



## IMPLEMENTASI DATA MINING UNTUK MENGANALISA POLA PENJUALAN MENGUNAKAN METODE K-MEANS PADA CV. HAVAS P2S2

Alfan Jamil<sup>a</sup>, Zaehol Fatah<sup>b</sup>

<sup>a</sup> Sistem Informasi, [alfanjami2003@gmail.com](mailto:alfanjami2003@gmail.com), Universitas Ibrahimy, Situbondo Jawa Timur

<sup>b</sup> Sistem Informasi, [zaeholfatah@gmail.com](mailto:zaeholfatah@gmail.com), Universitas Ibrahimy, Situbondo Jawa Timur

\*Korespondensi

### ABSTRACT

*Sales are an important aspect in the continuity of company operations, including CV. Hafas P2S2 which is engaged in the distribution of bottled drinking water products. In order to increase the effectiveness of marketing strategies and stock management, it is important to analyze sales patterns that occur. This research aims to implement data mining techniques using the K-means method to analyze sales patterns at CV. Memorize P2S2. The K-means method was chosen because of its ability to find associative relationships between items that are often purchased together in transactions. The data used in this research involves sales information recorded in the company system. The results of applying the K-means algorithm show that there are certain combinations of items that are often sold together, which provides valuable insights for companies in terms of stock management and marketing strategies. It is hoped that these findings can help CV. Hafas P2S2 in improving operational efficiency and maximizing profit potential by better understanding customer demand patterns. Thus, the implementation of data mining through the K-means method makes a significant contribution to data-based decision making in the company's sales sector.*

**Keywords:** Mining, K-Means, Sales Patterns, Associative Analysis, CV. Memorize P2S2.

### Abstrak

Penjualan merupakan aspek penting dalam kelangsungan operasional perusahaan, termasuk pada CV. Hafas P2S2 yang bergerak dalam distribusi produk air minum dalam kemasan. Dalam rangka meningkatkan efektivitas strategi pemasaran dan pengelolaan stok, penting untuk menganalisis pola penjualan yang terjadi. Penelitian ini bertujuan untuk mengimplementasikan teknik data mining menggunakan metode K-means untuk menganalisis pola-pola penjualan di CV. Hafas P2S2. Metode K-means dipilih karena kemampuannya dalam menemukan hubungan asosiatif antar item yang sering dibeli bersama dalam transaksi. Data yang digunakan dalam penelitian ini melibatkan informasi penjualan yang tercatat dalam sistem perusahaan. Hasil dari penerapan algoritma K-means menunjukkan adanya kombinasi item tertentu yang sering terjual bersamaan, yang memberikan wawasan berharga bagi perusahaan dalam hal pengelolaan stok dan strategi pemasaran. Temuan ini diharapkan dapat membantu CV. Hafas P2S2 dalam meningkatkan efisiensi operasional dan memaksimalkan potensi keuntungan dengan lebih memahami pola permintaan pelanggan. Dengan demikian, implementasi data mining melalui metode K-means memberikan kontribusi signifikan dalam pengambilan keputusan berbasis data pada sektor penjualan perusahaan.

**Kata Kunci:** Data Mining, K-Means, Pola Penjualan, Analisis Asosiatif, CV. Hafas P2S2

### 1. PENDAHULUAN

Perkembangan kehidupan manusia saat ini sangat dipengaruhi oleh kemajuan teknologi yang begitu cepat, sehingga hampir semua aspek kehidupan menjadi bergantung pada penggunaan teknologi[1]. Dalam industri distribusi, seperti yang dijalankan oleh CV. Hafas P2S2, yang bergerak di bidang penjualan air minum dalam kemasan, pemahaman terhadap pola penjualan menjadi hal yang sangat penting. Melalui

analisis yang mendalam terhadap data penjualan, perusahaan dapat memperoleh informasi yang relevan untuk merumuskan strategi pemasaran yang lebih baik, mengoptimalkan pengelolaan stok, dan meningkatkan keuntungan. Namun, seringkali data penjualan yang terkumpul dalam jumlah besar tidak dapat dimanfaatkan secara maksimal tanpa adanya teknik analisis yang tepat.

Air Minum Syifa p2s2 merupakan suatu perusahaan yang bergerak dalam usaha produksi air minum dalam kemasan[2]. Salah satu metode yang dapat digunakan untuk menggali pola dalam data transaksi adalah data mining. Memanfaatkan teknik data mining untuk memudahkan petugas dalam mengelola data penjualan dan menganalisis pola serta tren penjualan secara efisien.

