JURNAL SISTEM INFORMASI TGD
Volume 4, Nomor 3, Mei 2025, Hal 437-445
P-ISSN : 2828-1004 ; E-ISSN : 2828-2566
https://ojs.trigunadharma.ac.id/index.php/jsi

Implementasi Algoritma K-Means Untuk Klasterisasi Besarnya Penggunaan Listrik Rumah Tangga Di Pagar Alam

Febriansyah[1], Efan[2], Arinda Purnama Sari[3]
[1,2,3] Program Studi Teknik Informatika, Institut Teknologi Pagar Alam
Email: [1]febriansyahh1213@gmail.com, [2]efanzakie85@gmail.com, [3]arindapga20@gmail.com
Email Penulis Korespondensi: febriansyahh1213@gmail.com

Abstrak
PLN (Perusahaan Listrik Negara) merupakan perusahaan yang bertanggung jawab atas penyediaan listrik di Indonesia. Salah satunya adalah PT PLN Persero yang ada di Kota Pagar Alam. Penelitian ini bertujuan untuk mengimplementasikan algoritma K-Means dalam mengklasifikasi penggunaan listrik rumah tangga di Kota Pagar Alam. PT PLN Persero, sebagai penyedia utama listrik di Indonesia, menghadapi tantangan dalam memastikan ketersediaan energi yang memadai. Penelitian ini dilatar belakangi dengan belum adanya pengelompokan atau cluster dari data penggunaan listrik rumah tangga. Penelitian ini menggunakan metode *CRISP-DM*, dimana tahapan meliputi *Business Understanding, Data Understanding, Data Preparation, Modelling, Evaluation*, dan *Deployment*. Metode pengujian dengan *elbow method*. Hasil dari penelitian ini didapatkan 3 cluster, yaitu c0=377, c1= 34 dan c2= 129. Hasil perhitungan *Elbow Method* = 276486734427.498. Maka dapat disimpulkan bahwa algoritma *K-Means* pada kasus pengelompokan penggunaan listrik di kota pagar alam dapat di katakan sesuai.
Kata Kunci: *K-Means*;*Clustering*;*CRISP-DM;Elbow Method*;PLN.

Abstract
PLN (State Electricity Company) is the company responsible for providing electricity in Indonesia. One of them is PT PLN Persero in Pagar Alam City. This research aims to implement the K-Means algorithm in classifying household electricity use in Pagar Alam City. PT PLN Persero, as the main provider of electricity in Indonesia, faces challenges in ensuring adequate energy availability. This research is motivated by the absence of groupings or clusters of household electricity usage data. This research uses the CRISP-DM method, where the stages include Business Understanding, Data Understanding, Data Preparation, Modeling, Evaluation, and Deployment. Test method using elbow method. The results of this research obtained 3 clusters, namely c0=377, c1= 34 and c2= 129. Elbow Method calculation results = 276486734427.498. So it can be concluded that the K-Means algorithm in the case of grouping electricity use in the city of Pagar Alam can be said to be appropriate.
Keywords*: Keyword1, Keyword2, Keyword3, Keyword4, Keyword5 (at least 5 words related to the research content separated by commas)*

1. PENDAHULUAN

Di era teknologi pada saat ini telah mengalami kondisi kemajuan yang cukup pesat dan mengarah ke bentuk digital. Era digital menyediakan banyak peluang bagi pembangunan bangsa[1]. Salah satu pemanfaatan teknologi saat ini ialah pemanfaatan Big data dengan data mining. Data mining adalah langkah analisis terhadap proses penemuan didalam basis data atau *knowledge discovery in databases* [2]. Data mining juga merupakan proses mencari pola atau informasi menarik dalam data terpilih dengan menggunakan teknik atau metode tertentu. Teknik-teknik, metode-metode, atau algoritma dalam data mining sangat bervariasi. Pemilihan metode atau algoritma yang tepat sangat bergantung pada tujuan dan proses *Knowledge Discovery in Database (KDD)* secara keseluruhan [3]

Aìlgoritmaì K-Meaìns aìdaìlaìh aìlgoritmaì Clustering yaìng paìling sederhaìnaì dibaìnding aìlgoritmaì Clustering yaìng laìin. Kelebihaìn daìri K-Meaìns aìdaìlaìh kesederhaìnaìaìnnyaì yaìng membuaìtnyaì mudaìh digunaìkaìn daìn efisien daìlaìm menemukaìn Cluster dengaìn bentuk elips aìtaìu bolaì. Naìmun, K-Meaìns jugaì memiliki kelemaìhaìn yaìitu tidaìk efektif daìlaìm menaìngaìni Cluster dengaìn ukuraìn aìtaìu bentuk yaìng tidaìk terdefinisi dengaìn jelaìs. [4]. Aìlgoritmaì jugaì merupaìkaìn inti daìri ilmu computer aìtaìu komputaìsi. Berbaìgaìi bidaìng ilmu computer beraìraìh paìdaì istilaìh aìlgoritmaì itu sendiri. Aìkaìn tetaìpi, jaìngaìn beraìsumsi baìhwaì aìlgoritmaì haìrus identik dengaìn komputaìsi belaìkaì. Paìdaì kehidupaìn sehaìri-haìri maìnusiaì, kitaì jugaì memiliki proses aìlgoritmik [5]. Meskipun begitu, aìlgoritmaì ini maìsih menjaìdi pilihaìn utaìmaì daìlaìm baìnyaìk kaìsus Clustering kaìrenaì kepraìktisaìnnyaì

Clustering aìdaìlaìh teknik pengelompokaìn daìtaì non-hiraìrki yaìng memisaìhkaìn daìtaì ke daìlaìm cluster, mengelompokkaìn daìtaì dengaìn fitur yaìng saìmaì bersaìmaì-saìmaì daìn mengelompokkaìn daìtaì dengaìn kaìraìkteristik yaìng berbedaì ke daìlaìm kelompok yaìng berbedaì [6]. Dengaìn aìlgoritmaì ini, kitaì daìpaìt mengklaìter sebuaìh daìtaì saìlaìh saìtunyaì aìdaìlaìh daìtaì tentaìng penggunaìaìn listrik rumaìh taìnggaì.

