Rata-Rata	Peggunaan
1	GI Kelapa Gading	Disjaya	Sakura/H	6,21	...	8.10	6,57	0
2	GI Kelapa Gading	Disjaya	Sakura/L1	10,25	...	14.26	14,09	0
3	GI Kelapa Gading	Disjaya	Sakura/L2	17,64	...	30.95	26,37	1
4	GI Kelapa Gading	Disjaya	Adenium/H	8,1	...	7.86	8,63	0
5	GI Kelapa Gading	Disjaya	Adenium/L1	14,67	...	13.61	15,31	0
...	...	...	...	...	...	...	...	...
42	Gi Kelapa Gading	Disjaya	Aster / L2	10.93	...	19.26	13.818	0
SUNTER								
1	GI Sunter	Disjaya	ekonomi / H	14,621		14,66	14.659	0
2	GI Sunter	Disjaya	ekonomi /L1	22,342		22,78	23.019	1
3	GI Sunter	Disjaya	ekonomi /L2	32,591		29.25	35.850	1
4	GI Sunter	Disjaya	ekspor /H	7,497		7.62	7.549	0
5	GI Sunter	Disjaya	ekspor /L1	9,061		9.22	9.184	0
...	...	...	...	...		..	...	...
42	Gi Sunter	Disjaya	Niaga / L2	29.474	...	27.48	34.959	1

Data tersebut kemudian di simpan dalam format CSV. Selanjutnya, dilakukan *dataset splitting* atau biasa disebut pembagian data. Data ini dibagi menjadi data latih (*training*) dan data uji (*testing*) dengan perbandingan 60:40, dimana 60% menjadi data *training* dan sisanya sebagai data *testing*. Pendistribusian data dapat dilihat pada Tabel 6.

**Table 6.** Distribusi Data Training dan Data Testing

	Data training	Data testing
Jumlah data	84	56
Jumlah kolom	35	35



**Gambar 2.** Komposisi Data

Selanjutnya, dilakukan proses pengolahan data (processing data). Dalam hal ini, terdapat dua jenis transformasi yaitu numerik dan kategorikal, dapat dilihat pada gambar 2. Data numerik berisi fitur yang memiliki nilai numerik, seperti 1-31, dan rata-rata. Sementara itu, data kategorikal berisi fitur yang bersifat kategorikal seperti, gardu induk dan pemilik, yang kemudian diolah menggunakan teknik *one-hot encoding*.

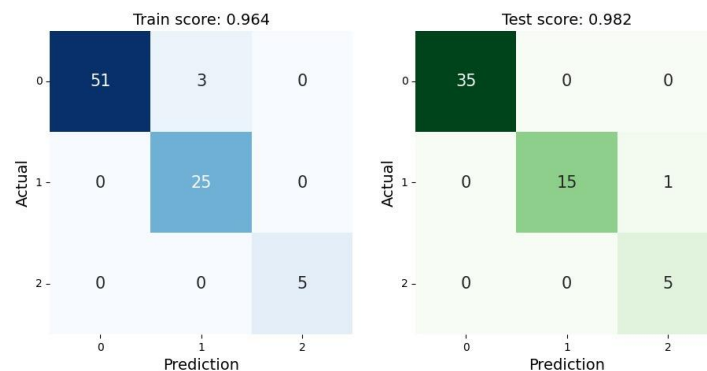
Hasil prediksi penggunaan listrik pada gardu induk kelapa gading dan gardu induk sunter menggunakan metode *Naïve Bayes* dapat dilihat pada tabel berikut.

**Tabel 7.** Hasil prediksi menggunakan data uji

GARDU INDUK	PEMILIK	NAMA	1	Dst.	31	RATA- RATA	PREDIKSI_ PENGUNAAN	TARGET_ PENGUNAAN
GI SUNTER	DISJAYA	LOGISTIK / L2	36.17	...	32.95	34.77	1	1
GI KELAPA GADING	PLN	TRAFO 2 / L2	224.82	..	173.35	165.68	2	2
GI KELAPA GADING	DISJAYA	KAKTUS / L1	12.72	..	9.76	-226.10	0	0
GI SUNTER	PLN	TRAFO 1 / L1	31.56	..	33.50	33.62	1	1
GI KELAPA GADING	DISJAYA	SEROJA / H	6.47	..	6.97	6.37	0	0
Dst..	..	..	..	..	..	..	..	..
GI SUNTER	DISJAYA	INVESTOR / H	0		0	0	0	0

