



## KLASIFIKASI KONSUMSI ENERGI INDUSTRI BAJA MENGGUNAKAN TEKNIK DATA MINING

Jajang Jaya Purnama<sup>1)</sup>, Sri Rahayu<sup>2\*)</sup>

<sup>1</sup>Program Studi Sistem Informasi, Universitas Nusa Mandiri

<sup>2</sup>Program Studi Informatika, Universitas Nusa Mandiri

Perumahan Griya Srimahi Indah Blok F9 no. 12, Srimahi, Tambun Utara, Bekasi

Email: <sup>1</sup>[jajang.jjp@nusamandiri.ac.id](mailto:jajang.jjp@nusamandiri.ac.id), <sup>2</sup>[sriahayu.rry@nusamandiri.ac.id](mailto:sriahayu.rry@nusamandiri.ac.id)

Email Penulis Korespondensi: [sriahayu.rry@nusamandiri.ac.id](mailto:sriahayu.rry@nusamandiri.ac.id)

### Abstract

*Human needs in fulfilling clothing, food and housing in today's life cannot be separated from the involvement of electrical energy. In several sectors of life, namely the household sector, industry, business, social, government office buildings, and public street lighting, electricity is needed. The energy consumption industry sector is relatively higher than other sectors, so it is necessary to control energy consumption, especially in the industrial sector. As a result, for a nation or region, forecasting the use of electrical energy becomes urgent and crucial. Research on this issue has emerged from various countries, for example, research from Korea on energy consumption prediction models for smart factories using a data mining algorithm that introduces and explores the steel industry energy consumption prediction model by producing the best model, namely Random Forest with an RMSE value of 7.33 in the test set. In addition, another study raised the title of an efficient energy consumption prediction model for an analytical data of industrial buildings in a smart city by presenting and exploring a predictive energy consumption model based on data mining techniques for a smart small-scale steel industry in South Korea using variables such as lagging and current. main reactive power, lagging power factor and leading current, carbon dioxide emission and load type. Research from Australia is also not left behind, discussing the prediction of industrial energy consumption using data mining techniques which presents and explores energy consumption prediction models using a data mining approach for the steel industry to show that the Random Forest model can best predict energy consumption and outperform other conventional algorithms in comparison. This study presents a classification of energy consumption in the steel industry, in order to know the pattern of using light loads, medium loads, and maximum loads using data mining techniques on public data that is already available on this matter, with the aim that energy users in the steel industry are wiser in using energy because you already know the pattern of each load. The methods used include Random Forest, Decision Tree, Naïve Bayes and Artificial Neural Networks resulting in accuracy of 91.13%, 90.50%, 70.97% and 75.56%, so that the classification method is the most suitable for use. In classifying industrial energy consumption on the steel industry energy consumption dataset, Random Forest.*

**Keyword:** Data Mining, Steel Industry Energy, Classification, Random Forest, Decision Tree.

### Abstrak

Kebutuhan manusia dalam memenuhi sandang, pangan dan papan pada kehidupannya saat ini tidak terlepas dari terlibatnya energi listrik. Pada beberapa sektor kehidupan, yaitu sektor rumah tangga, industri, bisnis, sosial, gedung kantor pemerintah, dan penerangan jalan umum membutuhkan energi listrik. Sektor industri konsumsi energi relatif lebih tinggi dibandingkan sektor lainnya, sehingga perlu adanya pengendalian konsumsi energi terutama di sektor industri. Akibatnya, untuk suatu Bangsa atau wilayah, prakiraan penggunaan energi listrik menjadi mendesak dan krusial. Penelitian mengenai hal ini muncul dari berbagai negara misalnya penelitian asal Korea tentang model prediksi konsumsi energi untuk smart factory menggunakan algoritma data mining yang memperkenalkan dan mengeksplorasi model prediksi konsumsi energi industri baja dengan menghasilkan model terbaik yaitu Random Forest dengan nilai RMSE 7,33 pada set pengujian. Selain itu, penelitian lain mengangkat judul model prediksi konsumsi energi yang efisien untuk suatu data analitik bangunan industri di kota pintar dengan menyajikan dan mengeksplorasi model konsumsi energi prediktif berdasarkan teknik penambangan data untuk industri baja skala kecil yang cerdas di Korea Selatan menggunakan variabel seperti lagging dan arus utama daya reaktif, faktor daya lagging dan arus terdapan, emisi karbon dioksida, dan jenis beban. Penelitian asal Australia juga tidak ketinggalan, membahas mengenai prediksi konsumsi energi industri menggunakan teknik data mining yang menyajikan dan mengeksplorasi model prediksi konsumsi energi menggunakan pendekatan data mining untuk industri baja hingga menunjukkan bahwa model Random Forest dapat memprediksi konsumsi energi terbaik dan mengungguli algoritma konvensional lainnya dalam perbandingan. Penelitian ini menyajikan klasifikasi konsumsi energi pada industri baja, agar dapat diketahui pola penggunaan beban ringan, beban sedang, dan



beban maksimum menggunakan teknik data mining pada data publik yang sudah tersedia mengenai hal tersebut, dengan tujuan para pengguna energi di industri baja lebih bijak dalam menggunakan energi karena telah mengetahui pola masing-masing beban. Dengan metode yang digunakan diantaranya Random Forest, Decision Tree, Naïve Bayes dan Artificial Neural Network menghasilkan akurasi berturut-turut yaitu 91,13%, 90,50%, 70,97% dan 75,56%, sehingga metode klasifikasi yang paling cocok digunakan dalam melakukan klasifikasi konsumsi energi industri pada dataset *steel industry energy consumption* adalah Random Forest.

