Vol. 11 No.1 Tahun 2025, pp 1-13
Jurnal Teknologi dan Manajemen Informatika
http:// http://jurnal.unmer.ac.id/index.php/jtmi
P-ISSN: 1693-6604 E-ISSN: 2580-8044

Optimasi K-Nearest Neighbor Dengan Particle Swarm Optimization Pada Klasifikasi Pelanggan Listrik Rumah Tangga Bersubsidi

Siti Andini Utiarahman [1], Hastuti Dalai [2], Ali S. Sabudi [3]

Universitas Ichsan Gorontalo, Indonesia

Article Info

Article History
Received: 01-11-2024
Revised : 28-11-2024
Accepted: 04-12-2024

Keywords
Data Mining;
Classification;
K-Nearest Neighbor;
Particle Swarm Optimization;
Electricity Subsidy.

✉ Corresponding Author
Siti Andini Utiarahman
Universitas Ichsan Gorontalo
siti_andini@unisan.ac.id

ABSTRACT

This research optimizes the K-Nearest Neighbor (KNN) method using Particle Swarm Optimization (PSO) for classifying household electricity subsidy power in Gorontalo Province. Using P3KE data with 98,859 records, this research aims to improve the accuracy of classifying 450 VA and 900 VA power for the electricity subsidy program. The research methodology includes data preprocessing, KNN implementation, parameter optimization using PSO, and model evaluation using a confusion matrix. The research results show an accuracy improvement of 1.3% from 83.53% to 84.83% after optimization. The optimized model showed an increase in precision for the 900 VA class from 0.58 to 0.71, although there was a decrease in recall from 0.32 to 0.25. For the 450 VA class, the model maintained a precision of 0.86 with an increase in recall from 0.95 to 0.98.

PENDAHULUAN

Listrik adalah salah satu kebutuhan dasar masyarakat yang memerlukan pengaturan dan penyediaan yang tepat oleh negara [1]. Perusahaan Listrik Negara (PLN), sebagai penyedia listrik nasional, mengatur distribusi listrik berdasarkan Peraturan Menteri Energi dan Sumber Daya Mineral (ESDM) Republik Indonesia Nomor 11 Tahun 2022 tentang Pelaksanaan Usaha Ketenagalistrikan. Salah satu bagian dari pengaturan ini adalah pemberlakuan sistem subsidi listrik, yang bertujuan untuk menjaga stabilitas ekonomi dan meringankan beban masyarakat [2]. Subsidi listrik menjadi salah satu bentuk intervensi pemerintah dalam membantu masyarakat kurang mampu agar dapat menikmati layanan listrik yang esensial [3].

Dalam program Subsidi Listrik Tepat Sasaran (SLTS), subsidi diberikan untuk Rumah Tangga (RT) miskin dengan daya listrik 450 Volt Ampere (VA) dan Rumah Tangga (RT) tidak mampu dengan 900 VA. Besaran subsidi yang diterima oleh RT tersebut bergantung pada konsumsi listrik bulanan mereka [4]. Rumah tangga katagori R-1/TR dengan daya 450 VA mendapat subsidi 100% (gratis) untuk pemakaian hingga 720 jam menyala, sementara dengan daya 900 VA mendapat subsidi sebesar 50% dengan ketentuan yang sama [5]. Penyaluran subsidi ini melibatkan anggaran yang sangat besar. Pada tahun 2023, alokasi anggaran untuk subsidi listrik mencapai 67,42 triliun rupiah dan diperkirakan meningkat menjadi 73,54 triliun rupiah pada tahun 2024 [6]. Oleh karena itu, efektivitas penyaluran daya subsidi listrik menjadi krusial agar anggaran dapat digunakan secara efektif dan tepat sasaran.

This is an open access article under the CC BY-SA license.
DOI: https://doi.org/10.26905/jtmi.v11i1.14661

Salah satu cara untuk meningkatakan efektivitas program SLTS adalah dengan melakukan klasifikasi daya subsidi listrik berdasarkan kebutuhan daya 450 VA atau 900 VA. Klasifikasi yang akurat dapat membantu pemerintah memperluas cakupan subsidi ke lebih banyak RT yang benar-benar memerlukan bantuan, sekaligus menghindari penyaluran yang salah sasaran. Namun, proses klasifikasi yang saat ini masih menghadapi sejumlah kendala, seperti kesalahan klasifikasi yang menyebabkan RT yang berhak menerima subsidi daya 450 VA justru memperoleh subsidi daya 900 VA dan sebaliknya.

Penelitian ini bertujuan untuk mengoptimalkan metode KNN dengan menggunakan PSO) dalam melakukan klasifikasi daya RT penerima subsidi listrik berdasarkan karakteristik sosial-ekonomi mereka. Data mining, sebagai bidang ilmu yang bertujuan untuk menemukan pola dan pengetahuan dari data yang besar, sering digunakan dalam penyelesaian masalah prediksi, termasuk klasifikasi [7]. KNN merupakan salah satu metode klasifikasi yang paling sederhana dan efektif [8]. KNN bekerja dengan cara mengukur jarak terpendek antara data baru dan data pelatihan, lalu mengelompokkan data baru tersebut berdasarkan tetangga terdekatnya. Jarak yang digunakan umumnya adalah *Euclidean distance* [9]. Meskipun sederhana, KNN memiliki kelemahan dalam hal pemilihan parameter yang tepat, yang dapat mempengaruhi akurasi hasil klasifikasi.

Untuk mengatasi masalah tersebut, PSO digunakan untuk mengoptimalkan pemilihan parameter KNN. PSO merupakan metode optimasi berbasis populasi yang meniru perilaku burung atau ikan dalam mencari makanan. PSO memiliki tiga parameter utama, yaitu bobot inersia *(inertia weight),* komponen *kognitif (cognitive component),* dan komponen sosial *(social component),* yang berperan penting dalam mempercepat konvergensi pencarian solusi optimal [10]. Dengan menggunakan PSO, diharapkan proses optimasi parameter KNN dapat dilakukan lebih cepat dan memperbaiki akurasi.

