

# Analisis Pengelompokan Gangguan TIK Pada Sistem Pencatatan Layanan Menggunakan Algoritma K-Means dan Metode Elbow

*Analysis of Incidents Segmentation Based on Service Recording System Data Using K-Means Algorithm and Elbow Method*

Arientawati<sup>1</sup>, Yuwan Jumaryadi<sup>2</sup>, Arief Wibowo<sup>3</sup>

<sup>1,2,3</sup>Magister Ilmu Komputer, Fakultas Teknologi Informasi, Universitas Budi Luhur

<sup>2</sup>Program Studi Sistem Informasi, Fakultas Ilmu Komputer, Universitas Mercu Buana

E-mail: <sup>1</sup>2111601866@student.budiluhur.ac.id, <sup>2</sup>yuwan.jumaryadi@mercubuana.ac.id,

<sup>3</sup>arief.wibowo@budiluhur.ac.id

## Abstrak

Pusintek merupakan unit pengelola tugas dan fungsi terkait TIK di instansi pemerintahan di Indonesia. Salah satu tugasnya berkaitan dengan fungsi penanganan gangguan TIK yang dilaporkan oleh Unit Pengguna. Pencatatan gangguan TIK menggunakan sistem pencatatan layanan dan gangguan berbasis *Customer Relationship Management* (CRM). Tujuan dari penelitian ini adalah untuk mengelompokkan jumlah laporan gangguan pada Unit Pelapor dan membantu divisi teknis dalam menganalisis seberapa sering gangguan yang dilaporkan tiap unit dan kategori gangguannya sehingga divisi teknis dapat merencanakan dan mempersiapkan pengambilan keputusan terkait tindak lanjut penyelesaian gangguan TIK pada instansi Unit Pelapor. Metode *Elbow* digunakan dalam proses penentuan jumlah cluster gangguan yang optimal. Pengujian dilakukan dengan membagi data gangguan TIK dan menghitung *cluster distance performance* sampai dengan 7 cluster. Tools yang digunakan untuk pengujian menggunakan *Rapidminer*. Hasil pengujian memperlihatkan bahwa penggunaan  $k=3$  akan menghasilkan cluster yang optimal untuk metode *K-Means Clustering* dan *Elbow*, dengan nilai *Average Centroid Distance* sebesar 69110.233 dan nilai *Davies Bouldin* sebesar 0.458.

Kata kunci: K-Means, Elbow, Gangguan TIK

## Abstract

*Pusintek is a task and function management unit related to ICT in government agencies in Indonesia. One of its duties relates to the function of handling ICT disturbances reported by the User Unit. The recording of ICT disturbances uses a Customer Relationship Management (CRM) based service and disturbance recording system. The purpose of this study is to classify the number of disturbance reports in the Reporting Unit and assist the technical division in analyzing how often disturbances are reported for each unit and the disturbance category so that the technical division can plan and prepare for decision-making related to the follow-up of ICT disturbance resolution at the Reporting Unit agency. The Elbow method is used in the process of determining the optimal number of disturbance clusters. The test was carried out by dividing the ICT interference data and calculating the cluster distance performance of up to 7 clusters. The tools used for testing use Rapidminer. The test results show that using  $k = 3$  will produce optimal clusters for the K-Means Clustering and Elbow method, with an Average Centroid Distance value of 69110.233 and a Davies Bouldin value of 0.458.*

Keywords: K-Means, Elbow, Gangguan TIK

## 1. PENDAHULUAN

Perkembangan teknologi informasi telah menghasilkan data yang cukup besar [1][2]. Data juga memiliki peran vital dalam pembuatan keputusan dalam suatu organisasi agar lebih bermakna [3][4]. Pusat Sistem Informasi dan Teknologi Keuangan (Pusintek) adalah Unit

Teknologi Informasi dan Komunikasi (TIK) yang memberikan layanan TIK kepada Unit Pelapor (Unit Pengguna) melalui beberapa jenis layanan di bidang TIK yang disusun dalam Katalog Layanan TIK. Apabila layanan TIK yang diberikan mengalami gangguan, maka tugas Pusintek adalah melakukan penyelesaian terhadap gangguan tersebut. Pusintek menerapkan IT *Service Management* (ITSM) dimana salah satu proses didalamnya yaitu *Incident Management* yang berfokus untuk memperbaiki kinerja layanan TIK organisasi atau instansi dalam waktu secepat mungkin agar dapat kembali normal dan tidak mengganggu jalannya bisnis organisasi.

