

PENERAPAN ALGORITMA K-MEANS CLUSTERING

Studi Kasus: Data Keluhan Pelanggan PT. PLN (Persero)

Sugrio Dwi Darmawan¹, Dewi Agushinta R.^{2*}, Murniyati³

Program Pascasarjana Magister Sistem Informasi¹

Sistem Informasi, Fakultas Ilmu Komputer dan Teknologi Informasi^{2,3}

Universitas Gunadarma

Email: sugrio.darmawan@gmail.com, dewiar@staff.gunadarma.ac.id

murni_rk@staff.gunadarma.ac.id

Received: February 10, 2022. **Revised:** February 25, 2022. **Accepted:** March 20, 2022. **Issue Period:** Vol.6 No.2 (2022), Pp. 327-340

Abstrak: Listrik merupakan kebutuhan masyarakat yang bisa dianggap kebutuhan sangat penting. Untuk menjaga kualitas pelayanan pelanggan perusahaan listrik Negara harus senantiasa menjaga mutu dan pelayanannya tidak terkecuali keluhan pelanggan. Penelitian ini membahas tentang penerapan algoritma *K-Means Clustering* pada data keluhan pelanggan. Data yang digunakan adalah data keluhan pelanggan selama tahun 2020. Analisis penelitian ini disajikan dalam 3 (tiga) tingkat *cluster* yaitu *cluster* keluhan tertinggi (C1), *cluster* keluhan menengah (C2), dan *cluster* keluhan rendah (C3). Data diolah menggunakan aplikasi Rapid Miner sebagai penyelesaiannya dan mendapatkan hasil akhir yaitu 2 jenis data keluhan berada di *cluster* tertinggi, 3 jenis keluhan di *cluster* menengah, dan 7 jenis keluhan di *cluster* terendah. Pengujian *cluster* menggunakan *Davies Bouldin Index* memiliki nilai sebesar 0,52 yang berarti antar anggota *cluster* memiliki kesamaan cukup baik. Hasil analisis dari penelitian ini dapat menjadi masukan untuk Perusahaan Listrik Negara agar menjadikan jenis laporan keluhan pelanggan dengan kategori tertinggi (C1) sebagai prioritas utama dalam meningkatkan pelayanan pelanggan guna menekan tingginya angka keluhan pada dua jenis laporan tertinggi yaitu PDPB dan Prabayar.

Kata kunci: Analisis; *Data Mining*; *K-Means*; *Clustering*; Keluhan Pelanggan

Abstract: Electricity is a basic need that plays a crucial role for the community. In order to maintain the quality of the customer service, Perusahaan Listrik Negara (PLN) is obligated to control the quality and service department, including the customer complaints. This study discusses the application of the *K-Means Clustering* algorithm to customer complaint data. The data is collected from customer complaint data during 2020. The analysis of this study uses 3 (three) cluster levels, which are the highest complaint type cluster (C1), the medium complaint type cluster (C2), and the low complaint type cluster (C3). The data is processed using the Rapid Miner application as a solution and resulted in two types of complaint data in the highest cluster, three types of complaints in the middle cluster, and seven types of complaints in the lowest cluster. Cluster testing using the *Davies Bouldin Index* has a value of 0.52 which means that the cluster members have fairly good similarities. The analysis result from this study can be valuable input for the State Electricity Company (PLN) to propose the customer complaint report with the highest category (C1) as a top priority in improving the customer service aspect to suppress the high number of complaints in the two highest types of reports, namely PDPB and Prepaid.

Keywords: Analysis; *Clustering*; Customer Complaints; *Data Mining*; *K-Means*



DOI: 10.52362/jisamar.v6i2.761

Ciptaan disebarluaskan di bawah [Lisensi Creative Commons Atribusi 4.0 Internasional](https://creativecommons.org/licenses/by/4.0/).

I. PENDAHULUAN

Seiring dengan kemajuan teknologi saat ini membuat manusia hidup dalam suasana yang nyaman dan praktis. Hal ini dimungkinkan dengan adanya energi listrik. Dengan energi listrik, berbagai jenis peralatan listrik dapat diubah menjadi energi putar, panas, cahaya, serta sinyal audio-video, sesuai kebutuhan.

Listrik merupakan kebutuhan masyarakat yang sangat penting. PT. PLN (Persero) sebagai BUMN yang menjalankan usaha distribusi energi listrik ke pelanggan dituntut untuk menjaga citra perusahaan (*corporate image*), jangan sampai menjadi implikasi negatif yang dapat mengurangi kepercayaan pelanggan kepada perusahaan, dan akhirnya akan mengurangi citra perusahaan di mata pelanggannya.

Citra perusahaan yang melekat pada benak pelanggan akan menambah kepuasan pelanggan yang mengakibatkan loyalitas pelanggan terhadap produk perusahaan [1] . Peningkatan penjualan, kesadaran dan image sebuah merek serta reputasi perusahaan mempengaruhi kepuasan pelanggan untuk membeli. Pada konteks ini, citra yang baik ditampilkan perusahaan kepada pelanggannya melalui reputasi atau merek maupun memberikan sikap dan kepercayaan yang tinggi dari pelanggan terhadap produk yang dapat membuat rasa puas dari pelanggan, dan akhirnya mengakibatkan atau meningkatkan pula loyalitas pelanggan tersebut.