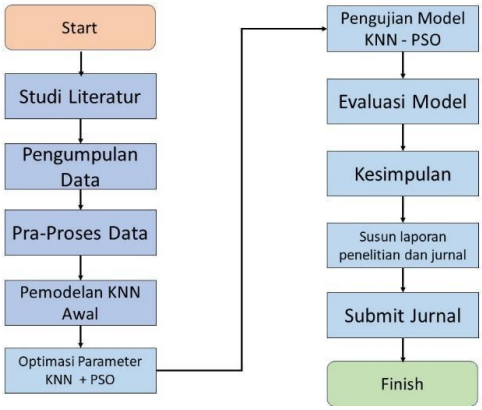
Beberapa penelitian terkait yang mendukung penelitian ini diantaranya penelitian Utiarahman SA dan Dalai, menggunakan metode *Analytical Hierachy Process* (AHP) untuk menentukan kelayakan RT penerima subsidi listrik gratis [1]. Menurut Gustian et al. membandingkan metode C4.5 yang dioptimalkan dengan *Genetic Algorithm* (GA) dan PSO untuk klasifikasi penerima subsidi listrik, dan menemukan bahwa metode C4.5-GA memiliki akurasi tertinggi, meskipun begitu cakupan datanya terbatas hanya pada satu desa saja [11]. Sedangkan Jannah et al. menggunakan metode KNN untuk mengklasifikasikan penerima subsidi di Desa Blang Payang. Namun hasil akurasi yang diperoleh hanya sebesar 66,6% masih diperlukan optimasi lebih lanjut untuk dapat meningkatkan akurasi [12]. Penelitian terakhir adalah oleh Hutahean et al. yang membandingkan kinerja KNN dan *Support Vector Machine* (SVM) dalam mengklasifikasikan RT penerima subsidi listrik di Provinsi Gorontalo. Dalam penelitian ini KNN menunjukkan kinerja lebih baik dengan akurasi 98,07% [13].

Penelitian-penelitian diatas menunjukkan bahwa masih terdapat ruang untuk meningkatkan akurasi klasifikasi, penelitian-penelitian tersebut juga hanya fokus pada klasifikasi penerima subsidi dan bukan penerima subsidi. Penelitian ini menawarkan kebaruan bahwa dataset ini merupakan data P3KE hasil pemuktahiran 2021-2022, yang mana RT dengan subsidi listrik dengan daya terpasang sama dengan 900 dan kurang dari 900 untuk mendefenisikan daya 450 VA. Selain itu, penelitian ini menerapkan optimasi KNN dengan PSO untuk memperbaiki akurasi klasifikasi, terutama melalui pemilihan parameter *k* yang optimal. Optimasi ini bertujuan menentukan jumlah tetangga terdekat terbaik dalam algoritma KNN, sehingga menghasilkan keseimbangan antara *overfitting* dan *underfitting.* Dengan metode ini, diharapkan pemerintah dapat mengidentifikasi daya penerima subsidi listrik dengan lebih tepat, sehingga program SLTS dapat berjalan efisien dan efektif. Implementasi metode yang diusulkan dalam penelitian ini diharapkan dapat memberikan kontribusi signifikan terhadap peningkatan efektivitas program SLTS. Dengan klasifikasi yang lebih akurat, dapat mengalokasikan sumber daya yang paling dibutuhkan. Selain itu, pendekatan ini juga membantu mengidentifikasi pola-pola sosial ekonomi yang mungkin terlewatkan dalam metode klasifikasi lain, memberikan wawasan berharga bagi pembuat kebijakan dalam merumuskan

strategi subsidi listrik yang lebih komprehensif di masa depan.

METODE

Penelitian ini merupakan penelitian kuantitatif dengan pendekatan eksperimental yang dilaksanakan selama 1 tahun periode tahun 2023-2024. Rincian komponen metodologi penelitian dapat dilihat pada gambar 1. Penelitian dilaksanakan di Provinsi Gorontalo dengan menggunakan data dari Dinas Energi dan Sumber Daya Mineral Provinsi Gorontalo. Data tersebut adalah periode pemuktahiran 2021-2022. Target atau sasaran dan subjek penelitian adalah Daya Rumah Tangga penerima subsidi listrik gratis di Provinsi Gorontalo. Populasi penelitian mencakup seluruh Rumah Tangga yang terdaftar dalam Data Penyasaran Percepatan Penghapusan Kemiskinan Ekstrem (P3KE) dengan data yang lengkap.



Gambar 1. Tahapan Penelitian

Tahap awal yang dilakukan adalah dengan melakukan studi literatur terkait data mining, model KNN, optimasi PSO dan klasifikasi, serta mencari informasi peraturan-peraturan dari Kementerian ESDM dan Gubernur. Tahap kedua melakukan pengumpulan data pelanggan listrik bersubsidi yang peneliti kumpulkan dari Dinas ESDM Provinsi Gorontalo. Data yang diberikan kepada peneliti berupa data hasil pemuktahiran basis data keluarga Indonesia Provinsi Gorontalo 2021-2022, yang disebut Data Penyasaran Percepatan Penghapusan Kemiskinan Ekstrem (P3KE). Tahap ketiga yaitu pra pemrosesan data. Diawali dari tahapan ini, peneliti menggunakan algoritma dioptimalkan menggunakan *library* sperti *scikit-learn* di *Phyton.* Sebelum pemrosesan data dilakukan analisis kesesuaian data berupa kesesuaian dan keterkaitan antar variabel pada data, pembersihan data dari data kosong, serta normalisasi data karena sebagian besar data merupakan data katagori. Tahap keempat yaitu pemodelan KNN Awal. Untuk validasi kinerja model menggunakan parameter *accuracy* dan *confusion matrix.* Tahap kelima melakukan optimasi parameter KNN dengan PSO dan tahap keenam pengujian model KNN dan PSO yang juga validasi kinerja model menggunakan parameter *accuracy* dan *confusion matrix*. Tahap ketujuh evaluasi model dengan membandingkan hasil model KNN awal dengan hasil KNN yang telah di optimalkan dengan PSO. Pada tahap evaluasi, performa model akan diukur untuk mengevaluasi akurasi dan kualitas data pelatihan yang digunakan. Pengujian akan dilakukan dengan *confusion matrix* yang memberikan pemahaman kesalahan klasifikasi secara mendalam melalui kalkulasi nilai *True Positive* (TP), *False Positive* (FP), *True Negative* (TN) dan *False Negative* (FN). Tahap kedelapan menarik kesimpulan dari hasil perbandingan klasifikasi dan optimasi. Tahap kesembilan membuat laporan penelitian dan artikel jurnal tahap terakhir adalah publikasi penelitian.

