

Penerapan Data Mining Menggunakan Metode K-Means Untuk Mengetahui Minat Customer Dalam Pembelian Merchandise Kpop

Untung Surapati^{1*}, Magfiratul Jannah²

^{1,2} Program Studi Teknik Informatika, Sekolah Tinggi Ilmu Komputer Cipta Karya informatika, Jakarta Timur, Indonesia

¹kisuro2003@gmail.com, ²fira.kolut@gmail.com

Abstrak– Teknik data mining dengan menggunakan metode k-means untuk mengidentifikasi minat pembelian merchandise K-Pop oleh customer. Metode k-means digunakan untuk membagi data customer menjadi beberapa kelompok berdasarkan preferensi pembelian merchandise K-Pop. Data yang digunakan dalam penelitian ini berasal dari transaksi pembelian di toko merchandise K-Pop. Data tersebut diolah menggunakan algoritma k-means untuk membangun kelompok-kelompok customer yang memiliki preferensi pembelian merchandise K-Pop yang sama. Analisis kemudian dilakukan untuk mengetahui kesamaan antara kelompok-kelompok tersebut dan untuk mengevaluasi kesesuaian kelompok dengan kategori merchandise K-Pop yang tersedia di toko merchandise K-Pop. Hasil dari penelitian ini menunjukkan bahwa teknik data mining dengan menggunakan metode k-means dapat memberikan informasi yang berguna dalam mengidentifikasi minat pelanggan terhadap pembelian merchandise K-Pop. Dengan menggunakan teknik ini, peneliti dapat menghasilkan kelompok-kelompok customer yang memiliki preferensi yang sama dalam pembelian merchandise K-Pop. Setelah data dianalisis dengan metode K-Means, ditemukan bahwa terdapat 2 kelompok minat pembelian yang berbeda. Kelompok pertama memiliki minat yang rendah dalam pembelian merchandise K-Pop, sementara kelompok kedua memiliki minat yang tinggi. Analisis ini membantu untuk memahami minat Customer dan memberikan informasi penting bagi toko merchandise K-Pop untuk meningkatkan penjualan dan strategi pemasaran.

Kata Kunci: K-Means, Merchandise K-Pop, Customer, Minat, Data Mining

Abstract–Data mining techniques using the k-means method to identify customers' interest in purchasing K-Pop merchandise. The k-means method is used to divide customer data into several groups based on purchasing preferences for K-Pop merchandise. The data used in this research comes from purchase transactions at K-Pop merchandise stores. This data is processed using the k-means algorithm to build groups of customers who have the same purchasing preferences for K-Pop merchandise. Analysis is then carried out to determine the similarities between these groups and to evaluate the group's suitability to the K-Pop merchandise categories available at the K-Pop merchandise store. The results of this research show that data mining techniques using the k-means method can provide useful information in identifying customer interest in purchasing K-Pop merchandise. By using this technique, researchers can produce groups of customers who have the same preferences in purchasing K-Pop merchandise. After the data was analyzed using the K-Means method, it was found that there were 2 different purchasing interest groups. The first group has low interest in purchasing Kpop merchandise, while the second group has high interest. This analysis helps to understand customer interests and provides important information for Kpop merchandise stores to improve sales and marketing strategies.

Keywords: K-Means, KPOP Merchandise, Customers, Interests, Data Mining

I. PENDAHULUAN

Industri musik K-Pop atau Korean K-Pop telah menjadi sangat populer di seluruh dunia dalam beberapa tahun terakhir, terutama di kalangan remaja dan anak muda. Seiring dengan pertumbuhan popularitas ini, permintaan untuk merchandise K-Pop juga meningkat secara signifikan. Namun, penjual merchandise K-Pop perlu memahami minat customer mereka dengan lebih baik untuk dapat menghasilkan produk yang sesuai dengan kebutuhan dan preferensi mereka. Dengan menganalisis data pembelian customer, penjual dapat mengidentifikasi item merchandise yang paling populer dan memastikan bahwa stok mereka selalu terisi dengan barang yang dibutuhkan. Ini juga dapat membantu mengurangi biaya yang terkait dengan memproduksi barang yang tidak diinginkan atau tidak terjual. Selain itu, data mining juga dapat [1].

Metode Clustering Algoritma K-Means, Maka tujuan penulis Adalah pengelompokan data menggunakan metode k-means untuk meminimalkan fungsi objektif yang diset dalam proses pengelompokan, yang pada umumnya berusaha meminimalkan variasi didalam suatu kelompok dan memaksimalkan variasi antar kelompok impulsan diantaranya adalah sebagai berikut Sistem yang telah dibangun ini dapat menganalisis data penjualan Bibit Unggul dengan menerapkan metode K-Means,Dapat menerapkan hasil perhitungan dengan metode K-Means dalam pengelolaan data persediaan Bibit Padi Pada PT.Sang Hyang Seri,dan Dapat mengimplementasikan 8 Metode K-Means dalam sebuah sistem yang nantinya akan digunakan untuk menganalisis data penjualan Bibit Padi[2]. Maka tujuan penulis adalah dengan metode K-Means clustering dimana metode ini cocok dalam mengelompokkan kualitas lulusan mahasiswa berdasarkan indeks prestasi kumulatif (IPK) dan program studi, sehingga bisa

