

Klasifikasi Beban Listrik dengan Machine Learning Menggunakan Metode K-Nearest Neighbor

Salma¹, Favian Dewanta², Muhammad Abdillah³

¹²Prodi S1 Teknik Telekomunikasi, Fakultas Teknik Elektro, Universitas Telkom

³Prodi S1 Teknik Elektro, Fakultas Teknik Elektro, Universitas Pertamina

¹²Telkom University, Jl. Telekomunikasi No.1, Terusan Buah Batu – Bojongsong, Sukapura, Dayeuhkolot, Bandung, Jawa Barat

³Universitas Pertamina, Jl. Teuku Nyak Arief, Simprug – Kby. Lama, Jakarta Selatan, Daerah Khusus Ibu Kota Jakarta
¹salmaa@student.telkomuniversity.ac.id, ²favian@telkomuniversity.co.id, ³m.abdillah@universitaspertamina.ac.id

ABSTRAK

Sistem pengenalan beban listrik merupakan peran yang penting dalam melakukan manajemen dan penghematan listrik. Sistem pengenalan beban listrik memiliki keandalan dalam memperoleh informasi yang relevan dari setiap beban listrik. Pada penelitian ini, akan dilakukan analisis karakteristik berbagai fitur beban listrik secara independen dan hubungan pencocokan antara fitur yang berbeda dan algoritma klasifikasi yang berbeda. Yaitu dengan melakukan perbandingan antara dua algoritma klasifikasi K-Nearest Neighbours dan Regresi Logistik Multinomial dari segi akurasi dan kecepatan proses dalam analisis. Sistem klasifikasi akan mengidentifikasi jenis, model, dan prasyarat yang tidak diketahui dari beban listrik dan mengelompokkannya. Karakteristik kelistrikan dari beban listrik yang akan diteliti antarlain besarnya tegangan dan arus root mean square, gelombang harmonisa, daya dan faktor daya dari variasi sample beban listrik yang berbeda. Hasil penelitian pada metode k-Nearest Neighbours didapatkan akurasi sebesar 99.619% sedangkan dengan metode Regresi Logistik Multinomial didapatkan akurasi sebesar 91.125%

Kata Kunci : Beban Listrik, Klasifikasi, K-Nearest Neighbours, Regresi Logistik Multinomial

ABSTRACT

The electrical load recognition system plays an important role in managing and saving electricity. In this study, the caharacteristic of various electrical load features in independent condition and the matching relationship between different features and different classification algorithm will be analyzed by doing a comparison between two classification algorithms, k-Nearest Neighbours and Multinomial Logistic Regression in terms of accuracy and speed of analysis process. The classification system will be identify unknown types, models and prerequisites of different electrical loads and classify them. The characteristics of the electrical load that will be analyzed include the magnitude of root mean square voltage and current, harmonic waves, power series, and power factor from variety of different electrical load samples. The results of the research on the k-Nearest Neighbors method obtained an accuracy of 99.619% while the Multinomial Logistics Regression method obtained an accuracy of 91.125%.

Keywords: Electrical Load, Classification, Machine Learning, k-Nearest Neighbours, Multinomial Logistic Regression

1 PENDAHULUAN

Sistem monitoring dan pengenalan beban listrik merupakan peran yang penting dalam melakukan manajemen dan penghematan listrik [1]. Sistem pengenalan beban listrik memiliki keandalan dalam memperoleh informasi yang relevan dari setiap beban listrik dengan menganalisis data meteran di area tertentu [2]. Dalam sistem pengenalan beban listrik, klasifikasi merupakan teknik yang penting dalam melakukan pengenalan. Umumnya, karena keterbatasan alat dalam akusisi data, fitur yang digunakan dalam melakukan klasifikasi adalah tegangan dan arus. Sehingga dengan itu, untuk

memastikan keakuratan dalam sistem pengenalan, sangat penting untuk melakukan ekstraksi fitur dalam data secara efektif untuk mendapatkan hasil klasifikasi yang optimal. Proses ekstraksi fitur pada beban listrik memiliki tujuan untuk menganalisis tiga jenis fitur, antarlain fitur fisik, fitur harmonik, dan hubungan tegangan dan arus [3]. Klasifikasi beban listrik dibagi menjadi dua, yaitu klasifikasi beban dengan kondisi berjalan secara independen atau satu dan klasifikasi dengan beberapa beban dengan kondisi berjalan secara bersamaan. Klasifikasi beban listrik independen hanya mengumpulkan data beban yang berjalan secara

independen sehingga mampu mengidentifikasi langsung jenis perangkat yang sedang berjalan. Dengan itu, klasifikasi beban listrik independen mampu memperdalam pemahaman mengenai perangkat yang berbeda dan memberikan referensi untuk klasifikasi beban listrik non-independen [4].

Konsep utama dari klasifikasi adalah memisahkan berbagai pola beban dan mengkategorikannya berdasarkan kelompok dengan pola yang sama atau lebih mirip satu sama lain berdasarkan berbagai algoritma pengelompokan. Berbagai algoritma telah diuji pada berbagai jenis data pola beban listrik diantaranya Decision Tree [5] dan [6], Random forest [7], Naïve Bayes [8], SVM [9], Principal Component Analysis (PCA) [10], dan Neural Network [11]. Beragam metode dalam penggunaan fitur telah diusulkan pada penelitian sebelumnya, antara lain fitur pengukuran tegangan dan arus [12], fitur gelombang arus [13], fitur gelombang harmonisa dari arus dan tegangan [14], fitur karakteristik beban listrik berdasarkan pola V-I trajectory [15], fitur beban listrik dengan pertimbangan sifat transien pada setiap beban dalam keadaan aktif dan non-aktif [16]. Pada penelitian [17] dilakukan identifikasi jenis beban harmonik menggunakan sampel data dari salah satu pembangkit listrik di Malaysia dengan permodelan Neural Network. Pada penelitian [18] dan [19], mengusulkan penggunaan distorsi daya dalam melakukan identifikasi dengan memperhatikan hubungan antara harmonisa tegangan, daya aktif, dan daya reaktif.

