



Klasifikasi Penggunaan Daya Listrik Rumah Tangga dengan Menggunakan Metode *Naive Bayes*

An Nisa Ziah Putri^{1*}, Dodo Zaenal Abidin², Errissya Rasywir³, Ibni Faiq Athallah⁴

^{1-2,4}Departemen Teknik Informatika, Fakultas Ilmu Komputer, Universitas Dinamika Bangsa, Indonesia

³Departemen Sistem Komputer, Fakultas Ilmu Komputer, Universitas Dinamika Bangsa, Indonesia

Email: annisaziahputri18@gmail.com^{1*}, dodozaenalabidin@gmail.com², errissya.rasywir@gmail.com³, iibni225@gmail.com⁴

*Penulis Korespondensi: annisaziahputri18@gmail.com

Abstract. *Data mining is a technique of several fields of science to find previously unknown relationships in the data warehouse so that it becomes an information that can be used later. The unwise use of electricity will of course have an impact on the high use of electricity, therefore it is expected that every community understands the effort to use electricity wisely. Therefore, authors perform analysis of data mining on these electrical usage data in order to know which is a small, medium and large category. The authors use data on electrical use questionnaire as much as 200 data which is then presented into the ARFF format. In performing author analysis using WEKA Tools. The method used is Naive Bayes classification method with the greatest percentage of accuracy obtained using the Use Training Set Correctly of 80.5%, using a 5-Fold Cross Validation Correctly of 75%, and using 10-Fold Cross Validation amounted to 74%. While the result of the selection of the attributes using the algorithm classifier attribute evaluation (ClassifierAttributeEval) is stated that the most influential attribute against the electrical power usage classification is Electronic Goods.*

Keywords: *Data Mining, Classification; Naive Bayes; Power Use; Weka*

Abstrak. *Data mining merupakan sebuah teknik dari beberapa bidang ilmu untuk menemukan hubungan yang tidak diketahui sebelumnya di dalam gudang data sehingga menjadi suatu informasi yang dapat digunakan kemudian. Penggunaan listrik yang tidak bijak tentu saja akan berdampak pada tingginya penggunaan listrik, untuk itu diharapkan setiap masyarakat memahami upaya dalam menggunakan listrik dengan bijak. Oleh karena itu penulis melakukan analisis data mining pada data-data penggunaan listrik tersebut agar dapat mengetahui mana daya listrik dengan kategori kecil, sedang dan besar. Penulis menggunakan data-data kuesioner penggunaan listrik sebanyak 200 data yang kemudian disajikan kedalam format arff. Dalam melakukan analisis penulis menggunakan alat bantu Tools WEKA. Metode yang di gunakan adalah metode klasifikasi *Naive Bayes* dengan persentasi akurasi terbesar diperoleh dengan menggunakan *Use Training Set Correctly* yaitu sebesar 80,5% , menggunakan *5-Fold Cross Validation Correctly* sebesar 75%, dan menggunakan *10-Fold Cross Validation* sebesar 74%. Sedangkan hasil seleksi atribut menggunakan algoritma *classifier attribute evaluation (ClassifierAttributeEval)* dinyatakan bahwa atribut yang paling berpengaruh terhadap klasifikasi penggunaan daya listrik adalah Barang-barang elektronik.*

Kata kunci: *Data Mining, Klasifikasi; Naive Bayes; Penggunaan Daya Listrik; Weka*

1. LATAR BELAKANG

Data Mining merupakan proses pengekstraksian informasi dari sekumpulan data yang sangat besar melalui penggunaan algoritma dan teknik penarikan dalam bidang statistik, pembelajaran mesin dan sistem manajemen basis data. Sedangkan *Naive Bayes* merupakan pengklasifikasian dengan metode probabilitas dan statistik yang dikemukakan oleh ilmuwan

Inggris Thomas Bayes, yaitu memprediksi peluang di masa depan berdasarkan pengalaman dimasa sebelumnya (Yos et al., 2016).

Listrik merupakan energi vital bagi keberlangsungan aktivitas manusia baik bagi individu, kelompok masyarakat, maupun dunia industri. Listrik telah menjadi kebutuhan pokok bagi masyarakat, baik dalam bidang industri maupun rumah tangga, karena berperan penting dalam menunjang aktivitas sosial dan ekonomi masyarakat (Kementerian Energi dan Sumber Daya Mineral Republik Indonesia, 2019).

Penggunaan energi listrik di rumah tangga umumnya dimanfaatkan untuk berbagai kebutuhan, seperti penerangan, alat setrika, hiburan, kipas angin, lemari es, serta pendingin ruangan (Sari et al., 2023). Penggunaan listrik yang tidak bijak dapat berdampak pada meningkatnya konsumsi energi listrik, yang pada akhirnya berkontribusi terhadap menipisnya ketersediaan energi listrik akibat kebutuhan yang lebih besar dibandingkan dengan persediaannya. Oleh karena itu, masyarakat diharapkan memiliki pemahaman mengenai upaya penggunaan listrik secara bijak dan efisien (Firdaus & Karim, 2025). Selain itu, setiap rumah tangga juga perlu mengetahui besarnya penggunaan listrik masing-masing agar dapat mengontrol dan menghemat konsumsi energi listrik secara efektif (Firdaus & Karim, 2025).

Dari 60 data penggunaan listrik rumah tangga yang diuji dengan metode naïve bayes, maka diperoleh hasil persentase 78,3333% untuk keakuratan prediksi, di mana dari 60 data penggunaan listrik rumah tangga yang diuji terdapat 47 data penggunaan listrik rumah tangga yang berhasil diklasifikasikan dengan benar. Untuk itu peneliti menggunakan metode Naïve Bayes untuk mengklasifikasi biaya penggunaan listrik rumah tangga dengan kriteria yang digunakan diantaranya Jumlah tanggungan keluarga, pendapatan per bulan, daerah tempat tinggal, luas rumah, industri rumah tangga, jenis rumah, daya listrik rumah tangga, dan barang-barang elektronik.