Salah satu teknik yang populer dalam data mining adalah metode K-Means, penerapan algoritma K-Means sangat sesuai karena terdapat hubungan antara berbagai item dalam penjualan peralatan kelengkapan depot air minum pada CV Hafas P2S2. Algoritma K-Means merupakan metode untuk mengekstraksi pola data melalui aturan asosiatif[3], yang digunakan untuk mengidentifikasi keterkaitan antara kombinasi item dalam transaksi.

Pentingnya penerapan data mining dalam bisnis modern tidak bisa dipandang sebelah mata. Hasil analisis yang dihasilkan dapat memberikan panduan bagi CV Hafas P2S2 dalam mengembangkan strategi pemasaran yang lebih efektif, serta meningkatkan efisiensi operasional. Dengan memahami pola penjualan, perusahaan dapat merespons perubahan pasar dengan lebih cepat dan akurat, sehingga dapat meningkatkan daya saing di pasar. Penggunaan algoritma K-Means melalui RapidMiner juga dapat membantu dalam mengidentifikasi pola-pola yang ada, termasuk dalam konteks CV Hafas P2S2. Dengan memanfaatkan alat ini, proses analisis data menjadi lebih akurat, memungkinkan untuk menemukan hubungan asosiatif antar item yang relevan dalam transaksi yang terjadi.

Penelitian ini bertujuan untuk mengeksplorasi implementasi data mining dalam menganalisis pola penjualan di CV Hafas P2S2. Melalui pendekatan yang sistematis, diharapkan perusahaan dapat mengambil keputusan yang lebih berbasis data, merumuskan strategi yang lebih tepat sasaran, dan meningkatkan kepuasan pelanggan. Dengan demikian, penelitian ini tidak hanya memberikan kontribusi bagi CV Hafas P2S2, tetapi juga bagi pengembangan praktik terbaik dalam penerapan data mining di sektor retail secara umum.

## **2. METODOLOGI PENELITIAN**

Dalam penelitian ini, metode yang digunakan untuk menganalisis pola penjualan di CV Hafas P2S2 adalah pendekatan data mining dengan fokus pada algoritma K-Means. Proses penelitian dapat dibagi menjadi beberapa langkah utama sebagai berikut:

### **2.1. Metode Pengumpulan Data**

Pada bagian ini merupakan suatu tata cara yang dilakukan untuk dapat mengumpulkan data terkait dengan penelitian yang sedang dilakukan[4]. Sedangkan metode pengumpulan data ini merupakan data yang benar dan sesuai dengan metode agar hasil yang diraih sesuai dengan tujuan penelitian awal atau hipotesis awal yang sudah ditentukan[5]. Data yang digunakan dalam analisis ini diambil dari sistem transaksi penjualan CV Hafas P2S2 selama periode tertentu. Data tersebut mencakup informasi tentang produk yang terjual, jumlah pembelian, waktu transaksi, dan informasi pelanggan. Data yang dikumpulkan kemudian dibersihkan untuk menghapus duplikasi dan kesalahan, sehingga hanya data yang relevan dan akurat yang digunakan dalam analisis.

### **2.2. Data Mining**

Data mining termasuk dalam ekstraksi informasi atau pola menarik dari data yang ada di database yang besar dalam jurnal ilmiah, data mining juga dikenal dengan nama *Knowledge Discovery in Database* (KDD)[6]. Data-data yang diolah dengan menggunakan teknik data mining juga mampu menghasilkan pengetahuan yang sesuai dengan harapan[7]. Proses dalam pembuatan pohon keputusan melibatkan perubahan data (tabel) menjadi sebuah model pohon, mengonversi model pohon tersebut menjadi aturan-aturan (rule), dan menyederhanakan aturan yang ada. Berikut adalah penjelasan dari proses KDD (Knowledge Discovery in Database) [8]:

### 2.2.1. Selection

Selection merupakan kegiatan untuk menyeleksi atau segmentasi data berdasarkan kriteria tertentu. Pada kegiatan ini, pengguna memilih data yang sesuai dari kesekian banyak data untuk diseleksi kembali pada tahap berikutnya.

### 2.2.2. Processing

Processing merupakan kegiatan pembersihan data, di mana atribut yang dianggap tidak mendukung dapat dibuang. Selain itu data dikonfigurasi ulang untuk menjamin format tetap konsisten.