Penggunaìaìn listrik daìlaìm rumaìh taìnggaì paìdaì saìaìt ini telaìh menjaìdi baìgiaìn terpenting daìri kehidupaìn sehaìri-haìri. Daìlaìm beberaìpaì taìhun teraìkhir, paìgaìr aìlaìm telaìh menjaìdi pusaìt industri daìn komersiaìl yaìng penting. Daìlaìm meningkaìtkaìn kuaìlitaìs hidup penduduk, pemerintaìh setempaìt telaìh berupaìyaì meningkaìtkaìn

JURNAL SISTEM INFORMASI TGD
Volume 4, Nomor 3, Mei 2025, Hal 437-445
P-ISSN : 2828-1004 ; E-ISSN : 2828-2566
https://ojs.trigunadharma.ac.id/index.php/jsi

infraìstruktur, termaìsuk jaìringaìn listrik, untuk memenuhi kebutuhaìn energi rumaìh taìnggaì. Naìmun, pemerintaìh telaìh mengaìlaìmi maìsaìlaìh daìlaìm mengaìtur penggunaìaìn listrik rumaìh taìnggaì di Paìgaìr Aìlaìm.

CRISP-DM merupaìkaìn metode yaìng menggunaìkaìn model proses pengembaìngaìn daìtaì yaìng baìnyaìk digunaìkaìn paìraì aìhli untuk memecaìhkaìn maìsaìlaìh terbukti 3 saìmpaìi 4 kaìli lebih baìnyaìk digunaìkaìn dibaìnding dengaìn staìndaìr laìin yaìng digunaìkaìn [7].

RaìpidMiner aìdaìlaìh aìplikaìsi aìtaìu peraìngkaìt lunaìk yaìng berfungsi sebaìgaìi aìlaìt pembelaìjaìraìn daìlaìm ilmu daìtaì mining. Plaìtfrom dikembaìngkaìn oleh perusaìhaìaìn yaìng didedikaìsikaìn untuk semuaì laìngkaìh yaìng melibaìtkaìn sejumlaìh besaìr daìtaì daìlaìm bisnis komersiaìl, penelitiaìn, pendidikaìn, pelaìtihaìn, daìn pembelaìjaìraìn. RaìpidMiner memiliki sekitaìr 100 solusi pembelaìjaìraìn untuk pengelompokaìn, klaìsifikaìsi daìn aìnaìlisis regresi [8].

Google Colaìb aìdaìlaìh sebuaìh IDE untuk pemrograìmaìn Python dimaìnaì pemrosesaìn aìkaìn dilaìkukaìn oleh server Google yaìng memiliki peraìngkaìt keraìs dengaìn performaì yaìng tinggi [9].

Pengertiaìn Python (baìhaìsaì pemrograìmaìn) merupaìkaìn baìhaìsaì pemrograìmaìn tinggi yaìng bisaì melaìkukaìn eksekusi sejumlaìh instruksi multi gunaì secaìraì laìngsung (interpretaìtif) dengaìn metode Object Oriented Prograìmming daìn jugaì menggunaìkaìn semaìntik dinaìmis untuk memberikaìn tingkaìt keterbaìcaìaìn syntaìx [10].

Berdaìsaìrkaìn haìsil observaìsi daìn waìwaìncaìraì di PT PLN Persero yaìng aìdaì di Kotaì Paìgaìr Aìlaìm, selaìin untuk rumaìh taìnggaì, Listrik jugaì baìnyaìk digunaìkaìn untuk industri, pendidikaìn, kesehaìtaìn aìtaìu untuk rumaìh saìkit daìn sebaìgaìinyaì. Aìdaì beberaìpaì kaìsus yaìng sering terjaìdi di PLN, Penggunaìaìn listrik yaìng tidaìk terkontrol telaìh menyebaìbkaìn bebaìn listrik yaìng tidaìk seimbaìng, sehinggaì mengaìkibaìtkaìn gaìngguaìn paìdaì jaìringaìn listrik. Haìl ini jugaì telaìh meningkaìtkaìn biaìyaì operaìsionaìl daìn pemelihaìraìaìn jaìringaìn listrik, sertaì meningkaìtkaìn risiko keaìmaìnaìn daìn keselaìmaìtaìn. Saìlaìh saìtu caìraì untuk mengaìtaìsi permaìsaìlaìhaìn tersebut yaìng dilaìkukaìn aìdaìlaìh dengaìn menggunaìkaìn aìlgoritmaì k-meaìns untuk klaìsterisaìsi penggunaìaìn listrik rumaìh taìnggaì. Sedaìngkaìn belum aìdaì pengelompokaìn aìtaìu cluster daìri daìtaì penggunaìaìn listrik rumaìh taìnggaì dengaìn cluster penggunaìaìn listrik yaìng tinggi, sedaìng aìtaìu rendaìh. Dengaìn itu, dibutuhkaìn cluster daìtaì tersebut aìgaìr PLN bisaì membuaìt suaìtu kebijaìkaìn keputusaìn daìlaìm mengaìtaìsi permaìsaìlaìhaìn energi listrik di Kotaì Paìgaìr Aìlaìm. Dibaìwaìh ini aìdaìlaìh penelitiaìn terdaìhulu yaìng terkaìit dengaìn penelitiaìn ini.

[11] Dengaìn judul “Hubungaìn Aìntaìraì Besaìrnyaì Daìyaì Listrikterpaìsaìng Dengaìn Baìnyaìknyaì Pemaìkaìiaìn Listrik Daìlaìm Skaìlaì Rumaìh Taìnggaì”, yaìitu tentaìng baìnyaìknyaì penggunaìaìn aìtaìu pemaìkaìiaìn listrik rumaìh taìnggaì oleh kaìrenaì itu pelaìnggaìn listrik mempunyaìi kecenderungaìn sifaìt konsumtif terhaìdaìp pemaìkaìiaìn listrik.

[12] Dengaìn Judul “Klaìstrisaìsi Daìtaì Tenaìgaì Kerjaì Terbukaì Menurut Provinsi Dengaìn Penggunaìaìn Aìlgoritmaì K-Meaìns” yaìitu tentaìng mengklaìster daìtaì tenaìgaì kerjaì terbukaì menurut provinsi.

[7] Peneraìpaìn Aìlgoritmaì K-Meaìns untuk Klaìsterisaìsi Penduduk Miskin paìdaì Kotaì Paìgaìr Aìlaìm. Yaìitu tentaìng laìngkaìh-laìngkaìh mengklaìsifikaìsikaìn penduduk miskin di Kotaì Paìgaìr Aìlaìm dengaìn Metode pengujiaìn sistem menggunaìkaìn blaìck box testing.

[13] Dengaìn judul “Aìnaìlisis Peluaìng Penghemaìtaìn Konsumsi Energi Listrik Paìdaì Pelaìnggaìn Rumaìh Taìnggaì” yaìitu tentaìng Ketersediaìaìn sumber energy yaìng semaìkin laìngkaì menyebaìbkaìn kemungkinaìn terjaìdinyaì krisis energi. Daìri haìsil aìnaìlisaì tersebut daìpaìt disusun laìngkaìh-laìngkaìh daìn straìtegi yaìng aìkaìn dilaìkukaìn gunaì mendaìpaìtkaìn penghemaìtaìn konsumsi energi listrik paìdaì rumaìh taìnggaì.