Hal inilah yang melatar belakangi penulis untuk melakukan penelitian guna memberi solusi terhadap masalah yang terjadi dengan mengangkat judul “**Klasifikasi Penggunaan Daya Listrik Rumah Tangga Dengan Metode Naïve Bayes**”. Hal ini peneliti lakukan untuk mengklasifikasikan penggunaan daya listrik tiap rumah tangga, apakah termasuk dalam kategori tinggi, sedang, atau rendah, agar Pengguna listrik dapat lebih bijak dalam menggunakan listrik.

2. KAJIAN TEORITIS

Data mining bukanlah suatu bidang yang sepenuhnya baru, melainkan berkembang dari berbagai disiplin ilmu yang telah mapan, seperti statistik, pembelajaran mesin, dan basis data. Data mining mengintegrasikan berbagai teknik untuk mengekstraksi informasi dan pola yang bermakna dari data dalam jumlah besar (Kotu & Deshpande, 2019).

Kebutuhan terhadap data mining muncul seiring dengan meningkatnya volume data yang tersimpan dalam basis data, yang berpotensi menghasilkan informasi dan pengetahuan yang berguna untuk mendukung pengambilan keputusan di berbagai bidang, seperti manajemen bisnis, kontrol produksi, dan kesehatan (Aggarwal, 2015).

Data mining merupakan bidang multidisiplin yang memadukan teknik dari pembelajaran mesin, pengenalan pola, statistik, dan sistem basis data untuk menemukan hubungan atau pola yang sebelumnya tidak diketahui dari kumpulan data berukuran besar (International Energy Agency, 2020). Klasifikasi dapat didefinisikan sebagai pekerjaan yang melakukan pelatihan/pembelajaran terhadap fungsi target f yang memetakan setiap set atribut (*fitur*) x kesatu dari sejumlah label kelas y yang tersedia (Aggarwal, 2015).

Klasifikasi didefinisikan sebagai bentuk analisis data untuk mengekstrak model yang akan digunakan untuk memprediksi label kelas (Aggarwal, 2015). Dari beberapa definisi diatas maka dapat disimpulkan bahwa klasifikasi merupakan pengelompokkan sampel berdasarkan ciri-ciri persamaan dan perbedaan dengan menggunakan variabel target sebagai kategori.

Bayesian classification adalah pengklasifikasian statistik yang dapat digunakan untuk memprediksi probabilitas keanggotaan suatu class. *Bayesian Classification* didasarkan pada teorema Bayes yang memiliki kemampuan klasifikasi serupa dengan *decision tree* dan *neural network*. *Bayesian Classification* terbukti memiliki akurasi dan kecepatan yang tinggi saat diaplikasikan ke dalam database dengan data yang besar (Aggarwal, 2015). Pada tahapan pengklasifikasian akan dihitung nilai probabilitas dari masing-masing label kelas yang ada terhadap masukan yang diberikan. Label kelas yang memiliki nilai probabilitas paling besar akan dijadikan sebagai label kelas data masukan tersebut. *Naïve Bayes* merupakan perhitungan teorema *Bayes* yang paling sederhana karena mampu mengurangi kompleksitas komputasi menjadi multiplikasi sederhana dari probabilitas. Selain itu, algoritma *Naïve Bayes* juga mampu menangani set data yang memiliki banyak atribut.

Persamaan dari *Naive Bayes* sebagai berikut:

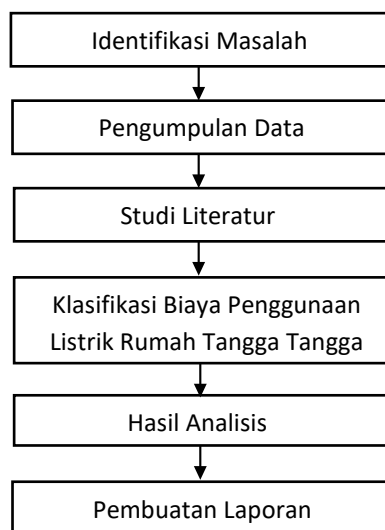
$$P(C_i | X) = \frac{P(X | C_i) P(C_i)}{P(X)}$$

Keterangan :

- X : Kriteria suatu kasus berdasarkan masukan
- C_i : Kelas solusi pola ke-i, dimana i adalah jumlah label kelas
- P(C_i|X) : Probabilitas kemunculan label kelas C_i dengan kriteria masukan X
- P(X|C_i) : Probabilitas kriteria masukan X dengan label kelas C_i
- P(C_i) : Probabilitas label kelas C_i

3. METODE PENELITIAN

Kerangka kerja penelitian merupakan tahapan-tahapan yang dilakukan selama mengerjakan penelitian. Kerangka kerja penelitian dibuat agar mempermudah pencapaian hasil penelitian, dapat menyelesaikan penelitian tepat waktu dan penelitian dapat berjalan sesuai dengan yang diharapkan. Adapun kerangka kerja penelitian yang digunakan dapat dilihat pada gambar 1.



Gambar 1. Kerangka Kerja Penelitian.

