



**KERJA PRAKTIK - SS 234758**

**OPTIMASI KLASIFIKASI POLA KONSUMSI LISTRIK  
PELANGGAN DI JAWA BARAT BERBASIS MACHINE  
LEARNING**

**Disusun Oleh:**

Dafa Wahyu Wicaksana

NRP 5003 21 1117

Husnan Ali Husnain

NRP 5003 21 1133

**Dosen Pembimbing:**

Santi Puteri Rahayu, M.Si, Ph.D.

Shofi Andari, Ph.D.

**PROGRAM STUDI SARJANA STATISTIKA  
DEPARTEMEN STATISTIKA  
FAKULTAS SAINS DAN ANALITIKA DATA  
INSTITUT TEKNOLOGI SEPULUH NOPEMBER  
SURABAYA  
2025**





**KERJA PRAKTIK - SS 234758**

**OPTIMASI KLASIFIKASI POLA KONSUMSI LISTRIK  
PELANGGAN DI JAWA BARAT BERBASIS MACHINE  
LEARNING**

**Disusun Oleh:**

Dafa Wahyu Wicaksana  
Husnan Ali Husnain

NRP 5003 21 1117  
NRP 5003 21 1133

**Dosen Pembimbing:**

Santi Puteri Rahayu, M.Si, Ph.D.  
Shofi Andari, Ph.D.

**PROGRAM STUDI SARJANA STATISTIKA  
DEPARTEMEN STATISTIKA  
FAKULTAS SAINS DAN ANALITIKA DATA  
INSTITUT TEKNOLOGI SEPULUH NOPEMBER  
SURABAYA  
2025**

*(Halaman ini sengaja dikosongkan)*

## LEMBAR PENGESAHAN I

**LAPORAN KERJA PRAKTIK**  
**Program Studi Sarjana Statistika**  
**Departemen Statistika**  
**Fakultas Sains dan Analitika Data**  
**Institut Teknologi Sepuluh Nopember Surabaya**

**Surabaya, 22 Januari 2025**

**Menyetujui,**

**Pembimbing I**

**Pembimbing II**

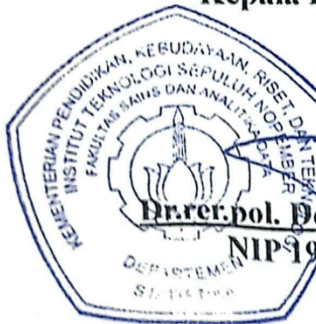


**Santi Puteri Rahayu, M.Si,**  
**Ph.D**  
**NIP 197501151999032003**



**Shofi Andari, Ph.D**  
**NIP 198712072014042001**

**Mengetahui,**  
**Kepala Departemen Statistika**  
**FSAD ITS**



  
**Dedy Dwi Prastyo, S.Si, M.Si**  
**NIP 19831204 200812 1 002**

*(Halaman ini sengaja dikosongkan)*

**LEMBAR PENGESAHAN II**

**LAPORAN KERJA PRAKTIK**  
**PT PLN ULP UP3 (Persero) Purwakarta Kota**

**Surabaya, 22 Januari 2025**

**Mengetahui,**  
**Pembimbing Lapangan**

**Dimas Arief Rahman Kurniawan**  
**NIP 9418257ZY**

*(Halaman ini sengaja dikosongkan)*



## KATA PENGANTAR

Puji syukur penulis panjatkan kehadiran Allah SWT yang telah melimpahkan rahmat serta hidayah-Nya, sehingga penulis dapat menyelesaikan laporan kerja praktik yang dilaksanakan di PT PLN UP3 (Persero) Purwakarta Kota. Laporan Kerja Praktik ini dapat terselesaikan dengan baik berkat dukungan dan bimbingan dari berbagai pihak. Oleh karena itu, penulis ingin menyampaikan terima kasih yang tulus kepada:

1. Bapak Dedy Dwi Prastyo, S.Si, M.Si selaku Kepala Departemen Statistika Fakultas Sains dan Analitika Data ITS.
2. Ibu Dr. Wibawati, S.Si, M.Si selaku Sekretaris Departemen Bidang Akademik dan Kemahasiswaan, Departemen Statistika FSAD ITS.
3. Ibu Santi Puteri Rahayu, M.Si., Ph.D, selaku dosen pembimbing I Kerja Praktik yang telah bersedia memberikan masukan kepada penulis.
4. Ibu Shofi Andari Ph.D, selaku dosen pembimbing II Kerja Praktik yang telah bersedia memberikan masukan kepada penulis.
5. Bapak/Ibu karyawan PT PLN UP3 (Persero) Purwakarta Kota, yang senantiasa menerima dan membimbing penulis selama masa Kerja Praktik berlangsung.
6. Orang tua serta keluarga penulis, yang selalu memberikan doa, motivasi, dan dukungan selama proses kerja praktik.
7. Semua pihak yang turut membantu dalam pelaksanaan kerja praktik yang tidak dapat penulis sebutkan satu per satu.

Semoga laporan ini dapat bermanfaat bagi para pembaca pada umumnya dan bagi penulis pada khususnya.

Surabaya, 22 Januari 2025  
Penulis

*(Halaman ini sengaja dikosongkan)*

## DAFTAR ISI

<b>LEMBAR PENGESAHAN I .....</b>	<b>iii</b>
<b>LEMBAR PENGESAHAN II.....</b>	<b>v</b>
<b>KATA PENGANTAR .....</b>	<b>vii</b>
<b>DAFTAR ISI .....</b>	<b>ix</b>
<b>DAFTAR GAMBAR .....</b>	<b>xi</b>
<b>DAFTAR TABEL .....</b>	<b>xii</b>
<b>DAFTAR LAMPIRAN.....</b>	<b>xiii</b>
<b>BAB I PENDAHULUAN .....</b>	<b>1</b>
1.1    Latar Belakang .....	1
1.2    Tujuan Kerja Praktik .....	2
1.2.1    Tujuan Umum.....	2
1.2.2    Tujuan Khusus.....	2
1.3    Manfaat Kerja Praktik .....	3
<b>BAB II GAMBARAN UMUM TEMPAT KERJA PRAKTIK DI PT PLN UP3 (PERSERO) PURWAKARTA KOTA .....</b>	<b>5</b>
2.1    Sejarah PT PLN (Persero) .....	5
2.2    Visi dan Misi PT PLN (Persero).....	6
2.3    Struktur Organisasi PT PLN UP3 (Persero) Purwakarta Kota .....	7
<b>BAB III PELAKSANAAN KERJA PRAKTIK.....</b>	<b>9</b>

3.1	Pelaksanaan Kerja Praktik.....	9
3.2	Langkah Analisis Tujuan Umum.....	9
3.3	Tinjauan Pustaka dan Metodologi Tujuan Khusus	9
3.3.1	Tinjauan Pustaka .....	9
3.3.2	Metodologi .....	19
<b>BAB IV HASIL KERJA PRAKTIK .....</b>		<b>25</b>
4.1	Observasi Data Pelanggan.....	25
4.2	Eksplorasi Data dan Data <i>Preprocessing</i> .....	28
4.1.1	Eksplorasi Data .....	28
4.1.2	Data <i>Preprocessing</i> .....	35
4.3	<i>Feature Selection</i> .....	42
4.4	Klasifikasi Random Forest, XGBoost, dan Logistic Regression.....	44
4.5	Evaluasi Model Terbaik.....	48
<b>BAB V KESIMPULAN DAN SARAN .....</b>		<b>51</b>
5.1	Kesimpulan.....	51
5.2	Saran.....	52
<b>DAFTAR PUSTAKA.....</b>		<b>55</b>
<b>LAMPIRAN.....</b>		<b>59</b>