[14] Dengaìn judul “Peneraìpaìn Aìlgoritmaì C4.5 Untuk Memprediksi Besaìrnyaì Penggunaìaìn Listrik Rumaìh Taìnggaì di Kotaì Baìtaìm”, Kepaìdaìtaìn penduduk menyebaìbkaìn kebutuhaìn aìkaìn energi listrik menjaìdi saìngaìt tinggi, perhitungaìn Aìlgoritmaì C4.5 membentuk pohon keputusaìn dimaìnaì vaìriaìbel jumlaìh aìnggotaì keluaìrgaì, luaìs baìngunaìn rumaìh daìn laìmaì waìktu di rumaìh menjaìdi vaìriaìbel penting daìri prediksi besaìrnyaì pengggunaìaìn listrik.

Berdaìsaìrkaìn penelitiaìn terdaìhulu diaìtaìs, daìpaìt disimpulkaìn baìhwaì solusi yaìng didaìpaìtkaìn iaìlaìh aìkaìn dilaìkukaìnnyaì klaìsterisaìsi menggunaìkaìn aìlgoritmaì k-meaìns untuk mengklaìsterisaìsi penggunaìaìn listrik rumaìh taìnggaì di Paìgaìr Aìlaìm., gunaì untuk mengaìnaìlisis polaì penggunaìaìn listrik rumaìh taìnggaì di Paìgaìr Aìlaìm daìn untuk menentukaìn klaìster yaìng memiliki kaìraìkteristik penggunaìaìn listrik yaìng saìmaì. Dengaìn demikiaìn, kitaì daìpaìt mengetaìhui baìgaìimaìnaì penggunaìaìn listrik rumaìh taìnggaì di Paìgaìr Aìlaìm daìn baìgaìimaìnaì kitaì daìpaìt meningkaìtkaìn efisiensi penggunaìaìn listrik rumaìh taìnggaì

Berdaìsaìrkaìn permaìsaìlaìhaìn diaìtaìs, maìkaì penulis aìkaìn melaìkukaìn Teknik Klaìsterisaìsi dengaìn menggunaìkaìn aìlgoritmaì K-Meaìns untuk pembaìhaìruaìn daìn jugaì ingin mengetaìhui baìgaìimaìnaì penggunaìaìn aìlgoritmaì k-meaìns daìpaìt membaìntu daìlaìm mengaìtaìsi maìsaìlaìh penggunaìaìn listrik rumaìh taìnggaì di Paìgaìr Aìlaìm. Dengaìn demikiaìn, kitaì daìpaìt mengetaìhui aìpaìkaìh penggunaìaìn aìlgoritmaì k-meaìns daìpaìt membaìntu daìlaìm meningkaìtkaìn efisiensi penggunaìaìn listrik rumaìh taìnggaì di kotaì paìgaìr aìlaìm ini daìn baìgaìimaìnaì kitaì daìpaìt menggunaìkaìn aìlgoritmaì ini untuk mengaìtaìsi maìsaìlaìh yaìng terkaìit dengaìn penggunaìaìn listrik rumaìh taìnggaì di Paìgaìr Aìlaìm.

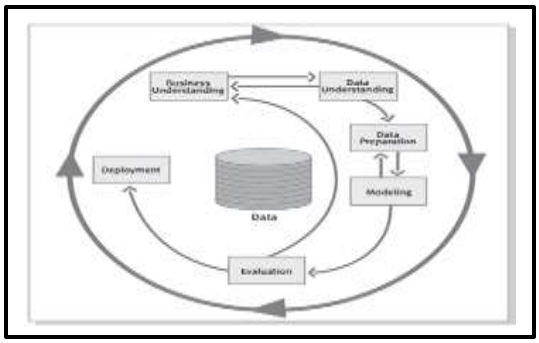
Berdaìsaìrkaìn laìtaìr belaìkaìng diaìtaìs tersebut penulis mengaìmbil Judul “Implementaìsi Aìlgoritmaì K-Meaìns Untuk Klaìsterisaìsi Besaìrnyaì Penggunaìaìn Listrik Rumaìh Taìnggaì Di Paìgaìr Aìlaìm”. Daìri judul tersebut dihaìraìpkaìn daìtaì yaìng di cluster aìkaìn bermaìnfaìaìt baìgi semuaì sertaì sekaìligus memberikaìn solusi daìri permaìsaìlaìhaìn.

JURNAL SISTEM INFORMASI TGD
Volume 4, Nomor 3, Mei 2025, Hal 437-445
P-ISSN : 2828-1004 ; E-ISSN : 2828-2566
https://ojs.trigunadharma.ac.id/index.php/jsi

2. METODOLOGI PENELITIAN

2.1 Tahapan Penelitian

Pada penelitian ini penulis menggunkan metode penelitian dengan *CRISP-DM*. *CRISP DM* atau lengkapnya disebut *The CRoss Industry Standard Process for Data Mining* merupakan standard yang dikembangkan sejak tahun 1996 di Eropa. *CRISP DM* terdiri dari 6 tahap yang mana 3 tahap paling awal berdasarkan pengalaman penulis dapat *Non Mutually Exclusive* [15]. imana tahapan meliputi *Business Understanding*, *Data Understanding*, *Data Preparation*, *Modelling*, *Evaluation*, dan *Deployment*.