Berdasarkan gambar kerangka kerja penelitian diatas, dapat diuraikan sebagai berikut:

A. Identifikasi Masalah

Pada tahap ini penulis mengidentifikasi masalah mengenai penggunaan listrik rumah tangga (Kementerian Energi dan Sumber Daya Mineral Republik Indonesia, 2019). Listrik merupakan salah satu bentuk energi yang banyak dibutuhkan (International

Energy Agency, 2020). Listrik juga menjadi kebutuhan pokok bagi masyarakat baik dalam bidang industri maupun rumah tangga. Penggunaan alat-alat listrik memerlukan arus listrik yang dihasilkan dari sumber energi (Suganthi & Samuel, 2019). Penggunaan listrik yang tidak bijak tentu saja akan berdampak pada tingginya penggunaan daya listrik. Setiap rumah tangga juga harus tahu besarnya biaya penggunaan listrik rumah tangga masing-masing. Dalam hal mengetahui besarnya penggunaan daya listrik ini maka diterapkan teknik klasifikasi data mining dengan metode *naïve bayes* (Aggarwal, 2015).

B. Pengumpulan Data

Pada tahap ini penulis melakukan pengumpulan data dan informasi yang dibutuhkan untuk mengklasifikasikan penggunaan daya listrik rumah tangga menggunakan metode *naïve bayes*. Pengumpulan data ini dilakukan dengan beberapa metode yaitu:

1) Angket (*Questioner*)

Metode pengumpulan data yang dilakukan dengan cara memberi seperangkat pertanyaan atau pernyataan tertulis kepada responden untuk dijawab (Creswell & Creswell, 2018). Dalam hal ini penulis menggunakan kuesioner tertutup yang dimana kuesioner ini merupakan kuesioner klasifikasi penggunaan daya listrik rumah tangga menggunakan metode *naïve bayes* yang telah menyediakan pilihan jawaban untuk dipilih oleh objek penelitian. Adapun data kuesioner di sebar dengan dua cara:

- a) *Online*, yaitu menggunakan google form sebagai media untuk menyebarkan kuesioner.
- b) *Offline*, yaitu dengan cara turun langsung ke lapangan untuk menyebarkan kuesioner.

Atribut dan pertanyaan yang digunakan dalam kuesioner diantaranya jumlah tanggungan keluarga, pendapatan per bulan, daerah tempat tinggal, luas rumah, industri rumah tangga, jenis rumah, daya listrik rumah tangga dan barang-barang elektronik.

2) Penelitian Kepustakaan (*Library Research*)

Untuk menunjang penelitian yang akan dilakukan, penulis melakukan *library research* yakni dengan mencari data-data dari buku maupun jurnal penelitian sejenis yang berhubungan dengan metode yang penulis gunakan dalam penelitian ini yaitu metode *Naive Bayes*.

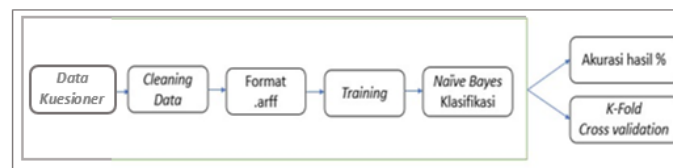
C. Studi Litelatur

Pada tahap ini penulis melakukan pencarian terhadap landasan-landasan teori yang diperoleh dari berbagai sumber seperti buku, jurnal ilmiah dan juga referensi lainnya untuk melengkapi penelitian baik mengenai konsep dan teori sehingga memiliki acuan yang baik dan relevan, antara lain (Boote & Beile, 2016):

- a. Implementasi metode klasifikasi *naive bayes* dalam memprediksi besarnya penggunaan listrik rumah tangga.
- b. Penerapan data mining dengan metode *naive bayes* untuk memprediksi kelulusan mahasiswa.
- c. Evaluasi kinerja akademik mahasiswa menggunakan algoritma *naive bayes*.
- d. Upaya menghemat pemakaian energi listrik.

D. Klasifikasi Penggunaan Daya Listrik

Pengklasifikasian penggunaan daya listrik dilakukan dengan beberapa tahapan. Adapun alur proses pengklasifikasian dapat dilihat pada gambar 2.



Gambar 2. Alur proses klasifikasi dan analisis *Naive Bayes*.

Berikut ini penjelasan mengenai tahapan-tahapan pada alur proses pengklasifikasian :

1) Data kuesioner

Data kuesioner merupakan data-data yang didapatkan dengan memberikan beberapa pertanyaan di dalamnya serta bersifat tertutup, yang kemudian ditujukan kepada masyarakat. Data yang dimaksud meliputi beberapa atribut yaitu jumlah tanggungan keluarga, jumlah tanggungan keluarga, pendapatan per bulan, daerah

tempat tinggal, luas rumah, industri rumah tangga, jenis rumah, daya listrik rumah tangga dan barang-barang elektronik.

2) *Cleaning Data*

Cleaning data merupakan preproses dimana pada tahap ini dilakukan pembersihan atau pemisahan data dari *noise* data, data yang tidak akurat, atau data yang salah guna menghasilkan data berkualitas (Kotu & Deshpande, 2019). Selain itu ada beberapa hal yang penulis lakukan dalam proses *cleaning* data yaitu sebagai berikut:

- a) Mengklasifikasikan kenaikan penggunaan listrik menjadi 3 kategori yaitu Kecil, Sedang, dan Besar.
- b) Kemudian hasil dari *cleaning* data ini disusun dalam bentuk format arff pada tools Notepad++.

3) *Format .arff*

WEKA mendukung beberapa format file dalam input-nya diantaranya adalah format arff (Witten et al., 2016). *Attribute-Relation File Format* (ARFF) merupakan tipe *file text* yang berisi berbagai *instance* data yang berhubungan dengan suatu set atribut data. kemudian data-data ini disusun dengan menggunakan tanda koma (,) sebagai pemisah data per-atributnya.

4) *Training*

Pada tahap ini dilakukan proses training oleh algoritma klasifikasi *naïve bayes* untuk membentuk sebuah model *classifier* pada data kuesioner yang sudah ada. Model ini merupakan representasi pengetahuan yang akan digunakan untuk prediksi kelas data baru yang belum pernah ada (Aggarwal, 2015).