**Kata Kunci:** Data Mining, Energi Industri Baja, Klasifikasi, Random Forest, Decision Tree.

## 1. PENDAHULUAN

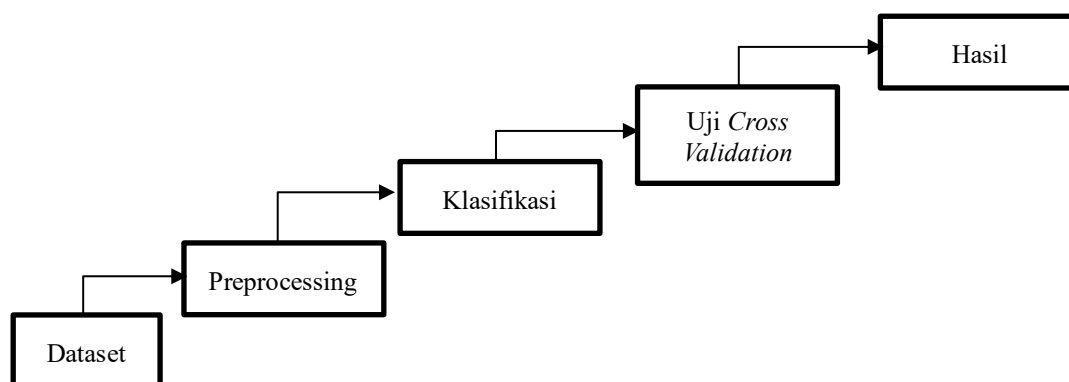
Kebutuhan manusia dalam memenuhi sandang, pangan dan papan pada kehidupannya saat ini tidak terlepas dari terlibatnya energi listrik. Pada beberapa sektor kehidupan, yaitu sektor rumah tangga, industri, bisnis, sosial, gedung kantor pemerintah, dan penerangan jalan umum membutuhkan energi listrik [1]. Akibatnya, untuk suatu Bangsa atau wilayah, prakiraan penggunaan energi listrik menjadi mendesak dan krusial [2]. Dari beberapa sektor yang disebutkan, sektor industri konsumsi energi relatif lebih tinggi dibandingkan sektor lainnya. Sehingga perlu adanya pengendalian konsumsi energi terutama di sektor industri.

Banyak negara-negara terus melakukan penelitian untuk mengoptimalkan konsumsi energi dan menyalurkan listrik dengan cara yang konsisten. Penelitian mengenai hal ini pernah dilakukan beberapa peneliti, diantaranya penelitian asal Korea tentang model prediksi konsumsi energi untuk *smart factory* menggunakan algoritma data mining oleh [3] yang memperkenalkan dan mengeksplorasi model prediksi konsumsi energi industri baja dengan menghasilkan model terbaik yaitu Random Forest dengan nilai RMSE 7,33 pada set pengujian. Selain itu, penelitian lain mengangkat judul model prediksi konsumsi energi yang efisien untuk suatu data analitik bangunan industri di kota pintar dengan menyajikan dan mengeksplorasi model konsumsi energi prediktif berdasarkan teknik penambangan data untuk industri baja skala kecil yang cerdas di Korea Selatan menggunakan variabel seperti lagging dan arus utama daya reaktif, faktor daya lagging dan arus terdepan, emisi karbon dioksida, dan jenis beban [4]. Selanjutnya penelitian asal Australia mengenai prediksi konsumsi energi industri menggunakan teknik data mining oleh [5] yang menyajikan dan mengeksplorasi model prediksi konsumsi energi menggunakan pendekatan data mining untuk industri baja hingga menunjukkan bahwa model Random Forest dapat memprediksi konsumsi energi terbaik dan mengungguli algoritma konvensional lainnya dalam perbandingan.

Dari beberapa penelitian yang pernah dilakukan dan permasalahan yang terjadi terhadap konsumsi energi industri, pada penelitian ini akan menyajikan klasifikasi konsumsi energi pada industri baja, agar dapat diketahui pola penggunaan beban ringan, beban sedang, dan beban maksimum menggunakan teknik data mining pada data publik yang sudah tersedia mengenai hal tersebut, dengan tujuan para pengguna energi di industri baja lebih bijak dalam menggunakan energi karena telah mengetahui pola masing-masing beban.

## 2. METODE PENELITIAN

Tahapan-tahapan metode penelitian yang dilakukan dalam penelitian ini digambarkan pada diagram alur seperti dibawah ini:



**Gambar 1.** Tahapan Metode Penelitian



**2.1 Dataset**

Dalam sebuah penelitian dataset menjadi bahan utama untuk diolah menggunakan algoritma, penelitian ini menggunakan dataset *steel industry energy consumption* yang diperoleh dari *website UCI Repository*. Dataset *steel industry energy consumption* dipublikasikan tahun 30 Maret 2021 dengan *instance* sebanyak 35040 dan memiliki 10 atribut diantaranya : *Industry Energy Consumption* , *Lagging Current Reactive Power* , *Leading Current Reactive Power* , *tCO2(CO2)* , *Lagging Current Power Factor* , *Leading Current Power Factor* , *Number of Seconds from Midnight* , *Week status* , *Day of week* , *Load Type* dengan detail informasinya diuraikan dalam tabel 1 berikut ini:

**Tabel 1. Dataset**

No	Data Variables	Type	Measurement
1	<i>Industry Energy Consumption</i>	<i>Continuous</i>	<i>kWh</i>
2	<i>Lagging Current Reactive Power</i>	<i>Continuous</i>	<i>kVarh</i>
3	<i>Lagging Current Reactive Power</i>	<i>Continuous</i>	<i>kVarh</i>
4	<i>tCO2(CO2)</i>	<i>Continuous</i>	<i>Ppm</i>
5	<i>Lagging Current Power Factor</i>	<i>Continuous</i>	<i>%</i>
6	<i>Leading Current Power Factor</i>	<i>Continuous</i>	<i>%</i>
7	<i>Number of Seconds from Midnight</i>	<i>Continuous</i>	<i>S</i>
8	<i>Week status</i>	<i>Categorical</i>	<i>(Weekend(0) or a Weekday(1))</i>
9	<i>Day of week</i>	<i>Categorical</i>	<i>Sunday, Monday ... Saturday</i>
10	<i>Load Type</i>	<i>Categorical</i>	<i>Light Load, Medium Load, Maximum Load</i>