## DAFTAR GAMBAR

<b>Gambar 2.1</b>	Logo PLN (Perusahaan Listrik Negara). ....	5
<b>Gambar 2.2</b>	Struktur Organisasi .....	7
<b>Gambar 3.1</b>	Diagram Alir.....	24
<b>Gambar 4.1</b>	Stan Meter Pelanggan .....	26
<b>Gambar 4.2</b>	Perhitungan Konsumsi Listrik .....	27
<b>Gambar 4.3</b>	Distribusi Kelas Normal vs Anomali .....	30
<b>Gambar 4.4</b>	Distribusi (a) Stan Meter, (b) Pemakaian Listrik, dan (c) Jam Nyala .....	32
<b>Gambar 4.5</b>	Boxplot (a) Stan Meter, (b) Pemakaian Listrik, dan (c) Jam Nyala .....	34
<b>Gambar 4.6</b>	Heatmap Korelasi Antar Variabel .....	34
<b>Gambar 4.7</b>	Boxplot (a) Stan Meter, (b) Jam Nyala, dan (c) Pemakaian Listrik .....	38
<b>Gambar 4.8</b>	Distribusi (a) Stan Meter, (b) Pemakaian Listrik, dan (c) Jam Nyala .....	40
<b>Gambar 4.9</b>	Distribusi Kelas Setelah SMOTE. ....	42
<b>Gambar 4.10</b>	Feature Importance Random Forest.....	42
<b>Gambar 4.11</b>	Kurva ROC-AUC untuk Model Random Forest.....	45
<b>Gambar 4.12</b>	Kurva ROC-AUC untuk Model XGBoost.	46
<b>Gambar 4.13</b>	Kurva ROC-AUC untuk Model Logistic Regression. ....	48
<b>Gambar 4.14</b>	Kurva ROC untuk Logistic Regression, Random Forest, dan XGBoost. ....	50

## DAFTAR TABEL

<b>Tabel 3.1</b> Struktur Data .....	20
<b>Tabel 3. 2</b> Variabel Penelitian .....	22
<b>Tabel 4. 1</b> Hasil Statistika Deskriptif .....	29
<b>Tabel 4. 2</b> Analisis Missing Value.....	36
<b>Tabel 4. 3</b> Analisis Data Duplikat .....	36
<b>Tabel 4. 4</b> Hasil Training Evaluasi Model .....	49
<b>Tabel 4. 5</b> Hasil Testing Evaluasi Model .....	49

## **DAFTAR LAMPIRAN**

<b>Lampiran 1</b>	Surat Penerimaan KP .....	59
<b>Lampiran 2</b>	Bukti Kegiatan Kerja Turun Lapangan 1 .....	60
<b>Lampiran 3</b>	Bukti Kegiatan Kerja Turun Lapangan 2 .....	61
<b>Lampiran 4</b>	Bukti Kegiatan Kerja Turun Lapang 3 .....	63





# **BAB I**

## **PENDAHULUAN**

### **1.1 Latar Belakang**

Dalam era modernisasi dan urbanisasi yang pesat, kebutuhan akan energi listrik di Indonesia, khususnya di wilayah Jawa Barat, terus meningkat. PT. PLN (Persero) sebagai perusahaan penyedia layanan kelistrikan memiliki tanggung jawab besar dalam memastikan ketersediaan energi listrik yang handal dan efisien bagi masyarakat. Salah satu tantangan utama yang dihadapi adalah pengelolaan konsumsi listrik pelanggan, terutama dalam mengidentifikasi perilaku konsumsi yang tidak wajar, yang dapat memengaruhi stabilitas sistem distribusi serta kinerja perusahaan secara keseluruhan.

Perilaku pelanggan yang tidak wajar, seperti penggunaan listrik yang melebihi batas daya atau konsumsi listrik yang tidak sesuai dengan pola normal, dapat menjadi indikator dari berbagai potensi masalah, termasuk kerugian energi, penggunaan ilegal, atau efisiensi yang rendah. Hal ini tidak hanya merugikan perusahaan dari segi finansial, tetapi juga menghambat upaya PLN dalam memberikan layanan yang optimal kepada seluruh pelanggan. Oleh karena itu, diperlukan langkah-langkah proaktif untuk mengidentifikasi dan mengelola perilaku konsumsi listrik yang tidak wajar tersebut.

Disinilah pentingnya pemanfaatan ilmu statistik dalam membantu PLN memahami pola perilaku konsumsi listrik pelanggan. Dengan menerapkan analisis klasifikasi perilaku pelanggan berbasis data, PLN dapat memperoleh wawasan yang lebih mendalam mengenai karakteristik pelanggan, sehingga memungkinkan pengambilan keputusan yang lebih cepat, tepat, dan berbasis data.

Kerja praktik yang dilakukan di PT. PLN ULP UP3 Purwakarta Kota ini bertujuan untuk mengamati dan mencatat data konsumsi listrik pelanggan melalui aktivitas lapangan serta mengolah data yang diperoleh menggunakan perangkat aplikasi

baca meter. Selain itu, kerja praktik ini juga bertujuan untuk mengaplikasikan estimasi model terbaik yang dapat mengklasifikasikan perilaku pelanggan berdasarkan pola konsumsi listrik mereka. Penelitian ini tidak hanya relevan untuk kebutuhan perusahaan dalam mendukung analisis pola konsumsi pelanggan, tetapi juga sejalan dengan visi PLN dalam meningkatkan efisiensi operasional melalui pengelolaan data yang akurat dan terstruktur. Hasil analisis diharapkan dapat memberikan rekomendasi strategis yang membantu PLN dalam mendeteksi anomali konsumsi, mengoptimalkan pelayanan, dan mendukung distribusi energi listrik yang lebih bertanggung jawab di wilayah Jawa Barat.

## **1.2 Tujuan Kerja Praktik**

Berdasarkan tujuan Pendidikan Studi Statistika Institut Teknologi Sepuluh Nopember (ITS) Surabaya yaitu membentuk sarjana yang memiliki keahlian dibidang statistika dengan kemampuan yang kreatif, inovatif, dan mandiri. Adapun tujuan dari KP ini meliputi tujuan umum dan khusus sebagai berikut.

### **1.2.1 Tujuan Umum**

Tujuan umum pelaksanaan Kerja Praktik ini adalah mengamati kegiatan statistik dan pencatatan data konsumsi listrik pelanggan melalui aktivitas lapangan, yang kemudian diolah dan dihitung menggunakan perangkat aplikasi baca meter untuk menghasilkan informasi akurat mengenai ketidakseimbangan fase listrik dan konsumsi listrik pelanggan.

### **1.2.2 Tujuan Khusus**

Tujuan khusus yang ingin dicapai dalam pelaksanaan KP ini sebagai berikut.

1. Mengevaluasi variabel signifikan yang memengaruhi konsumsi listrik pelanggan menggunakan Random Forest.
2. Mengaplikasikan serta mendapatkan model terbaik untuk klasifikasi pola konsumsi listrik pelanggan.

### **1.3 Manfaat Kerja Praktik**

Manfaat yang dihasilkan dari laporan ini adalah memberikan kontribusi signifikan bagi PT. PLN (Persero), khususnya unit layanan pelanggan UP3 Purwakarta Kota. Pengamatan dan pencatatan langsung melalui aktivitas lapangan menghasilkan data konsumsi listrik pelanggan yang akurat, yang kemudian diolah menggunakan perangkat aplikasi baca meter. Proses ini mendukung pengolahan data yang terstruktur, seperti identifikasi ketidakseimbangan fase listrik, rata-rata arus primer, dan pola konsumsi pelanggan. Selain itu, kerja praktik ini memberikan manfaat dalam mengaplikasikan model klasifikasi yang andal untuk mendeteksi anomali konsumsi listrik secara sistematis. Hasil analisis yang dihasilkan turut mendukung efisiensi operasional PLN dengan memberikan wawasan strategis untuk merancang peningkatan efisiensi distribusi listrik dan mengurangi potensi kerugian. Kerja praktik ini memberikan kesempatan untuk memperluas wawasan mengenai aplikasi ilmu statistik dalam dunia kerja nyata, meningkatkan keterampilan analisis data menggunakan perangkat lunak baca meter, serta memperdalam pemahaman tentang proses operasional di PT PLN UP3 (Persero) Purwakarta Kota. Pengalaman ini juga menjadi bekal penting untuk menghadapi tantangan di masa depan dalam bidang statistika dan analisis data. Selain itu, kerja praktik ini memberikan manfaat dalam mengevaluasi variabel signifikan yang memengaruhi pola konsumsi listrik pelanggan, serta dalam mengaplikasikan dan memahami model klasifikasi terbaik untuk menganalisis pola konsumsi tersebut. Proses ini tidak hanya memperdalam kemampuan penulis dalam pengolahan data dan penerapan metode statistik, tetapi juga mengasah kemampuan dalam menyelesaikan masalah kompleks berbasis data secara sistematis