5) *Naïve Bayes* Klasifikasi

Pada tahapan ini, dilakukan proses pengklasifikasian data menggunakan tools WEKA terhadap data kuesioner yang sudah ditesting menggunakan model *classifier* yang sudah dibentuk.

6) *Akurasi Hasil*

Akurasi Hasil menunjukkan kedekatan hasil pengukuran dengan nilai sesungguhnya.

7) *K-Fold Cross Validation*

Cross Validation merupakan salah satu teknik untuk menilai/memvalidasi keakuratan sebuah model yang dibangun berdasarkan dataset tertentu [13]. Salah satu metode *cross-validation* yang populer adalah *K-Fold Cross Validation*.

E. Hasil Analisis

Hasil dari analisis akan dinilai keakuratan data model yang sudah di dapatkan dengan perhitungan *naïve bayes* menggunakan teknik *Cross Validation*. Salah satu metode *cross-validation* yang populer dan akan digunakan adalah *K-Fold Cross Validation*. Dalam teknik ini dataset dibagi menjadi sejumlah K-buah partisi secara acak (Hastie et al., 2022). Kemudian dilakukan sejumlah K-kali eksperimen, dimana masing-masing eksperimen menggunakan data partisi ke-K sebagai data testing dan memanfaatkan sisa partisi lainnya sebagai data training. Dimana hasil dengan persentasi terbesar yang akan di ambil. Selain itu, bentuk visualisasi dari setiap atribut dapat dilihat pada *tools WEKA*.

F. Pembuatan Laporan

Setelah semua tahapan penelitian dilakukan, maka akan dibuat laporan sebagai dokumentasi penelitian agar dapat dimanfaatkan pada waktu yang akan datang. Baik oleh peneliti sendiri maupun peneliti lainnya.

4. HASIL DAN PEMBAHASAN

Berdasarkan hasil kuesioner yang sudah didapatkan, penulis memperoleh data-data penggunaan listrik sebanyak 200. Data kuesioner penggunaan listrik masing-masing dibagi menjadi 3 kelas yaitu sebanyak 148 kelas kecil, 46 kelas sedang, 6 kelas besar. Atribut yang digunakan pada seluruh data kuesioner penggunaan listrik yaitu, usia, jenis kelamin, jumlah tanggungan keluarga, status perkawinan kepala keluarga, pendapatan per bulan, pengeluaran per bulan, status tinggal, luas rumah, tipe rumah, industri rumah tangga, jenis rumah, pajak pbb dan daya listrik rumah tangga. Berikut adalah tabel data training yang akan digunakan untuk perhitungan *naïve bayes*.

Tabel 1. Data Training.

No	Usia	Jenis Kelamin	Jumlah Tanggungan Keluarga	Pendapatan per Bulan	Luas Rumah	Industri Rumah Tangga	Jenis Rumah	Daya Listrik Rumah Tangga
16	Remaja Akhir	L	Kecil	Sangat_Kecil	Sangat_Kecil	Tidak_Ada	Permanen	Kecil
17	Remaja Akhir	L	Kecil	Sangat_Kecil	Sedang	Tidak_Ada	Semi Permanen	Sedang
18	Remaja Akhir	P	Kecil	Besar	Sedang	Tidak_Ada	Semi Permanen	Sedang
19	Remaja Akhir	L	Sedang	Standar	Sedang	Tidak_Ada	Permanen	Kecil
20	Remaja Akhir	L	Kecil	Kecil	Sedang	Tidak_Ada	Permanen	Kecil
21	Remaja Akhir	P	Kecil	Besar	Kecil	Tidak_Ada	Permanen	Sedang
22	Remaja Akhir	P	Kecil	Kecil	Sedang	Tidak_Ada	Permanen	Sedang
23	Remaja Akhir	P	Sangat_Kecil	Sangat_Kecil	Sedang	Tidak_Ada	Permanen	Kecil
24	Remaja Akhir	P	Sedang	Standar	Sedang	Tidak_Ada	Semi Permanen	Sedang
25	Dewasa Awal	P	Sangat_Kecil	Sangat_Kecil	Sangat_Kecil	Ada	Semi Permanen	Kecil

