

Implementasi Algoritma K-Means dalam Menentukan Clustering pada Penilaian Kepuasan Pelanggan di Badan Pelatihan Kesehatan Pekanbaru

Aqshol Al Fahrozi¹, Fitri Insani^{2*}, Elvia Budianita³, Iis Afrianty⁴

¹Teknik Informatika, Fakultas Sains dan Teknologi, Universitas Islam Negeri Sultan Syarif Kasim Riau

¹Aqsholalfahrozi12@gmail.com, ²fitri.insani@uin-suska.ac.id,
³elvia.budianita@uinsuska.ac.id, ⁴iis.afrianty.@uin-suska.ac.id

Abstrak

Penelitian ini membahas implementasi algoritma K-Means dalam menentukan clustering pada penilaian kepuasan pelanggan di Badan Pelatihan Kesehatan Pekanbaru. Kepuasan pelanggan merupakan tingkat perasaan seseorang untuk mempresepsikan perbandingan antara kesannya konsumen terhadap tingkatan kinerja produk dan jasa dengan harapan pelanggan atau pembeli. Tujuan dari penelitian ini untuk melihat tingkat kepuasan pelanggan terhadap layanan Badan Pelatihan Kesehatan (Bapalkes) Pekanbaru menggunakan K-means clustering dan seberapa besar tingkat kepuasan pelanggan dengan metode K-means Clustering. Pada penelitian ini data yang digunakan adalah data pelanggan Balai Pelatihan Kesehatan dari tahun 2019 dan 2023. Data dikumpulkan melalui quisioner yang disebar melalui google form. Pembuatan model aturan data yang terkumpul menggunakan algoritma K-means dan perangkat lunak rapidminer. Dari hasil penelitian yang diperoleh menggunakan Algoritma K-Means dalam clustering data pelanggan dapat memberikan hasil segmentasi pelanggan yang sesuai dengan harapan, sehingga pihak Badan Pelatihan Kesehatan Pekanbaru dapat dengan mudah memahami karakteristik pelanggannya berdasarkan clusternya dan kepuasannya. Kemudian dengan metode elbow dan Davies Bouldin juga memberikan solusi pemilihan jumlah cluster yang tepat sehingga performanya lebih optimal dan menghasilkan hasil segmentasi pelanggan yang lebih akurat. Dari perhitungan algoritma K-means diperoleh nilai respon sangat dominan sebesar 259 yang menyatakan puas dan yang menyatakan tidak puas berjumlah 44 orang dari 303 pelanggan, sehingga algoritma K-means menggunakan uji sensitivitas dan spesifisitas sebanyak 86% menyatakan puas dan 14% menyatakan tidak puas terhadap pelayanan yang disediakan oleh Badan Pelatihan Kesehatan Pekanbaru.

Kata Kunci: K-Means, Clustering, Kepuasan Pelanggan, Data Mining

Abstract

This research discusses the implementation of the K-Means algorithm in determining clustering in customer satisfaction assessments at the Pekanbaru Health Training Agency. Customer satisfaction is the level of a person's feelings to perceive the comparison between the consumer's impression of the level of product and service performance and the customer's or buyer's expectations. The aim of this research is to see the level of customer satisfaction with the Pekanbaru Health Training Agency (Bapalkes) services using K-means clustering and how high the level of customer satisfaction is using the K-means Clustering method. In this research, the data used is Health Training Center customer data from 2019 and 2023. Data was collected through questionnaires distributed via Google form. Creating a rule model for the collected data using the k-means algorithm and rapidminer software. From the research results obtained using the K-Means algorithm in clustering customer data, it can provide customer segmentation results that are in line with expectations, so that the Pekanbaru Health Training Agency can easily understand the characteristics of its customers based on their clusters and their satisfaction. Then, using the elbow and Davies Bouldin methods, we also provide a solution for selecting the right number of clusters so that performance is more optimal and produces more accurate customer segmentation results. From the calculations of the k-means algorithm, it was obtained that the response value was very dominant at 259 who expressed satisfaction and 44 people who expressed dissatisfaction from 303 customers, so that the k-means algorithm used sensitivity and specificity tests, 86% expressed satisfaction and 14% expressed dissatisfaction with services provided by the Pekanbaru Health Training Agency.

Keywords : K-Means, Clustering, Customer Satisfaction, Data Mining

LATAR BELAKANG

Badan Pelatihan Kesehatan (Bapalkes) merupakan suatu satuan kerja yang bertanggungjawab kepada Dirjen Tenaga Kesehatan Kementerian Kesehatan RI yang memiliki wewenang dalam menjalankan penyelenggaraan pelatihan sumber daya manusia kesehatan. Dengan demikian Badan Pelatihan Kesehatan (Bapalkes) saling keterkaitan dengan pelanggan yang menggunakan layanan tersebut. Layanan kepuasan pelanggan merupakan prioritas yang harus dipenuhi. Kepuasan pelanggan memiliki arti perbandingan antara ekspektasi dan harapan pelanggan dengan apa yang dirasakan oleh pelanggan ketika memakai produk ataupun jasa tersebut. Apabila pelanggan merasakan bahwa performa produk sesuai ekspektasi atau melebihi ekspektasi, berarti pelanggan merasa puas. Jika kenyataan berbanding sebaliknya yaitu performa produk tidak sesuai ekspektasi pelanggan, berarti konsumen tidak merasa puas atas layanan yang diberikan (Gulto, 2020). Indrasari dalam Jurnal Ilmiah Bidang Sains dikatakan bahwasannya “kepuasan pelanggan adalah sifat emosional

seseorang atau tanggapan seseorang sesudah pembelian baik berupa kemarahan, kerunsingan, ketidakpuasan, netralitas, rasa senang atau gembira (Khairil, 2021)."

Kehadiran pelanggan menjadi sebagian faktor suatu perusahaan atau fasilitator layanan supaya tetap memiliki profit dan semakin berkembang. Dengan demikian agar tetap memberikan pelayanan yang baik diperlukan adanya evaluasi dan pengembangan setiap saat. Agar dapat melakukan hal tersebut pastinya dibutuhkan sebuah data, yakni dari data tersebut dapat dilihat dan dipotret gambaran di lapangan mengenai layanan yang diberikan (Radana, 2020)]. Sarana dan prasarana yang mendukung merupakan prioritas penting dalam meningkatkan kualitas layanan. Dimana perlunya tindakan yang sistematis dalam pelayanan, mulai dari proses input hingga tahap output. Agar dapat mengetahui tingkat kepuasan pelanggan dengan memakai kuisioner serta dapat mengetahui kualitas dan kesenjangan suatu layanan antara persepsi dengan harapan pelanggan diperlukan suatu metode. Peneliti melakukan clustering sebagai suatu metode analisa data yang dapat digunakan dalam memecahkan permasalahan yang berbentuk data. Pengolahan data pelanggan Badan Pelatihan Kesehatan (Bapalkes) perlu dilaksanakan agar dapat mengetahui informasi penting mengenai informasi baru terhadap penilaian ataupun pandangan pelanggan pada suatu instansi. Menurut Brown "kepuasan pelanggan merupakan suatu keadaan yang diperlukan, diinginkan dan diharapkan oleh konsumen terhadap sebuah jasa dan produk sepadan dengan penampilan dari jasa dan produk (Siregar, 2021).