*(Halaman ini sengaja dikosongkan)*

## **BAB II**

### **GAMBARAN UMUM TEMPAT KERJA PRAKTIK DI PT PLN UP3 (PERSERO) PURWAKARTA KOTA**

PLN merupakan salah satu perusahaan penjual jasa listrik di Indonesia. Dalam pelayanan pendistribusian kelistrikan PLN membagi-bagi fungsi unit induknya ke dalam beberapa unit induk berdasarkan pada sistem tenaga listrik yaitu pembangkitan, transmisi, dan distribusi. Selain itu ada juga unit induk atau pusat-pusat lain sebagai penunjang berlangsungnya perusahaan. Karena luasnya cakupan wilayah kerja PLN, maka PLN memiliki unit-unit di seluruh wilayah Indonesia yang mempunyai fungsi masing-masing sesuai dengan unit induknya.

#### **2.1 Sejarah PT PLN (Persero)**



**Gambar 2. 1** Logo PLN (Perusahaan Listrik Negara).

Pada akhir abad ke-19, industri pabrik gula dan ketenagalistrikan di Indonesia mulai berkembang ketika beberapa perusahaan Belanda yang bergerak di bidang pabrik gula dan teh mendirikan pembangkit listrik untuk keperluan mereka sendiri. Selama periode 1942-1945, pengelolaan perusahaan-perusahaan Belanda ini dialihkan ke Jepang setelah Belanda menyerah kepada pasukan Jepang pada awal Perang Dunia II. Setelah Jepang menyerah kepada Sekutu pada Agustus 1945, kekuasaan kembali beralih. Para pemuda dan buruh listrik memanfaatkan kesempatan

ini dengan menghadap Presiden Soekarno untuk menyerahkan perusahaan-perusahaan tersebut kepada Pemerintah Republik Indonesia. Pada 27 Oktober 1945, Presiden Soekarno membentuk Jawatan Listrik dan Gas di bawah Departemen Pekerjaan Umum dan Tenaga dengan kapasitas pembangkit listrik sebesar 157,5 MW (PLN, 2023).

Pada 1 Januari 1961, Jawatan Listrik dan Gas diubah menjadi BPU-PLN (Badan Pemimpin Umum Perusahaan Listrik Negara) yang mengelola listrik, gas, dan kokas. Namun, pada 1 Januari 1965, BPU-PLN dibubarkan dan dua perusahaan negara didirikan: Perusahaan Listrik Negara (PLN) untuk mengelola tenaga listrik dan Perusahaan Gas Negara (PGN) untuk mengelola gas. Pada tahun 1972, sesuai dengan Peraturan Pemerintah No. 18, status PLN ditetapkan sebagai Perusahaan Umum Listrik Negara dan sebagai Pemegang Kuasa Usaha Ketenagalistrikan (PKUK) dengan tugas menyediakan tenaga listrik bagi kepentingan umum (PLN, 2023). Sejak tahun 1994, status PLN berubah dari Perusahaan Umum menjadi Perusahaan Perseroan (Persero) dan tetap sebagai PKUK dalam menyediakan listrik bagi kepentingan umum (PLN, 2023).

PT. PLN UP3 Purwakarta Kota (Persero) adalah salah satu unit layanan pelanggan dari PT. PLN (Persero), yang berlokasi di Purwakarta, Jawa Barat. Perusahaan ini fokus pada pembangkitan, distribusi, dan penjualan listrik untuk memenuhi kebutuhan energi listrik masyarakat dan industri di wilayah Purwakarta dan sekitarnya. PT. PLN UP3 Purwakarta Kota (Persero) resmi didirikan pada 3 Oktober 2000 sebagai bagian dari upaya PLN untuk meningkatkan pelayanan listrik di Indonesia. Sejak saat itu, unit ini telah berkembang pesat, menambah kapasitas pembangkitan dan memperluas jangkauan layanan listriknya.

## **2.2 Visi dan Misi PT PLN (Persero)**

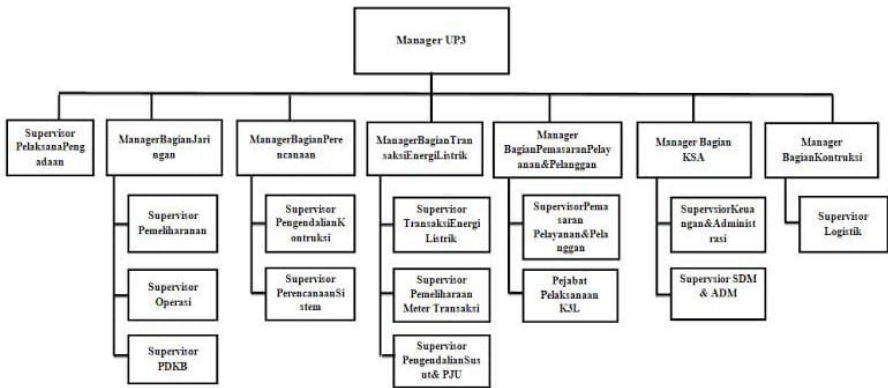
Adapun visi dan misi dari PT PLN (Persero) diantaranya dengan visi “Menjadi perusahaan listrik terkemuka se-Asia Tenggara dan menjadi pilihan utama pelanggan untuk solusi

energi.” (PLN, 2023). Misi dari PT PLN (Persero) antara lain sebagai berikut.

1. Menjalankan bisnis kelistrikan dan bidang terkait dengan berorientasi pada kepuasan pelanggan, anggota perusahaan, dan pemegang saham.
2. Menjadikan tenaga listrik sebagai media untuk meningkatkan kualitas kehidupan masyarakat.
3. Mengupayakan agar tenaga listrik menjadi pendorong kegiatan ekonomi.
4. Menjalankan kegiatan usaha yang berwawasan lingkungan.

**2.3 Struktur Organisasi PT PLN UP3 (Persero) Purwakarta Kota**

Dalam menjalankan kegiatan operasional dan pemeliharaan unit PT. PLN UP3 Purwakarta Kota (Persero), terdapat struktur organisasi yang memiliki fungsi untuk menjalankan kegiatan tersebut. Struktur organisasi pada PT. PLN UP3 Purwakarta Kota (Persero) ditampilkan pada Gambar 2.2.



**Gambar 2. 2** Struktur Organisasi

Struktur organisasi PT PLN UP3 Purwakarta Kota terdiri dari beberapa posisi utama yang saling berhubungan untuk memastikan operasional perusahaan berjalan dengan baik. Posisi tertinggi dipegang oleh Manager UP3 yang membawahi sejumlah divisi, termasuk Supervisor Pelaksana Pengadaan, Manager Bagian Jaringan, Manager Bagian Perencanaan, Manager Bagian Transaksi Energi Listrik, Manager Bagian Pemasaran, Pelayanan, dan Pelanggan, Manager Bagian KSA (Keuangan dan Administrasi), serta Manager Bagian Konstruksi. Setiap divisi memiliki supervisor yang bertanggung jawab untuk mendukung operasional spesifik, seperti Supervisor Pemeliharaan, Supervisor Operasi, Supervisor PDKB, Supervisor Pengendalian Kontruksi, Supervisor Perencanaan Sistem, Supervisor Transaksi Energi Listrik, Supervisor Pemeliharaan Meter Transaksi, Supervisor Pengendalian Sutet & PJU, Supervisor Pemasaran dan Pelayanan Pelanggan, Supervisor Keuangan dan Administrasi, Supervisor SDM & ADM, serta Supervisor Logistik. Struktur ini mencerminkan pembagian kerja yang terorganisir untuk mendukung tugas dan fungsi perusahaan.