Subsidi Listrik

Pemerintah Indonesia telah menerapkan kebijakan subsidi listrik untuk mengurangi biaya listrik bagi masyarakat berpenghasilan rendah. Tujuan dari subsidi ini adalah untuk memastikan bahwa kelompok masyarakat yang paling rentan tidak terbebani oleh biaya listrik yang tinggi, sehingga mereka dapat membelanjakan lebih banyak uang untuk kebutuhan dasar seperti makanan, pendidikan dan kesehatan [14]. Dalam pemberian subsidi listrik, agar

memastikan subsidi tarif listrik lebih akurat bagi RT pelanggan listrik dan mendukung perubahan dalam metode pemuktahiran data kesejahteraan sosial berbasis web, telah diterbitkan Peraturan Menteri Energi dan Sumber Daya Mineral Nomor 3 tahun 2024. Peraturan ini mengatur tentang subsidi tarif listrik bagi pelanggan rumah tangga PT. PLN (Persero).

Rumah Tangga yang menggunakan daya lebih dari 900 VA dan terdaftar dalam data dasar berhak mendapatkan subsidi listrik setelah kapasitas dayanya diturunkan menjadi 900 VA untuk katagori RT kecil pada tegangan rendah (R-1/TR). Selain itu, RT miskin dan tidak mampu dengan penggunaan daya diatas 900 VA juga layak menerima subsidi setelah daya mereka dikurangi menjadi 900 VA [15].

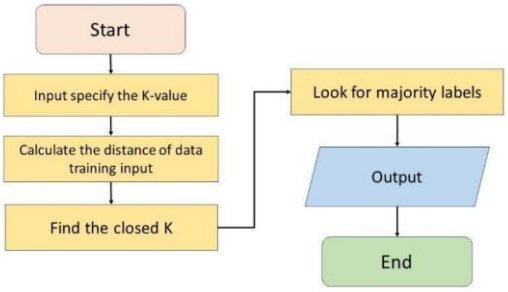
***K-Nearest Neighbor* (KNN)**

Algoritma KNN merupakan teknik klasifikasi data yang yang digunakan untuk menentukan kelompok suatu titik data berdasarkan kedekatannya dengan titik data lainnya. KNN adalah algoritma *lazy-learning* [16], beroperasi berdasarkan prinsip perhitungan jarak [17], yang hanya memperkirakan fungsinya secara lokal dan tidak menyelesaikan semua perhitungan sampai tahap klasifikasi. KNN bekerja dengan cara mencari k item yang paling mirip dari data latih untuk mengelompokkan objek pada data uji atau data baru. Langkah-langkah yang diperlukan untuk menghitung KNN adalah sebagai berikut [18]:

a. Menentukan parameter k
b. Menghitung jarak antara data latih dan data uji
c. Perhitungan jarak yang paling umum digunakan pada algoritma KNN adalah perhitungan jarak *Euclidean*. Rumusnya ditunjukkan pada persamaan 1.

$$euc = \sqrt{(\sum \quad (pi - qi)2\ n\ i = 1} \qquad (1)$$

Rumus jarak *Euclidean* ini untuk mengukur kedekatan antar titik data. Rumus ini menghitung akar kuadrat dari jumlah selisih kuadrat antara fitur-fitur yang bersesuaian dari dua titik data. [18].



Gambar 2. Tahapan *K-Nearest Neighbor* (KNN)

a. *Start* – tahap awal memulai algoritma
b. *Input specify the K-value* – tahap menentukan nilai k yang digunakan dalam algoritma. Nilai k menunjukkan berapa tetangga terdekat yang akan dipertimbangkan.
c. *Calculater the distance of data training input* – tahap menghitung jarak antara data input dengan semua data *training* yang ada. Biasanya menggunakan metode seperti *Euclidean Distance.*
d. *Find the closed k* – tahap mencari label mayoritas dari k tetangga terdekat yang telah ditemukan. Label yang paling banyak muncul akan dipilih.
e. *Output* – tahap mengeluarkan klasifikasi berdasarkan label mayoritas yang ditemukan.
f. *End* – tahap akhir algoritma selesai dijalankan.

***Particle Swarm Optimization* (PSO)**

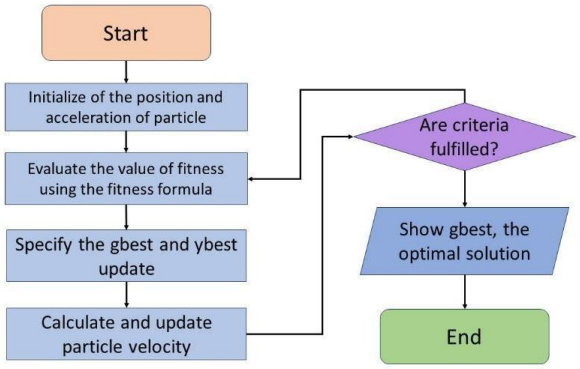
PSO merupakan sebuah evolusi algoritma relative baru bekerja seperti burung mencari makan [19]. PSO banyak dimanfaatkan untuk memecahkan masalah optimasi [20]. Rumus perhitungan PSO ditunjukkan pada persamaan 2 dan 3 [21].

$$v_j^{t+1} = w.v_j^t + c_1.r_1\left(pBest_j^t - x_j^t\right) + c_2.r_2\left(gBest_j^t - x_j^t\right) \quad (2)$$

$$x_j^{t+1} = x_j^t + v_j^{t+1} \quad (3)$$

Dimana :
v_j^t : kecepatan partikel j pada iterasi ke-t
x_j^t : posisi partikel j pada iterasi ke-t
w : bobot inersia
c_1 & c_2 : *learning rates*
r_1 & r_2 : bilangan acak yang bernilai antara 0 sampai 1
$pBest_j^t$: posisi terbaik dari partikel j pada iterasi ke-t
$gBest_j^t$: global optimal dari partikel j pada iterasi ke t

Tahapan algoritma PSO dalam optimasi ditunjukkan pada Gambar 3.