Dalam proses implementasinya, pengguna melaporkan gangguan TIK ke Pusintek melalui website pelayanan Pusintek yang sudah terintegrasi dengan sistem pencatatan layanan dan gangguan TIK berbasis *Customer Relationship Management* (CRM). Gangguan yang telah direkam tersebut menjadi tiket gangguan yang akan dieskalasi ke Service Desk Pusintek sebagai analisis gangguan layer 1 atau tingkat pertama. Apabila gangguan tersebut tidak dapat diselesaikan oleh *Service Desk* dikarenakan keterbatasan knowledge, maka *Service Desk* akan mengeskalasi atau meneruskan tiket tersebut ke tim teknis Incident Management sesuai area atau kategori gangguannya. Kemudian tiket gangguan tersebut diselesaikan oleh tim teknis sesuai norma waktu prioritas gangguan dan diteruskan kembali ke *Service Desk* dan kemudian *Service Desk* meneruskan tiket tersebut ke pengguna untuk menginformasikan bahwa gangguan telah diselesaikan.

Dari sistem pencatatan layanan dan gangguan TIK diatas, penulis melakukan penarikan data untuk melakukan analisis dengan mengklusterisasi Unit Pelapor mana saja yang melaporkan gangguan dan seberapa seringnya gangguan dilaporkan serta jenis kategori gangguannya. Penulis melakukan penarikan data 6 (enam) bulan terakhir, periode Juli-Desember 2022, dimana terdapat total 12865 record tiket gangguan yang dilaporkan unit pelapor serta kategori jenis gangguannya.

Penelitian ini mengembangkan algoritma K-Means Clustering dan digabungkan dengan metode Elbow. Algoritma K-Means Clustering diharapkan dapat mengelompokkan seberapa seringnya unit pengguna melaporkan gangguan dan jenis gangguan yang dilaporkan untuk membantu tim teknis melakukan pengambilan keputusan, menganalisis tren gangguan TIK dan kategori gangguannya serta merencanakan alokasi penempatan pegawai di tim teknis *Incident Management* sebagai langkah penanganan lebih lanjut [5].

Clustering merupakan Teknik data mining untuk mengelompokkan data yang memiliki kemiripan ke dalam beberapa kluster, di mana data yang memiliki kemiripan dimasukkan kedalam kluster yang sama, sedangkan untuk menghitung kesamaan diluar kluster diminimalkan dengan menggunakan perhitungan jarak [6]. Clustering dan algoritma K-Means dipilih karena kesederhanaan algoritma dan efisiensinya [7]. Selain itu juga bahwa clustering termasuk *unsupervised classification* [8]. Sedangkan metode elbow bertujuan untuk memilih nilai k yang kecil dan masih memiliki nilai *withinss* yang rendah [9].

Penelitian yang dilakukan oleh Asroni, dkk [10], menggunakan algoritma K-Means untuk mengelompokkan nilai akademik mahasiswa. Pada penelitian tersebut, K-Means digunakan untuk menentukan cluster dalam pemilihan mahasiswa terbaik yang akan mengikuti lomba. Atribut data pengujian yang digunakan dalam penelitian tersebut adalah NIM, nilai mata kuliah kalkulus 1, fisika dasar, algoritma dan pemrograman 1, dan IPK. Adapun jumlah mahasiswa dalam penelitian ini adalah 124 mahasiswa. Berdasarkan perhitungan terhadap 124 Mahasiswa dihasilkan 4 cluster, dimana cluster 0 sebanyak 9 mahasiswa (7%) dengan IPK = 0.51, cluster 1 sebanyak 28 mahasiswa (23%) dengan IPK = 3.41, cluster 2 sebanyak 40 mahasiswa (32%) dengan IPK = 3.30, dan cluster 3 sebanyak 47 mahasiswa dengan IPK = 3.89 (38%). Berdasarkan perhitungan tersebut maka cluster 1 yang bisa digunakan untuk memilih mahasiswa agar bisa mengikuti lomba. Selain itu, penelitian yang dilakukan oleh Wandana, dkk [11] melakukan klusterisasi untuk mengetahui pola penyebaran penyakit berdasarkan kelas layanan. Data yang digunakan dalam penelitian tersebut berasal dari data rekam medis pasien periode Oktober sampai Desember 2019. Berdasarkan hasil clustering menggunakan algoritma K-Means didapatkan 3 cluster dengan kriteria yang berbeda. Berdasarkan penelitian tersebut, penggunaan data mining sebagai basis data dan teknik analisis cocok untuk menggali berbagai informasi potensial yang ada dalam database. Data mining merupakan salah satu tahapan dalam analisis berdasarkan penemuan

pengetahuan dalam database [12]. Algoritma K-Means merupakan metode clustering dalam data mining yang dapat menemukan karakteristik-karakteristik dari data yang ada [13]. Akan tetapi K-Means memiliki kelemahan dalam hal penentuan jumlah cluster terbaik [14]. Untuk mengatasi kelemahan tersebut, maka digunakan metode Elbow untuk menentukan optimasi banyaknya cluster yang akan digunakan dalam perhitungan clustering dengan algoritma K-Means [15].

