PENERAPAN DATA MINING MENGGUNAKAN METODE K-MEANS CLUSTERING PADA PENJUALAN TAS DI ASIA TOSERBA CIREBON

Firdaus Sandova [1], Rudi Kurniawan [2], Tati Supratati [3]
[1,3] Teknik Informatika, [2] Rekayasa Perangkat Lunak, STMIK IKMI Cirebon
Jl. Perjuangan No. 10 Majasem Kec. Kesambi Kota Cirebon
Firdaussandova94@gmail.com

ABSTRAK

Asia Toserba adalah sebuah perusahaan besar yang berada di Kota Cirebon, dengan menjual berbagai banyak keperluan sehari-hari, diantaranya penjualan tas dengan berbagai macam merek yang berkualitas. Dikarenakan banyaknya produk tas yang dijual dari masing-masing merek sehingga memiliki daya saing yang kuat untuk menarik minat customer dengan berbagai macam strategi pemasaran. Penelitian ini bertujuan untuk menganalisis pola penjualan tas di Asia Toserba Cirebon dalam menganalisis barang yang banyak diminati dan kurang diminati. Pada penelitian ini data yang digunakan yaitu data penjualan tas merk Westpak Bag's, Subway, dan Cannon yang diambil pertiap bulan Januari – Juni 2023, dengan jumlah data mencapai 1.113 data, yang nantinya akan diolah menggunakan Software Rapid Miner. Tujuan dari penelitian ini adalah untuk menentukan berapa nilai Avarage Within Sentroid Distance dari masing-masing cluster dan nilai Davies Bouldin Index (DBI). Setelah itu Algoritma K-Means diterapkan untuk mengetahui jumlah cluster yang tepat atau sudah sangat baik menggunkan metode elbow method. Berdasarkan hasil penelitian yang telah dijelaskan, di dapatkan hasil nilai Avarage Within Centroid Distance dari 3 clustering yaitu k2 = 2007842385.954, k3 = 969761146.003, k4 = 554546997.018. Lalu jumlah cluster yang tepat atau sudah sangat baik dibuktikan dengan elbow method dengan hasil k3, dan nilai Davies Bouldin Index (DBI) yang mendekati nol pada saat k4 dengan nilai DBI k4 = 0.435. Dengan menggunakan Algoritma K-Means diharapkan dapat mengklasterisasi data penjualan tas di Asia Toserba Cirebon untuk mengatasi beberapa permasalahan dalam penjualan dan dapat mengoptimalkan penawaran produk dan pemasaran yang ditargetkan untuk memenuhi kebutuhan pelanggan secara lebih efektif.

Kata kunci : *K-Means Clustering, Data Mining, Penjualan Tas, Asia Toserba Cirebon, Analisis Penjualan, Rapid Miner*

1. PENDAHULUAN

Asia Toserba adalah sebuah perusahaan besar yang berada di Cirebon, dengan menjual berbagai banyak keperluan sehari-hari, diantaranya penjualan tas dengan berbagai merek yang berkualitas. Tas merek *Westpak Bag's, Subway, dan Cannon* tergolong tas yang kuat dan awet, karena sudah menggunakan bahan *polyester*, yang memberikan dampak positif bagi pelanggan di dalam ketahanan dan kualitas.

Seiring dengan persaingan yang semakin ketat, penjualan tas di Asia Toserba Cirebon mengalami banyak permintaan dari customer, yang menyebabkan penjualan tas di Asia Toserba Cirebon mengalami kesulitan dalam menganalisis barang mana yang banyak diminati dan kurang diminati, sehingga dibutuhkan Clustering data untuk mengelompokkan produk penjualan tas di Asia Toserba Cirebon.

K-means merupakan *algoritma unsupervised learning* yang berfungsi untuk mengelompokkan data ke dalam *cluster*. *Algoritma* ini dapat digunakan tanpa adanya label kategori pada data yang dimasukkan. Selain itu, *K-Means Clustering* termasuk dalam metode *non-hierarchy*. Metode ini bertujuan untuk mengelompokkan data ke dalam kelompok, di mana setiap kelompok menjelaskan karakteristik yang serupa, dan karakteristik tersebut berbeda dengan kelompok lainnya. Pengambilan sampel dalam *Cluster Sampling* dilakukan dengan cara memilih unit-unit populasi secara acak dari kelompok yang sudah ada, yang disebut sebagai *'cluster'*. Klasterisasi atau *clustering*, pada dasarnya, merupakan salah satu permasalahan yang memanfaatkan teknik *unsupervised learning*. Penerapan metode *K-Means Clustering* dapat diterapkan pada penjualan tas di Asia Toserba Cirebon. untuk mengetahui data penjualan tas mana yang paling banyak diminati dan kurang diminati. Potensi manfaat bagi penjualan tas di Asia Toserba Cirebon dengan menggunakan data mining dapat membantu mengelompokkan data besar dengan cepat serta memberikan kepuasan dalam mengambil keputusan bagi customer. Dalam produk tas westpak bag's, subway, dan cannon terdapat tantangan dalam menganalisis barang yang banyak diminati dan kurang diminati. Metode kuantitatif dapat membantu penjualan tas di Asia toserba Cirebon berupa pemecahan suatu masalah dalam bentuk penelitian. Dalam penelitian ini, data dikumpulkan menggunakan teknik wawancara. Untuk memperoleh data dan kebenaran informasi pada penjualan tas, dengan melakukan tanya jawab langsung dengan pegawai yang bekerja di Asia Toserba Cirebon.

