

SKRIPSI

**ANALISIS KONSUMSI LISTRIK ABNORMAL
MENGUNAKAN ALGORITMA *FUZZY C-MEANS* DAN
*DECISION TREE***

Disusun dan diajukan oleh:

**SILA FARSIDIA PUTRI
D121 191 045**



**PROGRAM STUDI SARJANA TEKNIK INFORMATIKA
FAKULTAS TEKNIK
UNIVERSITAS HASANUDDIN
GOWA
2024**

LEMBAR PENGESAHAN SKRIPSI

ANALISIS KONSUMSI LISTRIK ABNORMAL MENGGUNAKAN ALGORITMA FUZZY C-MEANS DAN DECISION TREE

Disusun dan diajukan oleh

Sila Farsidia Putri
D121191045

Telah dipertahankan di hadapan Panitia Ujian yang dibentuk dalam rangka Penyelesaian Studi Program Sarjana Program Studi Teknik Informatika Fakultas Teknik Universitas Hasanuddin Pada tanggal 27 Juni 2024 dan dinyatakan telah memenuhi syarat kelulusan

Menyetujui,

Pembimbing Utama,

Dr. Ir. Ingrid Nurtanio, M.T
NIP 196108131988112001

Pembimbing Pendamping,

Ir. Christoforus Yohannes, M.T
NIP 196007161987021002

Ketua Program Studi,



Prof. Dr. Ir. Indrabayu, ST., MT., M.Bus.Sys., IPM, ASEAN. Eng.
NIP 197507162002121004



PERNYATAAN KEASLIAN

Yang bertanda tangan dibawah ini ;

Nama : Sila Farsidia Putri
NIM : D121191045
Program Studi : Teknik Informatika
Jenjang : S1

Menyatakan dengan ini bahwa karya tulisan saya berjudul

Analisis Konsumsi Listrik Abnormal Menggunakan Algoritma Fuzzy C-Means dan Decision Tree

Adalah karya tulisan saya sendiri dan bukan merupakan pengambilan alihan tulisan orang lain dan bahwa skripsi yang saya tulis ini benar-benar merupakan hasil karya saya sendiri.

Semua informasi yang ditulis dalam skripsi yang berasal dari penulis lain telah diberi penghargaan, yakni dengan mengutip sumber dan tahun penerbitannya. Oleh karena itu semua tulisan dalam skripsi ini sepenuhnya menjadi tanggung jawab penulis. Apabila ada pihak manapun yang merasa ada kesamaan judul dan atau hasil temuan dalam skripsi ini, maka penulis siap untuk diklarifikasi dan mempertanggungjawabkan segala resiko.

Segala data dan informasi yang diperoleh selama proses pembuatan skripsi, yang akan dipublikasi oleh Penulis di masa depan harus mendapat persetujuan dari Dosen Pembimbing.

Apabila dikemudian hari terbukti atau dapat dibuktikan bahwa sebagian atau keseluruhan isi skripsi ini hasil karya orang lain, maka saya bersedia menerima sanksi atas perbuatan tersebut.

Sudiang, 30 Juni 2024

Yang Menyatakan



Sila Farsidia Putri



ABSTRAK

SILA FARSIDIA PUTRI. ANALISIS KONSUMSI LISTRIK ABNORMAL MENGGUNAKAN ALGORITMA FUZZY C-MEANS DAN DECISION TREE (dibimbing oleh Ingrid Nurtanio dan Christoforus Yohannes)

Energi listrik memiliki peran yang penting dalam mendukung aktivitas sehari-hari, khususnya di wilayah industri dan perumahan yang sangat bergantung pada ketersediaan energi ini. Di Indonesia, PT. PLN (Persero) sebagai BUMN penyedia energi listrik menghadapi tantangan dalam distribusi energi, termasuk konsumsi energi oleh pelanggan yang abnormal yang berpotensi merugikan instansi maupun pelanggan. Penelitian ini bertujuan untuk mengimplementasi penggunaan algoritma *Fuzzy C-Means* (FCM) dan menganalisis *Decision Tree* dalam mengidentifikasi dan mengklasifikasi karakteristik konsumsi energi listrik. Algoritma FCM digunakan untuk mengelompokkan karakteristik konsumsi energi listrik ke dalam kluster dan menggunakan algoritma *Decision Tree* untuk mengklasifikasikan karakteristik penggunaan listrik pelanggan. Data yang digunakan dalam penelitian ini adalah data penggunaan listrik pelanggan. Hasil *clustering* menggunakan FCM menghasilkan distribusi data ke dalam empat *cluster* yang berbeda dengan karakteristik dan rekomendasi spesifik. Temuan penting dari penelitian ini adalah *cluster* 4, mencakup hanya 2,5% dari dataset, yang menunjukkan kondisi tidak terbacanya arus yang mengalir dan rendahnya efisiensi penggunaan listrik. Sebanyak 27,8% dari dataset dikategorikan ke dalam *cluster* 1 yang dicirikan oleh tingginya ketidakseimbangan arus. *Cluster* 2, yang mencakup 8,7% dari dataset, mengalami masalah serupa dengan *cluster* 1 ditambah dengan rendahnya efisiensi penggunaan listrik. Sementara itu, mayoritas data berada pada *cluster* 3, yakni 60,8% yang menunjukkan kondisi normal. Sehingga perlu diadakan tindak lanjut untuk *cluster* 1, 2 dan 4. Hasil model klasifikasi menggunakan algoritma *decision tree* menunjukkan nilai akurasi 99,84% dengan rasio 60:40 dan 70:30. Berdasarkan penelitian yang telah dilakukan, algoritma FCM dapat mengelompokkan karakteristik penggunaan listrik dengan baik dan model klasifikasi yang dihasilkan dari algoritma *decision tree* menghasilkan akurasi yang baik pula.

Kata Kunci: *Fuzzy C-Means*, *Clustering*, *Decision Tree*, Akurasi, *Error rate*, Listrik



ABSTRACT

SILA FARSIDIA PUTRI. *ANALYSIS OF ABNORMAL ELECTRICITY CONSUMPTION USING FUZZY C-MEANS AND DECISION TREE ALGORITHMS* (supervised by Ingrid Nurtanio and Christoforus Yohannes)

Electricity plays a crucial role in supporting daily activities, particularly in industrial and residential areas that heavily depend on the availability of this energy. In Indonesia, PT. PLN (Persero), a state-owned electricity provider, faces challenges in energy distribution, including abnormal consumption by customers that can potentially harm both the institution and the customers. This research aims to implement the use of Fuzzy C-Means (FCM) and analyze Decision Tree algorithms in identifying and classifying the characteristics of electrical energy consumption. The FCM algorithm is used to group the characteristics of electrical energy consumption into clusters, and the Decision Tree algorithm is employed to classify the characteristics of customer electricity usage. The data used in this study are the electricity usage data of customers. Clustering results using FCM resulted in the data distribution into four distinct clusters with specific characteristics and recommendations. A significant finding of this research is cluster 4, comprising only 2.5% of the dataset, which shows unreadable current flow conditions and low electricity usage efficiency. As much as 27,8% of the dataset was categorized into cluster 1, characterized by high current imbalances. Cluster 2, which includes 8,7% of the dataset, faced similar issues as cluster 1, compounded by low electricity usage efficiency. Meanwhile, the majority of the data, 60,8%, fell into cluster 3, which indicated normal conditions. Therefore, follow-up actions are necessary for clusters 1, 2, and 4. The classification model results using the Decision Tree algorithm demonstrated high accuracy in scenarios using the entire dataset, with ratios of 60:40 and 70:30 producing an accuracy rate of 99,84% and an error rate of 0,16%. Based on the conducted research, the FCM algorithm can effectively group characteristics of electricity usage, and the classification model generated from the Decision Tree algorithm also shows good accuracy.

Keywords: Fuzzy C-Means, Clustering, Decision Tree, Accuracy, Error rate, Electricity



DAFTAR ISI

LEMBAR PENGESAHAN SKRIPSI	i
PERNYATAAN KEASLIAN.....	ii
ABSTRAK	iii
ABSTRACT	iv
DAFTAR ISI.....	v
DAFTAR GAMBAR	vii
DAFTAR TABEL.....	ix
DAFTAR SINGKATAN DAN ARTI SIMBOL	x
DAFTAR LAMPIRAN.....	xi
KATA PENGANTAR	xii
BAB I PENDAHULUAN	14
1.1 Latar Belakang	14
1.2 Rumusan Masalah	16
1.3 Tujuan Penelitian	16
1.4 Manfaat Penelitian	16
1.5 Batasan Masalah	16
1.6 Sistematika Penulisan	17
BAB II TINJAUAN PUSTAKA.....	18
2.1 <i>Data Mining</i>	18
2.2 <i>Clustering</i>	19
2.3 Algoritma <i>Fuzzy C-Means</i>	21
2.4 <i>Elbow Method</i>	23
2.6 <i>Box Plot</i>	23
2.7 <i>Principle Component Analysis</i>	24
2.8 Klasifikasi	24
2.9 Algoritma <i>Decision Tree</i>	26
2.10 Evaluasi Model.....	27
2.11 Normalisasi	29
2.12 Kelistrikan.....	29
2.13 Trafo.....	30
2.14 <i>Automatic Meter Reading (AMR)</i>	32
2.15 Abnormal	33
BAB III METODOLOGI PENELITIAN.....	34
3.1 Tempat dan Waktu Penelitian	34
3.2 Instrumen Penelitian	34
3.3 Tahapan Penelitian	34
3.4 Implementasi Algoritma <i>Clustering</i>	37
3.5 Interpretasi Hasil <i>Clustering</i>	38
3.6 Implementasi Algoritma Klasifikasi	39
BAB IV HASIL DAN PEMBAHASAN	41
4.1 Konsep Implementasi Algoritma <i>Clustering</i>	41
4.2 Implementasi Algoritma <i>Fuzzy C-Means</i> dalam <i>Clustering</i>	49
4.3 Pembahasan Hasil <i>Clustering</i> dengan 4 Cluster	53
4.4 Konsep Implementasi Algoritma <i>Decision Tree</i> untuk Klasifikasi	66
4.5 Analisis <i>Decision Tree</i> untuk Klasifikasi.....	70