## 2. METODE PENELITIAN

Data yang digunakan dalam penelitian ini berasal dari sistem pencatatan layanan dan gangguan TIK Pusintek (CRM). Penarikan data gangguan TIK dibatasi dalam rentang waktu selama 6 (enam) bulan terakhir, dari Juli sampai Desember tahun 2022. Gambaran tentang metode penelitian ditampilkan pada Gambar 1.



Gambar 1. Alur Penelitian

### 2.1 Identifikasi Masalah

Penelitian dimulai dari tahapan identifikasi permasalahan dimana terdapat kebutuhan untuk menentukan tingkat kesiapan tim teknis gangguan Pusintek yang menangani gangguan TIK di masing-masing area/kategori gangguan pada unit Pengguna. Karena itu dibutuhkan pengelompokan jumlah gangguan yang dilaporkan serta area/kategori gangguan TIK guna membantu tim teknis melakukan pengambilan keputusan, menganalisis tren gangguan TIK dan kategori gangguannya ataupun merencanakan alokasi penempatan pegawai di tim teknis Incident Management sebagai langkah penanganan lebih lanjut.

### 2.2 Studi Literatur

Tahapan studi literatur dilakukan dengan membaca beberapa artikel ilmiah yang membahas tentang penerapan data mining, serta tutorial cara perhitungan jumlah cluster dengan menggunakan Rapidminer, dan mencari beberapa penelitian yang berhubungan dengan clustering, analisis clustering dengan K-Means sehingga dapat memperbanyak sumber-sumber penelitian yang berkaitan.

### 2.3 Pengumpulan Data

Data-data pada penelitian ini didapatkan dengan melakukan penarikan data gangguan TIK yang dilaporkan pengguna dari sistem pencatatan layanan dan gangguan TIK Pusintek.

Adapun data yang digunakan dalam penelitian ini adalah gangguan TIK selama bulan Juli sampai Desember tahun 2022. Hasil penarikan data tersebut didapatkan jumlah total gangguan sebanyak 12825 tiket gangguan dengan 8 kategori atau gangguan seperti pada Tabel 1.

Tabel 1 Kategori Gangguan TIK

No	Kategori Gangguan	Jumlah Pelaporan Gangguan
1	Aplikasi	7386
2	Database	181
3	Hardware	696
4	Network	2779
5	Security	978
6	Software	263
7	Supporting Device	310
8	Layer 1 Service Desk	232
	Total	12825

Kemudian dari penarikan data tersebut didapatkan Unit Pengguna yang melaporkan gangguan selama enam bulan terakhir sebanyak 14 Unit Pengguna seperti pada Tabel 2.

Tabel 2 Unit Pelapor Gangguan TIK

No	Unit Pengguna	Jumlah Pelaporan Gangguan
1	BKF	58
2	BPPK	173
3	DJA	142
4	DJBC	2131
5	DJKN	622
6	DJP	5780
7	DJPB	1240
8	DJPK	59
9	DJPPR	141
10	ITJEN	135
11	LNSW	18
12	Sekretariat KNEKS	17
13	SETJEN	2298
14	STAF AHLI	11
	Total	12825

#### 2.4 Proses Perhitungan Metode Elbow

Agar dapat menghasilkan informasi mengenai cluster terbaik digunakan Metode Elbow dengan melihat persentase hasil perbandingan antara jumlah cluster (k) yang akan membentuk siku pada suatu titik [16][10]. Penggunaan metode Elbow terdapat dalam penentuan jumlah cluster terbaik yang biasanya digunakan pada algoritma K-Means Clustering. Adapun Persentase perhitungan yang dihasilkan akan digunakan sebagai pembanding antara jumlah cluster [9]. Adapun rumus SSE adalah [17]:

$$SSE = \sum_{k=1}^K \sum_{x_i \in S_k} \|x_i - c_k\|_2^2 \quad (1)$$

#### 2.5 K-Means

Algoritma dalam data mining untuk menemukan cluster dalam data adalah K-Means clustering, dimana jumlah *cluster* direpresentasikan sebagai k [18]. Suatu data dikelompokkan ke dalam satu kelompok atau cluster berdasarkan kemiripan atribut yang dimiliki dengan menerapkan pengukuran jarak [19].

