JURNAL KOMPUTER TEKNOLOGI INFORMASI SISTEM KOMPUTER
p-ISSN: 2962-3022 | e-ISSN: 2963-7104
Volume 5 No 1 Juni 2026
Hal: 200-211
http://ejurnal.lkpkaryaprima.id/index.php/juktisi
DOI: https://doi.org/10.62712/juktisi.v5i1.935

Pengelompokkan Jenis Surat Masuk di Dinas Komunikasi dan Informatika Menggunakan Metode K-Means Clustering

Sartika Siregar[1], Zulham[2*], Arif Rahman[3]

[1,2,3]Fakultas Teknik dan Ilmu Komputer, Rekayasa Perangkat Lunak, Universitas Dharmawangsa, Medan, Indonesia
Email: [1]sartikasiregar537@gmail.com, [2*]zulham@dharmawangsa.ac.id, [3]m.arif@dharmawangsa.ac.id
(*Email Corresponding Author: zulham@dharmawangsa.ac.id)
Received: 17 Maret 2026 | Revision: 30 Maret 2026 | Accepted: 1 April 2026

Abstrak

Pengelolaan administrasi surat masuk yang efektif merupakan salah satu faktor penting dalam meningkatkan kinerja dan pelayanan pada instansi pemerintah. Namun, pengolahan surat masuk secara manual sering kali kurang efisien karena jumlah data yang terus meningkat serta isi surat yang beragam, sehingga dapat menyebabkan kesulitan dalam proses pengarsipan, pencarian data, dan pengambilan keputusan. Oleh karena itu, diperlukan suatu metode yang mampu mengelompokkan data surat masuk secara otomatis. Salah satu teknik dalam data mining yang dapat digunakan adalah clustering dengan metode K-Means. Penelitian ini bertujuan untuk mengelompokkan surat masuk di Dinas Komunikasi dan Informatika Kota Medan berdasarkan kemiripan isi surat. Proses penelitian dilakukan melalui beberapa tahapan, yaitu preprocessing teks yang meliputi cleaning, tokenisasi, stopword removal, dan stemming, kemudian dilakukan pembobotan menggunakan metode TF-IDF sebelum dilakukan proses clustering menggunakan algoritma K-Means. Pengolahan data dilakukan menggunakan bahasa pemrograman Python pada platform Google Colaboratory (Google Colab). Hasil penelitian menunjukkan bahwa data surat masuk dapat dikelompokkan menjadi tiga klaster. Cluster pertama sebesar 3,9% berisi surat yang berkaitan dengan kegiatan perencanaan dan penyusunan dokumen strategis, cluster kedua sebesar 85,9% merupakan kelompok surat administrasi kepegawaian khususnya mengenai penunjukan jabatan fungsional, dan cluster ketiga sebesar 10,2% berisi surat yang berkaitan dengan kegiatan operasional dan kegiatan rutin instansi. Hasil pengelompokan ini menunjukkan bahwa sebagian besar surat masuk didominasi oleh administrasi kepegawaian. Dengan demikian, penerapan metode K-Means Clustering dapat membantu proses pengelompokan surat masuk secara lebih sistematis dan mendukung pengelolaan arsip yang lebih efektif dan efisien.

Kata Kunci: Data Mining, Clustering, K-Means Clustering, TF-IDF, Surat Masuk.

Abstract

Effective management of incoming mail administration is a crucial factor in improving performance and service delivery in government agencies. However, manual processing of incoming mail is often inefficient due to the ever-increasing volume of data and the diverse content, which can make archiving, data retrieval, and decision-making difficult. Therefore, a method capable of automatically grouping incoming mail data is needed. One data mining technique that can be used is K-Means clustering. This study aims to group incoming mail at the Medan City Communications and Informatics Office based on content similarity. The research process involved several stages: text preprocessing, including cleaning, tokenization, stopword removal, and stemming. Then, weighting was performed using the TF-IDF method, followed by clustering with the K-Means algorithm. Data processing was performed using the Python programming language on the Google Colaboratory (Google Colab) platform. The results showed that the incoming mail data could be grouped into three clusters. The first cluster, 3.9%, contains letters related to planning and strategic document preparation; the second cluster, 85.9%, is a group of personnel administration letters, specifically regarding the appointment to functional positions; and the third cluster, 10.2%, contains letters related to operational and routine agency activities. The results of this grouping indicate that most incoming letters are dominated by personnel administration. Thus, applying the K-Means Clustering method can help systematically group incoming letters and support more effective, efficient archive management.

***Keywords**: Data Mining, Clustering, K-Means Clustering, TF-IDF, Incoming Mail.*

1. PENDAHULUAN

Pengelolaan administrasi yang efektif merupakan salah satu faktor penting dalam meningkatkan kinerja dan pelayanan pada instansi pemerintah maupun swasta. Dalam era digital yang terus berkembang, pengolahan surat masuk secara manual menjadi kurang efektif karena jumlah data yang terus meningkat serta kompleksitas isi surat yang beragam. Kondisi ini dapat menyebabkan keterlambatan dalam pengarsipan, pencarian, dan pengambilan keputusan. Data Mining adalah serangkaian proses penambangan data dalam jumlah sangat besar untuk memperoleh informasi dari kumpulan data tersebut [1]. Informasi dihasilkan dengan mengekstraksi dan mencari pola-pola yang sangat penting dari suatu kumpulan data atau basis data. Salah satu teknik populer dalam data mining adalah

JURNAL KOMPUTER TEKNOLOGI INFORMASI SISTEM KOMPUTER
p-ISSN: 2962-3022 | e-ISSN: 2963-7104
Volume 5 No 1 Juni 2026
Hal: 200-211
http://ejurnal.lkpkaryaprima.id/index.php/juktisi
DOI: https://doi.org/10.62712/juktisi.v5i1.935

clustering, yaitu proses pengelompokan data ke dalam beberapa kelompok berdasarkan kemiripan karakteristik. Clustering membantu dalam mengidentifikasi pola dan struktur dalam data yang sebelumnya tidak terlihat atau tersembunyi. K-Means merupakan salah satu metode data clustering non-hierarki yang berusaha mempartisi data yang ada ke dalam satu atau lebih cluster atau kelompok, sehingga data yang memiliki karakteristik yang sama dikelompokkan ke dalam satu cluster dan data yang mempunyai karakteristik yang berbeda dikelompokkan ke dalam kelompok yang lainnya. K-Means Clustering bekerja dengan membagi data ke dalam beberapa cluster berdasarkan kemiripan atribut menggunakan metode partisi yang mengacu pada nilai rata-rata (means) setiap cluster [2][3]. Algoritma ini melakukan iterasi untuk menentukan pusat cluster (centroid) dan mengelompokkan data ke cluster dengan jarak terdekat terhadap centroid tersebut, Proses ini berulang hingga nilai centroid stabil dan tidak berubah.