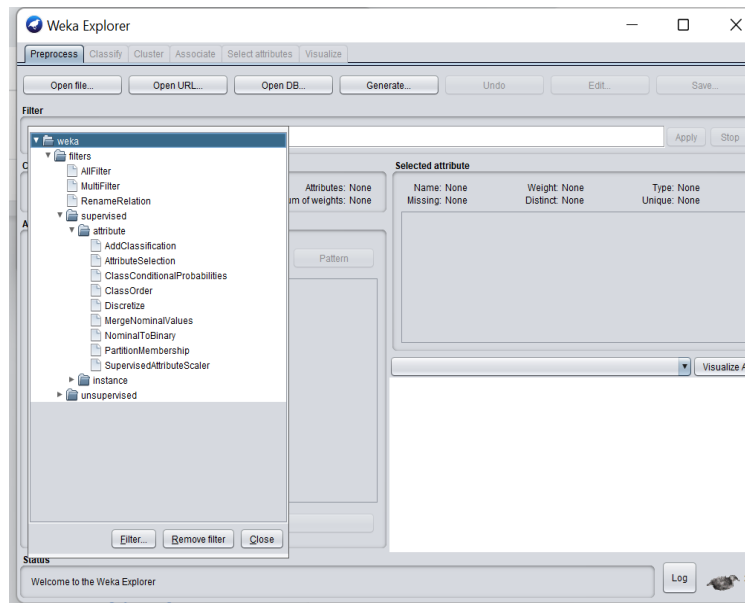
**2.2 Preprocessing**

Sebelum melakukan klasifikasi, agar mempermudah proses interpretasi, terlebih dahulu dilakukan *preprocessing*. Pada penelitian ini tahap *preprocessing* yang dilakukan adalah dengan memilih atribut menggunakan *tools* pada aplikasi data mining dan transportasi data menggunakan normalisasi, kemudian hasilnya digunakan untuk melakukan klasifikasi dan diambil *preprocessing* mana yang lebih baik, pada penelitian ini akan dibahas keduanya, dengan pengertian masing-masing adalah sebagai berikut:

- a. Proses seleksi atribut, merupakan tahapan untuk memilih atribut-atribut dalam dataset yang memiliki korelasi terhadap kelasnya [6]. *Feature selection* dapat didekati dengan beberapa metode, antara lain adalah *wrapper*, *filter*, dan *embedded*. Fitur yang biasanya digunakan dalam pemilihan atribut adalah nilai korelasi dan *mutual information*. *Correlation Feature Selection (CFS)* merupakan teknik yang menerapkan fitur korelasi dan *mutual*



*information*. CFS memilih atribut dengan menghitung korelasi antara atribut dan kelas serta korelasi antara atribut dan atribut lainnya, CFS akan memilih atribut dengan nilai korelasi dengan kelas yang sangat tinggi namun nilai korelasi antar atribut memiliki nilai yang kecil [7][8]. Gambar 2 dibawah ini merupakan tampilan dari aplikasi data mining, WEKA dalam melakukan seleksi atribut:



**Gambar 2.** Tampilan Seleksi Atribut pada WEKA

- b. Normalisasi adalah proses penskalaan nilai atribut dari data sehingga bisa terletak pada rentang tertentu [9]. Berikut beberapa tahap normalisasi yang dilakukan :
  1. *Min-Max Normalization* merupakan metode normalisasi dengan melakukan transformasi linier terhadap data asli sehingga menghasilkan keseimbangan nilai perbandingan antara data saat sebelum dan sesudah proses. Metode ini dapat menggunakan rumus sebagai berikut :
  2. *Z-score Normalization* merupakan metode normalisasi berdasarkan mean (nilai rata-rata) dan standard deviation (deviasi standar) dari data. Metode ini sangat berguna jika tidak diketahui nilai aktual minimum dan maksimum dari data. Rumus yang digunakan sebagai berikut :
  3. *Decimal scaling* merupakan metode normalisasi dengan menggerakkan nilai desimal dari data ke arah yang diinginkan. Formula yang digunakan sebagai berikut :

### 2.3 Klasifikasi

Klasifikasi dapat didefinisikan sebagai pekerjaan yang melakukan pelatihan/pembelajaran terhadap fungsi target yang memetakan setiap set atribut (fitur) kesatu jumlah label kelas yang tersedia, bisa juga diartikan pekerjaan menilai objek data untuk memasukkannya ke dalam kelas tertentu dari jumlah kelas yang tersedia [10]. Klasifikasi melakukan pembangunan model berdasarkan data latih yang ada, kemudian menggunakan model tersebut untuk mengklasifikasikan pada data yang baru, dengan tujuan sistem yang dibangun nantinya dapat melakukan klasifikasi semua data set dengan benar, keberhasilannya perlu dilakukan pengukuran setelah pola terbentuk [11]. Tahap Klasifikasi pada penelitian ini dilakukan menggunakan beberapa algoritma diantaranya:

- a. *Random Forest*, merupakan salah satu jenis algoritma klasifikasi yang terdiri dari lebih satu pohon keputusan yang setiap pohon keputusan dibentuk bergantung pada nilai-nilai vector acak sampel secara independen dan identik didistribusikan yang sama untuk semua pohon [12]. Metode ini merupakan salah satu metode klasifikasi yang sangat akurat digunakan dalam melakukan prediksi, bisa menangani inputan variabel yang sangat besar jumlahnya tanpa *overfitting*, dan membantu menghilangkan korelasi antara pohon keputusan seperti karakteristik *ensemble methods* [13].
- b. *Decision Tree*, merupakan algoritma pengambilan keputusan yang melakukan partisi rekursif atas ruang *instance*, sebuah pohon keputusan tipikal terdiri dari simpul internal, tepi dan simpul daun. Setiap simpul



internal disebut simpul keputusan yang mewakili tes pada atribut atau subset atribut, dan masing-masing edge diberi label dengan nilai spesifik atau rentang nilai atribut input [14]. Pengklasifikasi *Decision Tree* memperoleh akurasi yang serupa dan terkadang lebih baik jika dibandingkan dengan metode klasifikasi lainnya [15].