Hasil dari penelitian ini bisa memberikan dampak positif bagi penjualan tas di Asia Toserba Cirebon, dengan menerapkan metode K-Means Clustering untuk mengetahui data penjualan tas mana yang paling banyak diminati dan kurang diminati, serta memberikan kepuasan dalam mengambil keputusan bagi customer. Metode K-Means Clustering dapat

berkontribusi secara penting di bidang informatika, karena informatika seringkali terlibat dalam mengelola, menganalisis, dan memahami data yang besar dan rumit. K-means Clustering merupakan sebuah alat yang efektif dalam mengelompokkan data ke dalam kelompok-kelompok yang memiliki kemiripan. Dengan metode K-Means Clustering dapat memberikan pemahaman mengenali pola, struktur, atau kategori yang mungkin tersembunyi dalam data dan sulit terlihat tanpa analisis clustering.

Nama pertama nama mahasiswa, nama kedua dan ketiga adalah dosen pembimbing pertama dan kedua tanpa gelar, alamat email cukup satu saja gunakan email mahasiswa atau Corresponding Author.

2. TINJAUAN PUSTAKA

Penelitian dengan judul Penerapan metode K-Means Clustering dalam analisis penjualan toko Yana Sport menunjukkan bahwa meskipun proses penjualan dan pemasaran sudah menggunakan aplikasi E-commerce, namun strategi tambahan diperlukan untuk meningkatkan penjualan produk. Meskipun teknologi informasi berupa E-commerce memberikan potensi peningkatan penjualan, persaingan dalam dunia bisnis membuat pemilik toko harus mencari strategi yang efektif. Dalam penelitian ini, toko Yana Sport dibagi menjadi dua kelompok, yaitu laris terjual dan tidak laris terjual. Hasil analisis data menunjukkan bahwa Cluster 0 dengan nilai 110 dikategorikan sebagai produk yang tidak laris terjual, sedangkan Cluster 1 dengan nilai 21389 dikategorikan sebagai produk yang laris terjual. Data Performance menunjukkan bahwa strategi yang lebih fokus mungkin diperlukan untuk meningkatkan penjualan produk yang termasuk dalam Cluster 0 [1].

Pada penelitian Penerapan metode K-Means Clustering untuk menganalisis penjualan di toko Myam Hijab Penajam dilakukan dengan tujuan mengkategorikan produk hijab yang dijual di toko tersebut. Meskipun toko Myam Hijab khusus menjual hijab, tidak semua produk hijab tersebut memiliki tingkat penjualan yang sama. Data mengenai penjualan, pembelian barang, dan pengeluaran tak terduga di toko Myam Hijab terbilang kurang terstruktur. Metode K-Means diterapkan dengan mengelompokkan data stok hijab, dan dilakukan pemilihan 3 grup acak sebagai centroid awal. Setelah dilakukan proses tersebut dan data pada setiap kelompok stabil, hasil akhirnya menunjukkan bahwa terdapat 24 produk yang laku keras, 59 produk yang laris, dan 17 produk yang memiliki tingkat penjualan kurang optimal [2].

Pada Penelitian Analisis tingkat penjualan menu (makanan dan minuman) menggunakan Algoritma K-Means dilakukan dalam penelitian ini. Riset ini didasarkan pada pemahaman bahwa perencanaan menu merupakan aspek krusial dari strategi penjualan restoran, dan setiap menu memiliki tingkat penjualan yang berbeda. Data penelitian diperoleh dari catatan transaksi penjualan menu selama satu tahun. Hasil analisis menggunakan Algoritma K-Means membantu mengidentifikasi menu makanan dan minuman yang populer serta mengelompokkan menu berdasarkan tingkat penjualannya [3].

Pada Penelitian Penelitian tentang Analisis Cluster Data Transaksi Penjualan Minimarket Selama Pandemi Covid-19 menggunakan Algoritma K-Means menyatakan bahwa dampak buruk dari Covid-19 terlihat pada sektor ekonomi di Indonesia, terutama dari kerugian yang dialami oleh Minimarket Berkah Abadi berupa penurunan omset penghasilan. Untuk mengatasi hal ini, perlu dilakukan strategi penjualan guna meminimalisir kerugian. Analisis transaksi penjualan dapat menjadi sarana untuk menemukan kelompok produk dengan data penjualan tertinggi, sehingga manajemen stok dapat diatur dengan baik dan transaksi penjualan meningkat. Analisis ini dapat dilakukan dengan menerapkan algoritma K-Means, yang dapat mengelompokkan data berdasarkan karakteristik yang serupa. Pada penerapan algoritma ini pada 278,480 data transaksi, ditemukan tiga cluster data penjualan, yaitu cluster 2 dengan penjualan terbanyak mencakup 57 produk, cluster 1 dengan penjualan sedang mencakup 5 produk, dan sisanya adalah cluster 0 dengan penjualan terendah. Akurasi model klasterisasi yang dihasilkan, seperti yang terukur melalui confusion matrix, mencapai 87% [4].