Sebagai perusahaan yang sekaligus menyediakan produk/ jasa, PT. PLN (Persero) harus dapat memberikan kualitas pelayanan yang prima, yaitu pelayanan yang sangat baik dan melampaui harapan pelanggan sehingga dapat memuaskan para pelanggannya secara lebih baik, yang pada gilirannya dapat menumbuhkan loyalitas pelanggan, sehingga terjadi pembelian ulang [2].

PT. PLN (Persero) harus tanggap atas keluhan-keluhan konsumen seperti penCatatan meTer (Cater) oleh petugas yang tidak akurat, Perubahan Daya Pasang Baru (PDPB), pemuTusan sambungan (Tusbung), pemakaian tenaga listrik tidak sah, sistem penagihan yang kurang akurat, pelanggan selalu terancam dengan pemutusan listrik dan masih banyak lainnya. Banyak pelanggan listrik mengadukan keluhan yang tidak sedikit sehingga keluhan tersebut harus segera ditangani agar tidak menjadikan pengaruh atau pendapat buruk untuk Perusahaan Listrik Negara. Dengan meningkatkan pelayanan dan pembangunan di banyak sektor diharapkan dapat meningkatkan kualitas pelayanan perusahaan agar menekan banyaknya keluhan pelanggan pada suatu wilayah/ distribusi PLN saat ini sehingga pelanggan mendapatkan kepuasan dalam pelayanan dari Perusahaan Listrik Negara. Proses pengelompokkan data (*data clustering*) menjadi sangat penting dikarenakan peningkatan jumlah data yang cukup signifikan.

K-Means merupakan salah satu metode *data clustering* non-hirarki yang berusaha memartisi data yang ada ke dalam bentuk satu atau lebih *cluster/* kelompok. Metode ini memartisi data ke dalam *cluster/* kelompok sehingga data yang memiliki karakteristik sama dikelompokkan ke dalam satu *cluster* yang sama. Dalam rangka menemukan pola-pola dan informasi dari sekumpulan koleksi data yang besar, dapat dilakukan dengan menggunakan *data mining*. Secara sederhana *data mining* adalah penambangan atau penemuan informasi baru dengan mencari pola atau aturan tertentu dari sejumlah data yang sangat besar [3]. Dengan melihat permasalahan ini, maka perlu dilakukan penelitian terhadap keluhan-keluhan yang menjadi prioritas bagi manajemen PT. PLN (Persero) untuk mengetahui, menyelesaikan dan menekan angka keluhan pelanggan..

II. METODE DAN MATERI

2.1. Keluhan Pelanggan

Beberapa pandangan para ahli tentang pengertian keluhan. Keluhan pelanggan (*customer complaints*) adalah umpan balik (*feedback*) dari pelanggan yang ditunjukkan kepada perusahaan yang cenderung bersifat negatif. Umpan balik ini dapat dilakukan secara tertulis atau secara lisan [4]. Pengertian keluhan pelanggan adalah bentuk aspirasi pelanggan yang terjadi karena adanya ketidakpuasan terhadap suatu barang atau jasa. Macam-macam keluhan pelanggan pada dasarnya terbagi 2 (dua) yakni keluhan yang disampaikan lewat lisan dan keluhan yang disampaikan secara tertulis [5].

2.2. Data Mining

Perkembangan teknologi informasi telah memberikan kontribusi pada cepatnya pertumbuhan jumlah data yang dikumpulkan dan disimpan dalam basis data berukuran besar (*big data*). *Big data* adalah istilah yang menggambarkan volume data yang besar, baik data yang terstruktur maupun data yang tidak terstruktur. *Big data* memiliki potensi tinggi untuk mengumpulkan wawasan kunci dari informasi bisnis. *Big data* dapat



dianalisis untuk wawasan yang mengarah pada pengambilan keputusan dan strategi bisnis yang lebih baik. Sebuah metode atau teknik diperlukan untuk dapat merubah data tersebut menjadi sebuah informasi berharga atau pengetahuan yang bermanfaat untuk mendukung pengambilan keputusan. Suatu teknologi yang dapat digunakan untuk mewujudkannya adalah data mining. Belakangan ini data mining telah diimplementasikan ke dalam berbagai bidang, di antaranya dalam bidang bisnis atau perdagangan, bidang pendidikan, dan telekomunikasi. *Data mining* adalah kegiatan menemukan pola yang menarik dari data dalam jumlah besar, data dapat disimpan dalam database, *data warehouse*, atau penyimpanan informasi lainnya. *Data mining* berkaitan dengan bidang ilmu-ilmu lain, seperti *database system*, *data warehousing*, statistik, *machine learning*, *information retrieval*, dan komputasi tingkat tinggi. Selain itu, *data mining* didukung oleh ilmu lain seperti *Neural Network*, pengenalan pola, *spatial data analysis*, *image database*, *signal processing* [6]. *Data mining* didefinisikan sebagai proses menemukan pola-pola dalam data. Proses ini otomatis atau seringnya semi otomatis. Pola yang ditemukan harus penuh arti dan pola tersebut memberikan keuntungan, biasanya keuntungan secara ekonomi. Data yang dibutuhkan dalam jumlah besar [7].