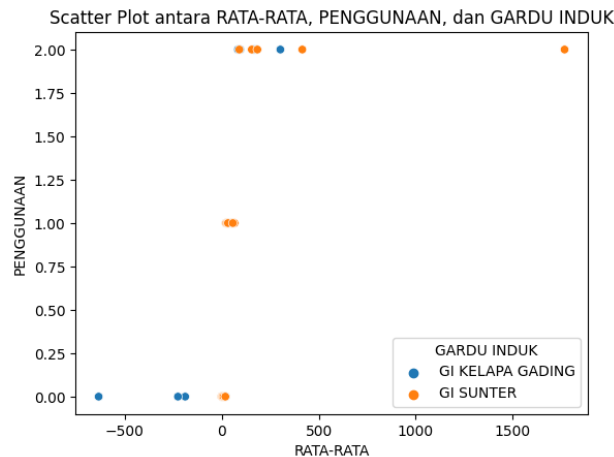
Berdasarkan data yang ditampilkan, nilai PREDIKSI\_PENGUNAAN dan TARGET\_PENGUNAAN telah sesuai dengan label sebenarnya. Oleh karena itu, hasil prediksi dapat dinyatakan telah akurat. Akurasi yang dihasilkan untuk data *training* sebesar 96.428%. Sementara akurasi yang dihasilkan untuk data *testing* sebesar 98.214%.

Konfusi (Confusion matrix) dapat dilihat pada gambar berikut.



**Gambar 3.** Confusion matrix

Dapat dilihat, Gambar 3 menunjukkan matriks konfusi pada data latih dengan skor 0.964 dan matriks konfusi pada data uji dengan skor 0.982. Terdapat tiga kelas yang digunakan dalam matriks konfusi ini, yaitu 0 untuk rendah, 1 untuk normal, dan 2 untuk tinggi. Pada matriks konfusi data latih, terdapat 84 data. Pada kelas rendah, terdapat 54 data, di mana 51 data terklasifikasi dengan benar dan 3 data keliru. Kemudian, pada kelas normal, terdapat 25 data yang terklasifikasi dengan benar, sementara pada kelas tinggi terdapat 5 data yang terklasifikasi dengan benar. Selanjutnya, pada matriks konfusi data uji, terdapat 56 data. Pada kelas rendah, terdapat 35 data yang terklasifikasi dengan benar. Kemudian, pada kelas normal, terdapat 16 data, di mana 15 data terklasifikasi dengan benar dan 1 data keliru. Sementara itu, pada kelas tinggi terdapat 5 data yang terklasifikasi dengan benar.