Pada Penelitian Penerapan metode Data Mining dengan Algoritma K-Means Clustering pada populasi ayam petelur di Indonesia menjadi fokus penelitian. Telur ayam merupakan jenis telur yang umumnya ditemui dan diminati oleh banyak orang, memenuhi kebutuhan masyarakat akan sumber protein hewani dan nutrisi sehari-hari. Penelitian ini melakukan analisis menggunakan Algoritma K-Means Clustering pada populasi ayam petelur di Indonesia. Sumber data mengenai populasi ayam ras petelur di Indonesia diperoleh dan dikumpulkan melalui situs web Badan Pusat Statistik Nasional, mencakup data dari tahun 2016 hingga 2020 dan melibatkan 34 provinsi. Data tersebut akan dikelompokkan ke dalam tiga cluster, yakni cluster populasi tinggi, sedang, dan rendah [5].

Pada Penelitian Penerapan teknik Data Mining pada clustering produksi daging sapi di Indonesia menggunakan Algoritma K-Means menjadi fokus penelitian ini. Sapi, sebagai hewan yang umum di Indonesia, memberikan manfaat besar bagi manusia melalui susu yang kaya gizi dan dagingnya yang menjadi sumber protein hewani penting. Meskipun begitu, produksi daging sapi di Indonesia masih belum mencukupi untuk memenuhi kebutuhan domestik. Oleh karena itu, pemerintah menyarankan agar masyarakat mengkonsumsi sumber protein lain, dan sebagian daging sapi masih harus diimpor untuk memenuhi kebutuhan protein harian. Penelitian ini membahas penerapan Data Mining, khususnya Algoritma K-Means, dalam melakukan clustering terhadap produksi daging sapi di Indonesia. Data penelitian diperoleh dari dokumen produksi daging sapi yang dikeluarkan oleh Badan Pusat Statistik

Nasional, mencakup data dari tahun 2009 hingga 2016 dari 34 provinsi. Penelitian ini bertujuan untuk mengelompokkan produksi daging sapi ke dalam tiga kelompok, yaitu tinggi, sedang, dan rendah [6]

Pada Penelitian Penerapan data mining untuk mengelompokkan produksi jagung menurut provinsi menggunakan Algoritma K-Means menjadi fokus penelitian ini. Permintaan terhadap jagung saat ini mengalami perkembangan yang signifikan, terlihat dari pertumbuhan pasar domestik yang pesat. Dalam konteks ini, peneliti bertujuan untuk meningkatkan produktivitas dan kualitas produksi jagung. Data yang akan digunakan dalam penelitian ini berasal dari Badan Pusat Statistik. Metode yang digunakan adalah Algoritma K-Means Clustering, dengan aplikasi Rapid miner sebagai alat bantu. Hasil dari penelitian ini akan mengelompokkan produksi jagung menjadi dua klaster, yaitu klaster tinggi dan klaster rendah. Hasil akhir menunjukkan bahwa terdapat dua provinsi yang termasuk dalam klaster tinggi, sementara 32 provinsi lainnya masuk ke dalam klaster rendah [7].

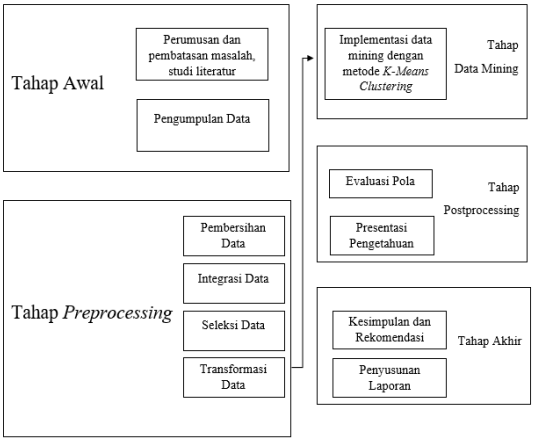
Pada Penelitian Penerapan Data Mining dalam menentukan prioritas pemesanan produk berdasarkan data penjualan barang menggunakan Algoritma Apriori menjadi fokus penelitian ini. Prioritas pemesanan produk sering kali bergantung pada preferensi konsumen, namun kesalahan dalam menentukan pemesanan dapat mengakibatkan kerugian bagi perusahaan. Untuk mengatasi masalah ini, diperlukan pendekatan yang tepat melalui analisis data penjualan. Dalam hal ini, Data Mining, sebagai metode pengolahan data, menggunakan Algoritma Apriori untuk menganalisis pola penjualan barang. Hal ini membantu perusahaan dalam proses penentuan pemesanan produk yang lebih efektif. Hasil penelitian menunjukkan bahwa dengan menggunakan Algoritma Apriori, ditemukan beberapa pola penjualan. Sebagai contoh, jika konsumen membeli Celana Biasa, kemungkinan besar juga akan membeli Celana Pendek dengan nilai support sebesar 10% dan nilai confidence sebesar 20%. Selain itu, kombinasi item lainnya yang ditemukan adalah jika membeli Celana Biasa, kemungkinan besar juga akan membeli Kaos Oblong, dengan nilai support 30% dan nilai confidence 60%. Selain itu, kombinasi item set terakhir yang ditemukan adalah jika konsumen membeli Celana Pendek, kemungkinan besar juga akan membeli Kaos Oblong, dengan nilai support 30% dan nilai confidence 60% [8].