Data training ini sudah melalui tahap *cleaning* data yang kemudian disusun sesuai kebutuhan dalam format excel, *Cleaning* data merupakan preproses dimana pada tahap ini dilakukan pembersihan atau pemisahan data dari *noise* data dan data yang tidak konsisten seperti data kuesioner penggunaan listrik yang tidak lengkap maka akan di hapus setelah itu data kuesioner penggunaan listrik disusun dalam format arff. Berikut data kuesioner penggunaan listrik dalam format arff :

```

1 @relation, penggunaanlistrik
2
3 @attribute Usia (Remaja_Awal, Remaja_Akhir, Dewasa_Awal, Lansia)
4 @attribute Jenis_Kelamin (L, P)
5 @attribute Jumlah_Tanggungan_Keluarga (Sangat_Kecil, Kecil, Sedang, Banyak, Sangat_Banyak)
6 @attribute Pendapatan_per_Bulan (Sangat_Kecil, Kecil, Standar, Besar, Sangat_Besar)
7 @attribute Daerah_Tempat_Tinggal (Pusat_Kota_Jambi, Desa_atau_Perkampungan, Pinggiran_Kota_Jambi, Luar_Kota_Jambi, Luar_Provinsi_Jambi)
8 @attribute Luas_Rumah (Sangat_Kecil, Kecil, Sedang, Besar, Sangat_Besar)
9 @attribute Industri_Rumah_Tangga (Ada, Tidak_Ada)
10 @attribute Jenis_Rumah (Semi_Permanen, Permanen, Tidak_Permanen)
11 @attribute Lampu (TRUE, FALSE)
12 @attribute Televisi (TRUE, FALSE)
13 @attribute Kipas_Angin (TRUE, FALSE)
14 @attribute Pompa_Air (TRUE, FALSE)
15 @attribute Mesin_Cuci (TRUE, FALSE)
16 @attribute Ac (TRUE, FALSE)
17 @attribute Rice_Cooker (TRUE, FALSE)
18 @attribute Kulkas (TRUE, FALSE)
19 @attribute Komputer (TRUE, FALSE)
20 @attribute Printer (TRUE, FALSE)
21 @attribute Penyedot_Debu (TRUE, FALSE)
22 @attribute Router_Wifi (TRUE, FALSE)
23 @attribute Microwave (TRUE, FALSE)
24 @attribute Blender (TRUE, FALSE)
25 @attribute Sound_Speaker (TRUE, FALSE)
26 @attribute Mixer (TRUE, FALSE)
27 @attribute Hair_Dryer (TRUE, FALSE)
28 @attribute Setrika (TRUE, FALSE)
29 @attribute Daya_Listrik_Rumah_Tangga (Kecil, Sedang, Besar)
30
31 @data
32 Remaja_Akhir, P, Kecil, Kecil, Luar_Provinsi_Jambi, Sedang, Tidak_Ada, Permanen, TRUE, TRUE, TRUE, TRUE, FALSE, FALSE, TRUE, FALSE, FALSE, TRUE, FALSE, FALSE, FALSE,
33 Remaja_Akhir, P, Banyak, Standar, Pinggiran_Kota_Jambi, Besar, Tidak_Ada, Semi_Permanen, TRUE, TRUE, TRUE, TRUE, TRUE, TRUE, TRUE, TRUE, TRUE, TRUE, TRUE, TRUE, FALSE, FALSE,
34 Remaja_Akhir, P, Kecil, Standar, Pusat_Kota_Jambi, Besar, Tidak_Ada, Permanen, TRUE, TRUE, TRUE, TRUE, FALSE, FALSE, TRUE, TRUE, TRUE, TRUE, TRUE, TRUE, TRUE, TRUE, TRUE, TRUE,
35 Remaja_Akhir, P, Kecil, Sangat_Kecil, Desa_atau_Perkampungan, Sangat_Kecil, Tidak_Ada, Permanen, TRUE, FALSE, TRUE, FALSE, TRUE, FALSE, TRUE, FALSE, FALSE, TRUE, TRUE, TRUE, TRUE,
36 Remaja_Akhir, L, Kecil, Sangat_Kecil, Pusat_Kota_Jambi, Kecil, Tidak_Ada, Permanen, TRUE, FALSE, TRUE, TRUE, FALSE, FALSE, TRUE, FALSE, TRUE, TRUE, TRUE, TRUE, TRUE, TRUE, TRUE,
37 Remaja_Akhir, P, Sangat_Kecil, Besar, Luar_Provinsi_Jambi, Sedang, Ada, Permanen, TRUE, FALSE, TRUE, TRUE, TRUE, TRUE, TRUE, TRUE, TRUE, TRUE, TRUE, TRUE, TRUE, TRUE, TRUE, TRUE,
38 Remaja_Akhir, L, Kecil, Kecil, Pusat_Kota_Jambi, Sedang, Tidak_Ada, Semi_Permanen, TRUE, TRUE, TRUE, TRUE, FALSE, FALSE, TRUE, TRUE, TRUE, TRUE, TRUE, TRUE, TRUE, TRUE, TRUE, TRUE,
39 Remaja_Akhir, P, Sedang, Sangat_Kecil, Pusat_Kota_Jambi, Sedang, Tidak_Ada, Permanen, TRUE, TRUE, TRUE, TRUE, FALSE, FALSE, FALSE, FALSE, TRUE, FALSE, FALSE, FALSE, TRUE, TRUE, TRUE, TRUE,

```

Gambar 3. Data penggunaan listrik dalam bentuk format arff.

A. Perhitungan Manual Naive Bayes

Penulis melakukan perhitungan dengan menggunakan persamaan *Teorema Bayes* berikut:

$$P(C_i | X) = \frac{P(X | C_i) \cdot P(C_i)}{P(X)}$$

Keterangan:

- X : Kriteria suatu kasus berdasarkan masukan
- C_i : Kelas solusi pola ke-i, dimana i adalah jumlah label kelas
- (C_i|X) : Probabilitas kemunculan label kelas C_i dengan kriteria masukan X
- P(X|C_i) : Probabilitas kriteria masukan X dengan label kelas C_i
- P(C_i) : Probabilitas label kelas C_i

Data *testing* yang digunakan di ambil dari data kuesioner penggunaan listrik yang sudah tidak memiliki *noise*. Berikut data testing yang akan digunakan untuk contoh perhitungan menggunakan *naive bayes* secara manual :

Tabel 2. Data Kuesioner.

No	Usia	Jenis Kelamin	Jumlah Tanggungan Keluarga	Pendapatan per Bulan	Luas Rumah	Industri Rumah Tangga	Jenis Rumah	Daya Listrik Rumah Tangga
16	Remaja Akhir	L	Kecil	Sangat_Kecil	Sangat_Kecil	Tidak_Ada	Permanen	?

Alur kerja algoritma *naive bayes* yaitu :

1) Menghitung jumlah kelas

Mengklasifikasikan daya listrik rumah tangga menjadi 3 kategori yaitu besar, sedang dan kecil. Langkah pertama yang harus dilakukan adalah menghitung jumlah kelas berdasarkan klasifikasi yang berbentuk (*prior probability*):

$$P(C_i)$$

$$P(\text{KelasDayaListrik} = \text{KECIL}) = \text{jumlah seluruh kelas besar dari total data} = 5/10$$

$$P(\text{KelasDayaListrik} = \text{SEDANG}) = \text{jumlah seluruh kelas sedang dari total data} = 5/10$$

$$P(\text{KelasDayaListrik} = \text{BESAR}) = \text{jumlah seluruh kelas kecil dari total data} = 0/10$$

2) Menghitung jumlah data per kelas/atribut yang ada

Cara menghitung jumlah data per atribut ini sama dengan cara menghitung jumlah kelas. Ada berapa banyak jumlah data kuesioner penggunaan listrik yang daya listrik rumah tangga dari kelas besar, sedang dan kecil. Begitu juga menghitung probabilitas untuk seluruh atribut.

