

# Analisis Pengaruh *Noise* pada Performa *K-Nearest Neighbors Algorithm* dengan Variasi Jarak untuk klasifikasi Beban Listrik

ARIS SURYA YUNATA<sup>1</sup>, ABDUL HALIM<sup>1</sup>, HILDA LUTHFIYAH<sup>2</sup>

<sup>1</sup>Departemen Teknik Elektro, Universitas Indonesia, Indonesia

<sup>2</sup>Badan Riset dan Inovasi Nasional, Indonesia

Email : aris.surya@ui.ac.id

Received 16 Mei 2024 | Revised 6 Juli 2024 | Accepted 9 Juli 2024

## ABSTRAK

*Teknik Non-Intrusive Load Monitoring (NILM) digunakan dalam pemantauan konsumsi energi. Variabel pengukuran yang digunakan yaitu Real Power dan Reactive Power. klasifikasi beban listrik menjadi acuan dalam mengurangi tagihan energi. Namun, data pengukuran sering kali terpengaruh oleh noise. Penelitian ini bertujuan untuk menganalisis pengaruh noise terhadap performa algoritma k-Nearest Neighbors (k-NN) dalam klasifikasi beban listrik. Berbagai tingkat noise secara random diberikan pada data pengukuran yang diperoleh. Selanjutnya, model k-NN dilatih dan dievaluasi dengan nilai k = 1 sampai 9 dan 15 tipe jarak. Hasil eksperimen menunjukkan bahwa penambahan noise pada data pengukuran secara signifikan mempengaruhi performa algoritma k-NN dalam mengklasifikasikan beban listrik. Pengaruh ini terlihat pada nilai akurasi tertinggi mayoritas pada k = 3 dan Tipe jarak Canberra menghasilkan nilai akurasi di atas rata-rata.*

**Kata kunci:** NILM, Real Power, Reactive Power, noise, k-NN

## ABSTRACT

*The Non-Intrusive Load Monitoring (NILM) technique is used in monitoring energy consumption. The measurement variables used are Real Power and Reactive Power. Electric load classification serves as a reference in reducing energy bills. However, measurement data is often affected by noise. This study aims to analyze the influence of noise on the performance of the k-Nearest Neighbors (k-NN) algorithm in electric load classification. Various levels of noise are randomly added to the obtained measurement data. Subsequently, the k-NN model is trained and evaluated with values of k = 1 to 9 and 15 distance types. The experimental results show that the addition of noise to the measurement data significantly affects the performance of the k-NN algorithm in classifying electric loads. This influence is observed in the highest accuracy values, mostly at k = 3, and the Canberra distance type yields accuracy values above average.*

**Keywords:** NILM, Real Power, Reactive Power, noise, k-NN

## 1. PENDAHULUAN

Di antara berbagai bentuk energi, listrik dianggap sebagai salah satu energi yang paling penting karena paling banyak digunakan oleh manusia. Akibatnya, kebutuhan listrik dalam beberapa tahun terakhir mengalami peningkatan yang signifikan. Penelitian **(McNeil, dkk, 2019)** memperkirakan beban listrik Indonesia sampai dengan tahun 2030 dapat meningkat tiga kali lipat menjadi 77,3 GW. Perangkat yang banyak mempengaruhi adalah *Air Conditioning* (AC), penerangan, dan lemari es. Penelitian lain menyebutkan pengaruh pandemi terhadap peningkatan beban listrik lebih besar 3,0 GJ dibandingkan sebelum pandemi, khususnya pada penggunaan AC dan kegiatan memasak **(Surahman, dkk, 2022)**.

Analisis statistik menunjukkan bahwa peningkatan pendapatan rumah tangga, luasan area rumah dan jumlah peralatan termasuk AC mempengaruhi peningkatan total konsumsi energi rumah tangga. Oleh karena itu, diperlukan pengelolaan listrik yang efektif dan efisien untuk mengurangi konsumsi dan biaya energi. Caranya dengan membangun jaringan listrik pintar (*Smart Grid*) yang dapat memberikan informasi total energi dikonsumsi dalam sebuah bangunan dan harga yang harus dibayarkan yang diberikan kepada pengguna. **(Santoso, dkk, 2018)** telah melakukan pengembangan sistem pemantauan energi khususnya untuk beban rumah tangga berbasis teknologi *Internet of Things* (IoT), sehingga dapat dilakukan pemantauan penggunaan energi listrik rumah tangga menggunakan aplikasi android di perangkat komunikasi telepon seluler (ponsel). Beberapa tahun ini telah dikembangkan sistem *Non-Intrusive Load Monitoring* (NILM) yang merupakan alternatif paling relevan untuk pemilahan energi, karena teknik ini menyediakan metode untuk memisahkan konsumsi individu untuk peralatan tertentu dan sering kali menggunakan alat pengukur pintar yang sudah digunakan **(Ruano, dkk, 2019)**.

Dalam dunia pengukuran *Smart Grid*, NILM maupun perangkat sejenis, keakuratan dan keandalan data sangat penting. Pengukuran konsumsi daya terbagi menjadi: pertama, sistem pengukuran pasif yaitu berdasarkan perolehan dan pemrosesan arus yang diserap oleh sistem. Kedua, sistem pengukuran aktif yang berdasarkan pengukuran respon terhadap sinyal frekuensi variabel yang disuntikkan ke dalam sistem **(Bucci, dkk, 2021)**. Pengukuran energi yang ada saat ini memberikan sangat sedikit atau tidak sama sekali informasi mengenai konsumsi energi masing-masing peralatan sehingga yang dapat dilakukan adalah dengan memilah seluruh beban listrik menjadi beban peralatan tersendiri dengan menggunakan *machine learning* **(Khan, dkk, 2019)**. Beberapa metode berbasis *machine learning* salah satunya *K-Nearest Neighbour* (k-NN) digunakan untuk memilah data *supervised* NILM **(Shabbir, dkk, 2024)**. k-NN merupakan metode ampuh tertua dan paling sederhana untuk klasifikasi menggunakan teknik klasifikasi berbasis memori dimana sebagian besar Jarak Euclidean diperlukan untuk algoritma k-NN **(C, dkk, 2019)**. Penelitian **(Li, dkk, 2021)** melakukan ekstraksi fitur pada sistem NILM berbasis waktu, yang secara langsung memengaruhi kinerja pengenalan model pengklasifikasi. Fitur bidang daya (P-Q) digunakan untuk mengidentifikasi peralatan listrik, dimana P adalah *Real Power* dan Q adalah *Reactive Power*.