Pada Penelitian Penerapan teknik clustering pada data berbagai merk mesin cuci bertujuan untuk mengelompokkan jenis-jenis mesin cuci. Dalam era modern, mesin cuci telah menjadi kebutuhan pokok bagi ibu rumah tangga, memberikan kemudahan dalam proses pencucian pakaian. Adanya berbagai tipe mesin cuci dengan keunggulan masing-masing menciptakan dilema bagi para ibu rumah tangga dalam memilih yang paling sesuai. Penelitian ini bertujuan untuk memberikan kemudahan bagi konsumen dalam memilih mesin cuci yang sesuai dengan kebutuhan mereka. Melalui penerapan Algoritma K-Means Clustering, data berbagai jenis merk mesin cuci berhasil dikelompokkan menjadi tiga kelompok. Kelompok pertama terdiri dari 5 data, kelompok kedua terdiri dari 8 data, dan kelompok ketiga terdiri dari 12 data. Pentingnya penentuan titik pusat awal (centroid) dalam proses clustering sangat mempengaruhi jumlah iterasi yang perlu dihitung [9].

Pada Penelitian Penerapan Data Mining Clustering menggunakan Algoritma K-Means pada data nasabah kredit bermasalah PT. BPR Milala merupakan upaya dalam menjawab kebutuhan perbankan terkait manajemen risiko kredit. Kehadiran kredit adalah unsur penting dalam aktivitas perbankan dan selalu memerlukan perhatian khusus terkait risiko yang mungkin timbul pada saat memberikan kredit. PT. BPR Milala menghadapi tantangan terkait kredit dan memerlukan suatu sistem yang dapat meningkatkan efisiensi dalam pengelompokkan dan mengurangi waktu penyelesaian data kredit bermasalah. Oleh karena itu, diperlukan suatu sistem berbasis komputer yang dapat mengelompokkan data kredit bermasalah, dan hal ini dapat dicapai melalui penerapan ilmu Data Mining. Sistem ini diharapkan dapat membantu dalam mengatasi masalah yang dihadapi oleh PT. BPR Milala dengan lebih cepat dan efisien dalam memproses data kredit bermasalah [10].

3. METODE PENELITIAN

Penelitian ini menggunakan metode kuantitatif yang berfokus pada analisis data berupa angka untuk menghasilkan informasi tambahan. Dalam penelitian ini, data numerik yang berasal dari penjualan tas diubah dan dianalisis menggunakan metode data mining K-Means Clustering melalui serangkaian langkah dalam proses Knowledge Discovery in Databases (KKD). Adapun tahapan atau langkah-langkah yang akan dilakukan seperti gambar 1 dibawah ini:



Gambar 1. Tahapan Penelitian

Berikut merupakan hal-hal yang perlu dilakukan dalam penelitian berdasarkan tahapan *knowledge discovery in database.*

1) Data Selection

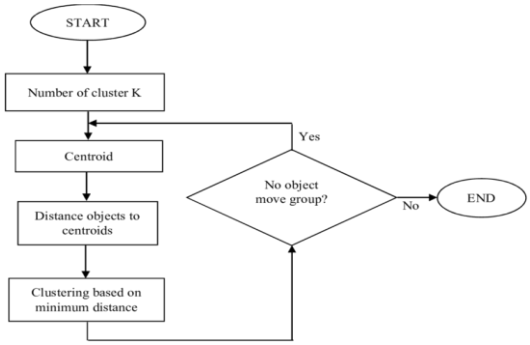
Data Selection Menyaring sekumpulan data mentah untuk menentukan variabel data apa saja yang akan diproses [3].

2) Prepocessing
Prepocessing merupakan tahap membersihkan data dari noise, missing value dan data yang tidak relevan [3].
3) Transformation
Transformation Mentransformasikan/ merubah data ke dalam bentuk yang sesuai dengan format pengolahan data pada algoritma data mining yang akan digunakan. Setiap algoritma data mining mengolah data dengan format tertentu [3].
4) Data Mining
Pada fase ini yang dilakukan adalah menerapkan algoritma atau metode pencarian pengetahuan. Ini adalah langkah penting di mana teknik kecerdasan diterapkan untuk mengekstrak pola informasi yang berpotensi berguna dari data yang dipilih [1].
5) Evaluation
Pada tahap evaluasi, akan diketahui apakah hasil daripada tahap data mining dapat menjawab tujuan yang telah ditetapkan. Untuk itu akan dilakukan profilisasi pada setiap cluster yang telah terbentuk, untuk diketahui karakteristik pada kelompok tersebut. Disamping itu untuk diketahui kesesuaian dengan jalur perminatan akan dilakukan analisis lebih lanjut untuk dihubungkan dengan atribut perminatan, Sehingga diharapkan mendapatkan informasi atau pola yang berguna sebagai acuan pemutakhiran data. Knowledge Tahap terakhir dari proses data mining adalah bagaimana memformulasikan keputusan atau aksi dari hasil analisis yang didapat [1].

