

**PERBANDINGAN IMPLEMENTASI K-NEAREST NEIGHBORS
DENGAN DAN TANPA SMOTE UNTUK DETEKSI
PENYALAHGUNAAN LISTRIK DI PLN UP3 DEMAK**

Usulan Penelitian Tesis S-2

Untuk memenuhi persyaratan memperoleh gelar Magister Teknik
Program Studi Magister Teknik Elektro



Diajukan oleh
RETNO SUPIYANTI
20602300023

**PROGRAM STUDI MAGISTER TEKNIK ELEKTRO
UNIVERSITAS ISLAM SULTAN AGUNG
SEMARANG
2024**

KATA PENGANTAR

Puji syukur penulis panjatkan kepada Tuhan Yang Maha Esa atas rahmat dan karunia-Nya sehingga saya dapat menyelesaikan tesis yang berjudul "Perbandingan Implementasi K-Nearest Neighbors dengan dan Tanpa SMOTE untuk Deteksi Penyalahgunaan Listrik di PLN UP3 Demak". Tesis ini disusun sebagai salah satu syarat untuk memperoleh gelar Magister Teknik pada Program Studi Magister Teknik Elektro.

Penyusunan tesis ini tidak lepas dari bantuan dan dukungan berbagai pihak. Oleh karena itu, pada kesempatan ini, penulis ingin menyampaikan terima kasih yang sebesar-besarnya kepada:

1. Dr. Ir. Novi Marlyana, ST., MT., IPU., ASEAN Eng, selaku Dekan Fakultas Teknik yang telah memberikan kesempatan dan fasilitas kepada saya untuk melaksanakan penelitian ini.
2. Prof. Dr.Ir. Sri Artini Dwi Prasetyowati, M.Si., selaku pembimbing utama yang telah memberikan bimbingan, saran, dan masukan berharga dalam penyusunan tesis ini.
3. Keluarga tercinta, yang selalu memberikan doa, dukungan moral, dan motivasi yang tiada henti.
4. Rekan-rekan di Program Studi Magister Teknik Elektro, yang telah memberikan bantuan dan kerjasama yang baik selama proses studi ini.
5. Semua pihak yang tidak dapat disebutkan satu per satu, yang telah membantu dan memberikan dukungan dalam penyelesaian tesis ini.

Penulis menyadari bahwa dalam penyusunan tesis ini masih terdapat kekurangan. Oleh karena itu, penulis mengharapkan kritik dan saran yang membangun dari pembaca untuk penyempurnaan tesis ini di masa mendatang.

Akhir kata, semoga tesis ini dapat memberikan manfaat dan kontribusi bagi perkembangan ilmu pengetahuan, khususnya di bidang teknik elektro.

Semarang, 10 Maret 2025



RETNO SUPIYANTI

LEMBAR PENGESAHAN

PERBANDINGAN IMPLEMENTASI K-NEAREST NEIGHBORS DENGAN DAN TANPA
SMOTE UNTUK DETEKSI PENYALAHGUNAAN LISTRIK DI PLN UP3 DEMAK

Yang dipersiapkan dan disusun oleh
Retno Supiyanti
20602300023

Telah dipertahankan di depan Dewan Penguji
Pada tanggal 7 Maret 2025

Susunan Dewan Penguji

Pembimbing



Prof. Dr. Ir. Sri Artini Dwi Prasetyowati, M.Si.
NIDN : 0620026501

Penguji Utama



Dr. Eka Nuryanto Budisusila, S.T., M.T.
NIDN : 0619107301

Penguji Kedua



Dr. Ir. Agus Adhi Nugroho, M.T.
NIDN : 0628086501

Testis ini telah diterima sebagai salah satu persyaratan
untuk memperoleh gelar Magister Teknik
Tanggal 7 maret 2025

Ketua Program Studi Magister Teknik Elektro



Prof. Dr. Ir. Sri Artini Dwi Prasetyowati, M.Si.
NIDN : 0620026501

SURAT PERNYATAAN KEASLIAN TUGAS AKHIR

Yang bertanda tangan dibawah ini:

Nama : Retno Supiyanti
NIM : 20602300023
Jurusan : Magister Teknik Elektro
Fakultas : Fakultas Teknologi Industri

Dengan ini saya menyatakan bahwa Tugas Akhir yang diajukan dengan judul **“PERBANDINGAN IMPLEMENTASI K-NEAREST NEIGHBORS DENGAN DAN TANPA SMOTE UNTUK DETEKSI PENYALAHGUNAAN LISTRIK DI PLN UP3 DEMAK”** adalah hasil karya sendiri, tidak pernah diajukan untuk memperoleh gelar kesarjanaan di perguruan tinggi lain maupun ditulis dan diterbitkan orang lain, kecuali secara tertulis diacu dalam daftar pustaka. Tugas Akhir ini adalah milik saya segala bentuk kesalahan dan kekeliruan dalam Tugas Akhir ini adalah tanggung jawab saya.

Demikian surat pernyataan ini saya buat dengan sadar dan penuh tanggung jawab.

Semarang, 10 Maret 2025

Yang Menyatakan



Retno Supiyanti

PERNYATAAN PERSETUJUAN UNGGAH KARYA ILMIAH

Saya yang bertanda tangan di bawah ini :

Nama	: Retno Supiyanti
NIM	: 20602300023
Program Studi	: Magister Teknik Elektro
Fakultas	: Fakultas Teknologi Industri

Dengan ini menyerahkan karya ilmiah berupa Tugas Akhir/Tesis dengan judul :

PERBANDINGAN IMPLEMENTASI K-NEAREST NEIGHBORS DENGAN DAN TANPA SMOTE UNTUK DETEKSI PENYALAHGUNAAN LISTRIK DI PLN UP3 DEMAK

dan menyetujuinya menjadi hak milik Universitas Islam Sultan Agung serta memberikan Hak Bebas Royalti Non-eksklusif untuk disimpan, dialihmediakan, dikelola dalam pangkalan data, dan dipublikasikannya di internet atau media lain untuk kepentingan akademis selama tetap mencantumkan nama penulis sebagai pemilik Hak Cipta.

Pernyataan ini saya buat dengan sungguh-sungguh. Apabila dikemudian hari terbukti ada pelanggaran Hak Cipta/Plagiarisme dalam karya ilmiah ini, maka segala bentuk tuntutan hukum yang timbul akan saya tanggung secara pribadi tanpa melibatkan pihak Universitas Islam Sultan Agung.

Semarang, 10 Maret 2025
Yang menyatakan,



Retno Supiyanti

DAFTAR ISI

KATA PENGANTAR	ii
DAFTAR ISI.....	vi
DAFTAR GAMBAR	viii
DAFTAR TABEL.....	ix
ABSTRAK	x
BAB I.....	1
PENDAHULUAN.....	1
1.1 Latar Belakang	1
1.2 Rumusan Masalah	2
1.3 Batasan Masalah.....	2
1.4 Tujuan Penelitian	2
1.5 Manfaat Penelitian.....	2
1.6 Keaslian Penelitian.....	3
1.7 Sistematika Penelitian	4
BAB II.....	5
TINJAUAN PUSTAKA DAN LANDASAN TEORI.....	5
2.1 Tinjauan Pustaka	5
2.2 Penertiban Pemakaian Tenaga Listrik P2TL.....	6
2.2.1 Prapemeriksaan Penertiban Pemakaian Tenaga Listrik	7
2.2.2 Golongan Pelanggaran Pemakaian Tenaga Listrik	8
2.3 Ukuran Gejala Pusat.....	10
2.3.1 Mean.....	11
2.3.2 Median.....	11
2.3.3 Modus.....	11
2.4 Normalisasi Data.....	12
2.5 Konsep Alogaritma SMOTE dan <i>K-Nearest Neighbor</i> untuk Klasifikasi	13
2.5.1 <i>K-Nearest Neighbor (K-NN)</i>	13
2.5.2 <i>Synthetic Minority Oversampling Technique (SMOTE)</i>	14
2.5.3 Evaluasi Kinerja Model.....	15
BAB III.....	17
METODOLOGI	17
3.1 Diagram Alur Penelitian.....	17
3.2 Pengumpulan Data	18
3.3 Teknik Pengolahan Data	19
3.3.1 Penentuan Nilai $K=8$	20
3.4 Perancangan Sistem dan Implementasi KNN	20

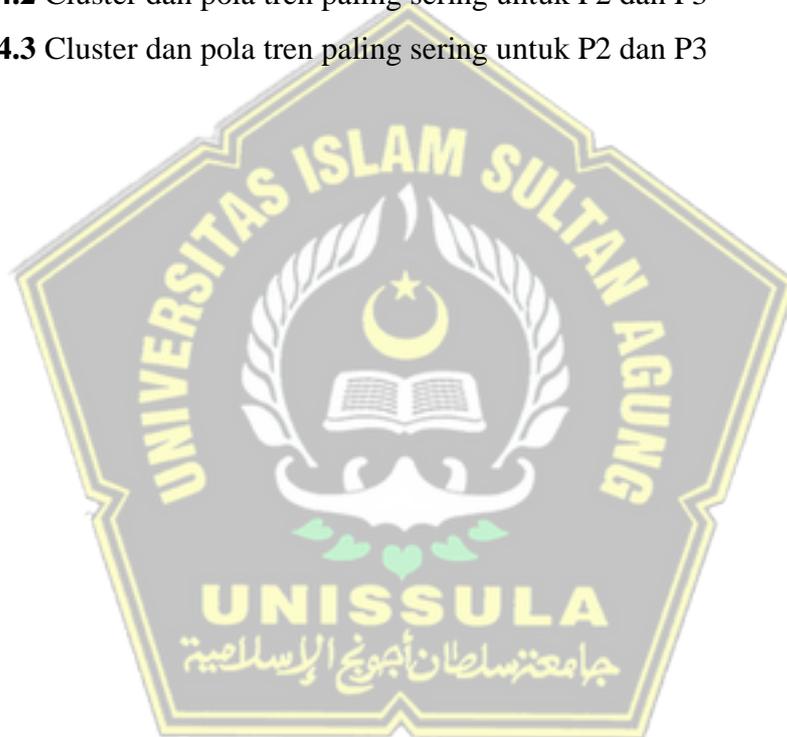
3.4.1	Pengumpulan dan Pra-pemrosesan Data.....	21
3.4.2	Pembagian Data.....	22
3.4.3	Penerapan SMOTE.....	23
3.4.4	Pelatihan Model KNN.....	23
3.4.5	Evaluasi Model.....	29
3.4.6	Clusters of Trend Patterns in 3D.....	29
3.4.7	Penghitungan Jarak dan Prediksi Kelas	31
3.4.8	Jumlah Data Penyalahgunaan.....	32
BAB IV		33
HASIL PENELITIAN DAN PEMBAHASAN		33
4.1	Deskripsi Data	34
4.2	Hasil Tanpa <i>Synthetic Minority Over-sampling Technique</i> (SMOTE)....	36
4.2.1	Preprocessing Data dan Implementasi K-Nearest Neighbors (KNN)	36
4.2.2	Distribusi dan Pola dalam Hasil Clustering	38
4.2.3	Pola Tren Tanpa SMOTE.....	40
4.3	Hasil dengan <i>Synthetic Minority Over-sampling Technique</i> (SMOTE) .	45
4.3.1	Preprocessing Data dan Implementasi K-Nearest Neighbors (KNN)	45
4.3.2	Distribusi dan Pola dalam Hasil Clustering dengan SMOTE	47
4.3.3	Pola Tren dengan SMOTE	49
4.4	Pembahasan.....	54
4.4.1	Kelebihan dan Kekurangan SMOTE Berdasarkan Hasil Analisis	54
4.4.2	Alasan SMOTE dan Penurunan Akurasi dengan Nilai k yang Lebih Baik	55
4.4.3	Pola Tren Pelanggan P2 dan P3 serta Kecenderungan Kecurangan ..	56
4.4.4	Hal yang Harus Diwaspadai Kedepan.....	56
BAB V.....		58
KESIMPULAN DAN SARAN.....		58
5.1	Kesimpulan.....	58
5.2	Saran.....	59
DAFTAR PUSTAKA		60
LAMPIRAN.....		62

DAFTAR GAMBAR

Gambar 3.1 Metode Pengerjaan	15
Gambar 3.2 Pembagian Train Test Split	22
Gambar 3.3 Nilai Terbaik K=1	24
Gambar 3.4 Akurasi Jarak Euclidean dan Tetangga terdekat	25
Gambar 4.1 Hardware Google Colab yang di gunakan	31
Gambar 4.2 Pola Konsumsi Listrik Pelanggan P2 dan P3	32
Gambar 4.3 Hasil Evaluasi K-Nearest Neighbors (KNN) tanpa SMOTE	33
Gambar 4.4 Hasil analisis clustering tanpa SMOTE	34
Gambar 4.5 Hasil Clustering berdasarkan P2 dan P3	35
Gambar 4.6 Tren Cluster untuk P2	36
Gambar 4.7 Klaster 0 P2	37
Gambar 4.8 Klaster 1 P2	37
Gambar 4.9 Klaster 2 P2	38
Gambar 4.10 Tren Cluster untuk P3	38
Gambar 4.11 Tren Cluster 0 untuk P3	39
Gambar 4.12 Tren Cluster 1 untuk P3	39
Gambar 4.13 Tren Cluster 2 untuk P3	40
Gambar 4.14 Hasil Evaluasi K-Nearest Neighbors (KNN)	41
Gambar 4.15 Hasil analisis clustering tanpa SMOTE	42
Gambar 4.16 Hasil Clustering berdasarkan P2 dan P3	43
Gambar 4.17 Tren Cluster untuk P2	44
Gambar 4.18 Klaster 0 P2	45
Gambar 4.19 Klaster 1 P2	45
Gambar 4.20 Klaster 2 P2	46
Gambar 4.21 Tren Cluster untuk P3	46
Gambar 4.22 Tren Cluster 0 untuk P3	47
Gambar 4.23 Tren Cluster 1 untuk P3	48
Gambar 4.24 Tren Cluster 2 untuk P3	48
Gambar 4.25 Perbandingan metode SMOTE	50

DAFTAR TABEL

Tabel 3.1 Data Pelanggan Tarif 450	16
Tabel 3.2 Data Pelanggan Tarif 900	16
Tabel 3.3 Data Pelanggan Tarif 2200	17
Tabel 3.4 Data Pelanggan Tarif 3500	18
Tabel 4.1 Perbandingan Google Colab dengan Matlab	30
Tabel 4.2 Cluster dan pola tren paling sering untuk P2 dan P3	35
Tabel 4.3 Cluster dan pola tren paling sering untuk P2 dan P3	43



ABSTRAK

PLN UP3 Demak adalah Unit Pelaksana Pelayanan Pelanggan yang mempunyai total pelanggan sebanyak 794.065 pelanggan dan energi listrik yang terjual sepanjang tahun 2023 sebanyak 1.889.320 MWh. Dalam kurun waktu tersebut terdapat susut energi sebanyak 199.147 MWh secara total baik teknis maupun non teknis. Pencarian target pelanggan dengan penyalahgunaan listrik di PLN UP3 Demak tersebut masih bersifat manual dan belum ada kajian khusus terkait hal tersebut. Penelitian ini membandingkan implementasi metode K-Nearest Neighbors (KNN) dalam mendeteksi penyalahgunaan listrik di PLN UP3 Demak dengan dan tanpa penggunaan Synthetic Minority Oversampling Technique (SMOTE). SMOTE digunakan untuk mengatasi ketidakseimbangan data dengan meningkatkan representasi kelas minoritas yang berhubungan dengan kasus penyalahgunaan listrik. Dua model KNN dianalisis, yaitu model dengan SMOTE untuk menangani data tidak seimbang dan model tanpa SMOTE yang mempertahankan data asli.

Hasil penelitian menunjukkan bahwa ROC AUC tanpa SMOTE adalah 0.50, sedangkan dengan SMOTE justru menurun menjadi 0.49. Ini menunjukkan bahwa keseimbangan data tidak selalu meningkatkan performa prediksi. Sebaliknya, model KNN tanpa SMOTE memiliki akurasi keseluruhan yang lebih baik karena tidak dipengaruhi oleh sampel sintesis yang dapat menyebabkan overfitting. Analisis pola tren menunjukkan bahwa perubahan konsumsi listrik, seperti peningkatan pada pelanggan dengan konsumsi rendah atau penurunan pada pelanggan dengan konsumsi tinggi, dapat menjadi indikator penyalahgunaan listrik. Penelitian ini memberikan wawasan yang berguna bagi PLN UP3 Demak dalam meningkatkan pengawasan terhadap potensi penyalahgunaan listrik dari histori pemakaian pelanggan yang sudah terkena P2TL (Penertiban Pemakaian Tenaga Listrik). Kegiatan P2TL ini berhasil menyelamatkan energi sebesar 8.250 MWh pada tahun 2023.

Kata kunci: K-Nearest Neighbors, SMOTE, P2TL, klasifikasi, ketidakseimbangan data.

ABSTRACT

PLN UP3 Demak is a Customer Service Implementation Unit with a total of 794,065 customers and electricity sold throughout 2023 of 1,889,320 MWh. During this period there was a total energy loss of 199,147 MWh, both technical and non-technical. The search for target customers with electricity misuse at PLN UP3 Demak is still manual and there has been no specific study regarding this matter.

This study compares the implementation of the K-Nearest Neighbors (KNN) method in detecting electricity abuse at PLN UP3 Demak, focusing on the use of and without the Synthetic Minority Oversampling Technique (SMOTE). SMOTE is applied to address class imbalance by enhancing the representation of the minority class related to electricity abuse cases. Two KNN models are analyzed: one with SMOTE to handle imbalanced data, and another without SMOTE to retain the original data.

The results of the study showed that the ROC AUC without SMOTE was 0.50, while with SMOTE it actually decreased to 0.49. This shows that data balance does not always improve prediction performance.. In contrast, the KNN model without SMOTE exhibits higher overall accuracy since it is not influenced by synthetic samples, which could lead to overfitting. Trend analysis reveals that changes in electricity consumption, such as increased usage among low-consumption customers or decreased usage among high-consumption customers, could indicate potential electricity abuse. This study provides valuable insights for PLN UP3 Demak to enhance surveillance of potential electricity misuse from the history of the customer's energy usage in violation which is get of Controlling the Use of Electric Power. This activities succeeded in saving 8,250 MWh of energy in 2023.

Keywords: K-Nearest Neighbors, SMOTE, Controlling the Use of Electric Power, classification, imbalanced data.

BAB I

PENDAHULUAN

1.1 Latar Belakang

PLN UP3 Demak merupakan Unit Pelaksana Pelayanan Pelanggan yang melayani ketersediaan tenaga listrik di wilayah Kabupaten Demak dan Kabupaten Grobogan. Dengan total pelanggan sebanyak 794.065 pelanggan dan 1.007 MVA daya tersambung dan sebanyak 1.889.320 MWh Energi listrik yang terjual sepanjang tahun 2023. Dalam kurun waktu tersebut terdapat susut energi sebanyak 199.147 MWh secara total baik teknis maupun non teknis dan dari kegiatan P2TL berhasil menyelamatkan energi sebesar 8.250 MWh.

P2TL (Penertiban Pemakaian Tenaga Listrik) merupakan kegiatan antisipasi atas kehilangan atau susut daya listrik yang diakibatkan oleh faktor yang sifatnya non-teknis seperti pencurian listrik. Kegiatan P2TL dilaksanakan berdasarkan pada Surat Keputusan Direksi PT PLN (Persero) Nomor : 0028.P/DIR/2023 tentang Penertiban Pemakaian Tenaga Listrik (P2TL)[1], yang merupakan turunan dari ketentuan dan perundang-undangan yang mengatur masalah kelistrikan yang berlaku saat ini, antara lain : Undang-undang No.30 tahun 2009 tentang Ketenagalistrikan [2], Peraturan Pemerintah No.23 tahun 2014 tentang perubahan atas peraturan pemerintah No.14 Tahun 2012 tentang Kegiatan Usaha Penyediaan Tenaga Listrik. Tahapan pelaksanaan penertiban pemakaian tenaga listrik meliputi 3 (tiga) tahap, yaitu :

1. Prapemeriksaan, P2TL meliputi kegiatan persiapan sebelum pelaksanaan P2TL.
2. Pemeriksaan P2TL meliputi kegiatan pelaksanaan P2TL.
3. Pascapemeriksaan, P2TL meliputi kegiatan tindak lanjut hasil pemeriksaan P2TL.

Pada tahapan prapemeriksaan P2TL antara lain : penetapan TO (Target Operasi), penyusunan jadwal pelaksanaan P2TL, koordinasi dengan institusi yang berwenang dalam penegakan hukum P2TL dan/atau pihak terkait lainnya dan penyiapan perlengkapan pelaksanaan P2TL.

Seiring berkembangnya ilmu pengetahuan dan teknologi, penetapan TO pada tahap prapemeriksaan P2TL dapat menggunakan pendekatan peramalan pelanggan/konsumen tenaga listrik yang melakukan pencurian tenaga listrik menggunakan algoritma *K-Nearest Neighbor*. Algoritma KNN merupakan metode yang digunakan untuk melakukan klasifikasi data berdasarkan jarak terpendek terhadap objek data[3]. Sehingga data pelanggan/konsumen tenaga listrik yang memiliki klasifikasi data dan jarak terpendek dengan klasifikasi pelanggan/konsumen yang melakukan pencurian listrik.

1.2 Rumusan Masalah

1. Apa metode yang dapat digunakan untuk mengelompokkan pelanggan berdasarkan kemungkinan melakukan pencurian listrik?
2. Bagaimana perbandingan model KNN untuk pendeteksi pola pencurian listrik antara dengan Synthetic Minority Oversampling Technique – selanjutnya disebut SMOTE atau tanpa SMOTE?

1.3 Batasan Masalah

Adapun penelitian ini dilakukan di wilayah PT PLN (Persero) UP3 Demak dengan menggunakan data pemakaian tenaga listrik pelanggan di Kabupaten Demak dan Kabupaten Grobogan.

1.4 Tujuan Penelitian

Tujuan penelitian ini adalah :

1. Mengidentifikasi klasifikasi pelanggan yang melakukan pencurian listrik dengan menggunakan algoritma KNN, sehingga PLN dapat dengan akurat menentukan daftar pelanggan yang terindikasi melakukan pencurian listrik untuk dijadikan Target Operasi P2TL.
2. Membandingkan akurasi, presisi, recall, dan F1-score model KNN dengan dan tanpa SMOTE dalam mendeteksi pola pencurian listrik.

1.5 Manfaat Penelitian

Penelitian ini diharapkan dapat memberikan manfaat untuk PLN dalam Penelitian ini memberikan wawasan tentang cara yang lebih efektif untuk mendeteksi pola pencurian listrik dengan memanfaatkan teknik pengelompokan pelanggan menggunakan algoritma KNN (K-Nearest Neighbors). Ini bermanfaat

bagi PT. PLN untuk lebih cepat mengidentifikasi pelanggan yang berpotensi melakukan pencurian, sehingga dapat mengurangi kerugian finansial dan meningkatkan efisiensi pemantauan.