2.3. Konsep Clustering

Clustering atau klasterisasi adalah salah satu alat bantu pada data mining yang bertujuan mengelompokkan obyek-obyek ke dalam *cluster-cluster*. *Cluster* adalah sekelompok atau sekumpulan obyek data yang similar satu sama lain dalam *cluster* yang sama dan dissimilar terhadap obyek-obyek yang berbeda *cluster*. Obyek akan dikelompokkan ke dalam satu atau lebih *cluster* sehingga obyek-obyek yang berada dalam satu *cluster* akan mempunyai kesamaan yang tinggi antara satu dengan lainnya. Obyek-obyek dikelompokkan berdasarkan prinsip memaksimalkan kesamaan obyek pada *cluster* yang sama dan memaksimalkan ketidaksamaan pada *cluster* yang berbeda. Kesamaan obyek biasanya diperoleh dari nilai-nilai atribut yang menjelaskan obyek data, sedangkan obyek-obyek data biasanya direpresentasikan sebagai sebuah titik dalam ruang multi dimensi [8].

2.4. Algoritma K-Means

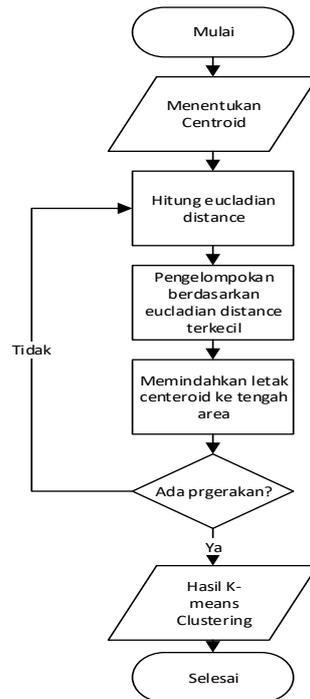
K-Means Clustering adalah suatu metode penganalisis data atau *metode data mining* yang melakukan proses pemodelan tanpa supervisi (*unsupervised*) dan merupakan salah satu metode yang melakukan pengelompokan data dengan sistem partisi. Ada dua jenis *data clustering* yang sering dipergunakan dalam proses pengelompokan data yaitu *hierarchical* dan *non-hierarchical*, dan *K-Means* merupakan salah satu metode data *clustering non-hierarchical* atau *partitional clustering*. Metode *K-Means Clustering* berusaha mengelompokkan data yang ada ke dalam beberapa kelompok, data dalam satu kelompok mempunyai karakteristik yang sama satu sama lainnya dan mempunyai karakteristik yang berbeda dengan data yang ada di dalam kelompok yang lain. Dengan kata lain, metode *K-Means Clustering* bertujuan untuk meminimalisasikan *objective function* yang diset dalam proses *clustering* dengan cara meminimalkan variasi antar data yang ada di dalam suatu *cluster* dan memaksimalkan variasi dengan data yang ada di *cluster* lainnya.

Algoritma *K-Means* bekerja dengan empat langkah, yaitu [9] :

1. Himpunan data yang akan diklasterisasi, dipilih sejumlah k objek secara acak sebagai *centroid* awal.
2. Setiap objek yang bukan *centroid* dimasukkan ke *cluster* terdekat berdasarkan ukuran jarak tertentu.
3. Setiap *centroid* diperbarui berdasarkan rata-rata dari objek yang ada di dalam setiap *cluster*.
4. Langkah kedua dan ketiga dilakukan secara diulang-ulang (diiterasi) sampai semua *centroid* stabil dalam arti semua *centroid* yang dihasilkan dalam iterasi saat ini sama dengan semua *centroid* yang dihasilkan pada iterasi sebelumnya.

Untuk memudahkan penentuan *cluster* dibutuhkan *flowchart* yang berfungsi sebagai rangkaian alur untuk mempermudah proses perhitungan berdasarkan data yang telah tersedia, seperti di gambar 1.





Gambar 1 Penentuan *Cluster* pada Algoritma *K-Means*

Berikut ini adalah tahapan algoritma *K-Means* :

1. Penentuan banyaknya *K*, *K* adalah jumlah *Cluster* yang akan dibentuk
2. Penentuan titik pusat *Cluster*

Dalam menentukan sebuah pusat *cluster* awal dilakukan pembangkitan bilangan *random* yang merepresentasikan urutan data input. Pusat awal *cluster* didapatkan dari data sendiri bukan dengan menentukan titik baru, yaitu dengan *random* pusat awal dari data.

3. Perhitungan jarak dengan pusat *cluster*

Untuk mengukur jarak antar data dengan pusat dengan *cluster* digunakan *Euclidian Distance*, algoritmanya adalah :

- a. Pemilihan nilai data dan nilai pusat *cluster*
- b. Penghitungan *Euclidian Distance* data dengan tiap pusat *cluster*

$$d(x_i, \mu_j) = \sqrt{(x_i - \mu_j)^2} \quad (1)$$

dengan

x_i : data kriteria

μ_i : *Centroid* pada *cluster* ke-*j*

4. Pengelompokan data

Jarak hasil perhitungan akan dilakukan perbandingan dan dipilih jarak terdekat antara data dengan pusat *cluster*. Jarak ini menunjukkan bahwa data tersebut berada dalam satu kelompok dengan pusat *cluster* terdekat. Adapun cara pengelompokan data tersebut adalah :

- a. Pemilihan nilai jarak tiap pusat *cluster* dengan data.
- b. Pencarian nilai jarak terkecil.
- c. Pengelompokan data dengan pusat *cluster* yang memiliki jarak terkecil.