Algoritma K-Means bekerja pada atribut numerik dan juga mempartisi data ke sejumlah kelompok [20]. Algoritma K-Means dimulai dengan memilih angka K secara acak serta pengambilan sebagian populasi sejumlah K untuk dijadikan sebagai titik pusat awal [21]. Euclidean Distance merupakan merupakan salah satu metode perhitungan jarak yang bisa digunakan dalam perhitungan K-Means. Berikut ini merupakan rumus perhitungan jarak menggunakan metode Euclidean [22]:

$$d_{\text{euc}}(x,y) = \sqrt{\sum_{j=1}^d (x_j - y_j)^2} \quad (2)$$

keterangan:

$x_j, y_j$  = nilai dari  $j$  atribut

Berikut ini merupakan prosedur dasar clustering K-Means [23]:

1. Tentukan jumlah cluster yang ingin dibentuk ( $k$ ).
2. Tentukan secara random  $k$  *centroid* awal.
3. Lakukan perhitungan jarak setiap data ke masing-masing *centroid*.
4. *Centroid* yang terdekat akan dipilih oleh setiap data.
5. Tentukan posisi *centroid* baru.
6. Jika posisi *centroid* baru berbeda dengan posisi *centroid* lama, maka ulangi langkah 3.

### 3. HASIL DAN PEMBAHASAN

#### 3.1 Hasil pemetaan data yang akan dihitung

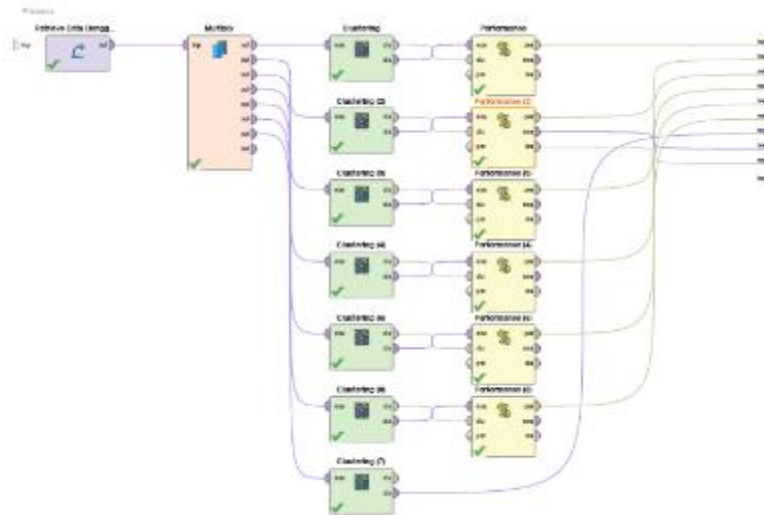
Dari 8 kategori gangguan yang didapat dipetakan kedalam 14 unit pelapor hasilnya didapat data yang akan diklusterisasikan seperti pada Tabel 3.

Tabel 3 Unit Pelapor Gangguan TIK

Unit Pelapor	Aplikasi	Database	Hardware	Network	Security	Software	Supporting Device	Layer 1 Service Desk	Grand Total
BKF	34	1	14	4	1	3	1	0	58
BPPK	97	5	0	54	9	4	1	3	173
DJA	69	2	35	13	4	12	5	2	142
DJBC	1380	30	36	257	249	57	78	44	2131
DJKN	396	23	14	117	41	22	4	5	622
DJP	3905	51	9	938	500	112	128	137	5780
DJPB	640	15	10	438	67	27	18	25	1240
DJPK	40	0	6	3	7	2	0	1	59
DJPPR	98	2	7	18	9	2	1	4	141
ITJEN	93	8	14	8	6	5	0	1	135
LNSW	15	0	0	1	1	1	0	0	18
Sekretariat KNEKS	15	0	0	0	0	2	0	0	17
SETJEN	599	44	547	926	84	14	74	10	2298
STAF AHLI	5	0	4	2	0	0	0	0	11

#### 3.2 Hasil perancangan perhitungan menggunakan Rapidminer

Data pemetaan jumlah gangguan per kategori beserta unit pelapor di *export* dan dilakukan perancangan perhitungan dengan rapidminer. Kemudian langkah selanjutnya adalah persiapan menghitung performance cluster terbaik dengan design perancangan perhitungannya seperti pada Gambar 2.