Dalam konteks pengelolaan surat masuk, penerapan clustering dapat dimanfaatkan untuk mengelompokkan surat berdasarkan tema, tingkat prioritas, sumber instansi, atau karakteristik isi surat. Pengelompokan ini penting karena surat yang diterima setiap hari umumnya memiliki variasi topik, tujuan, dan urgensi yang berbeda. Tanpa sistem klasifikasi yang terstruktur, petugas administrasi akan membutuhkan waktu lebih lama untuk memilah surat secara tepat. Akibatnya, proses disposisi surat kepada bagian terkait dapat mengalami keterlambatan. Dengan memanfaatkan algoritma K-Means, data surat masuk dapat diolah secara otomatis berdasarkan atribut tertentu yang telah ditentukan. Atribut tersebut dapat berupa tanggal surat, asal surat, perihal, jenis surat, maupun tingkat kepentingan. Hasil pengelompokan kemudian dapat membantu instansi dalam menyusun pola penanganan surat yang lebih sistematis dan terarah. Selain itu, klaster yang terbentuk juga memberikan gambaran mengenai dominasi jenis surat yang paling sering diterima dalam periode tertentu.

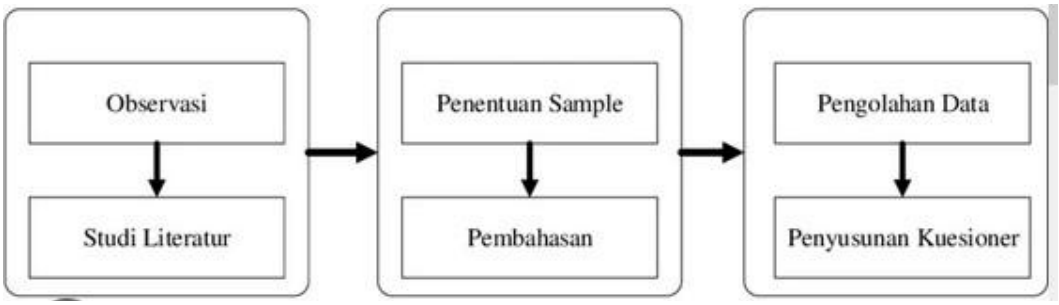
Penerapan metode K-Means pada data surat masuk tidak hanya berfungsi untuk pengelompokan, tetapi juga mendukung efisiensi kerja administrasi secara menyeluruh. Melalui hasil klaster, instansi dapat mengetahui kecenderungan distribusi surat dan menyesuaikan alur kerja sesuai kebutuhan. Informasi ini berguna untuk mempercepat proses pencatatan, pengarsipan, dan penelusuran kembali dokumen saat dibutuhkan. Di sisi lain, pengelompokan yang akurat dapat membantu pimpinan dalam mengambil keputusan yang lebih cepat karena surat dengan karakteristik serupa telah terkonsentrasi dalam kelompok tertentu. Sistem ini juga dapat mengurangi kesalahan manusia dalam proses pemilahan surat yang dilakukan secara manual. Dengan demikian, penggunaan data mining melalui algoritma K-Means menjadi solusi yang relevan untuk meningkatkan efektivitas pengelolaan surat masuk. Pendekatan ini sejalan dengan kebutuhan instansi modern yang menuntut kecepatan, ketepatan, dan efisiensi dalam pengelolaan administrasi. Oleh karena itu, penelitian mengenai penerapan K-Means Clustering pada pengolahan surat masuk menjadi penting untuk dilakukan.

2. METODOLOGI PENELITIAN

Penelitian ini dilakukan di Dinas Komunikasi dan Informatika, yang menjadi objek utama pengumpulan data surat masuk. Lokasi ini dipilih karena merupakan instansi yang secara rutin menerima berbagai jenis surat masuk yang akan dianalisis menggunakan metode K-Means Clustering agar dapat mengelompokkan jenis surat secara otomatis dan efisien [4].

Jenis data yang digunakan dalam penelitian ini adalah data kuantitatif berupa data surat masuk yang memuat atribut-atribut seperti nomor surat, tanggal surat, pengirim, jenis surat, dan isi ringkas. Sumber data primer diperoleh langsung dari arsip digital surat masuk di Dinas Komunikasi dan Informatika selama tahun 2025. Selain itu, data sekunder berupa laporan dan dokumen pendukung pengelolaan surat juga digunakan untuk mendukung analisis. Secara garis besar tahapan pada penelitian ini akan dibagi menjadi beberapa bagian yang utama yaitu, observasi, studi literatur, penentuan sampel, pembahasan, pengelohan data dan penyusunan kuesioner.

Tahapan pelaksanaan pada penelitian ini ditunjukan pada gambar 1 di bawah



Gambar 1. Tahapan Penelitian

Dari Gambar 1. Tahapan Penelitian dijelaskan bahwa penelitian memiliki beberapa bagian yang utama yaitu, observasi, studi literatur, penentuan sampel, pembahasan, pengelohan data dan penyusunan kuesioner[5].

JURNAL KOMPUTER TEKNOLOGI INFORMASI SISTEM KOMPUTER
p-ISSN: 2962-3022 | e-ISSN: 2963-7104
Volume 5 No 1 Juni 2026
Hal: 200-211
http://ejurnal.lkpkaryaprima.id/index.php/juktisi
DOI: https://doi.org/10.62712/juktisi.v5i1.935

1. Observasi: Observasi adalah tahap awal untuk mengamati secara langsung kondisi di lapangan. Tujuannya untuk memahami masalah yang terjadi, mengidentifikasi fenomena, serta mengumpulkan informasi awal yang relevan dengan topik penelitian.
2. Studi Literatur: Studi literatur dilakukan dengan mencari dan mempelajari referensi seperti buku, jurnal, artikel, dan penelitian sebelumnya. Tahap ini bertujuan untuk memperkuat dasar teori, mengetahui penelitian terdahulu, serta membantu merumuskan kerangka penelitian.
3. Penentuan Sampel: Pada tahap ini, peneliti menentukan siapa atau apa yang akan dijadikan sampel penelitian. Sampel dipilih dari populasi tertentu dengan metode tertentu (acak, purposive, dll.) agar dapat mewakili keseluruhan populasi.
4. Pengolahan Data: Setelah data terkumpul, langkah berikutnya adalah mengolah data tersebut. Proses ini meliputi pengelompokan, tabulasi, analisis statistik, hingga interpretasi data agar menjadi informasi yang bermakna.
5. Pembahasan: Pembahasan adalah tahap untuk menjelaskan hasil analisis data. Di sini peneliti mengaitkan hasil penelitian dengan teori atau penelitian sebelumnya, serta menjawab rumusan masalah yang telah ditentukan.
6. Penyusunan Kuesioner: Penyusunan kuesioner dilakukan untuk membuat instrumen pengumpulan data. Pertanyaan disusun secara sistematis, jelas, dan sesuai dengan tujuan penelitian agar responden dapat memberikan jawaban yang akurat.