5. Penentuan pusat *cluster* baru

Untuk mendapatkan pusat *cluster* baru bisa dihitung dari rata-rata nilai anggota *cluster* dan pusat *cluster*. Pusat *cluster* yang baru digunakan untuk melakukan iterasi selanjutnya, jika hasil yang didapatkan belum konvergen. Proses iterasi akan berhenti jika telah memenuhi maksimum iterasi yang dimasukkan oleh



user atau hasil yang dicapai sudah konvergen (pusat *cluster* baru sama dengan pusat *cluster* lama).

Algoritma penentuan pusat *cluster* :

- a. Pencarian jumlah anggota tiap *cluster*
- b. Penghitungan pusat baru dengan rumus:

$$\mu_j(t+1) = \frac{1}{N_{sj}} \sum_{j \in S_j} x_j \quad (2)$$

dengan :

$\mu_j(t+1)$: Centroid baru pada iterasi ke-1

N_{sj} : Banyak data pada *cluster* s_j

6. Kembali ke langkah 3, 4 dan 5.

Jika pada iterasi kedua, anggota *cluster* tidak ada yang berpindah ke *cluster* lain maka iterasi berhenti, tetapi jika ada anggota *cluster* yang berpindah ke *cluster* lain maka kembali ke langkah nomor 3, 4 dan 5. Iterasi berikutnya dilakukan sampai tidak ada anggota *cluster* yang berpindah ke *cluster* lain.

Hasil dari operasi *clustering* yang terbentuk selanjutnya akan dievaluasi menggunakan *Davies Bouldin Index* (DBI) dengan tahapan sebagai berikut:

1. *Sum of Square Within-cluster* (SSW)

Untuk mengetahui kohesi dalam sebuah *cluster* ke- i adalah dengan menghitung nilai dari *Sum of Square Within-cluster* (SSW). Kohesi didefinisikan sebagai jumlah dari kedekatan data terhadap titik pusat *cluster* dari sebuah *cluster* yang diikuti. Persamaan yang digunakan untuk memperoleh nilai *Sum of Square Within-cluster* adalah sebagai berikut :

$$SSW_i = \frac{1}{m_i} \sum_{j=i}^{m_i} d(x_j, c_i) \quad (3)$$

2. *Sum of Square Between-cluster* (SSB)

Perhitungan *Sum of Square Between-cluster* (SSB) bertujuan untuk mengetahui separasi antar *cluster*. Persamaan yang digunakan untuk menghitung nilai *Sum of Square Between-cluster* adalah :

$$SSB_{i,j} = d(c_i, c_j) \quad (4)$$

3. *Ratio* (Rasio)

Bertujuan untuk mengetahui nilai perbandingan antara *cluster* ke- i dan *cluster* ke- j . Untuk menghitung nilai rasio yang dimiliki oleh masing masing *cluster*, digunakan persamaan :

$$R_{ij} = \frac{SSW_i + SSW_j}{SSB_{i,j}} \quad (5)$$

4. *Davies Bouldin Index* (DBI)

Nilai rasio yang diperoleh dari persamaan (5) digunakan untuk mencari nilai *Davies Bouldin Index* (DBI) dengan menggunakan persamaan :

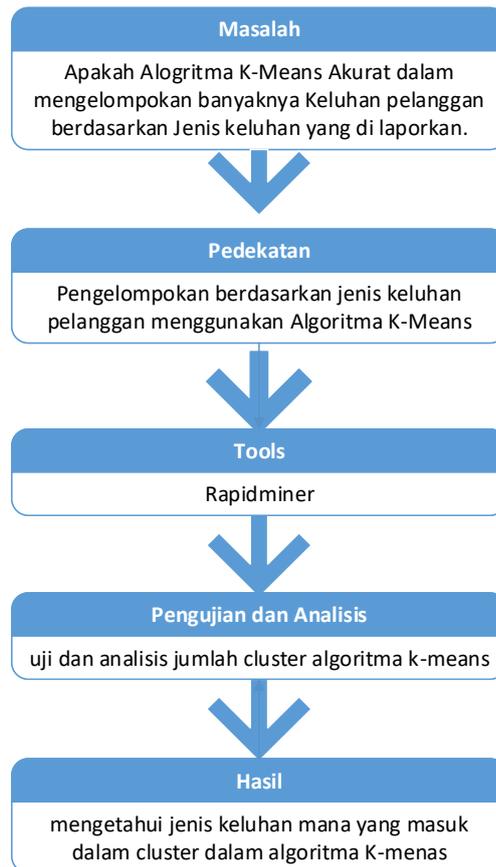
$$DBI = \frac{1}{k} \sum_{i=1}^k \max_{i \neq j} (R_{i,j}) \quad (6)$$

Dari persamaan tersebut, k adalah jumlah *cluster*. Semakin kecil nilai *Davies Bouldin Index* (DBI) yang diperoleh (non-negatif ≥ 0), maka semakin baik *cluster* yang diperoleh dari pengelompokan menggunakan algoritma *clustering* [10].

2.5. Kerangka Pemikiran

Kerangka penelitian disusun sesuai tujuan penelitian berdasarkan pada algoritma K-Means seperti gambar 2.