## VI. REFERENSI

- [1] A. M. Mappalotteng *et al.*, “Analisis Beban Pemakaian Daya Listrik Di Kabupaten Bantaeng Selama Masa Pandemi Covid-19 Electricity Load Analysis in Bantaeng Regency During the Covid-19 Pandemic,” vol. 19, no. 2, pp. 81–85, 2022.
- [2] G. Melinda and M. Longgom Nasution, “Peramalan Jumlah Konsumsi Energi Listrik di PT PLN (Persero) Rayon Bukittinggi Menggunakan Metode Arima,” *J. Math. UNP*, vol. 4, no. 2, p. 42, 2019.
- [3] SolarKita, “Kondisi Distribusi Listrik di Indonesia, Apakah Sudah Merata?,” *Kumparan*. 2021. [Online]. Available: <https://kumparan.com/solar-kita/kondisi-distribusi-listrik-di-indonesia-apakah-sudah-merata-1wOmOPmMJqk/full>
- [4] “12 Komponen Dan Peralatan Gardu Induk Beserta Fungsinya - Cara Ilmu.” [Online]. Available: <https://www.carailmu.com/2021/06/komponen-peralatan-gardu-induk.html>
- [5] A. Ahdiat, “Konsumsi listrik penduduk Indonesia naik pada 2022, Capai rekor baru,” *Databoks*, vol. 2021. p. 2023, 2023. [Online]. Available: <https://databoks.katadata.co.id/datapublish/2023/02/23/konsumsi-listrik-penduduk-indonesia-naik-pada-2022-capai-rekor-baru>
- [6] D. Dahri, F. Agus, and D. M. Khairina, “Metode Naive Bayes Untuk Penentuan Penerima Beasiswa Bidikmisi Universitas Mulawarman,” *Inform. Mulawarman J. Ilm. Ilmu Komput.*, vol. 11, no. 2, p. 29, 2016, doi: 10.30872/jim.v11i2.211.
- [7] D. Y. Fallo, “Metode Naïve Bayes Untuk Memprediksi Penggunaan Listrik Rumah Tangga,” *J. Pendidik. Teknol. Inf.*, vol. 1, no. 2, pp. 47–51, 2018, doi: 10.37792/jukanti.v1i2.11.
- [8] H. Tuhuteru and A. Iriani, “Analisis Sentimen Perusahaan Listrik Negara Cabang Ambon Menggunakan Metode Support Vector Machine dan Naive Bayes Classifier,” *J. Inform. J. Pengemb. IT*, vol. 3, no. 3, pp. 394–401, 2018, doi: 10.30591/jpit.v3i3.977.
- [9] Sri Diantika, Windu Gata, and Hiya Nalatissifa, “Komparasi Algoritma SVM Dan Naive Bayes Untuk Klasifikasi Kestabilan Jaringan Listrik,” *Elkom J. Elektron. dan Komput.*, vol. 14, no. 1, pp. 10–15, 2021, doi: 10.51903/elkom.v14i1.319.
- [10] F. V. Sari and A. Wibowo, “Analisis Sentimen Pelanggan Toko Online Jd.Id Menggunakan Metode Naïve Bayes Classifier Berbasis Konversi Ikon Emosi,” *J. SIMETRIS*, vol. 10, no. 2, pp. 681–686, 2019.
- [11] S. Yuwono, D. Diharto, and N. W. Pratama, “Manfaat Pengadaan Panel Surya dengan Menggunakan Metode On Grid,” *Energi & Kelistrikan*, vol. 13, no. 2, pp. 161–171, 2021, doi: 10.33322/energi.v13i2.1537.
- [12] “kWh (Kilowatt hour) - Definition, Formula and Calculation.”
- [13] W. W. W. Muttaqin, S. R. P. Muhammad Munsarif Green Ferry Mandias, S. A. H. A. F. Agung Widarman Wiranti Kusuma Hapsari, E. F. B. M. A. Pasnur and N. S. Suryani, *Pengenalan Data Mining*, Medan: Yayasan Kita Menulis, 2023.
- [14] Rayuwati, Husna Gemasih, and Irma Nizar, “IMPLEMENTASI ALGORITMA NAIVE BAYES UNTUK MEMPREDIKSI TINGKAT PENYEBARAN COVID,” *Jural Ris. Rumpun Ilmu Tek.*, vol. 1, no. 1, pp. 38–46, 2022, doi: 10.55606/jurritek.v1i1.127.
- [15] M. E. Lasulika, “Komparasi Naïve Bayes, Support Vector Machine Dan K-Nearest Neighbor Untuk Mengetahui Akurasi Tertinggi Pada Prediksi Kelancaran Pembayaran Tv Kabel,” *Ilk. J. Ilm.*, vol. 11, no. 1, pp. 11–16, 2019, doi: 10.33096/ilkom.v11i1.408.11-16.
- [16] S. Lestari, A. Akmaludin, and M. Badrul, “Implementasi Klasifikasi Naive Bayes Untuk Prediksi Kelayakan Pemberian Pinjaman Pada Koperasi Anugerah Bintang Cemerlang,” *PROSISKO J. Pengemb. Ris. dan Obs. Sist. Komput.*, vol. 7, no. 1, pp. 8–16, 2020, doi: 10.30656/prosisko.v7i1.2129.
- [17] B. P. Much Aziz Muslim, A. J. H. Eva Laily Harum Mawarni, S. H. R. Mirqotussa'adah and A. Nurzahputra, “Data Mining,” in *Algoritma C4.5*, Semarang, Much Aziz Muslim, 2019, p. 32.
- [18] D. I. Mulyana, H. H. Hartanto, and M. B. Yel, “Optimasi Klasifikasi Bunga Kantong Semar Dengan Menggunakan Algoritma Naïve Bayes, Data Augmentasi Dan PSO,” *Swabumi*, vol. 10, no. 2, pp. 167–174, 2022, doi: 10.31294/swabumi.v10i2.13620.
- [19] G. P. Kawani, “Implementasi Naive Bayes Untuk Menentukan Wadah Limbah B3 Sesuai Karakteristik,” *J. INISTA*, vol. 1, no. 2, pp. 73–81, 2019.
- [20] F. Fatmawati and N. Narti, “Perbandingan Algoritma C4.5 dan Naive Bayes Dalam Klasifikasi Tingkat Kepuasan Mahasiswa Terhadap Pembelajaran Daring,” *JTIM J. Teknol. Inf. dan Multimed.*, vol. 4, no. 1, pp. 1–12, 2022, doi: 10.35746/jtim.v4i1.196.
- [21] M. F. A. Saputra, T. Widiyaningtyas, and A. P. Wibawa, “Illiteracy classification using K means-naïve bayes algorithm,” *Int. J. Informatics Vis.*, vol. 2, no. 3, pp. 153–158, 2018, doi: 10.30630/joiv.2.3.129.
- [22] G. P. Kawani, “Implementasi Naive Bayes,” *J. Informatics, Inf. Syst. Softw. Eng. Appl.*, vol. 1, no. 2, pp. 73–81, 2019, doi: 10.20895/inista.v1i2.73.