Gambar 2. Desain Perhitungan Cluster Performance pada Rapidminer

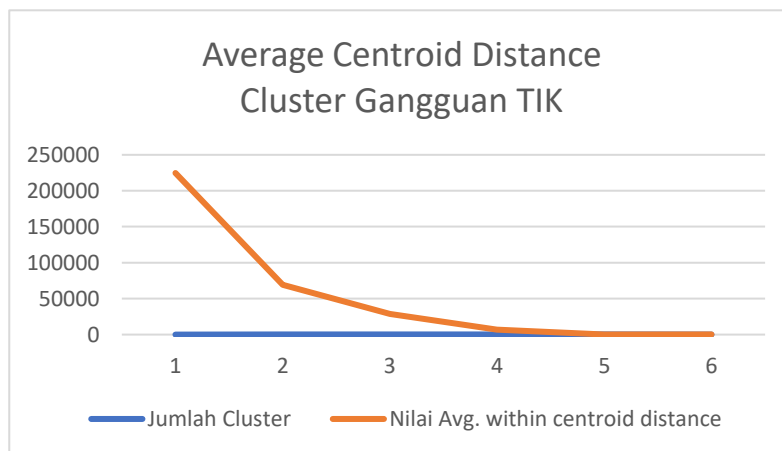
3.3 Penggunaan metode elbow untuk menentukan hasil cluster terbaik.

Penentuan cluster dilakukan dengan membandingkan dan mencari cluster terbaik dengan menggunakan perhitungan *cluster distance performance*. Penelitian ini mencoba untuk melakukan uji coba dengan membandingkan *cluster distance performance* dari mulai dua klaster ( $k = 2$ ) sampai dengan 7 klaster ( $k = 7$ ) dan mencari nilai rata-rata centroid distancenyanya seperti pada Tabel 4.

Tabel 4. Perhitungan *Cluster Distance Performance*

Uji Coba ke-	Jumlah Cluster	Nilai Avg. within centroid distance
1	2	224717.516
2	3	69110.233
3	4	28588.686
4	5	6939.087
5	6	150.238
6	7	265.482

Kemudian dari perhitungan cluster distance performance, maka dibuat grafik dari nilai rata-rata centroid distance dengan jumlah cluster yang diuji seperti pada Gambar 3.



Gambar 3. Grafik Perhitungan Performance Distance Cluster

Dapat dilihat pada Gambar 3 bahwa ada penurunan yang signifikan pada nilai  $K=2$  sampai

K=4, dan selanjutnya secara perlahan-lahan nilai K akan turun sampai nilai K tersebut stabil sehingga jumlah cluster k yang terbaik dalam penelitian ini adalah 3 cluster.

3.4 Hasil perhitungan metode K-Means

Setelah mendapatkan jumlah cluster yang terbaik diatas, maka penulis membagi jumlah laporan gangguan sebanyak 3 cluster. Selanjutnya adalah menjalankan proses perhitungan k = 3 dengan menggunakan Rapidminer, atribut yang dihitung adalah kategori gangguan dan unit pelapor.



Gambar 4. Rancangan Perhitungan k = 3

Berdasarkan gambar 3, dapat diperoleh hasil perhitungan seperti pada Tabel 5.

Tabel 5. Hasil Cluster

Cluster ke-	item
0	1
1	10
2	3
Total Item	14

Berdasarkan Tabel 5, didapatkan hasil pengelompokkan cluster yang beranggotakan beberapa item, dalam hal ini cluster adalah tingkat pelaporan gangguan dan unit pelapor. Cluster 0 sebanyak 1 item, kemudian cluster 1 sebanyak 10 items dan cluster 2 sebanyak 3 item. Hasil pengelompokkan cluster seperti pada Tabel 6.

