ANALISA PENERAPAN METODE *CLUSTERING K-MEANS* UNTUK PENGELOMPOKAN DATA TRANSAKSI KONSUMEN (Studi Kasus: Cv. Mitra Indexindo Pratama)

Masjunedi, Nana Suarna, Yudhistira Arie Wijaya
Program Studi Teknik Informatika S1, STMIK IKMI Cirebon
STMIK IKMI Cirebon, Jalan Perjuangan No. 10B Majasem, Kota Cirebon, Indonesia
Masjunnn417@gmail.com

ABSTRAK

Cv. Mitra indexindo pratama merupakan perusahaan yang bergerak di bidang penjualan dan jasa, Perusahaan juga menyediakan jasa perbaikan dan perawatan komputer untuk usaha/kantor yang membutuhkan dan juga melayani konsumen perorangan. Cv.Mitra Indexindo Pratama termasuk perusahaan perseorangan, karena semua sumber daya yang digunakan dalam operasi perusahaan adalah milik sendiri atau perseorangan, Data transaksi penjualan semakin lama semakin banyak sehingga akan terjadi penumpukan data, penumpukan data tsb kemudian akan di olah menggunakan aplikasi rapid miner. Akumulasi pengetahuan ini dapat digunakan untuk menemukan informasi yang berguna dan tersembunyi yang bisa digunakan untuk menerapkan strategi penjualan. Pengumpulan data tersebut dapat diproses oleh aplikasi rapidminer dengan menggunakan metode *clustering K-Means* dengan menggunakan 300 data transaksi konsumen, terhitung pengumpulan data dalam kurun waktu 2 bulan di cv mitra indexindo pratama. Tujuan dari penelitian ini adalah untuk mengelompokkan data transaksi konsumen menggunakan metode teknik clustering algoritma K-Means dengan membentuk 4 *cluster* berdasarkan *Measure Types Bregman Divergences* dengan parameter *SquaredEuclideanDistance* sampai menghasilkan nilai DBI terbaik. Hasil dari penelitian ini adalah hasil clustering K-means pada centroid akhir mengasilkan *Cluster*_2 (175.000.000), *Cluster*_0 (25.000.000), *Cluster*_1 (7.500.000) dan nilai terendah (0) pada cluster_3. Hasil pengujian kinerja dengan mengevaluasi Davies-Bouldin-Index (DBI) untuk efisiensi cluster, Berdasarkan hasil Davies-Bouldin Index (DBI) diperoleh selama pengujian data transaksi konsumen dengan program RapidMiner yaitu (K4) *Davies-Bouldin-Index*. (DBI) nilainya adalah 0,261.

***Kata kunci**: Data mining, Konsumen, Algoritm k-means*

1. PENDAHULUAN

Berbagai metode digunakan untuk pengelompokan data, termasuk metode berbasis partisi seperti metode hirarki dan metode *k-means*, menentukan *cluster* yang diketahui, dan menugaskan setiap objek data ke centroid terdekat berdasarkan metrik jarak. , Di antara algoritma Pengelompokan, K-Means adalah algoritm partisi yg paling terkenal dan banyak digunakan karena efisiensi dan kemudahan implementasinya, Pengelompokan data adalah teknik analisis data yang komprehensif dan mengklasifikasikan pola yang mendasari ke dalam kelompok atau cluster, Setiap *cluster* terdiri dari data yang mirip yang membuatnya sangat berbeda dari yang lain, Proses *clustering* menyederhanakan pemodelan layanan data dan banyak digunakan dalam berbagai aplikasi untuk menentukan jumlah *cluster* yang diketahui dan menetapkan setiap objek data ke centroid terdekat berdasarkan metrik jarak, Berdasarkan metrik ini, pengelompokan data dapat menyebabkan pola pengelompokan yang berbeda, yang sangat bergantung pada model jarak yang akan didefinisikan. digabungkan menjadi kelompok yang lebih umum atau secara berturut-turut dibagi menjadi kelompok yang lebih kecil[1]. K-Means adalah metode pengelompokan yang berhubungan dengan metode k-Medoids , yaitu membagi kumpulan data menjadi beberapa cluster yang dibentuk dari awal, metode k-means sangat bagus dan mudah diimplementasikan atau diterapkan**,** cukup cepat, mudah digunakan dan cocok untuk digunakan di berbagai aplikasi kecil dan menengah. Secara ilmiah, *k-means* merupakan salah satu teknik yang sangat penting dalam penerapan data mining.Teknik *k-means clustering* termasuk dalam teknik pengelompokan non-hierarkis yang bertujuan untuk mengelompokan objek berdasarkan identifikasi informasi yang akan dicluster[2]. Pengelompokan data adalah metode Data Mining yang memiliki karakter non-direktif (unsupervised), 2 jenis data yg sering dipakai pada pengelompokan data yang pertama adalah tipe klaster hierarkis dan yang kedua adalah data non-hierarkis. K-means Clustering merupakan teknik yang menggunakan data non-hierarki clustering dengan cara membagi data yang ada menjadi satu atau lebih cluster/kelompok. K-Method membagi data menjadi cluster/kelompok dengan karakteristik yang sama kemudian mengelompokkannya menjadi kelompok data, setelah itu data dengan karakteristik yang berbeda digabungkan menjadi kelompok data lainnya, Tujuan dari clustering data adalah untuk meminimalkan data fungsi objektif dari proses clustering dan memaksimalkan variasi antar cluster [3]. Clustering merupakan tahapan pengelompokan data menjadi beberapa kelompok (cluster), contohnya big data, digunakan sebagai data dimana informasi dari