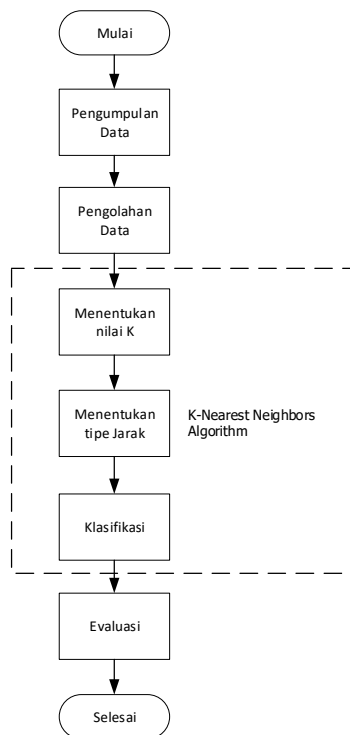
Pada penelitian **(Hidiyanto & Halim, 2020)** telah melakukan klasifikasi beban listrik dengan data AMPds2 dengan metode *machine learning* k-NN dengan variasi 9 jarak tanpa dipengaruhi *noise*. Padahal *noise* adalah salah satu tantangan utama yang sering dihadapi dalam pengukuran. *Noise* dihasilkan dari kesalahan pengukuran dimana selisih antara nilai sebenarnya dengan nilai terukur. Nilai sebenarnya adalah rata-rata dari jumlah pengukuran yang tak terhingga, dan nilai terukur adalah nilai presisi **(Mohajan, 2017)**. Maka pada penelitian ini difokuskan pada pengaruh *noise* pada data AMPds2 dan mengukur performa

algoritma k-NN melalui variasi 15 jarak. Dengan memahami dampak *noise* pada ketepatan metode NILM, penelitian ini bertujuan untuk mengevaluasi keefektifan k-NN dalam meningkatkan akurasi klasifikasi beban perangkat listrik yang diterapkan *noise*. Selain itu, diharapkan bahwa hasil penelitian ini tidak hanya memberikan wawasan mendalam tentang peran k-NN dalam menghadapi *noise* dalam data AMPds2, tetapi juga memberikan kontribusi terhadap pengembangan sistem pemantauan beban listrik yang lebih handal.

## 2. METODE

### 2.1 Gambaran Umum

Pada penelitian ini, kami menganalisa apakah *noise* mempengaruhi performa dari Algoritma k-NN dalam melakukan klasifikasi Beban Listrik. Oleh sebab itu kami melakukan langkah langkah penelitian sesuai dengan diagram alir sebagai berikut :

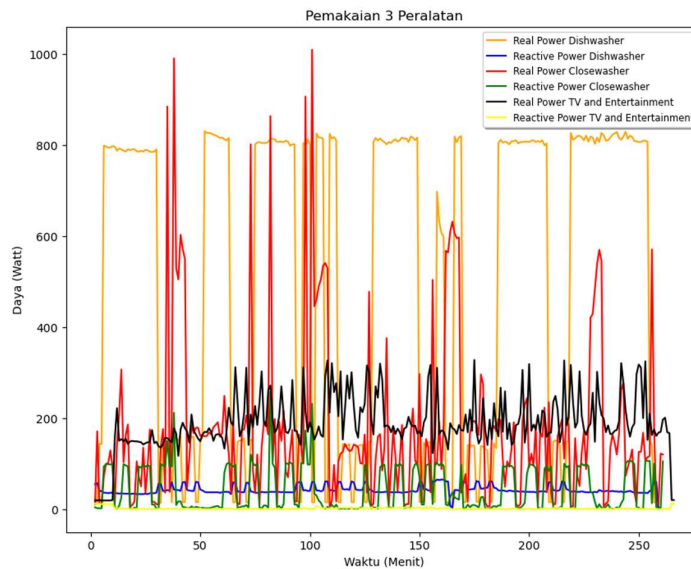


**Gambar 1. Diagram Alir Klasifikasi Beban Listrik**

Berdasarkan diagram alir pada Gambar 1 di atas, terlihat langkah – langkah penelitian dari awal sampai akhir. Penelitian dimulai dengan melakukan pengumpulan data AMPds2 yang diambil dari sumber secara online (**Makonin, 2016**). langkah selanjutnya yaitu pengolahan data dimana dilakukan pembagian data menjadi data *training* dan data *testing*. Kemudian untuk data *testing* dilakukan modifikasi data sebanyak 10 data dimana data *testing* yang baru merupakan data *testing* yang ditambahkan *random noise* dari 1 sampai 10 %. Setelah data *testing* terbentuk , langkah selanjutnya yaitu klasifikasi Beban Listrik dengan metode k-NN yang menggunakan bahasa pemrograman Python. Pada proses klasifikasi dengan metode k-NN, nilai k yang digunakan adalah dari 1 sampai 9 dan tipe perhitungan jarak adalah 15 tipe. Setelah proses klasifikasi dilakukan, langkah berikutnya yaitu melakukan evaluasi. Evaluasi dilakukan dengan membuat *confusion matrix*, hasil dari *confusion matrix* digunakan untuk menghitung nilai akurasi.

## 2.2 Pengumpulan Data

Dataset Almanac of Minutely Power (AMPds) berisi data satu tahun (**Makonin, dkk, 2013**). Sedangkan AMPds2 berisi data dua tahun yang mencakup tiga jenis konsumsi utama (listrik, air, dan gas alam) (**Makonin, dkk, 2016**). AMPds dan AMPds2 mempunyai 11 variabel pengukuran untuk 21 perangkat menggunakan interval pengukuran setiap satu menit. Dalam penelitian ini, Kami mengumpulkan 3 data peralatan dari data AMPds2: dataset The Almanac of Minutely Power (versi 2) yang dipublikasikan secara online (**Makonin, 2016**). Ketiga peralatan yang digunakan adalah *Dishwasher*, *Closewasher* dan *TV and Entertainment* dengan Variabel Pengukuran yang digunakan adalah *Real Power* dan *Reactive Power*. Gambar 2 adalah grafik pemakaian untuk 3 peralatan.