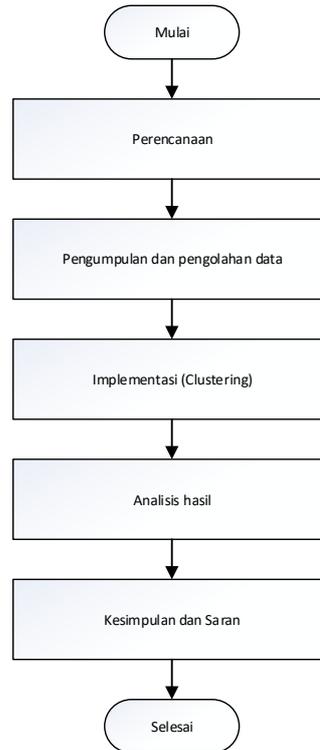
Gambar 2 Kerangka Penelitian

Dari kerangka pikir pada gambar 2 dapat dilihat bahwa dari masalah yang diangkat dari penelitian ini adalah keakuratan algoritma *K-Means* untuk menentukan banyaknya keluhan pelanggan per jenis laporan dengan pendekatan berdasarkan jenis laporannya seperti prabayar, cater, rekening, *invoice* dan sebagainya. *Tool* yang digunakan dalam melakukan *clustering* pada penelitian ini adalah *RapidMiner*. *Tool* ini dapat menjalankan algoritma *K-Means* serta pengujian *Davies Bouldin Index*. Pengujian menggunakan algoritma *K-Means* akan didapatkan jenis keluhan mana saja yang masuk ke dalam *cluster* tertinggi untuk dilakukan penindakan selanjutnya dari pihak terkait sehingga diharapkan dapat menekan banyaknya keluhan berdasarkan laporan tersebut.

2.6. Tahapan Penelitian

Tahapan yang dilakukan untuk penerapan metode *K-Means* untuk menentukan *cluster* pada jenis laporan keluhan pelanggan PT. PLN (Persero) dapat dilihat pada Gambar 3.





Gambar 3 Alur Metodologi Penelitianl

2.7. Tahap Pengumpulan Data dan Pengolahan Data

Implementasi *data mining* pada kasus klasifikasi wilayah/ distribusi terhadap banyaknya keluhan pelanggan sangat memerlukan data terkait sebagai penunjang analisis. Pada penelitian ini menggunakan sumber data sekunder yang diakses pada aplikasi internal perusahaan (aplikasi pengaduan dan keluhan terpadu). Data yang digunakan pada penelitian ini adalah data keluhan pelanggan berdasarkan jenis keluhannya pada bulan Januari sampai bulan Desember tahun 2020 di 22 wilayah/ distribusi PLN di Indonesia. Variabel yang digunakan adalah nilai jumlah keluhan pelanggan berdasarkan jenis keluhannya selama 12 bulan.

2.8. Tahap Clustering

Analisis *clustering* dalam prosedur klasifikasi data atau objek dalam suatu set data ke dalam sejumlah kelompok yang kesamaan datanya lebih besar dari kesamaan data kelompok lainnya. Analisis bertujuan untuk meminimalisir fungsi objektif yang diset dalam pemrosesan secara sederhana sehingga meminimalisir variasi pada suatu kelompok dan memaksimalkan variasi antar kelompok. Setelah semua data telah ditransformasi dengan menginisialisasikan ke dalam bentuk angka, maka data tersebut telah dapat dikelompokkan dengan menggunakan metode *K-Means*.

2.9. Tahap Evaluasi Clustering

Evaluasi clustering dilakukan dengan tujuan untuk mengetahui seberapa baik kualitas dari hasil *clustering*. Evaluasi hasil *clustering* yang digunakan adalah *Davies Bouldin Index* (DBI). Untuk mendapatkan nilai DBI, terlebih dahulu dihitung nilai *Sum of Square Within-cluster* (SSW), *Sum of Square Between-cluster* (SSB) dan *Ratio*.

2.10. Tahap Analisis Hasil

Pada tahap ini akan dianalisis data keluhan pelanggan berdasarkan laporan jenis keluhan. Data yang diperoleh diolah dengan menggunakan perhitungan bobot dari setiap indeks. Pada tahapan sebelumnya, data telah dikelompokkan ke dalam 3 (tiga) *cluster* yaitu tingkat jumlah keluhan pelanggan tinggi (C1), tingkat



jumlah keluhan pelanggan sedang (C2), dan jumlah keluhan pelanggan rendah (C3). Pada tahap inilah data akan dianalisis hasilnya.

III. PEMBAHASAN DAN HASIL

3.1. Hasil dan Pembahasan

Dalam proses *cluster*, sumber data utama yang digunakan adalah data rekap keluhan pelanggan per jenis laporan dari aplikasi pengaduan dan keluhan terpadu (APKT) dari bulan Januari sampai Desember tahun 2020. Data yang diterima dalam bentuk tabel pada file *excel* per bulan sehingga harus dilakukan pengumpulan data terlebih dahulu hingga 12 bulan. Tabel 1 merupakan hasil *export* data dari aplikasi bulan Januari 2020.

Jenis keluhan	Bulan											
	1	2	3	4	5	6	7	8	9	10	11	12
PDPB	9321	10389	12095	13967	12142	15459	23165	26945	29753	19578	18626	17603
Tusbung	5397	6479	5798	3517	4813	5906	5127	5675	4973	5400	5603	7231
Lain-Lain	268	255	295	263	171	274	361	376	938	3180	5217	11455
Cater	8820	8657	10127	14500	21159	19292	14005	9278	6711	5054	5005	6111
Pemakaian PTL Tidak Sah	534	578	616	434	363	465	578	533	551	556	544	505
Instalasi Listrik	9922	9296	9700	8191	6178	8447	956	7053	6733	7456	9281	11969
Rekening	670	585	705	1839	1782	2368	1678	1221	1661	831	812	834
Mutu dan Keandalan	3203	2861	3079	4148	4706	6335	4119	3527	3708	2498	2315	2222
Invoice	21	30	58	857	1171	1119	677	156	143	95	295	420
Non Transaksi	0	3	2	1	0	1	2	1	1	1	1	1
Prabayar	13574	11780	12034	11609	8933	12221	10414	11192	11855	13784	16128	19340
Calo atau Suap	25	28	24	26	21	25	27	28	40	30	18	21

Tabel 1 Data Keluhan Pelanggan Tahun 2020

Hasil akumulasi data berdasarkan kriteria yang telah ditentukan tertera pada tabel 2.