Tabel 6. Hasil Pengelompokkan Cluster

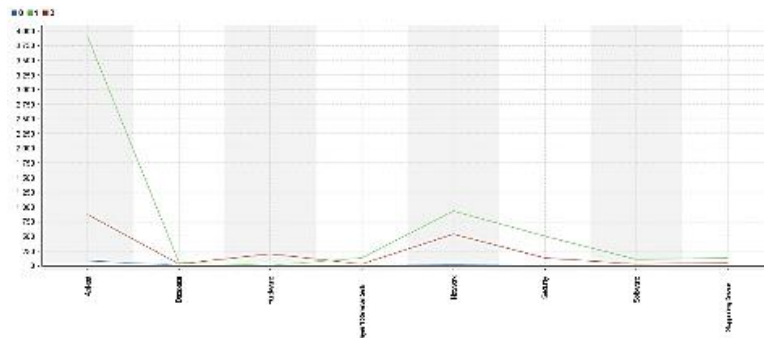
Id	Unit Pelapor	Cluster	Aplikasi	Database	Hardware	Network	Security	Software	Supporting Device	Layer 1 Service Desk	Grand Total
1	BKF	cluster_0	34	1	14	4	1	3	1	0	58
2	BPPK	cluster_0	97	5	0	54	9	4	1	3	173
3	DJA	cluster_0	69	2	35	13	4	12	5	2	142
4	DJBC	cluster_2	1380	30	36	257	249	57	78	44	2131
5	DJKN	cluster_0	396	23	14	117	41	22	4	5	622
6	DJP	cluster_1	3905	51	9	938	500	112	128	137	5780
7	DJPB	cluster_2	640	15	10	438	67	27	18	25	1240
8	DJPK	cluster_0	40	0	6	3	7	2	0	1	59
9	DJPPR	cluster_0	98	2	7	18	9	2	1	4	141
10	ITJEN	cluster_0	93	8	14	8	6	5	0	1	135
11	LNSW	cluster_0	15	0	0	1	1	1	0	0	18
12	Sekretariat KNEKS	cluster_0	15	0	0	0	0	2	0	0	17
13	SETJEN	cluster_2	599	44	547	926	84	14	74	10	2298
14	STAF AHLI	cluster_0	5	0	4	2	0	0	0	0	11

Hasil perhitungan centroid cluster dapat dilihat pada Tabel 7.

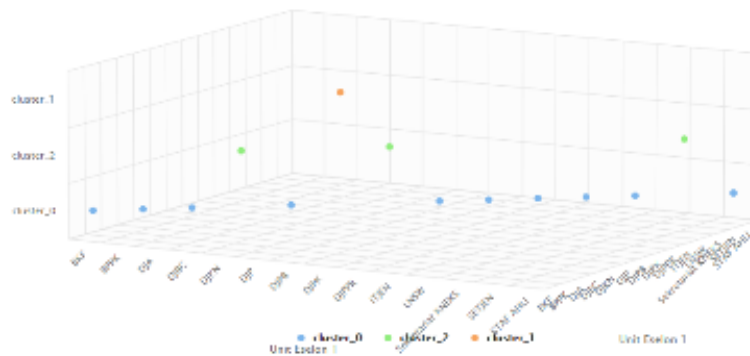
Tabel 7. Perhitungan Performance Distance Cluster

Attribute	Cluster_0	Cluster_1	Cluster_2
Aplikasi	86.2	3905	873
Database	4.1	51	29.667
Hardware	9.4	9	197.667
Layer 1 Service Desk	1.6	137	26.333
Network	22	938	540.333
Security	7.8	500	133.333
Software	5.3	112	32.667
Supporting Device	1.2	128	56.667

Grafik hasil pembagian plot cluster dan pembagian cluster dapat dilihat pada Gambar 5 dan Gambar 6.



Gambar 5. Plot Cluster di Rapidminer



Gambar 6. Grafik Scatter 3D

#### 4. KESIMPULAN DAN SARAN

Pada penelitian yang dilakukan menghasilkan 7 cluster dengan nilai dari *Average Within Centroid Distance* sebesar 54961.874 untuk total pelaporan gangguan. Berdasarkan hasil Clustering yang didapatkan diketahui adanya penurunan yang signifikan dari cluster  $k = 2$  sampai  $k = 3$  dengan nilai *Average Within Centroid Distance*  $k = 3$  sebesar 69110.233 menunjukkan bahwa cluster tersebut merupakan cluster yang terbaik. Hasil ini sama dengan penelitian yang dilakukan oleh Hartanti [15], dimana cluster optimal yang digunakan untuk Mengukur Kesiapan Siswa SMK Dalam Ujian Nasional adalah 3.

Hasil pengelompokan pada cluster 0 terdapat 10 Unit Pelapor yaitu BKF, BPPK, DJA, DJKN, DJPK, DJPPR, ITJEN, LNSW, Sekretariat KNEKS dan Staff Ahli. Pada cluster 1 terdapat satu Unit Pelapor yaitu DJP. Kemudian pada cluster 2 terdapat 3 Unit Pelapor yaitu DJPB, DJBC



dan SETJEN. Cluster 0 merupakan Unit Pengguna dengan tingkat pelaporan cenderung rendah dengan kisaran jumlah pelaporan sekitar 11 – 622 tiket pelaporan selama enam bulan terakhir, sedangkan cluster 2 merupakan Unit Pengguna dengan tingkat pelaporan menengah dengan kisaran jumlah pelaporan berkisar 1240 – 2298 tiket pelaporan selama enam bulan terakhir, dan cluster 1 merupakan Unit Pengguna dengan tingkat pelaporan cenderung tinggi dengan kisaran jumlah pelaporan sebesar 5780 tiket pelaporan selama enam bulan terakhir. Kemudian kategori gangguan dengan lima peringkat teratas dimulai dari pelaporan terbanyak yaitu gangguan terkait Aplikasi, Network/Jaringan, Security, Hardware dan Software.