satu kelompok memiliki karakteristik yang sama dan karakteristik yang berbeda dengan kelompok data lainnya.keuntungan dari *clustering* sangat besar, mudah diimplementasikan dan paling sering digunakan dalam pengelompokan[4]. Penelitian ini penting untuk di teliti dimaksudkan untuk mengetahui bahwa pengelompokan *K-means* telah terbukti secara efektif membedakan kelompok dalam klinis dataset, *K-means clustering* menghitung jarak antara sampel dan membentuk *cluster* dengan merepresentasikan dari sebuah hasil[5]. *K-means clustering* digunakan untuk mengklasifikasikan objek berdasarkan sekumpulan fitur ke dalam sejumlah k kelas, Klasifikasi objek dilakukan dengan cara meminimumkan jumlah kuadrat dari jarak antara objek dan *cluster* yang sesuai[6]. Mengenai algoritma k-means clustering, metode k-means memiliki Salah satu uji validitas internal adalah Davies-Bouldin Index (DBI), DBI didasarkan pada nilai-nilai kohesi dan separasi, di mana kohesi adalah jumlah kedekatan data dengan pusat massa klaster yang dipantau, Sedangkan separasi merupakan jarak antara pusat massa klaster[7].

2. TINJAUAN PUSTAKA

2.1. Penelitian Terdahulu

Penelitian menurut Yahya Novi Andi Cuhwanto, Dewi Agushinta R yang berjudul "Implementasi Data Mining Pemilihan Pelanggan Potensial Menggunakan Algoritma K-Means" Data miningadalah tahapan penggalian informasi penting dari data. Revolusi data mining dan konsep e-commerce (perdagangan elektronik) berkembang pesat, meningkatkan minat banyak pelaku bisnis dengan tujuan meningkatkan bisnis online. Informasi ini dapat digunakan untuk mengumpulkan Informasi berguna untuk membuat keputusan atau tindakan. diasumsikan Banyaknya pelanggan yang berbisnis menunjukkan loyalitas pelanggan terhadap toko online, bahkan perusahaan dapat memberikan hadiah kepada calon pelanggan. Perusahaan memahami pentingnya hubungan antara pelanggan setia dan keberhasilan bisnis perusahaan. Pengumpulan data, Pengolahan data awal, pengolahan data dan hasil penelitian. Teknik pengumpulan data sekunder dilakukan dengan menggunakan data yang tersedia dari toko online pasarayastore.com. Data yang digunakan adalah nama, pesanan saat ini dan uang yang dikeluarkan (Rp) dari periode (2018 -2021), Analisis data mining dilakukan dalam penelitian ini menggunakan Teknik Clustering Algoritm K-Means, mencari informasi dan mengklasifikasikan pelanggan yang melakukan bisnis. Sebagian besar data transaksi. aplikasi yang digunakan sebagai tester adalah Tanagra. Hasil studi ini mengelompokkan data menjadi pelanggan bisnis yang sering, sedang, dan jarang, Berdasarkan hasil perhitungan K-Means cluster, pelanggan dikelompokkan menjadi 3 cluster, yaitu 12 klien dengan rata-rata 11 event (cluster 1), 1403 klien dengan rata-rata 1 event (cluster 2) dan 68 klien dengan rata-rata event. rata-rata 3 transaksi (cluster 3), pelanggan potensial berhasil didapatkan, yaitu yang memiliki rata-rata transaksi dan pembelanjaan terbanyak di klaster pertama (C1) [8].