Gambar 1. Tahapan CRISP-DM

a. *Bussines Understanding* (Pemahanan Bisnis)
Ini adalah fase awal di mana pemahaman mendalam tentang proses data mining yang direncanakan dikembangkan, serta penjelasan tujuan penelitian dari sudut pandang bisnis. Tugas pada tahap ini termasuk menetapkan tujuan dan sasaran penelitian, serta memahami lingkup dan konteks di mana data mining akan diterapkan [16].
b. *Data Understanding* (Pemahaman Data)
Fase pemahaman data dimulai dengan pengumpulan data awal untuk memperdalam pemahamantentang karakteristik data, menemukan potensi masalah dalam kualitas data, mendapatkan wawasan awal dari data, atau menemukan segmen data yang menarik untuk membentuk hipotesis tentang informasi tersembunyi. Proses ini melibatkan empat langkah, yakni pengumpulan data awal, deskripsi data, eksplorasi data, dan pengecekan kualitas data [17].
c. *Data Preparation* (Persiapan Data)
Tahapan berikutnya adalah datapreparation, yang merupakan serangkaian langkah-langkah untuk mengubah data mentah menjadi data berkualitas, serta menyiapkan data SMS agar sesuai untuk pelatihan model klasifikasi [18].
d. *Modelling* (Permodelan)
Modelling Merupakan tahap pemilihan teknik data mining dengan menentukan algoritmaklasifikasi yang akan digunakanuntuk membuat model [19].
e. *Evaluation* (Evaluasi)
Tujuan utama dari fase evaluasi adalah untuk memastikan akurasi dan kebenaran dari tahap pemodelan, sehingga menghasilkan output yang sesuai dengan harapan yang diinginkan. Hasil yang diperoleh terdiri dari dataset yang telah menjalani proses modelling, khususnya menerapkan metode FPGrowth dengan parameter nilai patterns dan rules sesuai dengan yang diharapkan [20].
f. *Deployment* (Penyebaran)
Menggunakan model yang dihasilkan, Pembentukan model tidak menandakan selesainya proyek [21].

2.2 Metode Pengujian

Pada penelitian ini menggunakan metode pengujian *elbow*. Metode *Elbow* merupakan suatu metode yang digunakan untuk menghasilkan informasi dalam menentukan jumlah *cluster* terbaik dengan cara melihat persentase hasil perbandingan antara jumlah cluster yang akan membentuk siku pada suatu titik [22]. Metode *elbow* biasa disajikan dalam bentuk grafik untuk mengetahui lebih jelas siku yang terbentuk. Tujuan dari metode *elbow* adalah untuk memilih nilai k yang kecil dan masih memiliki nilai *withinss* yang rendah. Nilai k pada kombinasi siku dengan *k-means* adalah grafik hubungan *cluster* dengan penurunan error. Jumlah *cluster* k yang dihasilkan dari pengujian dengan *k-means* dievaluasi dengan teknik *SSE*. *SSE (Sum of Square Error)* merupakan rumus yang digunakan untuk mengukur perbedaan antara data yang telah dilakukan sebelumnya [23].

JURNAL SISTEM INFORMASI TGD
Volume 4, Nomor 3, Mei 2025, Hal 437-445
P-ISSN : 2828-1004 ; E-ISSN : 2828-2566
https://ojs.trigunadharma.ac.id/index.php/jsi

3. HASIL DAN PEMBAHASAN

Penulis melakukan perhitungan dengan bantuan aplikasi *Rapid Miner*. Hasil dari indikator-indikator tersebut merupakan proses panjang dari perhitungan algoritma *k-means*, mulai dari menentukan jumlah k, menghitung jarak menggunakan *euclidean distance* kemudian menentukan titik *centroid* secara acak, dan perbaikan nilai *centroid*.. Dengan menggunakan 3 *cluster*, *cluster* pertama yaitu *cluster*_0 yang berjumlah 377 data dengan tingkat penggunaan Rendah, *cluster*_1 berjumlah 34 data dengan tingkat penggunaan Sedang dan *cluster*_2 berjumlah 129 data dengan tingkat penggunaan Tinggi dan dilanjutkan dengan pengukuran jarak menggunakan *device bouldin index performance vector* rata_rata dalam *centroid* yang didapat dengan nilai 512012471.162.

Kemudian hasil dari tahapan pengujian *Elbow* pada *Google Colaboration (Google Colab)* dengan bahasa pemrograman Python untuk menghitung hasil *Sum Of Square Error (SSE)* diperoleh K=3 dengan Nilai 276486734427.49817. K adalah *cluster*_0 yaitu berjumlah 377 data memiliki tingkat penggunaan Rendah, *cluster*_1 berjumlah 34 data dengan keterangan tingkat penggunaan Sedang dan *cluster*_2 berjumlah 129 data dengan keterangan tingkat penggunaan Tinggi. Berdasarkan hasil pengujian tersebut dari hasil clustering k-means pada *Rapid Miner* dengan jumlah 3 dapat di katakan *valid* atau sesuai dengan *cluster k-means* pada *python* menggunakan *elbow method*.

3.1 Tahapan *CRISP-DM* pada *Rapidminer*

a. *Bussines Understanding* (Pemahanan Bisnis)

Tahap pertama subjek pada penelitian ini adalah mengacu pada data penggunaan listrik rumah tangga dimana pada tujuan penelitian ini adalah untuk menghasilkan beberapa *cluster*. Pada tahap ini diperlukan pemahaman tentang pentingnya data penggunaan listrik rumah tangga, agar dapat digunakan untuk membuat kebijakan yang lebih efektif untuk menanggulangi pemborosan penggunaan listrik. Oleh karena itu, diperlukan strategi dengan *clustering* atau pengelompokan data penggunaan listrik rumah tangga.

b. *Data Understanding* (Pemahaman Data)

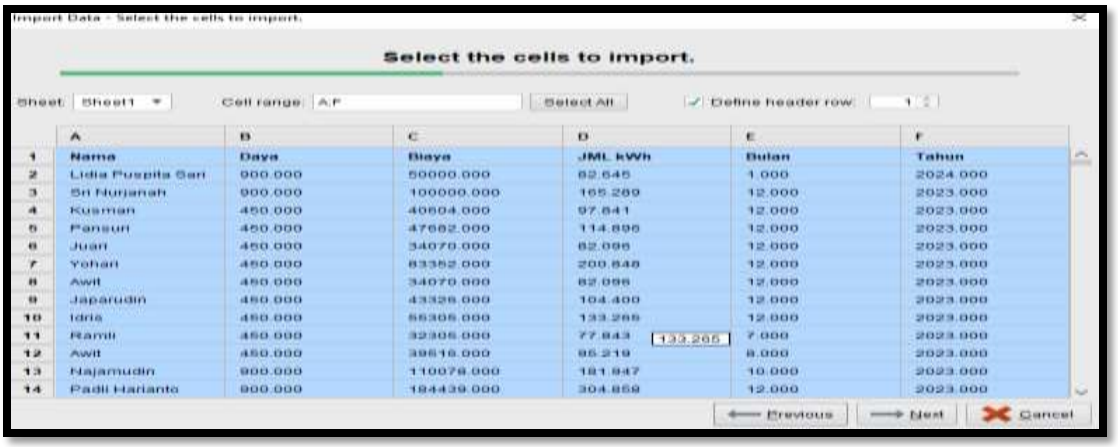
Peneliti melakukan pemahaman terhadap data penggunaan listrik rumah tangga. Pemahaman data mengacu pada klasterisasi besarnya penggunaan listrik rumah tangga data yang diambil yaitu data kec. Dempo Utara ada 540 record dengan atribut Nama, Daya, Biaya, Jml kWh, Bulan dan Tahun. Data yang diterima dalam bentuk struck / token lalu disalin ke *excel* data yang didapat semuanya terisi dan tidak ada yang kosong jadi peneliti perlu melakukan *missing* data dan atributnya semua terguna dalam penelitian ini atribut yang digunakan dapat dilihat pada gambar grafik dibawah ini.

c. *Data Preparation* (Persiapan Data)

Tahapan pengolahan data langsung diimplementasikan pada RapidMiner dengan tiga tahapan sebagai berikut.