**Gambar 2. Grafik Penggunaan 3 Peralatan**

## 2.3 Pengolahan Data

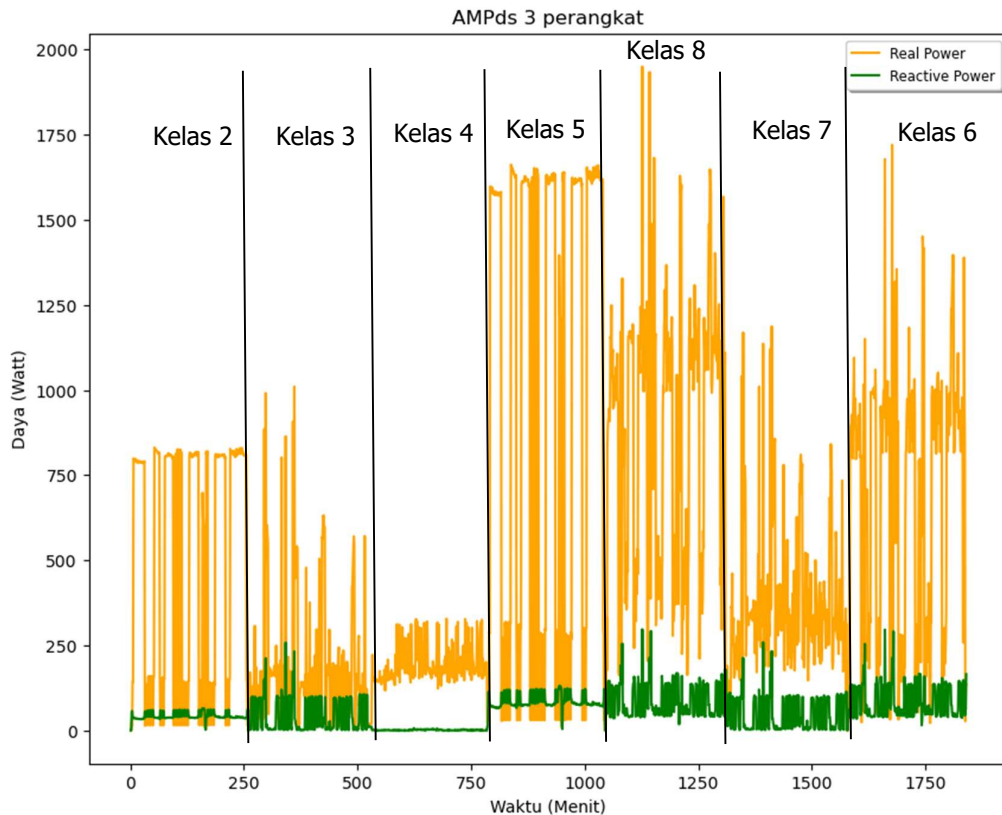
Dari data AMPds2 yang dikumpulkan, data kemudian diolah dengan kombinasi *On* dan *Off* dari ketiga perangkat secara waktu berdasarkan nilai *Real Power* dan *Reactive Power*. Berikut kombinasi peralatan yang dihasilkan ada 8 kelas

**Tabel 1. Klasifikasi Data AMPds2**

Kelas	Keterangan
1	Semua Peralatan Off
2	Dishwasher On
3	Clothes Washer On
4	TV On
5	Dishwasher dan TV On
6	Dishwasher dan Clothes Washer On
7	Clothes Washer dan TV On
8	Semua Peralatan On

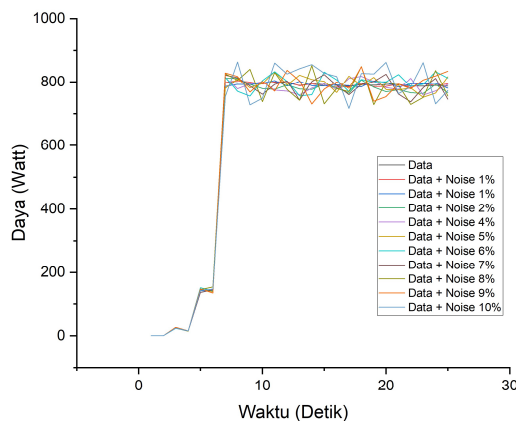
Dari pengolahan data diperoleh data yang akan dijadikan acuan untuk melakukan pengujian sebanyak 1840 data yang grafiknya ditunjukkan pada Gambar 3, yaitu grafik data dalam dua dimensi yang menunjukkan data yang tumpang tindih dari nilai *Real Power* dan *Reactive Power* yang serupa. (**Hidiyanto & Halim, 2020**).

Analisis Pengaruh *Noise* pada Performa K-Nearest Neighbors Algorithm dengan Variasi Jarak untuk Klasifikasi Beban Listrik



Gambar 3. Konsumsi Daya Gabungan

Dari data *testing*, di olah kembali untuk di berikan *rundom noise* (Fawzi, dkk, 2016) sebesar 1 sampai 10 persen dari nilai data. Dengan demikian didapatkan data baru sebanyak 10 data dengan variasi penambahan *noise* 1,2,3,4,5,6,7,8,9 dan 10 persen. Perbandingan Data baru yang dihasilkan dapat dilihat pada grafik dibawah ini (yang ditampilkan hanya 25 data pertama dari 1840 data) :



Gambar 4. Grafik Perbandingan Sampel Data *Testing* dengan *Noise*

Pada Gambar 4 menunjukan grafik daya terhadap waktu untuk perbandingan sampel data *testing* dengan *noise* dimana semakin besar *noise* yang diberikan mengakibatkan fluktuasi yang besar