Penelitian menurut Goldie Gunadi yang berjudul "Penerapan Algoritma K-Means Clustering Pada Analisis Transaks Penjualan Unit Gramedia Printing Jasa Cetak Print On Demand (POD)" Tujuan penelitian ini adalah melakukan analisis K-Means Clustering terhadap data transaksi penjualan jasa percetakan. menggunakan Aplikasi RapidMiner untuk mengelompokan data konsumen berdasarkan transaksi jenis produk cetak. Menggunakan metode kluster K-Means Hasil dari aplikasi menghasilkan 8 kelompok data pelanggan, kelompok terbesar yang mencapai 92,89% dari jumlah pelanggan. Berdasarkan hasil analisis, manajemen perusahaan dapat menentukan berbagai strategi bisnis untuk meningkatkan daya saing perusahaan [9].

2.2. *K-Means*

K-Means adalah algoritma data mining yang dapat digunakan untuk mengelompokkan/cluster data. Proses segmentasi pelanggan dimulai dengan menyiapkan data pelanggan, memilih data pelanggan yang akan dikelompokkan, menentukan jumlah cluster, dan menentukan titik tengah.) secara kebetulan. , menghitung jarak dari centroid, mengelompokkan data pelanggan berdasarkan jarak minimum, dan mentransfer data pelanggan. Jika data pelanggan dipindahkan, jarak dari pusat gravitasi dihitung lagi. Jika tidak ada data pelanggan yang ditransfer, proses selesai. [10]

3. METODE PENELITIAN

Metode penelitian yang digunakan dalam penelitian ini adalah metode eksperimen yang bertujuan untuk mempelajari pengaruh suatu variabel terhadap suatu variabel lainnya, atau menguji bagaimana pendekatan metodologi dapat mengetahui suatu kegiatan pembelajaran membuahkan hasil atau tidak dengan menggunakan algoritma untuk mengklasifikasikan data transaksi konsumen dengan menggunakan algoritma Clustering K-Means dengan tujuan untuk membantu pelaku usaha agar bisa memudahkan pengelompokan data transaksi konsumen di cv .mitra indexindo pratama.

3.1. Teknik Pengumpulan Data

Teknik pengumpulan data untuk penelitian ini berasal dari hasil kesepakatan bersama pada saat melakukan penelitian di cv. mitra indexindo Pratama dengan cara mengamati dan mendokumentasikannya di tempat penelitian.

3.2. Observasi

Observasi dilakukan langsung di lokasi penelitian untuk mendapatkan informasi yang dibutuhkan untuk penelitian dengan mengamati

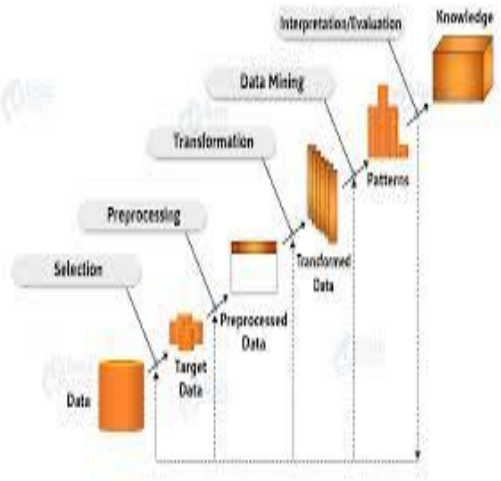
informasi yang sudah ada sebelumnya yang dikelola oleh perusahaan.

3.3. Dokumen

Data yang diperoleh yaitu data transaksi konsumen pada rentang bulan September 2022 yang dikelola oleh pihak perusahaan dalam file dokumen berbentuk table excel.

3.4. Metode Analisa Data

Metode analisis data yang dipakai dalam penelitian ini yaitu pendekatan data science yang menggunakan algoritma clustering K-Means pada fase knowledge discovery database untuk menganalisis data kejadian konsumsi. Penambangan data, atau yang sering disebut sebagai penemuan pengetahuan dalam basis data, adalah proses pencarian pengetahuan yang diterapkan pada sejumlah besar basis data. Tahapan Knowledge Discovery pada proses Database adalah sebagai berikut;



Gambar 1. Tahapan Proses KDD

Knowledge Discovery in Database (KDD) adalah tahapan yang menghasilkan informasi yang bermanfaat dalam database. Semua Langkah-langkah KDD biasanya terdiri dari beberapa langkah, yang pertama mencakup ruang lingkup aplikasi, yang kedua menentukan target data dari data mentah yang disimpan dalam database. , dan yang ketiga membersihkan dan memproses data [11].