### 2.2.3. Transformation

Transformation merupakan kegiatan yang mentransformasi data sehingga data dapat digunakan dan ditelusuri. Dengan kata lain, transformasi melakukan pemetaan data kompleks.

### 2.2.4. Interpretation and Evaluation

Interpretation and Evaluation merupakan proses interpretasi pola menjadi pengetahuan yang dapat digunakan untuk mendukung pengambilan keputusan.

Data Mining memiliki dua karakteristik utama, yaitu deskriptif dan prediktif. Sifat deskriptif dalam data mining berfungsi untuk menggambarkan keseluruhan data, yang memungkinkan pencarian pola yang paling mudah dipahami. Sebaliknya, data mining yang bersifat prediktif digunakan untuk melakukan peramalan dengan memanfaatkan model pengetahuan yang sudah ada.

## 2.3. Metode K-Means

K-Means adalah salah satu metode *clustering* yang paling populer dalam **unsupervised learning** (pembelajaran tanpa pengawasan) untuk mengelompokkan data ke dalam sejumlah kelompok (*cluster*). Metode ini bekerja dengan mempartisi data ke dalam beberapa klaster berdasarkan kedekatan jarak antar data.

## 2.4. Clustering

Pada tahapan ini akan dilakukan proses *clustering* dengan menentukan bobot- bobot nilai *cluster* dan yang nantinya akan dianalisis lebih lanjut untuk mendapatkan hasil yang sesuai. Tahap *clustering* ini menggunakan perangkat lunak *Rapidminer*[9].

Clustering merupakan upaya untuk mengelompokkan record, observasi, atau mengelompokkan ke dalam kelas yang memiliki kesamaan objek (Kusuma V. M., 2017). Pengklasteran berbeda dengan klasifikasi yang tidak adanya variable target dalam pengklasteran. Pengklasteran tidak digunakan untuk melakukan klasifikasi, mengestimasi, atau memprediksi nilai dari target. Pengklasteran digunakan untuk melakukan pembagian terhadap keseluruhan data menjadi kelompok – kelompok yang memiliki kemiripan.

## 2.5. Algoritma K-means

Metode K-means merupakan algoritma klusterisasi yang paling tua dan banyak digunakan dalam berbagai aplikasi kecil hingga menengah karena kemudahannya implementasinya. Ide dasar algoritma K-Means sangatlah sederhana, yaitu meminimalkan Sym of Squared Error (SSE) antara objek – objek data dengan empat Langkah.

## 2.6. Rapid Miner

RapidMiner merupakan perangkat lunak yang bersifat terbuka . RapidMiner menggunakan berbagai teknik deskriptif dan prediksi dalam memberikan wawasan kepada pengguna sehingga dapat membuat keputusan yang paling baik. Rapid Miner merupakan software yang berdiri sendiri untuk analisis data dan sebagai mesin data mining yang dapat di integrasikan pada produknya sendiri.

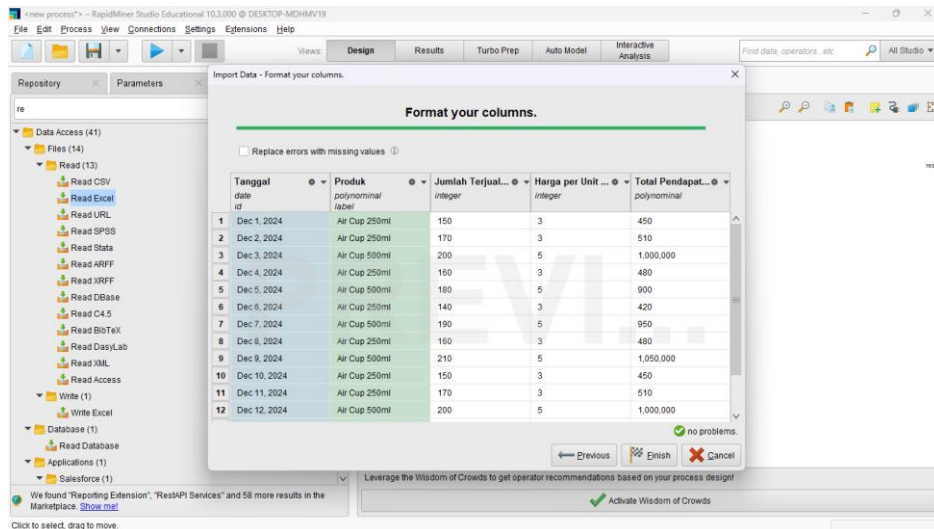
## 3. HASIL DAN PEMBAHASAN

Setelah menerapkan algoritma K-means pada data penjualan CV Hafas P2S2, sejumlah aturan asosiasi berhasil diidentifikasi. Peningkatan Strategi Pemasaran diharapkan pada CV Hafas P2S2 dapat memanfaatkan pola pembelian yang teridentifikasi untuk merancang strategi pemasaran yang lebih terarah dan efektif.