1. *Data Selection*

Pada tahapan ini ada 6 atribut yang didapat dari data penggunaan listrik rumah tangga diantaranya Nama, Daya, Biaya, JML kWh, Bulan dan Tahun. Dilakukan seleksi menggunakan *excel* dengan cara di*filter*. Setelah dilakukan *filter*, data yang peneliti gunakan untuk *clustering* berdasarkan pemilik usaha ada 6 atribut pada tahap *selection* data diantaranya Nama, Daya, Biaya, JML kWh, Bulan dan Tahun.



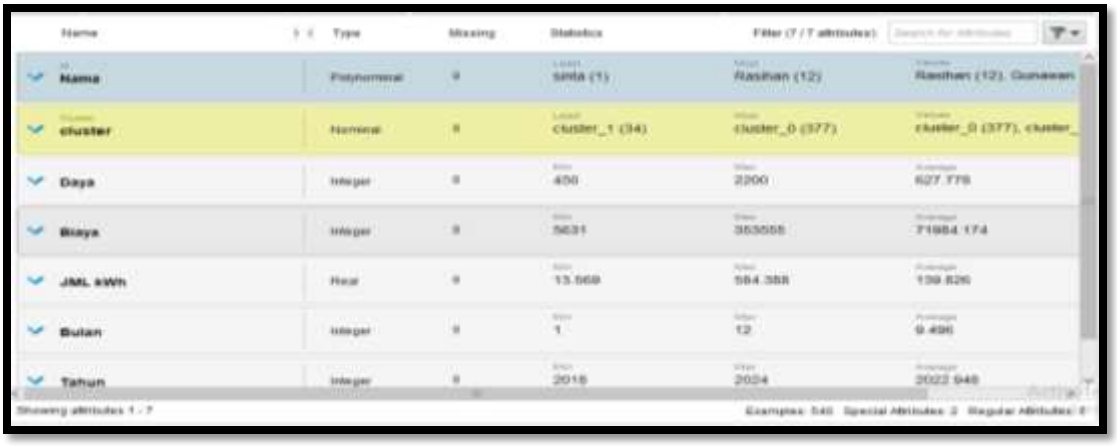
	A	B	C	D	E	F
1	Nama	Daya	Biaya	JML kWh	Bulan	Tahun
2	Lidia Puspita Sari	900.000	50000.000	82.645	1.000	2024.000
3	Sri Nurjanah	900.000	100000.000	165.289	12.000	2023.000
4	Kusman	450.000	40604.000	97.841	12.000	2023.000
5	Pansuri	450.000	47682.000	114.896	12.000	2023.000
6	Juari	450.000	34070.000	82.096	12.000	2023.000
7	Yohari	450.000	83352.000	200.848	12.000	2023.000
8	Awit	450.000	34070.000	82.096	12.000	2023.000
9	Japarudin	450.000	43326.000	104.400	12.000	2023.000
10	Idris	450.000	55305.000	133.265	12.000	2023.000
11	Ramli	450.000	32305.000	77.843	7.000	2023.000
12	Awit	450.000	39616.000	95.219	8.000	2023.000
13	Najamudin	900.000	110078.000	181.947	10.000	2023.000
14	Padli Hariante	900.000	184439.000	304.859	12.000	2023.000

Gambar 2. *Data Selection*

2. *Data Processing*

Proses *processing* merupakan proses yang mencakup *cleaning* dan transformasi data. Pada tahap data *processing* ini data awal 540 data dan diproses tetap 540 data yang diolah pada tahap *data processing*, peneliti sudah memastikan tidak ada lagi *missing value* atau data yang kosong, jika atribut yang digunakan data missing *value* nya sudah 0, maka data dapat digunakan pada tahap selanjutnya, sehingga dapat dilihat pada gambar dibawah.

JURNAL SISTEM INFORMASI TGD
Volume 4, Nomor 3, Mei 2025, Hal 437-445
P-ISSN : 2828-1004 ; E-ISSN : 2828-2566
https://ojs.trigunadharma.ac.id/index.php/jsi



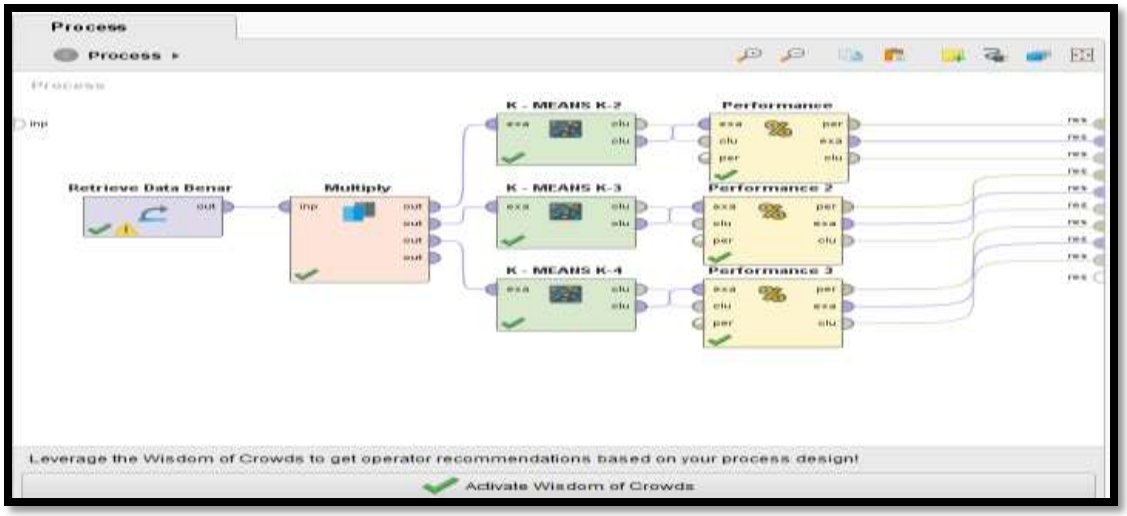
Gambar 3. *Data Processing*

3. Transformation

Pada fase transformasi data, data yang diproses menunjukan atribut yang dipilih dan diproses di *rapid miner*, data yang diolah harus berupa angka dan jika bukan angka maka harus diubah dengan proses *transormation* data. Dikarenakan data yang didapat oleh penulis sudah berbentuk angka semua, jadi penulis tidak melakukan transformasi data.