Dalam proses pencarian pola diperlukan Data Mining yang merupakan proses mencari informasi penting berupa pengetahuan baru. Pada data mining teknik, metode dan algoritma sangat bervariasi, salah satu metode yang terdapat pada data mining di penelitian ini yaitu *clustering*. *Clustering* adalah himpunan objek data yang mempunyai kemiripan antara satu sama lain dalam kelompok lain. Clustering juga lebih populer dikenal dengan analisis *clustering* adalah metode pengelompokan satu set benda fisik maupun abstrak kedalam satu kelas objek yang sama (Novianto, 2019). Sebahagian metode tersebut terdapat dalam clustering yakni metode K-means. Metode *K-means* dapat digunakan sebagai pengelompokan dalam jumlah yang cukup besar dan relative cepat serta efisien dalam perhitungan (Aulia, 2021). Algoritma *K-Means* adalah Algoritma yang sederhana dalam penerapannya serta mempunyai kinerja relatif mudah beradaptasi, cepat dan umum digunakan (Sembiring, 2022).

Dalam penggunaannya metode k-means dapat menentukan jumlah cluster yang akan dibuat, dengan demikian metode k-means dapat bersifat fleksibel. Metode k-means juga terdapat algoritma *K-means* yakni algoritma non hirarki yang berfondasi dari metode data clustering. Algoritma *K-means* berangkat dari pembentukan partisi cluster diawal kemudian dengan cara interaktif partisi cluster dibenahi sehingga tidak terjadi perubahan yang signifikan pada partisipasi cluster (Sulistiyawati, 2021). Setelah melewati proses algoritma K-Means, hasil pengelompokan data akan divalidasi menggunakan metode *Elbow*. Putu berpendapat mengenai "metode *Elbow* adalah suatu metode yang digunakan untuk menghasilkan informasi dalam menentukan jumlah cluster terbaik dengan cara melihat persentase

hasil perbandingan antara jumlah cluster (k) yang akan membentuk siku pada suatu titik (Winarta, 2021). Selain metode tersebut penelitian ini juga memakai metode *Davies Bouldin Index*, Uji Sensitivitas dan uji Statifisitas. Beberapa penelitian sebelumnya telah meneliti mengenai kepuasan pelanggan. Penelitian sebelumnya membahas mengenai analisis dan mengelola data dengan metode K-means dengan teknik clustering untuk data mining menghasilkan informasi produk asuransi yang lebih diminati para calon nasabah (Hutagalung, 2021).

Saat ini, terdapat sarana prasarana fasilitas pelayanan kesehatan yang menyebar di seluruh titik daerah Kota Pekanbaru untuk melaksanakan pelayanan kesehatan terhadap masyarakat Kota Pekanbaru. Dengan keberadaan sarana prasarana kesehatan baik yang dimiliki oleh pemerintah maupun swasta, masyarakat diharapkan dapat menikmati pelayanan kesehatan secara optimal dan masyarakat bebas menentukan pilihannya dalam memilih fasilitas kesehatan. Fasilitas pelayanan kesehatan yang sudah ada seperti Puskesmas, Rumah Sakit & Klinik menjadi tujuan masyarakat ketika hendak berobat. UU Republik Indonesia No. 36 tahun 2009 pasal 1 menyatakan Fasilitas Pelayanan Kesehatan adalah suatu alat dan atau tempat yang digunakan untuk menyelenggarakan upaya pelayanan kesehatan, baik promotif, preventif, kuratif maupun rehabilitatif yang dilakukan oleh pemerintah pusat, pemerintah daerah, dan/ atau masyarakat.

Penelitian sebelumnya yang telah dilakukan oleh Rohmawati (2014) terkait dengan algoritma K-Means mengenai pengimplementasian Algoritma K-Means dalam pengklasteran mahasiswa pelamar beasiswa yang menghasilkan akurasi data keseluruhan nilai purity lebih dari 80%. Penelitian ini bertujuan untuk mengelompokkan data pelamar beasiswa Bantuan Belajar Mahasiswa (BBM) sehingga diperoleh tiga kelompok yaitu mahasiswa yang berhak menerima beasiswa, mahasiswa yang dipertimbangkan untuk menerima beasiswa dan mahasiswa yang tidak berhak menerima beasiswa. Penelitian dalam bidang Kesehatan pernah dilakukan oleh Meisida (2014) untuk mengklasifikasikan penyakit karies gigi dan menunjukkan bahwa algoritma K-Means memiliki nilai eror yang kecil dengan melakukan bineralisasi data terlebih dahulu. Selanjutnya pada penelitian mengenai pengimplementasian Algoritma K-Means untuk mengelompokkan penyakit pada pasien di Puskesmas Kajeen dapat disimpulkan bahwa metode k-means berhasil dan dapat diterapkan dengan hasil pengujian pada sistem pengelompokkan penyakit tersebut dan mempunyai akurasi 62% dengan menggunakan data sebanyak 20% dari data training. Penelitian lainnya yang pernah dilakukan oleh Shedthi S, dkk (2017) dengan membandingkan algoritma K-Means dan Fuzzy C Means untuk segmentasi penyakit bercak pada daun tanaman Arecanut dan Iris yang menghasilkan kesimpulan bahwa algoritma K-Means menghasilkan segmentasi Gambar yang lebih jelas dan baik serta memiliki waktu yang relatif lebih cepat daripada Fuzzy C Means.