**Tabel 2. Formula Perhitungan Jarak**

No	Jarak	Keterangan
1	<i>Manhattan Distance</i> (MD)	$MD(x, y) = \sum_{i=1}^n  x_i - y_i $
2	<i>Lorentzian Distance</i> (LD)	$LD(x, y) = \sum_{i=1}^n \ln(1 +  x_i - y_i )$
3	<i>Canberra Distance</i> (CanD)	$CanD(x, y) = \sum_{i=1}^n \frac{ x_i - y_i }{ x_i  +  y_i }$
4	<i>Clark Distance</i> (ClaD)	$ClaD(x, y) = \sqrt{\sum_{i=1}^n \left( \frac{x_i - y_i}{ x_i  +  y_i } \right)^2}$
5	<i>Divergence Distance</i> (DivD)	$DivD(x, y) = 2 \sum_{i=1}^n \frac{(x_i - y_i)^2}{(x_i + y_i)^2}$
6	<i>Squared Chi Squared Distance</i> (SCSD)	$SCSD(x, y) = \sum_{i=1}^n \frac{(x_i - y_i)^2}{ x_i + y_i }$
7	<i>Average Distance</i> (AvgD)	$AvgD(x, y) = \sqrt{\frac{1}{n} \sum_{i=1}^n (x_i - y_i)^2}$
8	<i>Euclidean Distance</i> (ED)	$ED(x, y) = \sqrt{\sum_{i=1}^n  x_i - y_i ^2}$
9	<i>Mean Character Distance</i> (MCD)	$MCD(x, y) = \frac{\sum_{i=1}^n  x_i - y_i }{n}$
10	<i>Matusita Distance</i> (MatD)	$MatD(x, y) = \sqrt{\sum_{i=1}^n (\sqrt{x_i} - \sqrt{y_i})^2}$
11	<i>Squared Chord Distance</i> (SCD)	$SCD(x, y) = \sum_{i=1}^n (\sqrt{x_i} - \sqrt{y_i})^2$
12	<i>Hellinger Distance</i> (HeD)	$HeD(x, y) = \sqrt{2 \sum_{i=1}^n (\sqrt{x_i} - \sqrt{y_i})^2}$
13	<i>Squared Euclidean Distance</i> (SED)	$SED(x, y) = \sum_{i=1}^n (x_i - y_i)^2$
14	<i>Neyman Chi Squared Distance</i> (NCSD)	$NCSD(x, y) = \sum_{i=1}^n \frac{(x_i - y_i)^2}{x_i}$
15	<i>Pearson Chi Squared Distance</i> (PCSD)	$PCSD(x, y) = \sum_{i=1}^n \frac{(x_i - y_i)^2}{y_i}$

#### 2.4 K-Nearest Neighbors (K-NN)

K-Nearest Neighbors (K-NN) merupakan salah satu metode dalam *machine learning* yang digunakan untuk klasifikasi dan regresi. Metode ini bekerja dengan cara menemukan *k*-

*neighbors* (tetangga terdekat) dari suatu data point berdasarkan ukuran jarak, dan kemudian mengambil mayoritas kelas (untuk klasifikasi) atau nilai rata-rata (untuk regresi) dari tetangga tersebut (**Primartha, 2018**).

Berikut adalah langkah-langkah umum dalam metode K-NN:

1. Menentukan jumlah tetangga terdekat ( $k$ )
2. Menghitung jarak antara data point yang akan diprediksi Hitung Jarak
3. Menentukan kelas mayoritas
4. Menggunakan hasil mayoritas atau rata-rata untuk membuat prediksi.

Dalam penelitian ini penentuan nilai  $k$  menggunakan teknik *grid search* untuk mencoba berbagai nilai  $k$  dan memilih nilai yang memberikan hasil terbaik berdasarkan akurasi. Hasil teknik ini memberikan nilai  $k$  yang berbeda yaitu 1, 2, 3, 4, 5, 6, 7, 8 dan 9.

## 2.5 Perhitungan Jarak

Perhitungan jarak suatu data merupakan komponen terpenting dari hasil proses algoritme K-NN. Perhitungan jarak digunakan untuk mencocokkan antara data pelatihan dengan data baru sebagai pembandingan (**Wahyono, dkk, 2020**). Dalam tulisan ini, kami menggunakan 15 pengukuran jarak yang merupakan bagian dari 8 keluarga jarak yang berisi total 54 jenis jarak (**Prasath, dkk, 2019**). Pemilihan metrik jarak bergantung pada aplikasi spesifik dan sifat data untuk sebagian besar data kontinu yang mengutamakan jarak garis lurus sehingga memberikan fleksibilitas untuk menyesuaikan perhitungan jarak berdasarkan parameter, mempertimbangkan korelasi antara variabel dan skala dari data serta menjadikannya ideal untuk data multidimensional yang memiliki fitur berkorelasi. Tabel 2 menunjukkan formula perhitungan 15 jarak yang digunakan dalam penelitian ini.

## 2.6 Evaluation Metrics

*Evaluation metrics* atau metrik evaluasi, merujuk pada ukuran-ukuran yang digunakan untuk menilai sejauh mana model *machine learning* atau sistem memiliki kinerja yang baik dalam melakukan tugas tertentu (**Yang, dkk, 2018**). Metrik evaluasi membantu dalam memahami seberapa baik model dapat membuat prediksi atau klasifikasi dengan membandingkan hasil prediksi dengan nilai yang sebenarnya atau label yang diketahui.

		Aktual	
		Positif	Negatif
Prediksi	Positif	True Positif (TP)	False Positif (FP)
	Negatif	False Negatif (FN)	True Negatif (TR)

**Gambar 5. Confusion Matrix**

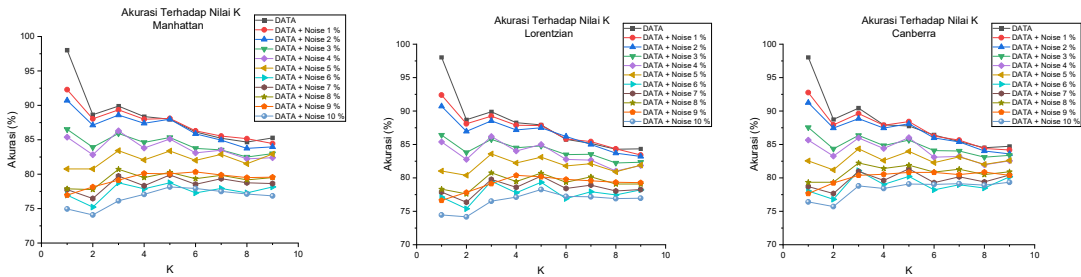
Gambar 5 adalah *Confusion Matrix* terdiri dari empat komponen utama yaitu True Positive (TP) adalah jumlah prediksi yang benar untuk kelas positif. True Negative (TN) adalah jumlah prediksi yang benar untuk kelas negatif. False Positive (FP) adalah jumlah prediksi yang salah di mana kelas negatif diprediksi sebagai positif dan False Negative (FN) adalah jumlah prediksi yang salah di mana kelas positif diprediksi sebagai negatif. Akurasi merupakan sebuah ukuran kinerja yang bisa dipakai untuk melihat keseluruhan kinerja dari sebuah *classifier*, yang mana didefinisikan sebagai perbandingan antara hasil tes yang benar terhadap keseluruhan data tes.