## **BAB III**

### **PELAKSANAAN KERJA PRAKTIK**

#### **3.1 Pelaksanaan Kerja Praktik**

Pelaksanaan KP dilaksanakan pada tanggal 15 Juli 2024 hingga 15 Agustus 2024. Lokasi pelaksanaan KP adalah sebagai berikut:

Tempat : PT PLN UP3 (Persero) Purwakarta Kota  
Alamat : Jl. Singawinata, Nagri Tengah, Kec.  
Purwakarta, Kabupaten Purwakarta, Jawa  
Barat 41114

Penempatan mahasiswa kerja praktik oleh PT PLN UP3 (Persero) Purwakarta Kota yaitu pada Unit Layanan Pelanggan.

#### **3.2 Langkah Analisis Tujuan Umum**

Berikut adalah langkah analisis yang dilakukan dalam menjawab tujuan umum.

1. Melakukan pengamatan statistik dan pencatatan data konsumsi listrik pelanggan secara langsung melalui aktivitas lapangan.
2. Mendokumentasikan hasil pengukuran arus primer, arus sekunder, dan parameter lain seperti daya dan rasio CT, yang diperlukan untuk analisis ketidakseimbangan fase dan konsumsi listrik.
3. Memasukkan data hasil pengamatan ke dalam perangkat aplikasi baca meter untuk melakukan perhitungan seperti ketidakseimbangan fase listrik, rata-rata arus primer, dan total konsumsi listrik pelanggan.

#### **3.3 Tinjauan Pustaka dan Metodologi Tujuan Khusus**

Teori dan metode yang digunakan untuk penyelesaian tugas pada Kerja Praktik ini adalah sebagai berikut.

##### **3.3.1 Tinjauan Pustaka**

Tinjauan pustaka memberikan dasar teoritis yang relevan untuk mendukung pendekatan yang digunakan dalam analisis

data pada kerja praktik ini. Landasan teori yang kuat diperlukan untuk memberikan arah yang jelas dalam setiap langkah analisis, mulai dari *pre-processing* data, algoritma pembelajaran mesin, hingga evaluasi model, sehingga hasil yang diperoleh dapat dipertanggungjawabkan secara akademis maupun aplikatif. Proses analisis ini dirancang untuk mengatasi masalah klasifikasi perilaku pelanggan listrik di PT PLN UP3 Purwakarta Kota dengan menggunakan metode yang efektif dan efisien. Dengan merujuk pada teori dan penelitian sebelumnya, kerja praktik ini diharapkan tidak hanya memenuhi kebutuhan analisis data di perusahaan, tetapi juga memberikan kontribusi dalam pengembangan aplikasi *data science* di bidang layanan kelistrikan.

#### A. Random Forest

Random Forest adalah salah satu metode pembelajaran mesin yang populer dan efektif untuk tugas klasifikasi dan regresi. Metode ini terdiri dari kumpulan pohon keputusan (*decision trees*) yang bekerja secara *ensemble* untuk meningkatkan akurasi prediksi dan mengurangi risiko *overfitting*. Random Forest dapat digunakan untuk mengklasifikasikan data pelanggan listrik berdasarkan berbagai fitur yang relevan.

Random Forest bekerja dengan membangun beberapa pohon keputusan dari subset data yang berbeda dan menggabungkan hasilnya untuk membuat prediksi akhir. Setiap pohon dalam hutan acak dibangun menggunakan sampel acak dari data pelatihan, dan pada setiap *node*, pemisahan terbaik dipilih dari subset acak dari fitur. Proses ini dikenal sebagai *bagging (bootstrap aggregating)* dan membantu dalam mengurangi varians model serta meningkatkan stabilitas dan akurasi prediksi (Breiman, 2001).

Salah satu keunggulan utama dari Random Forest adalah kemampuannya untuk menangani data yang tidak seimbang dan fitur yang hilang. Selain itu, Random Forest juga menyediakan ukuran pentingnya variabel (*variable importance*), yang dapat

digunakan untuk mengidentifikasi fitur yang paling berpengaruh dalam model. Dalam penelitian yang dilakukan oleh (Fadilah, 2018), Random Forest menunjukkan kinerja yang baik dalam mengklasifikasikan data yang tidak seimbang dengan menggunakan teknik sampling seperti *Random Over Sampling* (ROS) dan *Synthetic Minority Over-sampling Technique* (SMOTE) untuk menyeimbangkan data (Fadilah, 2018).

#### B. Gradient Boosting (XGBoost)

Gradient Boosting adalah algoritma pembelajaran mesin berbasis ensemble yang dirancang untuk meningkatkan akurasi prediksi melalui pendekatan iteratif. XGBoost, atau Extreme Gradient Boosting, merupakan implementasi dari algoritma Gradient Boosting yang diperkenalkan oleh Chen dan Guestrin (2016) dengan peningkatan signifikan dalam efisiensi, fleksibilitas, dan performa komputasi. XGBoost bekerja dengan membangun model prediksi secara bertahap, di mana setiap iterasi berfokus pada memperbaiki kesalahan dari model sebelumnya. Proses ini menghasilkan model akhir yang lebih akurat karena mampu mengatasi berbagai pola data yang kompleks.

Keunggulan utama XGBoost terletak pada kemampuannya menangani dataset besar dengan variabel yang beragam. Algoritma ini menggunakan pendekatan boosting untuk menggabungkan sejumlah pohon keputusan dengan bobot yang disesuaikan berdasarkan tingkat kesalahan prediksi. Hal ini memungkinkan XGBoost untuk menangkap hubungan non-linear antar variabel dan menghasilkan model yang lebih baik dibandingkan algoritma pembelajaran mesin lainnya. Selain itu, XGBoost memiliki fitur regulasi bawaan seperti *shrinkage* dan regularisasi L1/L2 yang membantu mencegah *overfitting*, menjadikannya sangat cocok untuk analisis dengan data yang kompleks dan potensi *noise* yang tinggi.

Dalam konteks kerja praktik ini, XGBoost dipilih karena sifatnya yang sangat adaptif terhadap pola data dan kemampuannya memberikan prediksi akurat pada data pelanggan

listrik. Dengan adanya variabel yang beragam, seperti konsumsi listrik, daya, dan jam nyala, XGBoost dapat mengidentifikasi interaksi antar variabel yang mungkin tidak terdeteksi oleh algoritma yang lebih sederhana, seperti Logistic Regression. Kecepatan dan efisiensi komputasi XGBoost juga menjadi pertimbangan penting, mengingat analisis dilakukan pada dataset yang cukup besar dengan variabel prediktor yang kompleks.

Selain itu, *XGBoost* menawarkan kemampuan untuk mengevaluasi pentingnya fitur (*Feature Importance*), yang tidak hanya membantu dalam membangun model tetapi juga memberikan wawasan mendalam tentang variabel mana yang memiliki pengaruh paling signifikan terhadap hasil klasifikasi. Hal ini sangat relevan dalam analisis perilaku pelanggan listrik, di mana mengetahui variabel kunci dapat membantu PT PLN UP3 Purwakarta Kota dalam merancang strategi layanan yang lebih efektif.

Dengan keunggulan-keunggulan tersebut, XGBoost menjadi pilihan yang tepat dalam analisis ini, terutama karena algoritma ini mampu menangani hubungan non-linear, menangkap pola data yang kompleks, dan memberikan hasil yang *robust* terhadap variasi data. Penggunaannya diharapkan dapat menghasilkan model prediktif yang akurat dan relevan untuk mendukung pengambilan keputusan berbasis data di perusahaan.

#### C. Logistic Regression

Logistic Regression adalah salah satu metode statistik yang sering digunakan untuk tugas klasifikasi, terutama untuk data dengan hasil biner (dua kelas). Metode ini bekerja dengan memodelkan hubungan antara variabel independen (prediktor) dan probabilitas keluaran (output) menggunakan fungsi logit. Logistic Regression memperkirakan parameter melalui metode maksimum likelihood estimation (MLE) untuk memastikan bahwa model memberikan probabilitas terbaik terhadap kelas target berdasarkan data yang diberikan (Hosmer, Lemeshow, & Sturdivant, 2013).