3.1. Algoritma K-Means

Langkah-langkah dalam Algoritma K-Means Clustering Adalah:

1. Mengidentifikasi jumlah kluster.
2. Menghitung nilai tengah (centroid) Dalam menghitung nilai tengah (centroid) pada iterasi awal, nilai awalnya dipilih secara acak. Namun, jika perlu menentukan nilai centroid sebagai bagian dari suatu iterasi.
3. Mengukur jarak antara pusat kluster dengan setiap titik objek.
4. Pengelompokan objek dilakukan dengan mempertimbangkan jarak minimum antara objek. Hasil nilai dalam keanggotaan data pada matriks jarak adalah 0 atau 1, di mana nilai 1 menunjukkan bahwa data tersebut dialokasikan ke dalam kluster, sementara nilai 0 menunjukkan bahwa data tersebut tidak dialokasikan ke dalam kluster tersebut.
5. Kembali ke langkah 2, ulangi proses tersebut sampai nilai centroid tetap dan anggota kluster tidak berpindah ke kluster lain.



Gambar 2. Flow Chart K-Means Clustering

4. HASIL DAN PEMBAHASAN

4.1. Data

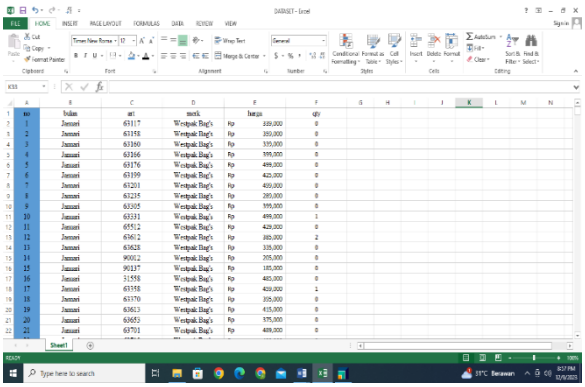
Pada penelitian ini dataset tentang penjualan tas di Asia Toserba Cirebon, dengan merk Westpak bag's, Subway dan Cannon sebanyak 1.113 data, pada periode Januari, Februari, Maret, April, Mei, dan Juni pada Tahun 2023. Berikut dapat dilihat pada gambar 3.

No	bulan	art	merk	harga	qty
1	Januari	63117	Westpak Bag's	Rp 339,000	0
2	Januari	63158	Westpak Bag's	Rp 359,000	0
3	Januari	63160	Westpak Bag's	Rp 339,000	0
191	Februari	63166	Westpak Bag's	Rp 399,000	0
192	Februari	63176	Westpak Bag's	Rp 499,000	0
193	Februari	63201	Westpak Bag's	Rp 499,000	0
365	Maret	31527	Westpak Bag's	Rp 375,000	0
367	Maret	31536	Westpak Bag's	Rp 309,000	0
538	April	31558	Westpak Bag's	Rp 485,000	0
539	April	31566	Westpak Bag's	Rp 349,000	0
540	April	62306	Westpak Bag's	Rp 489,000	0
702	Mei	31527	Westpak Bag's	Rp 375,000	1
703	Mei	31533	Westpak Bag's	Rp 445,000	0
1111	Juni	22317	Subway 50%	Rp 229,000	3
1112	Juni	ss 22322	Subway 50%	Rp 229,000	3
1113	Juni	22377	Subway 50%	Rp 229,000	3

Gambar 3. Dataset Penjualan

4.2. Data Cleaning

Setelah mendapatkan data yang relevan, agar data siap untuk diolah maka dilakukan data Cleaning untuk mendeteksi data ganda, dan data outlier seperti gambar 4.

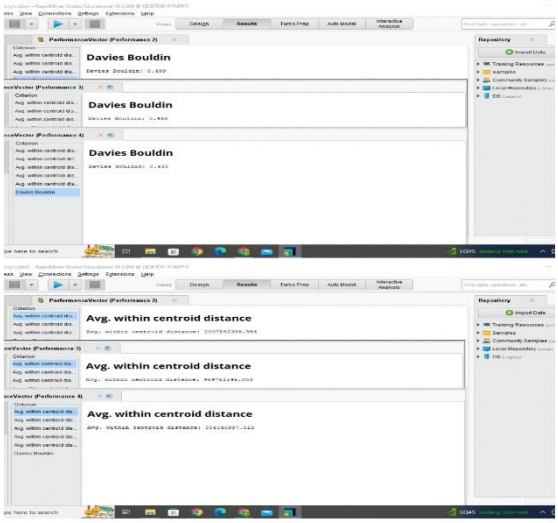


Gambar 4. Hasil Proses Data Preprocesing

4.3. Data Transformation

Setelah data clear dari data ganda dan lainnya, lalu dilakukan perubahan data sesuai dengan atribut yang ingin diteliti, dalam penelitian ini atribut yang

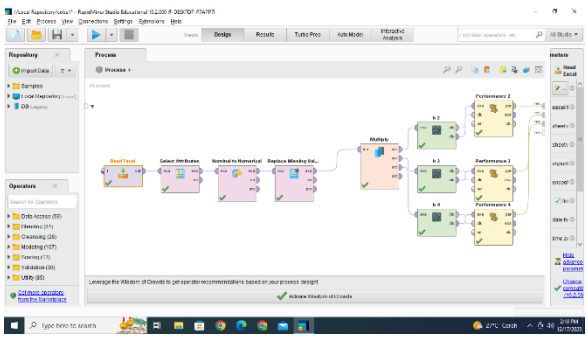
tidak ingin diteliti adalah nomer, selanjutnya atribut tersebut dihapus seperti gambar 5.