B. Perhitungan Data Testing

$$P(X | C_i)$$

- P (Usia | Kelas_DayaListrik)
 - P (Usia = Remaja_Akhir | Kelas_DayaListrik = KECIL = 4/5
 - P (Usia = Remaja_Akhir | Kelas_DayaListrik = SEDANG = 5/5
 - P (Usia = Remaja_Akhir | Kelas_DayaListrik = BESAR = 0/0
- P (Jenis_Kelamin | Kelas_DayaListrik)
 - P (Jenis_Kelamin = L | Kelas_DayaListrik = KECIL = 3/5
 - P (Jenis_Kelamin = L | Kelas_DayaListrik = SEDANG = 1/5
 - P (Jenis_Kelamin = L | Kelas_DayaListrik = BESAR = 0/0
- P (Jumlah_Tanggungan_Keluarga | Kelas_DayaListrik)
 - P (Jumlah_Tanggungan_Keluarga = Sangat_kecil | Kelas_DayaListrik = KECIL = 2/5
 - P (Jumlah_Tanggungan_Keluarga = Sangat_kecil | Kelas_DayaListrik = SEDANG = 4/5
 - P (Jumlah_Tanggungan_Keluarga = Sangat_kecil | Kelas_DayaListrik = BESAR = 0/0
- P (Pendapatan_per_Bulan | Kelas_DayaListrik)
 - P (Pendapatan_per_Bulan = Sangat_Kecil | Kelas_DayaListrik = KECIL = 3/5
 - P (Pendapatan_per_Bulan = Sangat_Kecil | Kelas_DayaListrik = SEDANG = 1/5
 - P (Pendapatan_per_Bulan = Sangat_Kecil | Kelas_DayaListrik = BESAR = 0/0
- P (Luas_Rumah | Kelas_DayaListrik)
 - P (Luas_Rumah = Sangat_Kecil | Kelas_DayaListrik = KECIL = 2/5
 - P (Luas_Rumah = Sangat_Kecil | Kelas_DayaListrik = SEDANG = 0/5
 - P (Luas_Rumah = Sangat_Kecil | Kelas_DayaListrik = BESAR = 0/0
- P (Industri_Rumah_Tangga | Kelas_DayaListrik)
 - P (Industri_Rumah_Tangga = Tidak_Ada | Kelas_DayaListrik = KECIL = 2/5
 - P (Industri_Rumah_Tangga = Tidak_Ada | Kelas_DayaListrik = SEDANG = 0/5
 - P (Industri_Rumah_Tangga = Tidak_Ada | Kelas_DayaListrik = BESAR = 0/0
- P (Jenis_Rumah | Kelas_DayaListrik)
 - P (Jenis_Rumah = Permanen | Kelas_DayaListrik = KECIL = 3/5
 - P (Jenis_Rumah = Permanen | Kelas_DayaListrik = SEDANG = 2/5
 - P (Jenis_Rumah = Permanen | Kelas_DayaListrik = BESAR = 0/0

1) Mengkalikan semua variabel kelas

Tabel 3. Hasil Perkalian Semua Kelas.

Nama kelas	Prob (Kelas)	Usia	Jenis Kelamin	Jumlah Tanggungan Keluarga	Pendapatan per Bulan	Luas Rumah	Industri Rumah Tangga	Jenis Rumah	Daya Listrik Rumah Tangga
KECIL	0,5	0,8	0,6	0,4	0,6	0,4	0,8	0,6	0,0110592
SEDANG	0,5	1	0,2	0,8	0,2	0	1	0,4	0
BESAR	0	0	0	0	0	0	0	0	0

2) Membandingkan *hasil perkelas*

Dilihat dari hasil perhitungan manual dengan sample sebuah data testing diatas, dengan data testing sebagai berikut :

Tabel 4. Hasil Testing.

No	Usia	Jenis Kelamin	Jumlah Tanggungan Keluarga	Pendapatan per Bulan	Luas Rumah	Industri Rumah Tangga	Jenis Rumah	Kelas Daya Listrik
16	Remaja Akhir	L	Kecil	Sangat_Kecil	Sangat_Kecil	Tidak_Ada	Permanen	Kecil

C. Analisis Atribut Pada Setiap Kelas

Berikut merupakan analisis dari semua atribut pada kelas, semua atribut di analisis dengan cara yang sama.

1) **Atribut Usia**

Dikelompokkan menjadi 4 yaitu remaja awal, remaja akhir, dewasa awal dan lansia.

Tabel 5. Analisis Atribut Usia.

Usia	KECIL	SEDANG	BESAR
Remaja Awal	1	0	0
Remaja Akhir	138	43	6
Dewasa Awal	9	2	0
Lansia	0	1	0

Tabel 4.5 Analisis diatas adalah analisis dari atribut usia. Diketahui bahwa Kelas KECIL terdapat (1 usia remaja awal, 138 usia remaja akhir, 9 usia dewasa awal dan 0 usia lansia), Kelas SEDANG terdapat (0 usia remaja awal, 43 usia remaja akhir, 2 usia dewasa awal dan 1 usia lansia) dan Kelas BESAR terdapat (0 usia remaja awal, 6 usia remaja akhir, 0 usia dewasa awal dan 0 usia lansia).