2.2.1 Analisis Karakteristik Hasil Clustering

Analisis karakteristik cluster dilakukan setelah proses K-Means mencapai kondisi stabil (konvergen). Pada tahap ini, hasil pengelompokan tidak hanya dilihat dari nilai jarak atau centroid, tetapi juga dianalisis berdasarkan atribut deskriptif lainnya untuk memahami pola dan kecenderungan setiap kelompok surat. Meskipun proses clustering hanya menggunakan atribut isi ringkas surat, atribut administratif seperti tanggal penerimaan dan sifat surat digunakan untuk memperkuat interpretasi hasil pengelompokan. Hal ini dilakukan untuk memastikan bahwa cluster yang terbentuk memiliki makna substantif dalam konteks pengelolaan surat masuk [6][7].

3. HASIL DAN PEMBAHASAN

2.1 Pengumpulan Data

Data yang diolah dalam penelitian ini merupakan data surat masuk pada Dinas Komunikasi dan Informatika Kota Medan. Data diperoleh melalui metode dokumentasi, yaitu dengan mengumpulkan arsip surat masuk digital yang tercatat dalam agenda surat masuk tahun 2025. Data tersebut memuat informasi penting seperti tanggal penerimaan surat, sifat surat, serta isi ringkas surat. Data surat masuk yang telah dikumpulkan selanjutnya digunakan sebagai objek penelitian untuk dianalisis dan dikelompokkan menggunakan metode K-Means Clustering [8][9]. Pada penelitian ini, digunakan 20 data surat masuk sebagai sampel penelitian. Contoh data surat masuk yang digunakan dapat dilihat pada Tabel 3.1.

Tabel 3.1 Data Surat Masuk

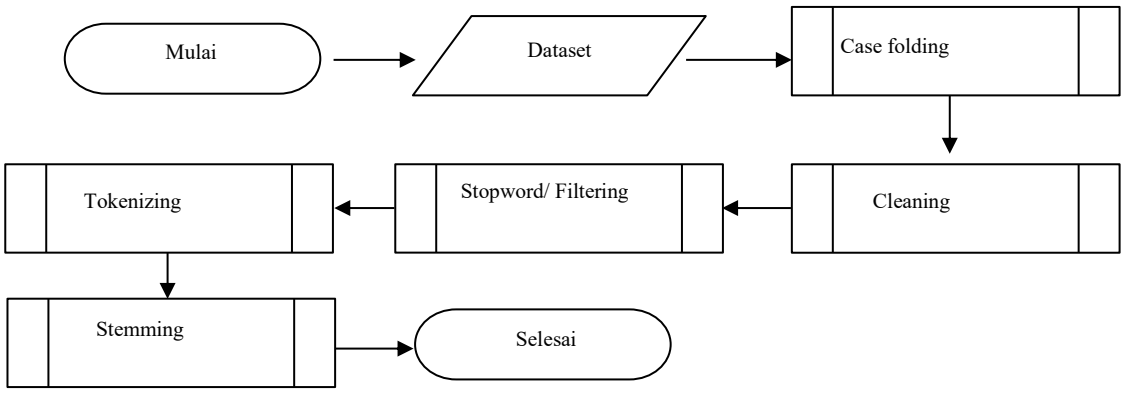
No.	Tanggal Penerimaan Surat	Sifat Surat	Isi Ringkas
1	02/01/2025	Biasa	Undangan Acara Open House Camat Medan Tuntungan dalam rangka Tahun Baru 2025
2	06/01/2025	Penting	Undangan rapat koordinasi hasil rapat internal Komisi-1 DPRD Kota Medan terhadap OPD yang menjadi mitra kerja Komisi-1 DPRD
3	08/01/2025	Biasa	Pelaksanaan penyampaian orientasi penyusunan dokumen perencanaan di Tahun 2025 kepada Perangkat Daerah
4	08/01/2025	Segera	Pemberian Penghargaan bagi SDM Keamanan Siber dan Sandi Tahun 2025
5	08/01/2025	Biasa	Penerbitan KPPD TA 2025
6	09/01/2025	Biasa	Tanggapan surat penunjukan pejabat fungsional perencana
7	09/01/2025	Biasa	Penawaran LCD Videowall, Led Videotron dan Interaktive Board
8	09/01/2025	Biasa	Rapat Pembahasan Ranperwal tentang pemenuhan kebutuhan sumber Daya Manusia melalui penggunaan tenaga alih Daya dilingkungan pemerintah kota Medan
9	09/01/2025	Biasa	Permohonan perbaikan handy talky (HT)

JURNAL KOMPUTER TEKNOLOGI INFORMASI SISTEM KOMPUTER
p-ISSN: 2962-3022 | e-ISSN: 2963-7104
Volume 5 No 1 Juni 2026
Hal: 200-211
http://ejurnal.lkpkaryaprima.id/index.php/juktisi
DOI: https://doi.org/10.62712/juktisi.v5i1.935

No.	Tanggal Penerimaan Surat	Sifat Surat	Isi Ringkas
10	09/01/2025	Biasa	Penyusunan perjanjian kinerja perangkat daerah di Lingkungan Pemerintah Kota Medan
11	10/01/2025	Biasa	Penawaran program factory visit
12	10/01/2025	Biasa	Undangan rapat pembahasan Finalisasi persiapan pelaksanaan penandatanganan Naskah perjanjian Hibah barang milik Daerah dan berita acara serah Terima antara pemerintah kota Medan dengan kantor kementerian Agama Kota Medan
13	10/01/2025	Biasa	Surat Panggilan Sidang
14	13/01/2025	Biasa	Mohon bantuan tenagapelaksanaan pembongkaran bangunan lokasi Jalan T.Amir Hamzah Kelurahan Helvetia Timur Kecamatan Medan Helvetia
15	13/01/2025	Biasa	Pemberian Penghargaan bagi SDM Keamanan Siber dan Sandi Tahun 2025 (Disposisi Sekda Kota Medan)
16	13/01/2025	Biasa	Permohonan Bantuan Sterilisasi
17	13/01/2025	Biasa	Mohon bantuan tenagapelaksanaan pembongkaran bangunan lokasi Jalan Bougenville Kelurahan Simpang Selayang Kecamatan Medan Tuntungan
18	13/01/2025	Biasa	Pengajuan kerja sama publikasi pemberitaan dan penayangan iklan pembangunan
19	14/01/2025	Biasa	Mohon bantuan tenaga pelaksanaan pembongkaran bangunan lokasi Jalan Ngumban Surbakti Gang Bahagia Kelurahan Tanjung Sari Kecamatan Medan Selayang
20	14/01/2025	Biasa	Mohon bantuan tenaga pelaksanaan pembongkaran bangunan lokasi Jalan Kayu Putih Sudut Gang Lingkungan VII Kelurahan Tanjung Mulia Hilir Kecamatan Medan Deli

2.2. Flowchart Preprocessing

Preprocessing bertujuan untuk menyusun teks tidak terstruktur menjadi teks terstruktur sehingga dapat digunakan untuk proses selanjutnya. Tahapan dalam preprocessing adalah case folding, cleansing, stopword/filtering, tokenizing dan stemming. Berikut ini merupakan flowchart preprocessing [10].



