PENERAPAN ALGORITMA *K-MEANS CLUSTERING* DALAM PENGELOMPOKAN DATA PENJUALAN SUPERMARKET BERDASARKAN CABANG (*BRANCH*)

Via Alvianatinova[1], Irfan Ali[2], Nining Rahaningsih[3], Agus Bahtiar[4]
[1,3] Komputerisasi Akuntansi, STMIK IKMI Cirebon
[2] Rekayasa Perangkat Lunak, STMIK IKMI Cirebon
[4] Sistem Informasi, STMIK IKMI Cirebon
Jl. Perjuangan No.10B, Karyamulya, Kota Cirebon, Jawa Barat, Indonesia
viaalvianatinova@gmail.com

ABSTRAK

Penjualan ritel, terutama dalam konteks supermarket, merupakan aspek krusial dalam operasional bisnis yang memerlukan pengelolaan data yang efisien. Penelitian ini dilakukan untuk mengeksplorasi implementasi algoritma K-Means clustering dalam mengelompokkan data penjualan dari berbagai cabang supermarket, dengan fokus utama pada peningkatan efisiensi operasional dan strategi penjualan. Dalam era digital saat ini, penjualan supermarket menghasilkan volume besar dan data penjualan yang kompleks setiap hari. Pengelolaan dan pemahaman data ini menjadi tantangan signifikan, terutama ketika terdapat banyak cabang yang tersebar luas. Algoritma K-Means clustering telah terbukti sebagai metode yang efektif dalam menyelesaikan permasalahan semacam ini. Metode ini memungkinkan pengelompokkan data penjualan ke dalam kelompok-kelompok yang memiliki karakteristik serupa, mempermudah analisis dan pengambilan keputusan. Studi ini mengumpulkan data historis penjualan dari berbagai cabang supermarket. Data diproses terlebih dahulu untuk memastikan kualitasnya sebelum menerapkan algoritma K-Means clustering. Hasil pengelompokan data dianalisis secara menyeluruh untuk mengidentifikasi pola penjualan utama. Analisis ini menjadi dasar untuk meningkatkan efisiensi operasional setiap toko, termasuk manajemen inventaris dan strategi penjualan. Tujuan penelitian ini adalah untuk memahami cara mengoptimalkan pengelolaan data penjualan supermarket menggunakan algoritma K-Means clustering. Hasilnya menunjukkan bahwa toko dapat dikelompokkan menjadi dua kelompok utama, yaitu kelompok cabang besar (cluster 0) dan kelompok cabang kecil (cluster 1). Penerapan algoritma K-Means clustering memungkinkan pengelompokkan data penjualan supermarket berdasarkan toko, memberikan kontribusi signifikan terhadap pemahaman dan pengelolaan data penjualan secara lebih efisien. Evaluasi model dengan indeks Davies Bouldin menghasilkan nilai sebesar 0.375, menunjukkan keberhasilan tinggi dalam mengelompokkan data.

Kata kunci : Algoritma *K-Means Clustering*, Data Penjualan, *Supermarke*t, Pengelompokan Data

1. PENDAHULUAN

Dalam era informasi yang berkembang pesat, data penjualan di sektor retail, termasuk supermarket, menjadi indikator penting dalam bisnis. Pengelolaan data penjualan yang efektif kini memanfaatkan Algoritma K-Means clustering. Algoritma ini membantu mengidentifikasi pola penjualan, mengoptimalkan inventaris, dan meningkatkan efisiensi operasional di berbagai cabang supermarket [1].

Beberapa supermarket masih menggunakan proses manual, menyebabkan kesalahan dan tidak efisien. Efisiensi operasional dan strategi penjualan yang tepat kritis dalam industri ini. Setiap cabang supermarket memiliki kebutuhan yang berbeda, memerlukan pendekatan untuk mengidentifikasi pola penjualan dan mengelompokkan data sesuai industri [2].

Penerapan algoritma K-Means pada data penjualan supermarket membantu mengidentifikasi kelompok penjualan yang tidak selalu terlihat langsung oleh manusia. Sebagai contoh, beberapa cabang mungkin memiliki pola penjualan serupa meskipun berlokasi geografis berjauhan, atau beberapa produk menunjukkan performa penjualan konsisten di berbagai cabang. Algoritma *K-Means Clustering* mengungkap struktur tersembunyi dalam data penjualan, memberikan wawasan berharga untuk pengambilan keputusan manajemen yang lebih cerdas [3].