1.6 Keaslian Penelitian

Sebagian besar penelitian yang ada menggunakan teknik pembelajaran mesin standar tanpa mengatasi masalah ketidakseimbangan kelas secara khusus. Penelitian ini menyoroti penggunaan teknik SMOTE yang dirancang untuk menangani ketidakseimbangan kelas dalam data penyalahgunaan tenaga listrik, yang merupakan pendekatan yang relatif baru dalam konteks ini. Dengan menerapkan SMOTE, penelitian ini berusaha mengatasi masalah ketidakseimbangan yang sering kali menyebabkan model pembelajaran mesin gagal mendeteksi kasus penyalahgunaan yang jarang terjadi[5].

Meskipun SMOTE telah digunakan dalam berbagai konteks, kombinasi teknik ini dengan algoritma K-NN untuk peramalan penyalahgunaan tenaga listrik menawarkan pendekatan baru. Penelitian ini mengeksplorasi bagaimana teknik SMOTE dapat meningkatkan efektivitas algoritma K-NN dalam mengklasifikasikan indikasi penyalahgunaan tenaga listrik, dengan mempertimbangkan karakteristik spesifik dari data listrik dan pola penyalahgunaan yang ada.

Penelitian ini difokuskan pada wilayah PLN UP3 Demak, yang menyediakan konteks lokal dan data spesifik yang belum banyak dieksplorasi dalam literatur. Penelitian ini mengeksplorasi bagaimana teknik SMOTE dan K-NN dapat diterapkan dalam konteks spesifik wilayah PLN UP3 Demak, sehingga menawarkan wawasan baru tentang permasalahan penyalahgunaan tenaga listrik di area tersebut.

Dengan menggabungkan teknik Synthetic Minority Oversampling Technique (SMOTE) dan K-Nearest Neighbors (K-NN) untuk analisis data penyalahgunaan tenaga listrik, penelitian ini memberikan kontribusi signifikan dalam metodologi deteksi penyalahgunaan. Teknik SMOTE meningkatkan performa model dengan mengatasi masalah ketidakseimbangan kelas, yang sering menghambat akurasi dalam mendeteksi kasus penyalahgunaan yang jarang terjadi. Selain itu, penelitian ini menawarkan aplikasi kontekstual yang unik dengan menerapkan teknik ini

dalam wilayah spesifik PLN UP3 Demak. Pendekatan ini tidak hanya memberikan solusi yang lebih efektif untuk mendeteksi penyalahgunaan tenaga listrik di area tersebut, tetapi juga berpotensi menjadi acuan bagi studi serupa di wilayah lain, memperluas pemahaman dan penerapan teknik pembelajaran mesin dalam konteks lokal.

Hasil dari penelitian ini diharapkan memberikan rekomendasi praktis untuk pihak PLN dalam mendeteksi dan menangani kasus penyalahgunaan tenaga listrik secara lebih efektif. Dengan metode yang diusulkan, diharapkan dapat meningkatkan akurasi dalam mendeteksi penyalahgunaan dan membantu dalam pengambilan keputusan yang lebih baik.

1.7 Sistematika Penelitian

Bab satu ini merupakan pendahuluan yang membahas mengenai latar belakang, perumusan masalah, tujuan penulisan, dan sistematika penulisan. Sementara Bab dua merupakan studi literatur dan dasar teori yang berisikan Penertiban Penertiban Pemakaian Tenaga Listrik (P2TL), jenis-jenis Pelanggaran P2TL dan metode *K-Nearest Neighbor*.

Bab tiga berisi tentang metode yang digunakan dalam pengambilan data Cluster pemakai energi yang melakukan pencurian listrik atau pelanggaran, Model Penelitian, Peralatan dan Material yang digunakan, Prosedur Penelitian, Metode Penelitian, *Software/Hardware* yang digunakan and Alur Tahapan Penelitian.

BAB II

TINJAUAN PUSTAKA DAN LANDASAN TEORI

2.1 Tinjauan Pustaka

Penelitian ini memiliki keaslian yang signifikan karena mengadopsi pendekatan yang berbeda dalam mengidentifikasi pelanggan yang melakukan pencurian listrik dengan menggunakan algoritma K-Nearest Neighbors (KNN). Meskipun terdapat beberapa penelitian sebelumnya yang menggunakan berbagai algoritma dan teknik dalam mendeteksi penyalahgunaan tenaga listrik, seperti Fuzzy Subtractive Clustering, Naïve Bayes, Random Forest, Support Vector Machine, dan teknik machine learning lainnya, tidak ada penelitian yang secara khusus mengkaji penggunaan KNN dalam konteks yang sama.

Sebagai contoh, penelitian oleh Muhamad Nagi menggunakan Fuzzy Subtractive Clustering untuk identifikasi target P2TL, sementara penelitian oleh Syahtriatna et al. menerapkan Naïve Bayes untuk analisis pencurian tenaga listrik. Penelitian lain, seperti yang dilakukan oleh Andika Saputra dan Fajar Nugroho, juga menggunakan algoritma yang berbeda untuk tujuan yang serupa. Namun, tidak ada riset yang secara khusus meneliti efektivitas KNN dalam klasifikasi pelanggan yang melakukan pencurian listrik di wilayah PLN. Berikut untuk Penjelasan detailnya :

1. Penelitian oleh Muhamad Nagi membahas penggunaan metode Fuzzy Subtractive Clustering untuk mengidentifikasi target operasi Penertiban Pemakaian Tenaga Listrik (P2TL) di PLN UP3 Semarang. Teknik ini meningkatkan efisiensi dalam mengidentifikasi pelanggan yang berpotensi menyalahgunakan tenaga listrik, sehingga operasi P2TL menjadi lebih tepat sasaran [6].
2. Penelitian oleh Syahtriatna, Rizki Novendra, dan Seragita Sabrina Tarigan membahas penerapan algoritma Naïve Bayes untuk mendeteksi dan menganalisis pencurian tenaga listrik di PT. PLN (Persero) Panam. Algoritma ini mengklasifikasikan pelanggan berdasarkan potensi pencurian dengan mempertimbangkan pola konsumsi energi dan variabel lainnya. Hasilnya menunjukkan bahwa Naïve Bayes efektif dalam membantu PLN mengidentifikasi pelanggan yang mencurigakan [7].

3. Penelitian oleh Andika Saputra membahas penggunaan algoritma Random Forest untuk memprediksi pelanggaran penggunaan tenaga listrik di PLN UP3 Cirebon. Metode ini mengklasifikasikan data pelanggan dan menunjukkan akurasi tinggi dalam mendeteksi potensi penyalahgunaan tenaga listrik [8].
4. Penelitian oleh Fajar Nugroho membahas penggunaan algoritma Support Vector Machine (SVM) untuk mendeteksi penyalahgunaan tenaga listrik di PLN UP3 Surabaya. SVM memisahkan data pelanggan yang melanggar dari yang tidak berdasarkan pola konsumsi energi, dan menunjukkan akurasi tinggi dengan tingkat false positive yang rendah [9].
5. Penelitian oleh Rina Wijayanti membahas deteksi anomali penggunaan energi listrik di PLN UP3 Bandung menggunakan pendekatan machine learning. Berbagai algoritma, seperti k-Means, Decision Tree, dan Neural Network, diterapkan untuk mengidentifikasi pola konsumsi yang tidak biasa. Hasilnya menunjukkan bahwa machine learning efektif dalam mendeteksi anomali yang mengindikasikan potensi penyalahgunaan tenaga listrik [10].
6. Penelitian oleh Dwi Santoso membahas penerapan algoritma K-Nearest Neighbors (K-NN) untuk klasifikasi pelanggaran penggunaan tenaga listrik di PLN UP3 Medan. K-NN mengelompokkan data pelanggan berdasarkan tingkat pelanggaran, dan hasil penelitian menunjukkan bahwa K-NN memiliki performa baik dalam mengklasifikasikan pelanggan, terutama dengan data yang tidak merata [11].
7. Penelitian oleh Budi Hartono mengevaluasi penggunaan teknik Synthetic Minority Oversampling Technique (SMOTE) untuk menyeimbangkan kelas dalam dataset prediksi penyalahgunaan tenaga listrik di PLN UP3 Yogyakarta. Penggunaan SMOTE berhasil meningkatkan akurasi model dengan memperbaiki distribusi kelas yang tidak seimbang, sehingga model lebih robust dalam mendeteksi pelanggan yang melanggar [12].

2.2 Penertiban Pemakaian Tenaga Listrik (P2TL)

Penertiban Pemakaian Tenaga Listrik – selanjutnya disingkat P2TL adalah rangkaian kegiatan yang meliputi perencanaan, pemeriksaan, tindakan dan penyelesaian yang dilakukan oleh PLN terhadap instalasi PLN dan/atau instalasi pemakai tenaga listrik dari PLN. Kebijakan penerpan P2TL dilaksanakan

berdasarkan pada Surat Peraturan Direksi PT PLN (Persero) Nomor : 0028.P/DIR/2023 tentang Penertiban Pemakaian Tenaga Listrik (P2TL)[1], yang merupakan turunan dari ketentuan dan perundang-undangan yang mengatur masalah kelistrikan yang berlaku saat ini, antara lain : Undang-undang No.30 tahun 2009 tentang Ketenagalistrikan [2], Peraturan Pemerintah No.23 tahun 2014 tentang perubahan atas peraturan pemerintah No.14 Tahun 2012 tentang Kegiatan Usaha Penyediaan Tenaga Listrik juga Undang-undang No.8 tahun 1999 tentang Perlindungan Konsumen. Serta Peraturan Menteri Energi dan Sumber Daya Mineral Nomor 45 tahun 2005 tentang Instalasi Ketenagalistrikan sebagaimana telah diubah dengan Peraturan Menteri Energi dan Sumber Daya Mineral Nomor 46 tahun 2006. Tahapan pelaksanaan penertiban pemakaian tenaga listrik meliputi 3 (tiga) tahap, yaitu :

1. Prapemeriksaan, P2TL meliputi kegiatan persiapan sebelum pelaksanaan P2TL.
2. Pemeriksaan P2TL meliputi kegiatan pelaksanaan P2TL.
3. Pascapemeriksaan, P2TL meliputi kegiatan tindak lanjut hasil pemeriksaan P2TL.

Peraturan Direksi PT PLN (PERSERO) Nomor : 0028.P/DIR/2023 tentang penerbitan pemakaian tenaga listrik bahwa untuk melaksanakan ketentuan Pasal 17 Peraturan Menteri Energi dan Sumber Daya Mineral Nomor 27 Tahun 2017 tentang Tingkat Mutu Pelayanan dan Biaya yang Terkait dengan Penyaluran Tenaga Listrik oleh PT Perusahaan Listrik Negara (Persero) sebagaimana telah diubah dengan Peraturan Menteri Energi dan Sumber Daya Mineral Nomor 18 Tahun 2019, perlu menetapkan pedoman terkait penerbitan pemakaian tenaga listrik[13].

2.2.1 Prapemeriksaan Penertiban Pemakaian Tenaga Listrik

1. Prapemeriksaan P2TL meliputi kegiatan antara lain :
 - a. Penetapan TO
 - b. Penyusunan jadwal pelaksanaan P2TL
 - c. koordinasi dengan institusi yang berwenang dalam penegakan hukum P2TL dan/atau pihak terkait lainnya
 - d. Penyiapan perlengkapan pelaksanaan P2TL

2. Penetapan TO sebagaimana dimaksud pada ayat (1) huruf a, dilakukan sebagai berikut:
 - a. TO ditetapkan oleh Pemberi Tugas atau Penanggung Jawab P2TL dalam rangka mencapai SO bulanan, triwulanan, semesteran, atau tahunan.
 - b. TO merupakan titik lokasi SO yang memuat antara lain data Pemakai Tenaga Listrik dan/atau lokasi sekitar dilakukannya pemeriksaan P2TL.
 - c. TO ditetapkan berdasarkan:
 - 1) Pemantauan dari Daftar Langgan yang Perlu Diperhatikan (DLPD), Daftar Pembacaan Meter (DPM) dan Daftar Pemakaian kwh (DPK);
 - 2) pemantauan terhadap pemakaian tenaga listrik Konsumen yang tidak wajar sesuai dengan karakteristik pemakaian listrik;
 - 3) kumpulan data dan informasi P2TL;
 - 4) data dan informasi lainnya yang diperoleh PLN dari kegiatan pelayanan konsumen;
 - 5) SO bulanan, triwulan, semesteran, atau tahunan;
 - 6) evaluasi data load profile terhadap kontinuitas penggunaan tenaga listrik pada Konsumen yang dibaca melalui AMR; dan/atau
 - 7) evaluasi *wiring* melalui diagram *phasor* pada Konsumen yang dibaca melalui AMR.
 - d. Dalam rangka menjaga kerahasiaan TO, penyerahan TO oleh Pemberi Tugas atau Penanggung Jawab P2TL kepada Pelaksana Lapangan P2TL dilakukan sebelum Pelaksana Lapangan P2TL berangkat ke lokasi pelaksanaan P2TL.

2.2.2 Golongan Pelanggaran Pemakaian Tenaga Listrik

Pelanggaran pemakaian tenaga listrik digolongkan menjadi [1] :

1. Pelanggaran Golongan I (P-I) merupakan pelanggaran yang mempengaruhi batas daya tetapi tidak memengaruhi pengukuran energi;
2. Pelanggaran Golongan II (P-II) merupakan pelanggaran yang mempengaruhi pengukuran energi tetapi tidak memengaruhi batas daya;
3. Pelanggaran Golongan III (P-III) merupakan pelanggaran yang mempengaruhi batas daya dan mempengaruhi pengukuran energi; dan

4. Pelanggaran Golongan IV (P-IV) merupakan pelanggaran yang dilakukan oleh Bukan Konsumen/Pelanggan.

Termasuk Pelanggaran Golongan P I yaitu dalam hal pada APP yang terpasang di pelanggan ditemukan 1 (satu) atau lebih fakta yang dapat mempengaruhi batas daya tetapi tidak mempengaruhi pengukuran energi, sebagai berikut :

- a. Segel milik PLN pada Alat Pembatas hilang atau rusak bukan karena korosi atau faktor alam lainnya atau tidak sesuai dengan aslinya, yang dibuktikan dengan hasil pemeriksaan lebih lanjut;
- b. Kemampuan Alat Pembatas menjadi lebih besar, antara lain dengan:
 - 1) Mengubah batas daya *Mini Circuit Breaker* (MCB);
 - 2) Mengubah *setting relay* Alat Pembatas; dan/atau
 - 3) Membalik *phasa* dengan netral;
- c. Alat Pembatas hilang, rusak atau tidak sesuai dengan aslinya;
- d. Khusus untuk Pelanggan yang menggunakan meter kVA maksimum:
 - 1) Segel pada meter kVA maksimum dan/atau perlengkapannya hilang atau rusak, bukan karena korosi atau faktor alam lainnya atau tidak sesuai dengan aslinya, yang dibuktikan dengan hasil pemeriksaan lebih lanjut; dan/atau
 - 2) Meter kVA maks dan/atau perlengkapannya hilang atau rusak bukan karena korosi atau faktor alam lainnya atau tidak sesuai dengan aslinya;
- e. Alat Pembatas terhubung langsung dengan kawat/kabel sehingga Alat Pembatas tidak berfungsi atau kemampuannya menjadi lebih besar; dan/atau
- f. Terjadi hal-hal lainnya dengan tujuan mempengaruhi batas daya tetapi tidak memengaruhi pengukuran energi.

Termasuk Golongan P II yaitu dalam hal pada APP yang terpasang di konsumen ditemukan satu atau lebih fakta yang dapat mempengaruhi pengukuran energi tetapi tidak memengaruhi batas daya sebagai berikut :

- a. Segel tera dan/atau segel milik PLN pada Alat Pengukur dan/atau perlengkapannya salah satu atau semuanya hilang/tidak lengkap, rusak atau putus bukan karena korosi atau faktor alam lainnya atau tidak sesuai dengan aslinya, yang dibuktikan dengan hasil pemeriksaan lebih lanjut;

- b. Alat Pengukur dan/atau perlengkapannya tidak berfungsi sebagaimana mestinya meski seluruh Segel milik PLN dan Segel Tera dalam keadaan lengkap dan baik.
- c. Alat Pengukur dan/atau perlengkapannya hilang atau tidak sesuai dengan aslinya;
- d. Ditemukan benda asing antara lain receiver remote dan/atau bekas jejaknya dalam kWhmeter sehingga kWhmeter tidak sesuai dengan aslinya, yang dibuktikan dengan hasil pemeriksaan lebih lanjut; dan/atau
- e. Terjadi hal lainnya yang memengaruhi pengukuran energi tetapi tidak memengaruhi batas daya.

Termasuk Golongan P III yaitu dalam hal pada APP dan instalasi tenaga listrik yang terpasang di konsumen ditemukan satu atau lebih fakta memengaruhi batas daya dan pengukuran energi sebagai berikut :

- a. Pelanggaran pemakaian tenaga listrik yang merupakan gabungan Pelanggaran Golongan I (P I) dan Pelanggaran Golongan II (P II);
- b. Menyambung langsung dari Instalasi PLN untuk disalurkan ke instalasi Konsumen tanpa melalui APP;
- c. Menyambung kembali tanpa izin PLN tenaga listrik yang telah dilakukan Pemutusan Sementara dengan pengambilan APP oleh PLN; dan/atau
- d. Terjadi hal lainnya yang memengaruhi batas daya dan pengukuran energi.

Termasuk Golongan P IV yaitu dalam hal ditemukan satu atau lebih fakta pemakaian tenaga listrik PLN tanpa PJBTL oleh Bukan Konsumen antara lain :

- a. Menyambung langsung dari Instalasi PLN;
- b. Pemakai tenaga listrik tidak terdaftar di dalam Data Induk Langgan (DIL) PLN atau Pemakai Tenaga Listrik yang identitas dan kode kedudukannya berbeda dan/atau;
- c. Pemakai tenaga listrik hasil *levering* dari pelanggaran P III

2.3 Ukuran Gejala Pusat

Ukuran gejala pusat adalah nilai-nilai yang menunjukkan pusat dari sekelompok data. Nilai ini terletak tepat atau di sekitar pertengahan data. Nilai-nilai statistika dapat berupa : mean, median, modus, rata-rata harmonis dan rata-rata-rata ukur dan lain-lain [20].

2.3.1 Mean

Mean kepanjangan dari *arithmetic mean* adalah rata-rata hitung atau rata-rata biasa. Nilai ini diperoleh dengan mencari rata-rata biasa. Apabila nilai setiap data sebesar X_i ($i=1,2,\dots, n$), maka :

$$Mean = \frac{\sum X_i}{n} \dots\dots\dots(2.1)$$

Apabila data yang dimiliki sudah disusun dalam distribusi frekuensi, maka besar mean dicari dengan rumus sebagai berikut :

$$Mean = \frac{\sum X_i f_i}{\sum f_i} \dots\dots\dots(2.2)$$

2.3.2 Median

Median adalah nilai yang membagi data menjadi dua bagian yang sama, apabila data diurutkan dari yang terkecil sampai dengan yang terbesar.

Besar median dicari dengan rumus sebagai berikut :

$$Median = L + C \left[\frac{f_1}{f_m} \right] \dots\dots\dots(2.3)$$

L = class boundary bawah dari kelas yang mengandung median, sama dengan pertengahan batas kelas atas kelas sebelum median dengan batas bawah kelas median

C = Class interval, atau panjang kelas

f_1 = perbedaan antara frekuensi kumulatif kelas sebelum median dengan letak median

f_m = frekuensi kelas median

2.3.3 Modus

Modus adalah bilangan atau sifat yang paling sering terjadi atau yang frekuensinya terbanyak. Nilai modus dapat dihitung dengan rumus sebagai berikut :

$$Modus = L + C \left[\frac{\Delta_1}{\Delta_1 + \Delta_2} \right] \dots\dots\dots(2.4)$$

Dimana :

Δ_1 = selisih frekuensi kelas modus dengan kelas sebelum modus

Δ_2 = selisih frekuensi kelas modus dengan kelas sesudah modus

2.4 Normalisasi Data

Terdapat 3 metode yang paling umum digunakan untuk normalisasi data pada *Data Analys* dan *Machine Learning*[14]:

1. *Min-Max Normalization*

Min-Max normalization merupakan metode normalisasi dengan melakukan transformasi linier terhadap data asli sehingga menghasilkan keseimbangan nilai perbandingan antar data saat sebelum dan sesudah proses. Metode ini dapat menggunakan rumus sebagai berikut :

$$x_{Normalized} = \frac{x - \min(x)}{\max(x) - \min(x)} \dots \dots \dots (2.5)$$

Dimana :

$X_{Normalized}$: Nilai x baru hasil normalisasi

X_{min} : Nilai Minimum dari x

X_{max} : Nilai Maksimum dari x

2. *Z-Score Normalization (StandardScaler)*

Z-score normalization merupakan metode normalisasi berdasarkan mean (nilai rata-rata) dan standard deviation (deviasi standar) dari data. Metode ini sangat berguna jika tidak diketahui nilai aktual minimum dan maksimum dari data. Rumus yang digunakan sebagai berikut :

$$x_{Normalized} = \frac{x - \text{mean}(x)}{\sigma(x)} \dots \dots \dots (2.6)$$

Dimana :

$X_{Normalized}$: Nilai x baru hasil normalisasi

$\text{mean}(x)$: Nilai rata-rata dari x

σx : Nilai standar deviasi dari x

3. *Decimal Scaling Normalization*

Decimal scaling merupakan metode normalisasi dengan menggerakkan nilai desimal dari data ke arah yang diinginkan. Setiap nilai data dibagi dengan nilai absolut maksimum data, sehingga menghasilkan nilai yang berkisar antara -1 hingga 1. Formula yang digunakan sebagai berikut :

$$x_{Normalized} = \frac{x}{10^j} \dots \dots \dots (2.7)$$

Dimana :

$X_{\text{Normalized}}$: Nilai x baru hasil normalisasi

10^j : Faktor skala

2.5 Konsep Algoritma SMOTE dan *K-Nearest Neighbor* untuk Klasifikasi

2.3.4 *K-Nearest Neighbor* (K-NN)

Metode *Nearest Neighbor* adalah pendekatan untuk mencari kasus dengan menghitung kedekatan antara kasus baru (*testing data*) dengan kasus lama (*training data*), yaitu berdasarkan pada pencocokan bobot dari sejumlah fitur yang ada[15].