- c. *Naive bayes*, merupakan sebuah pengklasifikasian probabilistik sederhana yang menghitung sekumpulan probabilitas dengan menjumlahkan frekuensi dan kombinasi nilai dari dataset yang diberikan, juga dikemukakan oleh ilmuwan Inggris Thomas Bayes, yaitu memprediksi peluang di masa depan berdasarkan pengalaman di masa sebelumnya [16].
- d. *Artificial Neural Network*, atau jaringan syaraf tiruan adalah model non-linear yang rumit dibangun dari komponen yang secara individu berperilaku mirip seperti model regresi yang dapat merepresentasikan sebuah grafik, dan beberapa sub-grafik tampaknya ada integritas yang sama dengan gerbang logika. struktur dari jaringan neuron atau saraf secara terperinci dirancang terlebih dahulu [17].

Aplikasi yang digunakan untuk pembelajaran mesin menggunakan aplikasi WEKA 3.8 pada sistem operasi Windows 11 dengan spesifikasi *personal computer* sebagai berikut :

**Tabel 2.** Spesifikasi *Personal Computer*

Processor	intel® core™ i7-8565U
CPU	1.80 GHz 1.99 GHz
RAM	8.00 GB

## 2.4 Uji *Cross Validation*

*Cross validation* adalah pengujian standar yang dilakukan untuk memprediksi *error rate* dan lain-lain [18]. *K-Fold Cross Validation* adalah salah satu dari jenis pengujian *cross validation* yang berfungsi untuk menilai kinerja proses sebuah metode algoritma dengan membagi sampel data secara acak dan mengelompokkan data tersebut sebanyak nilai  $K$  *k-fold*. Kemudian salah satu kelompok *k-fold* tersebut akan dijadikan sebagai data uji sedangkan sisa kelompok yang lain akan dijadikan sebagai data latih [19]. Penelitian ini menggunakan teknik *k-fold cross validation* dengan cara kerja 10 kali pengulangan 1 kali untuk pengujian sisanya pelatihan yang dilakukan berulang kali sebanyak 10 kali secara bergantian yang akan menghasilkan hasil *accuracy*, *TP Rate*, *FP Rate*, *Precision*, *Recall*, *F-Measure*, *ROC*.

## 3. HASIL DAN PEMBAHASAN

Dari tahapan metode penelitian yang dilakukan, kita mengetahui bahwa *preprocessing* dilakukan dengan 2 metode yaitu atribut selection dan normalisasi kemudian hasilnya masing-masing dilakukan pada 4 model klasifikasi. Pada penelitian ini, terdapat 3 kelas yang diklasifikasi yaitu beban ringan (a), beban sedang (b), dan beban maksimum (c). Hasil dari pengujian masing-masing algoritma ini akan dipaparkan pada pembahasan berikut:

### 3.1 Pengujian Algoritma dengan *Atribut Selection*

Hasil dari pengujian algoritma *Random Forest*, *Decision Tree*, *Naive Bayes*, dan *Artificial Neural Network* yang dihasilkan dari *preprocessing* menggunakan *atribut selection* disajikan pada poin-poin berikut ini, diantaranya:

#### 3.1.1 Hasil Pengujian Algoritma Klasifikasi *Random Forest*

Hasil pengujian algoritma *Random Forest* dengan penerapan *preprocessing* menggunakan *atribut selection* pada 3 kelas disajikan dengan perhitungan menggunakan *confusion matrix* pada tabel 3 kemudian dihitung dan menghasilkan akurasi sebesar 86,45%. *Confusion matrix* adalah sebuah tabel yang menyatakan jumlah data uji yang benar diklasifikasikan dan jumlah data uji yang salah diklasifikasikan [20].



**Tabel 3. Confusion Matrix**  
*Algoritma Random Forest Attribute Selection*

	a	b	c
a	16764	1304	4
b	479	7768	1449
c	298	1213	5761

$$\begin{aligned}
 \text{Akurasi} &= (TP+TN / TP+TN+FP+FN) \\
 &= 16764+7768+5761 / 16764 + 1304 + 4 + 479 + 7768 + 1449 + 298 + 1213 + 5761 \\
 &= 30.293 / 35040 \\
 &= 86,45\%
 \end{aligned}$$

3.1.2 Hasil Algoritma Klasifikasi *Decision Tree*

Hasil pengujian algoritma *Decision Tree* dengan penerapan *preprocessing* menggunakan *atribut selection* pada 3 kelas disajikan dengan perhitungan menggunakan *confusion matrix* pada tabel 4 kemudian dihitung dan menghasilkan akurasi sebesar 86,67%.

**Tabel 4. Confusion Matrix**  
*Algoritma Decision Tree Attribute Selection*

	a	b	c
a	16884	1184	4
b	545	7754	1397
c	331	1207	5734

$$\begin{aligned}
 \text{Akurasi} &= (TP+TN / TP+TN+FP+FN) \\
 &= 16884+7754+5734 / 16884 + 1184 + 4 + 545 + 7754 + 1397 + 331 + 1207 + 5734 \\
 &= 30.372 / 35040 \\
 &= 86,67\%
 \end{aligned}$$

3.1.3 Hasil Algoritma Klasifikasi *Naive Bayes*

Hasil pengujian algoritma *Naive Bayes* dengan penerapan *preprocessing* menggunakan *atribut selection* pada 3 kelas disajikan dengan perhitungan menggunakan *confusion matrix* pada tabel 5 kemudian dihitung dan menghasilkan akurasi sebesar 73,07%.