Gambar 3. Tahapan Algoritma PSO

a. *Start* – tahap awal memulai algoritma
b. *Initialize of the position and acceleration of particle* – tahap inisialisasi posisi awal dan percepatan dari setiap partikel dalam populasi. Ini adalah langkah pertama dimana partikel-partikel ditempatkan secara acak dalam ruang pencarian
c. *Evaluate the value of fitness using the fitness formula* – tahap mengevaluasi nilai fitness dari setiap partikel menggunakan fungsi objektif yang telah ditentukan. Ini untuk mengukur seberapa baik solusi yang dihasilkan .
d. *Specify the gbest and ybest update* – tahap memperbarui nilai *gbest* (global best) yaitu posisi terbaik yang ditemukan oleh seluruh kawanan, dan *pbest* (personal best) yaitu posisi terbaik yang pernah dicapai oleh masing-masing partikel.
e. *Calculate and update particle velocity* – tahap menghitung dan memperbarui kecepatan partikel berdasarkan *gbest* dan *pbest* yang telah ditentukan. Kecepatan ini akan mempengaruhi pergerakan partikel di iterasi berikutnya.
f. *Are criteria fulfilled?* – tahap pengecekan apakah kriteria penghentian algoritma sudah terpenuhi. Jika belum, proses akan kembali ke tahap evaluasi fitness.
g. *End* – tahap akhir algoritma selesai dijalankan.

Evaluasi Model

Evaluasi model merupakan langkah krusial setelah pembentukan model. Performa model diukur untuk mengevaluasi akurasi dan kualitas data pelatihan yang digunakan [22]. Ukuran evaluasi kinerja model didasarkan *confusion matrix* yang ditunjukkan pada Tabel 1.

Tabel 1. *Confusion Matrix* [23]

		Katagori Prediksi	
		Positive	***Negative***
Katagori Aktual	*Positive*	TP	FN
	Negative	FP	TN

Confusion matrix terdiri dari akurasi, *precision* dan *f1-score*. Akurasi dalam *confusion matrix* adalah presentase kebenaran dari prediksi pada data uji. *Precision* adalah ukuran proporsi prediksi *True Positive* (TP). Akurasi, *precision, recall* dan *F1-score* diperoleh sebagai persamaan yang ditunjukkan pada persamaan 4.

$$Accuracy = \frac{TP+TN}{TP+FP+TN+FN} \quad (4)$$

HASIL DAN PEMBAHASAN

Data dalam penelitian ini terdiri dari 98.859 *records* dengan 40 kolom variabel. Data ini dalam format *Excel.* Setelah melalui analisis kesesuaian variabel yang relevan dan mencerminkan aspek sosial-ekonomi dan kondisi kehidupan yang mempengaruhi kriteria daya subsidi listrik yang diterima, maka peneliti mengambil 22 variabel ditunjukkan pada Tabel 2.

Tabel 2. Variabel Data

x	*Variabel*	*Label*	*Field*
x1	Desil Kesejahteraan	1, 2, 3, 4	String
x2	Pekerjaan Kepala Keluarga	Nelayan, pedagang, pegawai swasta, pejabat negara, pekerja lepas, pensiunan, petani, PNS/TNI/Polri, tidak bekerja	Katagorikal
x3	Status Pekerjaan Kepala Keluarga	Berusaha dibantu buruh tetap, berusaha dibantu buruh tidak tetap, berusaha sendiri, karyawan/pegawai, pekerja bebas, pekerja keluarga/tidak dibayar, tidak ada	Katagorikal
x4	Jenis Kelamin KK	Laki-laki, perempuan	Katagorikal
x5	Pendidikan KK	Masih SD, masih SLTA, masih SLTP, tamat PT/akademi, Tamat SD, tamat SLTP, tamat SLTA, tidak tamat SD, tidak sekolah	Katagorikal
x6	Kepemilikan Rumah	Bebas sewa/menumpang, dinas, kontrak/sewa, lainnya, milik sendiri	Katagorikal
x7	Memiliki Simpanan uang/perhiasan/ternak/lainnya	Tidak, ya	Katagorikal
x8	Jenis Atap	Asbes/seng, bambu, beton, genteng, jerami/ijuk, kayu/sirap, lainnya	Katagorikal
x9	Jenis Dinding	Bambu, kayu/papan, seng, tembok, lainnya	Katagorikal
x10	Jenis Lantai	Bambu, kayu/papan, keramik/granit/marmer/ubin/ tegel, semen, tanah	Katagorikal
x11	Bahan Bakar Memasak	Arang/kayu, listrik/gas, minyak tanah	Katagorikal

x	*Variabel*	*Label*	*Field*
x12	Sumber Air	Air hujan, air kemasan, air permukaan, ledeng/PAM, sumur bor, sumur terlindung, sumur tak terlindung	Katagorikal
x13	Memiliki Fasilitas Buang Air Besar	Milik sendiri, umum	Katagorikal
x14	Resiko Stunting	Beresiko stunting, bukan target sasaran, tidak beresiko stunting	Katagorikal
x15	Penerima Bantuan Pangan Non Tunai	Tidak, ya	Katagorikal
x16	Penerima Bantuan Produktif Usaha Mikro	Tidak, ya	Katagorikal
x17	Penerima Bantuan Sosial Tunai	Tidak, ya	Katagorikal
x18	Penerima Program Keluarga Harapan	Tidak, ya	Katagorikal
x19	Penerima Sembako	Tidak, ya	Katagorikal
x20	Penerima Prakerja	Tidak, ya	Katagorikal
x21	Penerima KUR	Tidak, ya	Katagorikal
y	Daya Listrik Terpasang	=900 dan <900	Katagorikal

Pada Tabel 2 Informasi dataset menjelaskan informasi data yang terdapat 22 variabel dan 1 kelas target yaitu Daya Listrik Terpasang yang dibagi menjadi dua katagori: sama dengan 900 VA dan kurang dari 900 VA (didefenisikan sebagai 450 VA).