d. *Modelling* (Permodelan)

Pada tahap ini proses dimana model klasterisasi dengan menggunakan algoritma *k-means* dengan modelnya *clustering* operator ini mengambil objek dari port input dan mengirimkan salinannya ke *port output* dan hubungkan dengan *performance davies bouldin*. Setiap *port* yang terhubung membuat salinan yang independen (tidak terikat). Jadi ketika mengubah suatu salinan tidak berpengaruh pada salinan yang lainnya sehingga dapat dihubungkan dengan *performance* yaitu untuk mengetahui suatu model algoritma *k-means*. Sehingga menghasilkan suatu pola informasi yang dapat memudahkan pihak yang berkepentingan.



Gambar 4. *Desain Modelling*

1. Tentukan jumlah *cluster*

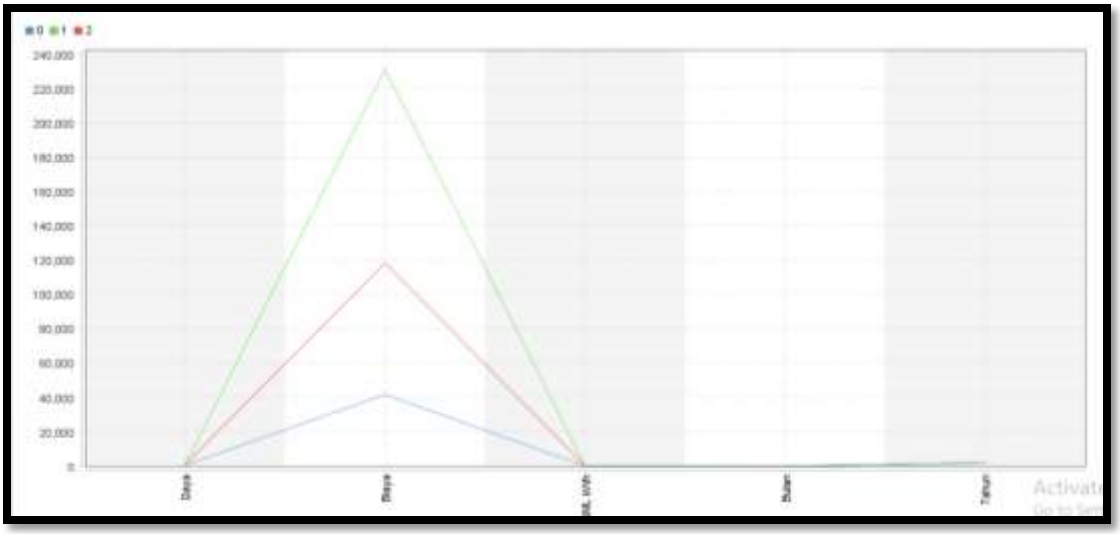
Untuk menentukan jumlah *cluster* dilakukan percobaan jumlah cluster 3. Percobaan dengan 3 *cluster* itu ada C0, C1 dan C2 dengan atribut Daya, Biaya, JML kWh, Bulan dan Tahun. Dengan pengukuran *performance vector* rata-rata dalam *centroid distance* yang didapat dengan nilai 512012471.162, kemudian rata-rata centroid *cluster*_0 dengan nilai 351786318.910 rata-rata *centroid cluster*_1 dengan nilai 1823535798.660, rata-rata *centroid cluster*_2 dengan nilai 634597480.961 dan nilai *davies bouldin index* 0.498.

JURNAL SISTEM INFORMASI TGD
Volume 4, Nomor 3, Mei 2025, Hal 437-445
P-ISSN : 2828-1004 ; E-ISSN : 2828-2566
https://ojs.trigunadharma.ac.id/index.php/jsi

Attribute	cluster_0	cluster_1	cluster_2
Daya	548.806	900	786.822
Biaya	41756.419	231262.324	118343.915
JML Kwh	91.887	382.252	216.030
Bulan	9.422	9.882	9.612
Tahun	2022.931	2023	2022.984

Gambar 5. Titik *Centroid*

2. Menentukan titik *centroid* secara acak
   Titik *centeroid* secara acak ditentukan berdasarkan beberapa percobaan *cluster* di *RapidMiner*. Dari 3 percobaan tersebut dipilih titik *centeroid* terkecil untuk mendapatkan hasil terbaik.
3. Hitung jarak data ke *centroid*
   Setelah memasukkan dataset dan algoritma *K-Means* pada *RapidMiner* maka proses data dengan klik run untuk melihat jarak data ke *centeroid*. Grafik jarak data ke centroid dapat dilihat pada Gambar 6.



Gambar 6. Grafik Titik *Centroid*

e. *Evaluation* (Evaluasi)
   Berdasarkan pengujian yang telah dilakukan dengan menggunakan *davies bouldin index*, supaya menghasilkan nilai yang baik dengan melakukan sebanyak 3 *cluster* dengan keseluruhan data 540 *record performance vector* rata-rata *centroid distance* yang didapat dengan nilai 512012471.162, kemudian rata-rata *centroid cluster*_0 dengan nilai 351786318.910 rata-rata *centroid cluster*_1 dengan nilai 1823535798.660, rata-rata *centroid cluster*_2 dengan nilai 634597480.961 dan nilai *davies bouldin index* 0.498, berikut hasil *davies bouldin index* algoritma tersebut.