Tantangan besar muncul dari jumlah data yang besar, variasi produk yang luas, dan preferensi konsumen yang beragam. Tanpa alat analisis yang tepat, pola penjualan yang relevan mungkin terlewat, mengakibatkan operasional yang tidak efisien dan efektif. Inilah sebabnya mengapa diperlukan alat analisis yang canggih untuk mengoptimalkan strategi penjualan [4]..

Penelitian ini bertujuan untuk menerapkan algoritma *K-Means clustering* dalam menganalisis data penjualan supermarket, dengan fokus mengidentifikasi pola penjualan yang mungkin terlewatkan oleh metode analisis tradisional. Tujuan utama adalah memberikan wawasan mendalam tentang preferensi konsumen, pola penjualan produk, dan performa cabang supermarket. Secara praktis, penelitian ini bertujuan untuk meningkatkan efisiensi operasional dan merumuskan strategi penjualan yang

lebih cerdas, dengan harapan hasil temuan dapat memberikan kontribusi positif bagi industri ritel, khususnya pada tingkat penjualan supermarket.

Hasil dari penelitian ini diharapkan memberikan kontribusi berharga bagi dunia akademik dengan menyajikan temuan-temuan yang dapat memperkaya literatur ilmiah di bidang analisis data penjualan. Secara praktis, penelitian ini juga diharapkan dapat memberikan panduan kepada pelaku industri ritel, khususnya penjualan supermarket, dalam mengoptimalkan operasional dan meningkatkan keberhasilan bisnis mereka.

2. TINJAUAN PUSTAKA

2.1. Data Mining

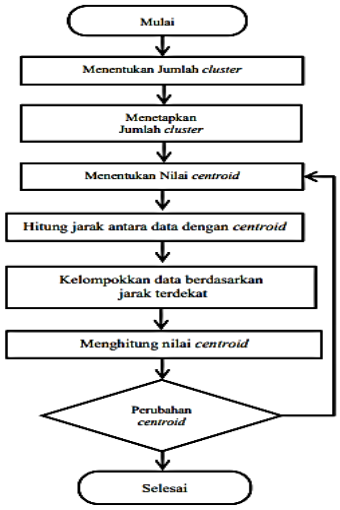
Data mining, sebagai proses ekstraksi informasi dari data besar, kompleks, dan tidak terstruktur, memungkinkan identifikasi pola yang sulit ditemukan secara manual. Algoritma K-Means clustering, sebagai teknik data mining, digunakan untuk mengelompokkan data penjualan supermarket dan mengungkap struktur tersembunyi dalam pola penjualan. Melalui pendekatan ini, penelitian ini bertujuan untuk mendapatkan wawasan yang lebih dalam tentang pola penjualan di sektor ritel dan merumuskan strategi penjualan yang lebih terarah. Oleh karena itu, landasan teori data mining memberikan dasar yang solid untuk memahami relevansi dan potensi algoritma *K-Means clustering* dalam mengolah informasi berharga dari data penjualan supermarket [5].

2.2. *Knowledge Discovery in Database* (KDD)

Knowledge Discovery in Database (KDD) mengacu pada proses ekstraksi pengetahuan yang bermanfaat dan berharga dari dalam basis data yang besar dan kompleks. Ini adalah pendekatan terintegrasi yang melibatkan langkah-langkah seperti pemilihan data, preprocessing, transformasi, data mining, evaluasi, dan interpretasi hasil [6], [7]. Tujuan utama dari KDD adalah untuk menemukan pola, tren, dan pengetahuan tersembunyi dalam data yang dapat memberikan wawasan untuk pengambilan keputusan. Proses KDD memanfaatkan berbagai teknik dan algoritma data mining untuk menggali informasi yang dapat digunakan untuk meningkatkan pemahaman tentang data, mengidentifikasi hubungan yang tidak terlihat secara langsung, dan merumuskan strategi atau keputusan yang lebih baik dalam berbagai bidang, termasuk bisnis, ilmu pengetahuan, dan penelitian [8].