Gambar 4. Grafik Atribut Usia.

Gambar 4.4 diatas adalah visualisasi dari atribut usia. Dengan sebaran Kelas KECIL terdapat (1 usia remaja awal, 138 usia remaja akhir, 9 usia dewasa awal dan 0 usia lansia), Kelas SEDANG terdapat (0 usia remaja awal, 43 usia remaja akhir, 2 usia dewasa awal dan 1 usia lansia) dan Kelas BESAR terdapat (0 usia remaja awal, 6 usia remaja akhir, 0 usia dewasa awal dan 0 usia lansia).

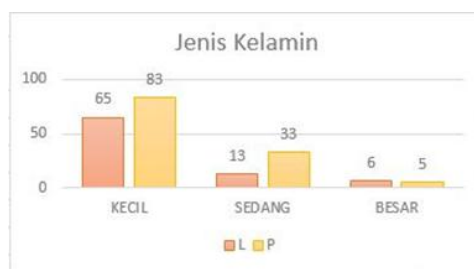
2) Atribut Jenis Kelamin

Dikelompokkan menjadi L dan P (Laki-laki dan Perempuan).

Tabel 6. Analisis Atribut Jenis Kelamin.

Jenis Kelamin	KECIL	SEDANG	BESAR
L	65	13	6
P	83	33	5

Tabel 4.6 Analisis diatas adalah analisis dari atribut jenis kelamin. Diketahui bahwa Kelas KECIL terdapat (65 Laki-laki dan 83 Perempuan), Kelas SEDANG terdapat (13 Laki-laki dan 33 Perempuan) dan Kelas BESAR terdapat (6 Laki-laki dan 5 Perempuan).



Gambar 5. Grafik Atribut Jenis Kelamin.

Gambar 4.5 diatas adalah visualisasi dari atribut jenis kelamin. Dengan sebaran data Kelas KECIL terdapat (65 Laki-laki dan 83 Perempuan), Kelas SEDANG terdapat (13 Laki-laki dan 33 Perempuan) dan Kelas BESAR terdapat (6 Laki-laki dan 5 Perempuan).

3) Atribut Jumlah Tanggungan Keluarga

Dikelompokkan menjadi 5 yaitu sangat kecil, kecil, sedang, banyak dan sangat banyak.

Tabel 7. Analisis Atribut Jumlah Tanggungan Keluarga.

Jumlah Tanggungan Keluarga	KECIL	SEDANG	BESAR
Sangat kecil	50	12	1
Kecil	57	19	4
Sedang	32	13	0
Banyak	8	1	1
Sangat Banyak	2	1	0

Tabel 4.7 Analisis diatas adalah analisis dari atribut jumlah tanggungan keluarga. Diketahui bahwa Kelas KECIL terdapat (50 jumlah tanggungan keluarga sangat kecil, 57 jumlah tanggungan keluarga kecil, 32 jumlah tanggungan keluarga sedang, 8 jumlah tanggungan keluarga banyak dan 2 jumlah tanggungan keluarga sangat banyak), Kelas SEDANG terdapat (12 jumlah tanggungan keluarga sangat kecil, 19 tanggungan keluarga kecil, 13 jumlah tanggungan keluarga sedang, 1 jumlah tanggungan keluarga banyak dan 1 jumlah tanggungan keluarga sangat banyak) dan Kelas BESAR terdapat (1 jumlah tanggungan keluarga sangat kecil, 4 jumlah tanggungan keluarga kecil, 0 jumlah tanggungan keluarga sedang, 1 jumlah tanggungan keluarga banyak dan 0 jumlah tanggungan keluarga sangat banyak).



Gambar 6. Grafik Atribut Jumlah Tanggungan Keluarga.

Gambar 4.6 diatas adalah visualisasi dari atribut jumlah tanggungan keluarga. Dengan sebaran data Kelas KECIL terdapat (50 jumlah tanggungan keluarga sangat kecil, 57 jumlah tanggungan keluarga kecil, 32 jumlah tanggungan keluarga sedang, 8 jumlah tanggungan keluarga banyak dan 2 jumlah tanggungan keluarga sangat banyak), Kelas SEDANG terdapat (12 jumlah tanggungan keluarga sangat kecil, 19 tanggungan keluarga kecil, 13 jumlah tanggungan keluarga sedang,

1 jumlah tanggungan keluarga banyak dan 1 jumlah tanggungan keluarga sangat banyak) dan Kelas BESAR terdapat (1 jumlah tanggungan keluarga sangat kecil, 4 jumlah tanggungan keluarga kecil, 0 jumlah tanggungan keluarga sedang, 1 jumlah tanggungan keluarga banyak dan 0 jumlah tanggungan keluarga sangat banyak).

4) Atribut Pendapatan per Bulan

Dikelompokkan menjadi 5 yaitu sangat kecil, kecil, standar, besar dan sangat besar.

Tabel 8. Analisis Atribut Pendapatan per Bulan.

Pendapatan per Bulan	KECIL	SEDANG	BESAR
Sangat Kecil	25	9	2
Kecil	42	8	0
Standar	47	13	2
Besar	24	11	1
Sangat Besar	10	5	1

Tabel 4.8 Analisis diatas adalah analisis dari atribut pendapatan per bulan. Diketahui bahwa Kelas KECIL terdapat (25 pendapatan sangat kecil, 42 pendapatan kecil, 47 pendapatan standar, 24 pendapatan besar dan 10 pendapatan sangat besar), Kelas SEDANG terdapat (9 pendapatan sangat kecil, 8 pendapatan kecil, 13 pendapatan standar, 11 pendapatan besar dan 5 pendapatan sangat besar) dan Kelas BESAR terdapat (2 pendapatan sangat kecil, 0 pendapatan kecil, 2 pendapatan standar, 1 pendapatan besar dan 1 pendapatan sangat besar).