Salah satu keunggulan Logistic Regression adalah kesederhanaannya dalam implementasi dan interpretasi hasil. Model ini memberikan pandangan yang jelas tentang hubungan antara setiap variabel prediktor dan kemungkinan keluaran melalui koefisien regresi. Selain itu, Logistic Regression tidak memerlukan banyak asumsi, seperti distribusi normal pada variabel prediktor, menjadikannya metode yang fleksibel untuk berbagai jenis data.

Dalam konteks kerja praktik ini, Logistic Regression digunakan sebagai baseline model. Model *baseline* bertujuan untuk memberikan gambaran awal performa klasifikasi tanpa kompleksitas tambahan yang dimiliki oleh algoritma yang lebih canggih, seperti Random Forest atau Gradient Boosting. Dengan membandingkan performa Logistic Regression dengan model lainnya, dapat dinilai sejauh mana model yang lebih kompleks memberikan peningkatan akurasi atau pemahaman yang lebih baik terhadap pola data pelanggan listrik.

Selain itu, Logistic Regression memiliki keunggulan dalam interpretasi hasil melalui probabilitas. Dalam analisis klasifikasi perilaku pelanggan listrik di PT PLN UP3 Purwakarta Kota, probabilitas yang dihasilkan Logistic Regression dapat memberikan informasi yang lebih rinci, seperti sejauh mana kemungkinan seorang pelanggan termasuk ke dalam kategori tertentu (misalnya, pelanggan yang memiliki kebiasaan tertentu dalam konsumsi listrik). Informasi ini dapat membantu perusahaan dalam merancang strategi yang lebih terarah untuk meningkatkan efisiensi layanan atau mengidentifikasi kebutuhan pelanggan.

Logistic Regression juga sangat efisien secara komputasi, menjadikannya pilihan yang ideal untuk digunakan dalam tahap awal analisis atau pada dataset dengan jumlah fitur yang relatif kecil. Dalam kerja praktik ini, Logistic Regression memberikan dasar yang kuat untuk membandingkan kinerja dengan model-model lain, seperti Random Forest dan Gradient Boosting, serta

membantu memvalidasi langkah-langkah pre-processing dan seleksi fitur yang telah dilakukan.

Dengan fleksibilitas, efisiensi, dan interpretasi hasil yang jelas, Logistic Regression menjadi komponen penting dalam rangkaian metodologi analisis. Meskipun model ini lebih sederhana dibandingkan algoritma lain, Logistic Regression tetap relevan untuk analisis awal, terutama untuk memahami struktur data dan memberikan kerangka dasar untuk membangun model yang lebih kompleks.

#### D. K-Fold Cross-Validation

K-Fold Cross-Validation adalah teknik validasi statistik yang bertujuan untuk mengukur kinerja model dengan membagi dataset menjadi  $k$  subset atau *folds* yang berukuran hampir sama. Dataset dibagi menjadi  $k$  bagian yang tidak tumpang tindih. Pada setiap iterasi, satu fold digunakan sebagai data validasi (uji), sementara  $k-1$  fold lainnya digunakan sebagai data pelatihan. Proses ini diulang sebanyak  $k$  kali, sehingga setiap fold bergantian menjadi data validasi satu kali. Hasil evaluasi model dari setiap iterasi dihitung dan dirata-rata untuk menghasilkan metrik performa yang lebih stabil dan representatif (Hastie, Tibshirani, & Friedman, 2009).

Keunggulan metode ini adalah memanfaatkan seluruh dataset untuk pelatihan dan pengujian, sehingga mengurangi kemungkinan bias akibat pembagian data secara acak. Selain itu, K-Fold Cross-Validation sangat cocok untuk dataset kecil karena memaksimalkan penggunaan data yang tersedia (James, Witten, Hastie, & Tibshirani, 2013).

Dalam laporan ini, K-Fold Cross-Validation digunakan untuk memastikan bahwa model yang dibangun memiliki kemampuan generalisasi yang baik pada pola konsumsi listrik pelanggan. Dengan menggunakan metode ini, model dapat dievaluasi secara menyeluruh pada setiap bagian dataset, sehingga performa yang dihasilkan lebih stabil dan tidak tergantung pada satu pembagian data tertentu. Pendekatan ini sangat relevan mengingat data konsumsi listrik pelanggan

memiliki pola yang beragam dan memerlukan evaluasi yang komprehensif.

E. Synthetic Minority Over-sampling Technique (SMOTE)

Synthetic Minority Over-sampling Technique (SMOTE) adalah metode pembelajaran mesin yang dirancang untuk menangani masalah ketidakseimbangan data, di mana jumlah kelas minoritas jauh lebih sedikit dibandingkan jumlah kelas mayoritas. Ketidakseimbangan data ini dapat menyebabkan algoritma pembelajaran mesin cenderung memprioritaskan kelas mayoritas, sehingga performa klasifikasi pada kelas minoritas menjadi buruk. SMOTE diperkenalkan oleh Chawla, Bowyer, Hall, & Kegelmeyer untuk meningkatkan representasi kelas minoritas melalui proses sintesis data baru berdasarkan interpolasi.

(2002)

Penggunaan SMOTE sangat relevan karena pendekatan ini membantu meningkatkan representasi kelas minoritas tanpa menambah data duplikasi, sehingga performa model dapat meningkat tanpa risiko *overfitting*. Selain itu, dengan menggunakan SMOTE, proses pembelajaran mesin menjadi lebih adil dalam mempertimbangkan pola dari kedua kelas, baik mayoritas maupun minoritas. Hal ini sangat penting dalam konteks klasifikasi perilaku pelanggan listrik, karena perusahaan perlu memahami seluruh segmen pelanggan, termasuk kelompok minoritas, untuk merancang strategi yang inklusif.

Penggunaan SMOTE dalam kerja praktik ini diharapkan dapat memberikan kontribusi signifikan pada akurasi klasifikasi data pelanggan listrik, terutama dalam memahami pola konsumsi dan anomali yang relevan untuk mendukung pengambilan keputusan berbasis data di PT PLN UP3 Purwakarta Kota.

F. Visualisasi Data

Visualisasi data adalah proses penyajian data dalam bentuk grafis atau visual untuk memudahkan pemahaman, interpretasi, dan analisis. Visualisasi data memainkan peran penting dalam mengungkap pola, tren, dan hubungan yang mungkin tidak

terlihat dalam bentuk numerik atau tekstual. Dengan menggunakan elemen visual seperti grafik, diagram, dan peta, visualisasi data membantu dalam menyederhanakan kompleksitas data dan membuat informasi lebih mudah diakses oleh berbagai pemangku kepentingan (Kurniawan, et al., 2023).

Salah satu manfaat utama dari visualisasi data adalah kemampuannya untuk menyampaikan informasi secara efektif dan efisien. Visualisasi yang baik dapat membantu dalam mengidentifikasi anomali, mengukur kinerja, dan mendukung pengambilan keputusan yang lebih baik. Dalam bidang bisnis, visualisasi data dapat digunakan untuk memantau kinerja operasional, mengidentifikasi peluang pasar, dan mengoptimalkan strategi bisnis. Selain itu, visualisasi data juga penting dalam konteks akademik dan penelitian, di mana representasi visual dari data dapat meningkatkan pemahaman dan komunikasi hasil penelitian (Knaflic, 2015)

#### G. Data Preprocessing

*Data preprocessing* adalah tahap penting dalam analisis data dan pembelajaran mesin yang bertujuan untuk mempersiapkan data mentah menjadi format yang lebih terstruktur dan siap untuk dianalisis. *Data preprocessing* melibatkan beberapa langkah utama yang dapat meningkatkan kualitas dan reliabilitas data yang digunakan. Langkah pertama adalah pembersihan data, yang mencakup identifikasi dan perbaikan atau penghapusan data yang tidak konsisten, hilang, atau duplikat. Proses ini penting untuk memastikan bahwa data yang digunakan adalah akurat dan bebas dari kesalahan yang dapat mempengaruhi hasil analisis (Hakim, 2021).