Gambar 2. Flowchart Preprocessing

2.3. Stemming

Stemming bertujuan untuk menangkap kata dasar yang dimiliki oleh kata kerja yang telah menerima imbuhan kata atau keterangan lain pada kata dasar. Dalam implementasinya, hasil stemming ini dicek pada daftar kata dasar yang ada. Berikut ini merupakan contoh penerapan stemming dapat dilihat pada Tabel 3.2.1

JURNAL KOMPUTER TEKNOLOGI INFORMASI SISTEM KOMPUTER
p-ISSN: 2962-3022 | e-ISSN: 2963-7104
Volume 5 No 1 Juni 2026
Hal: 200-211
http://ejurnal.lkpkaryaprima.id/index.php/juktisi
DOI: https://doi.org/10.62712/juktisi.v5i1.935

Tabel 1. Hasil Stemming

No.	Tokenizing	Stemming
1	[undangan, acara, open, house, camat, medan, tuntungan, rangka, tahun, baru]	[undang, acara, open, house, camat, medan, tuntungan, rangka, tahun, baru]
2	[undangan, rapat, koordinasi, hasil, rapat, internal, komisi, dprd, kota, medan, opd, mitra, kerja, komisi, dprd]	[undang, rapat, koordinasi, hasil, rapat, internal, komisi, dprd, kota, medan, opd, mitra, kerja, komisi, dprd]
3	[pelaksanaan, penyampaian, orientasi, penyusunan, dokumen, perencanaan, tahun, perangkat, daerah]	[laksana, sampai, orientasi, susun, dokumen, rencana, tahun, perangkat, daerah]
4	[pemberian, penghargaan, sdm, keamanan, siber, sandi, tahun]	[beri, hargai, sdm, aman, siber, sandi, tahun]
5	[penerbitan, kppd, ta]	[terbit, kppd, ta]
6	[tanggapan, surat, penunjukan, pejabat, fungsional, perencana]	[tanggap, surat, tunjuk, jabat, fungsional, rencana]
7	[penawaran, lcd, videowall, led, videotron, interaktive, board]	[tawar, lcd, videowall, led, videotron, interaktive, board]
8	[rapat, pembahasan, ranperwal, pemenuhan, kebutuhan, sumber, daya, manusia, penggunaan, tenaga, alih, daya, dilingkungan, pemerintah, kota, medan]	[rapat, bahas, ranperwal, penuhi, butuh, sumber, daya, manusia, guna, tenaga, alih, daya, lingkungan, pemerintah, kota, medan]
9	[permohonan, perbaikan, handy, talky, ht]	[mohon, baik, handy, talky, ht]
10	[penyusunan, perjanjian, kinerja, perangkat, daerah, lingkungan, pemerintah, kota, medan]	[susun, janji, kerja, perangkat, daerah, lingkungan, pemerintah, kota, medan]
11	[penawaran, program, factory, visit]	[tawar, program, factory, visit]
12	[undangan, rapat, pembahasan, finalisasi, persiapan, pelaksanaan, penandatanganan, naskah, perjanjian, hibah, barang, milik, daerah, berita, acara, serah, terima, pemerintah, kota, medan, kantor, kementerian, agama, kota, medan]	[undang, rapat, bahas, final, siap, laksana, tanda, naskah, janji, hibah, barang, milik, daerah, berita, acara, serah, terima, pemerintah, kota, medan, kantor, menteri, agama, kota, medan]
13	[surat, panggilan, sidang]	[surat, panggil, sidang]
14	[mohon, bantuan, tenaga, pelaksanaan, pembongkaran, bangunan, lokasi, jalan, tamir, hamzah, kelurahan, helvetia, timur, kecamatan, medan, helvetia]	[mohon, bantu, tenaga, laksana, bongkar, bangun, lokasi, jalan, tamir, hamzah, lurah, helvetia, timur, camat, medan, helvetia]
15	[pemberian, penghargaan, sdm, keamanan, siber, sandi, tahun, disposisi, sekda, kota, medan]	[beri, hargai, sdm, aman, siber, sandi, tahun, disposisi, sekda, kota, medan]
16	[permohonan, bantuan, sterilisasi]	[mohon, bantu, steril]
17	[mohon, bantuan, tenaga, pelaksanaan, pembongkaran, bangunan, lokasi, jalan, bougenville, kelurahan, simpang, selayang, kecamatan, medan, tuntungan]	[mohon, bantu, tenaga, laksana, bongkar, bangun, lokasi, jalan, bougenville, lurah, simpang, selayang, camat, medan, tuntungan]
18	[pengajuan, kerja, sama, publikasi, pemberitaan, penayangan, iklan, pembangunan]	[aju, kerja, sama, publikasi, berita, tayang, iklan, bangun]
19	[mohon, bantuan, tenaga, pelaksanaan, pembongkaran, bangunan, lokasi, jalan, ngumban, surbakti, gang, bahagia, kelurahan, tanjung, sari, kecamatan, medan, selayang]	[mohon, bantu, tenaga, laksana, bongkar, bangun, lokasi, jalan, ngumban, surbakti, gang, bahagia, lurah, tanjung, sari, camat, medan, selayang]

JURNAL KOMPUTER TEKNOLOGI INFORMASI SISTEM KOMPUTER

p-ISSN: 2962-3022 | e-ISSN: 2963-7104
Volume 5 No 1 Juni 2026
Hal: 200-211
http://ejurnal.lkpkaryaprima.id/index.php/juktisi
DOI: https://doi.org/10.62712/juktisi.v5i1.935

No.	Tokenizing	Stemming
20	[mohon, bantuan, tenaga, pelaksanaan, pembongkaran, bangunan, lokasi, jalan, kayu, putih, sudut, gang, lingkungan, kelurahan, tanjung, mulia, hilir, kecamatan, medan, deli.	[mohon, bantu, tenaga, laksana, bongkar, bangun, lokasi, jalan, kayu, putih, sudut, gang, lingkungan, lurah, tanjung, mulia, hilir, camat, medan, deli]

2.4. Implementasi K-means clustering

Pada tahap implementasi K-Means, analisis dan perancangan sistem yang telah dilakukan sebelumnya direalisasikan melalui dua pendekatan, yaitu perhitungan secara manual dan implementasi menggunakan sistem berbasis Python. Pada perhitungan manual, data surat masuk yang telah melalui tahapan preprocessing dan pembobotan TF-IDF terlebih dahulu diklasifikasikan berdasarkan variabel atau term yang terbentuk. Data yang telah dikelompokkan tersebut kemudian direpresentasikan ke dalam sejumlah variabel penelitian yang menjadi dasar dalam proses perhitungan K-Means [11].