didapat kelompokkelompok (cluster-cluster) yang sesuai, dengan tujuan membantu pihak perguruan tinggi dalam menganalisa kualitas lulusan anak didiknya dan potensi program studinya Dengan menggunakan algoritma KMeans , proses untuk mengelompokkan mahasiswa berdasarkan IPK dan Prodi nya lebih mudah dibandingkan dengan cara manual. [3]. Metode K-Means Pada Pengelompokkan Pengangguran Di Indonesia, Maka tujuan penulis adalah menggunakan teknik Data mining yakni Algoritma K-means, Metode K-means merupakan Metode Clustering yang berfungsi untuk memecahkan dataset menjadi berkelompok. Hal ini untuk membantu pemerintah dalam memperluas lapangan pekerjaan untuk mengembangkan dan meningkatkan ekonomi di setiap provinsi di Indonesia dan dapat memberikan masukanterhadap pemerintah. Khususnya provinsi - provinsi yang minim dalam lapangan pekerjaan di Indonesia yang berdampak pada pengangguran[4]. algoritma KMeans Clustering pernah dilakukan untuk. Dari hasil pengujian menggunakan data hasil produksi periode januari 2021, didapat hasil Cluster Produktifitas Tinggi terdiri atas 26 blok lahan, Cluster Produktifitas Sedang terdiri atas 35 blok lahan, dan Cluster Produktifitas Rendah terdiri atas 39 blok lahan. Aplikasi ini juga dapat berfungsi dengan baik dan benar dengan tingkat akurasi 100%[5]. Algoritma K-means. Data akan diolah dengan melakukan clushtering dalam 3 clushter yaitu clushter tingkat keluhan kesehatan tinggi, clushter tingkat keluhan kesehatan sedang dan rendah. [6]. Metode tersebut bertujuan untuk 12 mengelompokkan data mahasiswa seesuai dengan skill dan basic yang didominasi pada mata kuliah yang paling banyak diminati sebagai acuan dalam pengembangannya [7]. K-means merupakan salah satu metode pengelompokkan data non-hierarki yang mempartisi data yang ada ke dalam bentuk dua atau lebih kelompok. [8]. K-means merupakan salah satu metode pengelompokkan data non-hierarki yang mempartisi data yang ada ke dalam bentuk dua atau lebih kelompok. Metode ini mempartisi data ke dalam kelompok sehingga data berkarakteristik sama dimasukkan ke dalam satu kelompok yang sama dan 13 data yang berkarakteristik berbeda dikelompokkan ke dalam kelompok yang lain. Metode k-means berusaha mengelompokkan data yang ada ke dalam beberapa kelompok, dimana data dalam suatu kelompok mempunyai karakteristik yang berbeda dengan data yang ada di dalam kelompok yang lain. [9]. Hasil akhir dari penelitian ini adalah mengelompokkan data produk yang terjual untuk mengetahui data yang memiliki potensi atau kecenderungan pelanggan dalam membeli barang tersebut. Menggukan Metode K-Means Clustering yang dirancang dapat digunakan[10]. Adakalanya produk yang baru saja muncul menjadi produk yang lebih unggul dalam hal penjualan, sehingga peminatan suatu produk tidak mutlak diukur dari tahun produk tersebut dirilis. Faktor kendalanya diantaranya dari sisi pemasaran produk saat diluncurkan. Penawaran produkproduk dengan premi murah disertai benefit sesuai yang diinginkan nasabah. Namun perusahaan asuransi masih kesulitan dalam hal

mengelompokkan produk-produk unggulan yang banyak diminati calon nasabah. Untuk itu diperlukan teknik mengelompokkan produk-produk asuransi agar memudahkan perusahaan melihat produk-produk unggulan dan memilih produk yang sesuai dengan kebutuhan nasabahnya. Melakukan analisis dan mengolah data dengan metode K-Means dalam klasterisasi produk asuransi merupakan tujuan penelitian ini. Penerapan algoritma K-Means ini untuk membantu perhitungan nilai kemurniannya dari hasil clustering yang dilakukan sehingga klasterisasi produk asuransi sesuai dengan kebutuhan nasabahnya. Penerapan metode K-Means dengan teknik clustering untuk data mining menghasilkan informasi produk asuransi yang lebih diminati para calon nasabah[11]. Metode K-Means clustering digunakan agar mempermudah dalam menganalisis serta mengelompokkan data untuk mengetahui beberapa daerah rawan bencana di wilayah Purbalingga [12]. K-Means pada Rapidminer dilakukan dengan memasukkan data stok produk yaitu stok awal, stok terjual dan stok akhir yang akan menjadi Database pada Ms.Excel, data tersebut kemudian dikoneksikan ke dalam Tools Rapidminer, dan akan diolah dan dibentuk Kmeans. Setelah itu, Rapidminer akan menghasilkan produk mana yang sangat laris, laris, dan kurang laris[13].

Perusahaan perlu melakukan pengelompokan untuk setiap produk untuk mengetahui barang mana yang paling diminati dan kurang diminati oleh pelanggan. Penelitian ini bertujuan 18 menentukan tiga kelompok cluster dengan kemiripan produknya agar dijadikan rekomendasi bagi manajemen perusahaan dalam merencanakan stok barang. Pada penelitian ditentukan 3 cluster dengan cluster 1 merupakan produk yang paling laris, cluster 2 produk yang laris dan yang cluster 3 produk yang kurang laris. Penelitian ini mengambil 5 kategori produk dengan masing-masing kategori ada 10 jenis produk. Hasil perhitungan dengan menggunakan Ms. Excel menunjukkan bahwa cluster 1 mempunyai 18 anggota, untuk cluster 2 mempunyai 6 anggota dan cluster 3 mempunyai 26 anggota. Dari perhitungan dengan Rapidminer didapat bahwa anggota pada cluster 1 sebanyak 22 anggota, cluster 2 sebanyak 5 anggota dan cluster 3 sebanyak 23 anggota. Evaluasi cluster dengan DBI didapatkan hasil yang cukup baik dengan nilai 0,431. Pengukuran akurasi, recall dan precission hasil perhitungan Ms.Excel mendapatkan nilai masing-masing 62%, 67% dan 59%. Untuk perhitungan menggunakan Rapidminer mendapatkan nilai akurasi 64%, recall 81% dan precission 88%. Hasil perbandingan clustering membuktikan perhitungan menggunakan Rapidminer mendapatkan nilai yang accuracy, recall dan precission yang lebih tinggi[14].

Tujuan dalam hal ini membahas penerapan data mining menggunakan algortima K-Means Clustering dalam pengelompokan keluarga yang tergolong miskin, sederhana dan kaya yang dilihat dari segi jumlah pendapatan perbulan, pengeluaran rata-rata perbulan dan total aset yang dimiliki, serta jumlah anak yang dimiliki

yang digolongkan berdasarkan tingkatan SD, SMP, dan SMA untuk menentukan jumlah dana bantuan yang akan diberikan berdasarkan tingkatannya. Output yang dihasilkan akan berupa laporan 3 tabel yang dimana tiap tabel berisi data keluarga yang tergolong miskin, sederhana dan kaya, dimana golongan sederhana ini nantinya bisa ditinjau kembali apakah berhak atau tidak mendapatkan bantuan program Kartu Indonesia Pintar. Dari hasil pengujian dengan menggunakan 200 data keluarga didapatkan tingkat akurasi dari hasil penerapan KMeans Clustering ini sebesar 69%[15].