Pra Pemrosesan Data

Penelitian selanjutnya pada tahap pra pemrosesan data. Proses ini bertujuan untuk meningkatkan kemungkinan keberhasilan tahap pengolahan lebih lanjut [24], dataset yang digunakan terdiri dari 98.859 record dengan 22 variabel prediktor dan 1 variabel target berupa daya listrik terpasang. Variabel prediktor mencakup berbagai aspek sosial ekonomi seperti desil kesejahteraan, pekerjaan kepala keluarga, status kepemilikan rumah, dan berbagai indikator kesejahteraan lainnya.

Data *Cleaning*

Proses data *cleaning* dimulai dengan melakukan pengecekan terhadap nilai yang hilang *(missing value)* pada seluruh dataset. Hasil pemeriksaan menunjukkan bahwa tidak ditemukan adanya *missing value* dalam dataset, sehingga tidak diperlukan teknik imputasi data. Langkah selanjutnya adalah melakukan normalisasi data katagorikal menjadi numerik melalui proses *encoding*. Proses *encoding* ini sangat penting mengingat sebagian besar variabel dalam dataset berupa data katagorikal seperti jenis pekerjaan, tingkat pendidikan dan status kepemilikan rumah.

Transformasi data katagorikal menjadi numerik dilakukan menggunakan teknik label *encoding* dan *one-hot encoding* sesuai dengan karakteristik masing-masing variabel. Sebagai contoh untuk variabel ‘Desil Kesejahteraan’ yang memiliki nilai 1, 2, 3, dan 4 dilakukan label *encoding*, sementara untuk variabel ‘Jenis atap’ dan ‘Jenis dinding’ yang memiliki beberapa katagori tidak berurutan diterapkan *one-hot encoding*. Proses *encoding* ini esensial untuk memungkinkan algoritma KNN dan PSO dapat memproses data dengan efektif, mengingat kedua algoritma tersebut bekerja dengan data numerik. Data setelah *encoding* ditunjukkan pada Gambar 4.

```
# Data training setelah encoding
print(train_data.head())
```

```
   Desil Kesejahteraan  Pekerjaan Kepala Keluarga  \
0                    1                          7
1                    1                          5
2                    1                          5
3                    1                          5
4                    1                          5

   Status Pekerjaan Kepala Keluarga  Jenis Kelamin Kepala Keluarga  \
0                                 7                              0
1                                 7                              0
2                                 7                              0
3                                 0                              0
4                                 4                              0

   Pendidikan Kepala Keluarga  Kepemilikan Rumah  \
0                           5                  4
1                           6                  4
2                           7                  4
3                           8                  4
4                           8                  0

   Memiliki Simpanan Uang/Perhiasan/Ternak/Lainnya  Jenis Atap  Jenis Dinding  \
0                                                0           0              4
1                                                0           0              4
2                                                0           0              4
3                                                0           0              1
4                                                0           0              1

   Jenis Lantai  ...  Memiliki fasilitas Buang Air Besar  Resiko Stunting  \
0             4  ...                                   1                1
1             4  ...                                   2                1
2             4  ...                                   2                1
3             1  ...                                   2                0
4             1  ...                                   2                0

   Penerima BPNT  Penerima BPUM  Penerima BST  Penerima PKH  Penerima SEMBAKO  \
0              1              0             1             0                 1
1              0              0             0             0                 0
2              0              0             0             0                 0
3              0              0             0             0                 0
4              0              0             0             0                 0

    Penerima Prakerja  Penerima KUR  Daya Listrik Terpasang
0                   0             0                       0
1                   0             0                       0
2                   0             0                       0
3                   0             1                       0
4                   0             1                       0

[5 rows x 22 columns]
```

Gambar 4. Data Setelah E*ncoding*

Pembagian Data

Setelah proses data *cleaning* selesai, tahap selanjutnya adalah pembagian dataset menjadi data *training* dan data *testing.* Pembagian data menggunakan rasio 80:20, dimana 80% data (79,087 *records)* dialokasikan sebagai data *training* dan 20% sisanya (19,772 *records)* sebagai data *testing.* Pemilihan rasio ini didasarkan pada pertimbangan untuk menyediakan jumlah data yang cukup besar untuk proses pembelajaran model, sekaligus menyisakan porsi yang memadai untuk pengujian.

Implementasi Model KNN

Setelah tahap pra pemrosesan data, data *cleaning* dan membagi data selesai, langkah yang peneliti lakukan adalah mengimplementasikan algoritma KNN untuk klasifikasi daya listrik terpasang. Implementasi awal KNN dilakukan dengan menggunakan parameter *default* untuk memberikan *baseline* performa model sebelum dilakukan optimasi.

Analisis Hasil Klasifikasi KNN

Hasil implementasi model KNN terlihat pada Gambar 5. Gambar tersebut menunjukkan hasil klasifikasi menggunakan *confusion matrix*, dimana terlihat jelas perbedaan jumlah sampel antara kedua kelas. Kelas 0 (daya listrik < 900 VA) mendominasi dataset dengan 16.215 sampel atau sekitar 82% dari total data, sementara Kelas 1 (daya listrik = 900 VA) hanya memiliki 3.557 sampel atau 18% dari dataset.

```
Default KNN Classification Report:
              precision    recall  f1-score   support

           0       0.86      0.95      0.90     16215
           1       0.58      0.32      0.41      3557

    accuracy                           0.84     19772
   macro avg       0.72      0.63      0.66     19772
weighted avg       0.81      0.84      0.82     19772

Default KNN Accuracy: 0.835272101962371
```

Gambar 5. Hasil Klasifikasi Model KNN

Evaluasi performa model KNN awal menghasilkan akurasi keseluruhan sebesar 83,52%. Untuk kelas 0, model menunjukkan performa dengan *precision* 0,86, *recall* 0,95, dan *f1-score* yang mengindikasikan keseimbangan yang baik antara *precision* dan *recall.* Disisi lain, performa model pada kelas 1 menunjukkan hasil yang kurang optimal. *Precision* sebesar 0,58 dan *recall* 0,32.