Berikut metrik evaluasi Akurasi yang digunakan :

$$\text{Akurasi (Accuracy)} = \frac{TP+TN}{TP+TN+FP+FN} \quad (1)$$

### 3. HASIL DAN PEMBAHASAN

Pada simulasi yang kami lakukan, kami melakukan pengujian pada data *testing*. Data *testing* terdiri dari 1 data yang tidak diberikan *noise* dan 10 data yang diberikan *noise* 1 sampai 10 persen. variasi jarak yang digunakan adalah 15 tipe dan variasi k yang digunakan adalah 1 sampai 9 untuk setiap pengujian.

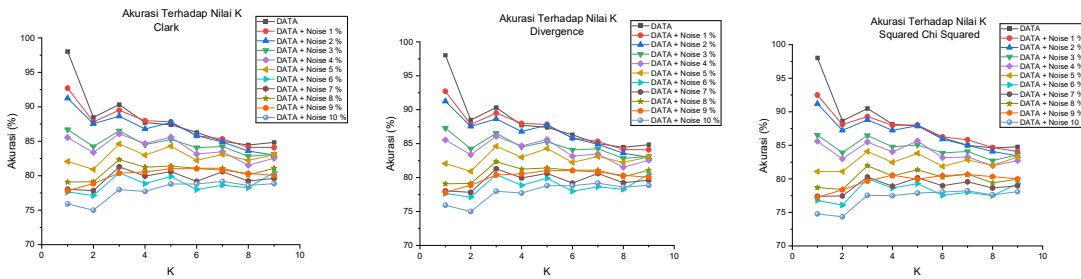


(a)

(b)

(c)

Gambar 6. Grafik Akurasi terhadap k (a) Manhattan, (b) Lorentzian dan (c) Canberra

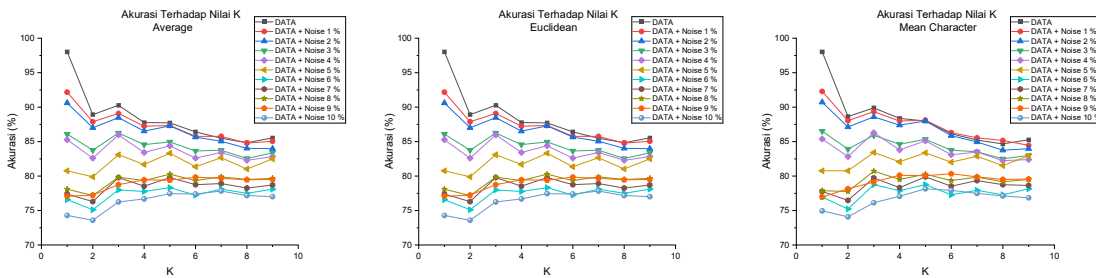


(a)

(b)

(c)

Gambar 7. Grafik Akurasi terhadap k (a) Clark, (b) Divergence dan (c) Squared Chi Squared



(a)

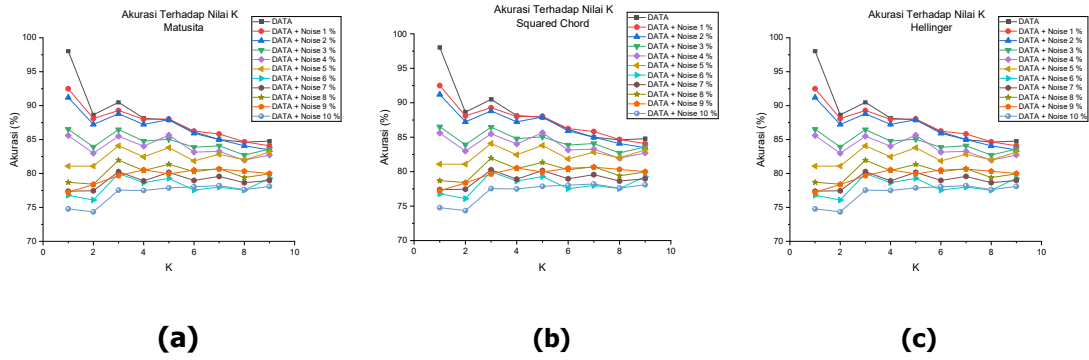
(b)

(c)

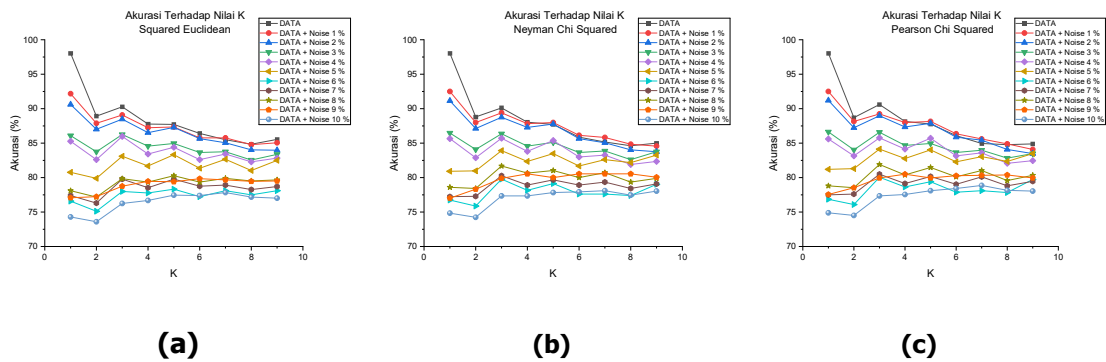
Gambar 8. Grafik Akurasi terhadap k (a) Average, (b) Euclidean dan (c) Mean Character