2.4.1. Data Surat Masuk Yang Sudah Dibobotkan Dengan TF-IDF

Data sample Hasil TF-IDF yang digunakan sebanyak 50 term data surat masuk untuk penerapan metode K-means clustering dan jumlah inisialisasi centroid K=3

2.4.2. Menentukan Titik Pusat Centroid

Pada tahap awal proses K-Means Clustering, penentuan pusat cluster (centroid) dilakukan secara acak (random). Dalam penelitian ini, centroid awal dipilih dari beberapa dokumen surat masuk yang dijadikan sebagai representasi awal masing-masing cluster. Misalnya, dokumen surat ke-3 dijadikan sebagai centroid awal untuk Cluster 1 (C1), dokumen surat ke-6 sebagai centroid awal untuk Cluster 2 (C2), dan dokumen surat ke-11 sebagai centroid awal untuk Cluster 3 (C3).
Cluster 1 (C1) : (0, 0, 0, 0, 0, 0, 0, 0, 2.435, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 2.099, 3.351, 3.351, 2.946, 3.351, 2.946, 2.946, 2.658, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0)
Cluster 2 (C2) : (0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 2.946, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 3.351, 2.946, 3.351, 3.351, 3.351, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0)
Cluster 3 (C3) : (0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 2.946, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 3.351, 3.351, 3.351, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0)

2.4.3. Perhitungan Jarak Pusat Centroid

Perhitungan jarak setiap data input terhadap masing- masing centroid digunakan rumus jarak Euclidean (Euclidean Distance) hingga ditemukan jarak yang paling dekat dari setiap data dengan centroid. Berikut merupakan contoh perhitungan Euclidean Distance dari semua data terhadap centroid pada cluster 1 [12]:

$$d1 = \sqrt{(D1_1 - C1_1) + (D1_2 - C1_2) + (D1_{...} - C1_{...}) + (D1_{50} - C1_{50})}$$
$$d1 = \sqrt{(2{,}658 - 0) + (2{,}946 - 0) + (\ldots - \cdots) + (0 - 0)}$$
$$= 12{,}193$$

$$d2 = \sqrt{(D2_1 - C2_1) + (D2_2 - C2_2) + (D2_{...} - C2_{...}) + (D2_{50} - C2_{50})}$$
$$d2 = \sqrt{(2{,}658 - 0) + (0 - 0) + (\ldots - \cdots) + (0 - 0)}$$
$$= 13{,}492$$

$$d3 = \sqrt{(D3_1 - C3_1) + (D3_2 - C3_2) + (D3_{...} - C3_{...}) + (D3_{50} - C3_{50})}$$
$$d3 = \sqrt{(0 - 0) + (0 - 0) + (\ldots - \cdots) + (0 - 0)}$$
$$= 0$$

$$d\ldots = \sqrt{(D\ldots_1 - C\ldots_1) + (D\ldots_2 - C\ldots_2) + (D\ldots_{...} - C\ldots_{...}) + (D\ldots_{50} - C\ldots_{50})}$$

$$d20 = \sqrt{(D20_1 - C20_1) + (D20_2 - C20_2) + (D20_{...} - C20_{...}) + (D20_{50} - C20_{50})}$$
$$d20 = \sqrt{(0 - 0) + (0 - 0) + (\ldots - \cdots) + (0 - 0)}$$
$$= 8{,}991$$

JURNAL KOMPUTER TEKNOLOGI INFORMASI SISTEM KOMPUTER
p-ISSN: 2962-3022 | e-ISSN: 2963-7104
Volume 5 No 1 Juni 2026
Hal: 200-211
http://ejurnal.lkpkaryaprima.id/index.php/juktisi
DOI: https://doi.org/10.62712/juktisi.v5i1.935

Hasil dari perhitungan Euclidean Distance cluster 1, cluster 2 dan cluster 3 pada iterasi 1 serta penentuan jarak terdekat dari masing-masing data ke centroid, yaitu jarak terdekat merupakan kelompok data dalam cluster tersebut dapat dilihat pada Tabel 3.3.3

Tabel 2. Hasil Iterasi 1

Dokumen	Jarak ke C1	Jarak ke C2	Jarak ke C3	Minimum	Anggota Cluster
D1	12,193	12,069	9,595	9,595	3
D2	13,492	12,930	10,657	10,657	3
D3	0	11,0479	9,2633	0	1
Dokumen	Jarak ke C1	Jarak ke C2	Jarak ke C3	Minimum	Anggota Cluster
D4	11,103	10,967	8,166	8,166	3
D5	10,527	9,797	6,509	6,509	3
D6	11,0479	0	8,4239	0	2
D7	12,377	11,762	8,209	8,209	3
D8	10,049	9,281	5,704	5,704	3
D9	8,782	7,892	2,946	2,946	3
D10	8,239	10,094	6,948	6,948	3
D11	9,2633	8,4239	0	0	3
D12	10,976	11,339	8,659	8,659	3
D13	9,263	7,322	4,166	4,166	3
D14	8,991	8,649	4,603	4,603	3
D15	11,462	11,331	8,648	8,648	3
D16	8,782	7,892	2,946	2,946	3
D17	9,461	9,136	5,465	5,465	3
D18	9,176	8,328	3,968	3,968	3
D19	8,991	8,649	4,603	4,603	3
D20	8,991	8,649	4,603	4,603	3

Pada iterasi pertama yang terdapat pada tabel diatas didapatkan kesimpulan bahwa anggota C1 adalah (D3), anggota C2 adalah (D6) sedangkan anggota C3 adalah (D1, D2, D4, D5, D7, D8, D9, D10, D11, D12, D13, D14, D15, D16, D17, D18, D19, D20). Pada iterasi kedua didapatkan kesimpulan bahwa anggota C1 adalah (D3), anggota C2 adalah (D6) sedangkan anggota C3 adalah (D1, D2, D4, D5, D7, D8, D9, D10, D11, D12, D13, D14, D15, D16, D17, D18, D19, D20). Hasil posisi cluster pada iterasi kedua sama dengan posisi iterasi pertama, maka proses dihentikan karena telah dinyatakan konvergen.