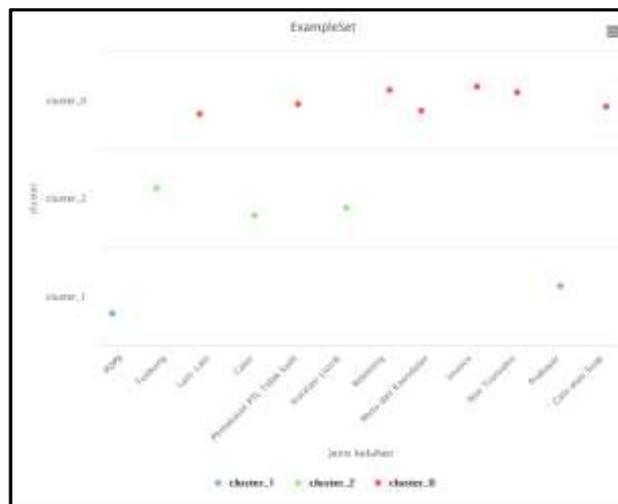
Tabel 2 Akumulasi Data Keluhan Pelanggan

Data Akumulasi	
Jenis keluhan	Jumlah Keluhan
PDPB	209043
Tusbung	65919
Lain-Lain	23053
Cater	128719
Pemakaian PTL Tidak Sah	6257
Instalasi Listrik	95182
Rekening	14986
Mutu dan Keandalan	42721



Data Akumulasi	
Jenis keluhan	Jumlah Keluhan
Invoice	5042
Non Transaksi	14
Prabayar	152864
Calo atau Suap	313

Nilai yang dihasilkan dari akumulasi data dari seluruh jenis keluhan, selanjutnya akan diproses dengan *software RapidMiner*. *Clustering* data menggunakan 3 *cluster* dengan menerapkan algoritma *K-Means*. Data yang telah diperoleh selanjutnya akan diuji ke *tool RapidMiner* dengan 3(tiga) *cluster* data seperti terlihat pada gambar 4.



Gambar 4 Hasil *Clustering* dengan Tool *RapidMiner*

3.2. Centroid Data

Pada penggunaan algoritma *K-Means clustering* dibutuhkan nilai titik pusat atau *centroid* yang digunakan sebagai data pusat dalam penentuan *cluster* yang diinginkan. Penentuan *cluster* dibagi ke dalam 3 tingkat yaitu *cluster* tingkat jenis keluhan pelanggan tinggi (C1), *cluster* tingkat jenis keluhan pelanggan menengah (C2), dan *cluster* tingkat jenis keluhan pelanggan rendah (C3). Penentuan nilai *centroid* awal dengan menggunakan nilai jumlah keluhan pelanggan pada 22 wilayah/ distribusi di wilayah kerja PT. PLN di Indonesia, sehingga diperoleh nilai *centroid* awal seperti tertera pada tabel 3.

Tabel 3 Centroid Awal (Iterasi 1)

Atribut	Nilai
C1 = Nilai <i>Max</i>	209043
C2 = Nilai <i>Average</i>	62009,42
C3 = Nilai <i>Min</i>	14

3.3. Clustering Data

Dengan penggunaan *centroid* awal, data dapat dikelompokkan menjadi 3 (tiga) *cluster* yang telah ditentukan sebelumnya. *Clustering* merupakan proses mengekstraksi data dengan cara mengambil nilai jarak terpendek dari hasil data yang diolah sehingga diperoleh nilai iterasi 1. Pada penelitian ini menghasilkan data



DOI: 10.52362/jisamar.v6i2.761

Ciptaan disebarluaskan di bawah [Lisensi Creative Commons Atribusi 4.0 Internasional](https://creativecommons.org/licenses/by/4.0/).

baru yakni keluhan pelanggan tinggi (C1) ada pada 2 jenis keluhan yaitu PDPB dan Prabayar, keluhan pelanggan sedang (C2) terdapat pada 4 jenis yaitu Tusbung, Cater, Instalasi Listrik serta Mutu dan Keandalan, dan keluhan pelanggan rendah (C3) terdapat pada 6 jenis keluhan yang diperoleh dari pengelompokan data pada iterasi 1 dengan menggunakan 3 (tiga) *cluster*. Hasil data lengkap dari perhitungan dan pengelompokan pada iterasi 1 dapat dilihat pada tabel 4 dan tabel 5.