4.6 Pembahasan Hasil Klasifikasi	74
BAB V KESIMPULAN DAN SARAN	81
5.1 Kesimpulan	81
5.2 Saran.....	81
DAFTAR PUSTAKA	83
LAMPIRAN	87



DAFTAR GAMBAR

Gambar 1 Proses dalam <i>data mining</i> (Han et al., 2012)	18
Gambar 2 <i>Box plot</i>	23
Gambar 3 Proses klasifikasi: tahap latih (Han et al., 2012)	25
Gambar 4 Proses klasifikasi: tahap uji (Han et al., 2012)	25
Gambar 5 Konsep <i>decision tree</i> (Han et al., 2012)	26
Gambar 6 <i>Confusion matrix</i> (Han et al., 2012)	27
Gambar 7 Trafo tiga fasa	30
Gambar 8 <i>Automatic meter reading</i>	33
Gambar 9 Tahapan penelitian	35
Gambar 10 Dataset mentah penggunaan listrik bulan Februari (a)	36
Gambar 11 Dataset mentah penggunaan listrik bulan Februari (b)	36
Gambar 12 Dataset mentah penggunaan listrik bulan Februari (c)	36
Gambar 13 <i>Flowchart</i> algoritma FCM	37
Gambar 14 <i>Flowchart</i> algoritma <i>decision tree</i>	39
Gambar 15 Jumlah optimal <i>cluster</i> menggunakan <i>elbow method</i>	50
Gambar 16 Menentukan nilai parameter awal	51
Gambar 17 Menentukan matriks partisi awal	51
Gambar 18 Menghitung pusat <i>cluster</i>	51
Gambar 19 Menghitung derajat keanggotaan	51
Gambar 20 Menghitung fungsi objektivitas	52
Gambar 21 Visualisasi hasil <i>clustering</i>	52
Gambar 22 Jumlah data tiap <i>cluster</i>	53
Gambar 23 Rata-rata parameter tiap <i>cluster</i>	54
Gambar 24 Distribusi nilai <i>voltage</i> pada ketiga fasa di <i>cluster 1</i>	54
Gambar 25 Distribusi nilai <i>current</i> pada ketiga fasa di <i>cluster 1</i>	55
Gambar 26 Distribusi nilai <i>power factor</i> , <i>active power</i> , <i>apparent power</i> , dan ketidakseimbangan <i>current</i> di <i>cluster 1</i>	56
Gambar 27 Distribusi nilai <i>voltage</i> pada ketiga fasa di <i>cluster 2</i>	57
Gambar 28 Distribusi nilai <i>current</i> pada ketiga fasa di <i>cluster 2</i>	58
Gambar 29 Distribusi nilai <i>power factor</i> , <i>active power</i> , <i>apparent power</i> , dan ketidakseimbangan <i>current</i> di <i>cluster 2</i>	58
Gambar 30 Distribusi nilai <i>voltage</i> pada ketiga fasa di <i>cluster 3</i>	60
Gambar 31 Distribusi nilai <i>current</i> pada ketiga fasa di <i>cluster 3</i>	61
Gambar 32 Distribusi nilai <i>power factor</i> , <i>active power</i> , <i>apparent power</i> , dan ketidakseimbangan <i>current</i> di <i>cluster 3</i>	61
Gambar 33 Distribusi nilai <i>voltage</i> pada ketiga fasa di <i>cluster 4</i>	63
Gambar 34 Distribusi nilai <i>current</i> pada ketiga fasa di <i>cluster 4</i>	63
Gambar 35 Distribusi nilai <i>power factor</i> , <i>active power</i> , <i>apparent power</i> , dan ketidakseimbangan <i>current</i> di <i>cluster 4</i>	64
Gambar 36 Membaca dataset <i>clustering</i>	70
Gambar 37 Skenario 1 pembagian dataset	71
Gambar 38 Skenario 2 pembagian dataset	71
Gambar 39 Skenario 3 pembagian dataset	71
Gambar 40 Pembagian dataset berdasarkan rasio	72
Gambar 41 Klasifikasi <i>decision tree</i>	72



Gambar 42 Fungsi <i>confusion matrix</i>	72
Gambar 43 Menjumlahkan banyak data sebenarnya	73
Gambar 44 Menjumlahkan hasil prediksi	73
Gambar 45 Menghitung nilai akurasi dan <i>error rate</i>	73
Gambar 46 Menghitung nilai <i>precision</i> , <i>recall</i> , dan <i>F1-Score</i>	73
Gambar 47 Hasil <i>confusion matrix</i> skenario 1.1	74
Gambar 48 Hasil <i>confusion matrix</i> skenario 1.2	75
Gambar 49 Hasil <i>confusion matrix</i> skenario 1.3	75
Gambar 50 Hasil <i>confusion matrix</i> skenario 2.1	76
Gambar 51 Hasil <i>confusion matrix</i> skenario 2.2	77
Gambar 52 Hasil <i>confusion matrix</i> skenario 2.3	77
Gambar 53 Hasil <i>confusion matrix</i> skenario 3.1	78
Gambar 54 Hasil <i>confusion matrix</i> skenario 3.2	79
Gambar 55 Hasil <i>confusion matrix</i> skenario 3.3	79



DAFTAR TABEL

Tabel 1 Parameter yang digunakan	42
Tabel 2 Contoh dataset.....	42
Tabel 3 Contoh dataset telah dinormalisasi	43
Tabel 4 Perhitungan jarak data terhadap <i>centroid</i>	47
Tabel 5 Perhitungan derajat keanggotaan tertinggi.....	48
Tabel 6 Nilai SSE pada setiap <i>cluster</i>	50
Tabel 7 Jumlah data tiap <i>cluster</i>	53
Tabel 8 Distribusi <i>voltage</i> ketiga fasa <i>cluster 1</i>	55
Tabel 9 Distribusi <i>current</i> ketiga fasa <i>cluster 1</i>	55
Tabel 10 Distribusi nilai <i>power factor</i> , <i>active power</i> , <i>apparent power</i> , dan ketidakseimbangan <i>current</i> di <i>cluster 1</i>	56
Tabel 11 Rangkuman interpretasi hasil analisis <i>cluster 1</i>	56
Tabel 12 Distribusi <i>voltage</i> ketiga fasa <i>cluster 2</i>	57
Tabel 13 Distribusi <i>current</i> ketiga fasa <i>cluster 2</i>	58
Tabel 14 Distribusi nilai <i>power factor</i> , <i>active power</i> , <i>apparent power</i> , dan ketidakseimbangan <i>current</i> di <i>cluster 2</i>	59
Tabel 15 Rangkuman interpretasi hasil analisis <i>cluster 2</i>	59
Tabel 16 Distribusi nilai <i>voltage</i> pada ketiga fasa di <i>cluster 3</i>	60
Tabel 17 Distribusi nilai <i>current</i> pada ketiga fasa di <i>cluster 3</i>	61
Tabel 18 Distribusi nilai <i>power factor</i> , <i>active power</i> , <i>apparent power</i> , dan ketidakseimbangan <i>current</i> di <i>cluster 3</i>	62
Tabel 19 Rangkuman interpretasi hasil analisis <i>cluster 3</i>	62
Tabel 20 Distribusi nilai <i>voltage</i> pada ketiga fasa di <i>cluster 4</i>	63
Tabel 21 Distribusi nilai <i>current</i> pada ketiga fasa di <i>cluster 4</i>	64
Tabel 22 Distribusi nilai <i>power factor</i> , <i>active power</i> , <i>apparent power</i> , dan ketidakseimbangan <i>current</i> di <i>cluster 4</i>	64
Tabel 23 Rangkuman interpretasi hasil analisis <i>cluster 4</i>	65
Tabel 24 Rekomendasi tingkat lanjut.....	66
Tabel 25 Dataset berlabel.....	67
Tabel 26 Contoh hasil <i>confusion matrix</i>	68
Tabel 27 Contoh hasil perhitungan evaluasi model	70
Tabel 28 Hasil evaluasi model skenario 1.1.....	74
Tabel 29 Hasil evaluasi model skenario 1.2.....	75
Tabel 30 Hasil evaluasi model skenario 1.3.....	76
Tabel 31 Hasil evaluasi model skenario 2.1.....	76
Tabel 32 Hasil evaluasi model skenario 2.2.....	77
Tabel 33 Hasil evaluasi model skenario 2.3.....	78
Tabel 34 Hasil evaluasi model skenario 3.1.....	78
Tabel 35 Hasil evaluasi model skenario 3.2.....	79
Tabel 36 Hasil evaluasi model skenario 3.3.....	80
Tabel 37 Perbandingan hasil evaluasi model tiap skenario	80