Langkah berikutnya adalah transformasi data, yang melibatkan konversi data mentah ke dalam format yang sesuai untuk analisis. Ini termasuk normalisasi, standardisasi, dan encoding data kategorikal. Normalisasi dan standardisasi membantu dalam mengurangi skala data sehingga setiap fitur memiliki kontribusi yang setara dalam analisis. Selain itu, reduksi dimensi juga merupakan bagian penting dari preprocessing, di



mana jumlah fitur dalam dataset dikurangi tanpa kehilangan informasi penting. Teknik seperti Principal Component Analysis (PCA) dan Correlation-Based Feature Selection (CFS) sering digunakan untuk mengidentifikasi dan mempertahankan fitur yang paling relevan (Gori, Sunyanto, & Fatta, 2024). Dengan melakukan preprocessing data yang tepat, kualitas dan akurasi model machine learning dapat ditingkatkan secara signifikan.

#### H. Evaluasi Model

Evaluasi model merupakan langkah penting dalam analisis data untuk memastikan kinerja algoritma yang digunakan dalam memprediksi target data. Menurut Gori, Sunyanto, & A.Fatta (2024) evaluasi model bertujuan untuk mengukur sejauh mana hasil prediksi model mendekati nilai sebenarnya berdasarkan data uji. Evaluasi ini tidak hanya memastikan keakuratan model, tetapi juga membantu membandingkan berbagai algoritma yang digunakan, sehingga dapat dipilih model terbaik yang sesuai dengan kebutuhan analisis.

Berbagai metrik evaluasi digunakan untuk menilai kinerja model klasifikasi, antara lain *confusion matrix*, akurasi, precision, recall, F1-score, *Area Under the Curve* (AUC), dan *Receiver Operating Characteristic* (ROC) curve. Metrik-metrik ini memberikan perspektif yang berbeda tentang seberapa baik model dapat memprediksi kelas target, terutama pada data yang tidak seimbang.

##### 1. Confusion Matrix

Confusion matrix adalah alat evaluasi dasar yang menggambarkan jumlah prediksi benar dan salah untuk masing-masing kelas. Menurut Provost, Fawcett, & Kohavi (1998), *matrix* ini memberikan wawasan tentang distribusi prediksi model, termasuk *True Positive* (TP), *True Negative* (TN), *False Positive* (FP), dan *False Negative* (FN). Dalam analisis pelanggan listrik di PT PLN UP3 Purwakarta Kota, *confusion matrix* digunakan untuk mengidentifikasi kesalahan model,

seperti salah prediksi pada pelanggan dengan konsumsi listrik anomali.

## 2. Akurasi

Akurasi adalah metrik yang mengukur proporsi prediksi benar terhadap total prediksi. Meskipun populer, akurasi dapat menyesatkan pada dataset yang tidak seimbang, di mana kelas mayoritas mendominasi hasil prediksi Saito & Rehmsmeier (2015). Oleh karena itu, akurasi digunakan dengan hati-hati dalam analisis ini, terutama untuk membandingkan performa baseline model seperti Logistic Regression dengan model yang lebih kompleks seperti Random Forest dan Gradient Boosting.

## 3. Precision dan Recall

Precision mengukur proporsi prediksi positif yang benar-benar positif, sedangkan recall mengukur proporsi data positif yang berhasil diprediksi dengan benar. Precision dan recall sangat penting dalam analisis ini, terutama untuk memastikan model dapat mengenali pelanggan dengan konsumsi listrik anomali (kelas minoritas) secara akurat. Dengan memadukan precision dan recall, metrik F1-score digunakan untuk memberikan keseimbangan antara keduanya.

## 4. F1-Score

F1-score adalah rata-rata harmonik dari precision dan recall, memberikan gambaran tentang kinerja model pada data yang tidak seimbang. Menurut Lipton, C. Elkan, & Naryanaswamy (2014), F1-score sangat relevan untuk kasus di mana kesalahan pada kelas minoritas memiliki dampak signifikan. Dalam analisis ini, F1-score membantu menilai model yang mampu mengenali pola konsumsi listrik pelanggan dengan baik.

## 5. AUC dan ROC Curve

AUC (*Area Under the Curve*) dan ROC (*Receiver Operating Characteristic*) curve digunakan untuk mengevaluasi

kemampuan model dalam membedakan antara kelas positif dan negatif. *ROC curve* menunjukkan *trade-off* antara *True Positive Rate* (TPR) dan *False Positive Rate* (FPR). AUC memberikan nilai agregat dari performa model pada semua tingkat *threshold*. Menurut Fawcett (2006), AUC adalah metrik yang kuat untuk menilai model pada dataset yang tidak seimbang, seperti analisis pelanggan listrik.

Dalam konteks kerja praktik ini, evaluasi model dilakukan untuk memastikan bahwa model tidak hanya akurat pada kelas mayoritas tetapi juga dapat menangkap pola dari kelas minoritas yang memiliki jumlah lebih sedikit. Penggunaan metrik seperti F1-score dan AUC membantu menghindari bias terhadap kelas mayoritas, yang sering terjadi pada dataset tidak seimbang.

Evaluasi ini juga digunakan untuk membandingkan algoritma yang digunakan, seperti Logistic Regression, Random Forest, dan Gradient Boosting. Dengan membandingkan hasil evaluasi, model terbaik dapat dipilih berdasarkan keseimbangan antara akurasi, interpretasi, dan kemampuan generalisasi.

Penggunaan metode evaluasi yang komprehensif dalam kerja praktik ini diharapkan dapat memberikan pemahaman yang mendalam tentang performa model yang digunakan, serta membantu PT PLN UP3 Purwakarta Kota dalam mengambil keputusan berbasis data yang lebih akurat dan andal.

### **3.3.2 Metodologi**

Metodologi dalam kerja praktik ini dirancang untuk memastikan bahwa seluruh proses analisis data dilakukan secara terstruktur dan sesuai dengan tujuan yang telah ditetapkan.

#### **A. Sumber Data**

Data yang digunakan merupakan data pemakaian listrik pelanggan PLN di Jawa Barat pada bulan Januari hingga Juni 2024 yang diperoleh dari PT PLN UP3 (Persero) Purwakarta Kota. Data ini mencakup informasi terkait konsumsi listrik

pelanggan dalam kurun waktu tersebut, meliputi variabel-variabel yang relevan untuk analisis, seperti jumlah konsumsi listrik, daya kontrak pelanggan, dan pola waktu penggunaan. Tabel 3.1 berikut menunjukkan struktur data yang digunakan dalam analisis ini, mencakup nama variabel, deskripsi, dan tipe data.

**Tabel 3.1** Struktur Data

<b>No</b>	<b>X<sub>1</sub></b>	<b>X<sub>2</sub></b>	<b>X<sub>3</sub></b>	<b>X<sub>4</sub></b>	<b>X<sub>5</sub></b>	<b>Y<sub>1</sub></b>
1	X <sub>1,1</sub>	X <sub>1,2</sub>	X <sub>1,3</sub>	X <sub>1,4</sub>	X <sub>1,5</sub>	Y <sub>1,1</sub>
2	X <sub>2,1</sub>	X <sub>2,2</sub>	X <sub>2,3</sub>	X <sub>2,4</sub>	X <sub>2,5</sub>	Y <sub>2,1</sub>
3	X <sub>3,1</sub>	X <sub>3,2</sub>	X <sub>3,3</sub>	X <sub>3,4</sub>	X <sub>3,5</sub>	Y <sub>3,1</sub>
...	...	...	...	...	...	...
n	X <sub>n,1</sub>	X <sub>n,2</sub>	X <sub>n,3</sub>	X <sub>n,4</sub>	X <sub>n,5</sub>	Y <sub>n,1</sub>

#### Keterangan

n = Jumlah observasi

X<sub>n,m</sub> = Nilai variabel X untuk pengamatan ke-n dan fitur ke-m

Y<sub>n</sub> = Hasil pengamatan atau output pada observasi ke-n.