**Tabel 5. Confusion Matrix**  
*Algoritma Naive Bayes Atribut Selection*

	a	b	c
a	15150	2235	687
b	1837	5315	2544
c	2544	1588	5141



$$\begin{aligned}
 \text{Akurasi} &= (TP+TN / TP+TN+FP+FN) \\
 &= 15150+5315+5141 / 15150 + 2235 + 687 + 1837 + 5315 + 2544 + 2544 + 1588 + 5141 \\
 &= 25.606 / 35040 \\
 &= 73,07\%
 \end{aligned}$$

3.1.4 Hasil Algoritma Klasifikasi *Artificial Neural Network*

Hasil pengujian algoritma *Artificial Neural Network* dengan penerapan *preprocessing* menggunakan *atribut selection* pada 3 kelas disajikan dengan perhitungan menggunakan *confusion matrix* pada tabel 6 kemudian dihitung dan menghasilkan akurasi sebesar 68,43%.

**Tabel 6. Confusion Matrix**  
Algoritma *Artificial Neural Network* Atribut Selection

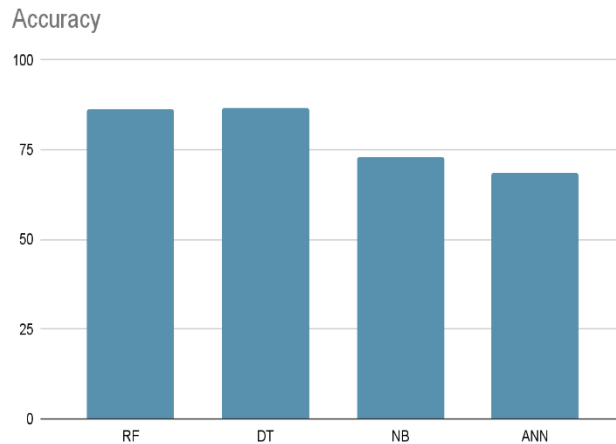
	a	b	c
a	15250	2082	740
b	2058	5687	1951
c	1049	3182	3041

$$\begin{aligned}
 \text{Akurasi} &= (TP+TN / TP+TN+FP+FN) \\
 &= 15250+5687+3041 / 15250 + 2082 + 740 + 2058 + 5687 + 1951 + 1049 + 3182 + 3041 \\
 &= 23.978 / 35040 \\
 &= 68,43\%
 \end{aligned}$$

Pada tabel 7 berikut ini disajikan hasil dari pengujian algoritma *Random Forest* (RF), *Decision Tree* (DT), *Naive Bayes* (NB), dan *Artificial Neural Network* (ANN) dengan *preprocessing* menggunakan *atribut selection*. Hasil pengujian ini diantaranya adalah mendapatkan nilai akurasi yang merupakan derajat ketepatan antara nilai yang diukur dengan nilai sebenarnya [21], *TP Rate* yang merupakan persentase dari kelas positif yang berhasil diklasifikasi sebagai kelas positif, *FP Rate* yang merupakan kelas negatif yang diklasifikasi sebagai kelas positif [22], *Precision* yang merupakan proporsi jumlah dokumen teks yang relevan terkendali diantara semua dokumen teks yang terpilih oleh sistem, *Recall* merupakan proporsi jumlah dokumen teks yang relevan terkendali diantara semua dokumen teks relevan yang ada pada koleksi [23], *F-Measure* adalah rata-rata harmoni dari presisi dan *recall* dan ROC Area menunjukkan metrik seberapa baik model dalam memprediksi [24]. ROC dapat digunakan untuk menganalisis model hasil klasifikasi [25].

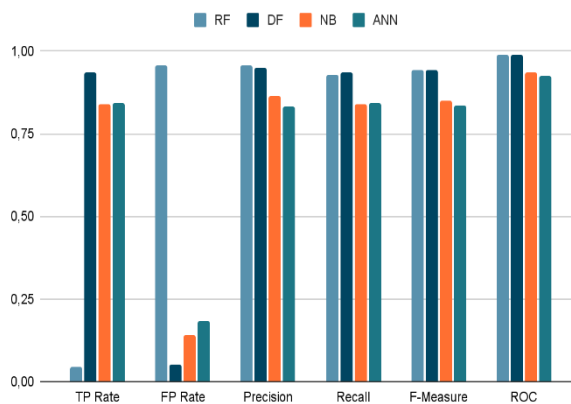
**Tabel 7. Hasil Pengujian Algoritma Atribut Selection**

Model	Accurate	TP Rate	FP Rate	Precision	Recall	F-Measure	ROC
RF	86,45 %	0,928	0,046	0,956	0,928	0,941	0,989
DT	86,67 %	0,934	0,052	0,951	0,934	0,942	0,988
NB	73,07 %	0,838	0,140	0,864	0,838	0,851	0,935
ANN	68,43%	0,844	0,183	0,831	0,844	0,837	0,925



**Gambar 3.** Grafik Akurasi Algoritma Klasifikasi dengan *Atribut Selection*

Pada gambar 3 disajikan grafik akurasi algoritma-algoritma yang digunakan dalam melakukan klasifikasi pada penelitian ini diantaranya *Random Forest* (RF), *Decision Tree* (DT), *Naive Bayes* (NB), dan *Artificial Neural Network* (ANN) dengan *preprocessing* menggunakan *atribut selection*. Akurasi hasil dari pembelajaran mesin pada dataset *steel industry energy consumption* adalah metode *Decision Tree* (DT) sebesar 86,67%.



**Gambar 4.** Grafik Hasil Pengujian Algoritma Klasifikasi dengan *Atribut Selection*

Pada gambar 4 disajikan grafik hasil pengujian berupa TP Rate, FP Rate, Precision, Recall, F-Measure, dan ROC algoritma-algoritma yang digunakan dalam melakukan klasifikasi pada penelitian ini diantaranya *Random Forest* (RF), *Decision Tree* (DT), *Naive Bayes* (NB), dan *Artificial Neural Network* (ANN) dengan *preprocessing* menggunakan *atribut selection*.