Berikut langkah-langkah proses penerapan metode KDD yang dilakukan yaitu:

a. Data *Selection*
   Tahapan pertama data selection yang dilakukan untuk memilih informasi yang digunakan untuk kepentingan penelitian. Proses seleksi ini dilakukan pada dataset transaksi konsumen pada rentang bulan September tahun 2022 yang terdiri dari 300 record dan 5 atribut. Pemrosesan data transaksi konsumen pada tahap pemilihan dilakukan dengan tools RapidMiner. Langkah pertama adalah mengimpor data transaksi konsumen ke dalam arsip, setelah itu aplikasi Rapid Miner akan membaca data tersebut menggunakan operator *Read Excel dan set role*.
b. Data *Preprocessing*
   Langkah selanjutnya data preprocessing, pada proses ini menggunakan operator *Select Atribute* untuk mengetahui data yang hilang. Apabila terdapat *missing value* maka akan dilakukan penghapusan data.
c. Data *Transformation*
   Pada tahapan data *transformation* digunakan untuk mengubah tipe data atribut yang semula *nominal* diubah menjadi tipe data *numerical* dengan tools *Nominal To Numerical* agar sesuai dengan tipe data yang dibutuhkan oleh algoritma *K-Means*.
d. Data Mining
   Data mining adalah proses menemukan pola atau informasi dalam kumpulan data terpilih selama proses transformasi data. Pengujian dengan dataset yang diberikan membutuhkan metode yang digunakan yaitu software K-Means Rapid Miner yang digunakan sesuai dengan tujuan penelitian, Langkah Langkah yang akan dilakukan adalah menghitung nilai (K= 2 3 4 5) dengan berdasarkan measure types BregmanDivergences dengan parameter SquaredEuclideanDistance.
e. Evaluation
   Tahapan *evaluation* ini menghasilkan pola-pola dari proses evaluasi guna memperkirakan tujuan yang diinginkan. Tahap ini juga menggunakan *tools* RapidMiner untuk menentukan nilai DBI terbaik dari algoritma *K-Means* dengan melakukan pengujian evaluasi menggunakan measure types *BregmanDivergences* dan parameter SquaredEuclideanDistance.

4. HASIL DAN PEMBAHASAN

4.1. Data

Pada penelitian ini dataset yang digunakan mengenai data transaksi konsumen yang diperoleh dari hasil kesepakatan bersama pada saat melakukan penelitian di cv. mitra indexindo pratama, dari proses pengumpulan dataset transaksi konsumen diperoleh sebanyak 300 *record*, yang berjumlah 5 atribut, diantaranya No, Tanggal, Alamat, Nama Pelanggan dan Jumlah. File yang di peroleh berupa dokumen *Microsoft Excel* dengan format xlsx dalam bentuk tabel, berikut adalah dataset transaksi konsumen yang diperoleh dari hasil pengumpulan data.

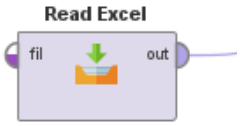
Tabel 1. Dataset Transaksi Konsumen Bulan September 2022

No	Tanggal	Alamat	Nama Pelanggan	Jumlah
1	Kamis, 01 september 2022	Cirebon	Regidon Computer	2500000
2	Kamis, 01 september 2022	Cirebon	Isp komputer	10400000
3	Kamis, 01 september 2022	Cirebon	Bpk dimas	25000000

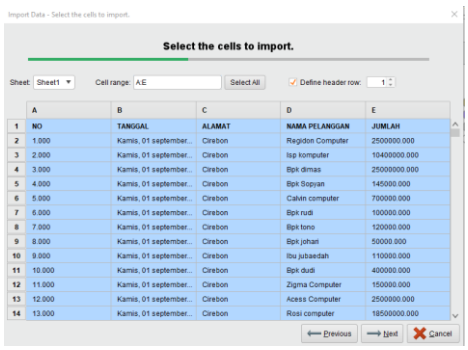
No	Tanggal	Alamat	Nama Pelanggan	Jumlah
4	Kamis, 01 september 2022	Cirebon	Bpk Sopyan	145000
5	Kamis, 01 september 2022	Cirebon	Calvin computer	700000
6	Kamis, 01 september 2022	Cirebon	Bpk rudi	100000
7	Kamis, 01 september 2022	Cirebon	Bpk tono	120000
8	Kamis, 01 september 2022	Cirebon	Bpk johari	50000
9	Kamis, 01 september 2022	Cirebon	Ibu jubaedah	110000
10	Kamis, 01 september 2022	Cirebon	Bpk dudi	400000

4.2. Data *Selection*

Langkah pertama dalam proses Knowledge Discovery in Database (KDD) adalah pemilihan data. Pada fase ini, tujuannya adalah untuk memilih data yang akan digunakan dalam proses analisis, langkah pertama yang dilakukan dengan cara *import* data melalui *tools* RapidMiner ke dalam *repository*, kemudian gunakan operator *read excel* untuk mengakses data yang telah disimpan pada *repository* dan memuatnya ke tab proses untuk melihat hasilnya. Berikut operator dan hasil dari *Read Excel* ditunjukan pada gambar 2 dan 3