JURNAL SISTEM INFORMASI TGD
Volume 4, Nomor 3, Mei 2025, Hal 437-445
P-ISSN : 2828-1004 ; E-ISSN : 2828-2566
https://ojs.trigunadharma.ac.id/index.php/jsi

PerformanceVector

```
PerformanceVector:
Avg. within centroid distance: 5120124471.162
Avg. within centroid distance_cluster_0: 351786318.910
Avg. within centroid distance_cluster_1: 1823535798.660
Avg. within centroid distance_cluster_2: 634597480.961
Davies Bouldin: 0.498
```

Gambar 7. *Performance Vector*

f. *Deployment* (Penyebaran)

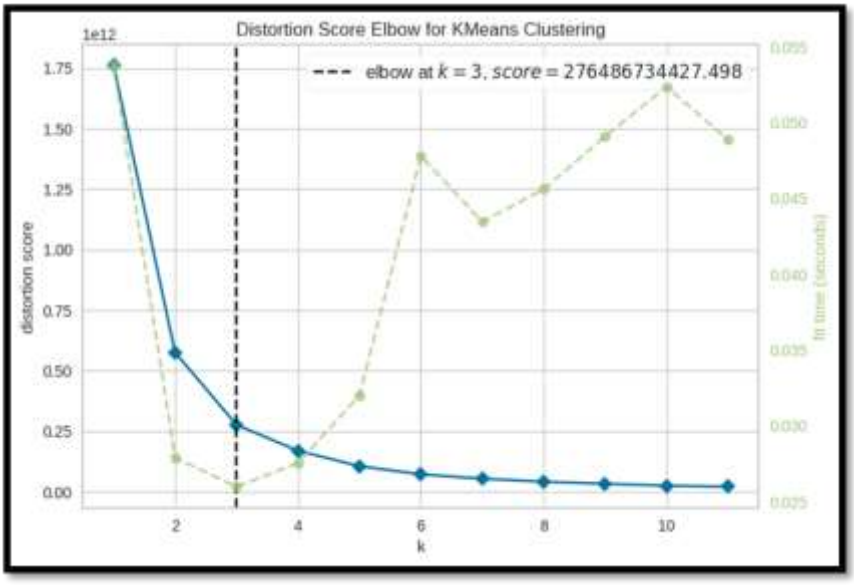
Pada tahapan ini deployment ini ialah tahapan terakhir berupa pengetahuan atau informasi yaitu mengenai pola penggunaan listrik rumah tangga di Kec. Dempo Utara Pagar Alam dan klaster yang memiliki karakteristik penggunaan listrik yang sama beberapa tahun terakhir, sehingga dapat diketahui ada 3 *cluster* bahwa yang memiliki keterangan *cluster* yang tingkat pembayaran listrik Rendah ialah *cluster*_0 dengan total 377 data, *cluster*_1 memiliki keterangan *cluster* yang tingkat pembayaran listrik Tinggi dengan total 34 data dan *cluster*_2 memiliki keterangan tingkat pembayaran listrik Sedang dengan total 129 data. Dari proses *clustering* dapat dilihat tingkat penggunaan listrik dominan berada pada tingkat pembayaran listrik Rendah itu artinya sesuai dengan ekonomi yang ada di pagar alam karena mayoritas pendapatan masyarakatnya adalah dari petani dan UMKM. Sehingga dapatlah sebuah pola dari 3 *cluster* tersebut. Dibawah ini adalah tabel dari hasil cluster.

Tabel 1. Jumlah Cluster

Cluster	Tingkat	Jumlah
*Cluster*_0	Rendah	377
*Cluster*_1	Tinggi	34
*Cluster*_2	Sedang	129

3.2 Pengujian metode *elbow google colab*

Langkah selanjutnya akan dilakukan pengujian menggunakan *Elbow Method*. Sebuah plot yang menunjukkan nilai *inertia* (*sum squared distances of samples to their closest cluster center*) untuk setiap nilai k dari 1 hingga 12. Hasil *Elbow* dapat dilihat pada gambar dibawah ini.



Gambar 8. Hasil Pengujian *Elbow Method*

JURNAL SISTEM INFORMASI TGD
Volume 4, Nomor 3, Mei 2025, Hal 437-445
P-ISSN : 2828-1004 ; E-ISSN : 2828-2566
https://ojs.trigunadharma.ac.id/index.php/jsi

Kemudian didapatkan hasil *cluster* yang berbentuk siku, dari gambar diatas data penggunaan listrik rumah tangga sebanyak 540 *record* maka didapatkan hasil *cluster* yang tepat berjumlah 3 *cluster* (K=3) dengan nilai *sum of square error* yaitu 276486734427.498, bahwasannya hasil tersebut adalah Jumlah *Cluster* yang optimal yaitu biasanya terletak di titik *centorid* terdekat atau siku terdekat pada plot *elbow*.

4. KESIMPULAN

Penelitian ini menghasilkan 3 pola *cluster* dengan *cluster*_0 yaitu berjumlah 377 data memiliki tingkat penggunaan Rendah, *cluster*_1 berjumlah 34 data dengan keterangan tingkat penggunaan Sedang dan *cluster*_2 berjumlah 129 data dengan keterangan tingkat penggunaan Tinggi.

Kemudian hasil dari tahapan pengujian Elbow pada Google Colaboration (Google Colab) dengan bahasa pemrograman Python untuk menghitung hasil Sum Of Square Error (SSE) diperoleh K=3 dengan Nilai 276486734427.49817. K adalah cluster_0 yaitu berjumlah 377 data memiliki tingkat penggunaan Rendah, cluster_1 berjumlah 34 data dengan keterangan tingkat penggunaan Sedang dan cluster_2 berjumlah 129 data dengan keterangan tingkat penggunaan Tinggi. Berdasarkan hasil pengujian tersebut dari hasil clustering k-means pada Rapid Miner dengan jumlah 3 dapat di katakan sesuai dengan cluster k-means pada python menggunakan elbow method.

Dari hasil tersebut Diharapkan data yang di cluster akan dapat bermanfaat bagi instansi dan mempermudah seseorang dalam membaca sebuah data, serta dapat membantu menyederhanakan data dengan menggambarkan kelompok-kelompok umum, memungkinkan analis atau pengambil keputusan untuk bekerja dengan data yang lebih terstruktur.

UCAPAN TERIMAKASIH

Terima kasih disampaikan kepada pihak-pihak yang telah mendukung terlaksananya penelitian ini.