Gambar 5. Hasil Proses Data Transformation

4.4. Data Model

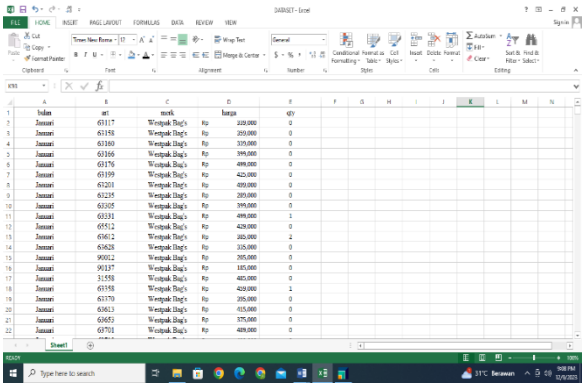
Pengelompokkan dilakukan menggunakan Aplikasi *Rapid Miner*, *rapid miner* merupakan *software*/perangkat lunak untuk pengolahan data. Dengan menggunakan prinsip dan *algoritma* data mining, rapid miner mengekstrak pola-pola dari data set yang besar dengan mengkombinasikan metode statistika, kecerdasan buatan dan database. Berdasarkan data gambar 6 tentang model *algoritma k-means* menjelaskan bahwa operator yang digunakan yaitu *read excel* dipergunakan untuk memanggil dataset. Kemudian *select attributes* digunakan untuk memilih atribut dari contoh dataset. Operator *nominal to numerical* digunakan mengubah tipe *non numeric* menjadi *numeric*. Selanjutnya operator *replace missing values* digunakan untuk menggantikan nilai yang hilang. Operator *multiply* digunakan untuk mengambil objek dari *port input* dan mengirimkan salinannya ke *port output.* Operator *clustering* digunakan untuk memodelkan dataset yang sudah ada. Dan yang terakhir operator *cluster distance performance* digunakan untuk evaluasi kinerja, memberikan daftar nilai kriteria kinerja.



Gambar 6. Model Proses K-Means Clustering

4.5. Hasil Evaluasi

Berdasarkan hasil implementasi *algoritma k-means* maka didapatkan hasil *performance* sebagai berikut:

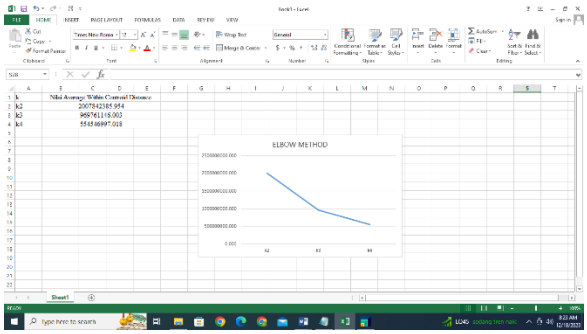


Gambar 7. Hasil Nilai *Avarage Within Centroid Distance*

Berdasarkan gambar 7 menjelaskan bahwa nilai yang di dapat adalah *Avarage Within Centroid Distance* dari 3 *clustering* yaitu k2 = 2007842385.954, k3 = 969761146.003, k4 = 554546997.018. Lalu jumlah *cluster* yang tepat atau sudah sangat baik dibuktikan dengan *elbow method* dengan hasil k3, dan nilai *Davies Bouldin Index* (DBI) yang mendekati nol pada saat k4 dengan nilai DBI k4 = 0.435.

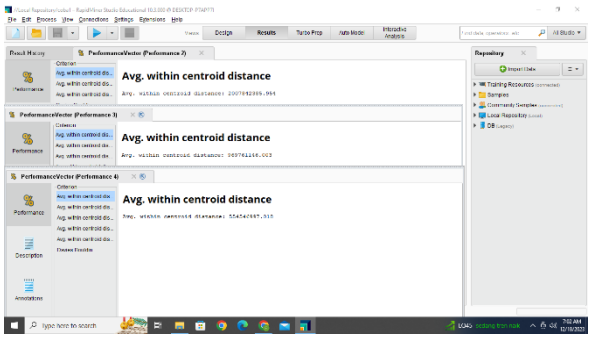
4.6. Pembahasan

Untuk mengoptimasi *k-means* digunakan teknik *Elbow Method* untuk menentukan nilai (k) yang optimum. Metode Elbow adalah suatu teknik yang digunakan untuk memperoleh informasi dalam menentukan jumlah cluster optimal dengan melihat persentase perbandingan antara jumlah cluster, yang dapat terlihat dari titik di mana grafik membentuk siku. Metode ini menghasilkan konsep dengan cara memilih nilai cluster dan menambahkannya ke model data untuk menentukan cluster terbaik. Selain itu, persentase perhitungan yang dihasilkan menjadi pembanding antara jumlah cluster yang ditambahkan. Perbedaan persentase dari setiap nilai cluster dapat dipresentasikan melalui grafik sebagai sumber informasi. Jika antara nilai cluster pertama dan nilai cluster kedua terbentuk sudut atau mengalami penurunan paling besar dalam grafik, maka nilai cluster tersebut dianggap sebagai yang terbaik.[11] Nilai (k) optimum diperoleh dari perbandingan nilai *avarege within centroid distance* dari ke-3 *clustering* yang sudah dibentuk, lalu nilai tersebut dimasukkan ke *Microsoft Excel*, setelah itu nilai *avarege within centroid distance* tersebut diseleksi, kemudian divisualisasi menggunakan kurva (x,y), seperti dilihat pada gambar 4.23 menunjukkan garis mengalami patahan yang membentuk *elbow* atau siku pada saat *k* = 3. Maka dengan menggunakan metode ini diperoleh nilai *k* optimal pada saat berada di *k* = 3 artinya penentuan jumlah *cluster* terbaik adalah 3 cluster.