Lingkup penelitian ini adalah melakukan analisis karakteristik berbagai fitur beban listrik secara independen dengan keadaan steady-state dan dilakukan klasifikasi. Pada kasus ini, mempunyai enam kategori di variabel responnya dan enam variabel prediktor yang mana merupakan data fitur. Penelitian ini akan dilakukan perbandingan metode klasifikasi yang berbeda untuk mengetahui metode mana yang paling cocok dalam mengklasifikasikan beban listrik antarlain dengan melakukan prediksi berdasarkan data label menggunakan algoritma K-Nearest Neighbours atau melakukan prediksi dengan menghitung probabilitas dari masing-masing variabel prediktor dengan menggunakan algoritma Regresi Logistik Multinomial. Kedua metode akan dievaluasi dari segi akurasi ketepatan dan kecepatan proses dalam analisis untuk mendapatkan hasil. Sistem klasifikasi akan mengidentifikasi jenis, model, dan prasyarat yang tidak diketahui dari beban listrik dan mengelompokannya. Fitur dari beban listrik yang akan diteliti antarlain besarnya tegangan dan arus root mean square, gelombang

harmonisa arus, seri daya dan faktor daya dari variasi sample beban listrik yang berbeda.

2 METODOLOGI

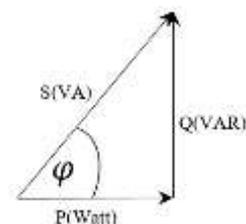
A. Harmonisa Beban Listrik

Harmonisa adalah gangguan yang terjadi pada sistem distribusi tenaga listrik yang disebabkan oleh terjadinya distorsi gelombang arus dan tegangan [20]. Didalam saluran tenaga listrik bolak-balik (AC), terdapat sinyal periodik dari tegangan atau arus pada beban non linier. Secara umum, distorsi harmonisa merupakan hal yang keberadaannya tidak diinginkan. Nilai harmonisa dapat diasumsikan dengan kuantitas dan sudut fasa yang acak dan dinyatakan dalam persamaan deret fourier [21].

$$f(t) = \frac{a_0}{2} + \sum_{n=1}^{\infty} \{a_n \cdot \cos(h\omega_0 t) + b_n \cdot \sin(h\omega_0 t)\}$$

Pada penelitian ini digunakan transformasi *Fast Fourier* (FFT) untuk mendapatkan fitur harmonisa arus dari suatu beban dengan mengubah fitur arus dari domain waktu menjadi domain frekuensi. Dengan menggunakan FFT akan didapatkan fitur dengan nilai yang berbeda untuk setiap sinyal dari beban listrik bergantung pada tingkat nonliniernya.

B. Daya Listrik



Gambar 1. Segitiga Daya

Gambar 1. memberi gambaran bahwa hubungan antar daya semu (S), daya aktif (P) dan daya reaktif (Q) dapat diekspresikan kedalam sebuah persamaan *phytagoras*.

$$S = \sqrt{P^2 + Q^2}$$

Antara daya semu (S) dan daya aktif (P) dipisahkan oleh sudut φ (Phi). Rasio antara daya semu (S) dan daya aktif (P) adalah nilai cosinus dari sudut φ (Phi). Apabila nilai daya reaktif semakin kecil (Q) maka besarnya daya semu (S) akan mendekati besarnya daya aktif (P) dan nilai $\cos \varphi$ akan semakin besar atau mendekati 1. Pada kasus dimana $\varphi = 0^\circ$, maka $\cos \varphi = 1$, daya semu (S) = daya aktif (P). artinya semua daya yang semu (nyata) dapat diartikan sebagai daya aktif, sebaliknya jika $\varphi = 90^\circ$, $\cos \varphi = 0$, daya semu (S) = daya reaktif (Q) yangmana dapat diartikan bahwa semua daya yang semu (nyata) yang diberikan tidak dapat dimanfaatkan dan hanya menjadi daya reaktif

saja. Dengan itu besar daya semu (S) merupakan resultan dari daya aktif dan daya reaktif (Q).

C. Algoritma *k-Nearest Neighbours*

Algoritma *k-Nearest Neighbours* bertujuan untuk mengklasifikasikan objek baru berdasarkan atribut dan sampel terlatih dari sebuah data dimana hasil dari sampel uji yang baru akan diklasifikasikan berdasarkan kategori dari *k-NN* setelah mengumpulkan *k-NN*, kemudian diambil mayoritas dari *k-NN* untuk dijadikan prediksi dari sample uji [22]. Untuk memutuskan titik dari set pelatihan, dilakukan prediksi dan pengamatan dengan memilih *k* terdekat yang menunjuk pada pengamatan. Algoritma *k-NN* diformulasikan sebagai berikut [23]:

$$Pr(Y = j|X = x_0) = \frac{1}{k} \sum_{i \in N_0} I(y_i = j)$$

Diasumsikan pada set data bahwa *X* adalah matriks fitur dari pengamatan dan *Y* adalah label kelas. Metode *k-NN* akan menghitung distribusi bersyarat dari *Y* terhadap *X* lalu mengklasifikasikannya ke kelas dengan probabilitas tertinggi. Parameter *k* adalah bilangan bulat positif yang merupakan pengamatan terdekat dengan pengamatan uji x_0 lalu memperkirakan probabilitas bahwa hasil pengamatan uji merupakan milik kelas *j* dengan menggunakan rumus (3) dimana N_0 adalah himpunan *k* terdekat dari pengamatan dan $I(y_i = j)$ merupakan variabel prediktor yang bernilai 1 jika pengamatan diberikan (x_i, y_i) didalam himpunan N_0 merupakan kelas *j* dan akan bernilai 0 jika bukan merupakan kelas *j*. setelah memperkirakan probabilitas, parameter *k* akan mengelompokkan pengamatan uji x_0 kedalam probabilitas dari yang terbesar.