Hasil penelitian ini dapat digunakan sebagai panduan agar Pusintek melakukan penempatan tim teknis gangguan TIK yang bertugas berdasarkan Unit Pelapor terbanyak (cluster 0) dan kategori gangguan terbanyak (Network) sehingga penanganan gangguan TIK tetap terjaga sesuai SLA yang telah dijanjikan sehingga dapat meningkatkan kepuasan pengguna terhadap pemenuhan penanganan gangguan TIK.

Saran yang diberikan untuk penelitian selanjutnya adalah rentang data penarikan dapat diperbesar yaitu satu atau dua tahun kebelakang agar lebih optimal. Keterbatasan penelitian ini adalah rentang penarikan data yang pendek disebabkan oleh lamanya waktu penarikan data, ukuran data yang besar dan keterbatasan waktu penelitian. Kemudian dalam pencarian cluster terbaik, dapat dilakukan ujicoba menggunakan alternatif metode lainnya selain metode Elbow antara lain seperti metode *Silhouette*, *Gap Statistic*, dan *Davies Bouldin index*.

#### DAFTAR PUSTAKA

- [1] Nurdin and R. N. Arifin, “Kontribusi Sistem Informasi Manajemen Berbasis Komputer Terhadap Kinerja Pegawai Pusdik Intelkam Polri Bandung,” *EduTech*, vol. 13, no. 2, p. 246, 2014, doi: 10.17509/edutech.v13i2.3107.
- [2] F. Febriantho, S. Samidi, G. Mikael, and E. Saputra, “Sistem Penentuan Paket Penjualan dengan Algoritma FP-Growth Serta Metode Up dan Cross Selling,” *J. Media Inform. Budidarma*, vol. 6, no. 4, p. 2269, 2022, doi: 10.30865/mib.v6i4.4800.
- [3] A. Nambiar and D. Mundra, “An Overview of Data Warehouse and Data Lake in Modern Enterprise Data Management,” *Big Data Cogn. Comput.*, vol. 6, no. 4, 2022, doi: 10.3390/bdcc6040132.
- [4] S. Suryani and M. Mustakim, “Estimasi Keberhasilan Siswa dalam Pemodelan Data Berbasis Learning Menggunakan Algoritma Support Vector Machine,” *Bull. Informatics Data Sci.*, vol. 1, no. 2, pp. 81–88, 2022.
- [5] I. Vhallah, Sumijan, and J. Santonyc, “Pengelompokan Mahasiswa Potensial Drop Out Menggunakan Metode Clustering K-Means,” *J. RESTI (Rekayasa Sist. dan Teknol. Informasi)*, vol. 2, no. 2, pp. 572–577, 2018.
- [6] R. Hidayati, A. Zubair, A. H. Pratama, and L. Indana, “Analisis Silhouette Coefficient pada 6 Perhitungan Jarak K-Means Clustering,” *Techno.Com*, vol. 20, no. 2, pp. 186–197, 2021.
- [7] M. Hermansyah, R. A. Hamdan, F. Sidik, and A. Wibowo, “Klasterisasi Data Travel Umroh di Marketplace Umroh.com Menggunakan Metode K-Means,” *J. Ilmu Komput.*, vol. 13, no. 2, p. 8, 2020, doi: 10.24843/jik.2020.v13.i02.p06.
- [8] Harwati, A. P. Alfiani, and F. A. Wulandari, “Mapping Student’s Performance Based on Data Mining Approach (A Case Study),” *Agric. Agric. Sci. Procedia*, vol. 3, pp. 173–177, 2015, doi: 10.1016/j.aaspro.2015.01.034.
- [9] N. H. Harani, C. Prianto, and F. A. Nugraha, “Segmentasi Pelanggan Produk Digital Service Indihome Menggunakan Algoritma K-Means Berbasis Python,” *J. Manaj. Inform.*, vol. 10, no. 2, pp. 133–146, 2020, doi: 10.34010/jamika.v10i2.2683.
- [10] A. Asroni and R. Adrian, “Penerapan Metode K-Means Untuk Clustering Mahasiswa Berdasarkan Nilai Akademik Dengan Weka Interface Studi Kasus Pada Jurusan Teknik Informatika UMM Magelang,” *Semesta Tek.*, vol. 18, no. 1, pp. 76–82, 2016, doi:

- 10.18196/st.v18i1.708.
- [11] J. Wandana, S. Defit, and S. Sumijan, “Klasterisasi Data Rekam Medis Pasien Pengguna Layanan BPJS Kesehatan Menggunakan Metode K-Means,” *J. Inf. dan Teknol.*, vol. 2, pp. 4–9, 2020, doi: 10.37034/jidt.v2i4.73.
  - [12] I. A. Nikmatun and I. Waspada, “Implementasi Data Mining untuk Klasifikasi Masa Studi Mahasiswa Menggunakan Algoritma K-Nearest Neighbor,” *J. SIMETRIS*, vol. 10, no. 2, pp. 421–432, 2019.
  - [13] R. Liu, “Data Analysis of Educational Evaluation Using K-Means Clustering Method,” *Comput. Intell. Neurosci.*, vol. 2022, no. Advanced Computational Intelligence for Smart Healthcare and Human Health Using Few-Shot Learning (Special Issue), pp. 1–10, 2022, doi: 10.1155/2022/3762431.
  - [14] L. Quiñones Huatangari, M. A. Cueva Ríos, and R. T. Huerta Camones, “Segmenting the eating behaviour of university students using the K-means algorithm,” *Bull. Electr. Eng. Informatics*, vol. 12, no. 4, pp. 2363–2371, 2023, doi: 10.11591/eei.v12i4.4543.
  - [15] N. T. Hartanti, “Metode Elbow dan K-Means Guna Mengukur Kesiapan Siswa SMK Dalam Ujian Nasional,” *J. Nas. Teknol. dan Sist. Inf.*, vol. 6, no. 2, pp. 82–89, 2020, doi: 10.25077/teknosi.v6i2.2020.82-89.
  - [16] D. A. I. C. Dewi and D. A. K. Pramita, “Analisis Perbandingan Metode Elbow dan Silhouette pada Algoritma Clustering K-Medoids dalam Pengelompokan Produksi Kerajinan Bali,” *Matrix J. Manaj. Teknol. dan Inform.*, vol. 9, no. 3, pp. 102–109, 2019, doi: 10.31940/matrix.v9i3.1662.
  - [17] A. A. Zaki Diab, M. A. Tolba, A. G. Abo El-Magd, M. M. Zaky, and A. M. El-Rifaie, “Fuel cell parameters estimation via marine predators and political optimizers,” *IEEE Access*, vol. 8, pp. 166998–167018, 2020, doi: 10.1109/ACCESS.2020.3021754.
  - [18] K. S. H. K. Al Atros, A. R. Padri, O. Nurdiawan, A. Faqih, and S. Anwar, “Model Klasifikasi Analisis Kepuasan Pengguna Perpustakaan Online Menggunakan K-Means dan Decision Tree,” *JURIKOM (Jurnal Ris. Komputer)*, vol. 8, no. 6, pp. 323–329, 2022, doi: 10.30865/jurikom.v8i6.3680.
  - [19] G. Feng, M. Fan, and Y. Chen, “Analysis and Prediction of Students’ Academic Performance Based on Educational Data Mining,” *IEEE Access*, vol. 10, pp. 19558–19571, 2022, doi: 10.1109/ACCESS.2022.3151652.
  - [20] A. J. M. S. Arockiam and E. S. Irudhayaraj, “Reclust: an efficient clustering algorithm for mixed data based on reclustering and cluster validation,” *Indones. J. Electr. Eng. Comput. Sci.*, vol. 29, no. 1, p. 545, 2022, doi: 10.11591/ijeecs.v29.i1.pp545-552.
  - [21] Y. Dharma Putra, M. Sudarma, and I. B. A. Swamardika, “Clustering History Data Penjualan Menggunakan Algoritma K-Means,” *Maj. Ilm. Teknol. Elektro*, vol. 20, no. 2, p. 195, 2021, doi: 10.24843/mite.2021.v20i02.p03.
  - [22] A. Wibowo, Moh Makruf, Inge Virdyna, and Farah Chikita Venna, “Penentuan Klaster Koridor TransJakarta dengan Metode Majority Voting pada Algoritma Data Mining,” *J. RESTI (Rekayasa Sist. dan Teknol. Informasi)*, vol. 5, no. 3, pp. 565–575, 2021, doi: 10.29207/resti.v5i3.3041.
  - [23] D. R. Ningrat, D. A. I. Maruddani, and T. Wuryandari, “Analisis Cluster Dengan Algoritma K-Means Dan Fuzzy C-Means Clustering Untuk Pengelompokan Data Obligasi Korporasi,” *J. Gaussian*, vol. 5, no. 4, pp. 641–650, 2016.