Metode *K-Nearest Neighbor* (K-NN) adalah suatu metode yang menggunakan algoritma supervised dimana hasil dari *query instance* yang baru diklasifikasikan berdasarkan mayoritas dari label *class* pada K-NN. Tujuan dari algoritma K-NN adalah mengklasifikasikan objek baru berdasarkan atribut dan training data. Algoritma K-NN bekerja berdasarkan jarak terdekat dari *query instance* ke training data untuk menentukan K-NN-nya. Salah satu cara untuk menghitung jarak dekat atau jauhnya tetangga menggunakan metode *euclidian distance*. Algoritma klasifikasi berbasis instance yang digunakan untuk menentukan label dari titik data baru berdasarkan label mayoritas dari k tetangga terdekatnya dalam dataset pelatihan. Dalam konteks ini, K-NN akan digunakan setelah proses oversampling dengan SMOTE untuk mengklasifikasikan apakah suatu kasus termasuk dalam kategori penyalahgunaan tenaga listrik atau bukan.

Langkah-langkah K-NN:

1. Tentukan Parameter K dengan *Cross-Validation*: *Cross-validation* adalah teknik evaluasi model yang digunakan untuk mengukur kemampuan generalisasi model pembelajaran mesin. Folds adalah cara membagi dataset menjadi beberapa bagian kecil. Dalam k-fold cross-validation, dataset dibagi menjadi k bagian (folds) yang sama besar. Setiap fold digunakan sebagai set validasi satu kali, sementara k-1 fold lainnya digunakan untuk melatih model. Proses ini diulang sebanyak k kali. Dalam teknik ini, dataset dibagi menjadi beberapa subset (folds), dan model dilatih dan divalidasi secara berulang kali pada kombinasi subset yang berbeda. Teknik yang umum digunakan adalah k-fold cross-validation, dimana dataset dibagi menjadi k bagian, dan model dilatih pada k-1 bagian dan divalidasi pada bagian yang tersisa. Proses ini diulangi k kali sehingga setiap bagian dataset digunakan sebagai set validasi

satu kali. Keuntungan utama cross-validation adalah kemampuannya untuk mengurangi bias dalam evaluasi model dan memberikan estimasi performa model yang lebih akurat.

2. Pengukuran Jarak menggunakan *Elbow Plot*: Elbow plot adalah metode visual untuk menentukan jumlah optimal tetangga (k) dalam algoritma KNN. Metode ini bekerja dengan mengukur akurasi model untuk berbagai nilai k dan memplot hasilnya. Nilai k yang optimal ditandai dengan adanya "sudut" atau "elbow" pada plot, yang menunjukkan titik di mana penambahan tetangga tidak lagi memberikan peningkatan yang signifikan dalam akurasi model. Dengan memilih nilai k yang berada di titik elbow, kita dapat meminimalkan overfitting dan underfitting pada model.

$$d(X_1, X_2) = \sqrt{\sum_{i=1}^n (x_{1i} - x_{2i})^2} \dots\dots\dots(2.8)$$

Dimana:

$d(X_1, X_2)$: jarak antara dua titik data.

x_{1i} dan x_{2i} : nilai atribut pada dimensi ke-I dari titik data X_1 dan X_2 .

3. Klasifikasi Berdasarkan Mayoritas: Mengklasifikasikan titik data baru dengan menentukan mayoritas label dari k tetangga terdekat.

Dalam aplikasi peramalan penyalahgunaan tenaga listrik, K-NN akan digunakan untuk memprediksi kemungkinan penyalahgunaan berdasarkan data yang telah dioversample. Dengan menggunakan dataset yang seimbang setelah SMOTE, K-NN akan lebih efektif dalam mengklasifikasikan apakah penggunaan tenaga listrik oleh pelanggan menunjukkan indikasi penyalahgunaan[16].

2.5.2 Synthetic Minority Oversampling Technique (SMOTE)

Synthetic Minority Oversampling Technique (SMOTE) adalah metode yang dirancang untuk menangani masalah ketidakseimbangan kelas dalam dataset, yang sering kali terjadi pada data dengan banyak kasus dari satu kelas dan sangat sedikit dari kelas lainnya. Pada kasus peramalan indikasi penyalahgunaan tenaga listrik, kelas minoritas biasanya adalah kasus penyalahgunaan yang jarang terjadi dibandingkan dengan kasus penggunaan yang sah. SMOTE membantu dalam menciptakan sampel sintetis dari kelas minoritas untuk memperbaiki

ketidakseimbangan ini, sehingga model dapat belajar dengan lebih baik dari kelas minoritas[5].

Langkah-langkah SMOTE:

1. Identifikasi Titik Data Minoritas: Menentukan titik data dari kelas minoritas yang akan ditingkatkan.
2. Pilih K-Tetangga Terdekat: Untuk setiap titik data minoritas, cari k tetangga terdekat di kelas minoritas yang sama. Jumlah tetangga k ini adalah parameter yang harus ditentukan.
3. Interpolasi untuk Membuat Sampel Sintetis: Sampel baru dibuat dengan interpolasi antara titik data minoritas yang ada dan tetangganya. Formula interpolasi adalah:

$$X_{new} = X_i + \lambda (X_{neighbor} - X_i) \dots\dots\dots(2.9)$$

Di mana:

- X_{new} : sampel sintetis yang baru.
- X_i : titik data minoritas yang ada.
- $X_{neighbor}$: titik data tetangga dalam kelas minoritas.
- λ : faktor acak antara 0 dan 1.

Dalam konteks peramalan penyalahgunaan tenaga listrik, SMOTE akan membantu menghasilkan data tambahan yang merepresentasikan pola-pola penyalahgunaan yang mungkin tidak terlihat dalam dataset asli. Ini memungkinkan model pembelajaran untuk lebih efektif mendeteksi penyalahgunaan tenaga listrik yang jarang tetapi kritikal.

2.5.3 Evaluasi Kinerja Model

Untuk menilai efektivitas model yang menggunakan SMOTE dan K-NN dalam mendeteksi penyalahgunaan tenaga listrik, beberapa metrik evaluasi digunakan:

1. Akurasi: Mengukur proporsi prediksi yang benar dari total prediksi.

$$Akurasi = \frac{Jumlah\ Prediksi\ Benar}{Total\ Prediksi} \dots\dots\dots(2.10)$$

2. Presisi: Mengukur proporsi prediksi positif yang benar dari total prediksi positif.

$$Presisi = \frac{Jumlah\ Prediksi\ Positif\ Benar}{Total\ Prediksi\ Positif} \dots\dots\dots(2.11)$$

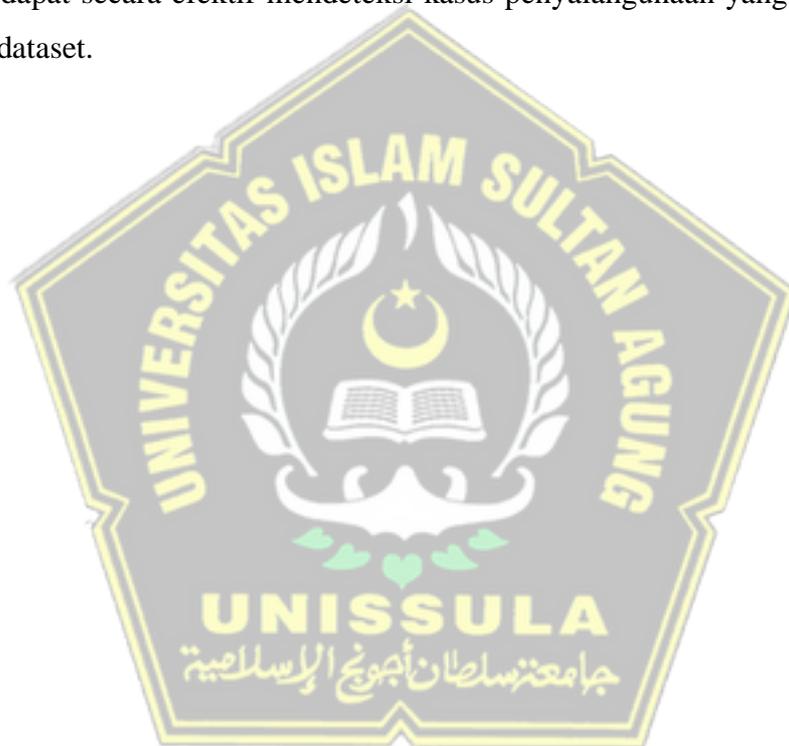
3. Recall: Mengukur proporsi prediksi positif yang benar dari total kejadian positif yang sebenarnya.

$$Recall = \frac{Jumlah\ Prediksi\ Positif\ Benar}{Total\ Kejadian\ Positif} \dots\dots\dots(2.12)$$

4. F1-Score: Rata-rata harmonis dari presisi dan recall, memberikan gambaran menyeluruh tentang kinerja model.

$$F1 - Score = 2 \times \frac{Presisi \times Recall}{Presisi + Recall} \dots\dots\dots(2.13)$$

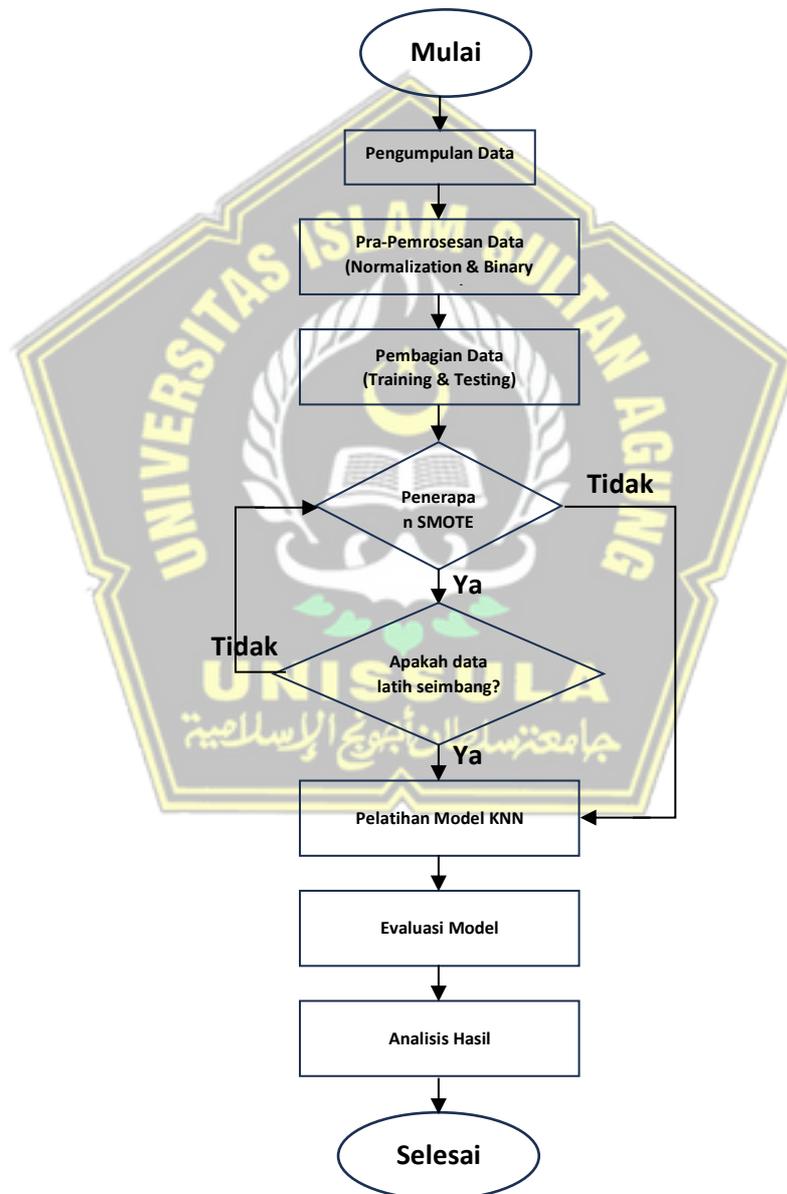
Relevansi dalam mengevaluasi seberapa baik model dalam mengidentifikasi penyalahgunaan tenaga listrik penting untuk memastikan bahwa model dapat secara efektif mendeteksi kasus penyalahgunaan yang jarang terjadi dalam dataset.



BAB III METODOLOGI

3.1 Diagram Alur Penelitian

Metodologi yang digunakan dalam penyusunan penelitian ini adalah metodologi studi pustaka, *K-Nearest Neighbor* (K-NN) untuk menemukan klasifikasi pemakai tenaga listrik yang melakukan pencurian listrik, yang mana data diambil dari PT. PLN (Persero) UP3 Demak.



Gambar 3.1 Metode Pengerjaan

3.2 Pengumpulan Data

Dalam penelitian ini dilakukan pengumpulan data di PT. PLN (Persero) UP3 Demak. Data yang diperlukan dibagi menjadi dua jenis, yaitu data training dan data testing. Data training mencakup Daftar Pemakaian kWh (DPK) pelanggan yang ditemukan melakukan pencurian listrik dari bulan Februari 2023 hingga April 2024. Sementara itu, data testing terdiri dari DPK pelanggan yang kedapatan melakukan pencurian listrik dari bulan Mei 2023 hingga Januari 2024. Pembagian data ini bertujuan untuk membangun dan menguji model yang dapat mendeteksi pola-pola pelanggaran penggunaan listrik [13].

Tabel 3.1 Data Pelanggan Daya 450

NO	IDPEL	NAMA	DAYA	BULAN 1	...	BULAN 15
1	523080056202	ABDUL MUNIF	R1/450	182	...	76
2	523070927861	SUPARMIN LASTRI	R1/450	129	...	117
3	523060418033	TRİYONO	R1/450	174	...	24
4	523060677412	SAKIMAN	R1/450	214	...	67
...
69	523061227711	TK PANGUDI LUHUR	S2/450	10	...	10

Tabel 3.1 berisi data pelanggan dengan tarif 450 VA dan memperlihatkan variasi konsumsi energi masing-masing. Data ini penting untuk memahami pola penggunaan listrik dalam kategori daya tertentu dan membantu dalam mengidentifikasi kemungkinan penyalahgunaan atau pencurian listrik. Dengan mempelajari variasi konsumsi energi, kita dapat lebih mudah mendeteksi pola yang tidak biasa atau mencurigakan.

Tabel 3.2 Data Pelanggan Daya 900

NO	IDPEL	NAMA	DAYA	BULAN -1	...	BULAN -15
1	523070752661	PARLI	R1/900	182	...	76
2	523060849281	MUHAMMAD ZUHRI	R1M/900	182	...	51
3	523060151452	AHMAT ALI	R1M/900	325	...	202
...
44	523080565378	RUSMAN	R1/900	150	...	111

Tabel 3.2 digunakan untuk menganalisis pola konsumsi listrik dan potensi penyalahgunaan pada kategori daya 900 VA. Perbedaan pola konsumsi energi pada daya ini, dibandingkan dengan daya 450 VA, dapat memberikan indikasi kebutuhan untuk analisis lebih mendalam.

Tabel 3.3 Data Pelanggan Daya 2200

NO	IDPEL	NAMA	DAYA	BULAN -1	...	BULAN -15
1	523060321684	YASIN AL FATAH	R1/2200	772	...	125

Tabel 3.3 ini memperlihatkan konsumsi energi pada daya 2200 VA dan penting untuk menganalisis indikasi penyalahgunaan.

Tabel 3.4 Data Pelanggan Daya 3500

NO	IDPEL	NAMA	DAYA	BULAN -1	...	BULAN -15
1	523060931785	SUKARJO S K M/PRAK	R2/3500	772	...	125

Tabel 3.4 mencakup pelanggan dengan daya 3500 VA, menunjukkan variasi konsumsi energi.

Analisis terhadap tabel ini bertujuan untuk mengidentifikasi kemungkinan adanya penyalahgunaan listrik pada kategori daya yang lebih besar serta potensi pelanggaran oleh konsumen dengan daya besar.

3.3 Teknik Pengolahan Data

Proses pengolahan data dimulai dengan melakukan normalisasi pada data pemakaian listrik menggunakan teknik Min-Max Scaler. Metode ini mengonversi nilai kWh ke dalam rentang 0 hingga 1, sehingga semua fitur berada pada skala yang seragam, menghindari bias pada perhitungan jarak. Setelah proses normalisasi, nilai kWh yang telah dinormalisasi kemudian diubah menjadi nilai biner. Nilai yang kurang dari 0.5 akan dikonversi menjadi 0, sedangkan nilai yang sama dengan atau lebih besar dari 0.5 akan dikonversi menjadi 1. Transformasi ini dilakukan untuk menyiapkan data dalam format yang sesuai untuk proses klasifikasi biner menggunakan algoritma K-Nearest Neighbors (KNN)[17].

3.3.1 Penentuan Nilai K=8

Untuk menghitung nilai $k = 8$ dalam algoritma K-Nearest Neighbors (KNN) dengan dataset yang diberikan, kita akan mengikuti langkah-langkah perhitungan manual menggunakan 8-fold cross-validation.

```
# Define the features and target variable
X = df[['KWH', 'KWH_Normalized', 'KWH_Binary']].values # or
include other relevant features
y = df['KWH_Binary'].values # replace with your target
variable if needed
# Define the range of k values to test
k_values = range(1, 11)
cross_val_scores = []
# Perform cross-validation for each k
for k in k_values:
    knn = KNeighborsClassifier(n_neighbors=k)
    scores = cross_val_score(knn, X, y, cv=5,
scoring='accuracy')
    cross_val_scores.append(np.mean(scores))
# Display the results
for k, score in zip(k_values, cross_val_scores):
    print(f'k={k}: Mean cross-validation accuracy =
{score:.4f}')
# Find the best k value
best_k = k_values[np.argmax(cross_val_scores)]
print(f'Best k: {best_k} with accuracy
{max(cross_val_scores):.4f}')
```

Program ini melakukan pemilihan parameter terbaik untuk algoritma K-Nearest Neighbors (KNN) dengan menggunakan teknik cross-validation. Data fitur `X` yang terdiri dari kolom `Bulan -1`, `Bulan -2`, `Bulan -3`, sampai `Bulan -15`, serta target variabel `y` diambil dari kolom `GOL_encoded`. Program kemudian menguji beberapa nilai k (jumlah tetangga terdekat) dari 1 hingga 20. Untuk setiap nilai k, model KNN dibangun dan dievaluasi menggunakan fold cross-validation, dan nilai akurasi rata-rata dihitung. Hasilnya ditampilkan, dan nilai k terbaik dipilih berdasarkan akurasi tertinggi.

3.4 Perancangan Sistem dan Implementasi KNN

Dalam perancangan dan implementasi sistem K-Nearest Neighbors (KNN), langkah pertama yang dilakukan adalah menentukan variabel fitur dan target yang akan digunakan[18]. Adapun langkah-langkah Proses KNN sebagai berikut :

1. Mengimpor dan Memproses Data:

Data diimpor ke dalam Python dan dilakukan encoding pada variabel kategorikal menggunakan *LabelEncoder*. Fitur-fitur (X) dan target (y) didefinisikan. Nilai yang hilang diimputasi menggunakan rata-rata (*SimpleImputer*), dan fitur dinormalisasi menggunakan *StandardScaler*.

2. Pembagian Data (70% Pelatihan, 30% Pengujian):

Fungsi *train_test_split* digunakan dengan parameter *test_size=0.3* untuk membagi data. 70% data digunakan untuk melatih model, dan 30% untuk menguji performa model.

3. Cross-Validation untuk Hyperparameter Tuning:

Model KNN diuji dengan berbagai nilai k (1-20) menggunakan validasi silang (5-fold cross-validation). Akurasi rata-rata dihitung untuk setiap nilai k, dan nilai k terbaik dipilih berdasarkan akurasi tertinggi.

4. Pelatihan Model dengan k Terbaik:

Model dilatih menggunakan nilai k yang dipilih dari langkah sebelumnya. Jarak Euclidean digunakan sebagai metrik jarak.

5. Evaluasi Model:

Prediksi dilakukan pada data pengujian (*X_test*). Laporan klasifikasi (*classification_report*) dan matriks kebingungan (*confusion_matrix*) digunakan untuk mengevaluasi performa model.

Akurasi dihitung sebagai:

$$Akurasi = \frac{\text{Jumlah Prediksi Benar}}{\text{Total Prediksi}} \dots\dots\dots(3.1)$$

Laporan klasifikasi memberikan metrik seperti presisi, recall, dan F1-score.

6. Visualisasi:

Grafik akurasi cross-validation terhadap nilai k dibuat untuk memvisualisasikan proses tuning. Matriks kebingungan divisualisasikan dalam bentuk heatmap.

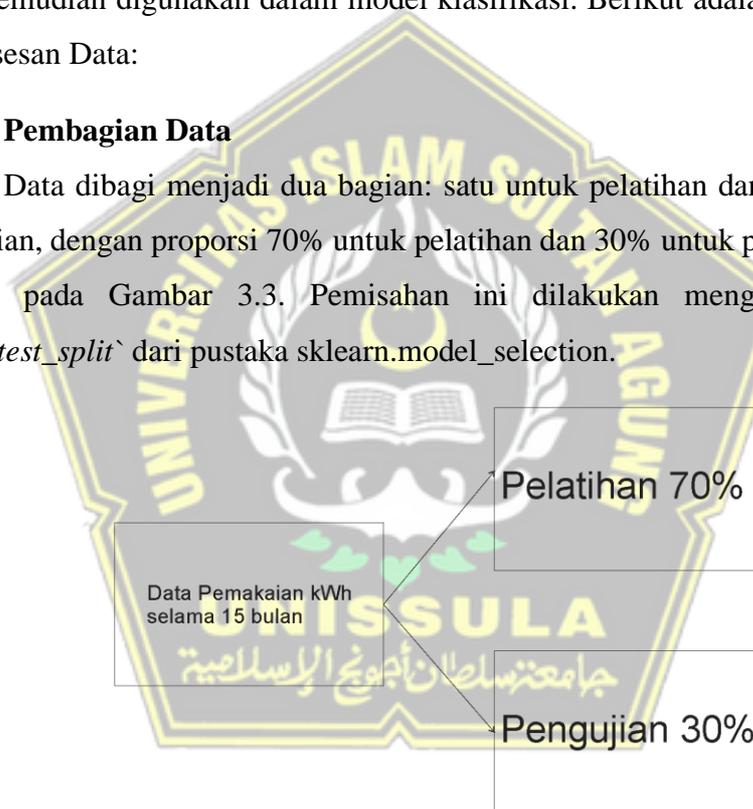
3.4.1 Pengumpulan dan Pra-pemrosesan Data

Sistem dimulai dengan memuat data dari file CSV menggunakan pustaka *pandas*, yang mencakup informasi tentang historis konsumsi listrik pelanggan (KWH) yang diperlukan untuk pengembangan model klasifikasi. Proses pra-pemrosesan data dilakukan dalam dua langkah utama. Pertama, nilai konsumsi

listrik dinormalisasi dengan menggunakan *MinMaxScaler* dari pustaka *sklearn.preprocessing*, yang mengubah rentang nilai KWH menjadi antara 0 dan 1. Setelah dilakukan normalisasi, nilai-nilai ini diubah menjadi format biner (0 atau 1) berdasarkan ambang batas 0.5, sehingga sesuai untuk proses klasifikasi biner. Langkah kedua melibatkan penentuan indikator penyalahgunaan, di mana nilai median KWH digunakan sebagai acuan. Jika nilai konsumsi listrik sama dengan atau melebihi median, indikator ini diatur ke 1; sebaliknya, jika nilai tersebut lebih rendah dari median, indikator diatur ke 0. Dengan demikian, data diubah menjadi label biner yang menandakan adanya atau tidak adanya penyalahgunaan listrik, yang kemudian digunakan dalam model klasifikasi. Berikut adalah hasil dari Pra-Pemrosesan Data:

3.4.2 Pembagian Data

Data dibagi menjadi dua bagian: satu untuk pelatihan dan satu lagi untuk pengujian, dengan proporsi 70% untuk pelatihan dan 30% untuk pengujian, seperti terlihat pada Gambar 3.3. Pemisahan ini dilakukan menggunakan fungsi `'train_test_split'` dari pustaka `sklearn.model_selection`.