Karena pada algoritma *k-NN* membutuhkan titik *k* terdekat maka akan dihitung jarak antar titik data input dengan titik lain. Dekat atau jauhnya titik biasanya dihitung berdasarkan jarak *euclidian*. Dimisalkan *x* adalah titik dengan koordinat $(x_1, x_2, x_3, \dots, x_p)$ dan *y* adalah titik dengan koordinat $(y_1, y_2, y_3, \dots, y_p)$ maka jarak antar keduanya didefinisikan sebagai berikut.

$$d(x, y) = \sqrt{\sum_{i=1}^p (x_i - y_i)^2}$$

D. Algoritma Regresi Logistik Multinomial

Regresi logistik multinomial merupakan regresi logistik yang digunakan saat variabel dependen mempunyai skala yang bersifat *polichotomous* atau multinomial [24]. Skala multinomial memiliki pengukuran lebih dari dua kategori. Untuk penelitian

ini digunakan metode regresi logistik dengan variabel dependen berskala nominal dengan enam kategori. Model regresi dengan skala nominal enam kategori digunakan kategori variabel hasil *Y* dikodekan 0, 1, 2, 3, 4, dan 5. Variabel *Y* akan diparametrisasi menjadi dua fungsi logit dan $Y=0$ akan digunakan sebagai pembanding antara keenam kategori. Untuk membentuk fungsi logit akan dibandingkan $Y=1, Y=2, Y=3, Y=4,$ dan $Y=5$ terhadap $Y=0$. Bentuk model regresi logistik multinomial dengan *n* merupakan variabel prediktor diformulasikan sebagai berikut.

$$P(Y = j | x) = \frac{e^{g_j(x)}}{\sum_{k=0}^5 e^{g_k(x)}}, j = 0,1,2,3,4,5$$

$$= \frac{e^{(\beta_{j0} + \beta_{j1}X_1 + \beta_{j2}X_2 + \dots + \beta_{jn}X_n)}}{\sum_{k=0}^5 e^{(\beta_{k0} + \beta_{k1}X_1 + \beta_{k2}X_2 + \dots + \beta_{kn}X_n)}} \quad (5)$$

Dengan $\beta_0 = 0$ maka $g_0(x) = 0$.

Keterangan: (3)

$P(Y = j | x)$ = Probabilitas bersyarat dari variabel respon *j* pada vektor *x*

X_n = Variabel independen

β_0 = Koefisien konstan

β_n = Koefisien dari variabel independen

$g_j(x)$ = Logit pada variabel respon *j*

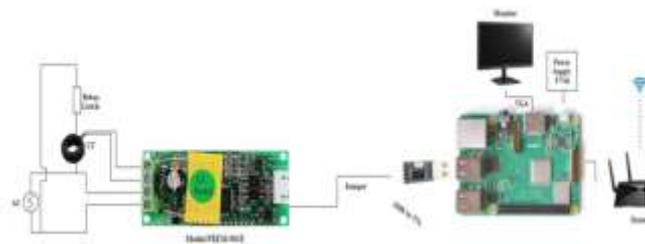
Umumnya, pada data yang berdimensi tinggi, penerapan algoritma regresi logistik sering terjadi *overfitting* akibat dari besarnya nilai parameter koefisien konstan (β_0) dan koefisien dari variabel independen (β). Maka dari itu, diterapkannya metode regularisasi. Terdapat dua jenis regularisasi yang dapat diterapkan pada metode klasifikasi regresi logistik, yaitu regularisasi pinalti l_1 dan pinalti l_2 yang keduanya diformulasikan sebagai berikut.

$$l_1(\beta_0, \beta) = C ||\{\beta_0, \beta\}||_1 = C(\beta_0 + \sum_j^p |\beta_j|) \quad (6)$$

$$l_2(\beta_0, \beta) = C ||\{\beta_0, \beta\}||_2 = C(\beta_0 + \sum_j^p |\beta_j|^2) \quad (7)$$

Dengan $j = 0,1,2,3,4,5$ dan *C* merupakan parameter regularisasi yang bernilai $C > 0$. Jenis regularisasi pada penelitian ini adalah regularisasi pinalti l_2 . Regularisasi l_2 akan membuat standar error dari estimasi parameter yang digunakan semakin mengecil [25] sehingga perubahan dari variasi parameter *C* nilainya menjadi lebih stabil jika dibandingkan dengan regresi logistik tanpa regularisasi.

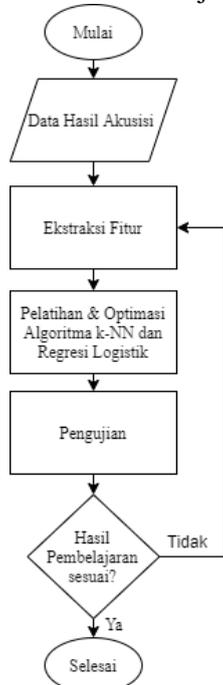
A. Sistem Perangkat Keras



Gambar 2. Sistem perangkat keras

Gambar 2 menjelaskan cara kerja sistem dalam melakukan akusisi data. Langkah pertama yang dilakukan adalah menghubungkan beban listrik dengan sumber tegangan. Set data yang digunakan dalam penelitian ini diambil dari pengukuran sensor PZEM-004T setiap detiknya selama 4000 detik. Data dari sensor di simulasikan oleh Raspberry-Pi menggunakan bahasa pemrograman python yang menampilkan data-data besaran dari beban listrik. Data tersebut akan ditampilkan pada serial monitor. Setelah itu, Raspberry-Pi melakukan pengiriman data ke database SQL server untuk melakukan pengumpulan dan penyimpanan set data dari setiap beban listrik. Pada set data terdapat beberapa informasi beban listrik disetiap detiknya. Fitur yang dihasilkan dari pengukuran sensor antara lain: waktu, tegangan (V), arus (A), daya aktif (P), energi (Kwh), frekuensi (Hz) dan faktor daya (PF).

B. Flowchart Proses Pembelajaran



Gambar 3. Flowchart proses pembelajaran

Gambar 3 memperlihatkan alur pembelajaran yang akan diteliti. Data hasil akusisi dari sensor akan

dilakukan ekstraksi fitur. Metode ekstraksi fitur akan sangat mempengaruhi dalam proses klasifikasi sehingga pada penelitian ini dikategorikan sebagai variabel bebas. Merujuk pada [3], penelitian ini akan menggunakan fitur beban dengan karakteristik steady-state, yaitu pada kondisi dimana beban listrik yang diuji dalam kondisi kerja yang stabil. Fitur yang digunakan pada karakteristik *steady-state* antara lain perubahan daya setiap waktu antara lain daya aktif (P), daya reaktif (Q), daya semu (S) dan faktor daya (PF), tegangan (V), arus (A) dan gelombang harmonisa arus. Pada penelitian ini digunakan transformasi Fast Fourier (FFT) untuk mendapatkan fitur harmonisa arus dari suatu beban dengan mengubah fitur arus dari domain waktu menjadi domain frekuensi. Dengan menggunakan FFT akan didapatkan fitur dengan nilai yang berbeda untuk setiap sinyal dari beban listrik bergantung pada tingkat nonliniernya.