Gambar 7. Grafik Atribut Pendapatan per Bulan.

Gambar 4.7 diatas adalah visualisasi dari atribut pendapatan per bulan. Dengan sebaran data Kelas KECIL terdapat (25 pendapatan sangat kecil, 42 pendapatan kecil, 47 pendapatan standar, 24 pendapatan besar dan 10 pendapatan sangat besar), Kelas SEDANG terdapat (9 pendapatan sangat kecil, 8 pendapatan

kecil, 13 pendapatan standar, 11 pendapatan besar dan 5 pendapatan sangat besar) dan Kelas BESAR terdapat (2 pendapatan sangat kecil, 0 pendapatan kecil, 2 pendapatan standar, 1 pendapatan besar dan 1 pendapatan sangat besar).

5. KESIMPULAN DAN SARAN

Kesimpulan yang dapat diambil dari hasil penelitian ini adalah sebagai berikut. Energi listrik dapat dimanfaatkan untuk melakukan aktivitas dengan manfaat yang sangat besar, dimana berbagai peralatan untuk memenuhi kebutuhan hidup dioperasikan menggunakan energi listrik. Masih banyak masyarakat yang belum memahami upaya dalam menggunakan listrik dengan bijak, sehingga diterapkan teknik klasifikasi menggunakan metode *Naive Bayes* (Tan et al., 2018).

Presentasi hasil akurasi klasifikasi *Naive Bayes* menunjukkan bahwa dengan menggunakan *Use Training Set* diperoleh 161 *Correctly Classified Instances*, 39 *Incorrectly Classified Instances*, dengan persentase akurasi 80,5 % dan 19,5 %. Menggunakan *5-Fold Cross Validation* diperoleh 150 *Correctly Classified Instances*, 50 *Incorrectly Classified Instances*, dengan persentase akurasi 75 % dan 25 %. Menggunakan *10-Fold Cross Validation* diperoleh 148 *Correctly Classified Instances*, 52 *Incorrectly Classified Instances*, dengan persentase akurasi 74 % dan 26 %. Hasil klasifikasi *Naive Bayes* dengan akurasi terbesar diperoleh dengan *Use Training Set*, yaitu 80,5 % *Correctly Classified Instances* dan 19,5 % *Incorrectly Classified Instances*.

Hasil dari seleksi atribut menggunakan algoritma *Classifier Attribute Evaluation* (*ClassifierAttributeEval*) menunjukkan bahwa atribut yang paling berpengaruh terhadap klasifikasi adalah barang-barang elektronik (Parmar et al., 2023).

Hasil penelitian ini diharapkan dapat diimplementasikan secara nyata oleh masyarakat agar lebih bijak dalam penggunaan listrik. Disarankan penelitian selanjutnya menggunakan atribut yang lebih banyak dan waktu pengumpulan data yang lebih lama untuk memperoleh hasil yang lebih akurat. Selain itu, data ini dapat digunakan dan diuji dengan algoritma maupun *tools* lain untuk keperluan klasifikasi.

DAFTAR REFERENSI

- Aggarwal, C. C. (2015). *Data Mining: The Textbook*. Springer.
- Boote, D. N., & Beile, P. (2016). Scholars Before Researchers: On The Centrality Of The Dissertation Literature Review In Research Preparation. *Educational Researcher*, 45(2), 82–92.
- Creswell, J. W., & Creswell, J. D. (2018). *Research Design: Qualitative, Quantitative, And Mixed Methods Approaches* (5th ed.). SAGE Publications.
- Firdaus, F., & Karim, S. A. (2025). Penghematan Dan Penggunaan Energi Listrik Bagi Warga Kelurahan Parang Bugisi Malino. *Prosiding Seminar Nasional Pengabdian Kepada Masyarakat*.
- Hastie, T., Tibshirani, R., & Friedman, J. (2022). *The Elements Of Statistical Learning: Data Mining, Inference, And Prediction* (2nd ed.). Springer.
- International Energy Agency. (2020). *World Energy Outlook 2020*. IEA.
- Kementerian Energi dan Sumber Daya Mineral Republik Indonesia. (2019). *Statistik Ketenagalistrikan Indonesia*.
- Kotu, V., & Deshpande, B. (2019). *Data Science: Concepts And Practice* (2nd ed.). Morgan Kaufmann.
- Parmar, M., Sonker, A., & Sejwar, V. (2023). A Comparative Analysis For Filter Based Feature Selection Techniques With Tree Based Classification. *International Journal on Recent and Innovation Trends in Computing and Communication*, 11(10s), 360–369.
- Sari, L., Alam, S., Surjati, I., & Astuti, P. (2023). Penyuluhan Penggunaan Listrik Untuk Kebutuhan Rumah Tangga Secara Efisien Di Kelurahan Jatibening Baru, Bekasi. *Jurnal Pengabdian Masyarakat Teknik*, 6(1), 27–32.
- Suganthi, L., & Samuel, A. A. (2019). Energy Models For Demand Forecasting---A Review. *Renewable and Sustainable Energy Reviews*, 33, 434–446.
- Tan, P.-N., Steinbach, M., & Kumar, V. (2018). *Introduction To Data Mining* (2nd ed.). Pearson.
- Witten, I. H., Frank, E., Hall, M. A., & Pal, C. J. (2016). *Data Mining: Practical Machine Learning Tools And Techniques* (4th ed.). Morgan Kaufmann.
- Yos, J. K. L., Km, S., & Mulia, T. (2016). *Penerapan Data Mining Dengan Metode Klasifikasi Naïve Bayes Untuk Memprediksi Kelulusan Mahasiswa Dalam Mengikuti English Proficiency Test (Studi Kasus: Universitas Potensi Utama)*.