2.3. *K-Means clustering*

K-Means Clustering merupakan sebuah metode dalam analisis data yang digunakan untuk mengelompokkan data ke dalam beberapa kelompok atau klaster berdasarkan kesamaan karakteristik tertentu [9]. Algoritma ini bertujuan untuk meminimalkan varians dalam setiap klaster dengan menempatkan data ke pusat klaster yang sesuai. Dalam konteks penelitian ini, *K-Means Clustering* digunakan untuk mengelompokkan data penjualan supermarket berdasarkan pola yang mungkin tidak terlihat secara langsung, membantu mengidentifikasi kelompok penjualan yang serupa atau memiliki karakteristik yang sama. Dengan penerapan K-Means Clustering, penelitian ini berusaha untuk mengungkap struktur tersembunyi dalam data penjualan dan memberikan wawasan yang lebih dalam mengenai preferensi konsumen dan performa produk di berbagai cabang supermarket. Dibawah ini merupakan gambar flowchart penelitian analisis data penjualan supermarket dengan menggunakan algoritma *K-Means Clustering* [10].



Gambar 1. K-Means Clustering

Untuk menentukan cluster menggunakan K-Means Clustering, langkah-langkahnya dapat dijelaskan sebagai berikut:

1) Pemilihan Jumlah Cluster (K)
   Tentukan jumlah cluster (K) yang diinginkan. Ini bisa berdasarkan pengetahuan awal atau menggunakan metode seperti elbow method untuk menentukan jumlah optimal cluster.
2) Inisialisasi Pusat Cluster
   Pilih secara acak K titik sebagai pusat awal dari setiap cluster.
3) Pengelompokkan Data
   Setiap titik data akan diberikan label cluster berdasarkan pusat cluster terdekat.
4) Perbaruan Pusat Cluster
   Hitung ulang pusat setiap cluster berdasarkan rata-rata dari semua titik data yang termasuk dalam cluster tersebut.
5) Iterasi
   Langkah 3 dan 4 diulang hingga tidak ada perubahan signifikan dalam pengelompokkan atau mencapai jumlah iterasi yang ditentukan.
6) Evaluasi Hasil
   Evaluasi kualitas cluster menggunakan metrik seperti inertia (sum squared distance antara setiap titik data dengan pusat clusternya), Davies-Bouldin Index, atau metrik lainnya.

2.4. Rapidminer Studio

RapidMiner merupakan sebuah platform open-source untuk data mining dan analisis prediktif yang menyediakan berbagai alat untuk merancang dan mengimplementasikan proses analisis data secara intuitif. RapidMiner memungkinkan peneliti untuk mengintegrasikan berbagai metode analisis data, termasuk algoritma clustering seperti K-Means, dalam suatu alur kerja yang terstruktur. Kelebihan RapidMiner terletak pada antarmuka pengguna yang ramah, memfasilitasi penggunaan algoritma kompleks tanpa memerlukan keterampilan pemrograman yang tinggi. Dalam penelitian ini [11], RapidMiner digunakan sebagai alat utama untuk menerapkan algoritma *K-Means Clustering* pada data penjualan supermarket, mempercepat dan menyederhanakan proses analisis data. Gambar flowchart penelitian juga mencakup langkah-langkah yang melibatkan RapidMiner sebagai platform utama dalam proses analisis data penjualan supermarket [12].

3. METODE PENELITIAN

3.1. Sumber Data

Sumber data dalam penelitian ini di ambil pada Website resmi Kaggel yaitu di *https://www.kaggle.com/datasets/aungpyaeap/supermarket-sales*. Data ini meliputi data Supermarket sales, waktu pengambilan data dilaksanakan pada tanggal 26 Juli 2023. Data sekunder adalah data yang dikumpulkan secara tidak langsung oleh pihak lain dengan sumber informasi melalui jurnal, buku, website, digunakan untuk kebutuhan data penelitian.

3.2. Teknik Pengumpulan Data

Pengumpulan data penjualan dapat dilakukan melalui analisis data historis atau ekstraksi data dari database penjualan dan pencatatan data transaksi. Teknik atau metode yang digunakan untuk mengumpulkan data adalah pengambilan data secara online (online riset). Pencarian data online adalah pencarian data yang dilakukan melalui internet atau website dengan tujuan memperoleh informasi berupa informasi teoritis dan data yang digunakan dalam penelitian ilmiah.[13]

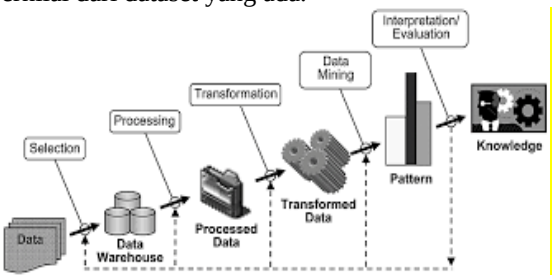
3.3. Tahapan Perancangan

Berikut adalah rancangan penelitian untuk metode Knowledge Discovery in Database (KDD):