Gambar 3.2 Pembagian Train Test Split

Fitur yang digunakan dalam analisis ini adalah nilai KWH yang telah dikonversi menjadi format biner, sedangkan targetnya adalah penanda yang menunjukkan ada atau tidaknya penyalahgunaan listrik.

```
from sklearn.model_selection import train_test_split
# Split the data into training and testing sets
X_train, X_test, y_train, y_test = train_test_split(X, y,
test_size=0.3, random_state=42)
```

Dalam proses pemisahan data, beberapa elemen penting harus diperhatikan. X dan Y adalah komponen utama, di mana X mewakili fitur yang digunakan untuk pelatihan model, seperti ['Bulan -1', 'Bulan -2', 'Bulan -3', 'Bulan -15'], sementara Y adalah target yang ingin diprediksi, yaitu GOL_encoded. Parameter **test_size=0.3** mengatur proporsi data yang akan digunakan untuk pengujian, yaitu 30%, dengan sisanya 70% untuk pelatihan. **random_state=42** memastikan pemisahan data yang konsisten setiap kali kode dijalankan, memungkinkan hasil yang dapat direproduksi. Fungsi `train_test_split` dari pustaka `sklearn.model_selection` digunakan untuk membagi dataset menjadi set pelatihan dan pengujian.

3.4.3 Penerapan SMOTE

Untuk mengatasi ketidakseimbangan kelas dalam data pelatihan, digunakan teknik *Synthetic Minority Oversampling Technique* (SMOTE). Teknik ini, yang diimpor dari pustaka `imblearn.over_sampling`, meningkatkan jumlah sampel dari kelas minoritas dengan membuat contoh baru secara sintetis. Ini bertujuan untuk menyeimbangkan dataset sehingga model KNN dapat dilatih dengan data yang lebih seimbang dan representatif[19].

3.4.4 Pelatihan Model KNN

Model K-Nearest Neighbors (KNN) dilatih menggunakan data yang telah dinormalisasi. Dalam proses ini, jumlah tetangga terdekat (k) diuji untuk berbagai nilai dari 1 hingga 20 menggunakan teknik *cross-validation* untuk menentukan nilai k terbaik. Model KNN kemudian dilatih dengan data pelatihan menggunakan nilai k yang terbaik, dan digunakan untuk melakukan prediksi pada data pengujian, berikut adalah contoh skrip yang digunakan:

```
from imblearn.over_sampling import SMOTE
from sklearn.neighbors import KNeighborsClassifier
from sklearn.model_selection import train_test_split
from sklearn.metrics import classification_report
# Inisialisasi dan latih model KNN
knn = KNeighborsClassifier(n_neighbors=5)
knn.fit(X_train_resampled, y_train_resampled)
# Prediksi dan evaluasi model
y_pred = knn.predict(X_test)
print('Classification Report:')
print(classification_report(y_test, y_pred))
```

Program ini menggunakan algoritma *K-Nearest Neighbors* (KNN) untuk membangun model klasifikasi berdasarkan data yang telah diimbangi dengan teknik *Synthetic Minority Oversampling Technique* (SMOTE). Pertama, model KNN dengan jumlah tetangga terdekat yang ditetapkan (dalam hal ini 5) dilatih menggunakan data yang telah diresampling ($\hat{X}_{train_resampled}$ dan $\hat{y}_{train_resampled}$). Setelah model dilatih, data uji (\hat{X}_{test}) digunakan untuk memprediksi hasil klasifikasi. Hasil prediksi kemudian dievaluasi menggunakan laporan klasifikasi, yang memberikan metrik seperti presisi, *recall*, dan *F1-score* untuk menilai kinerja model.

3.4.4.1 Perhitungan Nilai K=1

1. Bagi dataset menjadi 10 fold. Sebagai contoh Dataset dibagi sebagai berikut:
2. Latih Model: Gunakan data dari Fold 2, 3, 4, dan 5 untuk melatih model.
3. Hitung Jarak: Hitung jarak Euclidean antara data uji dan data latih.
4. Hitung Rata-rata Akurasi Misalkan kita menghitung jarak Euclidean antara data uji baris 1 dan data latih baris 10:
 - Fold 1= 1
 - Fold 2=1
 - Fold 3= 1
 - Fold 4= 1

Rata-rata Akurasi untuk k=10:

$$\text{Rata - rata Akurasi} = \frac{1+1+1+1+1+1+1+1+1+1}{10} = 1 \dots \dots \dots (3.2)$$

Dalam hal ini terlihat untuk hasil k=1, model KNN menghasilkan akurasi 100% karena setiap prediksi benar. berikut adalah gambar berdasarkan hasil ujicoba:

```
k=1: Mean cross-validation accuracy = 1.0000
k=2: Mean cross-validation accuracy = 1.0000
k=3: Mean cross-validation accuracy = 1.0000
k=4: Mean cross-validation accuracy = 1.0000
k=5: Mean cross-validation accuracy = 1.0000
k=6: Mean cross-validation accuracy = 1.0000
k=7: Mean cross-validation accuracy = 1.0000
k=8: Mean cross-validation accuracy = 1.0000
k=9: Mean cross-validation accuracy = 1.0000
k=10: Mean cross-validation accuracy = 1.0000
Best k: 1 with accuracy 1.0000
```

Gambar 3.3 Nilai Terbaik K=1

Hasil uji coba menunjukkan bahwa dengan $k = 1$, model KNN mencapai akurasi 100% karena semua prediksi tepat. Gambar 3.1 berikut ini menampilkan bagaimana $k = 1$ memberikan performa terbaik, menghasilkan setiap Prediksi yang akurat.

3.4.4.2 Hitung Jarak *Euclidean* dan *Tertangga Terdekat*

Perhitungan jarak Euclidean dalam penelitian ini telah diperbarui dengan menggunakan data konsumsi listrik dari pelanggan yang terdapat dalam dataset terbaru. Data ini mencakup riwayat konsumsi listrik selama 15 bulan terakhir serta informasi terkait daya dan tarif pelanggan.

Dalam contoh perhitungan ini, kita menggunakan $k = 8$ untuk menentukan tetangga terdekat berdasarkan nilai jarak Euclidean terkecil. Berikut adalah beberapa hasil perhitungan:

1. Jarak *Euclidean* untuk beberapa pelanggan dengan rumus sebagai berikut :

$$d(p, q) = \sqrt{\sum_{i=1}^n (q_i - p_i)^2} \dots\dots\dots(3.3)$$

Dimana :

- p, q adalah dua titik data dalam ruang n -dimensional
 - p_i dan q_i adalah nilai pada dimensi ke- i dari titik p dan q
 - $d(p, q)$ adalah jarak Euclidean antara dua titik
- Jarak antara pelanggan PARLI (IDPEL: 523070752661), ABDUL MUNIF (IDPEL: 523080056202) hingga pelanggan ke 8 TRIYONO (IDPEL: 523060418033) dihitung dengan menggunakan nilai kWh TS daya tersambung dikurangi total konsumsi listrik bulanan selama 15 bulan terakhir. Secara sederhana dirumuskan sebagai berikut :

$$\sqrt{(q_1 - p_1)^2 + (q_2 - p_2)^2 + \dots + (q_8 - p_8)^2} = d(p, q)$$

Dimana :

- q_n adalah nilai kWh TS sesuai daya tersambung
- p_n adalah nilai total pemakaian pelanggan selama 15 bulan terakhir

dengan rumusan yang sama maka diperoleh hasil perhitungan sebagai berikut :

$$\sqrt{(4957 - 1549)^2 + (2479 - 1225)^2 + \dots + (2479 - 961)^2} = 7.498,21$$

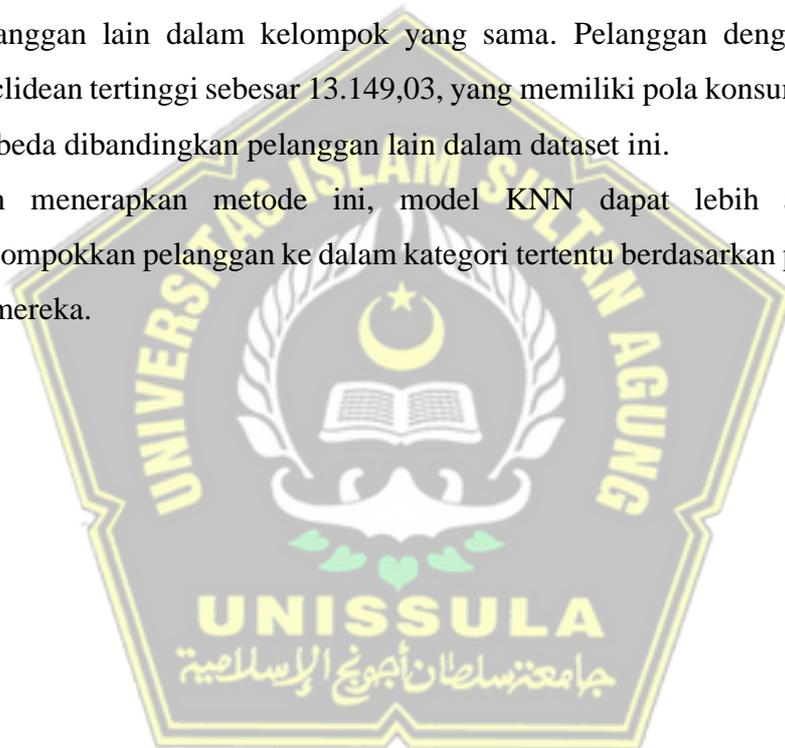
- Jarak antara pelanggan SUPARMIN LASTRI (IDPEL: 523070927861), AHMAT ALI (IDPEL: 523060151452) sampai YASIN AL FATAH (IDPEL : 523060321684) dan dihitung dengan rumus yang sama sebagai berikut:

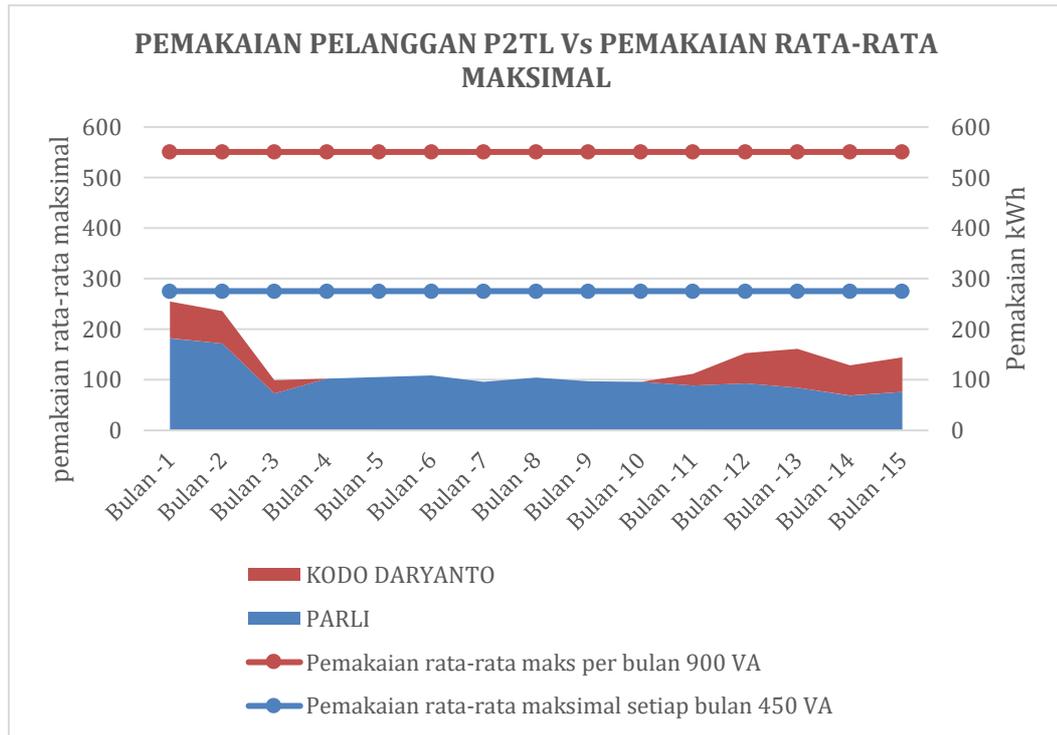
$$\sqrt{(2479 - 1933)^2 + (4957 - 3830)^2 + \dots + (12118 - 3251)^2} = 9.055,92$$

Demikian seterusnya hingga pelanggan yang terakhir ke 115.

2. Pemilihan Tetangga Terdekat: Dengan $k=8 = 8$, delapan pelanggan dengan jarak Euclidean terkecil dipilih sebagai referensi untuk prediksi klasifikasi.
3. Interpretasi Hasil nya Pelanggan dengan nilai jarak Euclidean terendah sebesar 4.020,67 cenderung memiliki pola konsumsi yang lebih mirip dengan pelanggan lain dalam kelompok yang sama. Pelanggan dengan nilai jarak Euclidean tertinggi sebesar 13.149,03, yang memiliki pola konsumsi yang lebih berbeda dibandingkan pelanggan lain dalam dataset ini.

Dengan menerapkan metode ini, model KNN dapat lebih akurat dalam mengelompokkan pelanggan ke dalam kategori tertentu berdasarkan pola konsumsi listrik mereka.



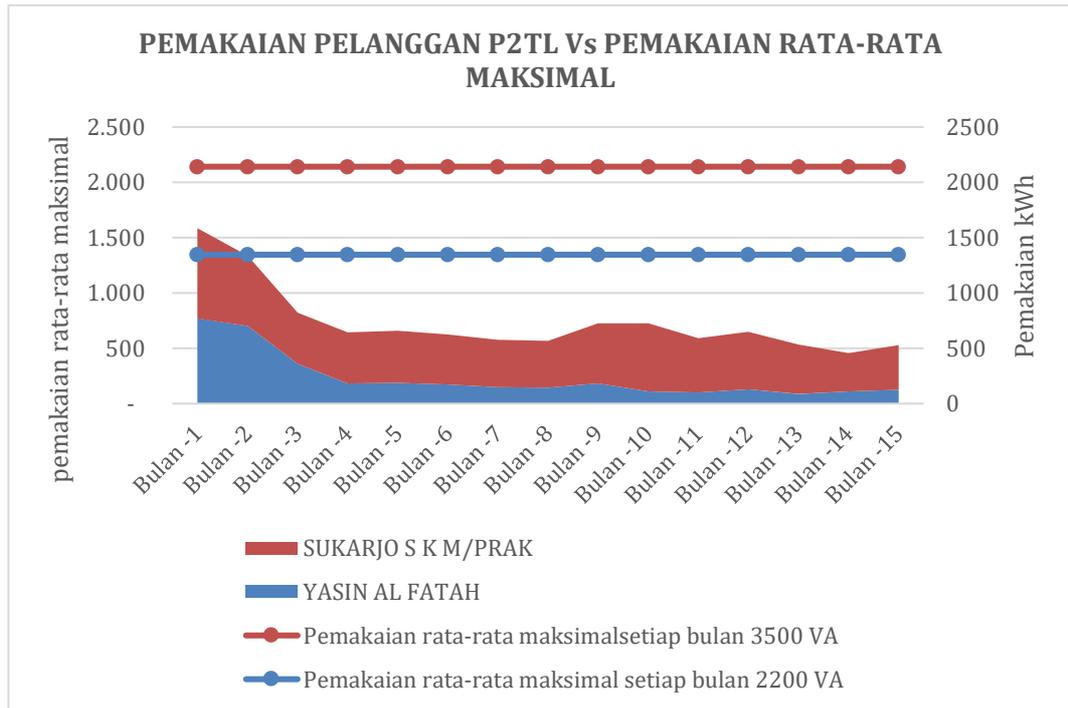


Gambar 3.4 Pemakaian Pelanggan P2TL Vs Pemakaian Rata-Rata Maksimal

Legenda:

- Area Biru dan Merah : Pemakaian pelanggan (PARLI & KODO DARYANTO).
- Garis Merah : Pemakaian rata-rata maksimal untuk daya 900 VA
- Garis Biru : Pemakaian rata-rata maksimal untuk daya 450 VA.

Menunjukkan pemakaian listrik pelanggan dengan daya 450 VA dan 900 VA dibandingkan dengan batas pemakaian rata-rata maksimal setiap bulan. Terlihat bahwa pemakaian pelanggan yang melakukan kecurangan cenderung berada jauh di bawah batas pemakaian rata-rata maksimal yang diperbolehkan. Pada bulan awal, pemakaian cukup tinggi namun mengalami penurunan setelah beberapa bulan. Pada bulan tertentu, terdapat kenaikan kembali pada pemakaian pelanggan.



Gambar 3.5 Perbandingan Pemakaian Pelanggan P2TL Vs Pemakaian Rata-Rata Maksimal

Legenda:

- Area Biru dan Merah : Pemakaian pelanggan (YASIN AL FATAH & SUKARJO S KM/PRAK).
- Garis Merah : Pemakaian rata-rata maksimal untuk daya 3500 VA.
- Garis Biru : Pemakaian rata-rata maksimal untuk daya 2200 VA.

Membandingkan pemakaian listrik pelanggan dengan daya 2200 VA dan 3500 VA dengan batas pemakaian rata-rata maksimal setiap bulan. Pemakaian pelanggan mengalami tren penurunan setelah bulan pertama dan mulai stabil di pertengahan periode. Pemakaian cenderung tetap berada di bawah batas maksimal yang diperbolehkan. Pada beberapa bulan terakhir, pemakaian mengalami sedikit kenaikan.

Maka dapat disimpulkan Pemakaian pelanggan umumnya lebih rendah dari batas rata-rata maksimal yang diizinkan sesuai dengan peraturan. Terdapat pola penurunan pemakaian setelah periode awal, yang mungkin disebabkan oleh perubahan pola konsumsi atau intervensi dari pihak terkait. Grafik ini menunjukkan bahwa mayoritas pelanggan tetap berada dalam batas pemakaian wajar sesuai daya terpasang.

3.4.5 Evaluasi Model

Untuk mengevaluasi kinerja model, beberapa metrik digunakan. Pertama, laporan klasifikasi (classification report) menyajikan informasi penting seperti precision, recall, f1-score, dan akurasi dari model. Metrik ini memberikan gambaran menyeluruh mengenai seberapa baik model dalam mengklasifikasikan data ke dalam kategori yang benar.

Kedua, Matriks kebingungan berfungsi untuk mengevaluasi kinerja model klasifikasi dengan memberikan rincian yang mendalam tentang kesalahan prediksi. Dalam matriks ini, **True Positives (TP)** adalah jumlah kasus yang benar-benar merupakan penyalahgunaan listrik dan diprediksi sebagai penyalahgunaan listrik oleh model. **True Negatives (TN)** adalah jumlah kasus yang benar-benar bukan penyalahgunaan listrik dan diprediksi sebagai bukan penyalahgunaan listrik oleh model. Sebaliknya, **False Positives (FP)** adalah kasus yang sebenarnya bukan penyalahgunaan listrik tetapi diprediksi sebagai penyalahgunaan listrik oleh model, sedangkan **False Negatives (FN)** adalah kasus yang sebenarnya merupakan penyalahgunaan listrik tetapi diprediksi sebagai bukan penyalahgunaan listrik oleh model. Akurasi model dihitung dari proporsi prediksi yang benar secara keseluruhan, **Precision** menilai akurasi prediksi positif, **Recall** mengukur seberapa baik model mendeteksi semua kasus positif, dan **F1 Score** memberikan ukuran keseimbangan antara Precision dan Recall. Dengan matriks ini, kita dapat memahami seberapa baik model dalam klasifikasi dan jenis kesalahan yang mungkin terjadi.

Ketiga, kurva Receiver Operating Characteristic (ROC) dibuat untuk menganalisis keseimbangan antara tingkat true positive dan false positive. Area di bawah kurva (AUC) dihitung untuk menilai seberapa baik model dapat membedakan antara kelas positif dan negatif. Kurva ROC memberikan pandangan tentang performa model dalam berbagai ambang batas keputusan.

3.4.6 Clusters of Trend Patterns in 3D

Data tersebut dinormalisasi menggunakan StandardScaler (*Z-Score Normalization*) untuk memastikan bahwa fitur berada pada skala yang seragam. Selanjutnya, algoritma K-Means digunakan untuk mengelompokkan data ke dalam lima cluster berdasarkan pola konsumsi listrik, dengan hasil pengelompokan

disimpan dalam kolom baru bernama 'Cluster'. Untuk mengklasifikasikan data baru ke dalam cluster yang ada, model K-Nearest Neighbors (KNN) dilatih menggunakan hasil cluster sebagai target, berikut adalah contoh dalam bentuk skrip:

```
# Load the data
df = pd.read_csv('data.csv')
# Define the feature columns (Bulan -1, Bulan -2, Bulan -3,
Bulan -4)
X = df[['Bulan -1', 'Bulan -2', 'Bulan -3', 'Bulan -4']]
# Scale the data using StandardScaler
scaler = StandardScaler()
X_scaled = scaler.fit_transform(X)
# Define the number of clusters (K)
K = 5
# Suppress the FutureWarning for n_init
warnings.filterwarnings("ignore", category=FutureWarning)
# Perform K-Means clustering
kmeans = KMeans(n_clusters=K, n_init=10)
kmeans.fit(X_scaled)
# Get the cluster labels
labels = kmeans.labels_
# Create a new column in the dataframe with the cluster labels
df['Cluster'] = labels
# Use KNN to classify new data points into the clusters
knn = KNeighborsClassifier(n_neighbors=5)
knn.fit(X_scaled, labels)
# Define a function to predict the cluster label for new data
def predict_cluster(new_data):
    new_data_scaled = scaler.transform(new_data)
    return knn.predict(new_data_scaled)
# Perform PCA to reduce dimensionality to 3D
pca = PCA(n_components=3)
X_pca = pca.fit_transform(X_scaled)
```

Selanjutnya, metode Principal Component Analysis (PCA) diterapkan untuk mereduksi dimensi data ke dalam ruang tiga dimensi, memungkinkan visualisasi cluster dalam bentuk grafik 3D. Grafik ini menampilkan bagaimana data dikelompokkan berdasarkan tiga komponen utama, memberikan pandangan yang jelas mengenai distribusi dan pemisahan cluster dalam ruang fitur. Visualisasi ini membantu dalam memahami struktur pola konsumsi listrik dan bagaimana data yang berbeda dipisahkan dalam ruang tiga dimensi.