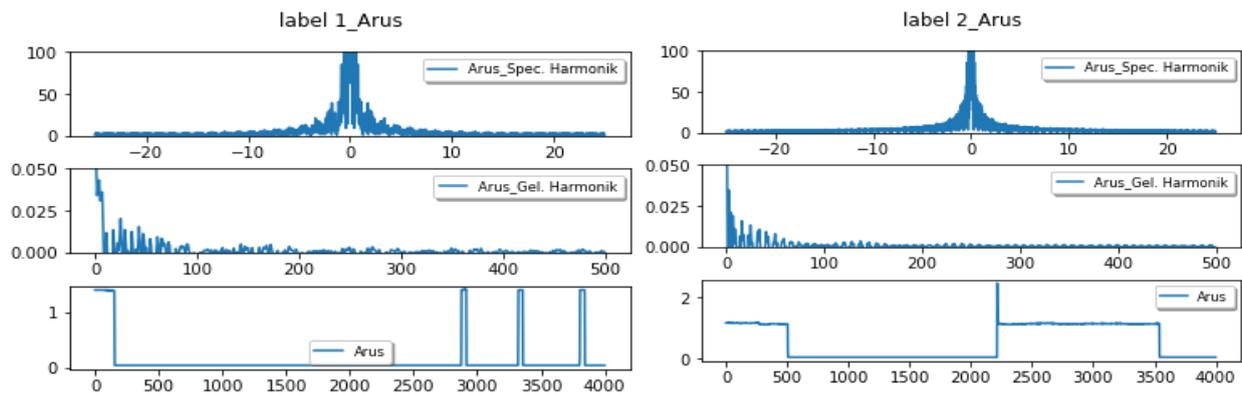
Pada perancangan ini akan dilakukan beberapa kali pengujian dengan metode berbeda dan optimalisasi parameter dari masing-masing algoritma yang akan diteliti. Pada penelitian ini diharapkan mampu mencapai akurasi minimal 90%. Dalam mewujudkan hasil yang optimal, maka dilakukan optimasi dari nilai parameter pada metode k-NN dan regresi logistik. Hasil akurasi diperoleh dari penggunaan parameter optimal dan metode validasi silang untuk mendapatkan hasil akurasi yang lebih pasti.

3 HASIL DAN PEMBAHASAN

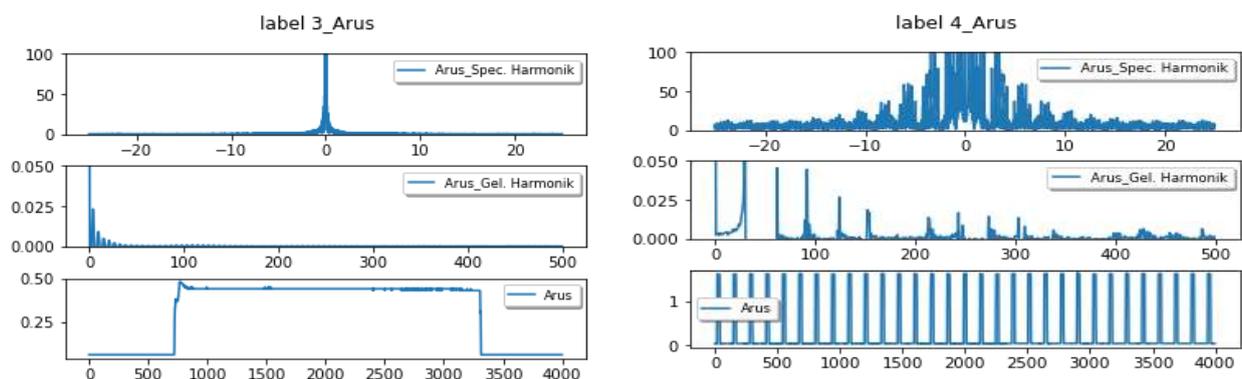
A. Pra-pemrosesan Data

Pada tahap awal akan dilakukan *data labelling*. Label yang dimaksudkan adalah target variabel yang akan diprediksi. Dalam melakukan *supervised learning*, label yang mana adalah tag kategori dari data adalah sebagai supervisornya. Pada proses ini akan dilakukan labelisasi pada masing-masing data beban dengan angka 1 hingga 6. Setelah melakukan *data labelling* lalu akan dilakukan pemeriksaan apakah terdapat kasus *missing value* pada data masing-masing fitur. Pada penelitian ini tidak ditemukan adanya *missing value* dari 24.000 data.

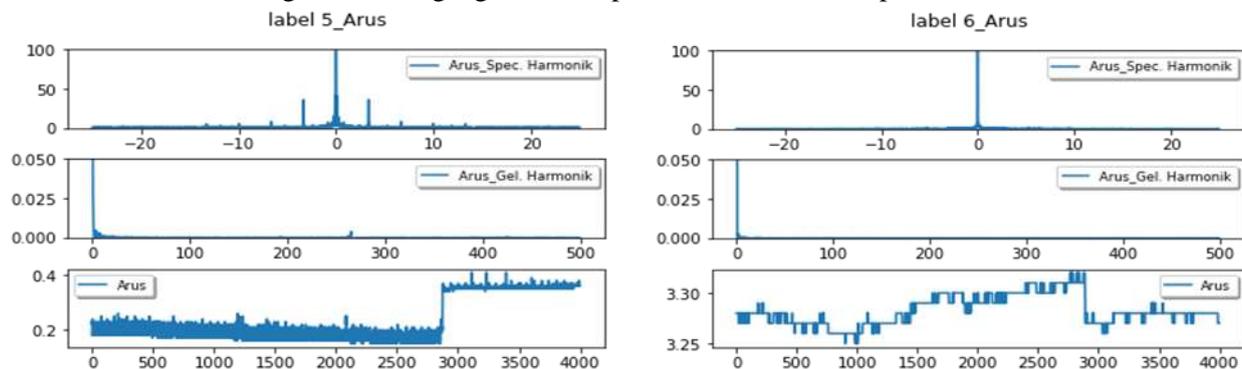
Selanjutnya dilakukan proses ekstraksi fitur. Pada penelitian ini, Fitur gelombang harmonik dihasilkan dari data sensor per detiknya. Frekuensi fundamental gelombang listrik bernilai 50 Hz dan satu siklus gelombang mewakili 1/60 detik.