Optimalisasi KNN menggunakan PSO

Untuk meningkatkan performa klasifikasi model KNN, penelitian ini mengimplementasikan algoritma PSO sebagai metode optimasi parameter. Proses optimasi di fokuskan pada pencarian nilai *k* optimal yang dapat menghasilkan akurasi klasifikasi terbaik. Implementasi PSO dimulai dengan menginisialisasi parameter-parameter kunci yang telah dikalibrasi secara cermat untuk memastikan eksplorasi ruang pencarian yang efektif. Parameter PSO yang digunakan meliputi jumlah partikel sebanyak 30, iterasi maksimum 100 *cognitive component* (c1) dan *social component* (c2) masing-masing sebesar 2.0, serta *inertia weight* (w) sebesar 0.7. Range pencarian nilai *k* ditetapkan antara 1 hingga 20, dengan fungsi fitness berupa akurasi klasifikasi. Setiap partikel dalam algoritma PSO merepresentasikan kandidat nilai *k* dalam bentuk integer dan bergerak dalam ruang pencarian untuk menemukan solusi optimal.

Kontribusi PSO dalam Optimasi Parameter KNN

PSO memberikan kontribusi signifikan dalam mengoptimalkan kinerja KNN melalui tiga aspek utama. Pertama, PSO mengimplementasikan mekanisme pencarian nilai *k* optimal secara otomatis dengan melakukan eksplorasi sistematis dalam range nilai *k* dari 1 hingga 20. Dalam proses ini, setiap partikel merepresentasikan kandidat nilai *k* yang berbeda, dimana algoritma secara dinamis menyesuaikan pencarian berdasarkan performa klasifikasi yang dihasilkan. Hal ini memungkinakan PSO untuk menemukan nilai *k* yang menghasilkan akurasi klasifikasi terbaik tanpa perlu mencoba setiap nilai *k* secara manual. Hasil optimasi ini dapat dilihat pada Gambar 7 yang menunjukan *confusion matrix* dari model KNN yang telah dioptimasi menggunakan PSO, dimana terlihat peningkatan performa klasifikasi dibandingkan dengan model KNN awal.

```
Optimal KNN (k=17) Classification Report:
              precision    recall  f1-score   support

           0       0.86      0.98      0.91     16215
           1       0.73      0.25      0.37      3557

    accuracy                           0.85     19772
   macro avg       0.79      0.61      0.64     19772
weighted avg       0.83      0.85      0.82     19772

Optimal KNN (k=17) Accuracy: 0.8482702812057455
```

Gambar 7. Hasil Klasifikasi Model KNN yang Dioptimasi dengan PSO

Kedua PSO menerapkan optimasi parameter yang adaptif melalui inisialisasi 30 partikel dengan posisi acak untuk mencakup ruang pencarian yang luas. Parameter *cognitive component* (c1) dan *social component* (c2) diset sebesar 2.0 untuk memastikan keseimbangan yang optimal antara eksplorasi individual setiap partikel dan pembelajaran kolektif dari *swarm.* Sementara itu, *inertia weight* (w) sebesar 0.7 berfungsi mengontrol momentum partikel, mencegah konvergensi prematur dan memastikan eksplorasi ruang pencarian yang menyeluruh.

Aspek Ketiga adalah proses konvergensi yang efisien, dimana PSO mencapai konvergensi pada iterasi ke-45 dari 100 iterasi maksimum yang ditetapkan. Dalam proses ini, PSO berhasil menemukan nilai *k* optimal sebesar 17 yang mneghasilkan peningkatan akurasi klasifikasi. Hal ini menunjukkan bahwa algoritma PSO berhasil menemukan solusi optimal secara konsisten dan efisien dalam mengoptimalkan parameter KNN. Ketiga aspek kontribusi PSO bekerja secara sinergis untuk mengoptimalkan parameter KNN, menghasilkan model klasifikasi yang akurat untuk menentukan daya listrik yang diterima oleh RT.

Analisis Komparatif Hasil Optimasi

Implementasi PSO dalam optimasi parameter KNN menunjukkan peningkatan performa yang dapat diukur secara kuantitaif melalui berbagai metrik evaluasi. Dari segi akurasi global, model yang dioptimasi dengan PSO mencapai peningkatan sebesar 1.3% dari 83,53% menjadi 84,83%. Peningkatan ini mengindikasikan bahwa optimasi parameter menggunakan PSO efektif dalam meningkatkan kemampuan klasifikasi model secara keseluruhan.

Analisis lebih detail pada metrik per kelas menunjukkan hasil yang bervariasi. Untuk kelas 0, model menunjukkan perbaikan yang konsisten dimana *precision* tetap stabil pada 0,86, sementara *recall* mengalami peningkatan dari 0,95 menjadi 0,98 menandakan kemampuan model yang lebih baik dalam mengindentifikasi kasus posistif. *F1-score* untuk kelas 0 juga mengalami peningkatan dari 0,90 menjadi 0,91 mengindikasikan keseimbangan yang lebih baik antara *precision* dan *recall.*

Pada kelas 1, hasil optimasi menunjukkan *trade-off* yang menarik. *Precision* mengalami peningkatan signifikan dari 0.58 menjadi 0.71, menunjukkan bahwa model yang dioptimasi lebih akurat dalam memprediksi kelas minoritas. Namun, terjadi penurunan *recall* dari 0,32 menjadi 0,25 mengindikasikan model menjadi lebih selektif dalam mengklasifikasikan model menjadi lebih selektif dalam mengklasifikasikan instance sebagai kelas 1. Konsekuansinya, *f1-score* untuk kelas 1 mengalami penyesuaian dari 0.41 menjadi 0.37, mencerminkan *trade-off* antara peningkatan *precision* dan penurunan *recall.*

Evaluasi Model

Kedua model KNN (k=5 dan optimasi PSO dengan k=17) dievaluasi menggunakan beberapa metrik kinerja klasifikasi. Evaluasi difokuskan pada akurasi keseluruhan, serta *precision, recall* dan *f1-score* untuk masing-masing kelas. Perbandingan antara kedua model ditunjukkan pada Tabel 3.