Analisis Pengaruh *Noise* pada Performa K-Nearest Neighbors Algorithm dengan Variasi Jarak untuk Klasifikasi Beban Listrik

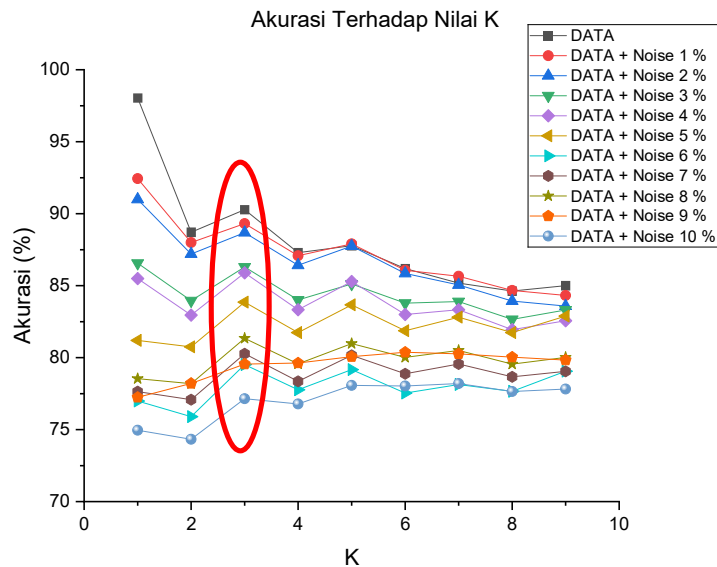


Gambar 9. Grafik Akurasi terhadap k (a) Matusita, (b) Squared chord dan (c) Hellinger



Gambar 10. Grafik Akurasi terhadap k (a) Squared Euclidean, (b) Neyman Chi Squared dan (c) Pearson Chi Squared

Gambar 6 sampai 10 merupakan grafik pola akurasi yang dihasilkan 15 tipe jarak. Dari ke 15 data akurasi yang dihasilkan, kemudian dibuat nilai rata-rata. Berikut grafik yang didapatkan dari nilai rata-rata.



Gambar 11. Grafik Akurasi Rata - Rata terhadap k

Pada Grafik di atas menunjukkan nilai akurasi terhadap nilai  $k$ . Dengan mengabaikan nilai  $k = 1$ , maka nilai akurasi tertinggi untuk setiap data *testing* adalah mayoritas pada  $k = 3$  (dapat dilihat pada grafik) dan semakin besar nilai  $k$  maka akurasi semakin rendah. Hal ini disebabkan karena nilai  $K$  yang besar menggabungkan lebih banyak tetangga dalam proses klasifikasi dan menjadi lebih kompleks karena mempertimbangkan banyak tetangga dalam klasifikasi. nilai  $K$  yang besar dapat mengaburkan keputusan klasifikasi dan mengakibatkan klasifikasi yang kurang tepat

Selain itu nilai akurasinya semakin kecil seiring dengan bertambahnya *noise*. Hal ini disebabkan oleh karakteristik K-NN yang sangat bergantung pada kedekatan atau jarak antara data *testing* dan data *training*. Ketika ada banyak *noise*, jarak antara data menjadi lebih tidak konsisten, sehingga kesalahan prediksi meningkat dan nilai akurasi menurun. *Noise* dapat menyebabkan model overfitting jika tidak diatasi dengan baik.

Berikut nilai akurasi untuk  $k = 3$  setiap 15 tipe pengukuran jarak

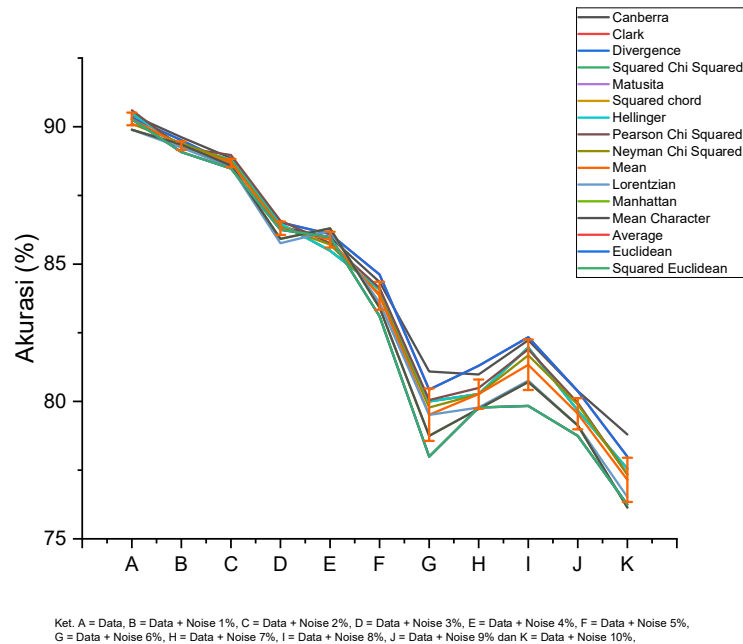
**Tabel 2. Akurasi dengan  $K = 3$  untuk 15 Tipe Jarak**