2.4.4. Analisis Karakteristik Hasil Clustering

Setelah diperoleh hasil akhir pengelompokan berdasarkan proses iterasi K-Means yang telah mencapai kondisi konvergen, dilakukan analisis karakteristik terhadap masing-masing cluster yang terbentuk [13]. Analisis ini dilakukan dengan memanfaatkan atribut pendukung, yaitu tanggal penerimaan surat, nomor dan tanggal surat, sifat surat, serta isi ringkas surat

1. Analisis cluster 1
   Cluster 1 terdiri dari 1 dokumen, yaitu D3. Berdasarkan isi ringkasnya, surat tersebut berkaitan dengan kegiatan orientasi penyusunan dokumen perencanaan Tahun 2025. Surat ini bersifat biasa dan diterima pada tanggal 08 Januari 2025. Secara tematik, cluster ini merepresentasikan surat yang berhubungan dengan kegiatan perencanaan dan penyusunan dokumen strategis. Karakteristik tersebut menunjukkan bahwa cluster ini memiliki fokus pada aktivitas administratif yang bersifat perencanaan awal tahun.
2. Analisis cluster 2
   Cluster 2 terdiri dari 1 dokumen, yaitu D6. Surat dalam cluster ini berisi tanggapan terkait penunjukan pejabat fungsional perencana. Surat tersebut bersifat biasa dan diterima pada tanggal 09 Januari 2025. Berdasarkan isi dan konteksnya, cluster ini dapat dikategorikan sebagai kelompok surat yang berkaitan dengan administrasi kepegawaian, khususnya mengenai penunjukan jabatan fungsional. Hal ini menunjukkan bahwa cluster 2 memiliki karakteristik tematik yang berbeda dari cluster lainnya.
3. Analisis cluster 3

JURNAL KOMPUTER TEKNOLOGI INFORMASI SISTEM KOMPUTER
p-ISSN: 2962-3022 | e-ISSN: 2963-7104
Volume 5 No 1 Juni 2026
Hal: 200-211
http://ejurnal.lkpkaryaprima.id/index.php/juktisi
DOI: https://doi.org/10.62712/juktisi.v5i1.935

Cluster 3 terdiri dari 18 dokumen dengan variasi isi yang lebih beragam dibandingkan cluster lainnya. Berdasarkan analisis isi ringkas surat, ditemukan pola dominan berupa surat yang berkaitan dengan kegiatan operasional instansi, seperti undangan rapat, pembahasan evaluasi kegiatan, permohonan bantuan teknis, pengadaan barang, serta perbaikan fasilitas. Kata-kata seperti "undangan", "rapat", "permohonan", "pengadaan", dan "perbaikan" muncul secara berulang pada beberapa dokumen dalam cluster ini. Kemunculan kata-kata tersebut menunjukkan adanya kemiripan tema antar surat sehingga sebagian besar dokumen tergabung dalam satu kelompok besar. Dengan demikian, Cluster 3 dapat dikategorikan sebagai kelompok surat operasional dan kegiatan rutin instansi.

2.5. Implementasi Berbasis Python

Penerapan metode K-Means Clustering untuk mengelompokkan surat masuk di Dinas Komunikasi dan Informatika diimplementasikan menggunakan bahasa pemrograman Python melalui platform Google Colaboratory (Google Colab). Platform ini digunakan karena berbasis cloud dan terintegrasi dengan Google Drive sehingga memudahkan proses penyimpanan dan pengolahan data [9][14][15].

Proses penerapan K-Means Clustering menggunakan Python melalui Google Colab dijelaskan sebagai berikut:

1. Dataset surat masuk yang awalnya tersimpan dalam format Microsoft Excel diekspor ke dalam format CSV agar dapat diproses dalam tahap clustering.
2. File dataset tersebut kemudian disimpan di Google Drive. Pada Google Colab dilakukan proses koneksi (mount) Google Drive agar sistem dapat mengakses file dataset secara langsung.
3. Setelah Google Drive terhubung, dataset dipanggil ke dalam lingkungan kerja Google Colab menggunakan path file yang sesuai untuk selanjutnya dilakukan proses pengolahan data.
4. Pada Google Colab dilakukan import library yang dibutuhkan, yaitu pandas untuk manipulasi dan pengelolaan data tabular, numpy untuk operasi perhitungan matematis, matplotlib untuk visualisasi data, scikit-learn untuk implementasi pembobotan TF-IDF dan algoritma K-Means Clustering, nltk untuk pemrosesan teks, Sastrawi untuk proses stemming bahasa Indonesia guna mereduksi kata ke bentuk dasar.
5. Dataset yang telah dipanggil dari Google Drive kemudian dapat dibaca dengan menampilkan isi data yang sudah di upload dengan nama file "data_surat.csv" dapat dilihat pada Gambar 3:

```
Data berhasil dibaca
  no tanggal_terima                  nomor_surat tanggal_surat    sifat  \
0  1     02-01-2025                   400.14.1.1    02-01-2025    Biasa
1  2     06-01-2025                400.14.6-049     02-01-2025  Penting
2  3     08-01-2025             000.7.2.4-017125    06-01-2025    Biasa
3  4     08-01-2025  137-BSSN-D1-SK.11.01-01-2025    06-01-2025   Segera
4  5     08-01-2025    900.1.3.6-56-BKAD-I-2-2025    06-01-2025    Biasa

                                    isi_ringkas
0  Undangan Acara Open House Camat Medan Tuntunga...
1  Undangan rapat koordinasi hasil rapat internal...
2  Pelaksanaan penyampaian orientasi penyusunan d...
3  Pemberian Penghargaan bagi SDM Keamanan Siber ...
4                          Penerbitan KPPD TA 2025
```

Gambar 4. Menampilkan Data CSV

6. Membuat Fungsi `preprocess_text` yang akan digunakan untuk membersihkan teks dengan tokenisasi, penghapusan stopwords bahasa Indonesia, dan proses stemming menggunakan sastrawi, menghasilkan teks yang telah diproses untuk analisis lebih lanjut, dapat dilihat pada Gambar 5:

JURNAL KOMPUTER TEKNOLOGI INFORMASI SISTEM KOMPUTER
p-ISSN: 2962-3022 | e-ISSN: 2963-7104
Volume 5 No 1 Juni 2026
Hal: 200-211
http://ejurnal.lkpkaryaprima.id/index.php/juktisi
DOI: https://doi.org/10.62712/juktisi.v5i1.935

```
[nltk_data] Downloading package punkt to /root/nltk_data...
[nltk_data]   Package punkt is already up-to-date!
[nltk_data] Downloading package stopwords to /root/nltk_data...
[nltk_data]   Package stopwords is already up-to-date!
                                          isi_ringkas  \
0  Undangan Acara Open House Camat Medan Tuntunga...
1  Undangan rapat koordinasi hasil rapat internal...
2  Pelaksanaan penyampaian orientasi penyusunan d...
3  Pemberian Penghargaan bagi SDM Keamanan Siber ...
4                          Penerbitan KPPD TA 2025

                                          clean_text
0  undang acara open house camat medan tuntung ra...
1  undang rapat koordinasi hasil rapat internal k...
2  laksana sampai orientasi susun dokumen rencana...
3                  beri harga sdm aman siber sandi
4                               terbit kppd ta
```