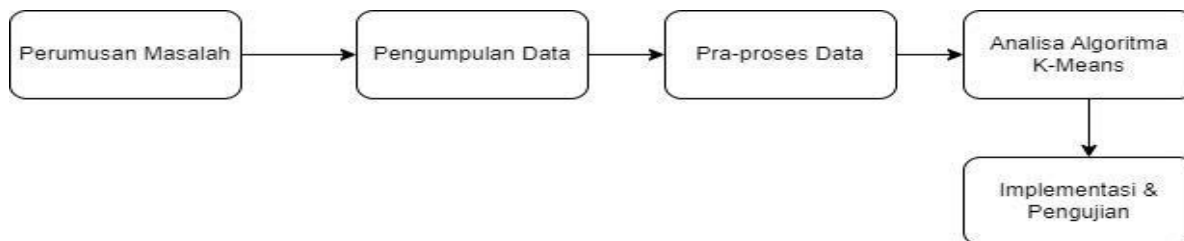
Maka pada penelitian ini dengan menerapkan algoritma K-Means akan mengelompokkan data responden terhadap kepuasan pelanggan di Badan Kesehatan Pekanbaru yang terdiri dari 1) keramahan pelayanan, 2) kredibilitas melayani masyarakat harus berpedoman pada prinsip/ tolok ukur, ketulusan, dan kejujuran dalam

menyajikan jasa pelayanan sesuai dengan kepentingan yang diharapkan, 3) dapat diakses dengan baik, dan dihubungi serta dinilai, 4) dukungan sarana prasarana yang memadai; dan 5) kemampuan menyajikan pelayanan sesuai dengan keinginan masyarakat (waktu, biaya, kualitas, kuantitas, moral, terampil, daya tanggap, keamanan, dan komunikasi yang baik (Along, 2020). Penelitian ini diharapkan dapat membantu pemerintah dan instansi terkait dalam mengetahui pola alternatif pengobatan masyarakat Pekanbaru berdasarkan cluster terbaik dengan menggunakan algoritma K-Means. Juga dapat dijadikan sebagai pertimbangan untuk Pemerintah dalam membuat kebijakan nasional terkait fasilitas pelayanan kesehatan khususnya di Kota Pekanbaru.

Berlatarbelakang teori diatas maka peneliti akan membahas tingkat kepuasan pelanggan atas pelayanan Badan Pelatihan Kesehatan (Bapelkes) Pekanbaru menggunakan *K-means Clustering* dan seberapa besar tingkat kepuasan pelanggan menggunakan metode *K-means Clustering*.

METODE

Metode penelitian yang digunakan oleh penulis dalam menerapkan algoritma k-means untuk menentukan kepuasan pelanggan terhadap layanan yang diberikan pihak Badan Pelatihan Kesehatan Pekanbaru, menggunakan metode penelitian yang sesuai pada Gambar 1.



Gambar 1. Metode penelitian

Perumusan Masalah

Langkah awal dalam penelitian ini adalah mencari perumusan masalah, diman peneliti mencari masalah dari berbagai sumber dari penitian sebelumnya, setelah masalahnya teridentifikasi, peneliti ini akan mempelajari masalah dari penelitian sebelumnya untuk mencari solusinya, penelitian digunakan untuk mencari kepuasan pelanggan menggunakan algoritma *K-Means Clustering*.

Pengumpulan Data

Pengumpulan data berfungsi sebagai informasi untuk penelitian ini. Data tersebut secara khusus meliputi atribut sebagai berikut : nama orang, pertanyaan 1-5. Jumlah data yang diambil sebanyak sebanyak 303 data pelanggan. Peneliti menggunakan data yang sudah terkumpul dari kuesioner yang dibagikan kepada pelanggan yang

sudah menggunakan layanan Balai Pelatihan Kesehatan Pekanbaru dari tahun 2019-2020 dan 2023.

Keterangan pertanyaan pada table 1 dibawah ini:

- 1) Saya tidak menunggu lama untuk mendapatkan pelayanan yang saya butuhkan
- 2) Saya paham penjelasan yang diberikan pegawai terkait apapun yang saya tanyakan
- 3) Saya merasa aman saat menggunakan fasilitas yang tersedia di Bapelkes
- 4) Pegawai Bapelkes bersikap ramah dan sopan dalam memeberikan pelayanan
- 5) Saya merasa puas terhadap fasilitas yang disediakan

Tabel 1 Data Hasil Kuesioner Pelanggan

No	Responden	Pertanyaan 1	Pertanyaan 2	Pertanyaan 3	Pertanyaan 4	Pertanyaan 5
1	R1	Setuju	Setuju	Setuju	Cukup Setuju	Kurang Setuju
2	R2	Setuju	Setuju	Setuju	Setuju	Setuju
3	R3	Cukup Setuju	Cukup Setuju	Cukup Setuju	Cukup Setuju	Cukup Setuju
.....						
.....						
301	R301	Setuju	Sangat Setuju	Sangat Setuju	Sangat Setuju	Sangat Setuju
302	R302	Setuju	Setuju	Setuju	Setuju	Setuju
303	R303	Setuju	Setuju	Setuju	Setuju	Setuju

Kualitas pelayanan dapat dinilai berdasarkan lima dimensi yang mencakup bukti fisik (*tangibles*), reliabilitas (*reliability*), daya tanggap (*responsiveness*), jaminan (*assurance*), dan empati (*emphaty*). Oleh sebab itu pendekatan sikap yang berkaitan dengan kepedulian; melayani dengan tindakan yang baik; dan adanya tujuan untuk memuaskan pelanggan dengan berorientasi pada standar layanan. Senada dengan hal maka Ibrahim juga berpendapat mengenai indikator layanan, diantaranya sebagai berikut:

- 1) Keramahan pelayanan
- 2) Kreadibilitas melayani masyarakat harus berpedoman pada prinsip/ tolok ukur, ketulusan, dan kejujuran dalam menyajikan jasa pelayanan sesuai dengan kepentingan yang harapan;
- 3) Dapat diakses dengan baik, dan dihubungi serta dinilai;
- 4) Dukungan sarana prasarana yang memandai; dan

- 5) Kemampuan menyajikan pelayanan sesuai dengan keinginan masyarakat (waktu, biaya, kualitas, kuantitas, moral, terampil, daya tanggap, keamanan, dan komunikasi yang baik (Along, 2020).

Selanjutnya Indikator kepuasan pelanggan menurut Kotler dan Keller menyatakan kunci untuk mempertahankan pelanggan adalah kepuasan konsumen. Indikator Kepuasan pelanggan dapat dilihat dari:

- 1) Expectations : Mengukur kualitas barang dan jasa yang diberikan sesuai dengan apa yang dijanjikan dan diharapkan.
- 2) Performance : Pelayanan atau kinerja yang baik dapat memberikan kepuasan bagi pelanggan (Murtiningsih, 2020).