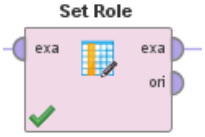


Gambar 2. Operator *Read Excel*



Gambar 3. *Hasil Read Excel*

Tahap selanjutnya masukan Operator *Set Role*, Operator set role ini merupakan langkah mengubah peran atribut seperti target id dll. Tampilan dari set role dapat di lihat pada gambar 4

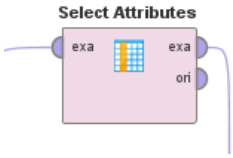


Gambar 4. Operator *Set Role*

4.3. Data *Preprocessing*

Pada langkah *selection* data yang tidak diperlukan dan perubahan peran atribut digunakan satu operator yaitu operator *Select Attributes*. Langkah-langkah untuk menggunakan operator adalah sebagai berikut:

Berfungsi untuk memilih data subset atribut yang ingin digunakan dari ExampleSet dan menghapus atribut lain yang tidak diperlukan. gambar Select Attributes operator ditunjukkan pada Gambar 5



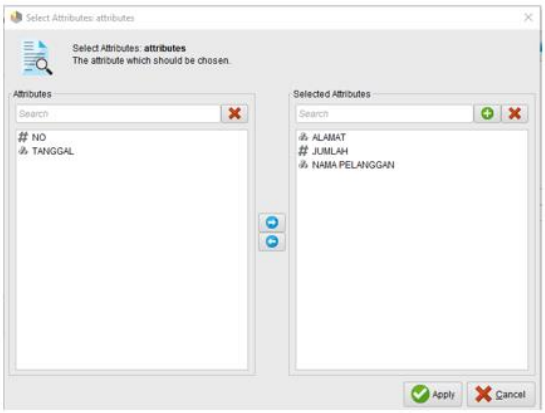
Gambar 5. Operator *Select Atributes*

Pada operator Select Attributes, proses selection dilakukan terhadap atribut yang akan dan tidak dipergunakan dalam riset yang ditunjukan pada table 2

Tabel 2. Tabel Atribut Data

No	Atribut yang tidak digunakan	Atribut yang digunakan
1	No	Alamat
2	Tanggal	Jumlah
	-	Nama Pelanggan

Atribut yang tidak dipakai pada proses riset ini adalah No, Tanggal. Petunjuk yang tertera tidak mempengaruhi proses Data Analisys karena atribut tersebut tidak terdapat pada data yang ingin di cluster dan digunakan sebagai nilai untuk mendapatkan nilai DBI terbaik. Setelah proses pemilihan atribut selesai, dari 5 atribut tersebut kini menjadi 3 atribut yang digunakan yaitu Alamat, Nama Pelanggan dan Jumlah, yang dapat digunakan dalam perhitungan data mining untuk mencari nilai DBI terbaik. Hasil yang diperoleh dari proses pemilihan operator ditunjukan pada gambar 6



Gambar 6 Proses Select Atributes

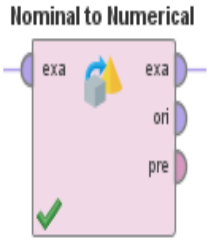
4.4. *Data Transformation*

Langkah selanjutnya data transformation, pada langkah ini dilakukan perubahan tipe atribut dataset transaksi konsumen yaitu tipe data nominal diubah menjadi tipe data *numeric*. Berikut tabel atribut data yang ditampilkan pada tabel 3

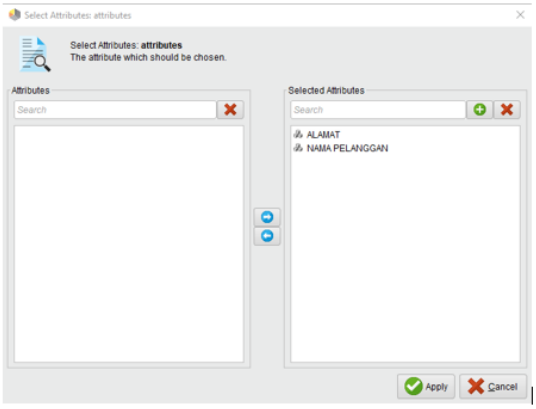
Tabel 3. Tipe Data Tabel

No	Tipe data	Tipe Data
1	Alamat	Numerik
2	Nama Pelanggan	Kategori
3	Jumlah	Kategori