Akurasi	Noise										
	0%	1%	2%	3%	4%	5%	6%	7%	8%	9%	10%
Jarak	0%	1%	2%	3%	4%	5%	6%	7%	8%	9%	10%
Manhattan	89.89	89.35	88.59	85.92	86.30	83.42	78.75	79.73	80.71	79.13	76.14
Lorentzian	89.89	89.24	88.53	85.76	86.20	83.59	79.51	79.78	80.76	79.13	76.52
Canberra	90.43	89.62	88.86	86.36	85.98	84.35	81.09	80.98	82.23	80.38	78.80
Clark	90.33	89.51	88.64	86.52	86.09	84.62	80.43	81.30	82.34	80.38	77.99
Divergence	90.33	89.51	88.64	86.52	86.09	84.62	80.43	81.30	82.34	80.38	77.99
Squared Chi Squared	90.49	89.29	88.80	86.47	85.49	84.08	80.00	80.27	81.96	79.67	77.55
Average	90.27	89.08	88.48	86.25	85.98	83.10	77.99	79.78	79.84	78.75	76.25
Euclidean	90.27	89.08	88.48	86.25	85.98	83.10	77.99	79.78	79.84	78.75	76.25
Mean Character	89.89	89.35	88.59	85.92	86.30	83.42	78.75	79.73	80.71	79.13	76.14
Matusita	90.49	89.29	88.80	86.47	85.49	84.08	80.00	80.27	81.96	79.67	77.55
Squared chord	90.49	89.29	88.80	86.47	85.49	84.08	80.00	80.27	81.96	79.67	77.55
Hellinger	90.49	89.29	88.80	86.47	85.49	84.08	80.00	80.27	81.96	79.67	77.55
Squared Euclidean	90.27	89.08	88.48	86.25	85.98	83.10	77.99	79.78	79.84	78.75	76.25
Neyman Chi Squared	90.11	89.40	88.75	86.36	85.71	83.91	79.78	80.27	81.68	79.89	77.34
Pearson Chi Squared	90.60	89.24	88.97	86.58	85.76	84.13	80.05	80.49	81.90	79.95	77.34
Mean	90.28	89.31	88.68	86.30	85.89	83.85	79.52	80.27	81.34	79.55	77.15
Standard Deviasi	0.23	0.15	0.15	0.24	0.29	0.51	0.95	0.53	0.92	0.57	0.80

Dari Tabel 2 terlihat bahwa nilai akurasi tertinggi untuk data tanpa *noise* adalah sebesar 90,60 % dengan menggunakan tipe jarak Pearson Chi Squared, untuk data dengan *noise* 1% adalah sebesar 89,62 % dengan menggunakan tipe jarak Canberra, untuk data dengan *noise* 2% adalah sebesar 88,97 % dengan menggunakan tipe jarak Pearson Chi Squared, untuk data dengan *noise* 3% adalah sebesar 86,58 % dengan menggunakan tipe jarak Pearson Chi Squared, untuk data dengan *noise* 4% adalah sebesar 86,58 % dengan menggunakan tipe jarak Manhattan dan Mean Character, untuk data dengan *noise* 5% adalah sebesar 84,62 % dengan menggunakan tipe jarak Clark dan Divergence, untuk data dengan *noise* 6% adalah sebesar 81,09 % dengan menggunakan tipe jarak Canberra, untuk data dengan *noise* 7% adalah sebesar 81,30 % dengan menggunakan tipe jarak Clark dan Divergence, untuk data dengan *noise* 8% adalah sebesar 82,34 % dengan menggunakan tipe jarak Clark dan Divergence, untuk data dengan *noise* 9% adalah sebesar 80,38 % dengan menggunakan tipe jarak Canberra, Clark dan Divergence sedangkan untuk data dengan *noise* 10% adalah sebesar 78,80 % dengan menggunakan tipe jarak Canberra.

## Analisis Pengaruh *Noise* pada Performa K-Nearest Neighbors Algorithm dengan Variasi Jarak untuk Klasifikasi Beban Listrik

Dari hasil perhitungan akurasi tiap tipe jarak, didapatkan nilai rata-rata dan nilai standar deviasi, berikut grafik nilai rata-rata yang diberikan nilai standar deviasi terhadap 15 tipe perhitungan jarak.



**Gambar 12. Grafik Perbandingan Akurasi terhadap Data untuk 15 tipe jarak**

Pada Gambar 12 terlihat untuk nilai Akurasi yang diatas rata – rata, tipe jarak Canberra menempati posisi teratas diikuti tipe jarak Clark dan Divergence, kemudian diurutan ketiga tipe jarak Squared Chi Squared, Matusita, Squared chord dan Hellinger. Di urutan keempat ada tipe jarak Pearson Chi Squared dan di urutan kelima tipe jarak Neyman Chi Squared. Sedangkan untuk nilai Akurasi yang dibawah rata – rata, di urutan pertama ada tipe jarak Lorentzian, diurutan kedua ada tipe jarak Manhattan dan Mean Character dan yang terbawah adalah tipe jarak Average, Euclidean dan Squared Euclidean

Berdasarkan Tabel 2, didapatkan nilai selisih akurasi dari tipe jarak Canberra dengan nilai rata – rata. untuk data tanpa *noise* sebesar 0,15%, untuk data dengan *noise* 1% sebesar 0,31%, untuk data dengan *noise* 2% sebesar 0,18%, untuk data dengan *noise* 3% sebesar 0,06%, untuk data dengan *noise* 4% sebesar 0,09%, untuk data dengan *noise* 5% sebesar 0,50%, untuk data dengan *noise* 6% sebesar 1,57%, untuk data dengan *noise* 7% sebesar 0,71%, untuk data dengan *noise* 8% sebesar 0,89%, untuk data dengan *noise* 9% sebesar 0,83% dan untuk data dengan *noise* 10% sebesar 1,65%.

## 4. KESIMPULAN

Makalah ini menjelaskan mengenai pengaruh *noise* terhadap performa dari k-NN dalam melakukan klasifikasi Beban Listrik. Dari hasil pengujian dapat disimpulkan sebagai berikut Pertama, algoritma k-NN yang didasarkan pada jumlah k. dimana semakin besar nilai k maka nilai akurasi yang dihasilkan semakin kecil. Kedua, telah diidentifikasi dengan tingkat *noise* memiliki dampak yang signifikan terhadap akurasi yang dihasilkan. Dari hasil pengujian dan analisis data yang dilakukan, dapat disimpulkan bahwa *noise* memiliki korelasi yang jelas dengan performa k-NN dimana semakin kecil *noise* yang diberikan ke data *testing* maka nilai

akurasi yang dihasilkan semakin besar dan semakin besar *noise* yang diberikan ke data *testing* maka nilai akurasi yang dihasilkan semakin kecil. Ketiga, dari ke 15 tipe pengukuran jarak yang digunakan, didapatkan bahwa Cambera merupakan tipe pengukuran jarak yang menghasilkan akurasi di atas rata rata dibandingkan dengan tipe pengukuran jarak lainnya.