Tabel 3. Perbandingan evaluasi model

Metrik	KNN (k=5)	KNN optimal PSO (k=17)
Akurasi	83.53%	84,83%
Precision kelas 0	0,86	0,86
Recall kelas 0	0,95	0,98
F1-score kelas 0	0,90	0,91
Precision kelas 1	0,58	0,71
Recall kelas 1	0,32	0,25
F1-score kelas 1	0,41	0,37
Macro Avg f1	0,66	0,64
Weighted Avg f2	0,82	0,82

Berdasarkan Tabel 3 dijelaskan bahwa akurasi keseluruhan KNN dengan optimasi PSO menunjukkan peningkatan akurasi sebesar 1,3% dibandingkan model KNN. Hal ini menunjukkan bahwa optimasi hyperparameter efektif meningkatkan kinerja keseluruhan. Kinerja kelas 0 dimana *precision* tetap konsisten (0,86) untuk kedua model. *Recall* meningkat sedikit dari 0,95 menjadi 0,98 dengan optimasi PSO, menunjukkan peningkatan dalam mengidentifikasi *instance* kelas 0. *F1-score* sedikit meningkat dari 0,90 menjadi 0,91 menandakan perbaikan kecil dalam keseimbangan *precision* dan *recall.*

Kinerja kelas 1 menunjukkaan hasil yang bervariasi, dimana *precision* meningkat signifikan dari 0,58 menjadi 0,71 mengindikasikan model optimasi lebih akurat dalam memprediksi kelas 1. Namun, terjadi penurunan *recall* dari 0,32 menjadi 0,25 menunjukkan

berkurangnya kemampuan model dalam mengenali instance kelas 1. *F1-score* sedikit menurun dari 0,41 menjadi 0,37 mencerminkan *trade-off* antara peningkatan *precision* dan penurunan *recall.*

Rata-rata metrik menunjukkan *Macro average f1-score* sedikit menurun dari 0,66 menjadi 0,64 menunjukkan sedikit penurunan dalam kinerja rata-rata kedua kelas. *Weighted average f1-score* tetap konsisten pada 0,82 menandakan bahwa kinerja keseluruhan, dengan mempertimbangkan ketidakseimbangan kelas relatif tidak berubah.

SIMPULAN DAN SARAN

Berdasarkan hasil penelitian yang dilakukan, dapat disimpulkan bahwa optimasi metode KNN menggunakan PSO berhasil meningkatkan akurasi klasifikasi daya subsidi listrik di Provinsi Gorontalo. Peningkatan akurasi sebesar 1.3% dari 83.53% menunjukkan efektifitas PSO dalam mengoptimalkan parameter KNN. PSO berhasil menemukan nilai *k* optimal yang menghasilkan performa klasifikasi lebih baik, khususnya dalam peningkatan *precision* pada kedua kelas. Untuk kelas 0, model mengalami peningkatan *recall* dari 0.95 menjadi 0.98 dan *f1-score* dari 0.90 menjadi 0.91. pada kelas 1, terjadi peningkatan *precision* yang signifikan dari 0.58 menjadi 0.71, meskipun terdapat *trade-off* berupa penurunan *recall* dari 0.32 menjadi 0.25. Hasil ini menunjukkan bahwa optimasi parameter menggunakan PSO mampu meningkatkan selektivitas model dalam melakukan klasifikasi, terutama untuk kelas 1. Untuk penelitian selanjutnya, disarankan untuk mengeksplorasi beberapa aspek pengembangan. Pertama, penerapan teknik penanganan ketidakseimbangan kelas seperti *resampling* dapat dipertimbangkan untuk meningkatkan perfoma klasifikasi. Kedua, penambahan fitur-fitur yang lebih diskriminatif dapat dieksplorasi untuk meningkatkan akurasi klasifikasi. Ketiga, implementasi metode *ensemble learning* dapat dipertimbangkan untuk meningkatkan *robustness* model. Selain itu, pengembangan sistem klasifikasi yang lebih komprehensif juga dapat dilakukan dengan mempertimbangkan faktor-faktor sosial ekonomi lainnya yang relevan dengan penentuan daya subsidi listrik. Hasil penelitian memberikan kontribusi pentng dalam upaya meningkatkan efektivitas program Subsidi Listrik Tepat Sasaran (SLTS). Dengan peningkatan akurasi klasifikasi, diharapkan penyaluran daya subsidi listrik dapat dilakukan dengan lebih tepat sasaran, memastikan bahwa RT benar-benar menerima daya sesuai kebutuhan. Hal ini pada gilirannya dapat membantu mengoptimalkan penggunaan anggaran subsidi listrik yang signifikan, serta mendukung upaya pemerintah dalam menjaga stabilitas ekonomi.

DAFTAR PUSTAKA

[1] S. A. Utiarahman and H. Dalai, “Penerapan Metode Analitycal Hierarchy Process dalam Sistem Pendukung Keputusan Kelayakan Rumah Tangga Penerima Listrik Gratis,” *Jurnal Riset Komputer*, vol. 9, no. 5, pp. 1659, 2022, doi: https://doi.org/10.30865/jurikom.v9i5.4857

[2] Kementerian ESDM, “PERMEN energi dan Sumber Daya Mineral tentang Pelaksanaan Usaha Ketenagalistrikan,” 2021.

[3] Kementerian ESDM, “PERMEN ESDM tentang Mekanisme Pemberian Subsidi Tarif Tenaga Listrik Untuk Rumah Tangga.” [Online]. Available: www.peraturan.go.id

[4] Siaran Pers, “Cek Golongan Tarif Listrik PLN,” https://web.pln.co.id/media/2022/06/cek-golongan-tarif-listrik-pln-di-sini-kamu-pelanggan-listrik-yang-dapat-subsidi

[5] K. Wisnubroto, “Ketentuan Baru Subsidi Listrik 2021,” https://indonesia.go.id/

[6] V. Setiawan, “Subsidi Listrik RI Tahun 2023 Tembus Rp67,42 Triliun,” https://www.cnbcindonesia.com/

[7] J. T. Samudra, R. Rosnelly, and Z. Situmorang, “Comparative Analysis of Support Vector Machine and Perceptron In The Classification of Subsidized Fuel Receipts,” *Jurnal Rekayasa Sistem dan Teknologi Informasi*, vol. 7, no. 3, pp. 652–656, 2023, doi: https://doi.org/10.29207/resti.v7i3.4731