Berikut adalah tahapan penyelasaian untuk proses Knowledge Discovery in Database (KDD) berdasarkan urutan yang disediakan:

1) Seleksi Data (Selection)
   Pilih data yang akan digunakan dari sekumpulan data operasional. Proses ini harus dilakukan sebelum memulai tahap penemuan informasi dalam KDD. Data yang dipilih disimpan dalam file terpisah dari basis data operasional.
2) Pemilihan Data (Pre-processing)
   Lakukan pre-processing pada data yang telah dipilih. Ini melibatkan penghapusan data duplikat, pengecekan konsistensi data, dan perbaikan kesalahan pada data. Proses enrichment juga dilakukan untuk memperkaya data dengan informasi eksternal.
3) Transformasi Data (Data Transformation)
   Setelah data dibersihkan, langkah berikutnya adalah transformasi data. Proses ini melibatkan konversi data mentah ke dalam format yang lebih sesuai untuk proses data mining.
4) Data Mining
   Terapkan proses data mining untuk mencari pola dan informasi menarik dalam data terpilih. Pengujian dilakukan menggunakan software Rapidminer.
5) Interpretation/Evaluation
   Evaluasi merupakan fase terakhir pada tahap ini. Hasil dari proses penambangan data dan pola informasi diinterpretasikan. Evaluasi kualitas dan potensi kegunaan pola tersebut untuk tujuan bisnis atau penelitian. Dalam penelitian ini, metode indeks Davies-Bouldin digunakan untuk menentukan jumlah cluster optimal. Semakin mendekati nilai 0 pada Davies-Bouldin Index, semakin baik cluster yang diperoleh.

Tahapan ini mencakup proses mulai dari pemilihan data hingga interpretasi hasil, memberikan gambaran keseluruhan tentang bagaimana KDD dapat diimplementasikan untuk mendapatkan wawasan yang bernilai dari dataset yang ada.



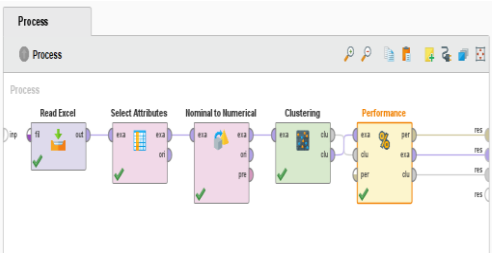
Gambar 2. Tahapan Metode KDD [13]

4. HASIL DAN PENGUJIAN

4.1. Hasil

4.1.1. Pemodelan Algoritma *K-Means clustering*

Menampilkan proses pengolahan data menggunakan metode K-Means dengan gambar dibawah ini.



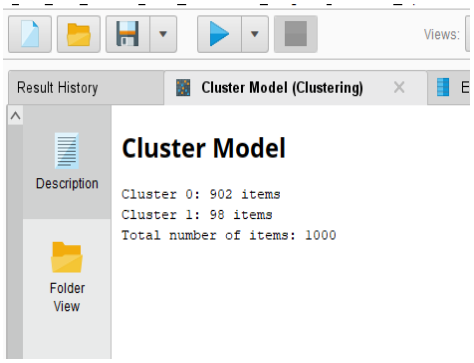
Gambar 3. Pemodelan algoritma *K-Means clustering*

Dalam pengolahan data menggunakan *K-Means Clustering* di RapidMiner, 5 operator

digunakan: Read Excel (membaca data Excel), Select Attribute (mengurangi data yang tidak dibutuhkan), Nominal to Numerical (mengubah atribut nominal menjadi numerik), Clustering (K-Means) (membagi data menjadi 2 klaster dengan 10 putaran maksimal), dan Performance (Cluster Distance Performance) (mengevaluasi kinerja clustering pada penjualan supermarket).