Atribut data yang diubah yaitu atribut Alamat, Nama Pelanggan, Jumlah menjadi numerik menggunakan operator *Nominal to* Numerical agar sesuai dengan tipe data yang dibutuhkan algoritma *K-Means* pada proses Analisa pengelompokan data. Tampilan operator *Nominal to Numerical* dan proses dari operator *Nominal to Numerical* ditampilkan pada gambar 7 dan 8

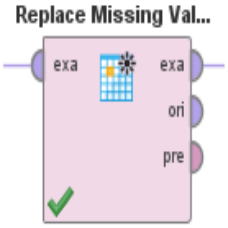


Gambar 7. Operator Nominal To Numerical



Gambar 8. Proses Pemilihan Atributes

berdasarkan tahapan ini dilakukan pemilihan atribute alamat dan nama pelanggan dengan menggunakan operator *replace missing values* untuk mengetahui apakah ada data yang hilang pada semua atribut, apabila terdapat atribut yang memiliki nilai kosong atau hilang maka akan dilakukan penghapusan data. Operator dari *replace missing value ditunjukan pada gambar 9*

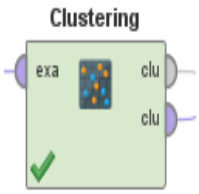


Gambar 9. Operator Replace Missing Value

4.5. *Data Mining*

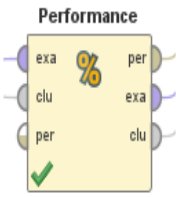
Data Mining merupakan mekanisme mencari model dan penjelasan oleh dataset yang terseleksi selama proses transformasi, Pemrosesan Data dengan Rapid Miner memiliki berbagai langkah yaitu sebagai berikut:

Masukan operator K-Means Clustering, operator ini cenderung memproses data yang dimasukkan sebelumnya untuk proses akumulasi berikutnya.



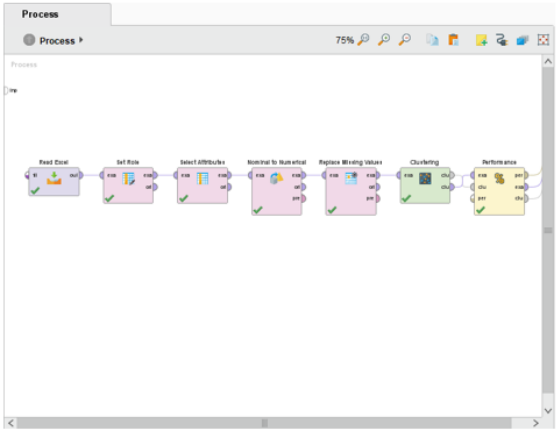
Gambar 10. Operator K-Means Clustering

Selanjutnya Masukan operator Cluster Distance Performance Tool dengan tujuan untuk mengevaluasi kinerja pada proses data mining yang berdasarkan cluster central.



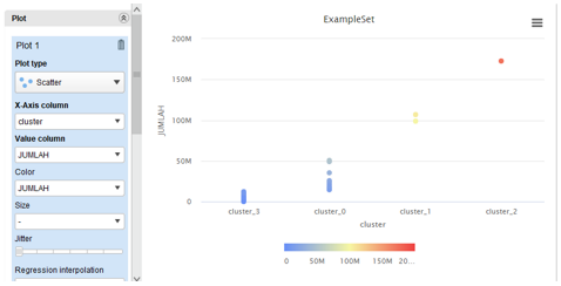
Gambar 11. Operator Cluster Distance Performance

Cluster Distance Performance digunakan untuk menentukan nilai (K = 2 3 4 5) dengan berdasarkan measure type Bregman Divergences dengan parameter Squared*Euclidean Distance*.



Gambar 12. Proses Clustering K-Means

Langkah-langkah algoritma K-Means, pertama pilih banyaknya sebuah cluster yang sudah ditentukan, banyak nya cluster yang digunakan adalah (K=2 3 4 5), proses ini mengelompokkan data berdasarkan kesamaan karakteristik dari masing-masing data. Proses clustering ditunjukkan pada Gambar 12



Gambar 13. Hasil Visualisasi Clustering

Visualisasi clustering algoritma K-means dari ExampleSet (K4) ditunjukkan pada gambar di atas, dimana jumlah *Cluster* yang sudah terbentuk adalah 4 Cluster, hubungan antar Cluster dengan jumlah ditunjukkan pada gambar 13