Tabel 4 Perhitungan Jarak Pusat *Cluster* Iterasi 1

Iterasi 1						
Jenis Keluhan	Jumlah Keluhan	C1	C2	C3	Jarak Minimum	
PDPB	209043	0	147033,58	209029	0	
Tusbung	65919	143124	3909,58	65905	3909,58	
Lain-Lain	23053	185990	38956,42	23039	23039	
Cater	128719	80324	66709,58	128705	66709,58	
Pemakaian PTL Tidak Sah	6257	202786	55752,42	6243	6243	
Instalasi Listrik	95182	113861	33172,58	95168	33172,58	
Rekening	14986	194057	47023,42	14972	14972	
Mutu dan Keandalan	42721	166322	19288,42	42707	19288,42	
<i>Invoice</i>	5042	204001	56967,42	5028	5028	
Non Transaksi	14	209029	61995,42	0	0	
Prabayar	152864	56179	90854,58	152850	56179	
Calo atau Suap	313	208730	61696,42	299	299	

Tabel 5 Hasil Pengelompokan Data Iterasi 1

Jenis Keluhan	C1	C2	C3
PDPB	1		
Tusbung		1	
Lain-Lain			1
Cater		1	
Pemakaian PTL Tidak Sah			1
Instalasi Listrik		1	
Rekening			1
Mutu dan Keandalan		1	
<i>Invoice</i>			1
Non Transaksi			1
Prabayar	1		
Calo atau Suap			1

Proses klusterisasi pada algoritma *K-Means* akan terus berlangsung sampai pengelompokan iterasi sebelum dan iterasi sesudahnya sama. Pada iterasi 2 dan seterusnya proses klusterisasi data akan berlangsung kembali dengan *centroid* baru yakni hasil dari perhitungan iterasi 1 atau iterasi sebelumnya. Hasil



pengelompokan data pada iterasi 1 dapat dilihat pada tabel 5 dan selanjutnya perhitungan *clustering* iterasi 2 dapat dilihat pada tabel 6 sampai tabel 8.

Tabel 6 *Centroid* Data Iterasi 2

Atribut	Nilai
C1 = Nilai <i>Max</i>	180953,50
C2 = Nilai <i>Average</i>	83135,25
C3 = Nilai <i>Min</i>	8277,50

Tabel 7 Perhitungan Jarak *Cluster* Pusat Iterasi 2

Iterasi 2					
Jenis Keluhan	JUMLAH	C1	C2	C3	Jarak Minimum
PDPB	209043	28089,50	125907,75	200765,50	28089,50
Tusbung	65919	115034,50	17216,25	57641,50	17216,25
Lain-Lain	23053	157900,50	60082,25	14775,50	14775,50
Cater	128719	52234,50	45583,75	120441,50	45583,75
Pemakaian PTL Tidak Sah	6257	174696,50	76878,25	2020,50	2020,50
Instalasi Listrik	95182	85771,50	12046,75	86904,50	12046,75
Rekening	14986	165967,50	68149,25	6708,50	6708,50
Mutu dan Keandalan	42721	138232,50	40414,25	34443,50	34443,50
<i>Invoice</i>	5042	175911,50	78093,25	3235,50	3235,50
Non Transaksi	14	180939,50	83121,25	8263,50	8263,50
Prabayar	152864	28089,50	69728,75	144586,50	28089,50
Calo atau Suap	313	180640,50	82822,25	7964,50	7964,50

Tabel 8 Hasil Pengelompokan Data Iterasi 2

Jenis keluhan	C1	C2	C3
PDPB	1		
Tusbung		1	
Lain-Lain			1
Cater		1	
Pemakaian PTL Tidak Sah			1
Instalasi Listrik		1	
Rekening			1
Mutu dan Keandalan			1
<i>Invoice</i>			1
Non Transaksi			1
Prabayar	1		



Jenis keluhan	C1	C2	C3
Calo atau Suap			1

Pengelompokkan data pada iterasi 2 berbeda hasilnya dengan pengelompokkan data iterasi 1. Pada iterasi 2 diperoleh hasil pengelompokkan data yaitu *cluster* keluhan pelanggan tinggi (C1) sebanyak 2 jenis yakni PDPB dan Prabayar, *cluster* keluhan pelanggan sedang (C2) sebanyak 3 jenis yakni Tusbung, Cater, dan Instalasi Listrik, sedangkan *cluster* keluhan pelanggan rendah (C3) ada pada 7 jenis keluhan lainnya. Proses klasterisasi harus tetap berlangsung kembali dengan mencari nilai *centroid* atau titik pusat untuk perhitungan iterasi selanjutnya dari hasil klasterisasi data pada iterasi 2. Proses penentuan jarak terdekat dan klasterisasi data pada iterasi 3 dapat dilihat pada tabel 9 sampai tabel 11.

Tabel 9 *Centroid* Data Iterasi 3

Atribut	Nilai
C1 = Nilai <i>Max</i>	180953,50
C2 = Nilai <i>Average</i>	96606,67
C3 = Nilai <i>Min</i>	13198

Tabel 10 Perhitungan Jarak *Cluster* Pusat Iterasi 3

Jenis Keluhan	JUMLAH	Iterasi 3			Jarak Minimum
		C1	C2	C3	
PDPB	209043	28089,50	112436,33	195845	28089,50
Tusbung	65919	115034,50	30687,67	52721	30687,67
Lain-Lain	23053	157900,50	73553,67	9855	9855,00
Cater	128719	52234,50	32112,33	115521	32112,33
Pemakaian PTL Tidak Sah	6257	174696,50	90349,67	6941	6941,00
Instalasi Listrik	95182	85771,50	1424,67	81984	1424,67
Rekening	14986	165967,50	81620,67	1788	1788,00
Mutu dan Keandalan	42721	138232,50	53885,67	29523	29523,00
<i>Invoice</i>	5042	175911,50	91564,67	8156	8156,00
Non Transaksi	14	180939,50	96592,67	13184	13184,00
Prabayar	152864	28089,50	56257,33	139666	28089,50
Calo atau Suap	313	180640,50	96293,67	12885	12885,00