4.1.2. *Cluster Model*

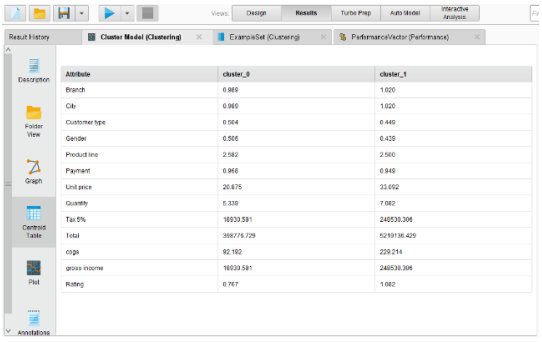
Distribusi jumlah anggota pada setiap klaster, seperti yang terlihat pada gambar di bawah, mencerminkan karakteristik unik dari masing-masing klaster. Analisis lebih lanjut terhadap pola distribusi ini dapat memberikan wawasan mendalam terkait perbedaan dan kesamaan antar klaster dalam konteks tertentu. Dengan demikian, memahami dinamika internal klaster dapat menjadi kunci untuk mengoptimalkan performa model dan memahami struktur data secara lebih komprehensif anggota tiap klaster ditunjukkan pada gambar dibawah ini.



Gambar 4. Cluster Model

Dari gambar diatas dapat dijelaskan bahwa untuk klaster 0 memiliki 902 anggota Branch Id dan klaster 1 memiliki 98 anggota Branch Id.

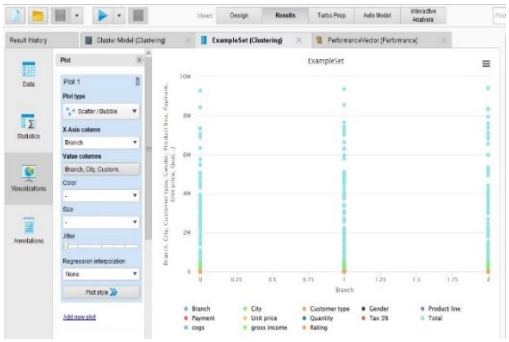
4.1.3. *Centroid Table*



Gambar 5. *Centroid Table*

Pada gambar 5. untuk mengetahui nilai rata rata pada atribut kode Branch, City, Cutomer Type, Gender, Produk Line, Unit price, Quantity, Tax 5%, Total, Payment, Cogs,, Gross Income, Rating. Merupakan centroid table yang dihasilkan oleh RapidMiner.

4.1.4. *Visualizations*

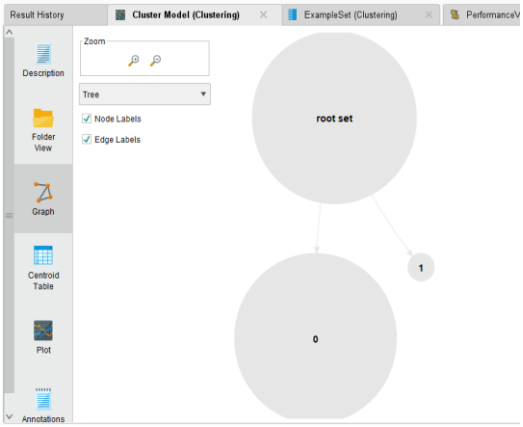


Gambar 6. Visualizations

Pada gambar 6. merupakan visualisasi data menggunakan Visualizations Scalter/ Bubble, di sana bisa melihat visualisasi data pada atribut Branch, pada gambar visualisasi Scalter/ Bubble di atas ada Branch A, Branch B dan Branch C.

4.1.5. *Graph*

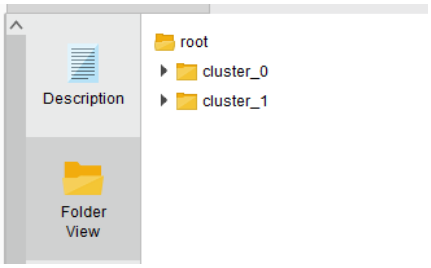
Gambar 7. di bawah ini adalah model yang di hasilkan dari data yang dimiliki dan menciptakan Rows pengetahuan.



Gambar 7. *Graph*

4.1.6. *Folder View*

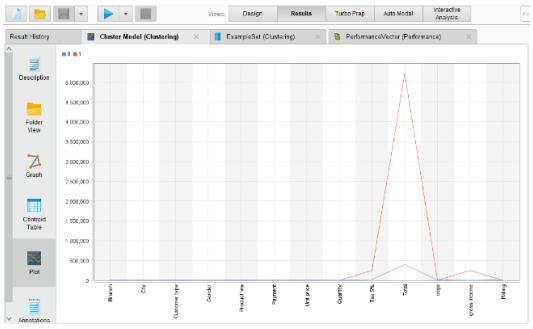
Pada gambar 8. di bawah terdapat Folder View untuk melihat jumlah anggota dan isi di dalam Cluster tersebut