II. METODE PENELITIAN

Pada bagian ini akan dijelaskan mengenai perancangan sistem dalam penelitian ini. Metode dalam penelitian ini dilakukan melalui beberapa tahap. Secara garis besar, alur penelitian dapat dijelaskan dibawah ini:

- Observasi dan wawancara Pada tahap ini akan dilakukan observasi langsung ke Toko Olypop_aai Store yang berlokasi di cakung kota jakarta untuk mendapatkan informasi terkait data yang dibutuhkan dalam menentukan minat customer dalam pembelian merchandise
- Wawancara Yaitu suatu bentuk penelitian yang dilakukan dengan mewawancarai pihak yang bersangkutan dalam organisasi untuk memperoleh data yang berhubungan dengan materi pembahasan.
- Studi literatur melakukan penelitian kepustakaan untuk mengumpulkan data dan informasi yang diperoleh dari perpustakaan, jurnal, artikel, dan internet yang berhubungan dengan penelitian ini

A. Data Mining

Data mining mengacu pada metode analisis bisnis yang melampaui hitungan, teknik deskriptif, pelaporan, dan metode berdasarkan aturan bisnis. Meskipun kami memperkenalkan visualisasi data, yang biasanya merupakan langkah pertama menuju analitik yang lebih maju, buku ini sebagian besar berfokus pada alat analisis data yang lebih canggih. Secara khusus, ini mencakup metode statistik dan pembelajaran mesin yang menginformasikan pengambilan keputusan, seringkali secara otomatis[1].

B. Metode K-Means

K-means menginisialisasi cluster mean dengan menghasilkan titik k secara acak di ruang data. Ini biasanya dilakukan dengan menghasilkan nilai secara seragam secara acak dalam rentang untuk setiap dimensi[2]. Berikut adalah langkah-langkah umum dalam algoritma k-means:

- Inisialisasi: Tentukan jumlah kluster yang diinginkan (k). Pilih secara acak k titik pusat awal (centroid) dari data.
- Alokasi: Hitung jarak antara setiap data dengan centroid dan alokasikan data ke kluster dengan centroid terdekat.
- Pembaruan centroid: Hitung ulang centroid baru untuk setiap kluster berdasarkan rata-rata dari data yang termasuk dalam kluster tersebut.

- Iterasi: Ulangi langkah kedua dan ketiga sampai tidak ada perubahan alokasi kluster atau mencapai kriteria penghentian yang ditentukan (misalnya, jumlah iterasi maksimum).

- Hasil: Setelah iterasi selesai, data akan terkluster menjadi k kelompok berdasarkan centroid yang dihasilkan.

C. Minat Customer

Untuk meningkatkan minat pelanggan, perusahaan dapat mengadopsi strategi pemasaran yang efektif, seperti melakukan riset pasar untuk memahami kebutuhan dan preferensi pelanggan, meningkatkan kualitas produk atau layanan, memberikan nilai tambah, memberikan pengalaman pengguna yang positif, serta berkomunikasi secara efektif dengan pelanggan melalui berbagai saluran komunikasi yang relevan. Penting bagi perusahaan untuk terus memantau dan memahami minat pelanggan, serta beradaptasi dengan perubahan tren dan preferensi pelanggan untuk tetap relevan dan kompetitif di pasar[3].

D. Pembelian

Pembelian adalah tindakan atau proses membeli atau memperoleh sesuatu dengan cara membayar harga yang ditetapkan. Dalam konteks ekonomi, pembelian merujuk pada aktivitas memperoleh barang atau jasa dengan menggunakan uang atau bentuk pembayaran lainnya[4].

E. Merchandise Kpop

Merchandise K-pop adalah produk-produk yang secara resmi dihasilkan dan didistribusikan oleh agensi atau perusahaan yang mengelola grup musik K-pop. Merchandise ini didesain dengan tema dan gambar yang berkaitan dengan grup atau anggota K-pop tertentu, dan biasanya mencakup berbagai jenis barang seperti album fisik, pakaian, aksesoris, poster, pernak-pernik, dan banyak lagi.

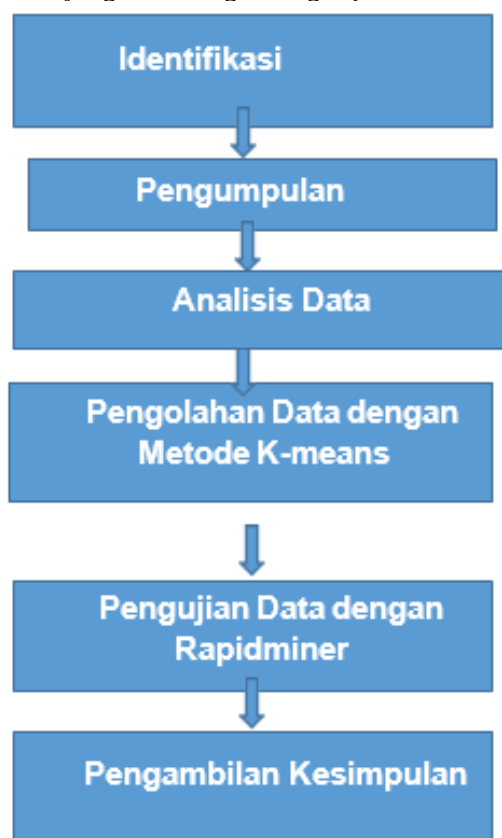
F. Rapidminer

RapidMiner merupakan perangkat lunak yang bersifat terbuka (open source). RapidMiner adalah sebuah solusi untuk melakukan analisis terhadap data mining, teks mining dan analisis prediksi. RapidMiner menggunakan berbagai teknik deskriptif dan prediksi alam memberikan wawasan kepada pengguna sehingga dapat membuat keputusan yang paling baik. RapidMiner sebelumnya bernama YALE (Yet Another Learning Environment), dimana versi awalnya mulai dikembangkan pada tahun 2001 oleh Ralf Klinkenberg, Ingo Mierswa, dan Simon Fischer di Artificial Intelligence Unit dari University of Dortmund. RapidMiner didistribusikan di bawah lisensi AGPL (GNU Affero General Public License) versi 3. Hingga saat ini telah ribuan aplikasi yang dikembangkan menggunakan RapidMiner di lebih dari 40 negara. RapidMiner sebagai software open source untuk data mining tidak perlu diragukan lagi karena software ini sudah terkemuka di dunia. RapidMiner menempati peringkat pertama sebagai Software data mining pada polling oleh KDnuggets, sebuah portal data-mining pada 2010-2011[7].