Pra-proses data

Pra-proses data merupakan salah satu langkah paling penting dalam proses data mining yang berkaitan dengan transformasi data. Data *Preprocessing* merupakan salah satu tahapan dalam melakukan mining data. Sebelum menuju ke tahap pemrosesan. Data mentah akan diolah terlebih dahulu. Data *Preprocessing* atau pra proses data biasanya dilakukan melalui cara eliminasi data yang tidak sesuai. Selain itu dalam proses ini data akan diubah dalam bentuk yang akan lebih dipahami oleh sistem. Pengertian lain menyebutkan bahwa data *preprocessing* adalah tahapan untuk menghilangkan beberapa permasalahan yang bisa mengganggu saat pemrosesan data. Hal tersebut karena banyak data yang formatnya tidak konsisten. Data *preprocessing* merupakan teknik paling awal sebelum melakukan data mining.

Tranformasi Data

Transformasi Data adalah proses perubahan data, menjadi data yang dapat diolah menggunakan algoritma yang akan dipakai apakah dalam bentuk numerik, klasifikasi dan lain- lain. Agar data dapat diolah dengan menggunakan algoritma K-Means, maka data yang berjenis data nominal seperti jawaban Kuesioner diinisialisasikan terlebih dahulu dalam bentuk angka.

Table 2. Pra-Proses data Jawaban

Jawaban	Transformasi
Sangat Setuju	5
Setuju	4
Cukup Setuju	3
Kurang Setuju	2
Tidak Setuju	1

Tabel 2 menjelaskan bahwa dalam melengkapi kuesioner yang mengharuskan mereka untuk menunjukkan tingkat persetujuannya terhadap serangkaian pertanyaan yang diberikan. Transformasi data adalah merubah skala data ke dalam bentuk lain

sehingga data memiliki distribusi yang diharapkan. Setiap data dilakukan operasi matematika yang sama pada data aslinya. Berarti kita merubah semua data untuk menjaga perbedaan perbedaan antar data relatif tetap. Jika data kita memiliki lebih dari satu variabel, maka kita memiliki lebih dari satu variabel, maka kita mentransformasikan semua variabel agar hubungan antar data tidak berubah.

Tabel 3. Hasil Pra-proses data kepuasan pelanggan Badan Pelatihan Kesehatan Pekanbaru

No	Responden	Pertanya an 1	Pertany an 2	Pertanya an 3	Pertanyaa n 4	Pertanyaan 5
1	R1	4	4	4	3	2
2	R2	4	4	4	4	4
3	R3	3	3	3	3	3
....						
..						
..						
....						
..						
.						
301	R301	4	5	5	5	5
302	R302	4	4	4	4	4
303	R303	4	4	4	4	4

Tabel 3 menjelaskan bahwa terdapat 303 responden yang memberikan jawaban terhadap jawaban kuesioner. Kuesioner yang disebarakan terdiri dari 5 pertanyaan mengenai kepuasan pelanggan Badan Pelatihan Kesehatan Pekanbaru. Secara keseluruhan jawaban responden berada pada kategori sangat setuju dan setuju. Hal ini berarti dari 5 pertanyaan kuesioner yang dibagikan kepada responden rata rata responden menyatakan sangat setuju dan setuju terhadap pertanyaan kepuasan pelanggan Badan Pelatihan Kesehatan Pekanbaru.

Analisa Algoritma K-Mean.

Penelitian ini menggunakan algoritma k-means untuk mengelompokkan data layanan berdasarkan berdasarkan kesamaan rating antar masing-masing layanan. Berikut tahapan dalam perhitungan algoritma k-means:

- 1) Menentukan nilai klaster pada data yang akan diproses
- 2) Memilih nilai sebagai centroid cluster awal secara acak
- 3) Menghitung jarak pusat klaster ke masing-masing data dengan rumus (Sinaga:2022):

$$d(x, y) = \sqrt{\sum^n 1(x_i - y_1)^2} \quad (1)$$

Dimana $d(x,y)$: Jarak objek x ke pusat cluster y

n : Jumlah Atribut

x_i : Objek Data

y_i : Data Cluster

4) Mengelompokkan data sesuai jarak terdekat antar objek ke pusat cluster, kemudian memperbarui titik centroid cluster pada iterasi selanjutnya (Rofiqo, 2018)].

5) Mengulangi langkah pada tahap ke3 sampai kluster tidak berubah.

K-Means Clustering Algoritma juga merupakan metode non-hierarchy. Metode Clustering Algoritma adalah mengelompokkan beberapa data ke dalam kelompok yang menjelaskan data dalam satu kelompok memiliki karakteristik yang sama dan memiliki karakteristik yang berbeda dengan data yang ada di kelompok lain. K-means merupakan salah satu algoritma yang bersifat *unsupervised learning*. K-Means memiliki fungsi untuk mengelompokkan data ke dalam data cluster.

Implementasi

Implementasi pada penelitian ini menggunakan aplikasi rapidminer. Aplikasi ini berguna untuk menghasilkan kluster terbaik pada setiap metode yang digunakan yakni pada metode metode *elbow*, metode DBI (*Davies Bouldin Index*) dan metode uji sensitivitas dan spesifisitas. Model Clustering merupakan salah satu model dalam data mining yang digunakan untuk melakukan proses pengelompokan dari kumpulan data ke dalam bentuk pengelompokan yang masing-masing hasil pengelompokan mempunyai kemiripan. Salah satu algoritma yang dapat digunakan untuk menemukan pengelompokan dalam data adalah dengan menggunakan algoritma K-Means. Clustering Algoritma (K-Means) memiliki tujuan untuk meminimalisasikan fungsi objective yang telah di set dalam proses clustering.

a) Metode Elbow

Metode elbow merupakan suatu metode penentuan jumlah kluster yang ideal dengan upaya menyesuaikan tingkatan kluster sehingga terbentuk siku pada titik tertentu. Maka dilakukanlah pengujian dengan menginputkan jarak rata-rata (Avg Distance). Avg Distance merupakan jarak rata-rata antara setiap titik K. Kemudian terbentuklah grafik Avg Distance yang memperlihatkan titik/garis yang membentuk siku. Maka dari itu dapat dilihat semakin besar titik sudut maka semakin tinggi pula akurasi dalam menentukan jumlah K pada K-Means Clustering (Humaira, 2023).