Gambar 8. *Foler view*

4.1.7. *Plot*

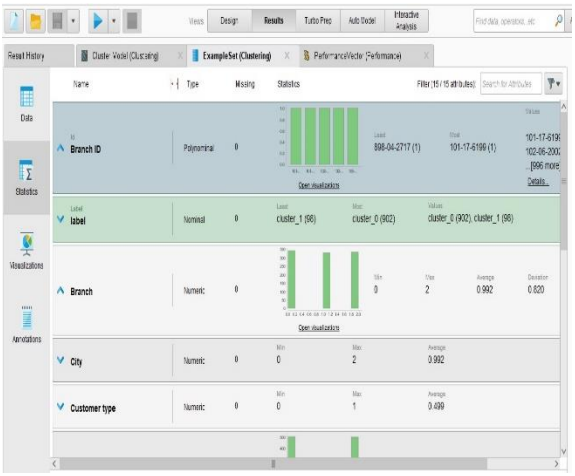
Dari tampilan plot bisa di lihat mana kelompok cluster tertinggi dan kelompok cluster terendah.



Gambar 9. *Plot*

4.1.8. *Statistics*

Pada tampilan Statistics di bawah ini bisa di lihat sebaran atribut data di masing masing Cluster dan bisa memilih atribut mana yang akan di lihat.



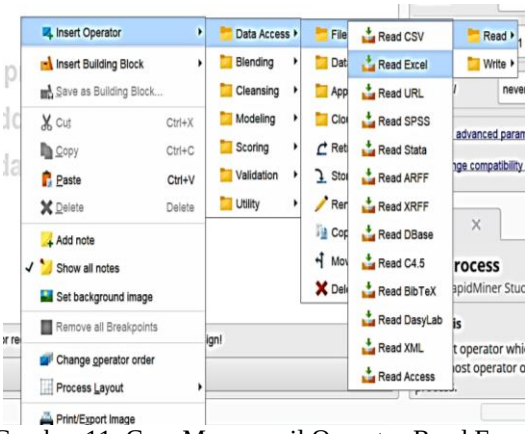
Gambar 10. Tampilan Statistics

4.2. Pengujian

4.2.1. Seleksi Data (*Data Selection*)

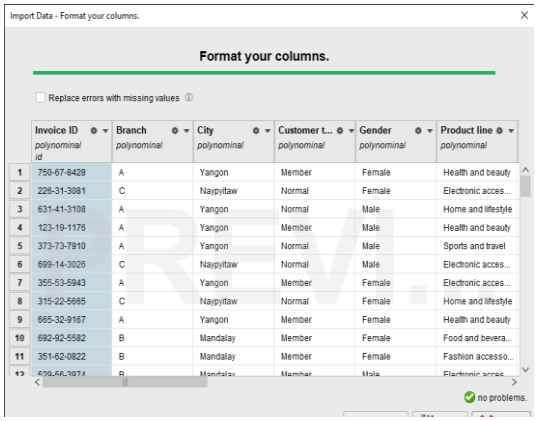
Pada tahap ini data yang digunakan pada penelitian ini yaitu data yang di peroleh dari Kaggle. Pada tanggal 26 Juni 2023, data ini terdapat 17 atribut yaitu Invoice ID, Branch, City, Cutomer Type, Gender, Produk Line, Unit price, Quantity, Tax 5%, Total, Date, Time, Payment, Cogs, Gross Margin Percentage, Gross Income, Rating. Banyaknya data yang diperoleh yaitu 1000 data.

Tahap pertama adalah menggunakan operator Read Excel kemudian dilakukan pemilihan data yang akan digunakan. Data yang sudah diproses dapat ditampilkan pada menu Results. Berikut tampilan dari proses operator Read Excel



Gambar 11. Cara Memanggil Operator Read Excel

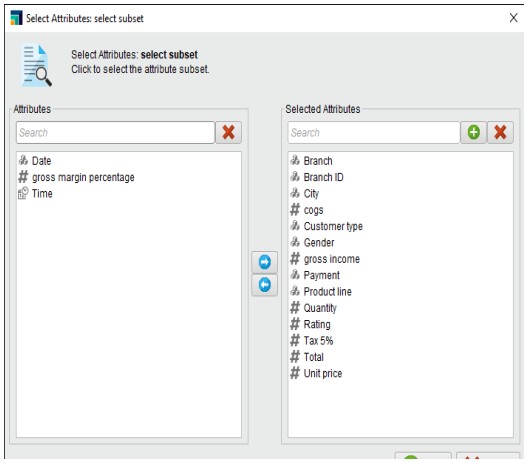
Selanjutnya pemilihan memasukan data yang akan di proses. Pada gambar di atas adalah pemilihan data yang akan di gunakan atau di proses pada Aplikasi RapidMiner. Setelah proses pemilihan data selsai maka akan muncul tampilan data mentah seperti gambar di bawah ini.