G. Penelitian Lapangan

Yaitu penelitian yang dilakukan pada perusahaan bersangkutan untuk memperoleh data yang berhubungan dengan penulisan dengan cara:

- Observasi dan wawancara Pada tahap ini akan dilakukan observasi langsung ke Toko Olypop_aini Store yang berlokasi di Cakung Kota Jakarta untuk mendapatkan informasi terkait data yang dibutuhkan dalam menentukan minat customer dalam pembelian merchandise Kpop.
- Wawancara Yaitu suatu bentuk penelitian yang dilakukan dengan mewawancarai pihak yang bersangkutan dalam organisasi untuk memperoleh data yang berhubungan dengan materi pembahasan.
- Studi literatur melakukan penelitian kepustakaan untuk mengumpulkan data dan informasi yang diperoleh dari perpustakaan, jurnal, artikel, dan internet yang berhubungan dengan penelitian ini:



Gambar 1. Alur Penelitian

Tahapan Penelitian Ini dapat digambarkan sebagai berikut

- Identifikasi Masalah Yaitu melakukan permasalahan yang terdapat di toko yaitu untuk mengetahui produk mana yang penjualan rendah dan penjualan tingkat tinggi sebagai strategi utama penyediaan stok merchandise Kpop di toko.
- Pengumpulan Data Mengumpulkan data yaitu menerima data stok merchandise Kpop
- Analisis Data Yaitu menganalisis data stok merchandise Kpop di toko Olypop_aini

- Pemrosesan Data Yaitu dengan mengolah data-data yang diberikan oleh toko Olypop_aini, kemudian data tersebut diolah dengan menggunakan metode K-means clustering yang dihitung dengan Excel untuk mengklasifikasikan produk yang penjualan rendah dan tinggi.
- Uji Data Yaitu dengan menguji data yang sudah diolah dengan RapidMiner
- Kesimpulan Yaitu menarik kesimpulan dari hasil perhitungan dan pengolahan data yang dihitung secara manual dengan data Rapidminer. Data yang digunakan dalam penelitian ini adalah data stok merchandise Kpop di toko Olypop_aini. Berikut data-data yang diolah, berupa stok merchandise Kpop:

H. Alat Penelitian

Dalam proses untuk menentukan minat customer, peneliti menggunakan Aplikasi Rapidminer untuk pengolahan data dan menggunakan aplikasi Excel dengan metode elbow method untuk menentukan nilai K berapa cluster yang paling bagus untuk data yang akan diolah.

$$d = |x - y| = \sqrt{\sum_{i=1}^n (x_i - y_i)^2}$$

I. Implementasi metode K-Means

Metode k-means adalah salah satu algoritma klustering yang populer dalam analisis data. Dalam melakukan pengujian ini kita menggunakan metode k-means untuk menentukan cluster pada data yang diolah. Langkah-langkah dalam mengcluster menggunakan metode K-Means adalah sebagai berikut :

- Tentukan nilai k nya sebagai jumlah kluster yang akan dibentuk.
- Tentukan titik k atau centroid secara acak
- Hitunglah jarak setiap data input masing – masing centroid menggunakan rumus jarak Euclidean (Euclidean Distance) sampai ditemukan jarak yang terdekat dari setiap data dengan centroid.

Berikut adalah persamaan Euclidean Distance :

(1)

- Mengklasifikasi data berdasarkan kedekatannya dengan centroid.
- Hitunglah kembali pusat kluster dengan anggota cluster yang sekarang. Pusat cluster ialah nilai rata-rata dari semua data objek dalam cluster tertentu.
- Hitung lagi setiap objek memakai pusat kluster yang baru. Jika pusat cluster tidak berubah lagi maka proses klustering selesai. Atau, kembali ke langkah nomor 3 sampai pusat kluster tidak berubah lagi.

J. Elbow Method

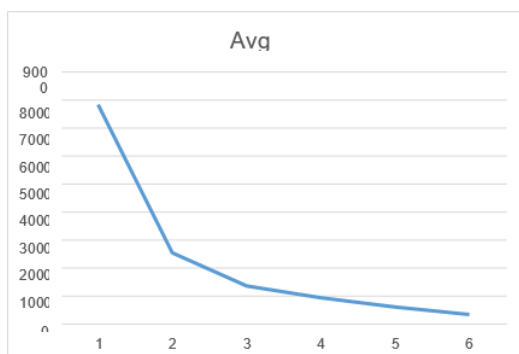
Metode Elbow (Elbow Method) adalah sebuah metode yang digunakan untuk menentukan jumlah optimal dari cluster dalam analisis klusterisasi (clustering).

Tabel 1. Data Avg.dist

No	Avg.within centroid distance
1	7834.311
2	2548.192
3	1365.086

4	948.419
5	615.086
6	349.813
7	7834.311

Data diatas adalah data yang diambil dari pegujian performancevector yang berupa centroid 2 sampai 7 untuk menentukan nilai k cluster yang bagus untuk data yang akan diolah.



Gambar 2. Elbow method.

Pada gambar 2 Elbow Method (siku) nilai k cluster yang paling bagus untuk dibentuk berapa cluster adalah 2, karna siku grafik berada di angka 2 dengan nilai Avg dist 2548.192.

K. Pengujian Menggunakan Excel

Dalam melakukan clustering, data yang diperoleh akan dihitung terlebih dahulu berdasarkan parameter yang ada yaitu transaksi dan jumlah penjualan seperti yang ditunjukkan pada gambar 3 berikut:

Nama Data	Transaksi	Penjualanm Produk
ALBUM	300	500
photobook	250	400
poster	150	250
Kaus dan Hoodie	50	50
Lighstick	150	150
tumbler	7	10
gantungan kunci	25	30
Clear bag	30	30
tempat pensil	15	20
Mug Custom	100	120
Kipas Tangan	50	80
Boneka	200	250
Spanduk Mini	20	20
Stiker	150	300
DVD	200	200

Gambar 3. Pengumpulan Data

Kemudian data tersebut akan masuk ketahapan clustering dengan menerapkan algoritma k-means.

- Menentukan Jumlah Cluster Menentukan jumlah cluster yang digunakan pada data penjualan sebanyak 2 cluster diantaranya penjualan rendah dan tingkat tinggi.
- Clustering Data Penentuan pusat awal cluster (centroid) ditentukan secara random atau acak yang diambil dari data yang ada. Nilai cluster 0 diambil dari data mug custom dan nilai cluster 1 pada data poster.

Centroid	TS	PP
C0	100	120
C1	150	250

Gambar 4. Penentuan Centroid Awal

- Menghitung jarak antara titik centroid dengan titik tiap objek dengan menggunakan Euclidian Distance. Adapun penghitungan centroid awal secara manual. Perhitungannya adalah sebagai berikut :

$$d = |x - y| = \sqrt{\sum_{i=1}^n (x_i - y_i)^2} \quad (2)$$

Nama Data	TS	PP	C1	C2	JARAK TERDEKAT	Cluster 1	Cluster 2
ALBUM	300	500	429,4182111	291,5475947	291,5475947	0	0
photobook	250	400	317,6476035	180,2775638	180,2775638	0	0
poster	150	250	139,2838828	0	0	0	0
stiker	150	300	186,8154169	50	50	0	0
Boneka	200	250	164,0121947	50	50	0	0
DVD	200	200	128,0624847	70,71067812	70,71067812	0	0
clear bag	30	30	114,0175425	250,5992817	114,0175425	1	1
Lighstick	150	150	58,30951895	100	58,30951895	1	1
tempat pensil	15	20	131,2440475	266,6927071	131,2440475	1	1
Mug Custom	100	120	0	139,2838828	0	1	1
Kipas Tangan	50	80	64,03124237	197,2308292	64,03124237	1	1
Kaus dan Hoodie	50	50	86,02325267	223,6067977	86,02325267	1	1
Spanduk Mini	20	20	128,0624847	264,1968963	128,0624847	1	1
Gantungan kunci	25	30	117,1537451	253,0316186	117,1537451	1	1
tumbler	7	10	144,0451318	279,3725112	144,0451318	1	1