Gambar 8. Hasil Optimasi Teknik *Elbow Method*

Dari Operator Cluster Distance Performance didapatkan hasil nilai Avarage Within Centroid Distance dari ke 3 clustering yang sudah diproses yaitu, k=2: 2007842385.954, k=3: 969761146.003, k=4: 554546997.018. seperti gambar 9.



Gambar 9. Nilai *Avarage Within Centroid Distance*

Davies bouldin index (DBI) adalah metric untuk mengevaluasi atau mempertimbangkan hasil algoritma clustering. Pertama kali diperkenalkan oleh David L. Davies dan Donald W. Bouldin pada tahun 1979. Dengan menggunakan DBI suatu cluster akan dianggap memiliki skema clustering yang optimal adalah yang memiliki DBI minimal.
Langkah-langkah perhitungan Davies Bouldin Index adalah Sebagai berikut :

Sum Of Square Within-Cluster (SSW) Untuk mengetahui kohesi dalam sebuh cluster ke-I salah satunya adalah dengan menghitung nilai dari Sum Of Square Within-Cluster (SSW). Dengan rumus sebagai berikut :

$$SSW_i = \frac{1}{mi} \sum_{j=i}^{mi} d(X_j, C_j) \quad \text{.....(1)}$$

Dimana :

mi = jumlah data dalam cluster ke-i
ci = centroid cluster ke-i
d(X_j,C_j) = jarak setiap data ke centroid i yang dihitung menggunakan jara jarak euclidiance.

Sum Of Square Between-Cluster (SSB) Perhitungan Sum Of Square Between-Cluster (SSB) bertujuan untuk mengetahui separasi atau jarak antar cluster. dengan rumus perhitungan sebagai berikut

$$SSB_{ij} = d\,(X_i, X_j) \quad \text{.....(2)}$$

Dimana :

d (X_i,X_j) = jarak antara data ke i dengan data ke j di cluster lain.

Ratio (Rasio)
Perhitungan rasio (Ri,j) ini bertujuan untuk mengetahui nilai perbandingan antara cluster ke-i dan cluster ke-j untuk menghitung nilai rasio yang dimiliki oleh masing-masing cluster. indeks I dan j merupakan merepresentasikan jumlah cluster, dimana jika terdapat 4 cluster maka terdapat indeks sebanyak 4 yaitu i,j,k dan l. untuk menentukan nilai rasio dengan rumus sebagai berikut :

$$R_{ij\ldots\ldots,n} = \frac{SSW_i + SSW_j + \cdots + SSW_n}{SSB_{i,j} + \cdots + SSB_{ni,nj}} \quad \text{.....(3)}$$

Dimana :

〖SSW〗_i = Sum Of Square Within-Cluster pada centroid i
〖SSB〗_(i,j) = Sum of Square Between Cluster data ke i dengan j pada cluster yang berbeda

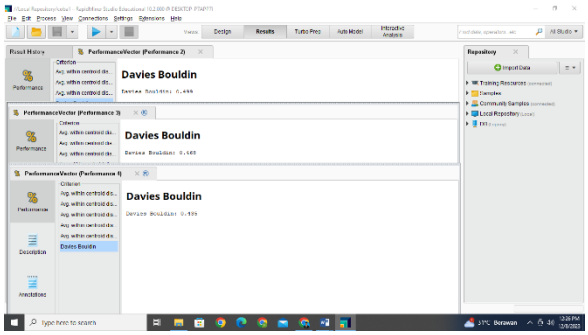
Pada rumus perhitungan 2.5 n akan berlanjut sejumlah cluster yang dipilih dengan syarat ni tidak sama dengan nj.

Davies Bouldin Index (DBI)
Nilai rasio yang diperoleh dari rumus 2.5 digunakan untuk mencari nilai DBI dengan menggunakan perhitungan sebagai berikut :

$$DBI = \frac{1}{k} \sum_{i=1}^{k} max_{\ i \neq j}(Ri, j, \ldots k) \quad \text{.....(4)}$$

Dimana, Ri,j merupakan ratio dari nilai SSW dan SSB melalui perhitungan rumus 2.5 dari perhitungan 2.6 maka dapat diketahui k adalah jumlah cluster. Dari perhitungan Davies Bouldin Index (DBI) dapat disimpulkan bahwa jika semakin kecil nilai Davies Bouldin Index (DBI) yang diperoleh (non negatif >= 0) maka cluster tersebut semakin baik.