Menurut Madhulatha dalam Merliana, Ernawati & Santoso disebutkan bahwa metode Elbow merupakan suatu metode yang digunakan untuk menghasilkan informasi dalam menentukan jumlah cluster terbaik dengan cara melihat persentase hasil perbandingan antara jumlah cluster yang akan membentuk siku pada suatu titik. Metode Elbow ini memberikan ide atau gagasan dengan cara memilih nilai cluster dan kemudian menambah nilai cluster tersebut untuk dijadikan model data dalam penentuan

cluster terbaik. Dan selain itu persentase perhitungan yang dihasilkan menjadi perbandingan antara jumlah cluster yang ditambah. Hasil persentase yang berbeda dari setiap nilai cluster dapat ditunjukkan dengan menggunakan grafik sebagai sumber informasinya. Jika nilai cluster pertama dengan nilai cluster kedua memberikan sudut dalam grafik atau nilainya mengalami penurunan paling besar maka nilai cluster tersebut yang terbaik (Muningsih, 2018).

b) Davies Bouldin Index

Davies Bouldin Index (DBI) adalah parameter yang dijumpai pada tahun 1979 oleh Donal W. Bouldin dan David L. Davies. cara mencari DBI adalah dengan menggunakan rapidminer untuk mendapatkan hasil cluster distance performance. DBI digunakan untuk menggambarkan algoritma K-Means. Semakin mendekati nilai 0 maka skema clustering tersebut akan semakin ideal (Muningsih, 2021).

Menurut Permatadevi, jika proses pengklasteran untuk masing-masing k selesai, maka untuk menentukan jumlah klaster yang terbaik dapat dilakukan penilaian menggunakan *Davies-Bouldin Index* (DBI). Pendekatan pengukuran ini bertujuan untuk memaksimalkan jarak antara klaster yang satu dengan yang lainnya dan pada waktu yang sama mencoba untuk meminimalkan jarak antara objek dalam sebuah klaster (Badruttamam:2020). *Davies Bouldin Index* (DBI) merupakan cara validasi cluster yang ditemukan oleh D. L. Davies. DBI adalah fungsi rasio dari jumlah distribusi didalam cluster untuk pemisahan antar cluster. Dalam penelitian ini, DBI digunakan untuk melakukan validasi data pada setiap cluster Hadiyattullah, 2020).

c) Uji Sensitivitas dan Spesifisitas

Sensitivitas dan spesifisitas adalah dua ukuran statistik yang paling umum digunakan untuk menilai kinerja tes alternatif terhadap standar emas.

$$Uji\ Sensitivitas = \frac{positif\ benar}{dxpositif\ benar + negatif\ palsu} \times 100\% \quad (4)$$

$$Uji\ Spesifitas = \frac{positif\ benar}{dxpositif\ benar + negatif\ palsu} \times 100\% \quad (5)$$

Sensitivitas, spesifisitas, dan nilai prediktif dapat digunakan untuk mengukur kinerja definisi kasus atau hasil tes diagnostik atau algoritma. Berbeda dengan sensitivitas dan spesifisitas, nilai prediksi bervariasi sesuai dengan prevalensi suatu kondisi dalam suatu populasi. Bahkan dengan tes diagnostik yang sangat spesifik, jika suatu penyakit jarang terjadi pada orang-orang yang dites, sebagian besar hasil tes positif akan menjadi positif palsu, dan nilai prediksi positifnya akan rendah (Nanda, 2020).

HASIL DAN PEMBAHASAN

Penerapan Algoritma K-Means

K-means merupakan salah satu algoritma yang bersifat unsupervised learning. K-Means memiliki fungsi untuk mengelompokkan data kedalam data cluster. Algoritma ini dapat menerima data tanpa ada label kategori. K-Means Clustering Algoritma juga merupakan metode non-hierarchy. Kelebihan dari metode k-means adalah mampu mengelompokkan data besar dengan sangat cepat, sedangkan kekurangan dari metode k-means adalah banyaknya gerombol harus ditentukan sebelumnya. Beberapa hal yang perlu diperhatikan dalam pengembangan metode K-Means dengan kernel trik ini adalah bahwa data pada feature space tidak lagi dapat didefinisikan secara eksplisit, sehingga penghitungan nilai membership function dan centroid tidak dapat dilakukan secara langsung. Pada penerapan algoritma k-means menggunakan tahapan:

- 1) Menentukan nilai kluster pada data yang akan diproses
- 2) Memilih nilai sebagai centroid cluster awal secara acak

Tabel 4. Cluster Awal

Data 9	Kluster 1	3	2	3	3	3
Data 14	Kluster 2	4	5	5	5	5

- 3) Menghitung jarak pusat kluster ke masing-masing data dengan menggunakan rumus 1 dan mendapatkan hasil:

$$DA_{1c1} = 2,8284 \quad DA_{2c1} = 2,8284$$

$$DA_{3c1} = 1$$

Dan selanjutnya sampai dengan DA_{303c1} . Kemudian untuk perhitungan jarak dari

Centroid ke-2 adalah sebagai berikut :

$$DA_{1c2} = 3,8729$$

$$DA_{2c2} = 2$$

$$DA_{3c2} = 4,1231$$

- 4) Mengelompokkan data sesuai jarak terdekat antar objek ke pusat kluster, kemudian memperbarui titik centroid kluster pada iterasi selanjutnya menggunakan rumus 2.

Pengelompokan dokumen berdasarkan karakteristik data yang ada pada dokumen dilakukan untuk memudahkan mengorganisir dokumen pada kebutuhan lebih lanjut. Algoritma K-Means clustering merupakan salah satu algoritma pengelompokan yang melakukan pengelompokan secara iteratif dengan melakukan partisi set data ke dalam sejumlah K cluster. Namun hasil pada K-Means clustering tersebut sangat ditentukan pada penentuan pusat cluster awal yang pada umumnya ditentukan secara random. Pada penelitian ini dilakukan penentuan pusat cluster awal K-Means untuk masalah pengelompokan dokumen dengan memanfaatkan algoritma genetika sebagai algoritma yang mengoptimasi pusat cluster awal K-Means tersebut.

Hasil perhitungan jarak/selisih antara setiap data dengan titik pusat Claster dapat dilihat pada Tabel 5 berikut ini;

Tabel 5. Perhitungan jarak/selisih

No	Responden	C1	C2	Jarak Terpendek	Cluster
1	R1	2,645751311	3,872983 35	2,64575131 1	CLUSTE R1
2	R2	2,828427125	2	2	CLUSTE R2
3	R3	1	4,123105 63	1	CLUSTE R1
.....					
.....					
..					
301	R301	4,69041576	0	0	CLUSTE R2
302	R302	2,828427125	2	2	CLUSTE R2
303	R303	2,828427125	2	2	CLUSTE R2

Beberapa hal yang perlu diperhatikan dalam pengembangan metode K-Means dengan kernel trik ini adalah bahwa data pada feature space tidak lagi dapat didefinisikan secara eksplisit, sehingga penghitungan nilai membership function dan centroid tidak dapat dilakukan secara langsung. Algoritma k-means dipilih karena metode klasterisasi dipakai membagi data menjadi beberapa klaster, yang mana data dengan kemiripan tinggi dikelompokkan menjadi satu cluster dan data dengan karakteristik beda dikelompokkan ke cluster berbeda. Keuntungan algoritma ini adalah mudah dipelajari dan sederhana karena pada dasarnya algoritma ini hanya menghitung jarak tiap data ke titik pusat cluster. Namun demikian, algoritma K-Means clustering tetap memiliki kelemahan, yaitu kita perlu menentukan jumlah clusternya terlebih dahulu. Maka ditemukanlah hasil cluster pada setiap data yang diproses.