DAFTAR SINGKATAN DAN ARTI SIMBOL

Lambang/Singkatan	Arti dan Keterangan
FCM	<i>Fuzzy C-Means</i>
D	Jarak
x_{ij}	Titik data ke- i dan atribut ke- j
c_{ij}	<i>Centroid cluster</i> ke- i dan atribut ke- j
N	Jumlah titik data
c	<i>Cluster</i>
$v_i^{(l)}$	Pusat <i>cluster</i> ke- i iterasi ke- l
m	<i>Fuzziness</i>
u_{ik}	Derajat keanggotaan data ke- k terhadap <i>cluster</i> ke- i
z_k	Titik data ke- k
D_{ik}	Jarak data ke- k (z_k) dan pusat <i>cluster</i> ke- i (v_i)
A	Matriks identitas
TP	<i>True Positive</i>
TN	<i>True Negative</i>
FP	<i>False Positive</i>
FN	<i>False Negative</i>
PF	<i>Power Factor</i>
CART	<i>Classification and Regression Trees</i>
p	Probabilitas



DAFTAR LAMPIRAN

Lampiran 1 Dataset Hasil <i>Clustering</i>	87
Lampiran 2 <i>Preprocessed Dataset</i>	87
Lampiran 3 Normalisasi <i>Preprocessed Dataset</i>	87
Lampiran 4 Dataset Hasil <i>Clustering</i>	87
Lampiran 5 Surat Penugasan Dosen Pembimbing.....	88
Lampiran 6 Surat Permohonan Pengambilan Data.....	89
Lampiran 7 Daftar Hadir Seminar Hasil.....	90
Lampiran 8 Berita Acara Seminar Hasil.....	91
Lampiran 9 Penerbitan Surat Penugasan Panitia/Penguji Seminar Hasil.....	92
Lampiran 10 Surat Penugasan Seminar Hasil.....	93
Lampiran 11 Daftar Hadir Ujian Skripsi.....	94
Lampiran 12 Berita Acara Ujian Skripsi.....	95
Lampiran 13 Usulan Susunan Panitia/Penguji Ujian Sarjana.....	96
Lampiran 14 Surat Penugasan Ujian Sarjana.....	97



KATA PENGANTAR

Puji dan syukur senantiasa penulis panjatkan kepada Tuhan Yang Maha Esa, yang telah melimpahkan rahmat dan karunia-Nya sehingga dapat menyelesaikan laporan tugas akhir yang berjudul “**Analisis Konsumsi Listrik Abnormal Menggunakan Algoritma *Fuzzy C-Means* dan *Decision Tree***” sebagai salah satu syarat untuk memperoleh gelar Sarjana Strata-1 pada Departemen Teknik Informatika, Fakultas Teknik, Universitas Hasanuddin.

Dalam proses penyelesaian laporan akhir ini tidak lepas dari bantuan, bimbingan serta dukungan dari berbagai pihak. Oleh karena itu, dengan segala kerendahan hati, penulis ingin mengucapkan rasa terima kasih sebesar-besarnya kepada:

1. Kedua orang tua penulis, Bapak drh. Soegiarto, M.Sc, Ph.D dan Almh. Ibu Dr. drh. Faizah, M.T.A serta saudari penulis, drh. Tiza, serta Keluarga Besar penulis yang telah mencurahkan begitu banyak kasih sayang serta motivasi kepada penulis untuk terus berkarya dan belajar dalam bentuk pemenuhan segala kebutuhan dan keperluan penulis baik secara moral maupun moril.
2. Ibu Dr. Ir. Ingrid Nurtanio, M.T selaku dosen pembimbing I dan Bapak Ir. Christoforus Yohannes, M.T selaku dosen pembimbing II, yang senantiasa menyediakan waktu, tenaga, pikiran dan bimbingan dalam mengarahkan penulis untuk menyelesaikan tugas akhir.
3. Ibu Dr. A. Ejah Umraeni, S.T., M.T. selaku dosen penguji I dan Ibu Tyanita Puti Marindah Wardhani ST., M.Inf. selaku dosen penguji II, yang telah memberikan masukan kepada penulis pada tugas akhir ini.
4. Segenap Dosen dan Staf Departemen Teknik Informatika, Fakultas Teknik, Universitas Hasanuddin yang telah membantu penulis selama masa perkuliahan.
5. Segenap pegawai PLN, Pak Luthfil, Pak Asdar, Elma dan Wafa, yang bersedia membantu dan meluangkan waktu untuk melakukan diskusi terkait tugas akhir.



6. Teman-teman semasa kuliah, Agil, Rahma, Yusrah, Artia, Brillli, Dita, Besse, Reskita, Fadhil, Rajab, Rayyan, Marcel, Arfani, Arfandy, Sabda serta rekan-rekan lab yang telah membantu dan memberikan semangat serta hiburan di masa-masa penyelesaian tugas akhir.
7. Sahabat-sahabat yang telah membantu penulis dalam proses penelitian, Artia, Brillli, Dita dan Gito yang telah membantu penulis dalam pengumpulan data, diskusi, dan memberikan dukungan dalam penyelesaian tugas akhir penulis.
8. Saudara-saudari S19NIFIER yang telah memberikan nasihat, bantuan dan semangat selama masa kuliah.
9. Sahabat-sahabatku Azki, Cica, Dona, Nindi, Nuqu dan Sasa yang senantiasa memberikan dukungan dan semangat untuk penyelesaian tugas akhir.
10. Serta pihak-pihak lain yang tidak disebutkan dan tanpa sadar telah menjadi inspirasi dan membantu penulis dalam menyelesaikan tugas akhir.

Penulis berharap semoga Tuhan Yang Maha Esa berkenan membalas segala kebaikan serta jasa dari semua pihak yang telah banyak membantu penulis dalam menyelesaikan tugas akhir ini. Penulis menyadari bahwa tugas akhir ini masih jauh dari kata sempurna, oleh karena itu penulis mengharapkan segala bentuk saran serta masukan yang membangun dari berbagai pihak. Semoga tugas akhir ini dapat memberi manfaat bagi para pembaca dan semua pihak. Aamiin.

Sudiang, 30 Juni 2024

Penulis,

Sila Farsidia Putri



BAB I PENDAHULUAN

1.1 Latar Belakang

Energi listrik memiliki peran yang sangat penting dalam kehidupan sehari-hari, karena sebagian besar peralatan rumah tangga dan peralatan lain yang dibutuhkan untuk mendukung aktivitas sehari-hari membutuhkan energi listrik. Wilayah industri dan perumahan contohnya, sangat bergantung pada ketersediaan energi listrik tersebut. Seiring dengan peningkatan penggunaan energi listrik, konsumsi energi listrik yang efisien dan efektif sangatlah penting (Herlambang et al., 2022). Di Indonesia, PT. PLN tidak hanya bertanggung jawab sebagai perusahaan BUMN yang menyediakan energi listrik bagi kepentingan umum, tetapi juga memastikan bahwa penyaluran dan penjualan energi listrik dilakukan dengan tingkat kualitas yang bagus. Khusus di wilayah kota Makassar Selatan, pendistribusian energi oleh PT PLN (Persero) UP3 Makassar Selatan harus memenuhi standar kualitas penyaluran energi yang baik serta digunakan oleh pelanggan dalam kondisi pemakaian yang normal, dengan pemakaian tegangan yang sesuai dalam kisaran variasi +5% dan -10% pada Jaringan Tegangan Rendah (JTR) 198 V sampai 231 V (Koerniawan et al., 2024). Serta standar ketidakseimbangan arus pada ketiga fasa diatur dalam std 446-1995 yaitu sebesar 5%-20% (Ginting, 2022). Standar nilai minimum untuk faktor daya yang ditetapkan oleh PLN berdasarkan peraturan SPLN 70-1 adalah $>0,85$ (Esye & Lesmana, 2021). Jika ditemukan nilai tegangan dan arus serta *power factor* yang tidak sesuai maka dapat merugikan baik dari sisi PLN maupun pelanggan.

Dalam melakukan proses distribusi energi listrik, perusahaan menghadapi masalah dalam teknis di sisi pelanggan, termasuk konsumsi energi listrik yang abnormal oleh pelanggan meliputi masalah yang terjadi karena kondisi peralatan atau infrastruktur pada sisi pelanggan yang mempengaruhi pengukuran atau kualitas energi listrik yang mereka terima (Fitriany, 2017). Kelainan konsumsi energi listrik merupakan salah satu masalah kerugian teknis yang dapat

ggu kualitas penyaluran energi ke pelanggan. Rugi teknis seperti n dalam instalasi listrik dari sisi pelanggan. Beberapa fitur yang umumnya n dalam mengidentifikasi kelainan konsumsi energi listrik oleh pelanggan



meliputi pengelompokan pola konsumsi energi beban listrik berdasarkan letak geografis yang berpotensi dapat mengidentifikasi rugi teknis (Buzau et al., 2019).