Gambar 5. Perhitungan Jarak dan Pengelompokan Data Iterasi Ke-1

Dari hasil perhitungan di atas didapatkan pusat cluster baru seperti sebagai berikut:

Centroi baru	TS	PP
C0	49,6667	56,6667
C1	208,333	316,667

Gambar 6. centroid baru literasi-1

Iterasi selanjutnya melakukan perhitungan lagi dengan menggunakan centroid Tabel yang baru Iterasi Ke-1

Nama Data	TS	PP	C1	C2	JARAK TERDEKAT	Cluster
ALBUM	300	500	509,1279036	204,9728979	204,9728979	0
photobook	250	400	397,5062543	93,16949906	93,16949906	0
poster	150	250	217,8177118	88,58454844	88,58454844	0
Boneka	200	250	244,903836	67,18548124	67,18548124	0
stiker	150	300	263,2069317	60,66758241	60,66758241	0
DVD	200	200	207,7126755	116,9639071	116,9639071	0
clear bag	30	30	33,13440642	337,6100644	33,13440642	1
lightstick	150	150	137,0324374	176,5801675	137,0324374	1
tempat pensil	15	20	50,46010525	354,1029354	50,46010525	1
Mug Custom	100	120	80,8984274	224,5303741	80,8984274	1
Kipas Tangan	50	80	23,33571416	284,7464759	23,33571416	1
Kaos dan Hoodie	50	50	6,674994798	310,1299011	6,674994798	1
Spanduk Mini	20	20	47,16519432	351,3980016	47,16519432	1
Gantungan kunci	25	30	36,32568727	340,2776644	36,32568727	1
tumbler	7	10	63,23149707	366,8508628	63,23149707	1

Gambar 7. centroid baru literasi-1

Karena pada Iterasi Ke-1 dan Ke-2 posisi cluster tidak berubah maka iterasi dihentikan dan hasil akhir yang diperoleh yaitu:

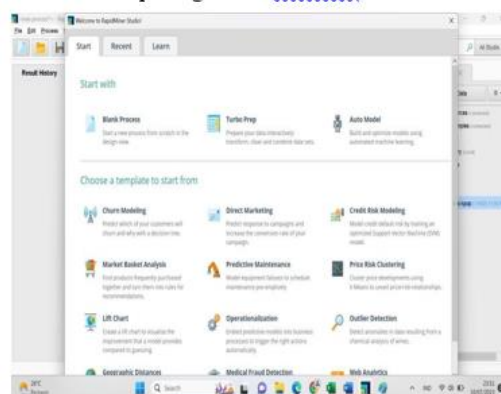
Hasil dari analisa algoritma K-means diatas dengan tujuan untuk mengetahui minat customer pada pembelian merchandise kpop yang terdapat pada toko Olypop_aii ini dihasilkan dari data yang telah dihitung dengan Ms excel.

Perhitungan dilakukan dengan terlebih dahulu membagi data yang ada kedalam beberapa cluster berdasarkan dari dua parameter yang ada yaitu transaksi dan jumlah penjualan. Karena pada Iterasi Ke-1 dan Ke-2 posisi cluster tidak berubah maka iterasi dihentikan dan hasil akhir yang diperoleh yaitu:

- C0 memiliki 6 data yang diartikan bahwa kelompok pertama adalah merchandise yang penjualan rendah.
- C1 memiliki 9 data yang diartikan bahwa kelompok kedua adalah merchandise dengan penjualan yang tinggi.

L. Pengujian Menggunakan Software Rapidminer

Pada Implementasi Dan Pengujian disini kita menggunakan sebuah Software RapidMiner Studio Versi10.1, dengan pengujian data menggunakan software kita akan membandingkan bagaimana hasil pengolahan data secara manual dengan hasil pengolahan data menggunakan sebuah software. Pada gambar 8 merupakan halaman utama dari aplikasi RapidMiner Studio v.10.1 saat pertama kali membuka aplikasi setelah proses loading selesai terlihat pada gambar berikut :

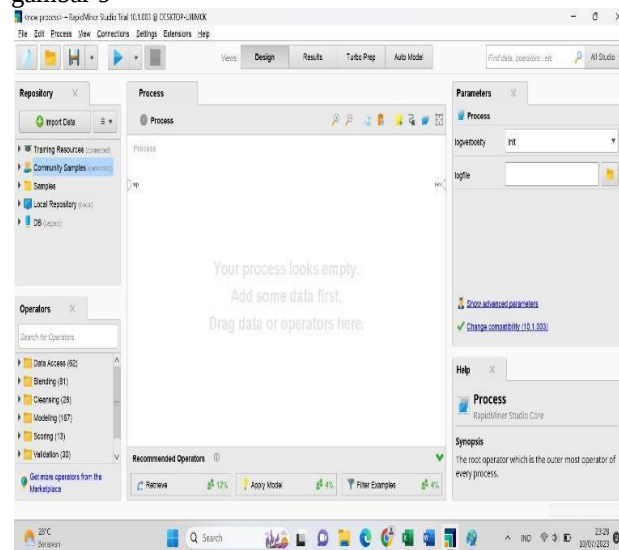


Gambar 8. Halaman Utama RapidMiner Studio v.10.1

Pada tampilan halaman utama ada lima menu yang akan digunakan yaitu :

- Start berfungsi untuk membuat halaman proses kerja Data Mining yang baru.
- Recent berfungsi untuk membuka proses yang sudah ada direpository sebelumnya.
- Learn berisikan petunjuk-petunjuk menggunakan RapidMiner Studio v.10.1.