Gambar 4. Gelombang arus terhadap domain waktu dan frekuensi pada beban dispenser dan freezerbox



Gambar 5. Gelombang arus dan tegangan terhadap waktu dan frekuensi pada beban kulkas dan setrika



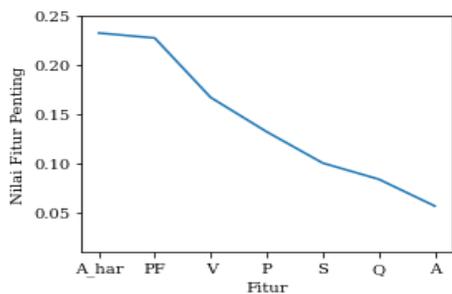
Gambar 6. Gelombang arus dan tegangan terhadap waktu dan frekuensi pada beban charger laptop dan hairdryer

Gambar 4 hingga Gambar 6 di atas merupakan grafik analisis magnitudo dan nilai fasa pada frekuensi dasar dari gelombang arus dalam domain waktu dan domain frekuensi. Kandungan harmonisa dengan domain frekuensi dihasilkan dari perhitungan amplitudo dan fasa menggunakan *Fast Fourier Transform*. Gambar tersebut menunjukkan bahwa masing-masing beban memiliki karakteristik tegangan dan arus yang unik.

Tahapan selanjutnya adalah proses seleksi fitur. Alasan dari perlunya melakukan proses seleksi fitur adalah karena setelah melakukan studi literatur dan dari beberapa kajian teori yang telah didapatkan [26]–[29], proses seleksi fitur dalam pra-pemrosesan data merupakan tahap yang penting dilakukan untuk mengurangi jumlah atribut atau fitur yang terkait

dalam melakukan prediksi karena atribut data yang berlebihan dan tidak relevan dapat berpengaruh pada proses dan hasil klasifikasi. Berdasarkan hasil dari seleksi fitur diharapkan mampu meningkatkan hasil akurasi klasifikasi serta mempercepat proses komputasi dalam melakukan klasifikasi data. Pada penelitian ini dilakukan seleksi fitur dengan melakukan pembobotan pada setiap fitur dengan perhitungan maksimal entropy [29]. Nilai entropy yang didapatkan digunakan untuk mendapatkan atribut terbaik. Lalu akan meranking nilai gain dari setiap atribut. Jika nilai gain atribut sangat rendah maka fitur tersebut tidak mempunyai pengaruh besar terhadap klasifikasi beban listrik sehingga fitur ini tidak diikuti dalam analisis lebih lanjut. Dalam menentukan entropy dari setiap atribut data pada

penelitian ini digunakan algoritma Decision Tree. Nilai dari fitur penting dapat dilihat pada Gambar 5.

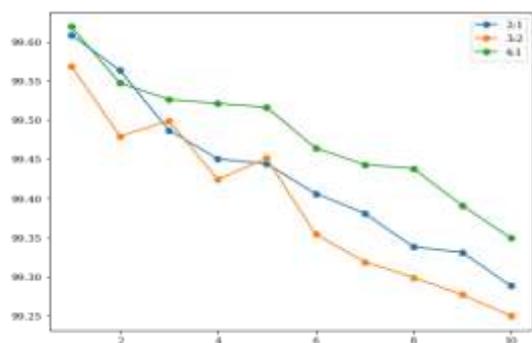


Gambar 7. Fitur Penting

Berdasarkan grafik diatas dapat dilihat bahwa nilai fitur penting untuk arus (A) menurun tajam. Hal ini menunjukkan bahwa fitur arus (A) tidak mempunyai pengaruh besar terhadap klasifikasi beban listrik sehingga variabel ini tidak diikuti dalam analisis lebih lanjut.

B. Optimasi Parameter *k-Nearest Neighbours*

Tahapan seleksi fitur pada set data telah dilakukan, kemudian akan dilakukan tahap evaluasi terhadap hasil klasifikasi menggunakan fitur yang terpilih menggunakan metode *k*-NN. diketahui bahwa nilai *k* harus lebih besar sama dengan satu sehingga range dari parameter *k* berada pada interval $(1 \leq k < \infty)$ [30] sehingga pada percobaan ini akan dilakukan optimasi parameter menggunakan nilai dari range interval tersebut. dengan nilai masing-masing parameter $k = \{1, 2, 3, 4, 5, 6, 7, 8, 9\}$ dan memprediksi salah satu nilai parameter yang paling optimal pada saat melakukan pelatihan. Untuk mendapatkan nilai akurasi yang lebih optimal, pada penelitian ini dilakukan validasi silang dengan nilai *n-fold* = 10 pada setiap level parameter yang ditentukan dengan menggunakan metode optimasi grid search untuk menjamin nilai optimum pada proses pengklasifikasian.