Identifikasi kelainan konsumsi energi listrik yang dilakukan selama ini pada PT. PLN UP3 Makassar Selatan tergolong manual seperti pemantauan data penggunaan listrik dari sistem *Automatic Meter Reading* (AMR) melalui *Microsoft Excel* dalam mengidentifikasi pola penggunaan yang abnormal. Namun model inspeksi secara manual ini tergolong tidak efisien dan rentan terhadap kesalahan manusia sehingga sulit untuk mempertahankan tingkat akurasi kelainan pada pola penggunaan beban listrik.

Dengan mengimplementasikan *clustering* dan klasifikasi pada dataset penggunaan energi listrik, identifikasi kelainan konsumsi oleh pelanggan dapat dilakukan secara efisien. Berdasarkan penelitian yang telah dilakukan oleh Hamdhani et al., 2022, teknik klasterisasi profil pelanggan dapat dilakukan menggunakan metode K-Means yang menghasilkan tiga profil klaster berbeda dimana masing-masing klaster memiliki kontribusi yang berbeda dalam kwh jual mulai dari 46,3%, 34,8%, dan 18,5% serta visualisasi hasil klaster menggunakan scatter plot. Hasil penelitian lain yang dilakukan oleh (Suliyanti et al., 2022), pendekatan algoritma *Fuzzy C-Means* dinilai lebih optimal pada klasifikasi histori data susut daya pelanggan AMR tidak wajar dibandingkan dengan K-Means.

Penelitian lainnya mengenai teknik klasterisasi telah dilakukan oleh Babaei et al., 2021, algoritma *decision tree* digunakan untuk mengklasifikasi pola konsumsi energi listrik dari peralatan listrik di smart home. Dimana model *decision tree* dilatih menggunakan data historis yang kemudian divalidasi menggunakan teknik cross-validation. Penelitian yang menggunakan algoritma serupa juga dilakukan oleh Lumbreras et al., 2023, dimana algoritma *decision tree* digunakan untuk mengklasifikasi pola harian penggunaan alat pemanas gedung yang juga menggunakan teknik cross-validation untuk memvalidasi model algoritma. Hasil dari kedua penelitian tersebut menunjukkan bahwa algoritma *decision tree* menunjukkan nilai akurasi yang tinggi utamanya untuk data yang bersifat kompleks.



asarkan uraian latar belakang di atas, penulis akan mengembangkan
 yang berfokus mengenai klasterisasi dan klasifikasi penggunaan energi
 oleh konsumen menggunakan algoritma *Fuzzy C-Means* dalam
 upokan data dengan menggunakan beberapa variabel pengukuran

penggunaan energi listrik untuk mengetahui ragam pola kelainan khususnya pada rugi non-teknis. Serta menggunakan algoritma *Decision Tree* untuk klasifikasi karakteristik penggunaan listrik pelanggan. Oleh karena itu, penulis mengangkat judul penelitian “Analisis Konsumsi Listrik Abnormal Menggunakan Algoritma *Fuzzy C-Means* dan *Decision Tree*”.

1.2 Rumusan Masalah

Adapun rumusan masalah yang menjadi dasar penelitian ini adalah:

1. Bagaimana implementasi algoritma FCM dalam *clustering* data penggunaan energi listrik pelanggan?
2. Bagaimana menganalisis kinerja algoritma *Decision Tree* dalam klasifikasi pada data uji pasca *clustering*?

1.3 Tujuan Penelitian

Berdasarkan rumusan masalah yang diuraikan, maka tujuan dari penelitian ini adalah:

1. Untuk mengimplementasikan algoritma FCM dalam *clustering* data penggunaan energi listrik pelanggan.
2. Untuk menganalisis kinerja algoritma *Decision Tree* dalam klasifikasi pada data uji pasca *clustering*.

1.4 Manfaat Penelitian

Adapun manfaat yang diharapkan dari penelitian ini adalah:

1. Untuk mempermudah pihak instansi dalam mengelompokkan karakteristik penggunaan listrik pelanggan yang terbentuk berdasarkan variabel-variabel pengukuran energi listrik dari hasil *clustering* menggunakan algoritma FCM.
2. Untuk mempermudah pihak instansi dalam mengkategorikan karakteristik penggunaan dalam konsumsi listrik secara efektif dan efisien untuk mempermudah pengambilan tindakan lebih lanjut.



Rumusan Masalah

Rumusan masalah pada penelitian ini adalah:

yang dianalisis dalam penelitian ini adalah dataset penggunaan energi listrik pelanggan industri selama tiga bulan.

2. Variabel-variabel yang digunakan dalam penelitian ini adalah:
 - a. *Voltage* fasa *L1 – L3*
 - b. *Current* fasa *L1 – L3*
 - c. *Power Factor Total*
 - d. *Active Power Total*
 - e. *Apparent Power Total*
 - f. *Current Imbalance*
3. Algoritma clustering yang digunakan adalah FCM
4. Menggunakan metode elbow dalam menentukan jumlah *cluster* yang optimal
5. Algoritma klasifikasi yang digunakan adalah *Decision Tree*
6. Menggunakan *confusion matrix*, *accuracy*, *error rate*, *precision*, *recall*, dan *F1-Score* sebagai metode evaluasi model klasifikasi

1.6 Sistematika Penulisan

BAB I PENDAHULUAN

Bab ini menjelaskan mengenai latar belakang masalah, rumusan masalah, tujuan penelitian, manfaat penelitian, batasan masalah, dan sistematika penulisan.

BAB II TINJAUAN PUSTAKA

Bab ini membahas landasan teori yang membangun kerangka berpikir serta membantu menganalisis dan menyelesaikan masalah yang diteliti serta metode dan variabel lain yang digunakan dalam penelitian.

BAB III METODOLOGI PENELITIAN

Bab ini membahas mengenai tahapan penelitian, instrumen penelitian, pengumpulan dan pengolahan data, perancangan sistem, dan penerapan algoritma *clustering* dan klasifikasi.

BAB IV HASIL DAN PEMBAHASAN

Bab ini membahas tentang pembahasan hasil dari implementasi algoritma.

BAB V PENUTUP

Bab ini berisi tentang kesimpulan yang diperoleh berdasarkan hasil penelitian yang telah dilakukan serta saran-saran untuk pengembangan lebih lanjut.

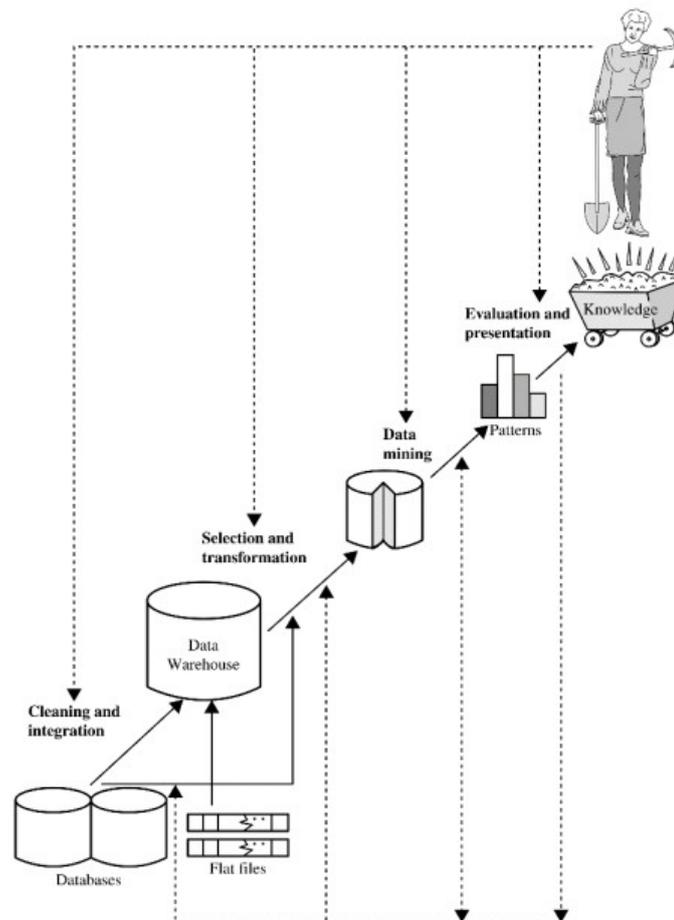


BAB II TINJAUAN PUSTAKA

2.1 Data Mining

Data mining adalah studi tentang mengumpulkan, membersihkan, memproses, menganalisis, dan mendapatkan pengetahuan yang berguna dari data. Terdapat variasi yang luas dalam hal domain masalah, aplikasi, formulasi, dan representasi data pada penerapan dunia nyata. Sehingga istilah *data mining* adalah istilah yang luas untuk menggambarkan berbagai aspek pemrosesan data (Larose & Larose, 2014).