### 3.2 Pengujian Algoritma dengan Normalisasi

Hasil dari pengujian algoritma *Random Forest*, *Decision Tree*, *Naive Bayes*, dan *Artificial Neural Network* yang dihasilkan dari *preprocessing* menggunakan normalisasi disajikan pada poin-poin berikut ini, diantaranya:

#### 3.2.1 Hasil Pengujian Algoritma Klasifikasi *Random Forest*

Hasil pengujian algoritma *Random Forest* dengan penerapan *preprocessing* menggunakan normalisasi pada 3 kelas disajikan dengan perhitungan menggunakan *confusion matrix* pada tabel 8 kemudian dihitung dan menghasilkan akurasi sebesar 91,13%.





**Tabel 8. Confusion Matrix**  
*Algoritma Random Forest Normalisasi*

	a	b	c
a	17688	341	43
b	332	8147	1217
c	84	1088	6100

$$\begin{aligned}
 \text{Akurasi} &= (TP+TN / TP+TN+FP+FN) \\
 &= 17688+8147+6100 / 17688 + 341 + 43 + 332 + 8147 + 1217 + 84 + 1088 + 6100 \\
 &= 31.935 / 35040 \\
 &= 91,13\%
 \end{aligned}$$

3.2.2 Hasil Algoritma Klasifikasi *Decision Tree*

Hasil pengujian algoritma *Decision Tree* dengan penerapan *preprocessing* menggunakan normalisasi pada 3 kelas disajikan dengan perhitungan menggunakan *confusion matrix* pada tabel 9 kemudian dihitung dan menghasilkan akurasi sebesar 86,67%.

**Tabel 9. Confusion Matrix**  
*Algoritma Decision Tree Normalisasi*

	a	b	c
a	17539	459	74
b	387	7960	1349
c	106	952	6214

$$\begin{aligned}
 \text{Akurasi} &= (TP+TN / TP+TN+FP+FN) \\
 &= 17539+7960+6214 / 17539+ 459 + 74 + 387 + 7960 + 1349 + 106 + 952 + 6214 \\
 &= 31.713 / 35040 \\
 &= 90,50\%
 \end{aligned}$$

3.2.3 Hasil Algoritma Klasifikasi *Naive Bayes*

Hasil pengujian algoritma *Naive Bayes* dengan penerapan *preprocessing* menggunakan normalisasi pada 3 kelas disajikan dengan perhitungan menggunakan *confusion matrix* pada tabel 10 kemudian dihitung dan menghasilkan akurasi sebesar 70,97%.

**Tabel 10. Confusion Matrix**  
*Algoritma Naive Bayes Normalisasi*

	a	b	c
a	15047	1759	1266
b	1215	3698	4783
c	639	510	6123



$$\begin{aligned}
 \text{Akurasi} &= (TP+TN / TP+TN+FP+FN) \\
 &= 15047+3698+6123 / 15047+ 1759 + 1266 + 1215 + 3698 + 4783 + 639 + 510 + 6123 \\
 &= 24.868 / 35040 \\
 &= 70,97\%
 \end{aligned}$$

3.2.4 Hasil Algoritma Klasifikasi *Artificial Neural Network*

Hasil pengujian algoritma *Artificial Neural Network* dengan penerapan *preprocessing* menggunakan normalisasi pada 3 kelas disajikan dengan perhitungan menggunakan *confusion matrix* pada tabel 11 kemudian dihitung dan menghasilkan akurasi sebesar 75,56%.

**Tabel 11. Confusion Matrix**  
Algoritma *Artificial Neural Network* Normalisasi

	a	b	c
a	16319	1211	542
b	1289	5657	2750
c	512	2257	4503

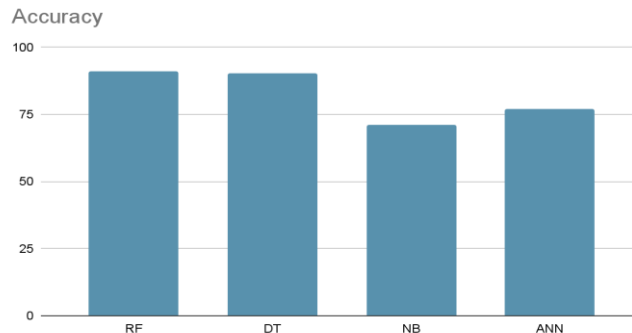
$$\begin{aligned}
 \text{Akurasi} &= (TP+TN / TP+TN+FP+FN) \\
 &= 16319+5657+4503 / 16319+ 1211 + 542 + 1289 + 5657 + 2750 + 512 + 2257 + 4503 \\
 &= 26.479 / 35040 \\
 &= 75,56\%
 \end{aligned}$$

Pada tabel 12 berikut ini disajikan hasil dari pengujian algoritma *Random Forest* (RF), *Decision Tree* (DT), *Naive Bayes* (NB), dan *Artificial Neural Network* (ANN) dengan *preprocessing* menggunakan normalisasi. Hasil pengujian ini diantaranya adalah mendapatkan nilai akurasi, *TP Rate*, *FP Rate*, *Precision*, *Recall*, *F-Measure*, dan *ROC*.