Tabel 6. Hasil Cluster

Cluster	Hasil
Cluster1	44
Cluster2	259

Tabel 6 menjelaskan hasil pada penilaia kepuasan pelanggan Badan Pelatihan Kesehatan Pekanbaru. Hasil akhir ini didapat melalui beberapa tahapan. Proses

klasterisasi dimulai dengan menginisialisasi jumlah kluster yang diinginkan. Kemudian, data yang akan diklasterisasi dimasukkan ke dalam model K-Means dan dilakukan iterasi untuk memperbarui posisi centroid dan label kluster hingga mencapai konvergensi. Setelah pelatihan selesai, model K-Means dapat digunakan untuk memprediksi label kluster dari data baru.

Hasil Pengujian

4.1.1. Hasil Pengujian

Menyesuaikan hasil perhitungan manual dengan Rapidminer 10.0 merupakan langkah selanjutnya setelah mendapatkan hasil akhir. RapidMiner merupakan software yang berdiri sendiri untuk analisis data dan sebagai mesin data mining yang dapat diintegrasikan pada produknya sendiri. RapidMiner ditulis dengan menggunakan bahasa java sehingga dapat bekerja di semua sistem operasi. Bahwa disimpulkan data yang digunakan adalah benar. Hal ini ditunjukkan dengan evaluasi perhitungan manual dan menggunakan Rapidminer 10.0, keduanya menghasilkan hasil yang sama.

Cluster Model

```
Cluster 0: 44 items
Cluster 1: 259 items
Total number of items: 303
```

Gambar 2. Menunjukkan tampilan cluster model yang ada di Rapidminer:

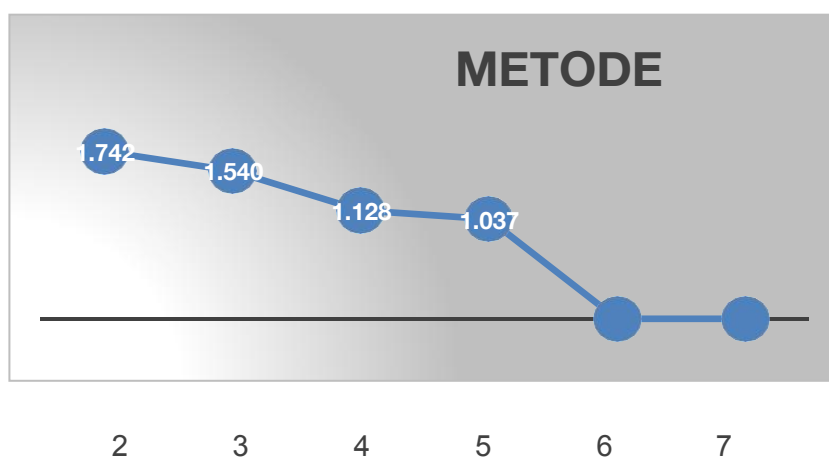
Gambar 2 menjelaskan bahwa dari data pelanggan yang memiliki potensi penilaian baik/setuju yakni sekitar 259 orang dan 44 orang dengan potensi penilaian kurang baik/kurang setuju. Peneliti dapat menginput data yang valid dengan menggunakan RapidMiner dari K hingga n dalam pengujian, untuk menyederhanakan dan mempercepat proses clustering. Peneliti melakukan uji coba sampai K7 dengan menggunakan data yang sama.

Tabel 7. Menentukan banyaknya K

Klaster	Avg Distance
2	1.742
3	1.540
4	1.128
5	1.059
6	0.956

7	0.918
---	-------

Clustering merupakan metode segmentasi data yang sangat berguna dalam prediksi dan analisa masalah bisnis tertentu. K Means clustering berguna untuk mengelompokkan data yang besar. Pengelompokan kumpulan data besar secara manual bisa memakan waktu dan rentan terhadap kesalahan. Dengan K Means clustering, proses grouping data dapat berjalan secara otomatis sehingga perusahaan bisa fokus pada analisis dan interpretasi hasil saja. Pada Tabel 7 didapatkan K percobaan sebanyak K7 dengan masing-masing Avg Distancenya, untuk dapat melihat posisi titik dari setiap nilai *Avg Distance* dibuatkan sebuah diagram linier yang dapat dilihat pada Gambar 3 berikut;



Gambar 3. Grafik Elbow Methode

Dapat dilihat pada Gambar 3 . titik x (6) dengan titik y bernilai 0,956 memiliki titik yang paling mendekati siku-siku sedangkan pada titik x (2) sampai x (4) memiliki titik yang tumpu. Sesuai dengan metode *elbow* dimana K terbaik ketika memiliki bentuk paling siku. Maka didapatkan titik yang paling siku ketika titik berada pada y (*Avg Distance*) dengan nilai y berjumlah 0,956 dengan nilai x (K) bernilai 6. Setelah mendapatkan banyaknya K terbaik dengan metode *elbow*, kemudian peneliti mengeksekusi data yang sama dengan melakukan clustering dengan K sebanyak n pada RapidMiner sehingga membuat kelompok *CLUSTER* 2, 3,4,5,6 dan 7.

Maka didapatkan hasil pada cluster 1 memiliki tingkat sangat puas dan puas, cluster 2 memiliki tingkat puas, cluster 3 memiliki tingkat puas dan cukup puas, cluster 4 memiliki tingkat cukup puas dan kurang puas, cluster 5 sangat puas dan puas, cluster 6 memiliki tingkat sangat puas dan puas. Pada percobaan kedua ini peneliti telah membagi data kedalam 7 kelompok sesuai dengan jarak dengan titik senternya. Namun, hasil *Davies-Bouldin Index* (DBI) masih bernilai negatif seperti Tabel 8 berikut.