Data mining, atau yang dikenal juga sebagai *Knowledge Discovery from Data* (KDD), dapat juga diartikan sebagai proses untuk menemukan pola yang mewakili pengetahuan baru yang tersimpan secara implisit dalam basis data besar seperti gudang data, web, dan repositori informasi massif lainnya (Han et al., 2012).



Gambar 1 Proses dalam *data mining* (Han et al., 2012)

gunaan *data mining* telah banyak diterapkan di berbagai aplikasi. Salah satu penerapannya adalah *data mining* pada aliran jaringan komputer untuk



mendeteksi penyusupan berdasarkan anomali aliran pesan yang dapat ditemukan melalui pengelompokan data atau dengan membandingkan pola yang sering muncul saat ini dengan yang ada pada waktu sebelumnya (Han et al., 2012).

Sebagai rangkaian proses, *data mining* dapat dibagi menjadi beberapa tahap sebagai berikut (Han et al., 2012):

1. *Data Cleaning and Integration* (Pembersihan dan Integrasi Data)

Data cleaning merupakan proses yang penting. Proses ini melibatkan pengisian nilai yang kosong, meminimalisir *noise*, mengidentifikasi *outlier*, serta melakukan perubahan terhadap data yang tidak konsisten. Proses ini juga melibatkan *data integration*, yakni penggabungan data dari beberapa tempat penyimpanan atau sumber data untuk membantu dalam meningkatkan akurasi.

2. *Data Selection* (Pemilihan Data)

Proses *data selection* merupakan proses dalam pemilihan data yang relevan pada sumber data sesuai dengan tujuan yang akan dicapai.

3. *Data Transformation* (Transformasi Data)

Data pada tahap transformasi data akan diubah diubah atau diselaraskan sehingga proses pengembangan data yang dihasilkan dapat menjadi lebih efisien, dan pola-pola yang ditemukan dapat lebih mudah dipahami.

4. *Data Mining*

Proses implementasi metode *data mining* dilakukan untuk mendapatkan pola dalam data.

5. *Pattern Evaluation* (Evaluasi Pola)

Tidak semua pola yang dihasilkan dari implementasi metode *data mining* menarik. Sehingga diperlukan proses *pattern evaluation* untuk menemukan pola yang menarik berdasarkan ukuran-ukuran yang subjektif.

6. *Knowledge Presentation* (Presentasi Pengetahuan)

Proses *knowledge presentation* merupakan teknik memvisualisasikan data untuk menyajikan pengetahuan dari hasil *data mining* untuk dipresentasikan.



tering

ik *clustering* merupakan sebuah proses yang menganggap data *tuple* objek yang dimana objek-objek tersebut dibagi menjadi beberapa kelompok

atau *subset* yang kemudian disebut sebagai *cluster*. *Cluster* tersebut terbentuk dari adanya kesamaan karakteristik antar anggota dalam satu *cluster* dan berbeda dari objek-objek dalam *cluster* lain. Kesamaan karakteristik ini didefinisikan dengan seberapa dekat jarak antar objek yang dapat dihitung berdasarkan fungsi jarak, seperti persamaan *Euclidean distance* (1), yang mengukur seberapa dekat objek-objek dalam ruang. Kualitas dari sebuah *cluster* dapat direpresentasikan oleh diameter *cluster* yakni jarak maksimum antara dua objek dalam *cluster* tersebut yang dapat membantu menggambarkan sejauh mana objek-objek dalam *cluster* berdekatan satu sama lain (Han et al., 2012).

$$D(x_{ij}, c_{ij}) = \sqrt{(x_{ij} - c_{ij})^2 + (x_{ij} - c_{ij})^2} \quad (1)$$

Analisis *cluster* dibagi menjadi beberapa metode, yaitu *hard clustering* dan *soft clustering*. Metode *hard clustering* berkaitan dengan pengelompokan data dimana setiap data hanya ditempatkan dalam satu *cluster* saja, contohnya algoritma *K-Means*. Sedangkan *soft clustering* berkaitan dengan pengelompokan data dimana satu data dapat ditempatkan dalam beberapa *cluster* sekaligus, contohnya adalah algoritma *Fuzzy C-Means* (Petrus et al., 2019).

Saat ini metode *clustering* dibagi menjadi empat kategori yang berbeda. Pertama, terdapat metode *partitioning* yang menentukan jumlah *cluster* lalu menggunakan teknik iteratif untuk memindahkan objek dari satu *cluster* ke *cluster* lainnya, dengan tujuan untuk meningkatkan partisi. Kedua, terdapat metode *hierarchical* dengan membuat dekomposisi hierarkis dari sekumpulan objek data yang diberikan, yang dapat diklasifikasikan sebagai aglomeratif (*bottom-up*) atau divisif (*top-down*) berdasarkan bagaimana dekomposisi hierarkis terbentuk. Ketiga, metode *density-based* mengelompokkan objek berdasarkan kepadatan objek dengan objek sekitar. Terakhir, metode *grid-based* memetakan ruang objek menjadi sejumlah petak yang membentuk grid, lalu melakukan pengelompokan berdasarkan struktur grid tersebut.

Analisis *clustering* telah diterapkan di beberapa bidang seperti bidang *business intelligence*, pengenalan pola gambar, pencarian web, dan keamanan. Analisis ; juga dapat digunakan sebagai langkah dalam pra-pemrosesan data.



es akhir dari *clustering* adalah melakukan evaluasi untuk menilai 1 atau kesesuaian penerapan analisis *clustering* pada sebuah kumpulan

menilai kualitas yang dihasilkan dari Teknik *clustering*. Proses evaluasi

clustering meliputi menilai pola struktur pengelompokan pada sebuah kumpulan data, penilaian dalam penentuan jumlah *cluster* yang optimal, dan menilai kualitas dari *clustering* yang terbentuk.

2.3 Algoritma *Fuzzy C-Means*

Fuzzy C-Means, atau yang dikenal sebagai FCM, merupakan salah satu metode *clustering partitioning* yang tidak bersifat hirarkis, dan merupakan bagian dari metode *K-Means*. Metode *K-Means* merupakan salah satu metode yang membagi data menjadi satu atau lebih *cluster* dengan derajat atau tingkat keanggotaan yang bervariasi. Tingkat keanggotaan suatu data dalam suatu *cluster* ditentukan oleh derajat keanggotaannya.

FCM adalah salah satu algoritma *clustering* yang menentukan derajat keanggotaan tiap titik data dalam setiap *cluster*. Algoritma ini pertama kali dikenalkan oleh Jim Bezdek pada tahun 1981. Algoritma ini menggunakan model pengelompokan *fuzzy*, sehingga setiap data dapat menjadi anggota dari semua *cluster* yang terbentuk dengan derajat keanggotaan yang berkisar antara 0 hingga 1 (Bezdek et al., 1984).

Konsep dasar *Fuzzy C-Means* pertama-tama melibatkan penentuan pusat *cluster* yang mewakili lokasi rata-rata dari setiap *cluster*. Pada awalnya, posisi pusat *cluster* masih belum akurat. Setiap data memiliki derajat keanggotaan untuk setiap *cluster*. Melalui iterasi yang berulang dengan memperbaiki pusat *cluster* dan derajat keanggotaan data, *cluster* secara bertahap akan mendekati posisi yang optimal. Proses iterasi ini didasarkan pada pengurangan nilai fungsi objektif yang menggambarkan jarak antara titik data dengan pusat *cluster* yang bobotnya dipengaruhi oleh derajat keanggotaan titik data tersebut (Sanusi et al., 2020).

$$1 < c < N \quad (2)$$

Algoritma FCM diinisialisasi dengan menentukan jumlah *cluster*, disimbolkan dengan c , dengan batasan bahwa jumlah *cluster* berkisar antara 1 hingga N , di mana N merupakan jumlah total titik data, persamaannya dapat dilihat pada persamaan lain itu, terdapat juga parameter eksponen bobot m yang nilainya lebih dari mengatur tingkat *fuzziness* dalam penentuan *cluster*, dan toleransi error ϵ menentukan kapan proses iterasi harus berhenti. Algoritma ini memulai



operasinya dengan matriks partisi yang diinisialisasi secara acak (Sanusi et al., 2020).