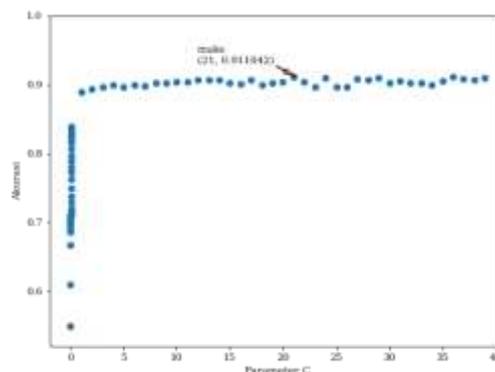


Gambar 8. Grafik akurasi terhadap parameter k dan rasio data

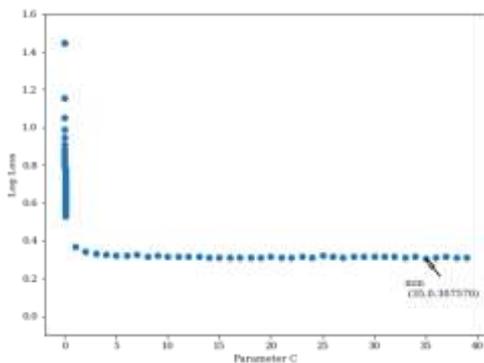
Gambar 8 menunjukkan tingkat akurasi dari metode *k*-NN dengan menggunakan rasio data latih dan data uji yang berbeda dengan masing-masing data latih sebanyak 16000, 14400, dan 19200 data. Grafik diatas menunjukkan akurasi tertinggi pada $k = 1$ dengan rasio data 4:1 yang mana jumlah data latihnya sebanyak 19200 data dengan akurasi sebesar 99.609%. Nilai akurasi mengalami penurunan untuk setiap nilai parameter *k* diatas 1.

C. Optimasi Parameter Regresi Logistik Multinomial

Tahapan selanjutnya adalah melakukan modelling dengan menggunakan pendekatan Regresi Logistik. Setelah dilakukan ekstraksi dan seleksi fitur dari masing-masing data beban, untuk mendapatkan fitur yang paling baik atau bahkan berguna untuk mengatasi ambiguitas dalam klasifikasi pada penelitian ini akan menggunakan regularisasi. Regularisasi berfungsi sebagai seleksi fitur yang secara otomatis memilih serangkaian fitur yang paling baik dan yang relevan dari kumpulan kandidat label yang akan diprediksi. Parameter yang digunakan dalam penggunaan model ini adalah parameter *C*. Maksud dari parameter *C* adalah untuk menunjuk invers dari kekuatan regulasi yang mana harus bernilai positif. Parameter *C* dipilih dalam skala logaritmik dengan range $(1e^{-4}, 1e^4)$. Diketahui bahwa nilai parameter yang *C* lebih kecil akan menentukan regularisasi yang lebih kuat [31]. Lalu model akan melakukan prediksi probabilitas sesuai dengan banyaknya label yang digunakan. Pada proses optimasi ini akan dihitung akurasi dari berbagai kombinasi nilai parameter yaitu parameter *C* dan memprediksi salah satu nilai parameter yang paling optimal pada saat melakukan pelatihan. Pada penelitian ini digunakan nilai parameter *C* dari 0.0001 hingga 40 dengan penalti l_2 . Hasil perhitungan dari parameter *C* terhadap data latih ditunjukkan pada Gambar 7. dan Gambar 8.

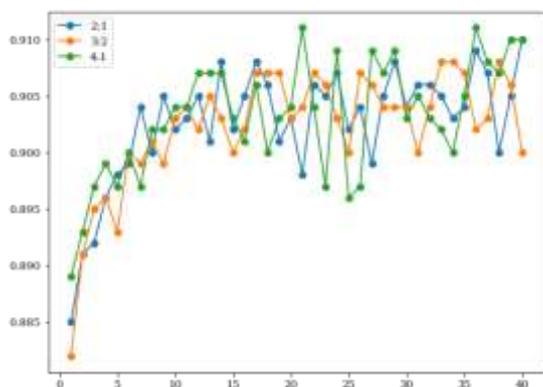


Gambar 9. Nilai akurasi terhadap nilai C pada set data pelatihan



Gambar 10. Nilai *Log Loss* terhadap nilai *C* pada set data pelatihan

Pada grafik diatas menjelaskan hasil akurasi dari data pelatihan berdasarkan nilai *C* pada metode Regresi Logistik. Ditunjukkan bahwa hasil mencapai akurasi maksimal menggunakan parameter $C=21$ dengan nilai akurasi mencapai 91.104%. Namun nilai *log loss* paling kecil didapatkan saat menggunakan parameter $C=35$ daripada penggunaan nilai parameter *C* yang lainnya, yaitu sebesar 0.30757. Pada setiap perubahan nilai parameter pada rentang 1 hingga 40 didapatkan nilai akurasi dengan selisih yang sedikit satu sama lain. Sehingga dengan itu, dalam percobaan ini nilai regularisasi parameter akan ditentukan dengan metode validasi silang dan optimasi grid search dengan memproses parameter *C* lalu mengidentifikasi hasil yang paling optimum dengan validasi silang sebanyak 10-folds. Diketahui bahwa sedikit perubahan dari nilai parameter *C* akan memberikan nilai logloss atau estimasi error yang berbeda. Akurasi akan mencapai hasil yang optimal jika nilai loglossnya kecil. Sehingga dengan itu perlu untuk mempertimbangkan nilai loglossnya disamping dengan akurasi terhadap nilai parameter *C*.



Gambar 11. Grafik akurasi terhadap parameter *C* dan rasio data

Gambar 11 merepresentasikan tingkat akurasi dari metode Regresi Logistik dengan menggunakan rasio data latih dan data uji yang berbeda dengan masing-masing data latih sebanyak 16000, 14400, dan 19200 data menggunakan metode validasi silang. Hasil menunjukkan ternyata metode pemilihan parameter *C* dalam validasi silang menghasilkan kinerja pengujian yang lebih baik daripada parameter yang dipilih diluar validasi silang. Akurasi tertinggi didapatkan pada rasio data 4:1 yang mana jumlah data latihnya sebanyak 19200 data. Nilai parameter regularisasi dipilih secara otomatis dengan validasi silang sebanyak 10-folds dan akurasi tertinggi ditunjukkan pada penggunaan parameter *C* dengan nilai 35, yaitu sebesar 91,125%. Dapat disimpulkan bahwa semakin banyak data pelatihan yang digunakan maka akan semakin akurat hasil dari pengujian.