3.4.7 Penghitungan Jarak dan Prediksi Kelas

Program ini melatih model K-Nearest Neighbors (KNN) dengan menggunakan nilai `k` yang dipilih melalui teknik cross-validation. Untuk setiap nilai `k`, model menghitung jarak dari setiap titik data ke tetangga terdekat di data pelatihan. Jarak ini kemudian digunakan untuk mengklasifikasikan data baru dengan menentukan kelas mayoritas di antara tetangga terdekat tersebut. Proses ini memungkinkan model untuk membuat keputusan klasifikasi berdasarkan kedekatannya dengan data pelatihan yang ada. Hasil dari teknik ini membantu dalam memahami bagaimana model menggunakan informasi jarak untuk melakukan prediksi kelas.

```
# Fungsi untuk menghitung jarak Euclidean antara dua titik
def euclidean_distance(x1, x2):
    return np.sqrt(np.sum((x1 - x2) ** 2))
# Fungsi KNN untuk memprediksi kelas
def knn_predict(X_train, y_train, X_test, k):
    predictions = []
    for test_point in X_test:
        # Hitung jarak antara test_point dan data di X_train
        distances = [euclidean_distance(test_point, x_train)
                     for x_train in X_train]
        # Ambil k jarak terkecil
        k_indices = np.argsort(distances)[:k]
        # Ambil label yang sesuai dengan k data terdekat
        k_nearest_labels = [y_train[i] for i in k_indices]
        # Tentukan prediksi berdasarkan label yg paling umum
        most_common = Counter(k_nearest_labels).most_common(1)
        predictions.append(most_common[0][0])
    return predictions
```

Fungsi ini terdiri dari dua bagian utama: pertama, `euclidean_distance` yang digunakan untuk menghitung jarak *Euclidean* antara dua titik data, dengan cara menjumlahkan kuadrat selisih antara koordinat masing-masing titik, kemudian mengambil akar kuadratnya. Kedua, fungsi `knn_predict` yang memprediksi kelas dari titik uji (`X_test`) berdasarkan data latih (`X_train` dan `y_train`). Untuk setiap titik uji, fungsi menghitung jarak Euclidean ke semua titik latih, kemudian memilih `k` titik dengan jarak terdekat. Berdasarkan label dari `k` titik terdekat ini, fungsi menentukan prediksi kelas dengan memilih label yang paling sering muncul di antara mereka, dan menyimpan prediksi tersebut dalam daftar hasil.

3.4.8 Jumlah Data Penyalahgunaan

Berikut adalah program Python yang menghitung jumlah data untuk kategori penyalahgunaan listrik (biner 1) dan kategori non-penyalahgunaan (biner 0) berdasarkan kolom `KWH_Binary` dalam dataset:

```
import pandas as pd
# Load the data from CSV
data = pd.read_csv('table.csv')
# Check the total number of rows
total_rows = len(data)
# Check for missing values in 'KWH_Binary'
missing_values = data['KWH_Binary'].isnull().sum()
# Print the results
print(f"Total rows in the dataset: {total_rows}")
print(f"Missing values in 'KWH_Binary' column:
{missing_values}")
# Ensure 'KWH_Binary' column exists
if 'KWH_Binary' in data.columns:
    binary_counts = data['KWH_Binary'].value_counts()
    # Extract counts for each binary class
    misuse_count = binary_counts.get(1, 0)
    non_misuse_count = binary_counts.get(0, 0)
    # Print the results
    print(f"Total data indicating misuse (KWH_Binary = 1):
{misuse_count}")
    print(f"Total data indicating no misuse (KWH_Binary = 0):
{non_misuse_count}")
else:
    print("The column 'KWH_Binary' does not exist in the
dataset.")
```

Program ini menggunakan pustaka `pandas` untuk memproses data dari file `CSV`. Setelah memuat data menggunakan `pd.read_csv()`, program memeriksa keberadaan kolom `GOL_encoded` dalam dataset. Selanjutnya, `value_counts()` digunakan untuk menghitung jumlah data dalam setiap kategori di kolom tersebut, dengan `binary_counts.get(1, 0)` dan `binary_counts.get(0, 0)` memberikan jumlah kategori **pelanggaran (1)** dan **non-pelanggaran (0)** masing-masing, sambil memastikan nilai default 0 jika kategori tidak ada. Akhirnya, program menampilkan jumlah data untuk setiap kategori, memberikan gambaran jelas tentang distribusi data dalam kolom `GOL_encoded`.

BAB IV

HASIL PENELITIAN DAN PEMBAHASAN

Pada bab ini, akan dibahas hasil penelitian mengenai perbandingan implementasi algoritma K-Nearest Neighbors (KNN) untuk prediksi penyalahgunaan listrik di PLN UP3 Demak. Penelitian ini bertujuan untuk mengetahui pola penyalahgunaan listrik, khususnya pola P2 dan P3, serta membandingkan hasil prediksi menggunakan teknik SMOTE dan tanpa SMOTE.

Data yang digunakan dalam penelitian ini bersumber dari PT PLN UP3 Demak dan mencakup informasi pemakaian listrik pelanggan yang kedapatan melakukan penyalahgunaan listrik selama periode Februari 2023 hingga April 2024. Data ini meliputi pemakaian kWh atau tagihan listrik selama 15 bulan terakhir.

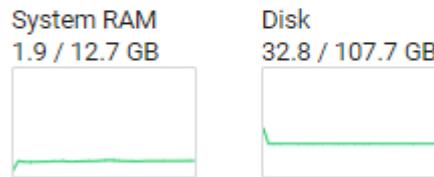
Google Colab dipilih karena gratis, memiliki kemampuan komputasi yang tinggi dengan GPU, dan mendukung Python, yang sangat berguna untuk machine learning.

Tabel 4.1 Perbandingan Google Colab dengan Matlab

Kriteria	Google Colab	MATLAB
Aksesibilitas	Gratis, Berbasis Cloud	Lisensi Berbayar, Desktop
Kemampuan Komputasi	GPU/TPU Akses Mudah	GPU, Memerlukan Konfigurasi
Ekosistem	Python, Pustaka Machine Learning	MATLAB, Toolboxes Khusus
Kolaborasi	Mudah Berbagi via Google Drive	Terbatas pada Pengguna MATLAB
Reproduksibilitas	Sangat Baik (Notebooks)	Baik, Namun Perlu Pengaturan

Pada Tabel 4.1 nampak memudahkan kolaborasi dan memastikan hasil penelitian bisa diulang, sehingga lebih praktis untuk eksperimen dan analisis data yang rumit dibandingkan MATLAB.

Python 3 Google Compute Engine backend
Showing resources from 6:03 PM to 7:21 PM



Gambar 4.1 Hardware Google Colab yang di gunakan

Pada gambar 4.1 menunjukkan spesifikasi hardware yang digunakan pada Google Colab, termasuk dukungan GPU yang memungkinkan pemrosesan data secara cepat dan efisien

4.1 Deskripsi Data

Dataset yang digunakan dalam penelitian ini mencakup data penggunaan listrik dari 115 pelanggan yang telah teridentifikasi melakukan pelanggaran. Pelanggaran ini dikelompokkan ke dalam dua kategori utama, yaitu pelanggaran golongan P2 (P2) dan golongan P3 (P3). Setiap pelanggan diidentifikasi berdasarkan variabel ID pelanggan (`IDPEL`) dan diklasifikasikan sesuai dengan kategori pelanggaran mereka (`GOL`).

Data ini mencakup penggunaan listrik selama 15 bulan terakhir, dengan variabel `Bulan -1` hingga `Bulan -15`, yang merekam konsumsi listrik dalam kilowatt-hour (kWh) setiap bulannya. Selain itu, dataset ini juga mencakup informasi tambahan seperti alamat pelanggan (`ALAMAT`), daya terpasang (`DAYA`), dan total penggunaan kWh (`KWH`). Pola tren konsumsi listrik dari setiap pelanggan selama periode tersebut dianalisis dan ditampilkan dalam kolom `Trend Pattern`.

	NOAGENDA	IDPEL	NAMA	GOL	\
0	P2TL/52554/20240103/00001	523070752661	PARLI	P2	
1	P2TL/52552/20240104/00007	523080056202	ABDUL MUNIF	P2	
2	P2TL/52554/20240117/00001	523071318622	KODO DARYANTO	P3	
3	P2TL/52551/20240117/00003	523060401763	SULKAN B SANUSI	P3	
4	P2TL/52551/20240102/00001	523060016994	TMP RATNA BINTARU	P3	
..	
110	P2TL/52553/20240426/00010	523071674042	JASMIAH	P2	
111	P2TL/52551/20240430/00010	523060046701	KAMJAWI	P3	
112	P2TL/52554/20240422/00001	523071627360	SUWARTI	P2	
113	P2TL/52552/20240426/00014	523080797535	MURTADHO	P2	
114	P2TL/52552/20240130/00022	523080565378	RUSMAN	P3	

	ALAMAT	DAYA	KWH	Bulan -1	Bulan -2	Bulan -3	...	\
0	JA MBANGAN BARAT	R1/900	4,957	182	172	73	...	
1	DS JAWONG KEMBAN	R1/450	2,479	212	179	107	...	
2	DK MADOH BANGSRI	R1/450	2,479	73	64	26	...	
3	NG AWEN RT1/3 WED	R1/450	2,479	77	65	62	...	
4	JL PEMUDA	S2/450	2,479	230	192	84	...	
..	
110	DK PONDOK DEMPEL	R1/450	2,479	103	103	111	...	
111	KA TONSARI	R1M/900	4,957	169	400	400	...	
112	DK JETAK P KULON	R1/900	4,957	228	207	214	...	
113	DS BRAMBANG RT.02	R1/450	2,479	44	42	43	...	
114	TL OGOREJO RT 1/4 K	R1/900	4,957	150	157	145	...	

	Bulan -7	Bulan -8	Bulan -9	Bulan -10	Bulan -11	Bulan -12	Bulan -13	\
0	96	105	97	96	89	93	84	
1	64	57	64	63	82	51	68	
2	0	0	0	0	23	60	77	
3	78	68	89	79	91	21	80	
4	132	107	74	84	99	76	60	
..	
110	95	94	90	89	95	71	102	
111	120	77	50	62	80	36	85	
112	140	132	171	164	162	154	128	
113	47	47	45	50	100	1	49	
114	155	152	102	145	137	114	141	

	Bulan -14	Bulan -15	Trend Pattern
0	69	76	[Decreasing, Decreasing, Increasing, Increasin...
1	58	50	[Decreasing, Decreasing, Decreasing, Increasin...
2	60	68	[Decreasing, Decreasing, Decreasing, Stable, S...
3	40	43	[Decreasing, Decreasing, Increasing, Decreasin...
4	45	0	[Decreasing, Decreasing, Decreasing, Increasin...
..
110	80	78	[Stable, Increasing, Decreasing, Decreasing, I...
111	36	77	[Increasing, Stable, Decreasing, Decreasing, D...
112	111	141	[Decreasing, Increasing, Decreasing, Decreasin...
113	29	46	[Decreasing, Increasing, Increasing, Increasin...
114	139	111	[Increasing, Decreasing, Decreasing, Increasin...

[115 rows x 23 columns]

Gambar 4.2 Pola Konsumsi Listrik Pelanggan P2 dan P3

Pada Gambar 4.2 ini telah melalui proses preprocessing, termasuk imputasi nilai yang hilang, normalisasi, dan encoding variabel kategorikal, untuk memastikan kualitas data dan kesiapannya untuk analisis lebih lanjut. Data ini memberikan wawasan penting mengenai pola konsumsi listrik yang mencurigakan dan membantu dalam prediksi serta identifikasi penyalahgunaan listrik di wilayah yang diteliti.

Berikut untuk penjelasan Perhitungan Jarak *Euclidean* Menggunakan $k = 8$ pada temuan nilai K terdekat:

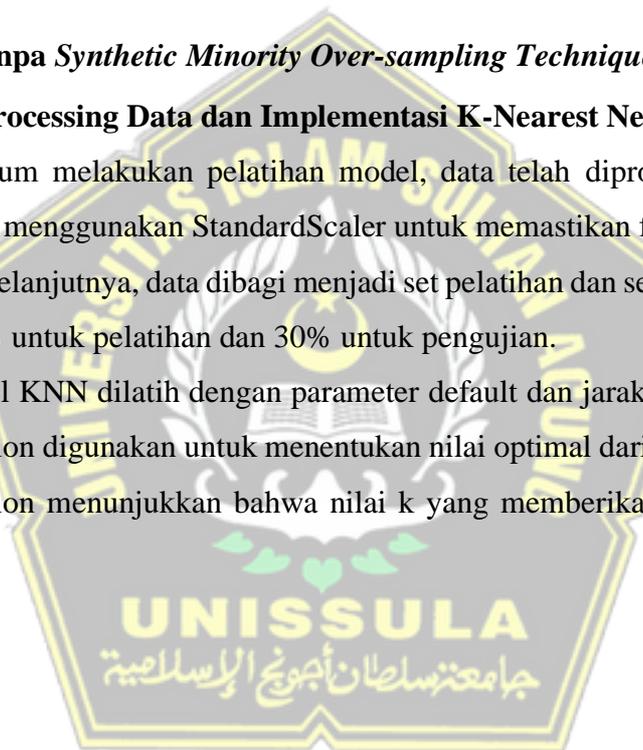
- Jarak IDPEL1 dan IDPEL8 $\sqrt{(4957-1549)^2} + \sqrt{(2479-1225)^2} + \sqrt{(2479-451)^2} + \sqrt{(2479-979)^2} + \sqrt{(2479-1195)^2} + \sqrt{(4957-1124)^2} + \sqrt{(4957-711)^2} + \sqrt{(2479-9961)^2} = 7498,21$
- Jarak IDPEL9 dan IDPEL16 $\sqrt{(2479-1933)^2} + \sqrt{(4957-3830)^2} + \sqrt{(2479-1464)^2} + \sqrt{(2479-2945)^2} + \sqrt{(2479-2684)^2} + \sqrt{(2479-1431)^2} + \sqrt{(2479-443)^2} + \sqrt{(2479-3521)^2} = 9055,92$
- Jarak IDPEL17 dan IDPEL24 $\sqrt{(4957-1887)^2} + \sqrt{(19278-7516)^2} + \sqrt{(2479-748)^2} + \sqrt{(4957-2471)^2} + \sqrt{(4957-2599)^2} + \sqrt{(2479-1432)^2} + \sqrt{(2479-1741)^2} + \sqrt{(4957-1999)^2} = 13149,03$
-
- Jarak IDPEL 113 dan IDPEL115 $\sqrt{(4957-2494)^2} + \sqrt{(2479-694)^2} + \sqrt{(4957-2102)^2} + \sqrt{(0-0)^2} = 4171,76$

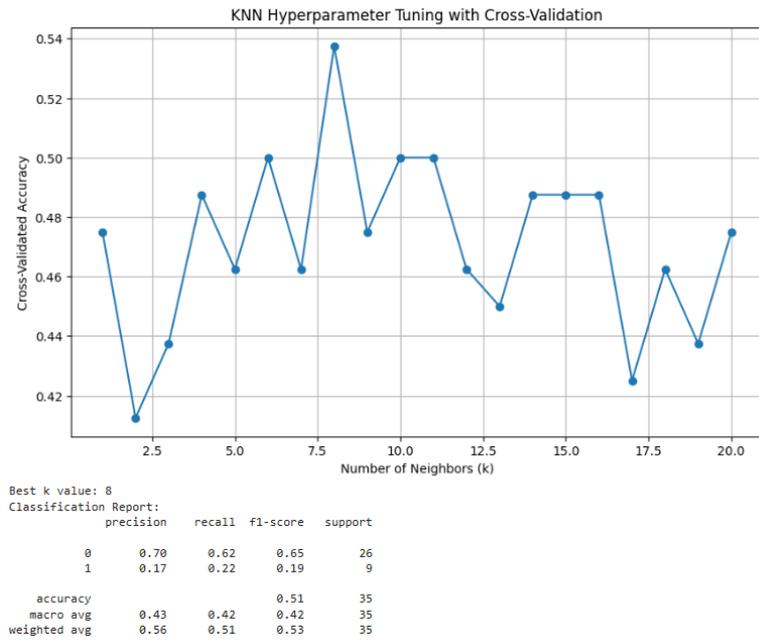
4.2 Hasil Tanpa *Synthetic Minority Over-sampling Technique* (SMOTE)

4.2.1 Preprocessing Data dan Implementasi K-Nearest Neighbors (KNN)

Sebelum melakukan pelatihan model, data telah diproses dengan teknik standardisasi menggunakan StandardScaler untuk memastikan fitur memiliki skala yang sama. Selanjutnya, data dibagi menjadi set pelatihan dan set pengujian dengan proporsi 70% untuk pelatihan dan 30% untuk pengujian.

Model KNN dilatih dengan parameter default dan jarak Euclidean. Teknik cross-validation digunakan untuk menentukan nilai optimal dari parameter k. Hasil cross-validation menunjukkan bahwa nilai k yang memberikan performa terbaik adalah 8.

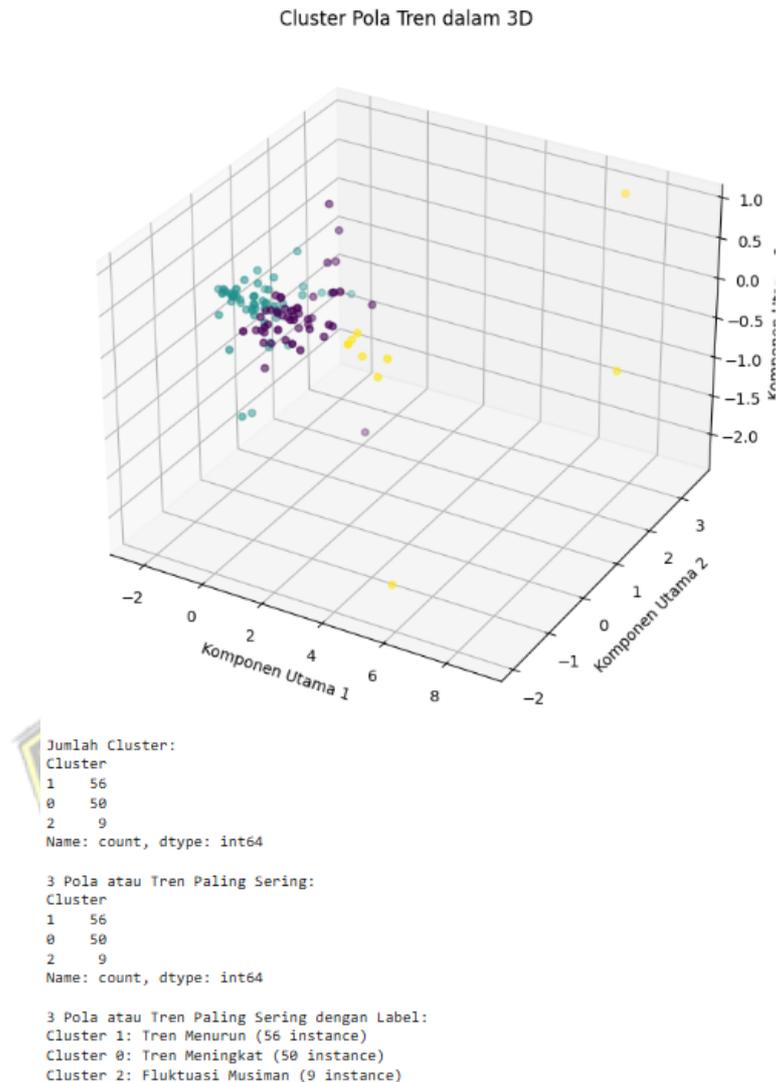




Gambar 4.3 Hasil Evaluasi K-Nearest Neighbors (KNN) tanpa SMOTE

Hasil evaluasi pada gambar 4.3 model K-Nearest Neighbors (KNN) dengan nilai k terbaik yaitu 8 menunjukkan performa yang bervariasi antara dua kelas dalam dataset. Model berhasil mengidentifikasi kasus tanpa pelanggaran (kelas 0) dengan baik, menghasilkan precision 0.70, recall 0.62, dan f1-score 0.65. Namun, untuk kelas pelanggaran (kelas 1), model kurang efektif, dengan precision 0.17 dan recall 0.22, serta f1-score yang rendah yaitu 0.19. Akurasi keseluruhan model adalah 0.51.

4.2.2 Distribusi dan Pola dalam Hasil Clustering



Gambar 4.4 Hasil analisis clustering tanpa SMOTE

Hasil analisis clustering pada gambar 4.4 menunjukkan distribusi jumlah instance yang signifikan antara cluster-cluster yang dihasilkan. Cluster 0, yang diberi label sebagai "Tren Meningkat," memiliki jumlah instance terbanyak dengan 50 kasus, menandakan bahwa sebagian besar data cenderung menunjukkan pola peningkatan penggunaan atau tren yang positif. Diikuti oleh Cluster 1, yang diberi label sebagai "Tren Menurun," dengan 56 kasus, menunjukkan bahwa ada sejumlah data yang menunjukkan penurunan penggunaan atau tren negatif. Cluster 2, dengan label "Fluktuasi Musiman," memiliki jumlah kasus yang jauh lebih sedikit, yaitu 9 instance, menandakan bahwa pola fluktuasi musiman hanya muncul dalam sedikit data. Distribusi ini mengindikasikan bahwa sebagian besar data cenderung

menunjukkan pola peningkatan atau penurunan, sementara fluktuasi musiman relatif jarang terjadi dalam dataset ini.

```

Jumlah Cluster untuk P2:
Cluster
2    37
0    25
1     8
Name: count, dtype: int64

3 Pola atau Tren Paling Sering untuk P2:
Cluster
2    37
0    25
1     8
Name: count, dtype: int64

Jumlah Cluster untuk P3:
Cluster
0    22
2    18
1     5
Name: count, dtype: int64

3 Pola atau Tren Paling Sering untuk P3:
Cluster
0    22
2    18
1     5
Name: count, dtype: int64

```

Gambar 4.5 Hasil Clustering berdasarkan P2 dan P3

Pada Gambar 4.5 dalam kedua dataset P2 dan P3, pola tren meningkat adalah pola yang paling umum diidentifikasi, dengan Cluster 1 mewakili pola ini sebagai cluster terbesar.