Ketika iterasi dimulai, algoritma menghitung *cluster centroid* secara berkala. Secara khusus, setiap pusat *cluster* v_i iterasi ke l didefinisikan sebagai bobot rata-rata dari semua titik data. Bobot di sini adalah derajat keanggotaan u_{ik} dari setiap titik data k terhadap *cluster* i , yang dinaikkan ke pangkat m , dan dapat dituliskan melalui persamaan (3) berikut :

$$v_{i^{(l)}} = \frac{\sum_{k=1}^N (u_{ik}^{(l-1)})^m z_k}{\sum_{k=1}^N (u_{ik}^{(l-1)})^m} \quad (3)$$

Selanjutnya, algoritma akan menghitung jarak antara setiap titik data dan titik-titik pusat *cluster*. Jarak kuadrat, D_{ik}^2 , antara suatu titik data z_k dan pusat *cluster* v_i dinilai dengan menggunakan matrik bobot, dimana bobotnya diwakili oleh matriks A :

$$D_{ikA}^2 = (z_k - v_i^{(l)})^T A (z_k - v_i^{(l)}) \quad (4)$$

Dengan menggunakan jarak-jarak tersebut, algoritma *Fuzzy C-Means* akan memperbaharui derajat keanggotaan u_{ik} . Untuk setiap titik data, derajat keanggotaannya terhadap suatu *cluster* tertentu dihitung ulang dengan menggunakan persamaan (5) berikut :

$$u_{ik}^{(l)} = \frac{1}{\sum_{j=1}^c \left(\frac{D_{ikA}}{D_{jkA}} \right)^{\frac{2}{m-1}}} \quad (5)$$

Persamaan tersebut memastikan bahwa titik data yang lebih dekat dengan pusat *cluster* akan memiliki derajat keanggotaan yang lebih tinggi terhadap *cluster* tersebut, sementara yang berjarak lebih jauh akan memiliki derajat keanggotaan yang lebih rendah. Proses ini juga mencakup langkah validasi, jika jarak yang dihitung, D_{ikA}^2 , bernilai positif untuk semua *cluster*, maka persamaan (5) akan diterapkan. Jika tidak, maka nilai keanggotaan akan diatur menjadi 0 atau 1, memastikan bahwa nilai-nilai tersebut dinormalisasi dengan benar di semua *cluster*.



akhir, menghitung fungsi objektif untuk memberikan ukuran kualitas upokkan terbaru berdasarkan jarak terdekat antara titik data dan pusat menggunakan rumus (6) berikut :

$$J = \sum_{k=1}^N \sum_{l=1}^c (u_{ij})^m D_{ikA}^2 \quad (6)$$

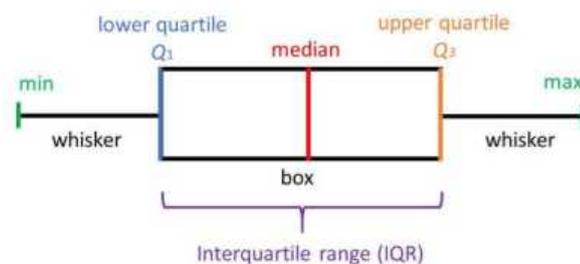
Seluruh proses, mulai dari perhitungan pusat *cluster* hingga pembaharuan derajat keanggotaan, dilakukan secara berulang hingga mencapai nilai toleransi error ϵ . Dengan terus-menerus mempertajam pusat-pusat *cluster* dan derajat keanggotaan, algoritma FCM bertujuan untuk menemukan solusi pengelompokan yang bersifat *fuzzy* dengan meminimalkan bobot jarak titik data terhadap pusat-pusat *cluster* masing-masing (Rahakbauw et al., 2017).

2.4 Elbow Method

Elbow method merupakan suatu metode yang membandingkan perbandingan hasil *sum square error* (SSE) dari setiap *cluster* dan mendapatkan jumlah *cluster* terbaik dari perubahan nilai SSE yang signifikan atau membentuk siku (*elbow*). Oleh karena itu, metode ini dapat digunakan untuk menentukan jumlah *cluster* yang optimal (Nugraha et al., 2023).

2.6 Box Plot

Box plot atau *whisker plot* adalah salah satu metode populer yang digunakan untuk secara statistik memvisualisasikan persebaran data.



Gambar 2 *Box plot*

(sumber: <https://www.simplypsychology.org/wp-content/uploads/box-whisker-plot.jpg>)

Gambar 2 menunjukkan lima angka ringkasan, diantaranya minimum, kuartil pertama (Q1), median, kuartil ketiga (Q3), dan maksimum. Semakin kecil ukuran *box* yang terbentuk, maka variasi data semakin rendah. Sedangkan semakin besar terbentuk, maka semakin tinggi variasi data (Sheshank, 2022).



2.7 Principle Component Analysis

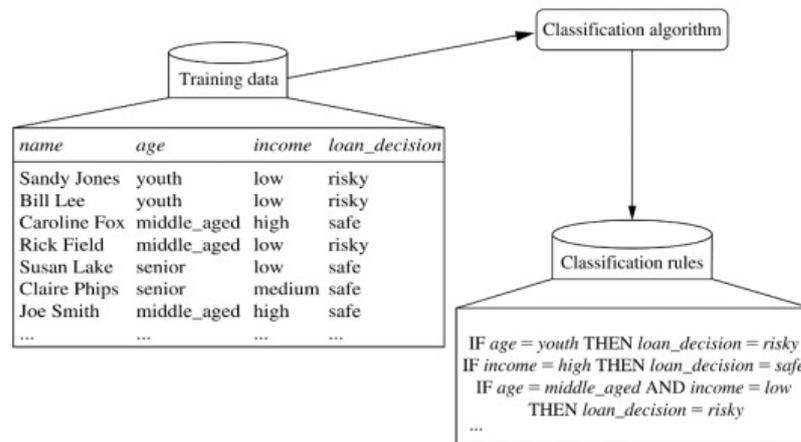
Principle Component Analysis (PCA) merupakan teknik yang digunakan untuk mengurangi dimensi jumlah variabel yang saling terkait, sambil mempertahankan sebanyak mungkin informasi (variasi) yang ada dalam sebuah dataset. Visualisasi menggunakan teknik ini digunakan dengan tujuan memvisualisasikan pola atau struktur dalam data yang berukuran besar. PCA melakukan transformasi terhadap data asli sehingga menghasilkan sejumlah komponen utama (*principal components*) yang merupakan kombinasi linier dari variabel asli. Komponen utama ini diurutkan berdasarkan banyaknya variansi data yang dapat dijelaskan, dengan komponen pertama (PCA1) menjelaskan variansi terbesar, diikuti oleh komponen kedua (PCA2), dan seterusnya (Hintze, 2007).

2.8 Klasifikasi

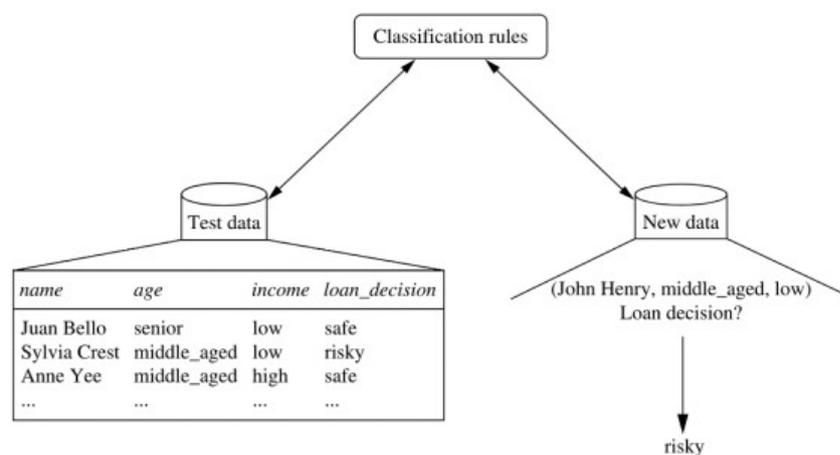
Teknik klasifikasi adalah teknik yang umum digunakan pada *data mining* untuk mengekstraksi model yang menggambarkan berbagai kelas data. Model ini, yang disebut sebagai *classifier*, bertugas untuk memprediksi label kelas kategorikal (diskrit dan tidak berurutan). Misalnya, sebuah model klasifikasi bisa dirancang untuk mengklasifikasikan aplikasi pinjaman bank sebagai “aman” atau “beresiko”. Analisis seperti ini dapat membantu memberikan pemahaman yang lebih baik tentang data secara keseluruhan. Klasifikasi dapat diimplementasikan dalam berbagai bidang, termasuk deteksi penipuan, pemasaran target, prediksi kinerja, manufaktur, dan diagnosis medis. Dalam proses ini, model atau *classifier*, dirancang untuk memprediksi label kelas kategorikal, seperti “aman” atau “beresiko”, untuk data aplikasi peminjaman; “ya” atau “tidak” untuk data pemasaran; atau “perawatan A”, “perawatan B”, atau “perawatan C” untuk data medis. Kategori ini diwakili dengan nilai-nilai diskrit yang tidak memiliki urutan (Han et al., 2012).

Proses klasifikasi data melibatkan dua tahap, yaitu tahap pembelajaran di mana model klasifikasi dibuat dan tahap klasifikasi yang menggunakan model tersebut untuk memprediksi label kelas data yang diberikan. Proses ini digambarkan akan contoh data aplikasi peminjaman yang telah disederhanakan untuk diilustrasi pada Gambar 3.





Gambar 3 Proses klasifikasi: tahap latih (Han et al., 2012)



Gambar 4 Proses klasifikasi: tahap uji (Han et al., 2012)

Dalam tahap pertama, *classifier* dibuat untuk menggambarkan kumpulan kelas data yang telah ditentukan. Ini adalah tahap pembelajaran, dimana algoritma klasifikasi mengembangkan *classifier* dengan menganalisis dari set pelatihan yang terdiri dari tupel database dan label kelas yang terkait dengan data tersebut. Sebuah tupel, X , digambarkan sebagai vektor atribut n -dimensi, $X = (x_1, x_2, \dots, x_n)$, yang menggambarkan pengukuran dari atribut database masing-masing A_1, A_2, \dots, A_n . Setiap tupel, X , diasumsikan masuk dalam kelas yang telah ditentukan sebelumnya, yang ditunjukkan oleh atribut label kelas yang terpisah. Atribut label kelas ini bersifat diskrit dan tidak berurutan. Tupel-tupel yang menjadi bagian dari set pelatihan disebut sebagai tupel pelatihan dan dipilih secara acak dari database yang sedang dianalisis.