D. Evaluasi Performansi

Terdapat beberapa pertimbangan utama dalam menentukan baik buruknya suatu model berdasarkan dari peformansinya. Peforma sistem algoritma klasifikasi machine learning umumnya dapat dievaluasi menggunakan confusion matrix. Dari perhitungan confusion matrix selanjutnya akan diperoleh nilai akurasi, presisi, sensitivitas dan *F1 score* dari setiap label atau beban yang diteliti.

Tabel 1. Komputasi metode *k-Nearest Neighbours* pada rasio data set berbeda

Evaluasi	2:1	3:2	4:1
Waktu Pelatihan	0:00:01.697	0:00:01.135	0:00:02.249
Waktu Pengujian	0:00:00.724	0:00:00.720	0:00:00.380

Tabel 1. memperlihatkan nilai komputasi dan akurasi pada penggunaan proporsi data pengujian yang berbeda-beda. Proporsi data pengujian 4:1 memberikan akurasi yang lebih tinggi daripada rasio data yang lainnya dan membutuhkan waktu yang sedikit pada proses pengujian sehingga penggunaan metode ini membutuhkan komputasi yang relatif rendah.

Tabel 2. Performasi metode *k-Nearest Neighbours* dalam proses pengujian

Beban	2:1			3:2			4:1		
	Presisi (%)	Sensitivitas (%)	<i>FI-Score</i> (%)	Presisi (%)	Sensitivitas (%)	<i>FI-Score</i> (%)	Presisi (%)	Sensitivitas (%)	<i>FI-Score</i> (%)
Dispenser	99.775	99.699	99.737	99.813	99.813	99.813	99.626	100.00	99.813
Freezerbox	99.324	99.250	99.287	99.251	99.375	99.313	99.001	99.125	99.063
Kulkas	99.475	99.475	99.475	99.625	99.625	99.625	99.377	99.750	99.563
Setrika	99.624	99.400	99.512	99.687	99.500	99.593	99.748	99.000	99.373
C.Laptop	99.626	99.850	99.738	99.813	99.875	99.844	99.875	99.750	99.812
Hairdryer	99.850	100.00	99.925	99.813	99.813	99.813	99.875	99.875	99.875

Dengan demikian, diketahui bahwa model mampu memprediksi dengan baik menggunakan fitur-fitur yang digunakan. Hasil mencapai akurasi maksimal pada percobaan proporsi data pelatihan dan pengujian sebesar 4:1. Dari hasil prediksi beban, ditunjukkan hasil bahwa model pada ketiga rasio proporsi data mampu mengenali beban *hairdryer* dengan sangat akurat.

Tabel 3 memperlihatkan pada percobaan menggunakan metode Regresi Logistik membutuhkan waktu yang lama dalam proses pengujian dibandingkan dengan metode *k-NN* sehingga penggunaan metode ini membutuhkan komputasi besar untuk melakukan proses pengujian.

Tabel 3. Komputasi metode Regresi Logistik rasio data set berbeda

Evaluasi	2:1	3:2	4:1
Waktu Pelatihan	0:00:35.617	0:00:20.721	0:00:27.397
Waktu Pengujian	0:00:00.004	0:00:00.003	0:00:00.003

Tabel 4. Performasi metode Regresi Logistik dalam proses pengujian

Beban	2:1			3:2			4:1		
	Presisi (%)	Sensitivitas (%)	<i>FI-Score</i> (%)	Presisi (%)	Sensitivitas (%)	<i>FI-Score</i> (%)	Presisi (%)	Sensitivitas (%)	<i>FI-Score</i> (%)
Dispenser	82.187	95.874	88.504	81.990	94.750	87.910	81.415	96.375	88.266
Freezerbox	82.972	63.241	71.775	85.763	63.250	72.806	88.927	64.250	74.601
Kulkas	91.951	89.122	90.514	91.917	91.687	91.802	92.181	91.375	91.777
Setrika	86.676	95.577	90.909	85.299	95.375	90.056	87.273	96.000	91.429
C.Laptop	98.276	98.276	98.276	98.424	97.563	97.991	98.379	98.625	98.502
Hairdryer	99.700	99.625	99.662	99.625	99.625	99.625	99.625	99.625	99.625

Dengan demikian, diketahui bahwa model mampu memprediksi dengan baik menggunakan fitur-fitur yang digunakan namun tidak sebaik metode k -NN. Dari hasil prediksi beban, ditunjukkan hasil bahwa model pada ketiga rasio proporsi data mampu mengenali beban *hairdryer* dengan baik.

4 KESIMPULAN

Kedua metode yang digunakan sama-sama memiliki akurasi klasifikasi yang baik, tetapi juga memiliki sensitivitas pada fitur tertentu yang berpeluang untuk terjadi overfitting dan fenomena lainnya. Oleh karena itu, perlu adanya optimasi parameter inti dari metode k -NN dan Regresi Logistik, melakukan percobaan konsistensi set data yang berbeda dan melakukan validasi silang untuk mendapatkan hasil klasifikasi yang optimal. Hasil menunjukkan bahwa metode k -NN memiliki akurasi yang lebih tinggi dari pada metode Regresi Logistik sehingga penggunaan metode k -NN cocok untuk mengklasifikasikan peristiwa langka pada data.