Gambar 12. Tampilan data mentah

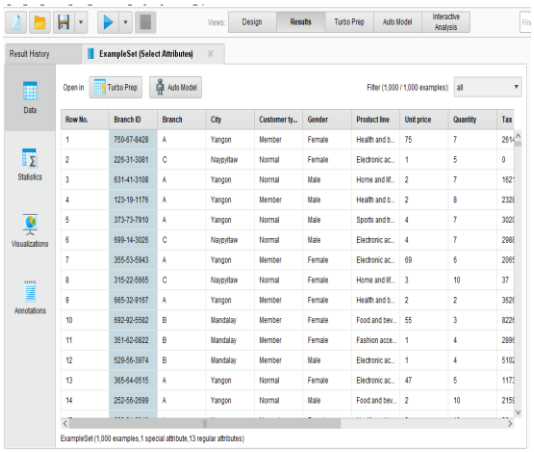
4.2.2. Pemilihan data (*Preprocessing*)

Tahapan ini merupakan cara untuk mengolah data dari sebelumnya data mentah menjadi data yang siap untuk diproses lebih lanjut. Tahapan dari Preprocessing adalah yang dilakukan yaitu Data Reduction. Data Reduction adalah pengurangan dari suatu data yang diambil tetapi hal tersebut tidak akan mengubah hasil analisis dari data itu sendiri. Atribut yang tidak dibutuhkan pada data Supermarket salles ini ada 3 atribut yaitu atribut kode Date, Time dan Gross Margin Percentage . Sehingga atribut yang digunakan untuk dianalisis lebih lanjut memiliki 14 atribut yaitu kode Invoice ID, Branch, City, Cutomer Type, Gender, Produk Line, Unit price, Quantity, Tax 5%, Total, Payment, Cogs,, Gross Income, Rating. . Operator Select Attribute akan digunakan dalam proses ini.



Gambar 13. Proses seleksi atribut

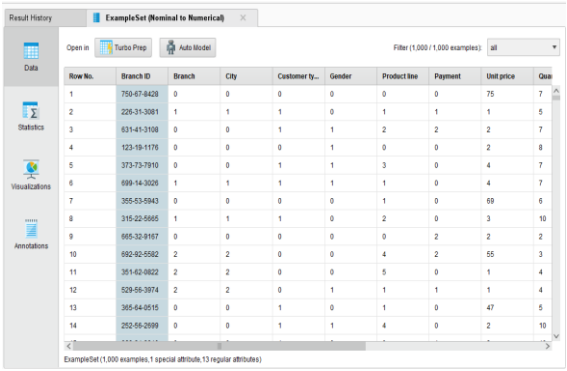
Dibawah ini adalah tampilan data yang sudah di proses dalam Select Attribute. Di sana Atribut yang tidak di pakai adalah Date, Gross margin percentage dan Time.



Gambar 14. Hasil dari Data Reduction

4.2.3. Transformasi data (*Transformation*)

Transformasi data adalah proses mengubah dataset dengan tujuan agar data lebih sesuai dengan proses pemodelan data mining. Dalam proses ini menggunakan algoritma *K-Means Clustering* dimana atribut dari suatu data harus berupa numerik. Oleh karena itu, data yang memiliki atribut nominal harus diubah menjadi numerik menggunakan operator Nominal to Numerical yaitu sebagai berikut.



Gambar 14. Hasil dari Transformasi data

Pada gambar diatas telah dilakukan transformasi data pada atribut Branch, City, Cutomer Type, Gender, Produk Line, Payment dari sebelumnya bentuk nominal menjadi numerik agar dapat diproses menggunakan algoritma K-Means Clustering.