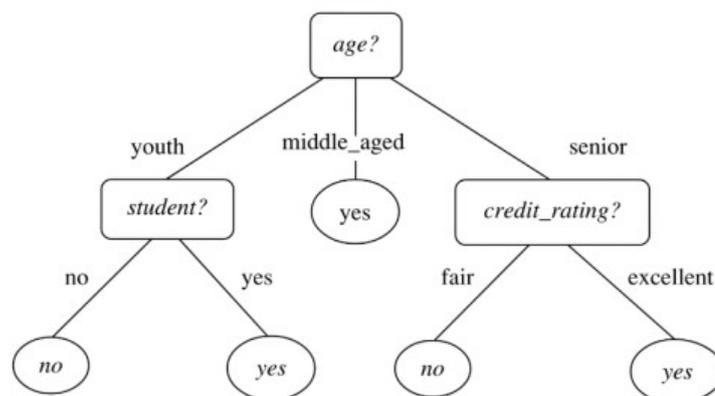
tahap kedua, ilustrasi Gambar 4, model digunakan untuk klasifikasi. pertama adalah mengestimasi akurasi prediktif dari klasifikasi. Jika set pada tahap pembelajaran digunakan untuk mengukur akurasi, maka akan terlihat terlalu baik karena *classifier* cenderung melampaui batas

kecocokan. Sehingga digunakan set tes yang terdiri dari tupel tes dan label kelas yang terkait yang independen dari tupel pelatihan, artinya tupel-tupel tes tidak digunakan dalam membangun *classifier*.

Akurasi *classifier* terhadap set tes diukur dari persentase tupel tes yang diklasifikasikan secara benar. Label kelas dari setiap tupel tes dibandingkan dengan prediksi kelas dari pengklasifikasi yang telah dipelajari. Jika akurasi dari *classifier* dianggap memadai, maka *classifier* tersebut dapat digunakan untuk memprediksi klasifikasi tupel data yang belum memiliki label kelas.

2.9 Algoritma *Decision Tree*

Decision tree adalah sebuah metode yang strukturnya menyerupai *flow chart*, di mana setiap *internal node* menandakan sebuah uji coba terhadap suatu atribut, setiap *branch* mewakili hasil dari uji coba tersebut, dan setiap *leaf node* menyimpan label kelas. Simpul paling atas dalam struktur adalah *root node* (Koli, 2023). Sebuah contoh struktur *decision tree* ditunjukkan pada Gambar 5. Struktur tersebut mengilustrasikan konsep pembelian komputer, dengan tujuan memprediksi apakah pelanggan di toko Z akan membeli komputer. *Internal node* ditandai dengan persegi panjang dan *leaf node* ditandai dengan bentuk oval (Han et al., 2012).



Gambar 5 Konsep *decision tree* (Han et al., 2012)

Decision tree diimplementasikan dalam klasifikasi. Ketika ada tupel, X , dengan label kelas yang belum diketahui, nilai atribut dari tupel ini diuji melalui *decision* apakah jalur akan ditelusuri dari *root node* hingga *leaf node*.



tahun 1984, sekelompok statistikawan menerbitkan buku *Classification and Regression Trees* (CART), yang menjelaskan pembuatan *decision tree* biner. Mereka mengadopsi pendekatan *greedy*, yakni tanpa mengecek kembali, dimana

decision tree dibuat dengan cara rekursif dari atas ke bawah dengan membagi data menjadi subset yang lebih kecil. *Gini index* digunakan dalam metode CART, dimana *Gini index* digunakan untuk mengukur impuritas suatu data D atau mengukur seberapa heterogen sebuah kumpulan data.

$$Gini(D) = 1 - \sum_{i=1}^m p_i^2 \quad (7)$$

Dimana p_i adalah probabilitas dari dataset D yang termasuk ke dalam kelas C_i atau label dan dihitung sebagai $|C_i, D| / |D|$. Penjumlahan dilakukan untuk m kelas.

2.10 Evaluasi Model

Setelah membangun model klasifikasi, perlu dilakukan evaluasi untuk memperkirakan akurasi *classifier* dalam melakukan klasifikasi. Terdapat beberapa metrics evaluasi yang dapat dilakukan dalam mengevaluasi model, diantaranya *confusion matrix*, akurasi, *error rate*, *precision*, *recall* dan *F1-Score*. Terdapat empat terminologi yang umum digunakan dalam mengevaluasi model klasifikasi: TP (*True Positives*) mengarah pada tupel bernilai positif yang diklasifikasikan dengan benar oleh *classifier*, TN (*True Negatives*) mengarah pada tupel bernilai negatif yang diklasifikasikan dengan benar, FP (*False Positives*) mengarah pada tupel bernilai negatif yang keliru diklasifikasikan sebagai positif, dan FN (*False Negatives*) yang merupakan kebalikan dari FP dimana sampel positif keliru diklasifikasikan sebagai negatif (Awoyele, 2020).

2.10.1 Confusion Matrix

Salah satu evaluasi yang digunakan untuk menilai performa model klasifikasi adalah *confusion matrix* yang berisi ringkasan tabular dari jumlah prediksi yang benar dan salah dari sebuah *classifier*. TP dan TN menandakan *classifier* berhasil mengidentifikasi hasil dengan benar, sedangkan FP dan FN menandakan *classifier* salah dalam mengidentifikasi hasil.

		Predicted class	
		yes	no
Actual class	yes	TP	FN
	no	FP	TN

Gambar 6 *Confusion matrix* (Han et al., 2012)



Confusion matrix berisi matriks $m \times m$, dimana $m \geq 2$. Baris m dan kolom m pertama mengindikasikan jumlah tupel dari kelas aktual i yang diklasifikasikan oleh model sebagai kelas j pada entri, $CM_{i,j}$. Agar *classifier* memiliki akurasi model yang baik, idealnya sebagian tupel akan direpresentasikan sepanjang diagonal *confusion matrix*, dari entri $CM_{1,1}$ hingga entri $CM_{m,m}$, dengan sisa entri bernilai 0 atau mendekati 0.

2.10.2 Akurasi

Akurasi dari sebuah *classifier* adalah persentase tupel yang diklasifikasikan dengan benar oleh *classifier*. Akurasi dihitung menggunakan persamaan berikut,

$$accuracy = \frac{TP + TN}{TP + FP + TN + FN} \quad (8)$$

2.10.3 Error Rate

Error rate dari sebuah *classifier* adalah persentase tupel yang diklasifikasikan salah oleh *classifier*. *Error rate* dihitung menggunakan persamaan berikut,

$$error\ rate = \frac{FP + FN}{TP + FP + TN + FN} \quad (9)$$

2.10.4 Precision

Precision juga disebut dengan ukuran ketepatan atau keakuratan dari hasil positif yang diklasifikasikan dengan persamaan (10). *Precision* merupakan persentase tupel yang dilabeli positif oleh model yang sebenarnya positif.

$$precision = \frac{TP}{TP + FP} \quad (10)$$

2.10.5 Recall

Recall adalah ukuran kelengkapan yang mengukur seberapa baik model dalam melabeli tupel dengan label positif yang persamaannya dapat dilihat pada persamaan (11).

$$recall = \frac{TP}{TP + FN} \quad (11)$$



2.9.5 F1-Score

F1-Score adalah metrik yang menghitung nilai rata-rata harmonik dari *precision* dan *recall* dengan nilai yang berkisar antara 0 hingga 1, dimana 1 adalah hasil terbaik. Persamaan dari penghitungan metrik ini dapat dilihat pada persamaan (12).

$$F1 - Score = \frac{2 \times precision \times recall}{precision + recall} \quad (12)$$

2.11 Normalisasi

Satuan pengukuran yang digunakan dapat mempengaruhi analisis data. Untuk mengurangi ketergantungan pada pilihan unit ukuran, data harus dinormalisasi atau distandarisasi. Proses normalisasi data berguna untuk memberikan bobot yang seragam pada semua data atribut. Salah satu metode normalisasi yang sering digunakan adalah metode *min-max scaler*.

Normalisasi *min-max* mengubah data asli melalui sebuah transformasi linier. Misalnya, jika min_A dan max_A merupakan nilai minimum dan maksimum dari atribut A , metode ini akan mengkonversi nilai v_i dari A sampai v_o dalam rentang $[new_min_A, new_max_A]$ melalui persamaan (13).

$$v'_i = \frac{v_i - min_A}{max_A - min_A} (new_max_A - new_min_A) + new_min_A \quad (13)$$

2.12 Kelistrikan

Kelistrikan adalah cabang ilmu fisika yang mempelajari fenomena muatan listrik, arus, medan, dan gaya yang dihasilkan dari interaksi antar muatan. Ini mencakup pembangkitan, distribusi, dan pemanfaatan energi listrik dalam berbagai aplikasi mulai dari peralatan rumah tangga hingga teknologi industri. Dalam kelistrikan, konsep-konsep seperti tegangan dan arus perlu dipahami untuk memahami bagaimana konsep dari kelistrikan (The Engineering Mindset, 2017).