Tabel 12. Hasil Pengujian Algoritma

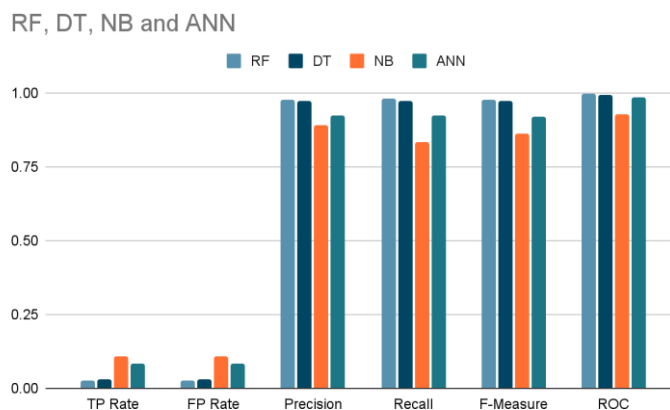
Model	<i>Accurate</i>	<i>TP Rate</i>	<i>FP Rate</i>	<i>Precision</i>	<i>Recall</i>	<i>F-Measure</i>	<i>ROC</i>
RF	91.13%	0.979	0.025	0.977	0.979	0.978	0.998
DT	90.50%	0.971	0.029	0.973	0.971	0.972	0.992
NB	70.97%	0.833	0.109	0.890	0.833	0.860	0.928
ANN	75.56%	0.903	0.106	0.901	0.903	0.902	0.968

Pada gambar 5 disajikan grafik akurasi algoritma-algoritma yang digunakan dalam melakukan klasifikasi pada penelitian ini diantaranya *Random Forest* (RF), *Decision Tree* (DT), *Naive Bayes* (NB), dan *Artificial Neural Network* (ANN) dengan *preprocessing* menggunakan normalisasi. Akurasi hasil dari pembelajaran mesin pada dataset *steel industry energy consumption* didapatkan akurasi terbesar adalah *random forest* 91.13 %



**Gambar 5.** Grafik Akurasi Algoritma Klasifikasi dengan Normalisasi

Pada gambar 6 disajikan grafik hasil pengujian berupa *TP Rate*, *FP Rate*, *Precision*, *Recall*, *F-Measure*, dan *ROC* algoritma-algoritma yang digunakan dalam melakukan klasifikasi pada penelitian ini diantaranya *Random Forest* (RF), *Decision Tree* (DT), *Naive Bayes* (NB), dan *Artificial Neural Network* (ANN) dengan *preprocessing* menggunakan normalisasi.



**Gambar 6.** Grafik Hasil Pengujian Algoritma Klasifikasi dengan Normalisasi

#### 4. KESIMPULAN

Sektor industri konsumsi energi relatif lebih tinggi dibandingkan sektor lainnya, sehingga perlu adanya pengendalian konsumsi energi terutama di sektor industri. Hal ini dapat diklasifikasi jenis bebannya dari faktor-faktor yang mendukung, beberapa parameter sangat dominan dalam hasil pengklasifikasian dari dataset *steel industry energy consumption* yang dipublikasikan oleh *UCI Repository* pada tahun 2021 yang terdiri dari 10 atribut dengan *class load type* yang bersifat kategori, dari pola yang didapat dari penelitian dihasilkan akurasi 91.13% paling besar menggunakan model *Random Forest* dengan *preprocessing* normalisasi dibandingkan dengan *preprocessing attribute selection* dan algoritma klasifikasi lain yaitu *Decision Tree*, *Naive Bayes*, dan *Artificial Neural Network*. Untuk penelitian selanjutnya diharapkan dapat menyajikan pembahasan mengenai implementasi dari pola yang sudah terbangun.

#### UCAPAN TERIMA KASIH

Rasa syukur kami panjatkan kepada Allah SWT. Terima kasih kami haturkan kepada Sathishkumar V E dari *Department of Information and Communication Engineering, Suncheon National University, Suncheon, Republic of Korea* yang telah menyediakan dataset publik yang kami gunakan dalam penelitian ini.

#### DAFTAR PUSTAKA

- [1] L. Garci Reyes, "Analisis kebutuhan listrik dan penambahan pembangkit listrik," *J. Chem. Inf. Model.*, vol. 53,