Tabel 4.2 Cluster dan pola tren paling sering untuk P2 dan P3

Golongan	Cluster	Jumlah Pelanggan	Pola Tren
P2	1	8	Tren Menurun
	0	25	Tren Meningkatkan
	2	37	Fluktuasi Musiman
P3	1	5	Tren Menurun
	0	22	Tren Meningkatkan
	2	18	Fluktuasi Musiman

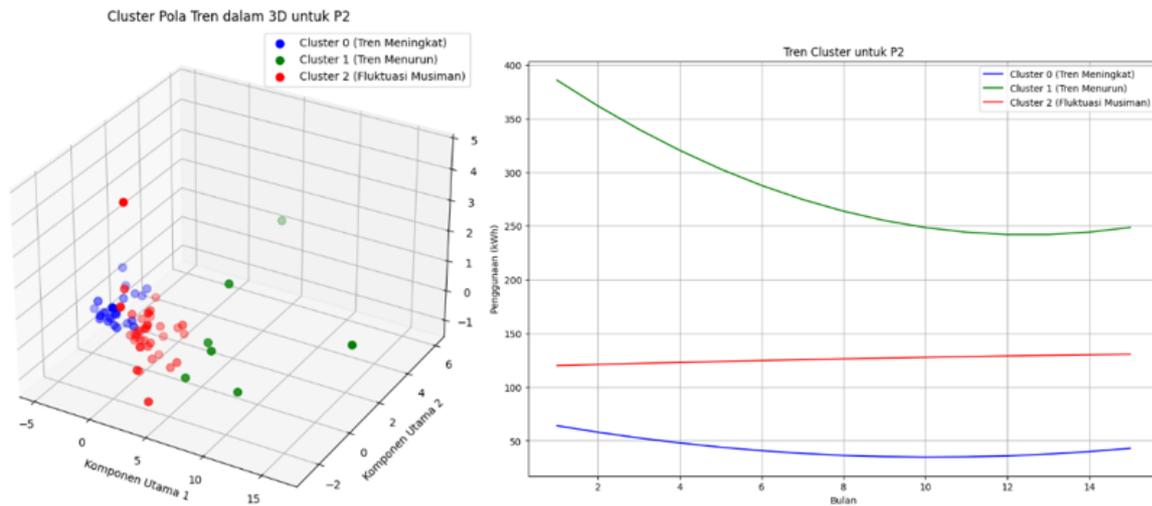
Berdasarkan tabel 4.2 Untuk dataset P2, **Cluster 2** adalah yang paling umum dengan 37 instance, menunjukkan pola fluktuasi musiman yang dominan. **Cluster 0** dengan 25 instance menunjukkan pola tren meningkat, dan **Cluster 1** yang memiliki 8 instance, menggambarkan Tren Menurun, yang sangat jarang ditemukan.

Pada dataset P3, **Cluster 0** paling banyak dengan 22 instance, menandakan pola tren meningkat. **Cluster 2** dengan 18 instance menunjukkan pola fluktuasi musiman, sedangkan **Cluster 1** dengan 5 instance menunjukkan pola tren menurun, yang juga jarang terjadi.

4.2.3 Pola Tren Tanpa SMOTE

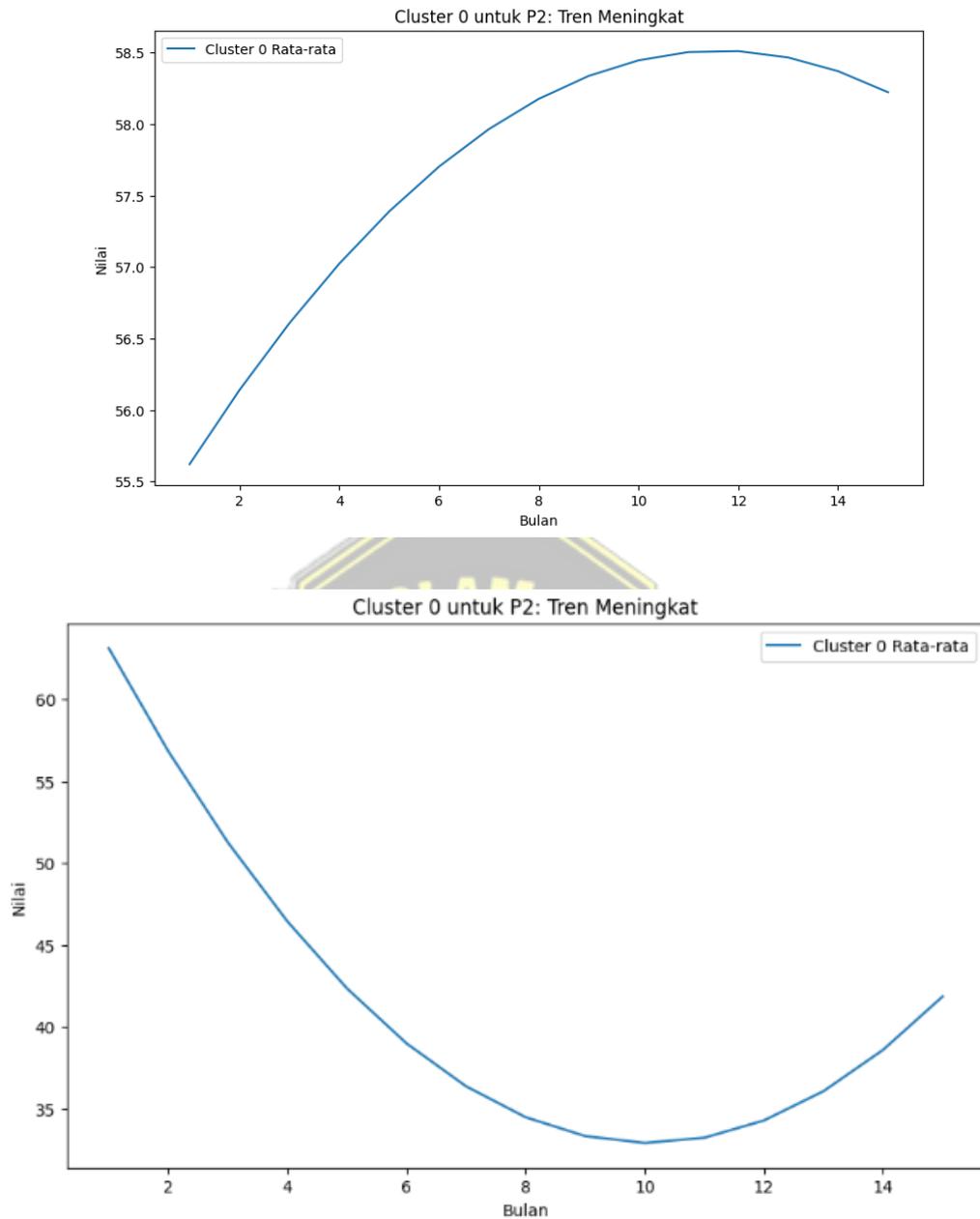
4.2.3.1 Pola Tren P2 tanpa SMOTE

Secara keseluruhan, pola tren meningkat adalah yang paling umum di kedua dataset, sedangkan fluktuasi musiman jarang terjadi.



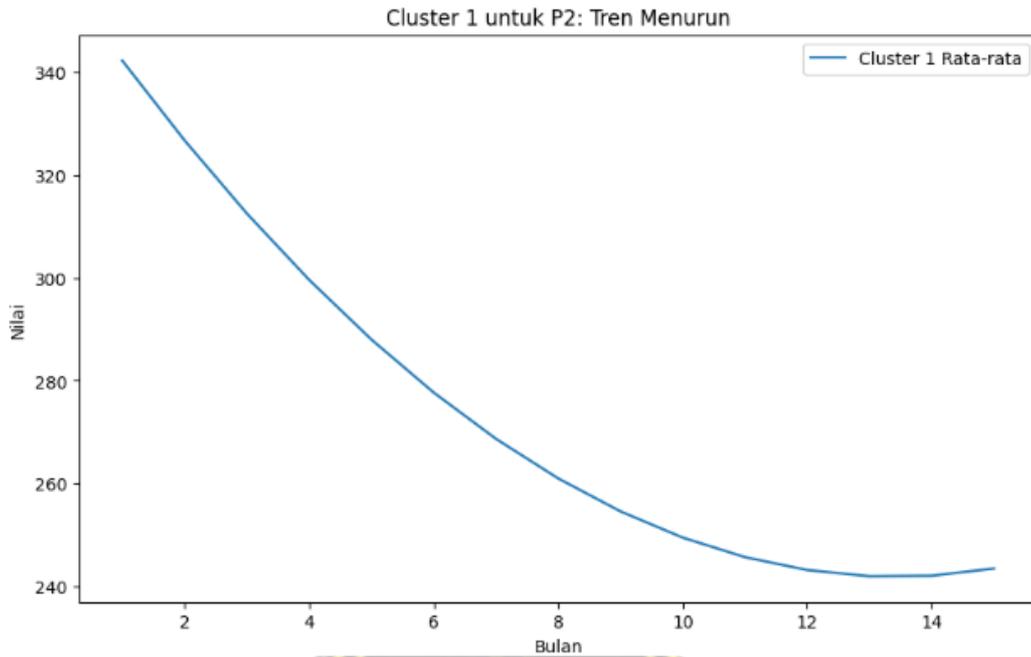
Gambar 4.6 Tren Cluster untuk P2

Pada tren cluster P2 Pelanggan dengan pemakaian kwh diatas 300 perbulannya terjadi penurunan dan hal ini baik walaupun sempat mengalami kenaikan pada 4 bulan pertama. Sedangkan tren cluster P2 Pelanggan dengan pemakaian kwh ± 100 cenderung stabil dan mengalami peningkatan, kemudian tren cluster P2 Pelanggan dengan pemakaian kwh di bawah 100 perbulannya mengalami penurunan.



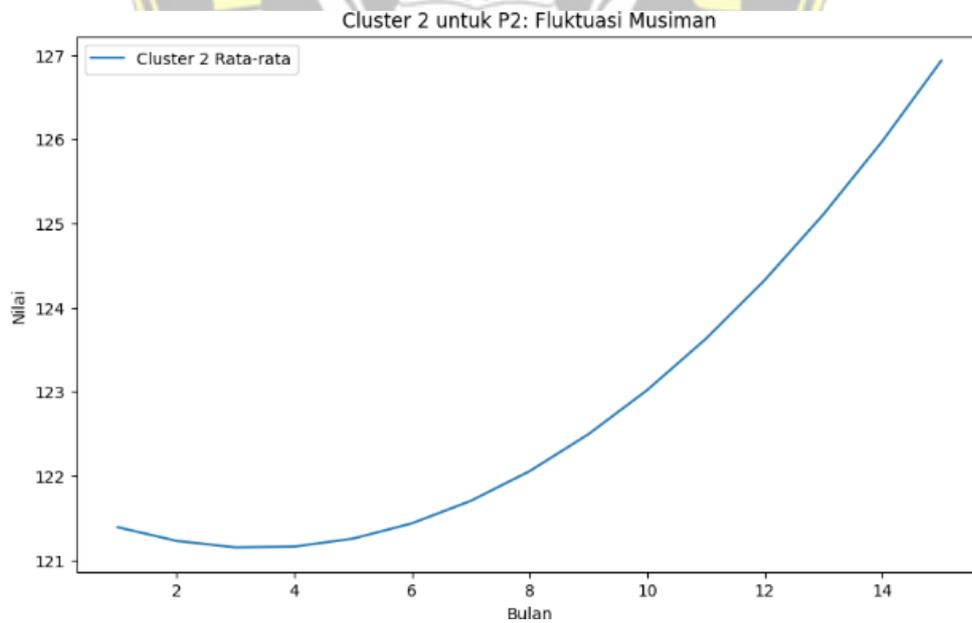
Gambar 4.7 Klaster 0 P2

Pada Gambar 4.7 sebanyak 37 pelanggan P2 cenderung melakukan kecurangan dengan penggunaan kwh listrik 35 sampai 65 dan rata2 tren ini cenderung menurun.



Gambar 4.8 Klaster 1 P2

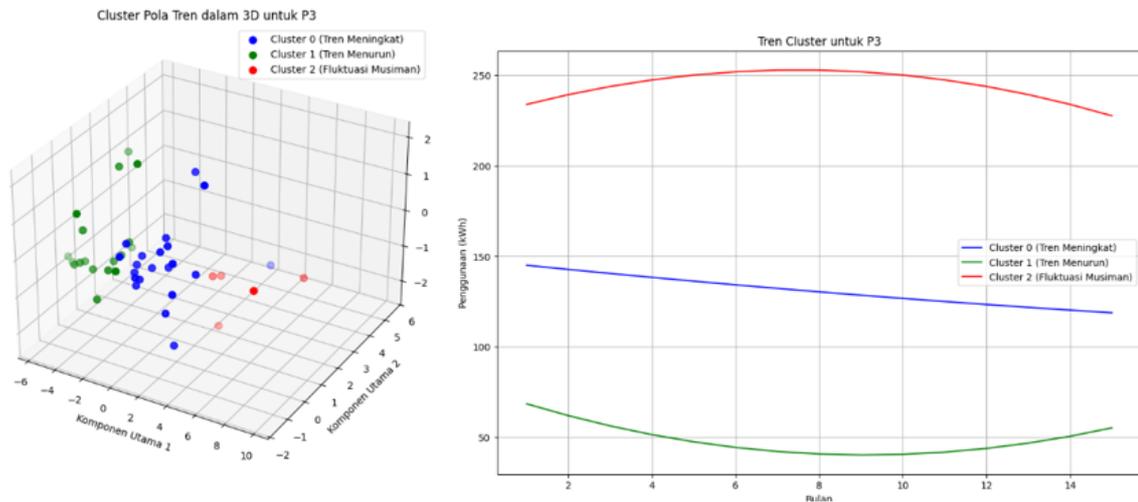
Pada Gambar 4.8 sebanyak 25 pelanggan P2 cenderung melakukan kecurangan dengan penggunaan kwh listrik 340 turun sampai 240 dan rata2 tren ini cenderung turun.



Gambar 4.9 Klaster 2 P2

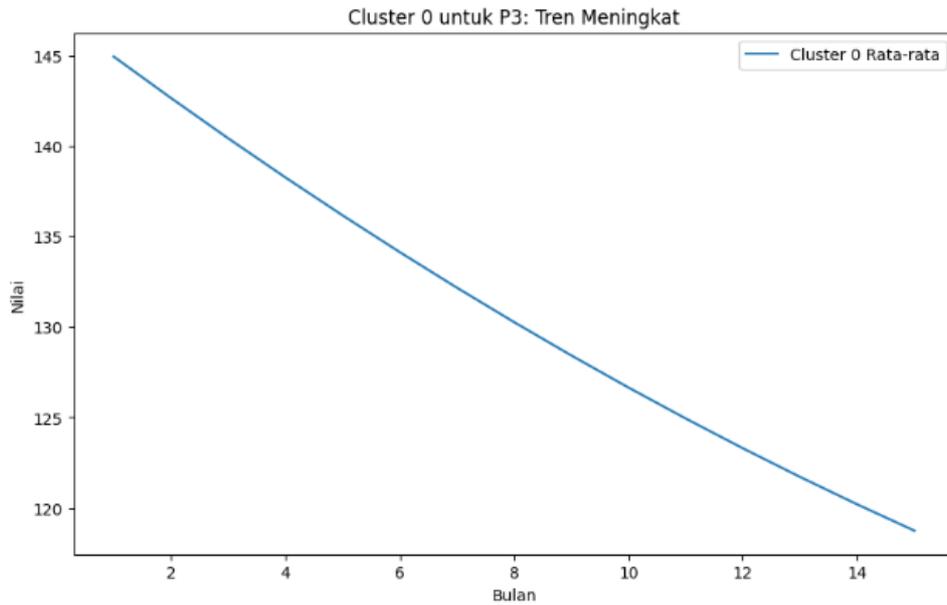
Pada Gambar 4.9 sebanyak 8 pelanggan P2 cenderung melakukan kecurangan dengan penggunaan kwh listrik 121 sampai 127 dan rata2 tren ini cenderung naik, sebelum naik signifikan selama 8 bulan Akhir.

4.2.3.2 Pola tren P3 tanpa SMOTE



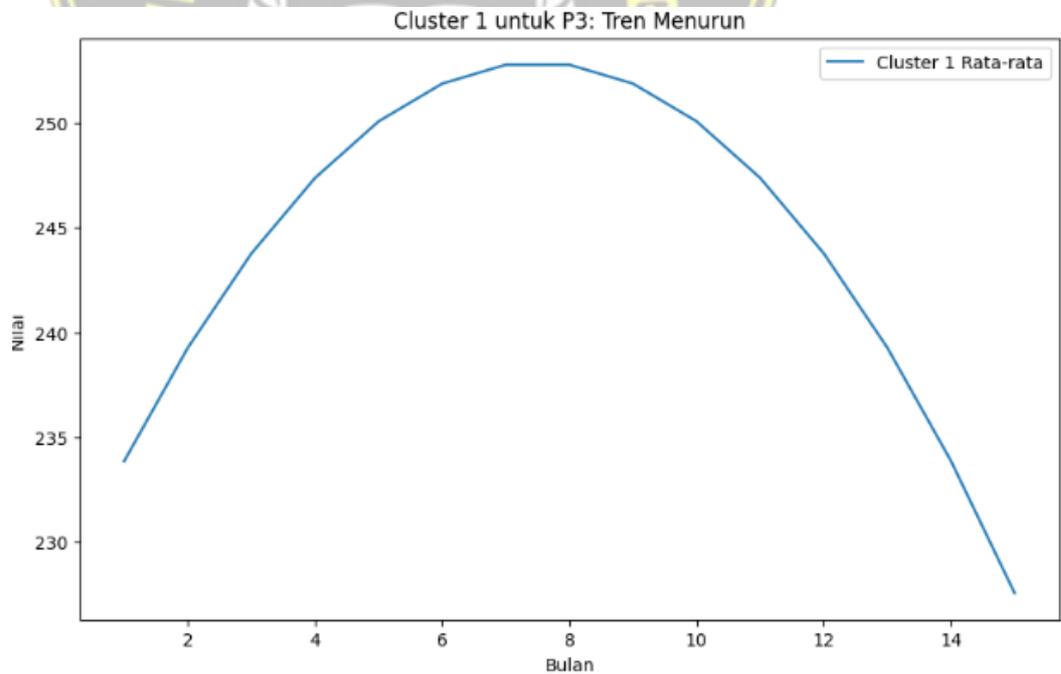
Gambar 4.10 Tren Cluster untuk P3

Pada tren cluster P3 Pelanggan dengan pemakaian kwh diatas 235 perbulannya terjadi penurunan pemakaian kwh selama 15 bulan terakhir. Sedangkan tren cluster P3 Pelanggan dengan pemakaian kwh 125 juga mengalami penurunan selama 15 bulan terakhir walaupun mengalami penurunan pada 6 bulan pertama, kemudian tren cluster P3 pelanggan dengan pemakaian kwh di bawah 75 perbulannya cenderung mengalami penurunan dan ini baik.



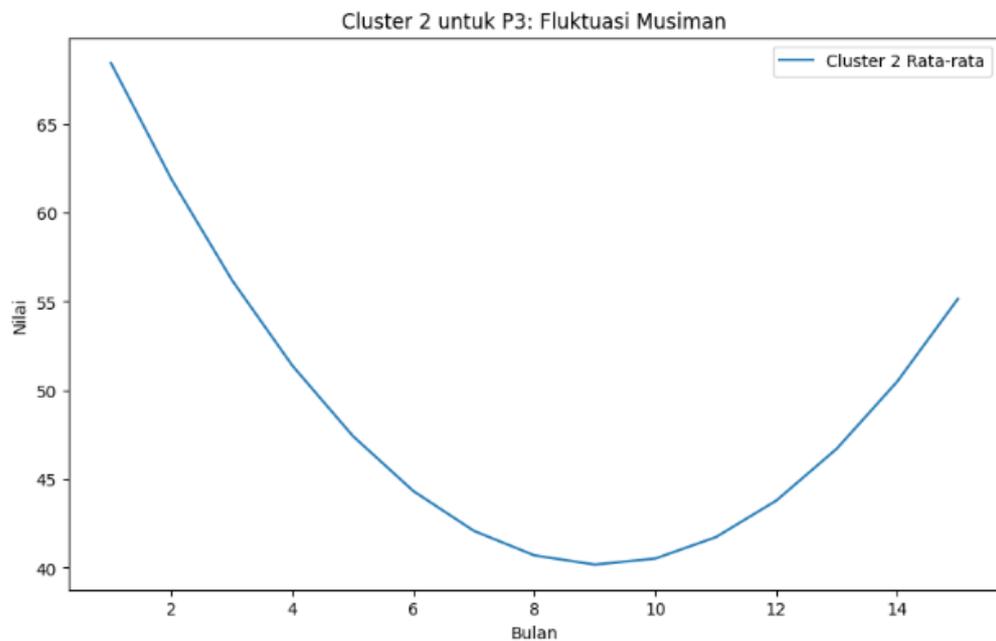
Gambar 4.11 Tren Cluster 0 untuk P3

Pada Gambar 4.11 sebanyak 22 pelanggan P3 cenderung melakukan kecurangan dengan penggunaan kwh listrik 145 turun ke 120 dan rata2 tren ini cenderung turun dan menunjukkan hal yang baik.



Gambar 4.12 Tren Cluster 1 untuk P3

Pada Gambar 4.12 sebanyak 5 pelanggan P2 cenderung melakukan kecurangan dengan penggunaan kwh listrik 233 sampai 156 dan rata2 tren ini cenderung turun, sebelum naik signifikan selama 8 bulan pertama.



Gambar 4.13 Tren Cluster 2 untuk P3

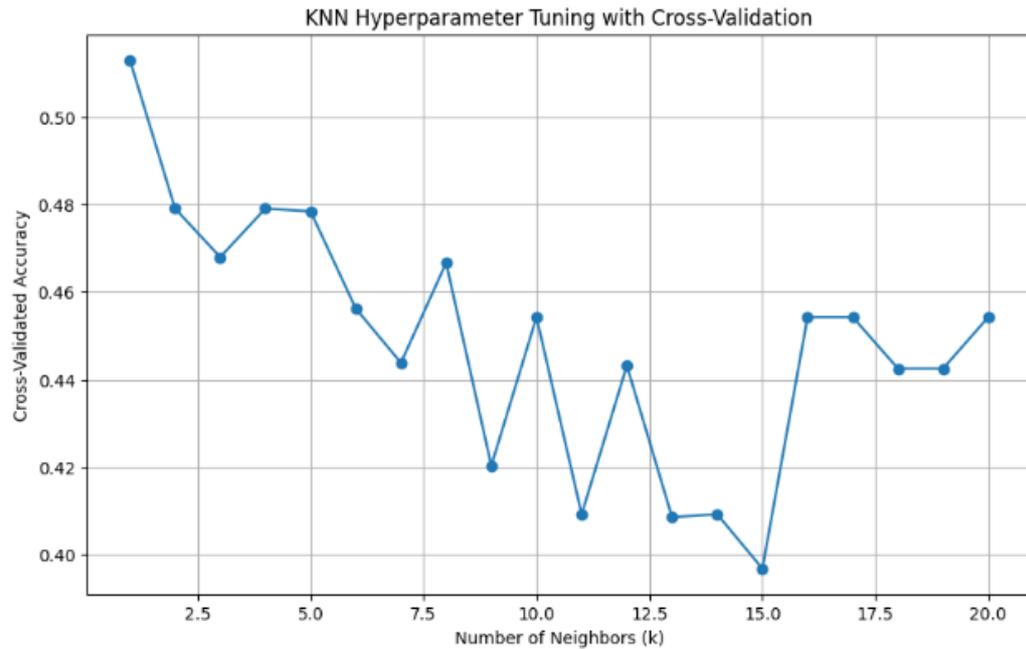
Pada Gambar 4.1 sebanyak 18 pelanggan P3 cenderung melakukan kecurangan dengan penggunaan kwh listrik 40 sampai 70 dan rata-rata tren ini cenderung turun, sebelum naik signifikan selama 4 bulan akhir.