Cluster Model

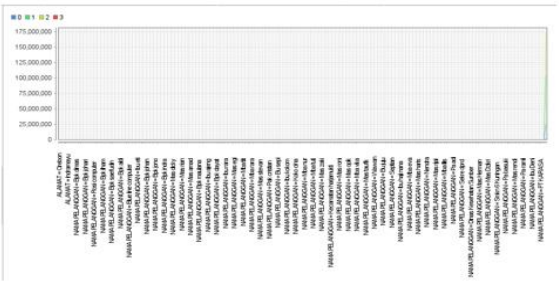
```
Cluster 0: 21 items
Cluster 1: 2 items
Cluster 2: 1 items
Cluster 3: 276 items
Total number of items: 300
```

Gambar 14. Hasil Cluster Model

Hasil cluster set (K4) menunjukkan bahwa Cluster 0 memiliki 21 data, Cluster 1 memiliki 2 data, Cluster 2 memiliki 1 data dan cluster 3 memiliki 276 data. Total data keseluruhan adalah 300 data.

Tabel 4. Hasil Centroid Akhir

Attribute	Cluster_ 0	Cluster_ 1	Cluster_ 2	Cluster_ 3
Alamat	1	1	1	0.999
Nama Pelanggan	0.816	1	0	0.928
Jumlah	249033 8095	102480 500	173000 000	180866 6667



Gambar 15. Hasil Grafik Plot K-Means Clustering

Hasil *K-Means Clustering* pada centroid akhir, hijau cluster 2, biru muda cluster 1, biru cluster 0, dan merah cluster 3, nilai teratas yaitu 175.000.000 pada cluster 2 hijau, kedua 7.500.000 pada cluster 1 biru muda, ketiga 25.000.000 pada cluster 0 biru dan yang terendah yaitu 0 pada cluster_3.

4.6. *Evaluation*

Pada titik ini, tes RapidMiner dijalankan untuk menguji kemampuan Algoritma *K-Means* dengan *Davis-Bouldin Index (DBI).*

PerformanceVector

```
PerformanceVector:
Avg. within centroid distance: -13799922569287.137
Avg. within centroid distance_cluster_0: -108457974331067.100
Avg. within centroid distance_cluster_1: -15598550250001.000
Avg. within centroid distance_cluster_2: -0.000
Avg. within centroid distance_cluster_3: -6634645685991.776
Davies Bouldin: -0.261
```

Gambar 16. Hasil Uji Performance

Hasil pengujian Performance menggunakan mengevaluasi *Davis-Bouldin Index (DBI)* pada proses clustering, dari hasil implementasi menggunakan program RapidMiner saat menguji data transaksi konsumen berdasarkan hasil Davis-Bouldin Index (DBI)) yang didapat merupakan (K4) Davies Bouldin Index. (DBI) nilainya 0,261.

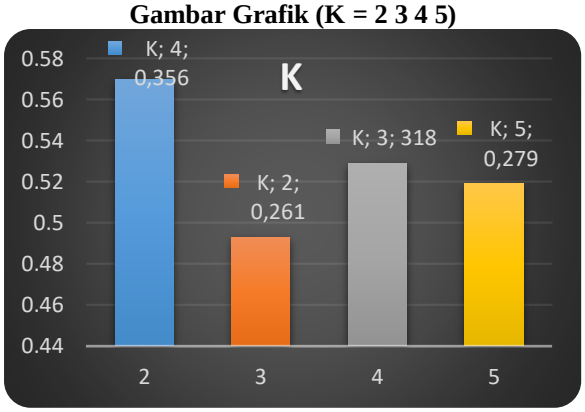
Tabel 5. Hasil Davies Bouldin Index (DBI)

Measure Type	Cluster Set	DBI
Bregman Divergences	2	0,261
	3	0,318
	4	0,356
	5	0,279

Dalam proses clustering, diperoleh beberapa nilai uji performance berdasarkan Davies-Bouldin (DBI) untuk setiap nilai K, berdasarkan skor *Davis-Bouldin*

index (DBI) yang didapat ialah (K4) dengan (DBI) skor. 0,356, (K3) DBI Score 0,318, (K5) dengan DBI Score 0,279 dan (K2) dengan DBI Score 0,261, maka dapat disimpulkan dari hasil analisis cluster bahwa (K2) merupakan skor terbaik karena nilai Indeks Davies Bouldin (DBI) mendekati nol dan karenanya merupakan nilai terbaik untuk Indeks Davies Bouldin (DBI).

Gambar Grafik (K = 2 3 4 5)



Gambar 4. 17 Grafik (K = 2 3 4 5) nilai Davies-Buldin Index (DBI) setiap cluster.