5. KESIMPULAN

Berdasarkan hasil pengolahan data menggunakan beberapa perangkat lunak data mining, termasuk aplikasi RapidMiner, penelitian ini berhasil mengidentifikasi dua kelompok cluster dengan jumlah cabang yang signifikan berbeda di setiap cluster. Proses pengelompokan ini menghasilkan kelompok cabang besar (Cluster 0) dengan 902 item, dan kelompok cabang kecil (Cluster 1) dengan 98 item. Davies Bouldin Index (DBI) sebesar 0,375 memberikan gambaran evaluatif tentang kualitas pemisahan antar kelompok. Saran untuk penelitian selanjutnya dapat difokuskan pada analisis lebih lanjut terkait karakteristik masing-masing cluster, seperti pola penjualan dominan dan profil konsumen. Integrasi faktor eksternal dan penggunaan metrik evaluasi clustering yang lebih komprehensif juga dapat memperkaya pemahaman tentang performa penjualan cabang supermarket.

DAFTAR PUSTAKA

[1] I. Adityo, M. 1, and J. Heikal, "Customer Clustering Using The K-Means Clustering Algorithm in Shopping Mall in Indonesia," Management Analysis Journal, 2022. [Online]. Available: http://maj.unnes.ac.id

[2] M. Benri, H. Metisen, and S. Latipa, "ANALISIS CLUSTERING MENGGUNAKAN METODE K-MEANS DALAM PENGELOMPOKKAN PENJUALAN PRODUK PADA SWALAYAN FADHILA," in Jurnal Media Infotama, vol. 11, no. 2, 2015.

[3] Y. Darmi et al., "PENERAPAN METODE CLUSTERING K-MEANS DALAM PENGELOMPOKAN PENJUALAN PRODUK," in Jurnal Media Infotama, vol. 12, no. 2, 2016.

[4] E. Prasetyo, Data Mining: Konsep dan Aplikasi menggunakan MATLAB, Andi, 2012.

[5] E. Febrianty, L. Awalina, and W. I. Rahayu, "Optimalisasi Strategi Pemasaran dengan Segmentasi Pelanggan Menggunakan Penerapan K-Means Clustering pada Transaksi Online Retail," Jurnal Teknologi Dan Informasi (JATI), vol. 13, 2023. [Online]. Available: https://doi.org/10.34010/jati.v13i2

[6] M. L. J. M. Frederic Lardinois, "Google is acquiring data science community Kaggle," TechCrunch, Mar. 8, 2017. [Online]. Available: https://techcrunch.com/2017/03/07/google-is-acquiring-data-science-community-kaggle/

[7] M. Harahap, Y. Lubis, and Z. Situmorang, "Analisis Pemasaran Bisnis dengan Data Science : Segmentasi Kepribadian Pelanggan

berdasarkan Algoritma K-Means Clustering," Data Sciences Indonesia (DSI), vol. 1, no. 2, 2022. [Online]. Available: https://doi.org/10.47709/dsi.v1i2.1348

[8] P. S. Hasugian, "PENERAPAN DATA MINING UNTUK KLASIFIKASI PRODUK MENGGUNAKAN ALGORTIMA K-MEANS (STUDI KASUS : TOKO USAHA MAJU BARABAI)," Jurnal Mantik Penusa, vol. 2, no. 2, pp. 191–198, 2018.

[9] J. Mantik and N. Hidayati, "Implementation of data mining to determine stock inventory at kenza grocery stores using the k-means clustering method," in Jurnal Mantik, vol. 6, no. 3, 2022. [Online]. Available: [Provide the link if available]

[10] M. Mirbod and H. Dehghani, "Smart Trip Prediction Model for Metro Traffic Control Using Data Mining Techniques," Procedia Computer Science, vol. 217, pp. 72–81, 2022. https://doi.org/10.1016/j.procs.2022.12.203

[11] A. Nur Khomarudin, "Teknik Data Mining : Algoritma K-Means Clustering," 2003. [Online]. Available: https://agusnkhom.wordpress.com

[12] D. H. Nusti et al., "Application of K-Means Clustering Algorithm in Grouping Inventory Data at Putra Shop," JURNAL KOMITEK, vol. 1, no. 1, 2021. [Online]. Available: https://doi.org/10.53697/jkomitek.v1i1

[13] S. Rahayuni, I. Gunawan, and I. O. Kirana, "Material Sales Clustering Using the K-Means Method," JOMLAI: Journal of Machine Learning and Artificial Intelligence, vol. 1, no. 1, 2022. [Online]. Available: https://doi.org/10.55123/jomlai.v1i1.177

.