2.12.1 Voltage



tage atau tegangan adalah gaya yang menyebabkan muatan listrik dalam sebuah konduktor. Gaya yang diperlukan untuk membuat muatan mengalir dalam sebuah konduktor disebut dengan beda potensial atau yang

dikenal sebagai tegangan dengan simbol “V”. Satuan ukuran dari tegangan adalah volt yang juga disimbolkan sebagai “V”.

2.12.2 Current

Current atau arus listrik merupakan aliran elektron bebas dalam konduktor dari satu atom ke atom berikutnya pada arah yang sama. Aliran elektron ini disebut sebagai arus yang ditandai dengan simbol “I”. Satuan ukuran dari arus adalah ampere yang disimbolkan dengan “A”.

2.13 Trafo

Trafo merupakan perangkat listrik yang berfungsi untuk menghantarkan listrik dari sumber daya induk ke konsumen. Trafo dapat mengubah tegangan dan arus dari suatu tingkatan ke tingkatan lainnya pada aliran listrik AC yang menggunakan prinsip induksi elektromagnetik (Wildi, 2002). Secara umum, trafo terdiri atas dua kumparan tembaga yang terisolasi satu sama yang lain yang dikenal dengan kumparan primer dan kumparan sekunder.



Gambar 7 Trafo tiga fasa

(sumber: <https://studyelectrical.com/2019/11/three-phase-transformer.html>)

Kumparan primer akan terhubung ke pemasok daya energi, sedangkan kumparan sekunder akan terhubung dengan perangkat konsumen. Ketika arus AC mengalir melalui kumparan primer, akan menghasilkan medan magnet pada sekitar kumparan, yang kemudian akan mempengaruhi elektron dan menginduksi tegangan pada kumparan sekunder. Trafo dapat dirancang untuk meningkatkan maupun



kan tingkat tegangan sesuai dengan kebutuhan, dikenal dengan *step-up* *ier* dan *step-down transformer*, dengan memanfaatkan jumlah lilitan yang berbeda pada masing-masing kumparan dengan satuan yang n adalah *volt-amp* (The Engineering Mindset, 2022).

Setelah tegangan listrik disesuaikan oleh trafo, energi listrik tersebut kemudian siap digunakan oleh konsumen dari skala rumah tangga hingga industri besar. Konsep dasar trafo yang telah dijelaskan juga diterapkan pada trafo tiga fasa. Trafo tiga fasa terdiri dari tiga set kumparan yang masing masing beroperasi dalam tiga fasa yang berbeda, sehingga memberikan keluaran yang lebih stabil dan efisien untuk distribusi tenaga listrik skala besar pada konsumen industri. Dari sisi pelanggan, penggunaan energi listrik tidak hanya berfokus pada penerimaan energi dari trafo, tetapi juga bagaimana konsumen mengolah dan memanfaatkan energi tersebut secara efisien. Penerapan teknologi seperti *Automatic Meter Reading* (AMR) digunakan dalam manajemen dan monitoring penggunaan listrik di sisi pelanggan.

Terdapat beberapa konsep dasar seperti *voltage*, *current*, *apparent power*, *active power*, *power factor*, dan ketidakseimbangan beban pada fasa-fasa yang menentukan kinerja efisiensi sebuah sistem listrik.

2.13.1 Voltage 3 Fasa

Pada trafo 3 fasa Jaringan Tegangan Rendah (JTR) 198 V sampai 231 V, standar tegangan yang digunakan oleh pelanggan dalam kondisi pemakaian normal berada pada kisaran variasi +5% dan -10 (Koerniawan et al., 2024).

2.13.2 Current 3 Fasa

Current pada trafo 3 fasa memberikan informasi tentang keadaan sistem listrik seperti keseimbangan beban, kondisi peralatan dan keamanan sistem. Tidak terbaca atau terukurnya arus dapat mengindikasikan peralatan *wiring* meter yang rusak atau indikasi lain yang perlu tindakan lebih lanjut (Sutura et al., 2022).

2.13.3 Apparent Power

Apparent power atau daya semu adalah daya yang dihasilkan dari perkalian antara daya dan arus dengan satuan *volt-amp* (VA) yang juga merupakan kombinasi antara *active power* dan *reactive power*. *Apparent power* juga mengindikasikan banyaknya penggunaan listrik yang digunakan oleh pelanggan (The Engineering 2022).



2.13.4 Active Power

Active power merupakan energi yang terpakai atau daya yang sebenarnya dipakai oleh konsumen pada aliran listrik AC dengan satuan W. Semakin tinggi nilai *active power* pada *apparent power*, maka akan menghasilkan nilai *power factor* yang mendekati 1 (The Engineering Mindset, 2022).

2.13.5 Power Factor

Power factor merupakan rasio antara energi yang digunakan atau *active power* dan total energi atau *apparent power*. Nilai *power factor* berada antara sama dengan 0 hingga kurang dari sama dengan 1. Semakin mendekati 1 maka semakin efisien penggunaan energi yang dikonsumsi. Standar nilai minimum untuk *power factor* yang ditetapkan oleh PLN berdasarkan peraturan SPLN 70-1 adalah >0,85 (Esye & Lesmana, 2021).

2.13.6 Ketidakseimbangan Beban

Kondisi ketidakseimbangan beban merupakan keadaan dimana syarat seimbang pada ketiga fasa tidak terpenuhi. Keadaan tidak seimbang arus-arus fasa dapat dinyatakan dengan koefisien a, b dan c sebagai berikut (Rumakat & Fauziah, 2021).

$$a = \frac{I_R}{I_{rata-rata}} \quad (14)$$

$$b = \frac{I_S}{I_{rata-rata}} \quad (15)$$

$$c = \frac{I_T}{I_{rata-rata}} \quad (16)$$

$$\% \text{ ketidakseimbangan} = \frac{(|a - 1| + |b - 1| + |c - 1|)}{3} \times 100\% \quad (17)$$

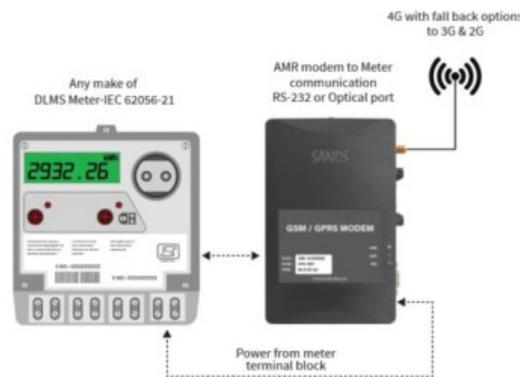
Adapun standar nilai ketidakseimbangan beban pada ketiga fasa diatur dalam standar 1995 yaitu sebesar 5%-20% (Ginting, 2022).



Automatic Meter Reading (AMR)

Automatic Meter Reading atau yang dikenal sebagai AMR adalah sistem an atau pengambilan data hasil pengukuran meter elektronik secara

terpusat dan otomatis dari jarak jauh melalui media komunikasi khusus dan perangkat lunak yang mendukung pengolahan data. Parameter yang dibaca pada umumnya terdiri dari pencatatan, pengukuran tertinggi, pengukuran sesaat dan *load profile*. Fungsi utama dari sistem ini adalah mengumpulkan data yang parameternya telah ditentukan sebelumnya untuk pemantauan pasokan energi kepada pelanggan dan penggunaan energi pelanggan. Data hasil pembacaan tersebut dikumpulkan dan disimpan ke dalam database yang kemudian dapat diakses untuk melakukan analisa (Wiharja & Albahar, 2018).



Gambar 8 *Automatic meter reading*

(sumber: <https://www.sandsindia.com/automatic-electric-meter-reading-amr/>)

Adapun beberapa parameter yang tercatat pada pencatatan *load profile* seperti tegangan pada tiga fasa, arus pada tiga fasa, total energi semu yang terpakai (*apparent power*), total energi yang terpakai (*real power*), total reaktif energi yang terpakai (*reactive power*), hingga *power factor*.

2.15 Abnormal

Dalam berbagai konteks, kata abnormal merujuk pada sesuatu yang menyimpang dari apa yang dianggap standar, normal, atau biasa. Khususnya dalam distribusi listrik, kata tersebut mengacu pada situasi atau kondisi yang berbeda dari standar Perusahaan PT PLN (Persero), atau yang dikenal sebagai SPLN (Standar Perusahaan Listrik Negara). SPLN mencakup berbagai peraturan, pedoman, instruksi, cara pengujian dan spesifikasi teknik yang dirancang untuk memastikan operasi dan keamanan sistem distribusi listrik. Sejak tahun 1976, PT PLN sudah

lirampungkan lebih dari 246 buah standar, diantaranya 99 standar bidang (Funan & Utama, 2020).