4.3 Hasil dengan Synthetic Minority Over-sampling Technique (SMOTE)

4.3.1 Preprocessing Data dan Implementasi K-Nearest Neighbors (KNN)

Sebelum pelatihan model, data yang tidak seimbang diperbaiki dengan menggunakan teknik Synthetic Minority Over-sampling Technique (SMOTE). SMOTE digunakan untuk menghasilkan sampel sintesis dari kelas minoritas sehingga data menjadi lebih seimbang.

Setelah menerapkan SMOTE, model KNN dilatih kembali dengan parameter yang sama seperti pada eksperimen sebelumnya. Nilai k yang optimal tetap 1 berdasarkan hasil cross-validation.



```

Best k value: 1
Classification Report:
      precision    recall  f1-score   support

   0       0.72      0.50      0.59         26
   1       0.24      0.44      0.31          9

 accuracy          0.49         35
 macro avg         0.48         35
 weighted avg      0.60         35

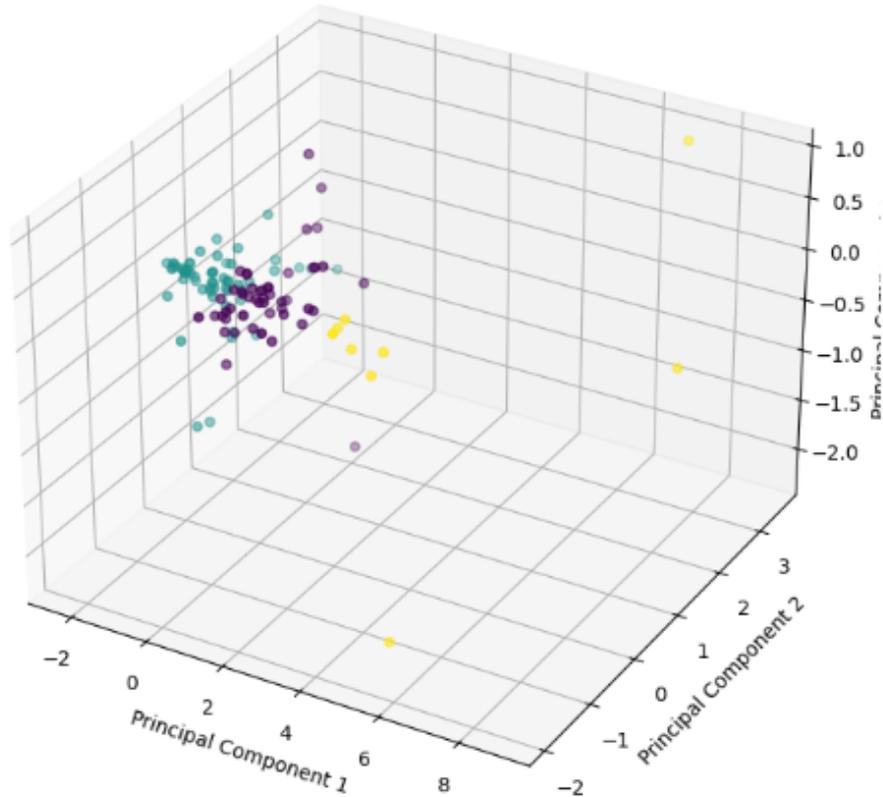
```

Gambar 4.14 Hasil Evaluasi K-Nearest Neighbors (KNN)

Hasil evaluasi model K-Nearest Neighbors (KNN) dengan nilai k terbaik yaitu **1** menunjukkan performa yang bervariasi antara dua kelas dalam dataset. Model berhasil mengidentifikasi kelas 0 (No Violation) dengan precision 0.72 dan recall 0.50, yang menunjukkan bahwa model cukup baik dalam mengenali kasus tanpa pelanggaran. Namun, untuk kelas 1 (Violation), model kurang efektif dengan precision 0.24 dan recall 0.44, yang berarti model sering kali tidak berhasil mengidentifikasi kasus pelanggaran secara akurat. Akurasi keseluruhan model adalah 0.49, menandakan bahwa model memiliki performa yang moderat. Rata-rata precision, recall, dan f1-score untuk semua kelas (macro average) adalah sekitar 0.45, menunjukkan kinerja yang mendekati seimbang namun masih perlu perbaikan, terutama dalam mendeteksi pelanggaran.

4.3.2 Distribusi dan Pola dalam Hasil Clustering dengan SMOTE

Clusters of Trend Patterns in 3D



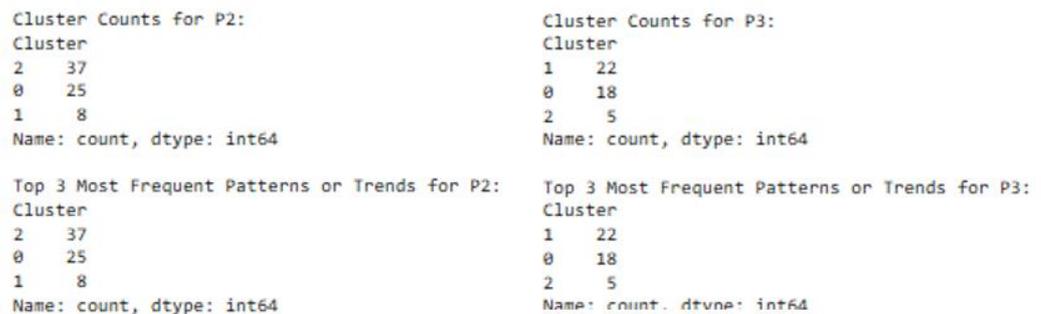
```
Cluster Counts:  
Cluster  
1    57  
0    49  
2     9  
Name: count, dtype: int64  
  
Top 3 Most Frequent Patterns or Trends:  
Cluster  
1    57  
0    49  
2     9  
Name: count, dtype: int64  
  
Top 3 Most Frequent Patterns or Trends with Labels:  
Cluster 1: Decreasing Trend (57 instances)  
Cluster 0: Increasing Trend (49 instances)
```

Gambar 4.15 Hasil analisis clustering tanpa SMOTE

Dalam analisis clustering pada gambar 4.15 untuk dataset, hasilnya menunjukkan tiga pola utama:

- **Cluster 1** (Decreasing Trend): Ini adalah pola yang paling sering ditemukan dengan 57 instance. Artinya, sebagian besar data menunjukkan tren penurunan yang konsisten.
- **Cluster 0** (Increasing Trend): Dengan 49 instance, pola ini menunjukkan bahwa data yang menunjukkan tren peningkatan juga cukup signifikan, meskipun kurang umum dibandingkan dengan tren penurunan.
- **Cluster 2** (Seasonal Fluctuations): Pola ini hanya ditemukan pada 9 instance, menunjukkan bahwa fluktuasi musiman jarang terjadi dalam dataset ini.

Secara keseluruhan, pola penurunan adalah yang paling dominan, sementara tren peningkatan dan fluktuasi musiman lebih jarang ditemukan.



```

Cluster Counts for P2:
Cluster
2    37
0    25
1     8
Name: count, dtype: int64

Cluster Counts for P3:
Cluster
1    22
0    18
2     5
Name: count, dtype: int64

Top 3 Most Frequent Patterns or Trends for P2:
Cluster
2    37
0    25
1     8
Name: count, dtype: int64

Top 3 Most Frequent Patterns or Trends for P3:
Cluster
1    22
0    18
2     5
Name: count, dtype: int64

```

Gambar 4.16 Hasil Clustering berdasarkan P2 dan P3

Pada hasil clustering P2 dan P3 sama dengan dan tanpa menggunakan SMOTE di karenakan penggolongan pada tabel yang sudah terdaftar pada dataset. Berikut adalah tabel yang merangkum jumlah cluster dan pola atau tren yang paling sering ditemukan dalam dataset P2 dan P3:

Tabel 4.3 Cluster dan pola tren paling sering untuk P2 dan P3

Golongan	Cluster	Jumlah Instance	Deskripsi Pola/Tren
P2	1	8	Tren Menurun
	0	25	Tren Meningkatkan
	2	37	Fluktuasi Musiman
P3	1	22	Tren Menurun
	0	18	Tren Meningkatkan
	2	5	Fluktuasi Musiman

Dataset P2: Dalam dataset ini, tren menurun (Cluster 1) cukup sedikit dengan 8 instance. Tren meningkat (Cluster 0) cukup sering muncul dengan 25

instance. Namun, fluktuasi musiman (Cluster 2) paling sering terjadi dengan 37 instance.

Dataset P3: Di dataset ini, tren menurun (Cluster 1) menjadi yang paling sering terjadi dengan 22 instance. Tren meningkat (Cluster 0) berada di urutan kedua dengan 18 instance. Fluktuasi musiman (Cluster 2) jarang terjadi, hanya ditemukan pada 5 instance.

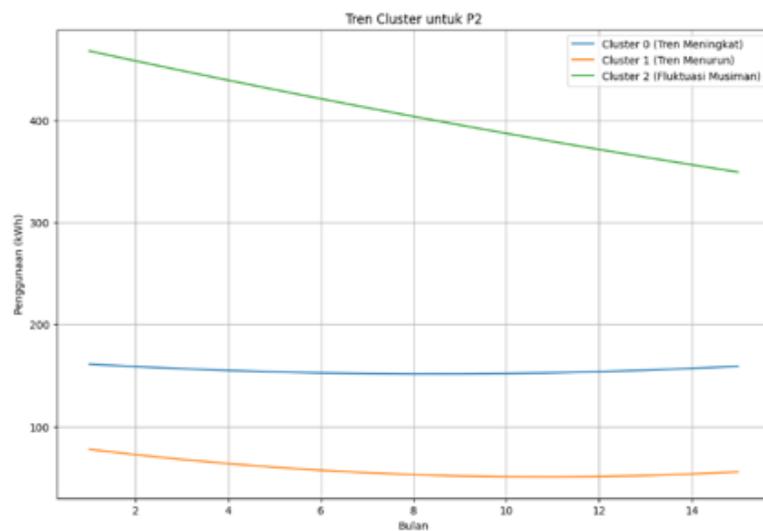
Pada dataset P2, pola tren meningkat dan fluktuasi musiman adalah pola yang paling dominan, sedangkan pola tren menurun sangat jarang ditemui.

Pada dataset P3, pola tren menurun dan meningkat adalah pola yang paling dominan, di sisi lain pola fluktuasi musiman jarang ditemukan.

4.3.3 Pola Tren dengan SMOTE

4.3.3.1 Pola Tren P2 dengan SMOTE

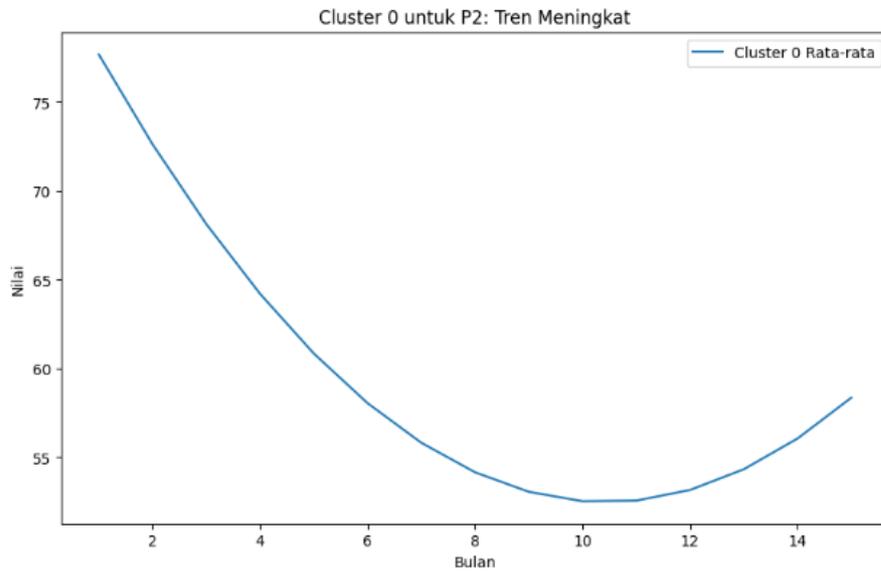
Secara keseluruhan, pola tren meningkat adalah yang paling umum di kedua dataset, sedangkan fluktuasi musiman jarang terjadi.



Gambar 4.17 Tren Cluster untuk P2

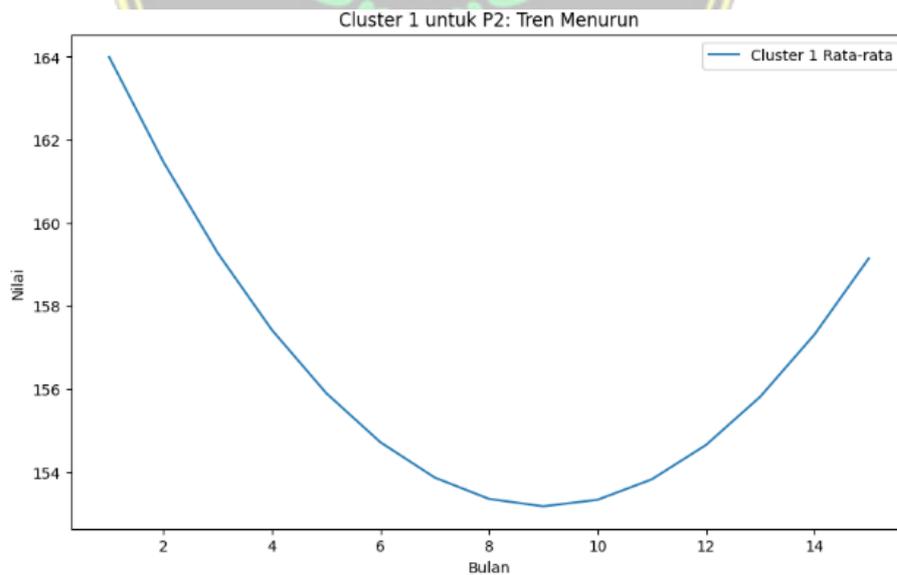
Pada tren cluster P2 Pelanggan dengan pemakaian kwh diatas 300 perbulannya terjadi penurunan dan hal ini baik. Sedangkan tren cluster P2 Pelanggan dengan pemakaian kwh ± 150 cenderung stabil dan terjadi penurunan,

kemudian tren cluster P2 Pelanggan dengan pemakaian kwh di bawah 100 perbulannya mengalami penurunan.



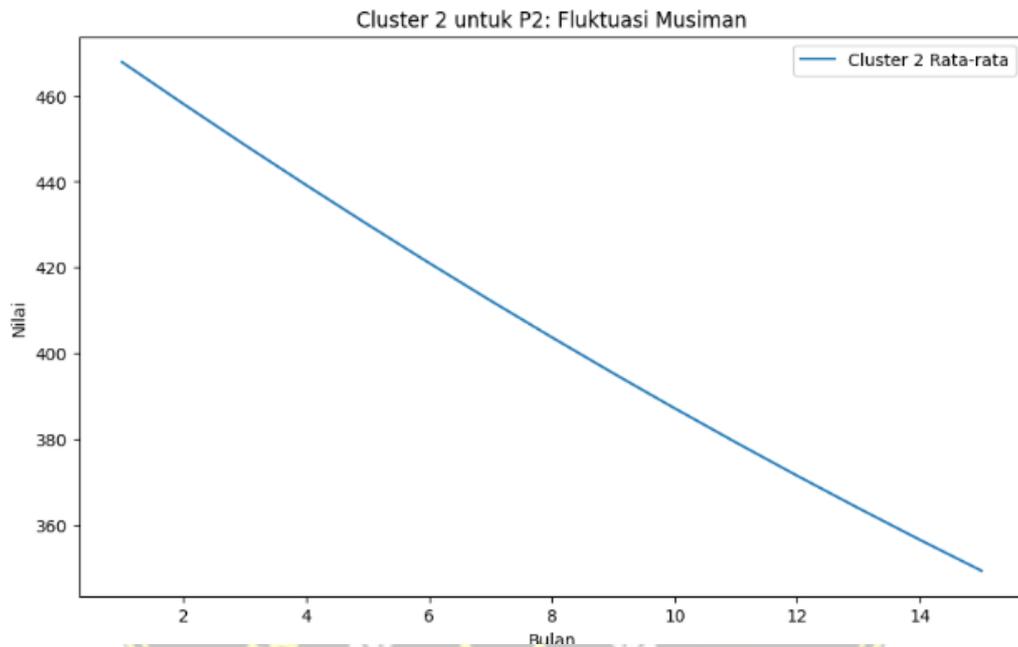
Gambar 4.18 Klaster 0 P2

Pada Gambar 4.18 sebanyak 38 pelanggan P2 cenderung melakukan kecurangan dengan penggunaan kwh listrik 80 dan rata2 tren ini cenderung tidak mengalami perubahan namun terjadi penurunan selama 8 bulan dan kemudian kenaikan pemakaian kwh.



Gambar 4.19 Klaster 1 P2

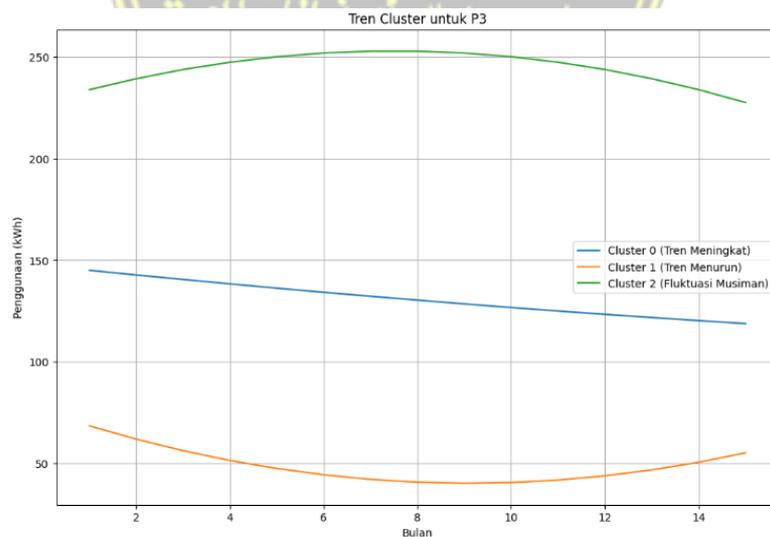
Pada Gambar 4.19 sebanyak 30 pelanggan P2 cenderung melakukan kecurangan dengan penggunaan kwh listrik 154 naik sampai 164 dan rata2 tren ini cenderung menurun setelah 4 bulan terakhir mengalami kenaikan.



Gambar 4.20 Klaster 2 P2

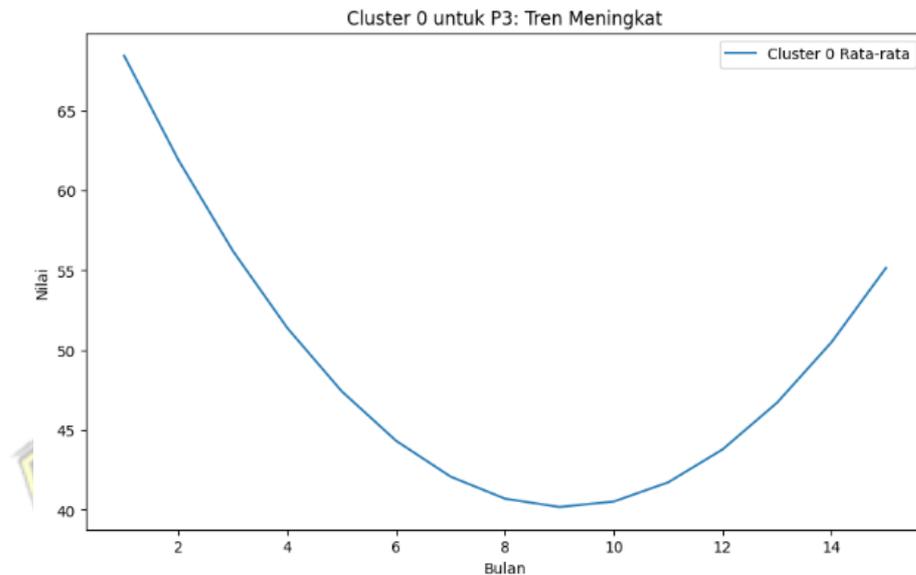
Pada Gambar 4.20 sebanyak 5 pelanggan P2 cenderung melakukan kecurangan dengan penggunaan kwh listrik 460 turun sampai 360 dan rata2 tren ini cenderung turun.

4.3.3.2 Pola tren P3 tanpa SMOTE



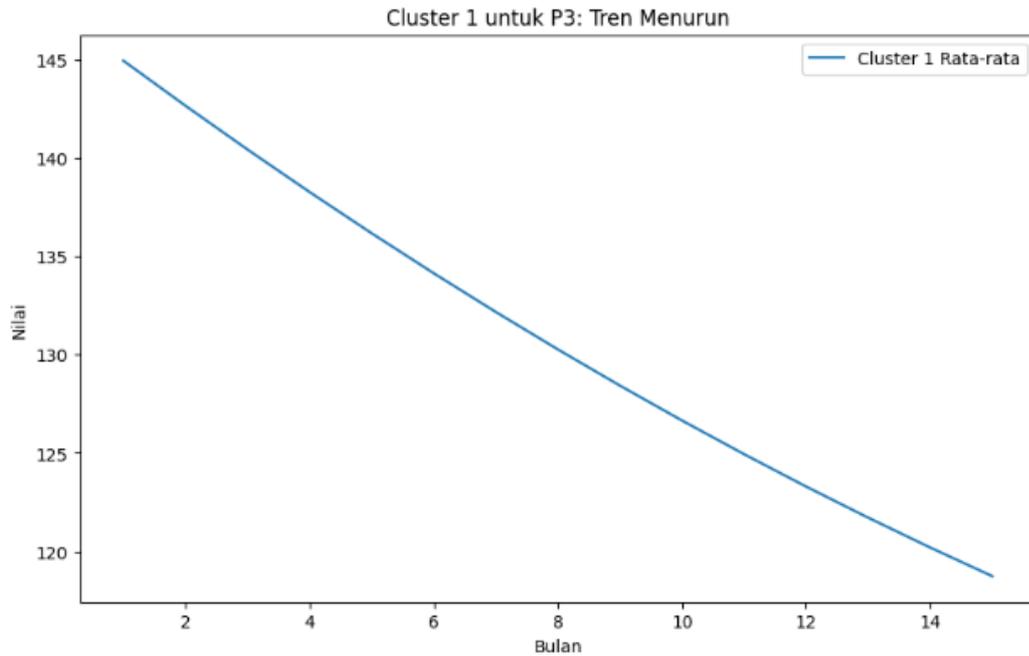
Gambar 4.21 Tren Cluster untuk P3

Pada tren cluster P3 Pelanggan dengan pemakaian kwh diatas 225 perbulannya terjadi kenaikan pemakaian kwh selama 5 bulan awal dari setelahnya mengalami penurunan. Sedangkan tren cluster P3 Pelanggan dengan pemakaian kwh 125 juga mengalami penurunan selama 15 bulan terakhir, kemudian tren cluster P3 pelanggan dengan pemakaian kwh di bawah 75 perbulannya cenderung mengalami penurunan dan setelah 4 bulan terakhir mengalami kenaikan.