4.7. Pembahasan

Berdasarkan hasil penelitian dengan algoritma K-Means menggunakan measure type Bregman Divergences dengan parameter Squared*Euclidean Distance*

dari algoritma K-Means memberikan hasil uji Davies-Bouldin Index (DBI). *Data Mining* adalah proses menghasilkan pola yang menarik dan tersembunyi di kumpulan data besar yang tersimpan didalam basis data, seperti data were house dan fasilitas penyimpanan data yang lain [12].

5. KESIMPULAN DAN SARAN

Menurut proses hasil *data mining yang sudah* diterapkan pada data transaksi konsumen dengan memakai algoritma K-Means bisa disimpulkan bahwa: Algoritma K-Means bisa diimplementasikan ditahap proses cluster dengan memakai software Rapid Miner dan cluster test dengan menentukan nilai (K = 2 3 4 5) pada proses clustering dengan hasil uji kinerja berdasarkan measure type Bregman Divergences dengan parameter Squared*Euclidean Distance*. Algoritma *K-means*, menghasilkan nilai *Davis-Bouldin Index* (DBI) yang didapat yaitu (K2) menghasilkan 4 cluster dengan nilai *Davies Bouldin Index* (DBI) 0,261.

DAFTAR PUSTAKA

[1] Andi Cuhwanto, Y. N., & R, D. A. (2021). Implementasi Data Mining Pemilihan Pelanggan Potensial Menggunakan Algoritma K-Means. *PETIR*, *15*(1), 48–56. https://doi.org/10.33322/petir.v15i1.1358.

[2] Clayman, C. L., Srinivasan, S. M., & Sangwan, R. S. (2020). K-means clustering and principal components analysis of microarray data of L1000 landmark genes. *Procedia Computer Science*, *168*, 97–104. https://doi.org/10.1016/j.procs.2020.02.265

[3] Darlinda, D., & Utamajaya, J. N. (2022). Sistem Pendukung Keputusan Penerima Beasiswa Program Indonesia Pintar Menggunakan Metode Algoritma K-Means Clustering. *JURIKOM (Jurnal Riset Komputer)*, *9*(2), 167. https://doi.org/10.30865/jurikom.v9i2.3971

[4] Farouq, K., & Susilo, H. (n.d.). *KLASTERISASI VIRUS COVID-19 DI WILAYAH KABUPATEN LAMONGAN DENGAN METODE K-MEANS CLUSTERING*.

[5] Gunadi, G. (2022). PENERAPAN ALGORITMA K-MEANS CLUSTERING UNTUK MENGANALISA TRANSAKSI PENJUALAN JASA CETAK PADA UNIT PRINT ON DEMAND (POD) PERCETAKAN GRAMEDIA. *Infotech: Journal of Technology Information*, *8*(2), 117–126. https://doi.org/10.37365/jti.v8i2.148

[6] Javidan, S. M., Banakar, A., Vakilian, K. A., & Ampatzidis, Y. (2023). Diagnosis of grape leaf diseases using automatic K-means clustering and machine learning. *Smart Agricultural Technology*, *3*, 100081. https://doi.org/10.1016/j.atech.2022.100081

[7] Kurniadi, D., & Sugiyono, A. (2020). Pengelompokkan Data Akademik Menggunakan Algoritma K-Means Pada Data Akademik Unissula. *TRANSFORMTIKA*, *18*(1), 93–101.

[8] Shrifan, N. H. M. M., Akbar, M. F., & Isa, N. A. M. (2022). An adaptive outlier removal aided k-means clustering algorithm. *Journal of King Saud University* - Computer *and Information Sciences*, *34*(8), 6365–6376. https://doi.org/10.1016/j.jksuci.2021.07.003

[9] Suriani, L. (2020). Pengelompokan Data Kriminal Pada Poldasu Menentukan Pola Daerah Rawan Tindak Kriminal Menggunakan Data Mining Algoritma K-Means Clustering. *Jurnal Sistem Komputer Dan Informatika (JSON)*, *1*(2), 151. https://doi.org/10.30865/json.v1i2.1955

[10] Tanjung, F. A., Windarto, A. P., Fauzan, M., Studi, M. P., Informasi, S., & Tunas Bangsa, S. (n.d.). *Penerapan Metode K-Means Pada Pengelompokkan Pengangguran Di Indonesia*. *6*, 61–74. https://tunasbangsa.ac.id/ejurnal/index.php/jurasik

[11] Triyandana, G., Putri, L. A., & Umaidah, Y. (2022). Penerapan Data Mining Pengelompokan Menu Makanan dan Minuman Berdasarkan Tingkat Penjualan Menggunakan Metode K-Means. In *Journal of Applied Informatics and Computing (JAIC)* (Vol. 6, Issue 1). http://jurnal.polibatam.ac.id/index.php/JAI