#### B. Variabel

Analisis ini melibatkan sejumlah variabel yang terdiri dari variabel dependen (target) dan variabel independen (prediktor). Variabel dependen yang digunakan adalah Y, yang merepresentasikan klasifikasi perilaku pelanggan, seperti pelanggan dengan konsumsi listrik normal atau anomali. Dimana Indikasi untuk mengelompokkan pelanggan ke dalam kelas 0 (Normal) atau kelas 1 (Anomali) didasarkan pada beberapa kriteria yang relevan dengan pola konsumsi listrik. Pelanggan masuk ke dalam kelas 0 jika konsumsi listriknya stabil dan berada dalam batas wajar sesuai daya kontrak, tanpa fluktuasi besar antarperiode. Ketidakseimbangan fasa pada arus primer (Ip) antar

fasa R, S, dan T juga rendah, dengan persentase ketidakseimbangan (%R, %S, %T) masih dalam toleransi tertentu. Selain itu, alat Stan Meter berfungsi dengan baik, dan pola konsumsi pelanggan konsisten serta sesuai dengan penggunaan normal. Sementara itu, pelanggan dikategorikan ke dalam kelas 1 jika menunjukkan konsumsi listrik yang tidak wajar, seperti konsumsi yang jauh lebih tinggi atau lebih rendah dari kapasitas daya kontrak. Ketidakseimbangan fasa yang signifikan, dengan nilai persentase ketidakseimbangan melebihi ambang batas toleransi, juga menjadi indikator utama. Fluktuasi konsumsi listrik yang tidak konsisten, gangguan atau kejanggalan pada pembacaan Stan Meter, serta indikasi penggunaan listrik ilegal atau masalah teknis pada alat meter, dapat menjadi alasan lain untuk memasukkan pelanggan ke dalam kelas 1. Pengelompokan ini bertujuan untuk mendeteksi pelanggan yang berpotensi menyebabkan gangguan sistem atau memerlukan evaluasi lebih lanjut, sehingga mendukung efisiensi pengelolaan energi listrik. Sementara itu, variabel independen terdiri dari lima variabel utama, yaitu  $X_1$  (Daya),  $X_2$  (Stan Meter),  $X_3$  (Jam Nyala),  $X_4$  (Pemakaian Listrik), dan  $X_5$  (Tarif). Variabel-variabel ini dipilih karena memberikan kontribusi penting dalam membangun model klasifikasi untuk menganalisis pola konsumsi listrik pelanggan. Model klasifikasi yang didasarkan pada variabel-variabel ini memungkinkan deteksi anomali konsumsi dengan akurasi lebih tinggi, sehingga mempermudah identifikasi pelanggan yang memiliki pola pemakaian tidak wajar. Deteksi dini terhadap anomali ini dapat membantu perusahaan dalam mengambil tindakan proaktif, seperti inspeksi lapangan atau evaluasi alat meter, untuk mengurangi potensi gangguan pada sistem distribusi listrik. Selain itu, pemahaman mendalam terhadap pola konsumsi juga mendukung perencanaan strategis dalam pengelolaan energi listrik, termasuk optimalisasi distribusi

daya dan perbaikan infrastruktur. Deskripsi rinci dari masing-masing variabel disajikan dalam tabel berikut.

**Tabel 3. 2** Variabel Penelitian

<b>Variabel</b>	<b>Nama</b>	<b>Deskripsi</b>
X <sub>1</sub>	Daya	Kapasitas daya pelanggan dalam satuan Volt-Ampere (VA)
X <sub>2</sub>	Stan Meter	Catatan kumulatif pelanggan dalam kilowatt/jam (kWh)
X <sub>3</sub>	Jam Nyala	Total jam nyala listrik pelanggan dalam jam
X <sub>4</sub>	Pemakaian Listrik	Total pemakaian listrik pelanggan dalam kilowatt/jam (kWh)
X <sub>5</sub>	Tarif	Jenis tarif pelanggan berdasarkan kategori
Y	Klasifikasi	Klasifikasi perilaku pelanggan (0: normal, 1: anomali)

### C. Langkah Analisis Tujuan Khusus

Adapun langkah-langkah dalam analisis klasifikasi data pelanggan listrik menggunakan bantuan software *Python* untuk menjawab tujuan khusus.

#### a. Evaluasi Variabel

Evaluasi variabel dilakukan untuk menentukan faktor-faktor yang signifikan memengaruhi pola konsumsi listrik pelanggan. Berikut adalah langkah untuk mengevaluasi variabel signifikan.

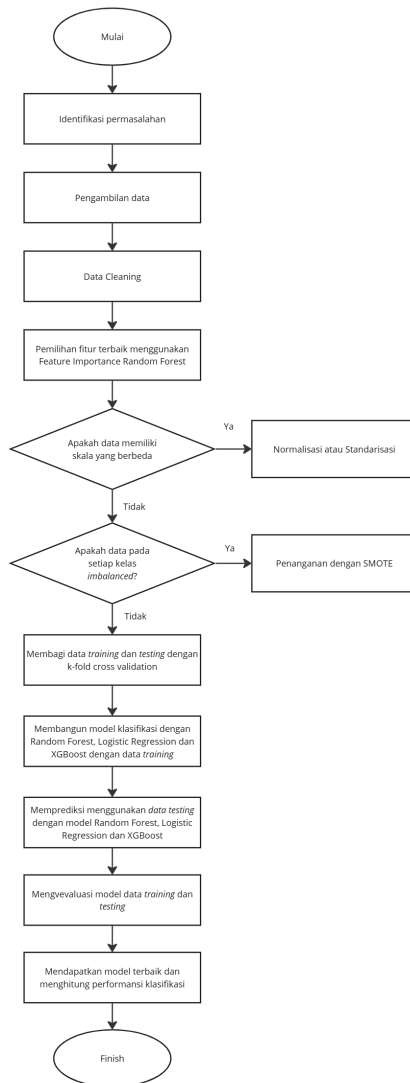
1. Identifikasi permasalahan perilaku pelanggan listrik.
2. Pengambilan data pemakaian listrik pelanggan dari sistem informasi PT PLN UP3 Purwakarta Kota.
3. *Pre-processing* data (pembersihan data dari *missing value* dan *duplicate value*).
4. Eksplorasi data untuk memahami karakteristik variabel penelitian, seperti distribusi, tren, dan pola data.
5. Pemilihan fitur terbaik pada variabel prediktor menggunakan *Feature Importance Random Forest*.

#### b. Mencari Model Terbaik

Tahap ini bertujuan untuk mengidentifikasi model terbaik dalam mengklasifikasikan pola konsumsi listrik pelanggan.

1. Identifikasi data yang sudah di evaluasi variabel
2. Penanganan data *imbalanced* dengan SMOTE
3. Pembagian data *training* dan *testing*.
4. Membangun model *Random Forest*, *Logistic Regression*, dan *Gradien Boosting (XGBoost)* untuk klasifikasi perilaku pelanggan berdasarkan fitur terpilih.
5. Mengevaluasi kebaikan model dengan menggunakan confusion matrix, skor akurasi, precision, recall, F1-score, AUC, dan kurva ROC.

Diagram alir dari langkah analisis yang dirancang untuk menjawab tujuan khusus ditunjukkan pada Gambar 3.1.



mini

**Gambar 3.1** Diagram Alir.





























































## **BAB V**

### **KESIMPULAN DAN SARAN**

#### **5.1 Kesimpulan**

1. Observasi data pelanggan dilakukan melalui aktivitas lapangan, termasuk pengumpulan data dari alat Stan Meter di lokasi pelanggan. Data yang diperoleh memberikan informasi aktual terkait konsumsi listrik, yang kemudian diverifikasi menggunakan sistem informasi PLN. Observasi ini juga mengidentifikasi potensi ketidakseimbangan fasa dan kondisi alat yang memerlukan pemeliharaan.
2. Evaluasi variabel signifikan yang memengaruhi klasifikasi konsumsi listrik pelanggan menunjukkan bahwa variabel Stan Meter, Jam Nyala, dan Pemakaian Listrik memiliki pengaruh terbesar terhadap hasil klasifikasi. Stan Meter, sebagai catatan kumulatif konsumsi listrik pelanggan, memberikan kontribusi tertinggi dalam analisis dengan skor Feature Importance di atas 0.6. Variabel Jam Nyala mencerminkan durasi penggunaan listrik pelanggan, sementara Pemakaian Listrik menunjukkan total konsumsi listrik dalam periode tertentu. Kedua variabel ini juga memberikan pengaruh signifikan terhadap klasifikasi. Sebaliknya, variabel seperti Tarif dan Daya menunjukkan pengaruh yang lebih rendah, sehingga dihapus untuk meningkatkan efisiensi model. Hasil evaluasi ini memberikan wawasan penting bagi PLN untuk fokus pada variabel-variabel utama dalam mengelola konsumsi listrik pelanggan dan mendeteksi anomali secara lebih efektif.
3. Model analitik berbasis machine learning untuk klasifikasi pola konsumsi listrik pelanggan berhasil dikembangkan menggunakan Random Forest, XGBoost, dan Logistic