Gambar 5. Menampilkan Hasil Preprocessing TF-IDF

7. Langkah-langkah ini melakukan konversi teks ke dalam representasi *Term Frequency* (TF) menggunakan `*TfidfVectorizer*` dari scikit-learn dengan parameter `*use_idf=False*` dan `*norm=None*`, yang menghasilkan matriks TF. Selanjutnya, data TF tersebut diubah menjadi data frame menggunakan pandas dengan kolom-kolom yang merepresentasikan setiap term dan frekuensinya dalam dokumen, dan akhirnya ditampilkan dalam bentuk data frame dengan pesan "TF per Term" dapat dilihat pada Gambar 6:

```
===== MATRKS TF =====
   abadi  academy  acara  access  acquisition  action  ada  adam  adaptor  \
0      0        0      1       0            0       0    0     0        0
1      0        0      0       0            0       0    0     0        0
2      0        0      0       0            0       0    0     0        0
3      0        0      0       0            0       0    0     0        0
4      0        0      0       0            0       0    0     0        0

   adat  ...  yatim  yayasan  yemika  yose  zakat  zhafira  ziarah  zikir  \
0     0  ...      0        0       0     0      0        0       0      0
1     0  ...      0        0       0     0      0        0       0      0
2     0  ...      0        0       0     0      0        0       0      0
3     0  ...      0        0       0     0      0        0       0      0
4     0  ...      0        0       0     0      0        0       0      0

   zonasi  zoom
0       0     0
1       0     0
2       0     0
3       0     0
4       0     0

[5 rows x 1824 columns]
```

Gambar 6. Nilai *term frequency* Setiap Dokumen

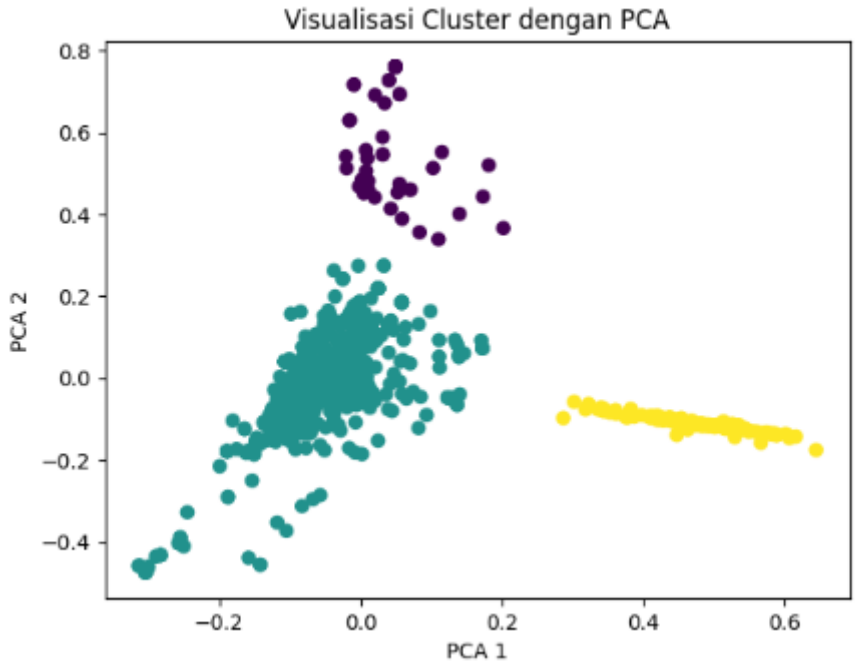
8. Menghitung matriks Term Frequency-Inverse Document Frequency (TF-IDF) dari teks yang telah dibersihkan, yang dapat digunakan untuk mewakili bobot relatif dari setiap kata dalam dokumen dapat dilihat pada Gambar 7:

JURNAL KOMPUTER TEKNOLOGI INFORMASI SISTEM KOMPUTER
p-ISSN: 2962-3022 | e-ISSN: 2963-7104
Volume 5 No 1 Juni 2026
Hal: 200-211
http://ejurnal.lkpkaryaprima.id/index.php/juktisi
DOI: https://doi.org/10.62712/juktisi.v5i1.935

```
===== MATRKS TF-IDF =====
   abadi  academy     acara  access  acquisition  action  ada  adam  adaptor  \
0    0.0      0.0  0.291807     0.0          0.0     0.0  0.0   0.0      0.0
1    0.0      0.0  0.000000     0.0          0.0     0.0  0.0   0.0      0.0
2    0.0      0.0  0.000000     0.0          0.0     0.0  0.0   0.0      0.0
3    0.0      0.0  0.000000     0.0          0.0     0.0  0.0   0.0      0.0
4    0.0      0.0  0.000000     0.0          0.0     0.0  0.0   0.0      0.0

   adat  ...  yatim  yayasan  yemika  yose  zakat  zhafira  ziarah  zikir  \
0   0.0  ...    0.0      0.0     0.0   0.0    0.0      0.0     0.0    0.0
1   0.0  ...    0.0      0.0     0.0   0.0    0.0      0.0     0.0    0.0
2   0.0  ...    0.0      0.0     0.0   0.0    0.0      0.0     0.0    0.0
3   0.0  ...    0.0      0.0     0.0   0.0    0.0      0.0     0.0    0.0
4   0.0  ...    0.0      0.0     0.0   0.0    0.0      0.0     0.0    0.0

   zonasi  zoom
0     0.0   0.0
1     0.0   0.0
2     0.0   0.0
3     0.0   0.0
4     0.0   0.0

[5 rows x 1824 columns]
```

Gambar 7. Nilai TF-IDF Setiap Dokumen

9. Algoritma K-Means dalam library scikit-learn digunakan untuk melakukan pengelompokan data menjadi tiga claster berdasarkan representasi vektor TF-IDF (`X`) yang dihasilkan sebelumnya, dengan `n_init` menentukan jumlah kali algoritma dijalankan dengan inisialisasi yang berbeda, dan hasilnya disimpan dalam objek `kmeans`.
10. Menggunakan Principal Component Analysis (PCA) dari scikit-learn untuk mereduksi dimensi data vektor TF-IDF (`X`) menjadi dua dimensi, kemudian hasil transformasi PCA disimpan dalam variabel `X_pca`. Ini bertujuan untuk memvisualisasikan hasil clustering dengan menggunakan scatter plot dalam dua dimensi.
11. Selanjutnya dihasilkan scatter plot menggunakan dua komponen utama hasil reduksi dimensi PCA, dengan warna titik ditentukan oleh hasil pengelompokan K-Means dapat dilihat pada Gambar 8:



Gambar 8. Visualisasi Hasil Clustering

12. Menggabungkan hasil clustering K-Means dengan isi ringkas surat dalam data frame `cluster_results`, kemudian tabel tersebut diurutkan berdasarkan klasternya dan ditampilkan sebagai hasil clustering K-Means dalam bentuk tabel dapat dilihat pada Gambar 9:

JURNAL KOMPUTER TEKNOLOGI INFORMASI SISTEM KOMPUTER
p-ISSN: 2962-3022 | e-ISSN: 2963-7104
Volume 5 No 1 Juni 2026
Hal: 200-211
http://ejurnal.lkpkaryaprima.id/index.php/juktisi
DOI: https://doi.org/10.62712/juktisi.v5i1.935

	clean_text	cluster
0	undang acara open house camat medan tuntung ra...	2
1	undang rapat koordinasi hasil rapat internal k...	2
2	laksana sampai orientasi susun dokumen rencana...	2
3	beri harga sdm aman siber sandi	2
4	terbit kppd ta	2

Gambar 9. Hasil Cluster

13. Program ini menggunakan pandas untuk menghitung jumlah anggota dalam setiap cluster dari data, menyusun informasinya dalam dataframe, mengurutkannya berdasarkan indeks cluster, dan menampilkan tabel hasilnya ke layar dengan format yang rapi menggunakan fungsi `tabulate` dapat dilihat pada Gambar 10:
14.

```
===== JUMLAH ANGGOTA CLUSTER =====
+----+-----------+----------+
|    |   Cluster |   Jumlah |
+====+===========+==========+
|  2 |         1 |       50 |
+----+-----------+----------+
|  0 |         2 |     1107 |
+----+-----------+----------+
|  1 |         3 |      132 |
+----+-----------+----------+
```

Gambar 10. Total Anggota Cluster

Pada nilai k = 3 di dapatkan hasil pada kelompok pertama terdapat 50 surat dan pada kelompok kedua terdapat 1107 surat sedangkan untuk kelompok ketiga terdapat 132 surat.

4. KESIMPULAN

Dari hasil penelitian yang telah dilakukan dengan metode K-Means Clustering untuk mengelompokkan surat masuk di Dinas Komunikasi dan Informatika, maka dapat disimpulkan bahwa : Surat masuk di Dinas Komunikasi dan Informatika Kota Medan lebih banyak kelompok surat yang berkaitan dengan administrasi kepegawaian, khususnya mengenai penunjukan jabatan fungsional dan hasil pengelompokan surat masuk dengan metode K-Means Clustering menunjukkan bahwa sekitar 85,9% kelompok surat administrasi kepegawaian khususnya mengenai penunjukan jabatan fungsional, sekitar 10,2% kelompok surat operasional dan kegiatan rutin instansi dan sekitar 3,9% surat kegiatan perencanaan dan penyusunan dokumen strategis.

REFERENCES

[1] J. Informasi, D. A. Fakhri, and S. Defit, “Optimalisasi Pelayanan Perpustakaan terhadap Minat Baca Menggunakan Metode K-Means Clustering,” vol. 3, 2021, doi: 10.37034/jidt.v3i3.137.
[2] W. Ananda *et al.*, “PENERAPAN ALGORITMA K-MEANS CLUSTERING DALAM,” vol. 6, no. 2, pp. 861–867, 2022.
[3] P. Algoritma and D. M. Dan, “Perbandingan algoritma dbscan-k means dan k means untuk pengelompokan madrasah aliyah provinsi jawa timur,” 2023.
[4] M. A. Nasution and M. Safii, “ALGORITMA K - MEANS DALAM PENGELOMPOKAN SURAT KELUAR DI KANTOR KEMENTERIAN,” vol. 4, pp. 61–71, 2024.
[5] U. M. Riau, . “3 No. 1,” no. 1, 2024.
[6] A. P. Ulasan, D. F. Rahayu, A. Manoar, H. Pardede, and S. Ramadani, “Jurnal Publikasi Ilmu Komputer dan Pengelompokan Data Warga dalam Pengurusan Surat Keterangan Berdasarkan Tujuan dengan Menggunakan Metode Clustering,” 2025.
[7] M. A. Khowarizmi, “Algoritma Mean Shift untuk Menentukan Segmentasi Pelanggan pada Penjualan Toko Online,” vol. 3, pp. 1–7, 2021.
[8] I. Rusydi and N. Hidayah, “APPLICATION OF DATA MINING IN GROCERY SALES USING THE FP-GROWTH ALGORITHM,” pp. 676–695.
[9] R. Rambu, S. Anawoli, A. A. Pekuwali, and P. A. R. L. Lede, “Development System in Letter Archiving Based on Object Oriented Programming Model System Development dalam Pengarsipan Surat Berbasis Model Object Oriented Programming,” vol. 4, no. April, pp. 463–471, 2024.
[10] N. Sari, H. H. Handayani, and A. M. Siregar, “Implementasi Clustering Data Kasus Covid 19 Di Indonesia Menggunakan Algoritma K-Means,” vol. 11, no. 1, pp. 7–12, 2023.

JURNAL KOMPUTER TEKNOLOGI INFORMASI SISTEM KOMPUTER
p-ISSN: 2962-3022 | e-ISSN: 2963-7104
Volume 5 No 1 Juni 2026
Hal: 200-211
http://ejurnal.lkpkaryaprima.id/index.php/juktisi
DOI: https://doi.org/10.62712/juktisi.v5i1.935

[11] J. V. Santoti, J. Jocelyn, and H. Irsyad, "Implementasi Term Frequency - Inverse Document Frequency dan Cosine Similarity untuk Analisis Kemiripan Deskripsi Produk Halal," vol. 03, no. 1, pp. 44–52, 2025.
[12] J. Nasional, I. Komputer, E. T. Naldy, F. Teknik, I. Komputer, and U. B. Darma, "Penerapan Data Mining Untuk Analisis Daftar Pembelian Konsumen Dengan Menggunakan Algoritma Apriori Pada Transaksi Penjualan Toko Bangunan MDN," vol. 2, no. 2, pp. 89–101, 2021.
[13] G. David and P. Maramis, "Arsip Surat Masuk Dan Keluar Pada Kejaksaan Tinggi Sulawesi Utara Dengan Algoritma K - Means Berbasis Web," 2025.
[14] P. G. Sindanglaut, "IMPLEMENTASI ALGORITMA K-MEANS DALAM OPTIMALISASI PENGELOMPOKAN SURAT MASUK DI," vol. 13, no. 1, 2025.
[15] L. Rusdiana and V. C. Hardita, "Algoritma K-Means dalam Pengelompokan Surat Keluar pada Program Studi Teknik Informatika STMIK Palangkaraya K-Means," vol. 9, 2023.