DAFTAR PUSTAKA

- [1] H. Liu, C. Yu, and H. Wu, "Potential Applications of Smart Device Recognition in Industry," in *Smart Device Recognition*, Springer, 2021, pp. 259–294.
- [2] M. Zeifman and K. Roth, "Nonintrusive appliance load monitoring: Review and outlook," *IEEE Trans. Consum. Electron.*, vol. 57, no. 1, pp. 76–84, 2011.
- [3] H. Liu, C. Yu, and H. Wu, *Smart Device Recognition: Ubiquitous Electric Internet of Things*. Springer Nature, 2020.
- [4] Y. Beck and R. Machlev, "Harmonic loads classification by means of currents' physical components," *Energies*, vol. 12, no. 21, p. 4137, 2019.
- [5] J. Lin, X. Ding, D. Qu, and H. Li, "Non-intrusive load monitoring and decomposition method based on decision tree," *J. Math. Ind.*, vol. 10, no. 1, pp. 1–14, 2020.
- [6] H. Kang and H. Kim, "Household appliance classification using lower odd-numbered harmonics and the bagging decision tree," *IEEE Access*, vol. 8, pp. 55937–55952, 2020.
- [7] W. Guohua, Y. Diping, Y. Jiyao, Z. Wenhua, D. Peng, and X. Yiqing, "Research on non-intrusive load monitoring based on random forest algorithm," in *2020 4th International Conference on Smart Grid and Smart Cities (ICSGSC)*, 2020, pp. 1–5.
- [8] P. Adjei, N. S. Sethi, C. P. de Souza, and M. A. Capretz, "Energy Disaggregation using Multilabel Binarization and Gaussian Naive Bayes Classifier," in *2020 11th IEEE Annual Ubiquitous Computing, Electronics & Mobile Communication Conference (UEMCON)*, 2020, pp. 0093–0100.
- [9] A. F. M. Jaramillo, D. M. Laverty, J. M. Del Rincón, P. Brogan, and D. J. Morrow, "Non-Intrusive Load Monitoring Algorithm for PV Identification in the Residential Sector," in *2020 31st Irish Signals and Systems Conference (ISSC)*, 2020, pp. 1–6.
- [10] R. Machlev, D. Tolkachov, Y. Levron, and Y. Beck, "Dimension reduction for NILM classification based on principle component analysis," *Electr. Power Syst. Res.*, vol. 187, p. 106459, 2020.
- [11] G. Zhou, Z. Li, M. Fu, Y. Feng, X. Wang, and C. Huang, "Sequence-to-sequence load disaggregation using multiscale residual neural network," *IEEE Trans. Instrum. Meas.*, vol. 70, pp. 1–10, 2020.
- [12] R. Medico *et al.*, "A voltage and current measurement dataset for plug load appliance identification in households," *Sci. Data*, vol. 7, no. 1, pp. 1–10, 2020.
- [13] M. H. Jopri, "Naïve bayes and linear discriminate analysis based diagnostic analytic of harmonic source identification," *Indones. J. Electr. Eng. Comput. Sci.*, vol. 20, no. 3, 2020.
- [14] Y. Li, Y. Yang, K. Sima, B. Li, T. Sun, and X. Li, "Non-intrusive load monitoring based on harmonic characteristics," *Procedia Comput. Sci.*, vol. 183, pp. 776–782, 2021.
- [15] A. Faustine and L. Pereira, "Multi-label learning for appliance recognition in NILM using Fryze-current decomposition and convolutional neural network," *Energies*, vol. 13, no. 16, p. 4154, 2020.
- [16] H. Wu and H. Liu, "Non-intrusive load transient identification based on multivariate LSTM neural network and time series data augmentation," *Sustain. Energy Grids Netw.*, vol. 27, p. 100490, 2021.
- [17] S. Y. Wong, K. S. Yap, and X. Li, "Classification of the harmonic load types using multi-layer extreme learning machine," 2018.

- [18] C. I. Garcia, F. Grasso, A. Luchetta, M. C. Piccirilli, L. Paolucci, and G. Talluri, "A comparison of power quality disturbance detection and classification methods using CNN, LSTM and CNN-LSTM," *Appl. Sci.*, vol. 10, no. 19, p. 6755, 2020.
- [19] A. E. Lazzaretti *et al.*, "A multi-agent NILM architecture for event detection and load classification," *Energies*, vol. 13, no. 17, p. 4396, 2020.
- [20] H. B. Utomo, S. Sunarto, and E. Fadhlurohman, "Analisa Perancangan Filter Pasif dan Efisiensi Transformator Daya Akibat Harmonisa dengan Menggunakan Matlab," in *Prosiding Industrial Research Workshop and National Seminar*, 2021, vol. 12, pp. 324–330.
- [21] D. Antono, A. Wasono, and L. J. D. Atmanto, "HARMONISA ARUS DAN TEGANGAN PADA JARINGAN LISTRIK AKIBAT PEMBEBANAN AIR CONDITION (AC) INVERTER TIGA FASA," *ReTII*, 2015.
- [22] K. Eliyen, H. Tolle, and M. A. Muslim, "K-Nearest Neighbor Untuk Klasifikasi Penilaian Pada Virtual Patient Case," *J. Arus Elektro Indones.*, vol. 3, no. 1, pp. 15–18, 2017.
- [23] "K-Nearest Neighbors (KNN)," *K-Nearest Neighbors (KNN)*, 2020. <https://stats.libretexts.org/@go/page/2483>
- [24] L. Etaati, "Introduction to Machine Learning," in *Machine Learning with Microsoft Technologies*, Springer, 2019, pp. 3–14.
- [25] T. Widhianingsih, "Klasifikasi Data Berdimensi Tinggi dengan Metode Ensemble Berbasis Regresi Logistik dalam Permasalahan Drug Discovery," Institut Teknologi Sepuluh Nopember, 2018.
- [26] B. N. Sari, "Implementasi teknik seleksi fitur information gain pada algoritma klasifikasi machine learning untuk prediksi performa akademik siswa," *SEMNASTEKNOMEDIA ONLINE*, vol. 4, no. 1, pp. 2-9–55, 2016.
- [27] M. Z. F. Nasution, "Seleksi Fitur Dataset Menggunakan Principal Component Analysis (PCA) untuk Meningkatkan Akurasi Klasifikasi Decision Tree C4. 5," 2018.
- [28] I. G. N. D. Adnyana, R. M. Arjuna, A. N. Indrainsi, and D. S. Prasvita, "Pengaruh Seleksi Fitur Pada Algoritma Machine Learning Untuk Memprediksi Pembatalan Pesanan Hotel," *Senamika*, vol. 2, no. 1, pp. 551–558, 2021.
- [29] J. Novaković, "Toward optimal feature selection using ranking methods and classification algorithms," *Yugosl. J. Oper. Res.*, vol. 21, no. 1, 2016.
- [30] P. Hall, B. U. Park, and R. J. Samworth, "Choice of neighbor order in nearest-neighbor classification," *Ann. Stat.*, vol. 36, no. 5, pp. 2135–2152, 2008.
- [31] R. Tanone and A. B. Emmanuel, "Prediksi Not Operational Transaction Menggunakan Logistic Regression pada Bank XYZ di Kota Kupang," *AITI*, vol. 17, no. 1, pp. 42–55, 2020.