Dari Operators Cluster Distance Performance diperoleh juga nilai Davies-Bouldin Index (DBI) dari ke-3 clustering yaitu, k=2: 0.499, k=3: 0.468, k=4 0.435, dilihat dari hasil tersebut nilai Davies-Bouldin Index (DBI) yang paling mendekati nol ialah pada saat k = 4: 0.435, sehingga tingkat akurasi dari hasil cluster termasuk baik, karena nilai DBI sudah mendekati nol seperti digambar 10.



Gambar 10. Nilai *Davies Bouldin*

5. KESIMPULAN DAN SARAN

Berdasarkan penelitian yang telah diuraikan diatas maka dapat disimpulkan hasil dari penelitian "Penerapan data mining menggunakan metode k-means clustering pada penjualan tas di Asia Toserba Cirebon" adalah nilai avarage within centroid distance dari ke 3 clustering dibuktikan pada operator cluster distance performance diperoleh *k* = 2: 2007842385.954, *k* = 3: 969761146.003, *k* = 4: 554546997.018. Jumlah cluster yang sudah ideal pada penelitian ini dibuktikan dengan Elbow Method adalah pada saat k = 3. Nilai davies bouldin index (DBI) yang mendekati nol adalah pada saat k = 4 dengan nilai davies bouldin k = 4: 0.435.

DAFTAR PUSTAKA

[1] A. Nugraha, O. Nurdiawan, and ..., "PENERAPAN DATA MINING METODE K-MEANS CLUSTERING UNTUK ANALISA PENJUALAN PADA TOKO YANA SPORT," *JATI (Jurnal Mhs. …*, 2022, [Online]. Available: https://ejournal.itn.ac.id/index.php/jati/article/view/5755

[2] G. Aprilianur and E. L. Hadisaputro, "Penerapan Data Mining Menggunakan Metode K-Means Clustering Untuk Analisa Penjualan Toko Myam Hijab Penajam," *JUPITER (Jurnal Penelit. Ilmu …*, 2022, [Online]. Available: https://jurnal.polsri.ac.id/index.php/jupiter/article/view/4684

[3] H. Syahputra, "Clustering Tingkat Penjualan Menu (Food and Beverage) Menggunakan Algoritma K-Means," *J. KomtekInfo*, vol. 9, pp. 29–33, 2022, doi: 10.35134/komtekinfo.v9i1.274.

[4] I. S. Mangku Negara, P. Purwono, and I. A. Ashari, "Analisa Cluster Data Transaksi Penjualan Minimarket Selama Pandemi Covid-19 dengan Algoritma K-means," *JOINTECS (Journal Inf. Technol. Comput. Sci.*, vol. 6, no. 3, p. 153, 2021, doi: 10.31328/jointecs.v6i3.2693.

[5] E. Ramadanti and M. Muslih, "Penerapan Data Mining Algoritma K-Means Clustering Pada Populasi Ayam Petelur Di Indonesia," *Rabit J. Teknol. dan Sist. Inf. Univrab*, vol. 7, no. 1, pp. 1–7, 2022, doi: 10.36341/rabit.v7i1.2155.

[6] H. Pandiangan, "Penerapan Data Mining Dalam Clustering Produksi Daging Sapi Di Indonesia Menggunakan Algoritma K-Means," *J. Comput. Networks, Archit. High Perform. Comput.*, vol. 1, no. 2, pp. 37–44, 2019, doi: 10.47709/cnapc.v1i2.239.

[7] N. Erlangga, S. Solikhun, and I. Irawan, "Penerapan Data Mining Dalam Mengelompokan Produksi Jagung Menurut Provinsi Menggunakan Algoritma K-Means," *KOMIK (Konferensi Nas. Teknol. Inf. dan Komputer)*, vol. 3, no. 1, pp. 702–709, 2019, doi: 10.30865/komik.v3i1.1681.

[8] A. Triayudi and A. Iskandar, "Penerapan Data Mining Dalam Penentuan Proritas Pemesanan Produk Berdasarkan dengan Data Penjualan Barang Menggunakan Algoritma Apriori," *J. Comput. Syst. …*, vol. 4, no. 1, pp. 25–30, 2022, doi: 10.47065/josyc.v4i1.2523.

[9] W. A. Wahyuni and S. Saepudin, "Penerapan Data Mining Clustering Untuk Mengelompokkan Berbagai Jenis Merk Mesin Cuci," *Semin. Nas. Sist. …*, pp. 306–313, 2021, [Online]. Available: https://sismatik.nusaputra.ac.id/index.php/sismatik/article/view/35%0Ahttps://sismatik.nusaputra.ac.id/index.php/sismatik/article/download/35/31

[10] R. Hasibuan Budiansyah, H. Hafizah, and R. Mahyuni, "Penerapan Data Mining Clustering Dengan Menggunakan Algoritma K-Means Pada Data Nasabah Kredit Bermasalah PT. BPR Milala," *J-SISKO TECH (Jurnal Teknol. Sist. Inf. dan Sist. Komput. TGD)*, vol. 5, no. 1, p. 7, 2022, doi: 10.53513/jsk.v5i1.4767.

[11] V. A. Ekasetya and A. Jananto, "Klusterisasi Optimal Dengan Elbow Method Untuk Pengelompokan Data Kecelakaan Lalu Lintas Di Kota Semarang," *J. Din. Inform.*, vol. 12, no. 1, pp. 20–28, 2020, doi: 10.35315/informatika.v12i1.8159.