Tabel 8. Hasil DBI Tiap Kelompok

No	Clust er	DBI
1	k2	-1.041
2	k3	-1.417
3	k4	-1.201
4	k5	-1.163
5	k6	-1.308
6	K7	-1.187

Dilihat dari Tabel 8 bahwa DBI dengan nilai terbaik yaitu nilai yang paling mendekati nol, maka didapatkan DBI dengan nilai -1.037 namun nilai tersebut masih negatif. Untuk menghasilkan nilai yang positif pada pemodelan menggunakan parameter *maximize* sehingga mengubah nilai negatif yang dihasilkan dari *Davies Bouldin Index* (DBI). Parameter ini menentukan apakah kinerja harus dimaksimalkan. Jika benar, hasilnya tidak dikalikan dengan minus satu. Pada kualitas parameter dimaksimalkan untuk setiap kualitas dari masing-masing kelompok, maka parameter *maximize* perlu disertakan. Dengan ini *maximize* dilakukan untuk mengoptimalkan hasil dari DBI, sehingga hasil yang didapatkan menjadi seperti pada Tabel 9 berikut.

Tabel 9. Hasil DBI Setelah Maksimal

No	Cluster	DBI
1	k2	1.041
2	k3	1.417
3	k4	1.201
4	k5	1.163
5	k6	1.308
6	K7	1.187

Setelah menggunakan parameter *maximize* dapat dilihat nilai negatif kini berubah menjadi nilai positif. Untuk lebih mengoptimalkan nilai DBI perlu dilakukan *normalize* karena parameter *normalize* digunakan untuk membuat skala agar sesuai dengan rentang tertentu. Hal tersebut penting ketika terdapat data dengan unit dan skala yang berbeda. Misalnya, menggunakan jarak *euclidean*, semua atribut harus berada pada skala yang sama untuk perbandingan yang adil. Normalisasi berguna untuk membandingkan atribut dengan ukuran berbeda. Operator ini menormalkan atribut yang dipilih. Tersedia empat metode pemulihan. Parameter ini menentukan apakah hasil harus dinormalisasi. Jika benar, kriteria akan dibagi dengan jumlah fitur. Maka hasil yang didapatkan ketika menggunakan parameter *maximize* dan *normalize* akan tampak pada Tabel 10 berikut.

Tabel 10. Hasil DBI Setelah Maksimal dan Normalisasi

No	Cluster	DBI
1	k2	0.208
2	k3	0.283
3	k4	0.229
4	k5	0.233
5	k6	0.262
6	K7	0.237

Dapat dilihat pada Tabel 10 bahwa setelah menggunakan parameter *maximize* dan *normalize* hasil nilai DBI semakin optimal. Setelah melakukan tahap penentuan banyaknya cluster dan mengelola hasil DBI agar semakin optimal, didapatkan jumlah items dari masing-masing kelompok dengan nilai sesuai dengan cluster model pada Gambar 4 berikut:

Cluster Model

```
Cluster 0: 180 items
Cluster 1: 9 items
Cluster 2: 21 items
Cluster 3: 3 items
Cluster 4: 57 items
Cluster 5: 14 items
Cluster 6: 19 items
Total number of items: 303
```

Gambar 4. Hasil Clustering pada Aplikasi Rapidminer

Gambar 4 menunjukkan Cluster1 mempunyai 10 items, Cluster2 mempunyai 39 items, Cluster3 mempunyai 15 items, Cluster4 mempunyai 179 items, Cluster5 mempunyai 33 item, dan Cluster 6 mempunyai 27 items. Clustering adalah proses partisi satu set objek data ke dalam himpunan bagian yang disebut dengan cluster. Maka itu, metode clustering ini berguna untuk menemukan kelompok yang tidak dikenal dalam data. Pengklasteran objek data merupakan salah satu cara untuk mempermudah dalam membaca data, terutama data berdimensi tinggi. Obyek-obyek data berada dalam satu klaster jika mempunyai kesamaan yang tinggi, dan sebaliknya, berada pada klaster berbeda jika menunjukkan ketidaksamaan.

Hasil Metode Algoritma K-Means Menggunakan Sensitivitas dan Spesifitas

Menentukan nilai positif benar, dan negatif palsu dimana positif benar adalah nilai yang tinggi dan menyatakan puas pada variabelnya dan negatif palsu adalah nilai yang rendah dan menyatakan tidak puas pada variable. Menentukan nilai negatif benar, dan positif palsu dimana negatif benar adalah nilai yang rendah dan menyatakan tidak puas pada variable lain dan negatif benar adalah nilai yang tinggi dan menyatakan puas pada variable lain. Pengujian sensitivitas metode algoritma K-Means bisa mengklarifikasikan tingkat kepuasan pelanggan dengan variable 86%. Sedangkan, uji

spesifisitas metode algoritma K- Means dapat mengklarifikasikan tidak puasnya pelanggan dengan variable 16% dengan menggunakan rumus 4 dan 5.

SIMPULAN

Dari hasil penelitian yang diperoleh menggunakan Algoritma K-Means dalam clustering data pelanggan dapat memberikan hasil segmentasi pelanggan yang sesuai dengan harapan, sehingga pihak Badan Pelatihan Kesehatan dapat dengan mudah memahami karakteristik pelanggannya berdasarkan clusternya dan kepuasannya. Kemudian dengan metode *elbow* dan *Davies Boildin* juga memberikan solusi pemilihan jumlah cluster yang tepat sehingga performanya lebih optimal dan menghasilkan hasil segmentasi pelanggan yang lebih akurat. Perhitungan metode algoritma K-Means memperoleh nilai respons yang paling dominan sebesar 259 menyatakan puas dan sebesar 44 menyatakan tidak puas dari 303 pelanggan sehingga dinyatakan dalam rumus Hasil Metode Algoritma K-Means menggunakan Uji Sensitivitas dan Spesifisitas sekitar 86% menyatakan puas dan 14% menyatakan tidak puas atas pelayanan yang disediakan pihak Badan Pelatihan Kesehatan pekanbaru