[8] A. D. W. M. Sidik, I. H. Kusumah, A. Suryana, M. Artiyasa, and A. P. Junfithrana, "Gambaran Umum Metode Klasifikasi Data Mining," *Fidelity: Jurnal Teknik Elektro*, vol. 2, no. 2, pp. 34–38, 2020, doi: https://doi.org/10.52005/fidelity.v2i2.111

[9] A. Setiawan, "Perbandingan Penggunaan Jarak Manhattan, Jarak Euclid, dan Jarak Minkowski dalam Klasifikasi Menggunakan Metode KNN pada Data Iris," *Jurnal Sains dan Edukasi Sains*, vol. 5, no. 1, pp. 28–37, 2022, doi: https://doi.org/10.24246/juses.v5i1p28-37

[10] M. M. Baharuddin, H. Azis, and T. Hasanuddin, "Analisis Performa Metode K-Nearest Neighbor Untuk Identifikasi Jenis Kaca," *ILKOM Jurnal Ilmiah*, vol. 11, no. 3, pp. 269–274, 2019, doi: https://doi.org/10.33096/ilkom.v11i3.489.269-274

[11] D. Gustian, F. Sembiring, R. Amelia, E. Nurhasanah, S. Waelah, and N. Anggraeni, "Comparison Data Mining based on Optimization Algorithms in Receiving Electricity Subsidies," in *6th International Conference on Computing, Engineering, and Design, ICCED 2020*, 2020. doi: https://doi.org/10.1109/ICCED51276.2020.9415844

[12] M. Jannah, S. Salsabila, N. Faiza, and Mutasar, "Classification of Receiving Electricity Subsidy Assistance in Blang Panyang Village Using the K-NN (K-Nearest Neighbor) Method," *Journal of Advanced Computer Knowledge and Algorithms*, vol. 1, no. 1, pp. 1-4, 2024, doi: https://doi.org/10.29103/jacka.v1i1.14530

[13] Y. M. Hutahaean and A. W. Wijayanto, "Klasifikasi Rumah Tangga Penerima Subsidi Listrik di Provinsi Gorontalo Tahun 2019 dengan Metode K-Nearest Neighbor dan Support Vector Machine," *Jurnal Sistem dan Teknologi Informasi* , vol. 10, no. 1, p. 63, 2022, doi: https://doi.org/10.26418/justin.v10i1.51210

[14] Kementerian Energi dan Sumber Daya Mineral, "Peraturan Menteri Energi dan Sumber Daya Mineral Nomor 8 Tahun 2023 tentang Perubahan Kelima Atas Peraturan Menteri Energi dan Sumber Daya Mineral Nomor 28 Tahun 2016 tentang Tarif Tenaga Listrik Yang Disediakan Oleh PT Perusahaan Listrik Negara (Persero)," https://jdih.esdm.go.id/index.php/web/result/1561/detail

[15] V. K. Ningsih and S. Syalikha, "Implementasi Subsidi Listrik untuk Mendorong Pencapaian SDGs Tujuan 7," *Journal of Economics, Assets, and Evaluation*, vol. 2, no. 3, 2025, doi: https://doi.org/10.47134/jeae.v2i3.366

[16] S. Zhang, "Cost-sensitive KNN classification," *Neurocomputing*, vol. 391, pp. 234–242, 2020, doi: https://doi.org/10.1016/j.neucom.2018.11.101

[17] H. Basri, H. Darwis, F. Umar, and A. History, "Klasifikasi Daun Herbal Menggunakan K-Nearest Neighbor dan Convolutional Neural Network dengan Ekstraksi Fourier Descriptor," *Jurnal Teknologi dan Manajemen Informatika*, vol. 9, no. 2, pp. 79–90, 2023, doi: https://doi.org/10.26905/jtmi.v9i2.10350

[18] M. A. Imron and B. Prasetiyo, "Improving Algorithm Accuracy K-Nearest Neighbor Using Z-Score Normalization and Particle Swarm Optimization to Predict Customer Churn," *J. Soft Comput. Explor.*, vol 1, no. 1, pp. 56–62, 2020, doi: https://doi.org/10.52465/joscex.v1i1.7

[19] Z. Yong, Y. Li-juan, Z. Qian, and S. Xiao-yan, "Multi-objective optimization of building energy performance using a particle swarm optimizer with less control parameters," *Journal of Building Engineering*, vol. 32, 2020, doi: https://doi.org/10.1016/j.jobe.2020.101505

[20] H. T. Rauf, U. Shoaib, M. I. Lali, M. Alhaisoni, M. N. Irfan, and M. A. Khan, "Particle Swarm Optimization with Probability Sequence for Global Optimization," *IEEE Access*, vol. 8, pp. 110535–110549, 2020, doi: https://doi.org/10.1109/ACCESS.2020.3002725

[21] Y. P.Sugandhi, B. Warsito, and A. R. Hakim, "Prediksi Harga Saham Harian Menggunakan Cascade Forward Neural Network (CFNN) dengan Particle Swarm Optimization (PSO)," *Statistika*, vol. 19, no. 2, pp. 71–82, 2019, doi: http://dx.doi.org/10.29313/jstat.v19i2.4878

[22] B. F. Rochman, A. Rahim, and T. A. Y. Siswa, "Optimasi Algoritma KNN dengan Parameter K dan PSO Untuk Klasifikasi Status Gizi Balita," *Jurnal Media Informatika Budidarma*, vol. 8, no. 3, pp. 1609–1616, 2024, doi: https://doi.org/10.30865/mib.v8i3.7841

[23] H. Rasmita Ngemba *et al.*, "Penerapan Algoritma K-Nearest Neighbor untuk Menentukan Potensi Ekspor Komoditas Pertanian di Provinsi Sulawesi Tengah," *Jurnal Teknologi dan Manajemen Informatika*, vol. 9, no. 2, pp. 151–160, 2023, doi: https://doi.org/10.26905/jtmi.v9i2.10235

[24] F. Wibowo *et al.*, "Klasifikasi Tanaman Beringin (Ficus Bernjamina) Berdasarkan Citra Daun Menggunakan Algoritma K-Nearest Neighbors," *Jurnal Teknologi dan Manajemen Informatika*, vol. 7, no. 2, pp. 131–138, 2021, doi: https://doi.org/10.26905/jtmi.v7i2.6758