### 3.1. Alur Data Mining

#### 3.1.1. Data Set

Data yang digunakan pada penelitian adalah data pola penjualan pada bulan february 2024 sampai dengan April 2024 pada cv. havas. Data berikut adalah data *sheet* yang digunakan sebagai data penelitian. Berikut adalah datasheet dari pola penjualan pada cv. havas.

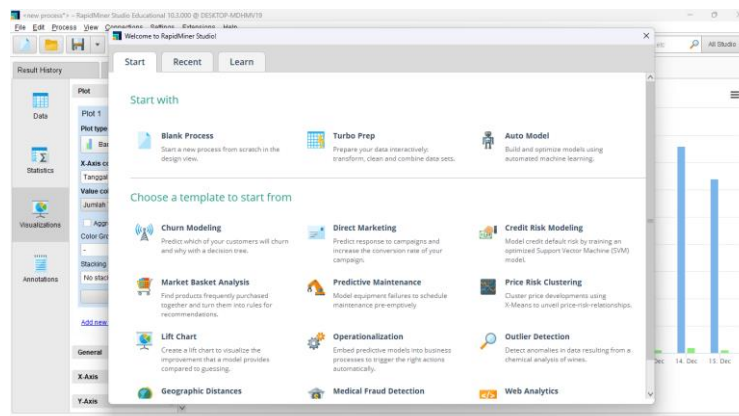


Tanggal	Produk	Jumlah Terjual	Harga per Unit	Total Pendapatan
1 Dec 1, 2024	Air Cup 250ml	150	3	450
2 Dec 2, 2024	Air Cup 250ml	170	3	510
3 Dec 3, 2024	Air Cup 500ml	200	5	1,000,000
4 Dec 4, 2024	Air Cup 250ml	160	3	480
5 Dec 5, 2024	Air Cup 500ml	180	5	900
6 Dec 6, 2024	Air Cup 250ml	140	3	420
7 Dec 7, 2024	Air Cup 500ml	190	5	950
8 Dec 8, 2024	Air Cup 250ml	160	3	480
9 Dec 9, 2024	Air Cup 500ml	210	5	1,050,000
10 Dec 10, 2024	Air Cup 250ml	150	3	450
11 Dec 11, 2024	Air Cup 250ml	170	3	510
12 Dec 12, 2024	Air Cup 500ml	200	5	1,000,000

Gambar 1. Data Set

#### 3.1.2. Implementasi Dan Pengujian

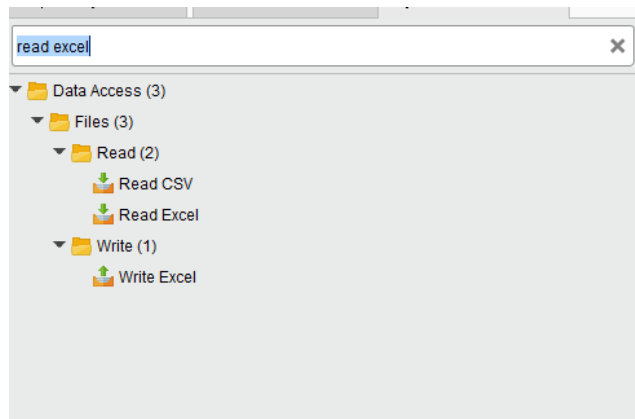
Dalam implementasi dan pengujian disini menggunakan sebuah *software* Rapidminer Studio versi 10.3, dengan pengujian data menggunakan *software* akan melakukan membandingkan bagaimana hasil pengolahan data secara manual dengan hasil pengolahan data menggunakan *software*. Gambar 2 merupakan tampilan halaman utama dari aplikasi *rapidminer* studio versi 10.3.