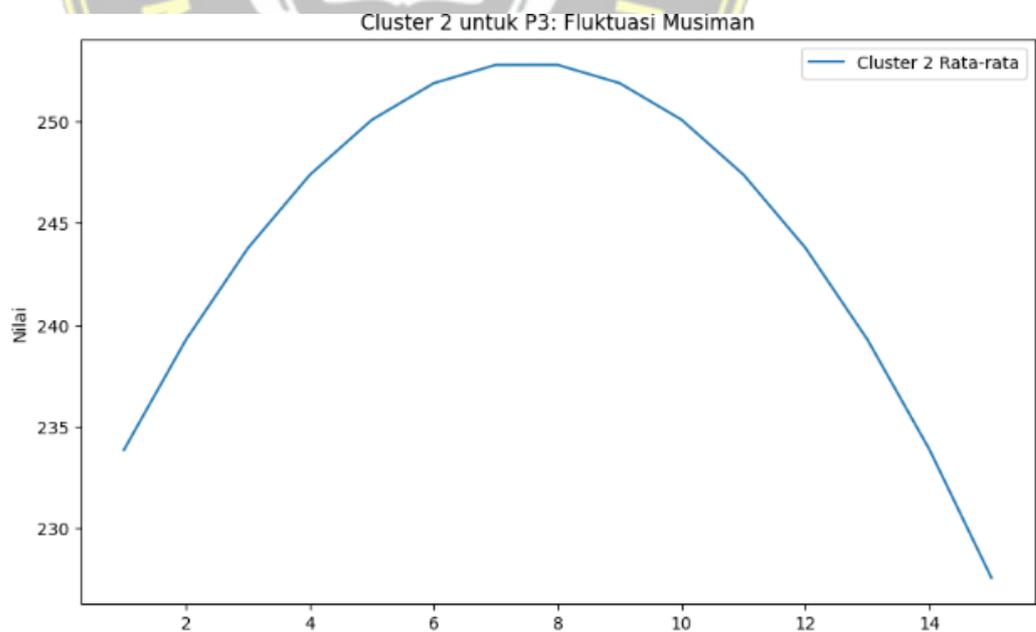
Gambar 4.22 Tren Cluster 0 untuk P3

Pada Gambar 4.22 sebanyak 22 pelanggan P3 cenderung melakukan kecurangan dengan penggunaan kwh listrik 40 naik ke 60 pada 4 bulan akhir dan rata2 tren ini cenderung turun.



Gambar 4.23 Tren Cluster 1 untuk P3

Pada Gambar 4.23 sebanyak 18 pelanggan P3 cenderung melakukan kecurangan dengan penggunaan kwh listrik 120 sampai 145 dan rata-rata tren ini cenderung turun dan ini baik.



Gambar 4.24 Tren Cluster 2 untuk P3

Pada Gambar 4.24 sebanyak 5 pelanggan P3 cenderung melakukan kecurangan dengan penggunaan kwh listrik 235 naik ke sampai 250 dan rata-rata tren ini cenderung turun selama 5 bulan terakhir.

4.4 Pembahasan

4.4.1 Kelebihan dan Kekurangan SMOTE Berdasarkan Hasil Analisis

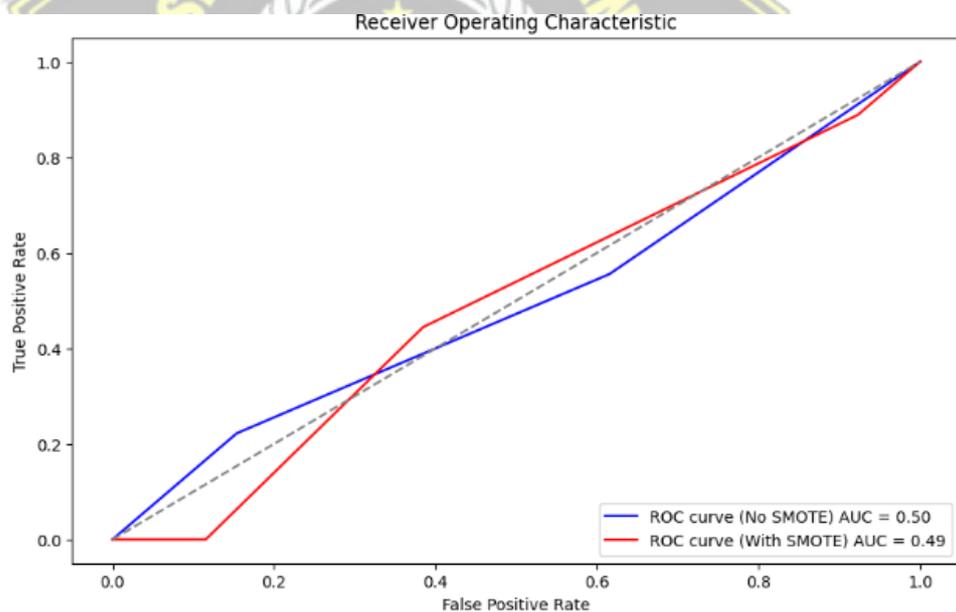
Kelebihan Menggunakan SMOTE: Penggunaan SMOTE (Synthetic Minority Over-sampling Technique) memiliki beberapa keunggulan utama dalam menangani data yang tidak seimbang. SMOTE efektif dalam menyeimbangkan dataset dengan menambahkan sampel sintetis untuk kelas minoritas, seperti yang diterapkan pada dataset P2 dan P3. Dengan data yang lebih seimbang, model KNN dapat lebih baik dalam mengenali pola konsumsi listrik pada kelas minoritas, sehingga meningkatkan sensitivitas dalam mendeteksi potensi penyalahgunaan listrik. Meskipun akurasi keseluruhan model mengalami penurunan setelah menerapkan SMOTE (dari 0.51 tanpa SMOTE menjadi 0.49 dengan SMOTE), teknik ini meningkatkan recall untuk kelas pelanggaran (kelas 1) yaitu dari 0.22 menjadi 0.44. Hal ini menunjukkan bahwa model lebih mampu mengidentifikasi lebih banyak kasus penyalahgunaan listrik, meskipun dengan konsekuensi peningkatan false positives. Selain itu, dalam analisis clustering, SMOTE menghasilkan distribusi data yang lebih merata, memungkinkan pola tren yang lebih jelas terlihat.

Kekurangan Menggunakan SMOTE: Namun, SMOTE juga memiliki beberapa keterbatasan. Salah satu kelemahan utama adalah potensi overfitting pada kelas minoritas, karena data sintetis yang dihasilkan tidak selalu mencerminkan variasi nyata dalam data asli. Selain itu, penerapan SMOTE tidak selalu meningkatkan akurasi model secara keseluruhan. Dalam analisis yang dilakukan, akurasi model dengan SMOTE justru lebih rendah dibandingkan tanpa SMOTE, yang dapat disebabkan oleh ketidakseimbangan semu yang diciptakan oleh sampel sintetis. SMOTE juga meningkatkan kompleksitas komputasi karena membutuhkan waktu tambahan untuk menghasilkan sampel sintetis dan melatih model dengan dataset yang lebih besar. Dalam kasus dataset yang sangat besar, hal ini bisa menjadi tantangan dalam hal efisiensi pemrosesan. Selain itu, meskipun recall meningkat, precision untuk kelas minoritas justru sedikit menurun, yang berarti

model lebih sering salah mengklasifikasikan pelanggan sebagai pelanggar meskipun sebenarnya bukan. Dengan demikian, meskipun SMOTE membantu dalam meningkatkan deteksi kasus penyalahgunaan listrik, penggunaan teknik ini harus dipertimbangkan dengan matang terutama terkait trade-off antara recall dan precision serta dampaknya terhadap performa model secara keseluruhan.

4.4.2 Alasan SMOTE dan Penurunan Akurasi dengan Nilai k yang Lebih Baik

Fokus pada Kelas Minoritas: SMOTE digunakan untuk memperbaiki performa model pada kelas minoritas, tetapi hasil evaluasi menunjukkan bahwa AUC tanpa SMOTE sebesar 0.5, sedangkan dengan SMOTE justru menurun menjadi 0.49. Ini mengindikasikan bahwa SMOTE tidak memberikan peningkatan dalam mendeteksi kelas minoritas dalam kasus ini. Nilai k yang lebih rendah setelah SMOTE menunjukkan bahwa model KNN membutuhkan lebih sedikit tetangga untuk mencapai performa terbaik dalam dataset yang lebih seimbang.



Gambar 4.25 Perbandingan metode SMOTE

Trade-off Precision dan Recall: Pada Gambar 4.13, kasus tanpa SMOTE menunjukkan akurasi lebih tinggi dibandingkan dengan setelah penerapan SMOTE. Biasanya, SMOTE meningkatkan recall kelas minoritas, tetapi dalam hasil ini, justru terjadi penurunan AUC. Hal ini menunjukkan bahwa penambahan sampel

sintetis mungkin menyebabkan distribusi data yang kurang representatif, sehingga model mengalami kesulitan dalam mendeteksi pola sebenarnya.

Relevansi dalam Konteks Bisnis: Dalam konteks deteksi penyalahgunaan listrik, peningkatan recall biasanya dianggap menguntungkan karena memungkinkan identifikasi lebih banyak pelanggan potensial. Namun, dalam penelitian ini, SMOTE tidak memberikan keuntungan yang diharapkan, karena tidak meningkatkan AUC dan justru dapat menyebabkan model mendeteksi lebih banyak false positives tanpa peningkatan kemampuan prediksi secara keseluruhan. Oleh karena itu, metode lain seperti tuning parameter KNN atau pendekatan balancing data yang berbeda mungkin perlu dipertimbangkan.

4.4.3 Pola Tren Pelanggan P2 dan P3 serta Kecenderungan Kecurangan

Pola Tren P2: Pelanggan dalam kelompok P2 menunjukkan beberapa pola tren utama, termasuk tren menurun, tren meningkat, dan fluktuasi musiman. Berdasarkan hasil clustering. Sebanyak 37 pelanggan P2 memiliki konsumsi listrik antara 35 hingga 65 kWh dan tren ini cenderung menurun. Sebanyak 25 pelanggan P2 dengan konsumsi listrik antara 340 hingga 240 kWh juga menunjukkan tren menurun. Sebanyak 8 pelanggan P2 memiliki konsumsi listrik antara 121 hingga 127 kWh, dengan tren yang meningkat secara signifikan selama 8 bulan terakhir.

Pola Tren P3: Pelanggan dengan konsumsi di atas 235 kWh per bulan mengalami penurunan konsumsi selama 15 bulan terakhir. Pelanggan dengan konsumsi 125 kWh juga mengalami penurunan selama 15 bulan terakhir, meskipun terjadi sedikit penurunan dalam 6 bulan pertama. Pelanggan dengan konsumsi di bawah 75 kWh per bulan cenderung mengalami penurunan konsumsi, yang dianggap sebagai hal yang baik.

Setelah penerapan SMOTE, hasil analisis menunjukkan bahwa pola tren tetap serupa dengan yang tanpa SMOTE. Namun, model mengalami sedikit penurunan performa dalam deteksi kecurangan, dengan nilai AUC yang turun dari 0.5 tanpa SMOTE menjadi 0.49 dengan SMOTE.

4.4.4 Hal yang Harus Diwaspadai Kedepan

Berdasarkan hasil evaluasi, penggunaan SMOTE tidak selalu meningkatkan performa model. ROC AUC tanpa SMOTE adalah 0.50, sedangkan dengan

SMOTE justru mengalami sedikit penurunan menjadi 0.49. Hal ini menunjukkan bahwa metode oversampling ini tidak memberikan keuntungan signifikan dalam meningkatkan deteksi kecurangan.

Kecurangan pada Pola Meningkat: Pola tren meningkat, terutama pada pelanggan dengan konsumsi listrik yang rendah, tetap perlu diawasi dengan ketat. Peningkatan konsumsi yang tidak wajar dalam cluster ini dapat menjadi indikasi penyalahgunaan listrik yang lebih canggih, terutama setelah periode penurunan yang diikuti dengan kenaikan drastis..

Fluktuasi Musiman: Walaupun jarang terjadi, pola fluktuasi musiman tetap harus diwaspadai. Jika terdapat pola konsumsi yang tidak konsisten dalam periode tertentu, bisa jadi ini merupakan indikasi manipulasi konsumsi listrik yang berbasis musim atau strategi penghindaran deteksi.



BAB V

KESIMPULAN DAN SARAN

5.1 Kesimpulan

Berdasarkan analisis dan pembahasan yang telah dilakukan, dapat diambil beberapa kesimpulan sebagai berikut :

1. Dalam penelitian ini implementasi K-Nearest Neighbors (KNN) tanpa SMOTE untuk mendeteksi penyalahgunaan listrik dari histori pemakaian pelanggan di PLN UP3 Demak lebih baik dan akurat dibandingkan dengan menggunakan SMOTE. Penggunaan SMOTE tidak selalu meningkatkan akurasi model dan performa prediksi. Meskipun SMOTE digunakan untuk menyeimbangkan data dengan sampel sintetis untuk kelas minoritas, hasil evaluasi menunjukkan bahwa ROC AUC tanpa SMOTE adalah 0.50, sedangkan dengan SMOTE justru menurun menjadi 0.49.
2. Tren konsumsi pelanggan menunjukkan pola tertentu yang perlu diawasi. P2: Sebanyak 37 pelanggan P2 memiliki konsumsi 35-65 kWh dengan tren menurun, 25 pelanggan P2 dengan konsumsi 340-240 kWh juga menunjukkan tren menurun, sedangkan 8 pelanggan P2 dengan konsumsi 121-127 kWh mengalami kenaikan yang signifikan selama 8 bulan terakhir. P3: Pelanggan dengan konsumsi di atas 235 kWh/bulan mengalami penurunan konsumsi selama 15 bulan terakhir, sedangkan pelanggan dengan konsumsi di bawah 75 kWh/bulan juga mengalami penurunan yang dianggap baik.
3. Pelanggan dengan konsumsi rendah tetapi meningkat tajam perlu diawasi lebih lanjut. 30 pelanggan mengalami kenaikan konsumsi mencurigakan setelah penurunan signifikan. Pelanggan P3 dengan konsumsi 40-70 kWh menunjukkan kenaikan setelah 4 bulan terakhir, yang dapat menjadi indikasi penyalahgunaan listrik.
4. Pola fluktuasi musiman jarang terjadi tetapi tetap perlu diperhatikan. Walaupun jumlahnya lebih sedikit, fluktuasi musiman bisa menjadi tanda adanya manipulasi konsumsi listrik berdasarkan musim atau strategi penghindaran deteksi.

5.2 Saran

1. Evaluasi kembali penggunaan SMOTE untuk balancing data. Berdasarkan hasil yang diperoleh, SMOTE tidak meningkatkan deteksi penyalahgunaan listrik dan bahkan menyebabkan penurunan ROC AUC dari 0,50 menjadi 0,49. Oleh karena itu, disarankan untuk mengeksplorasi metode balancing lain, seperti ADASYN (Adaptive Synthetic Sampling) atau cost-sensitive learning. Tuning lebih lanjut pada model KNN dan eksplorasi algoritma lain. Lakukan hyperparameter tuning pada nilai k untuk menemukan parameter yang lebih optimal. Pertimbangkan untuk menggunakan algoritma lain seperti Random Forest, XGBoost, atau Neural Networks, yang mungkin lebih efektif dalam menangani data tidak seimbang.
2. Tingkatkan pemantauan dan deteksi penyalahgunaan listrik secara real-time. Implementasikan sistem pemantauan otomatis untuk mendeteksi pola konsumsi yang mencurigakan secara lebih cepat. Fokuskan pengawasan pada pelanggan dengan konsumsi rendah yang mengalami lonjakan tiba-tiba, karena pola ini sering dikaitkan dengan praktik pencurian listrik. Analisis lebih mendalam terhadap pola konsumsi pelanggan. Perlu dilakukan studi tambahan untuk memahami lebih lanjut faktor-faktor yang menyebabkan kenaikan konsumsi mendadak. Gunakan metode clustering yang lebih canggih seperti DBSCAN atau Gaussian Mixture Model untuk mengidentifikasi pola penyalahgunaan yang lebih kompleks.

DAFTAR PUSTAKA

- [1] PLN, “Peraturan Direksi PT. PLN (Persero) Nomor.0028.P/DIR/2023 tentang Penertiban Pemakaian Tenaga Listrik.”
- [2] Presiden RI, “Undang-undang Republik Indonesia Nomor 30 tahun 2009 tentang Ketenagalistrikan.”
- [3] S. Sumarlin, “Implementasi Algoritma K-Nearest Neighbor Sebagai Pendukung Keputusan Klasifikasi Penerima Beasiswa PPA dan BBM,” *Jurnal Sistem Informasi Bisnis*, vol. 5, no. 1, pp. 52–62, Apr. 2015, doi: 10.21456/VOL5ISS1PP52-62.
- [4] M. Danny, A. Muhidin, and A. Jamal, “Application of the K-Nearest Neighbor Machine Learning Algorithm to Predict Sales of Best-Selling Products,” *Brilliance: Research of Artificial Intelligence*, vol. 4, no. 1, pp. 255–264, Jun. 2024, doi: 10.47709/brilliance.v4i1.4063.
- [5] G. A. Pradipta, R. Wardoyo, A. Musdholifah, I. N. H. Sanjaya, and M. Ismail, “SMOTE for Handling Imbalanced Data Problem: A Review,” *2021 6th International Conference on Informatics and Computing, ICIC 2021*, 2021, doi: 10.1109/ICIC54025.2021.9632912.
- [6] D. J. Won, I. Y. Chung, J. M. Kim, S. Il Moon, J. C. Seo, and J. W. Choe, “Development of power quality monitoring system with central processing scheme,” *Proceedings of the IEEE Power Engineering Society Transmission and Distribution Conference*, vol. 2, no. SUMMER, pp. 915–919, 2002, doi: 10.1109/PESS.2002.1043496.
- [7] T. Zhang, “Solving large scale linear prediction problems using stochastic gradient descent algorithms,” *Proceedings, Twenty-First International Conference on Machine Learning, ICML 2004*, pp. 919–926, 2004, doi: 10.1145/1015330.1015332.
- [8] R. N. Singarimbun, “Adaptive Moment Estimation Untuk Meminimalkan Kuadrat Error pada Algoritma Backpropagation,” 2019, Accessed: Aug. 06, 2024. [Online]. Available: <https://repositori.usu.ac.id/handle/123456789/20255>
- [9] M. S. Iqbal, M. F. A. Limon, M. M. Kabir, M. K. M. Rabby, M. J. A. Soeb, and M. F. Jubayer, “A hybrid optimization algorithm for improving load frequency control in interconnected power systems,” *Expert Syst Appl*, vol. 249, Sep. 2024, doi: 10.1016/J.ESWA.2024.123702.
- [10] M. Soori, B. Arezoo, and R. Dastres, “Optimization of energy consumption in industrial robots, a review,” *Cognitive Robotics*, vol. 3, pp. 142–157, Jan. 2023, doi: 10.1016/J.COGR.2023.05.003.
- [11] D. Granados-Lieberman, R. J. Romero-Troncoso, R. A. Osornio-Rios, A. Garcia-Perez, and E. Cabal-Yepez, “Techniques and methodologies for power quality

- analysis and disturbances classification in power systems: A review,” *IET Generation, Transmission and Distribution*, vol. 5, no. 4, pp. 519–529, Apr. 2011, doi: 10.1049/IET-GTD.2010.0466.
- [12] N. A. Enoch, P. G. George, and J. Aning, “Predicting the Remaining Lifetime of Distribution Transformers using Machine Learning,” *American Journal of Engineering and Applied Sciences*, vol. 13, no. 4, pp. 627–638, Apr. 2020, doi: 10.3844/AJEASSP.2020.627.638.
- [13] PLN, “INFORMASI P2TL.” Accessed: Jan. 25, 2024. [Online]. Available: <https://web.pln.co.id/pelanggan/informasi-p2tl#:~:text=Penertiban Pemakaian Tenaga Listrik– selanjutnya,pemakai tenaga listrik dari PLN.>
- [14] D. A. Nasution, H. H. Khotimah, and N. Chamidah, “Perbandingan Normalisasi Data untuk Klasifikasi Wine Menggunakan Algoritma K-NN,” *CESS (Journal of Computer Engineering, System and Science)*, vol. 4, no. 1, pp. 78–82, Jan. 2019, doi: 10.24114/CESS.V4I1.11458.
- [15] L. Zhang and D. Jánošík, “Enhanced short-term load forecasting with hybrid machine learning models: CatBoost and XGBoost approaches,” *Expert Syst Appl*, vol. 241, p. 122686, May 2024, doi: 10.1016/J.ESWA.2023.122686.
- [16] Y.-S. Kim, M. K. Kim, N. Fu, J. Liu, J. Wang, and J. Srebric, “Investigating the Impact of Data Normalization Methods on Predicting Electricity Consumption in a Building Using different Artificial Neural Network Models.,” *Sustain Cities Soc*, p. 105570, Jun. 2024, doi: 10.1016/J.SCS.2024.105570.
- [17] J. Taub, M. Elliot, and J. Sakshaug, “A Study of the Impact of Synthetic Data Generation Techniques on Data Utility using the 1991 UK Samples of Anonymised Records.”
- [18] S. F. Nasim, S. Aziz, A. Qaiser, U. Kulsoom, and S. Ahmed, “Fault Detection and Fault Diagnosis in Power System Using AI: A Review,” *Sir Syed University Research Journal of Engineering & Technology*, vol. 14, no. 1, pp. 27–32, Apr. 2024, doi: 10.33317/SSURJ.598.
- [19] T. Kataray *et al.*, “Integration of smart grid with renewable energy sources: Opportunities and challenges – A comprehensive review,” *Sustainable Energy Technologies and Assessments*, vol. 58, p. 103363, Aug. 2023, doi: 10.1016/J.SETA.2023.103363.
- [20] Subagyo. Pangestu, “Statistika Terapan Edisi 2011”, Edisi Kedua, Cetakan ketiga, Juli 2012 : BPFE-Yogyakarta:2012.