Berikut ini adalah tampilan New Process yang akan digunakan untuk memulai proses kerja baru terlihat pada gambar 9

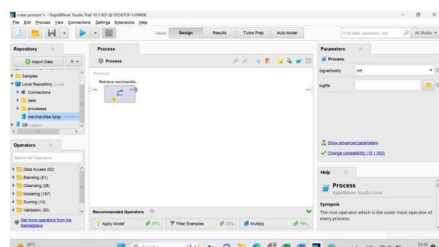


Gambar 9. New Process RaidMiner Studio v.10.1

Tampilan new process adalah untuk membuat halaman kerja pada RapidMiner Studio v.10.1 Import data dilakukan untuk memasukkan data yang akan diuji dalam bentuk format .xls atau .xlsx. Berikut adalah cara untuk melakukan import file Microsoft Excel. Untuk membuat mengimport data yang akan diproses, maka dilakukan New process. Setelah selesai maka akan tampil read excel pada RapidMiner Studio v.10.1, Seperti terlihat pada Gambar 10. Setelah itu lanjutkan dengan mengklik Tab my computer untuk melakukan menginputkan data dari Microsoft Excel (dengan format .xls, atau .xlsx).



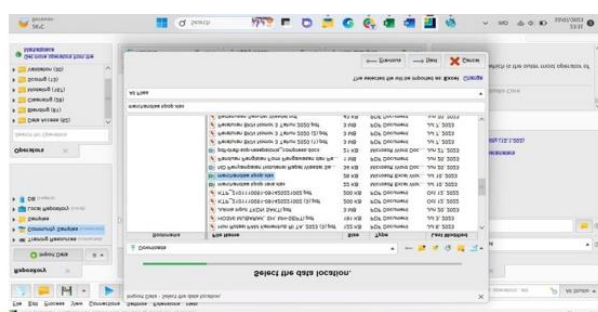
Gambar 10. Tampilan Import Data



Gambar 11. Tampilan Operator Baru Retrieve Excel.

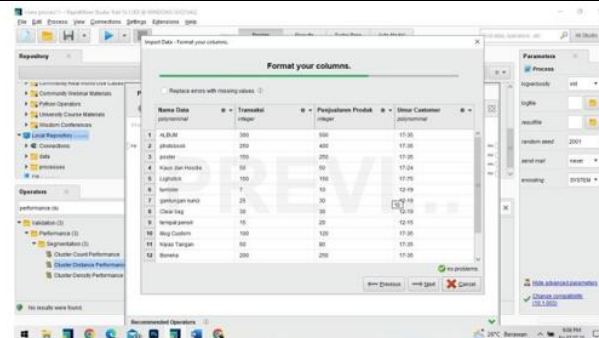
Dalam melakukan import data pada aplikasi RapidMiner Studio v.10.1 terdiri dari 4 tahap, adapun langkah-langkahnya sebagai berikut:

- Data Import – Step 1 of 4 Tahap pertama adalah mencari lokasi file yang sudah dibuat sebelumnya dengan format .xlsx . atau xls, pilih dan kemudian simpan. Pada pengujian ini data yang akan diuji disimpan dengan nama merchandise kpop.xlsx lalu pilih diklik, seperti Gambar 12.



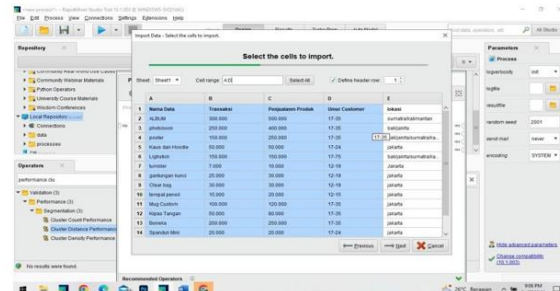
Gambar 12. Tampilan Data Import 1 of 4

- Data Import – Step 2 of 4 Kemudian pilih next maka lanjut ke tahap berikutnya yaitu impor step 2 of 4. Pada tahap ini akan dipilih Sheet yang berisikan sumber data pada Microsoft Excel yang digunakan. Maka akan terlihat data import seperti terlihat pada gambar 13



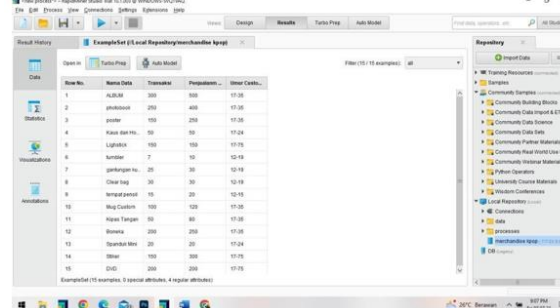
Gambar 13. Tampilan Data Import Step 2 of 4

- Data Import 3 of 4, tahap ini merupakan tahap penentuan tipe data dan atribut. Sebenarnya RapidMiner Studio akan memberikan tipe data yang tepat secara otomatis. Namun, jika kita merasa tipe data yang diberikan RapidMiner Studio tidak cocok, kita bisa mengubahnya / processing data kemudian klik tombol Next.



Gambar 14. Tampilan Data Import Step 2 of 4

- Data Import Step 4 of 4 Pada Step 4 Tahap import data selesai dengan memilih Finish, maka pada Main Process terlihat sebuah operator baru Excel yang sudah berisi file merchandise kpop.xlsx yang di import langsung dari file excel. Data tersebut siap dilakukan pengujian.



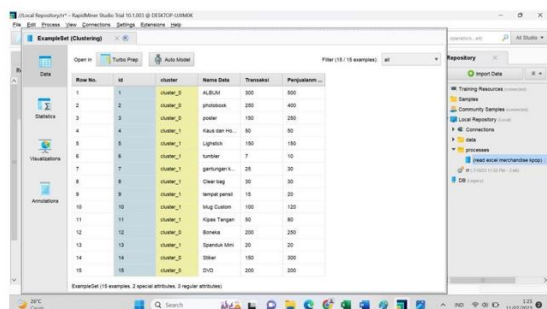
Gambar 14. Tampilan Data Import Step 2 of 4

III. HASIL DAN PEMBAHASAN

A. Tahapan Metode K-Mens

Pada gambar 15 di atas dilakukan pengaturan nilai k, di mana k merupakan nilai yang digunakan untuk menentukan jumlah cluster yang akan dibentuk. Di sini jumlah cluster yang akan dibentuk adalah sebanyak 2 cluster sesuai tingkatan. Hubungkan data dengan clustering

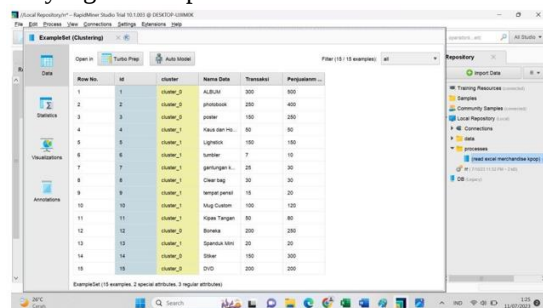
kmeans untuk mengetahui output ke arah result. Apabila sudah terhubung maka tentukan jumlah cluster dan maksimal runs yang diberikan. Terakhir klik Tombol Play. Selanjutnya melakukan pengaturan algoritma K-Means, yang diatur pada menu Parametere Clustering K-Means.