DAFTAR PUSTAKA

- A. Along, "Kualitas Layanan Administrasi Akademik di Politeknik Negeri Pontianak," *J. Ilm. Adm. Publik*, vol. 006, no. 01, pp. 94–99, 2020, doi: 10.21776/ub.jiap.2020.006.01.11.
- A. Badruttamam, S. Sudarno, and D. A. I. Maruddani, "PENERAPAN ANALISIS KLASSTER K-MODES DENGAN VALIDASI DAVIES BOULDIN INDEX DALAM MENENTUKAN KARAKTERISTIK KANAL YOUTUBE DI INDONESIA (Studi Kasus: 250 Kanal YouTube Indonesia Teratas Menurut Socialblade)," *J. Gaussian*, vol. 9, no. 3, pp. 263–272, 2020, doi: 10.14710/j.gauss.v9i3.28907.
- A. Hadiyatullah, D. Arifianto, and H. A. Al Faruq, "Pengelompokan Kecamatan Berdasarkan Hasil Produksi Komoditas Unggulan Di Kabupaten Jember Menggunakan Algoritma Partitioning Around Medoids," pp. 1–9, 2020.
- A. P. Nanda, D. E. H. Pramono, and S. Hartati, "Menentukan Tingkat Kepuasan Mahasiswa Terhadap Pelayanan Akademik Menggunakan Metode Algoritma K-Means," *J. Sist. Inf. dan Telemat.*, vol. 11, no. 1, pp. 23–28, 2020.
- A. Sulistiyawati and E. Supriyanto, "Implementasi Algoritma K-means Clustering dalam Penentuan Siswa Kelas Unggulan," *J. Tekno Kompak*, vol. 15, no. 2, p. 25, 2021, doi: 10.33365/jtk.v15i2.1162.
- A. Winarta and W. J. Kurniawan, "Optimasi Cluster K-means Menggunakan Metode Elbow pada Data Pengguna Narkoba dengan Pemrograman Python," *J. Tek. Inform. Kaputama*, vol. 5, no. 1, pp. 113–119, 2021.
- E. Muningsih and S. Kiswati, "Sistem Aplikasi Berbasis Optimasi Metode Elbow Untuk Penentuan Clustering Pelanggan," *Joutica*, vol. 3, no. 1, p. 117, 2018, doi: 10.30736/jti.v3i1.196.
- E. Muningsih, I. Maryani, and V. R. Handayani, "Penerapan Metode K-Means dan Optimasi Jumlah Cluster dengan Index Davies Bouldin untuk Clustering Propinsi

- Berdasarkan Potensi Desa,” *J. Sains dan Manaj.*, vol. 9, no. 1, pp. 95–100, 2021, [Online]. Available: <https://ejournal.bsi.ac.id/ejurnal/index.php/evolusi/article/view/10428/4839>
- E. Murtiningsih, R. Susanti, R. Susanti, D. R. Indriastuti, and D. R. Indriastuti, “Analisis Pengaruh Kualitas Pelayanan Dan Harga Terhadap Loyalitas Pelanggan Dengan Kepuasan Pelanggan Sebagai Variabel Mediasi,” *J. Ekon. Dan Kewirausahaan*, vol. 20, no. 1, pp. 74–95, 2020, doi: 10.33061/jeku.v20i1.4366.
- J. Hutagalung and F. Sonata, “Penerapan Metode K-Means Untuk Menganalisis Minat Nasabah,” *J. Media Inform. Budidarma*, vol. 5, no. 3, p. 1187, 2021, doi: 10.30865/mib.v5i3.3113.
- Khairil and D. A. Syafutra, “Penilaian Kepuasan Pelanggan Dengan Aplikasi Survei Pada Pdam Kota Bengkulu,” *Teknosia J. Ilm. Bid. Sains-Teknologi*, vol. 1, no. 1, pp. 16–21, 2021, doi: 10.33369/teknosia.v1i1.15542.
- Meisida, Novita dkk. 2014. K-Means Untuk Klasifikasi Penyakit Karies Gigi. KLIK Vol. 01No.01 ISSN: 2406- 7857
- M. F. M. Gultom, Dedek Kurniawan; Arif, “Determinasi Kepuasan Pelanggan Terhadap Loyalitas Pelanggan Melalui Kepercayaan Dedek,” *MANEGGIO J. Ilm. Magister Manaj.*, vol. 3, no. 2, pp. 273– 282, 2020, [Online]. Available: <http://jurnal.umsu.ac.id/index.php/MANEGGIO>
- M. G. Pradana, A. C. Nurcahyo, and P. H. Saputro, “Penerapan Metode K-Means Klustering untuk Menentukan Kepuasan Pelanggan K-Means Clustering Method to Determine Customer Satisfaction,” *Citec J.*, vol. 7, no. 1, pp. 42–50, 2020.
- M. M. Sinaga, “Implementasi Data Mining Untuk Mengelompokkan Kepuasan Belajar Siswa di Sekolah SMA Swasta Pencawan Medan Menggunakan Metode K-Means Clustering,” *J. Inf. Technol. ...*, vol. 1, pp. 117–125, 2022.
- M. Siregar, “Analisis Kepuasan Pelanggan Ompu Gende Coffee Medan,” *J. Divers.*, vol. 7, no. 1, pp. 114–120, 2021, doi: 10.31289/diversita.v7i1.5190.
- N. Rofiqo, A. P. Windarto, and D. Hartama, “Penerapan Clustering Pada Penduduk Yang Mempunyai Keluhan Kesehatan Dengan Datamining K-Means,” *KOMIK (Konferensi Nas. Teknol. Inf. dan Komputer)*, vol. 2, no. 1, pp. 216–223, 2018, doi: 10.30865/komik.v2i1.929.
- P. Penerimaan and M. Baru, “Yunita, Penerapan Data Mining Menggunakan Algoritma K-Means Clustering Pada Penerimaan Mahasiswa Baru (Studi Kasus : Universitas Islam Indragiri) 1 238,” vol. 7, no. September, pp. 238–249, 2018.
- R. NOVIANTO, “Penerapan Data Mining menggunakan Algoritma K-Means Clustering untuk Menganalisa Bisnis Perusahaan Asuransi,” *JATISI (Jurnal Tek. Inform. dan Sist. Informasi)*, vol. 6, no. 1, pp. 85–95, 2019, doi: 10.35957/jatisi.v6i1.150.
- Rohmawati N, Defiyanti S, Jajuli M. Implementasi Algoritma K-Means dalam Pengklasteran Mahasiswa Pelamar Beasiswa. Jurnal Ilmiah Teknologi Informasi Terapan. 2015; 01(02): 62- 68.
- S. Aulia, “Klasterisasi Pola Penjualan Pestisida Menggunakan Metode K-Means Clustering (Studi Kasus Di Toko Juanda Tani Kecamatan Hutabayu Raja),” *Djtechno J. Teknol. Inf.*, vol. 1, no. 1, pp. 1–5, 2021, doi:

10.46576/djtechno.v1i1.964.

- S. N. Br Sembiring, H. Winata, and S. Kusnasari, "Pengelompokan Prestasi Siswa Menggunakan Algoritma K-Means," *J. Sist. Inf. Triguna Dharma (JURSI TGD)*, vol. 1, no. 1, p. 31, 2022, doi: 10.53513/jursi.v1i1.4784.
- T. Humaira Mufidah, M. Risky, Z. Anshori, and D. Jenderal Penyelenggaraan Haji dan Umrah Kementerian Agama, "Ciledug Raya No.99, RT.10/RW.2, Petukangan Utara," *J. Media Infotama*, vol. 19,no. 1, pp. 3–4, 2023.