## DAFTAR RUJUKAN

- Bucci, G., Ciancetta, F., Fiorucci, E., Mari, S., & Fioravanti, A. (2021). Measurements for non-intrusive load monitoring through machine learning approaches. *Acta IMEKO*, *10*(4), 90–96. [https://doi.org/10.21014/acta\\_imeko.v10i4.1184](https://doi.org/10.21014/acta_imeko.v10i4.1184)
- C, S., Kumar. M., & Karuppasamy, I. (2019). Design and Implementation of Non-Intrusive Load Monitoring using Machine Learning Algorithm for Appliance Monitoring. *2019 IEEE International Conference on Intelligent Techniques in Control, Optimization and Signal Processing (INCOS)*, (pp. 1–6). <https://doi.org/10.1109/INCOS45849.2019.8951312>
- Fawzi, A., Moosavi-Dezfooli, S., & Frossard, P. (2016). Robustness of classifiers: from adversarial to random *noise*. Conference on Neural Information Processing Systems (NIPS).
- Hidiyanto, F., & Halim, A. (2020). K-NN Methods with Varied k, Distance and *Training* Data to Disaggregate NILM with Similar Load Characteristic. *Asia Pacific Conference on Research in Industrial and Systems Engineering (APCORISE)*, (pp. 93–99). <https://doi.org/10.1145/3400934.3400953>
- Iqbal, H. K., Malik, F. H., Muhammad, A., Qureshi, M. A., Abbasi, M. N., & Chishti, A. R. (2021). A critical review of state-of-the-art non-intrusive load monitoring datasets. *Electric Power Systems Research*, *192*, 106921. <https://doi.org/10.1016/j.epsr.2020.106921>
- Khan, M. M. R., Siddique, Md. A. B., & Sakib, S. (2019). Non-Intrusive Electrical Appliances Monitoring and Classification using K-Nearest Neighbors. *2019 2nd International Conference on Innovation in Engineering and Technology (ICIET)*, (pp. 1–5). <https://doi.org/10.1109/ICIET48527.2019.9290671>
- Li, Y., Wang, H., Yang, J., Wang, K., & Qi, G. (2021). A non-intrusive load monitoring algorithm based on multiple features and decision fusion. *Energy Reports*, *7*, 1555–1562. <https://doi.org/10.1016/j.egyr.2021.09.087>
- Makonin, S., Ellert, B., Bajić, I. V., & Popowich, F. (2016). Electricity, water, and natural gas consumption of a residential house in Canada from 2012 to 2014. *Scientific Data*, *3*. <https://doi.org/10.1038/sdata.2016.37>

- Makonin, S., Popowich, F., Bartram, L., Gill, B., & Bajic, I. V. (2013). AMPds: A public dataset for load disaggregation and eco-feedback research. *2013 IEEE Electrical Power & Energy Conference*, (pp. 1–6). <https://doi.org/10.1109/EPEC.2013.6802949>
- Makonin, Stephen. (2016). AMPds2: The Almanac of Minutely Power dataset (Version2). Retrieved from [dataverse.harvard.edu/dataset.xhtml?persistentId=doi:10.7910/DVN/FIE0S4](http://dataverse.harvard.edu/dataset.xhtml?persistentId=doi:10.7910/DVN/FIE0S4).
- McNeil, M. A., Karali, N., & Letschert, V. (2019). Forecasting Indonesia's electricity load through 2030 and peak demand reductions from appliance and lighting efficiency. *Energy for Sustainable Development*, *49*, 65–77. <https://doi.org/10.1016/j.esd.2019.01.001>
- Mohajan, H. K. (2017). Two Criteria for Good Measurements in Research: Validity and Reliability Two Criteria for Good Measurements in Research: Validity and Reliability. *Munich Personal RePEc Archive*, *17*(3), 58–82.
- Prasath, V. B. S., Abu Alfeilat, H. A., Hassanat, A. B. A., Lasassmeh, O., Tarawneh, A. S., Alhasanat, M. B., & Eyal Salman, H. S. (2019). Effects of Distance Measure Choice on K-Nearest Neighbor Classifier Performance: A Review. *Big Data*, *7*(4), 221–248. <https://doi.org/10.1089/big.2018.0175>
- Primartha, R. (2018). Belajar Machine Learning Teori dan Praktik. Penerbit Informatika.
- Ruano, A., Hernandez, A., Ureña, J., Ruano, M., & Garcia, J. (2019). NILM techniques for intelligent home energy management and ambient assisted living: A review. *Energies* *12*(11). MDPI AG. <https://doi.org/10.3390/en12112203>
- Santoso, H. B., Prajogo, S., & Mursid, S. P. (2018). Pengembangan Sistem Pemantauan Konsumsi Energi Rumah Tangga Berbasis Internet of Things (IoT). *ELKOMIKA: Jurnal Teknik Energi Elektrik, Teknik Telekomunikasi, & Teknik Elektronika*, *6*(3), 357. <https://doi.org/10.26760/elkomika.v6i3.357>
- Shabbir, N., Vassiljeva, K., Nourollahi Hokmabad, H., Husev, O., Petlenkov, E., & Belikov, J. (2024). Comparative Analysis of Machine Learning Techniques for Non-Intrusive Load Monitoring. *Electronics*, *13*(8), 1420. <https://doi.org/10.3390/electronics13081420>
- Surahman, U., Hartono, D., Setyowati, E., & Jurizat, A. (2022). Investigation on household energy consumption of urban residential buildings in major cities of Indonesia during COVID-19 pandemic. *Energy and Buildings*, *261*. <https://doi.org/10.1016/j.enbuild.2022.111956>
- Wahyono, W., Trisna, I. N. P., Sariwening, S. L., Fajar, M., & Wijayanto, D. (2020). Comparison of distance measurement on k-nearest neighbour in textual data classification. *Jurnal*

*Teknologi dan Sistem Komputer*, 8(1), 54–58.

<https://doi.org/10.14710/jtsiskom.8.1.2020.54-58>

Yang, C. C., Soh, C. S., & Yap, V. V. (2018). A systematic approach in appliance disaggregation using k-nearest neighbours and naive Bayes classifiers for energy efficiency. *Energy Efficiency*, 11(1), 239–259. <https://doi.org/10.1007/s12053-017-9561-0>