- no. 9, pp. 1689–1699, 2013.
- [2] Z. X. Wang, Q. Li, and L. L. Pei, “A seasonal GM(1,1) model for forecasting the electricity consumption of the primary economic sectors,” *Energy*, vol. 154, pp. 522–534, 2018, doi: 10.1016/j.energy.2018.04.155.
- [3] C. Shin, J. Park, and Y. Cho, “An Energy Consumption Prediction Model for Smart Factory Using Data Mining Algorithms,” *KIPS Trans. Softw. Data Eng.*, vol. 9, no. 5, pp. 153–160, 2020.
- [4] S. V E, C. Shin, and Y. Cho, “Efficient energy consumption prediction model for a data analytic-enabled industry building in a smart city,” *Build. Res. Inf.*, vol. 49, no. 1, pp. 127–143, 2021, doi: 10.1080/09613218.2020.1809983.
- [5] S. V E, J. Lim, M. Lee, K. Cho, J. Park, and C. Shin, “Industry Energy Consumption Prediction Using Data Mining Techniques,” *Int. J. Energy, Inf. Commun.*, vol. 11, no. 1, pp. 7–14, 2020, doi: 10.21742/ijeic.2020.11.1.02.
- [6] V. Bolón-Canedo, N. Sánchez-Marroño, and A. Alonso-Betanzos, “Feature selection and classification in multiple class datasets: An application to KDD Cup 99 dataset,” *Expert Syst. Appl.*, vol. 38, no. 5, pp. 5947–5957, 2011, doi: 10.1016/j.eswa.2010.11.028.
- [7] I. N. T. Wirawan and I. Eksistyanto, “Penerapan Naive Bayes Pada Intrusion Detection System Dengan Diskritisasi Variabel,” *JUTI J. Ilm. Teknol. Inf.*, vol. 13, no. 2, p. 182, 2015, doi: 10.12962/j24068535.v13i2.a487.
- [8] D. Pramadhana, “Klasifikasi Penyakit Diabetes Menggunakan Metode CFS dan ROS dengan Algoritma J48 Berbasis Adaboost,” *Edumatic J. Pendidik. Inform.*, vol. 5, no. 1, pp. 89–98, 2021, doi: 10.29408/edumatic.v5i1.3336.
- [9] D. A. Nasution, H. H. Khotimah, and N. Chamidah, “Perbandingan Normalisasi Data untuk Klasifikasi Wine Menggunakan Algoritma K-NN,” *Comput. Eng. Sci. Syst. J.*, vol. 4, no. 1, p. 78, 2019, doi: 10.24114/cess.v4i1.11458.
- [10] R. E. Putri, Suparti, and R. Rahmawati, “Perbandingan Metode Klasifikasi Naïve Bayes Dan K-Nearest Neighbor Pada Analisis Data Status Kerja Di Kabupaten Demak Tahun 2012,” *J. Gaussian*, vol. 3, no. 4, pp. 831–838, 2014.
- [11] B. A. Mukhtar, N. A. Setiawan, and T. B. Adji, “Analisis Perbandingan Tingkat Akurasi Algoritma Naive Bayes Classifier dengan Correlated-Naive Bayes Classifier,” *Semin. Nas. Teknol. Inf. dan Multimed. 2015*, pp. 49–54, 2015.
- [12] Gde Agung Brahmana Suryanegara, Adiwijaya, and Mahendra Dwifabri Purbolaksono, “Peningkatan Hasil Klasifikasi pada Algoritma Random Forest untuk Deteksi Pasien Penderita Diabetes Menggunakan Metode Normalisasi,” *J. RESTI (Rekayasa Sist. dan Teknol. Informasi)*, vol. 5, no. 1, pp. 114–122, 2021, doi: 10.29207/resti.v5i1.2880.
- [13] T. N. Nuklianggraita, A. Adiwijaya, and A. Aditsania, “On the Feature Selection of Microarray Data for Cancer Detection based on Random Forest Classifier,” *J. Infotel*, vol. 12, no. 3, pp. 89–96, 2020, doi: 10.20895/infotel.v12i3.485.
- [14] H. M. Nawawi, S. Rahayu, M. J. Shidiq, and J. J. Purnama, “Algoritma C4.5 Untuk Memprediksi Pengambilan Keputusan Memilih Deposito Berjangka,” *J. Techno Nuasa Mandiri*, vol. 16, no. 1, pp. 65–72, 2019.
- [15] J. J. Purnama, H. M. Nawawi, S. Rosyida, Ridwansyah, and Risandar, “Klasifikasi Mahasiswa Her Berbasis Algoritma Svm Dan Decision Classification of Her Students Based on Svm Algorithm and,” *J. Teknol. Inf. dan Ilmu Komput.*, vol. 7, no. 6, pp. 1253–1260, 2020, doi: 10.25126/jtiik.202073080.
- [16] S. Rahayu, J. Jaya Purnama, H. Mahmud Nawawi, F. Septia Nugraha, I. Komputer, and S. Nusa Mandiri Jakarta, “Algoritma Naive Bayes Classifier Untuk Memprediksi Gejala Autism Spectrum Disorders Pada Anak-Anak,” *Semin. Nas. Rekayasa dan Teknol.*, vol. 27, no. November, 2019.
- [17] I. Ali and L. Sularto, “Optimasi Parameter Artificial Neural Network Menggunakan Algoritma Genetika Untuk Prediksi Kelulusan Mahasiswa,” *J. ICT Inf. Commun. Technol.*, vol. 18, no. 1, pp. 54–59, 2019, doi: 10.36054/jict-ikmi.v18i1.52.
- [18] J. J. Purnama, S. Rahayu, S. Nurdiani, T. Haryanti, and N. A. Mayangky, “Analisis Algoritma Klasifikasi Neural Network Untuk Diagnosis Penyakit Diabetes,” *IJCIT (Indonesian J. Comput. Inf. Technol.)*, vol. 5, no. 1, 2020, doi: 10.31294/ijcit.v5i1.6391.
- [19] D. Cahyanti, A. Rahmayani, and S. A. Husniar, “Analisis performa metode Knn pada Dataset pasien pengidap Kanker Payudara,” *Indones. J. Data Sci.*, vol. 1, no. 2, pp. 39–43, 2020, doi: 10.33096/ijodas.v1i2.13.
- [20] M. J. Shidiq, S. Rahayu, and F. S. Nugraha, “Klasifikasi Diagnosis Melahirkan Dengan Metode Sesar Menggunakan Neural Network,” *J. Pilar Nusa Mandiri*, vol. 15, no. 2, pp. 157–162, 2019, doi: 10.33480/pilar.v15i2.602.
- [21] E. Emawati, W. Andriatna, and S. Syarofah, “Emma Emawati, Wendi Andriatna, Siti Syarofah 2017,” vol. IV, no.



- 2, pp. 14–19, 2017.
- [22] G. A. Buntoro, “Analisis Sentimen Calon Gubernur DKI Jakarta 2017 Di Twitter,” *Integer J.*, vol. 2, no. 1, pp. 32–41, 2017, [Online]. Available: <https://t.co/jrvaMsgBdH>
- [23] L. A. Andika, P. A. N. Azizah, and R. Respatiwan, “Analisis Sentimen Masyarakat terhadap Hasil Quick Count Pemilihan Presiden Indonesia 2019 pada Media Sosial Twitter Menggunakan Metode Naive Bayes Classifier,” *Indones. J. Appl. Stat.*, vol. 2, no. 1, p. 34, 2019, doi: 10.13057/ijas.v2i1.29998.
- [24] A. Primajaya and B. N. Sari, “Random Forest Algorithm for Prediction of Precipitation,” *Indones. J. Artif. Intell. Data Min.*, vol. 1, no. 1, pp. 27–31, 2018.
- [25] S. Dewi, “Komparasi 5 Metode Algoritma Klasifikasi Data Mining Pada Prediksi Keberhasilan Pemasaran Produk Layanan Perbankan,” *Techno Nusa Mandiri*, vol. 13, no. 1, pp. 60–66, 2016.