Regression. Model Random Forest memberikan performa terbaik dengan akurasi sebesar 92.2% dan AUC 0.985, menunjukkan kemampuannya dalam membedakan pola konsumsi listrik normal dan tidak wajar. Proses pengembangan model melibatkan langkah-langkah penting, seperti preprocessing data dengan metode SMOTE untuk menangani ketidakseimbangan data, seleksi fitur menggunakan Feature Importance, serta evaluasi model berdasarkan metrik akurasi, precision, recall, dan F1-score. Hasil pengembangan ini memberikan dasar yang kuat untuk mendukung deteksi otomatis perilaku konsumsi listrik pelanggan secara lebih efisien dan akurat.

## **5.2 Saran**

Berdasarkan hasil optimasi klasifikasi pola konsumsi listrik pelanggan, berikut beberapa saran yang dapat diterapkan oleh pihak PLN untuk mendukung pengelolaan dan pengambilan keputusan yang lebih efektif.

1. PLN disarankan untuk mengembangkan sistem berbasis machine learning yang mampu mendeteksi secara real-time konsumsi listrik pelanggan yang tidak wajar. Model Random Forest yang terbukti memiliki performa terbaik dapat diintegrasikan ke dalam sistem ini untuk memberikan deteksi dini terhadap anomali konsumsi listrik.
2. Dengan hasil analisis yang menunjukkan bahwa variabel Stan Meter, Jam Nyala, dan Pemakaian Listrik adalah faktor utama yang memengaruhi klasifikasi, PLN dapat memprioritaskan pengumpulan dan analisis lebih lanjut terhadap data ini. Hal ini dapat membantu dalam optimalisasi monitoring dan pengambilan keputusan terkait efisiensi konsumsi listrik.

3. PLN disarankan untuk melakukan evaluasi berkala terhadap model yang digunakan, seperti Random Forest, untuk memastikan keandalannya dalam berbagai kondisi data. Dengan pembaruan model setiap enam bulan atau berdasarkan kebutuhan, akurasi prediksi dan performa klasifikasi dapat terus terjaga.

diharapkan PLN dapat meningkatkan efisiensi dalam mengelola konsumsi listrik pelanggan, mendeteksi perilaku anomali secara lebih cepat, serta memberikan layanan yang lebih baik dan efektif bagi pelanggan.

*(Halaman ini sengaja dikosongkan)*

## DAFTAR PUSTAKA

- Breiman, L. (2001, October). Machine Learning. *Random Forests*, 45, 5-32.
- Chawla, N., Bowyer, K., Hall, L., & Kegelmeyer, W. (2002). SMOTE: Synthetic Minority Over-sampling Technique. *Journal of Artificial Intelligence Research*, 16, 321-357.  
doi:<https://doi.org/10.1613/jair.953>
- Chen, T., & Guestrin, C. (2016). XGBoost: A scalable tree boosting system. Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining,, (pp. 785-794).  
doi:<https://doi.org/10.1145/2939672.2939785>.
- Dwiyanto, A. (2023). *Statistika Deskriptif: Pengertian, Fungsi, dan Jenisnya*.
- Fadilah, L. (2018). Klasifikasi Random Forest Pada Data Imbalanced. *Program Studi Matematika Fakultas Sains Dan Teknologi Universitas Islam Negeri Syarif Hidayatullah*.
- Fawcett, T. (2006). An introduction to ROC analysis. *Pattern Recognition Letters*, 861-874.
- Firdaus, Z. Z. (2012). Pengaruh Unit Produksi, Parkerin dan Dukungan Keluarga Terhadap Kesiapan Kerja Siswa SMK. 2.  
doi:<http://dx.doi.org/10.21831/jpv.v2i3.1045>
- Gori, M., Sunyanto, H., & A.Fatta. (2024). Advanced Evaluation Metrics for Imbalanced Datasets. *Journal of Data Science and Applications*, 45-60.

- Gori, T., Sunyanto, A., & Fatta, H. A. (2024). PREPROCESSING DATA DAN KLASIFIKASI UNTUK PREDIKSI KINERJA AKADEMIK SISWA. *Jurnal Teknologi Informasi dan Ilmu Komputer (JTIK)*, 11. doi:10.25126/jtiik.20241118074
- Hakim, B. (2021). ANALISA SENTIMEN DATA TEXT PREPROCESSING PADA DATA MINING DENGAN MENGGUNAKAN MACHINE LEARNING. *Journal of Business and Audit Information Systems*, 4. doi:http://dx.doi.org/10.30813/jbase.v4i2.3000
- Hastie, T., Tibshirani, R., & Friedman, J. (2009). *The Elements of Statistical Learning*. Springer.
- Hosmer, D., Lemeshow, S., & Sturdivant, R. (2013). *Applied Logistic Regression* (3rd ed.). John Wiley. doi:https://doi.org/10.1002/9781118548387
- James, G., Witten, D., Hastie, T., & Tibshirani, R. (2013). *An Introduction to Statistical Learning*. Springer.
- Knaflic, C. N. (2015). *Storytelling with Data : A Data Visualization Guide for Business Professionals*.
- Kurniawan, J., Hartoto, Ahyani, A. Z., Hikmah, Ridwan, M., Amane, A. P., . . . Hozairi. (2023). *ANALISIS DAN VISUALISASI DATA*. (E. Damayanti, Ed.) Bandung, Indonesia: Widina Bhakti Persada Bandung.
- Lipton, Z. C., Elkan, C., & Naryanaswamy, B. (2014). Thresholding Classifiers to Maximize F1 Score.
- Martias, L. D. (2021, Januari - Juni). STATISTIKA DESKRIPTIF SEBAGAI KUMPULAN



- INFORMASI. *FIHRIS: Jurnal Ilmu Perpustakaan dan Informasi*, 16. doi:<https://doi.org/10.14421/fhrs.2021.161.40-59>
- PLN. (2023). *Profil Perusahaan PT PLN UP3 Purwakarta Kota*. Retrieved from <https://web.pln.co.id/tentang-kami/profil-perusahaan>
- Provost, F., Fawcett, T., & Kohavi, R. (1998). The Case Against Accuracy Estimation for Comparing Induction Algorithms. *Proceedings of the Fifteenth International Conference on Machine Learning*, 445–453.
- Saito, T., & Rehmsmeier, M. (2015). The Precision-Recall Plot Is More Informative than the ROC Plot When Evaluating Binary Classifiers on Imbalanced Datasets. *PLoS ONE*.
- Statistika, D. (2023). *BUKU PANDUAN AKADEMIK SARJANA STATISTIKA 2018-2023*. Surabaya: Departemen Statistika ITS. Retrieved from [https://www.its.ac.id/statistika/wp-content/uploads/sites/43/2020/08/7\\_Agust\\_Buku-Panduan-Akademik-Jurusan-Statistika-2019-REVISI-1.pdf](https://www.its.ac.id/statistika/wp-content/uploads/sites/43/2020/08/7_Agust_Buku-Panduan-Akademik-Jurusan-Statistika-2019-REVISI-1.pdf)
- Wardani, F. S., I., C. D., & Susantiningrum. (2019, November). Analisis Pelaksanaan Prarktik Kerja Lapangan Dalam Mempersiapkan Siswa Memasuki Dunia Kerja Di Era Revolusi Industri 4.0. *Studi Kasus pada Siswa Kelas XI Kompetensi*, 3. Retrieved from <http://jurnal.fkip.uns.ac.id/index/php/jikap>

*(Halaman ini sengaja dikosongkan)*