Gambar 15. Tampilan untuk proses K-means

B. Running Data

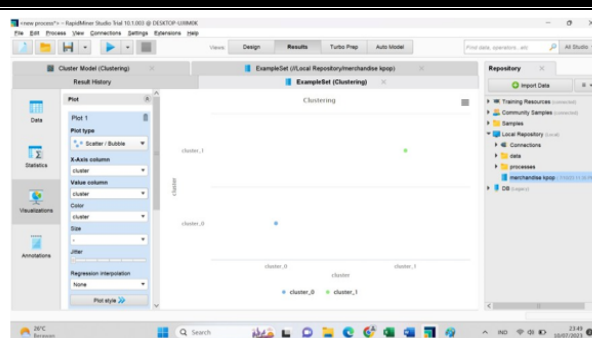
Setelah proses running, akan tampil Example Set, seperti yang terlihat pada Gambar 16.



Gambar 16. Tampilan Running Data

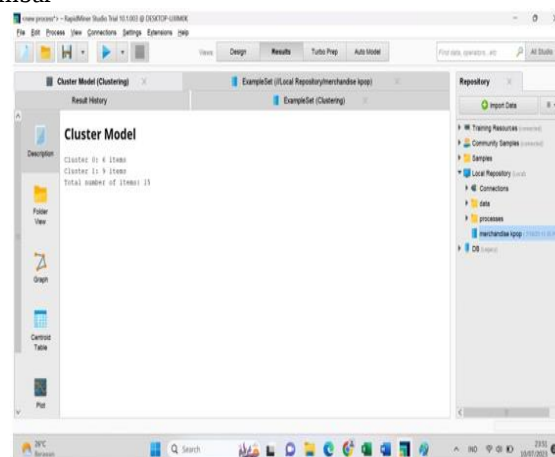
Pada bagian Example Set terdapat beberapa bentuk hasil cluster yaitu: Data View, Statistics View, visualizations, dan Annotations. Untuk Cluster model terdapat Description, Folder View, Graph View, Centroid Table, Centroid Plot View dan Annotations. Pada Hasil Pengujian Data terdapatlah beberapa output yang dihasilkan oleh software RapidMiner v.10.1 yaitu :

- ExampleSet Pada ExampleSet dapat dilihat beberapa tampilan hasil cluster, yaitu Data View. Data View merupakan tampilan hasil cluster data secara keseluruhan sesuai dengan data yang telah diinputkan
- Visualizations merupakan tampilan grafik hasil pengelompokan atau cluster sampel data merchandise kpop dengan 2 cluster.

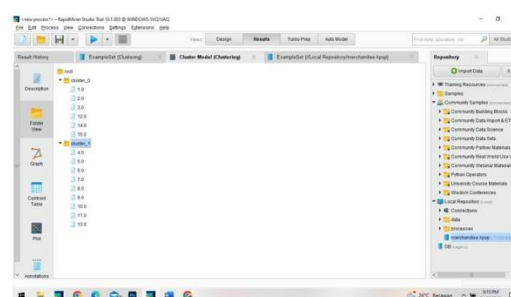


Gambar 16. Tampilan Scatter pada Visualizations

Cluster Model (Clustering) Pada Cluster Model (Clustering) dapat dilihat beberapa tampilan hasil cluster, yaitu description yang merupakan tampilan hasil pengelompokan berdasarkan cluster dan jumlah anggotanya. Tampilan description akan terlihat pada gambar. Tampilan Text View pada akan terlihat pada gambar



Gambar 17. Tampilan Cluster Model



Gambar 18. Tampilan Folder View Root Cluester

Folder View merupakan tampilan data bagian-bagian cluster secara keseluruhan. Pada penelitian ini diperoleh hasil 2 cluster. Cluster(0) merupakan cluster yang rendah dan Cluster (1) merupakan cluster yang tinggi. Cluster(0) memiliki data 6 dan cluster (1) memiliki data 9. Cluster 0 masuk kategori tingkat minat yang rendah dalam pembelian merchandise kpop sedangkan Cluster 1 masuk kategori tingkat tinggi dalam pembelian merchandise kpop. Data yang masuk cluster 0 adalah album, photobook,

poster, boneka, dvd, dan stiker sedangkan data yang masuk cluster 1 adalah kaus dan hoodie, lighstick, tumbler, gantungan kunci, clear bag, tempat pensil, mug custom, kipas tangan dan panduk mini. Pada perhitungan excel dan rapid miner hasil clusternya sama dan data yang termasuk dalam cluster juga sama.

IV. KESIMPULAN

Berdasarkan hasil dari analisa dan pengolahan data pada pembahasan diatas, dapat ditarik kesimpulan sebagai berikut: Hasil dari penelitian ini menunjukkan bahwa teknik data mining dengan menggunakan metode k-means dapat memberikan informasi yang berguna dalam mengidentifikasi minat pelanggan terhadap pembelian merchandise KPOP. Dengan menggunakan teknik ini, peneliti dapat menghasilkan kelompok-kelompok customer yang memiliki preferensi yang sama dalam pembelian merchandise KPOP. Metode data mining dalam menganalisis minat customer data dianalisis dengan metode K-Means dan pengujian excel dan software rapidminer ditemukan bahwa terdapat 2 kelompok minat pembelian yang berbeda. Kelompok pertama memiliki minat yang rendah dalam pembelian merchandise Kpop, sementara kelompok kedua memiliki minat yang tinggi. Cluster(0) merupakan cluster minat rendah dalam pembelian merchandise kpop dan cluster 1 memiliki minat tinggi. C0 memperoleh 6 itemss dan C1 memperoleh 9 items. Analisis ini membantu untuk memahami minat Customer dan memberikan informasi penting bagi toko merchandise Kpop untuk meningkatkan penjualan dan strategi pemasaran.