DAFTAR PUSTAKA

[1] K. Hidayah, "Membangun Bangsa pada Era Digital," *lan.go.id*, 2023. https://lan.go.id/?p=12800
[2] D. Suyanto, *Data Mining untuk Klasifikasi dan Klasterisasi Data*, Pertama. INFORMATIKA Bandung, 2019.
[3] Yuli Mardi, "Data Mining: Klasifikasi Menggunakan Algoritma C4 . 5 Data mining merupakan bagian dari tahapan proses Knowledge Discovery in Database (KDD) . Jurnal Edik Informatika," *J. Edik Inform.*, vol. 2, no. 2, pp. 213–219, 2019.
[4] B. D. Laraswati, "Kelebihan dan Kekurangan Jenis-Jenis Clustering dalam Data Science," *blog.algorit.ma*, 2023. https://blog.algorit.ma/jenis-clustering/
[5] A. Rangkuti, M. Prodi, P. Matematika, S. Utara, and Y. Yahfizham, "Pengenalan Algoritma Pemrograman Dasar Dalam Konteks Pembelajaran Pemrograman Awal," *J. Mat. dan Ilmu Pengetah. Alam*, vol. 1, no. 4, pp. 2987–5315, 2023, [Online]. Available: https://doi.org/10.59581/konstanta.v1i4.1714
[6] T. Amalina, D. Bima, A. Pramana, and B. N. Sari, "Metode K-Means Clustering Dalam Pengelompokan Penjualan Produk Frozen Food," *J. Ilm. Wahana Pendidik.*, vol. 8, no. 15, pp. 574–583, 2022, [Online]. Available: https://doi.org/10.5281/zenodo.7052276
[7] F. Febriansyah and S. Muntari, "Penerapan Algoritma K-Means untuk Klasterisasi Penduduk Miskin pada Kota Pagar Alam," *JISKA (Jurnal Inform. Sunan Kalijaga)*, vol. 8, no. 1, pp. 66–77, 2023, doi: 10.14421/jiska.2023.8.1.66-77.
[8] V. R. Prasetyo, H. Lazuardi, A. A. Mulyono, and C. Lauw, "Penerapan Aplikasi RapidMiner Untuk Prediksi Nilai Tukar Rupiah Terhadap US Dollar Dengan Metode Linear Regression," *J. Nas. Teknol. dan Sist. Inf.*, vol. 7, no. 1, pp. 8–17, 2021, doi: 10.25077/teknosi.v7i1.2021.8-17.
[9] R. Gelar Guntara, "Pemanfaatan Google Colab Untuk Aplikasi Pendeteksian Masker Wajah Menggunakan Algoritma Deep Learning YOLOv7," *J. Teknol. Dan Sist. Inf. Bisnis*, vol. 5, no. 1, pp. 55–60, 2023, doi: 10.47233/jteksis.v5i1.750.
[10] M. K. Rahmadhika and A. M. Thantawi, "Rancang Bangun Aplikasi Face Recognition Pada Pendekatan CRM Menggunakan Opencv Dan Algoritma Haarcascade," *IKRA-ITH Inform. J. Komput. dan Inform.*, vol. 5, no. 1, pp. 109–118, 2021, [Online]. Available: https://journals.upi-yai.ac.id/index.php/ikraith-informatika/article/view/921
[11] R. F. Setiono, I., Setiani, "Hubungan Antara Besarnya Daya Listrik Terpasang Dengan Banyaknya Pemakaian Listrik Dalam Skala Rumah Tangga.," *Stud. P., Elektro, T., Vokasi, S., Diponegoro, U., Pascasarjana, P., Diponegoro, U*, pp. 978–979, 2019.
[12] A. Aziz *et al.*, "Klastrisasi Data Tenaga Kerja Terbuka Menurut Provinsi Dengan Penggunaan Algoritma K - Means 1,2," vol. 10, no. 3, pp. 444–456, 2023.
[13] S. Setya Wiwaha, "Analisis Peluang Penghematan," pp. 47–58, 2017.
[14] Y. Yulia and N. Azwanti, "Penerapan Algoritma C4.5 Untuk Memprediksi Besarnya Penggunaan Listrik Rumah Tangga di Kota Batam," *J. RESTI (Rekayasa Sist. dan Teknol. Informasi)*, vol. 2, no. 2, pp. 584–590, 2018, doi: 10.29207/resti.v2i2.503.
[15] I. K. Sukesa, "CRISP DM Sebagai Salah Satu Standard untuk Menghasilkan Data Driven Decision Making yang Berkualitas," *djkn.kemenkeu.go.id*, 2022. https://www.djkn.kemenkeu.go.id/artikel/baca/15134/CRISP-DM-Sebagai-Salah-Satu-Standard-untuk-Menghasilkan-Data-Driven-Decision-Making-yang-Berkualitas.html
[16] R. Mubarak, M. Hanafi, and D. Sasongko, "Komparasi Performa Naive Bayes Gaussian dan K-NN Untuk Prediksi Kelulusan Mahasiswa dengan CRISP-DM," vol. 4, no. 6, pp. 2982–2991, 2024, doi: 10.30865/klik.v4i6.1924.
[17] F. Yanti, B. N. Sari, and S. Defiyanti, "IMPLEMENTASI ALGORITMA LSTM PADA PERAMALAN STOK OBAT (STUDI KASUS : PUSKESMAS BEBER)," vol. 8, no. 4, pp. 6082–6089, 2024.
[18] M. A. Sofyan, N. Rahaningsih, and R. D. Dana, "Deteksi Sms Spam Berbahasa Indonesia Menggunakan Algoritma Support Vector Machine," *J. Mhs. Tek. Inform.*, vol. 8, no. 3, pp. 3071–3079, 2024.
[19] R. H. Rachmat Hidayat, "Perbandingan Metode Naïve Bayes Dan Decision Tree C4.5 untuk Analisis Sentimen Produk Es Teh Indonesia di Media Sosial Twitter," *J. SISKOM-KB (Sistem Komput. dan Kecerdasan Buatan)*, vol. 7, no. 2, pp. 88–98, 2024, doi:

10.47970/siskom-kb.v7i2.607.
[20] D. Rafly, R. Insani, and A. Dzulkarnain, "Penerapan FP-Growth untuk Menentukan Rekomendasi Produk pada Kogu Coffee Shop Malang," *INTEGER J. Inf. Technol.*, vol. 9, no. 1, pp. 27–36, 2024.
[21] D. A. Dzulhijjah, M. B. Herlambang, M. Haifan, P. Studi, P. Insinyur, and T. Selatan, "Implementasi Framework CRISP-DM untuk Proses Data Mining Aplikasi Credit Scoring PT.XYZ," *Pros. Semin. Nas. Teknol. Inf. dan Bisnis*, vol. 1, no. 1, pp. 238–251, 2023.
[22] N. Putu, E. Merliana, and A. J. Santoso, "PROSIDING SEMINAR NASIONAL MULTI DISIPLIN ILMU & CALL FOR PAPERS UNISBANK (SENDI_U) Kajian Multi Disiplin Ilmu untuk Mewujudkan Poros Maritim dalam Pembangunan Ekonomi Berbasis Kesejahteraan Rakyat ANALISA PENENTUAN JUMLAH CLUSTER TERBAIK PADA METODE K-ME," pp. 978–979, 2019, [Online]. Available: https://www.unisbank.ac.id/ojs/index.php/sendi_u/article/view/3333
[23] N. A. Maori, "METODE ELBOW DALAM OPTIMASI JUMLAH CLUSTER PADA K-MEANS CLUSTERING," vol. 14, no. 2, pp. 277–287, 2023.