Tabel 11 Hasil Pengelompokkan Data Iterasi 3

Jenis Keluhan	C1	C2	C3
PDPB	1		
Tusbung		1	
Lain-Lain			1
Cater		1	
Pemakaian PTL Tidak Sah			1



Jenis Keluhan	C1	C2	C3
Instalasi Listrik		1	
Rekening			1
Mutu dan Keandalan			1
Invoice			1
Non Transaksi			1
Prabayar	1		
Calo atau Suap			1

Pada iterasi 3, proses klusterisasi atau pengelompokkan data yang telah dilakukan dengan menggunakan 3 cluster dari iterasi 3 bernilai sama dengan pengelompokkan data pada iterasi 2. Maka proses klusterisasi data pada data keluhan pelanggan berdasarkan jenis laporan berhenti.

3.4. Davies Bouldin Index (DBI)

Untuk mendapatkan nilai *Davies Bouldin Index*, terlebih dahulu dihitung nilai *Sum of Square Within-cluster*, *Sum of Square Between-cluster* dan *Ratio*. Setelah melakukan analisis perhitungan evaluasi cluster menggunakan DBI didapatkan nilai DBI sebesar 0.524. Hasil ini sama dengan hasil perhitungan yang dilakukan oleh tool RapidMiner.

IV. KESIMPULAN

Berdasarkan metode yang digunakan yaitu algoritma *K-Means*, didapatkan hasil cluster dengan jenis laporan keluhan tertinggi adalah PDPB dengan jumlah keluhan sebesar 209.043 dan Prabayar dengan jumlah keluhan sebesar 152.864. Cluster dengan jenis laporan keluhan sedang adalah Tusbung dengan jumlah keluhan sebesar 65.919, Cater dengan jumlah keluhan sebesar 128.719, dan Instalasi Listrik dengan jumlah keluhan sebesar 95.182. Cluster dengan jenis laporan rendah terdapat pada 7 jenis lainnya yaitu Lain-Lain, Pemakaian PTL Tidak Sah, Rekening, Mutu dan Keandalan, Invoice, Non Transaksi dan Calo atau Suap. Hasil analisis dari penelitian ini dapat dijadikan sebagai masukan bagi Perusahaan Listrik Negara (PLN), agar menjadikan jenis laporan keluhan pelanggan dengan kategori cluster tertinggi sebagai prioritas utama dalam meningkatkan pelayanan pelanggan guna menekan tingginya angka keluhan pada dua jenis laporan tertinggi yaitu PDPB dan Prabayar. Hasil pengujian dengan nilai *Davies Bouldin Index* sebesar 0,524 menunjukkan bahwa antar anggota cluster memiliki kesamaan yang cukup baik dengan DBI dikatakan baik jika nilainya mendekati angka 0.

REFERENSI

- [1] C. Fornell (1992) "A National Customer Satisfaction Barometer: The Swedish Experience". *Journal of Marketing*. Volume 56 Issue 1, 6–21. doi: 10.1177/002224299205600103.
- [2] Tjiptono (2006) *Strategi Pemasaran*. Yogyakarta: Andi Offset.
- [3] Davies and Paul Beynon (2004) *Database Systems Third Edition*. New York: Palgrave Macmillan.
- [4] S. J. Bell and J. A. Luddington (2006) "Coping with customer complaints". *Journal of Service Research*, Volume 8 No. 3, 221–233. doi: 10.1177/1094670505283785.
- [5] Kotler and Philip (2005) *Manajemen Pemasaran*. Jakarta: PT Indeks Kelompok Gramedia.
- [6] J. Han and M. Kamber. (2006). *Data Mining Concepts and Techniques Second Edition*. San Francisco: Morgan Kaufmann
- [7] I. H. Witten, E. Frank and M. A. Hall (2011) *Data Mining: Practical Machine Learning Tools and Techniques*. 3rd Edition, USA: Elsevier. ISBN 978-0-12-374856-0



DOI: 10.52362/jisamar.v6i2.761

Ciptaan disebarluaskan di bawah [Lisensi Creative Commons Atribusi 4.0 Internasional](https://creativecommons.org/licenses/by/4.0/).

- [8] Baskoro, H. (2010) "Implementasi Algoritma K-Means Menggunakan Data Penyewaan Alat Berat Untuk Melakukan Estimasi Nilai Outcome". Skripsi Program S1 Fakultas Ilmu Komputer. Jakarta: Universitas Pembangunan Nasional "Veteran".
- [9] Suyanto (2017) *Data Mining Untuk Klasifikasi Dan Klasterisasi Data*. Bandung: Informatika.
- [10] A. Bates and J. Kalita (2016) "Counting *Clusters* in Twitter Posts". In Proceedings of the Second International Conference on Information and Communication Technology for Competitive Strategies (ICTCS '16). Association for Computing Machinery, New York, NY, USA, Article 85, 1–9. <https://doi.org/10.1145/2905055.2905295>



DOI: 10.52362/jisamar.v6i2.761

Ciptaan disebarluaskan di bawah [Lisensi Creative Commons Atribusi 4.0 Internasional](https://creativecommons.org/licenses/by/4.0/).