UCAPAN TERIMA KASIH

Saya ingin mengucapkan terima kasih yang sebesar-besarnya atas bimbingan, dukungan, dan dedikasi Bapak Untung Surapati dalam membantu pembuatan jurnal ini. Tanpa panduan dan dorongan yang berharga dari Anda, pencapaian ini tidak akan mungkin terwujud. Saya sangat berterimakasih atas waktu dan pengetahuan yang Anda luangkan untuk membimbing saya dalam menyusun jurnal ini. Dalam setiap pertemuan dan diskusi, Bapak telah memberikan wawasan yang berharga, saran yang berharga, dan arahan yang tepat. Bimbingan Bapak telah memainkan peran penting dalam membantu saya memahami proses penelitian, mengembangkan metodologi yang tepat, dan menganalisis data dengan benar. Selain itu, terima kasih juga karena telah memberikan koreksi dan umpan balik konstruktif dalam setiap tahap penulisan jurnal. Pengamatan dan penilaian Bapak yang mendalam telah membantu saya untuk meningkatkan kualitas tulisan dan memperbaiki kelemahan yang ada. Saya sangat menghargai kesabaran dan ketelitian yang Anda tunjukkan dalam membimbing saya menuju hasil akhir yang memuaskan. Lebih dari sekadar menjadi pembimbing, Bapak juga menjadi inspirasi bagi saya. Kepedulian dan semangat Bapak terhadap penelitian dan pengembangan ilmu pengetahuan telah mendorong saya untuk terus

belajar dan berkembang. Saya merasa sangat beruntung dan bersyukur dapat belajar dari Bapak, seorang akademisi yang berkompeten dan berdedikasi. Terima kasih karena telah memberikan kesempatan kepada saya untuk terlibat dalam penelitian ini dan memperluas pengetahuan serta wawasan saya. Saya meyakini bahwa karya tulis ini tidak hanya akan memberikan manfaat bagi saya secara pribadi, tetapi juga akan memberikan kontribusi yang berarti dalam perkembangan bidang ilmu yang kita geluti. Sekali lagi, terima kasih yang tak terhingga atas semua bimbingan, dorongan, dan kepercayaan yang Bapak berikan kepada saya. Saya sangat beruntung dan bangga dapat belajar di bawah arahan Bapak. Semoga keberhasilan ini menjadi bukti nyata dari dedikasi dan komitmen Bapak dalam membentuk generasi muda yang berkualitas dan berkontribusi positif dalam dunia akademik.

V. REFERENSI

- [1] F. Amin, D. S. Anggraeni, and Q. Aini, "Penerapan Metode K-Means dalam Penjualan Produk Souq.Com," *Appl. Inf. Syst. Manag.*, vol. 5, no. 1, pp. 7–14, Apr. 2022, doi: 10.15408/aism.v5i1.22534.
- [2] H. Annur, "Penerapan Data Mining Menentukan Strategi Penjualan Variasi Mobil Menggunakan Metode K-Means Clustering (Studi Kasus Toko Luxor Variasi Gorontalo)," 2019.
- [3] A. Asroni, H. Fitri, and E. Prasetyo, "Penerapan Metode Clustering dengan Algoritma K-Means pada Pengelompokan Data Calon Mahasiswa Baru di Universitas Muhammadiyah Yogyakarta (Studi Kasus: Fakultas Kedokteran dan Ilmu Kesehatan, dan Fakultas Ilmu Sosial dan Ilmu Politik)," *Semesta Tek.*, vol. 21, no. 1, 2018, doi: 10.18196/st.211211.
- [4] H. Effendi, A. Syahrial, S. Prayoga, and W. D. Hidayat, "Penerapan Metode K-Means Clustering untuk Pengelompokan Lahan Sawit Produktif pada PT Kasih Agro Mandiri APPLICATION OF K-MEANS CLUSTERING METHOD FOR GROUPING PRODUCTIVE PALM OIL LAND AT PT KASIH AGRO MANDIRI," *TEKNOMATIKA*, vol. 11, no. 02, pp. 1–5, 2021.
- [5] E. Fammaldo and L. Hakim, "PENERAPAN ALGORITMA K-MEANS CLUSTERING UNTUK PENGELOMPOKAN TINGKAT KESEJAHTERAAN KELUARGA UNTUK PROGRAM KARTU INDONESIA PINTAR," 2018.
- [6] K. Fatmawati and A. P. Windarto, "DATA MINING: PENERAPAN RAPIDMINER DENGAN K-MEANS CLUSTER PADA DAERAH TERJANGKIT DEMAM BERDARAH DENGUE (DBD) BERDASARKAN PROVINSI," 2018. [Online]. Available: <https://www.depkes.go.id/>.

-
- [7] D. I. T. Hijab, "PENERAPAN DATA MINING MENGGUNAKAN ALGORITMA K-MEANS UNTUK MENGETAHUI MINAT CUSTOMER," vol. 15, no. 2, pp. 241–246, 2019, doi:10.33480/pilar.v15i2.650.
 - [8] J. Hutagalung and F. Sonata, "Penerapan Metode K-Means Untuk Menganalisis Minat Nasabah Asuransi," vol. 5, pp. 1187–1194, 2021, doi: 10.30865/mib.v5i3.3113.
 - [9] J. Nasir, "PENERAPAN DATA MINING CLUSTERING DALAM MENGELOMPOKAN BUKU DENGAN METODE K-MEANS," J. SIMETRIS, vol. 11, no. 2, 2020.
 - [10] V. Novita Sari and D. Maharani, "PENERAPAN METODE K-MEANS CLUSTERING DALAM MENENTUKAN PREDIKAT KELULUSAN MAHASISWA UNTUK MENGANALISA KUALITAS LULUSAN," vol. IV, no. 2, pp. 133–140, 2018.
 - [11] S. Nurajizah and A. Salbinda, "Penerapan Data Mining Metode K-Means Clustering Untuk Analisa Penjualan Pada Toko Fashion Hijab Banten," J. Tek. Komput. AMIK BSI, vol. 7, no. 2, 2021, doi: 10.31294/jtk.v4i2.
 - [12] R. Oktavia et al., "Penerapan Metode Algoritma K-means Dalam Pengelompokan Angka Harapan Hidup Saat Lahir Menurut Provinsi." [Online]. Available: www.bps.go.id/
 - [13] A. Prasatya, R. R. A. Siregar, and R. Arianto, "Penerapan Metode K-Means Dan C4.5 Untuk Prediksi Penderita Diabetes," PETIR, vol. 13, no. 1, pp. 86–100, Mar. 2020, doi: 10.33322/petir.v13i1.925.
 - [14] D. I. Ramadhani, O. Damayanti, O. Thaushiyah, and A. R. Kadafi, "Penerapan Metode K-Means Untuk Clustering Desa Rawan Bencana Berdasarkan Data Kejadian Terjadinya Bencana Alam," JURIKOM (Jurnal Ris. Komputer), vol. 9, no. 3, p. 749, Jun. 2022, doi: 10.30865/jurikom.v9i3.4326.
 - [15] N. Rofiqo, A. P. Windarto, and D. Hartama, "KOMIK (Konferensi Nasional Teknologi Informasi dan Komputer) PENERAPAN CLUSTERING PADA PENDUDUK YANG MEMPUNYAI KELUHAN KESEHATAN DENGAN DATAMINING K-MEANS", [Online]. Available: <http://ejurnal.stmik-budidarma.ac.id/index.php/komik>