Gambar 2. Tampilan New Proses Rapidminer

#### 3.1.3. Read Excel

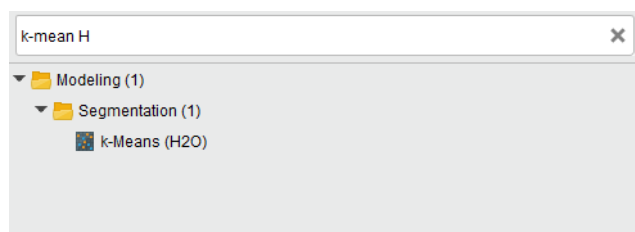
Selanjutnya masuk kedalam panel operator kemudian ketik Read Excel, lalu drag operator read excel agar operator tersebut muncul di panel proses dan bisa digunakan untuk memasukkan data excel yang telah melalui tahap data cleansing dan data transformation. Seperti yang terlihat pada gambar 3 berikut:



Gambar 3. Read Excel

3.1.4. K-Means

Langkah selanjutnya adalah memasukkan model yang akan digunakan untuk proses clustering data mining dengan cara masuk ke menu operator kemudian pilih K-Means seperti pada gambar 4 berikut:



Gambar 4. Memilih Pemodelan Clustering

3.1.5. Data View

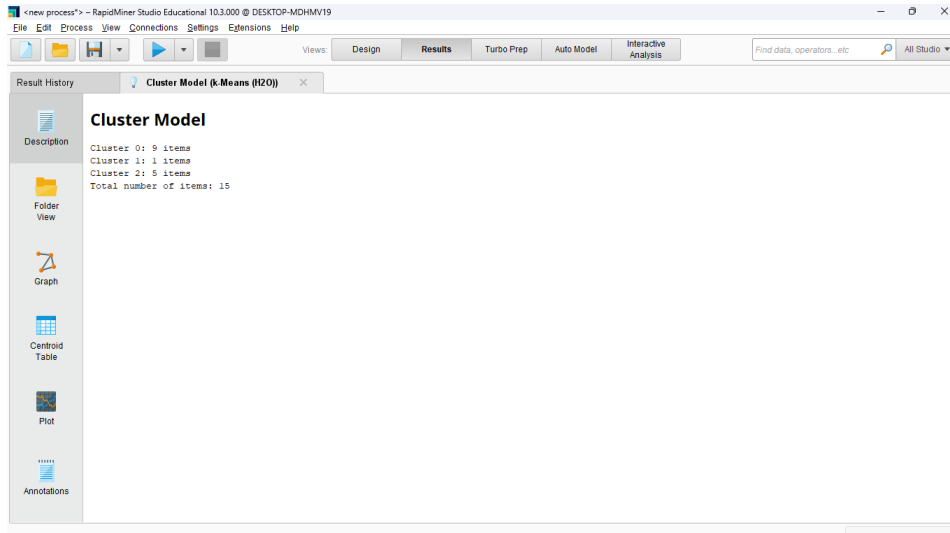
Pada gambar 5 adalah hasil *import* data yang berasal data data *excel* yang kemudian di proses sehingga menghasilkan data view hasil clusterisasi penjualan

Row No.	Tanggal	Produk	cluster	Jumlah Terj..	Harga per U..	Total Penda..
1	Dec 1, 2024	Air Cup 250ml	cluster_0	150	3	450
2	Dec 2, 2024	Air Cup 250ml	cluster_0	170	3	510
3	Dec 3, 2024	Air Cup 500ml	cluster_2	200	5	1,000,000
4	Dec 4, 2024	Air Cup 250ml	cluster_0	160	3	480
5	Dec 5, 2024	Air Cup 500ml	cluster_1	180	5	900
6	Dec 6, 2024	Air Cup 250ml	cluster_0	140	3	420
7	Dec 7, 2024	Air Cup 500ml	cluster_2	190	5	950
8	Dec 8, 2024	Air Cup 250ml	cluster_0	160	3	480
9	Dec 9, 2024	Air Cup 500ml	cluster_2	210	5	1,050,000
10	Dec 10, 2024	Air Cup 250ml	cluster_0	150	3	450
11	Dec 11, 2024	Air Cup 250ml	cluster_0	170	3	510
12	Dec 12, 2024	Air Cup 500ml	cluster_2	200	5	1,000,000
13	Dec 13, 2024	Air Cup 250ml	cluster_0	180	3	540
14	Dec 14, 2024	Air Cup 500ml	cluster_2	190	5	950
15	Dec 15, 2024	Air Cup 250ml	cluster_0	160	3	480

Gambar 5. Tampilan Hasil Cluster Pada Data View

3.1.6. Cluster Model

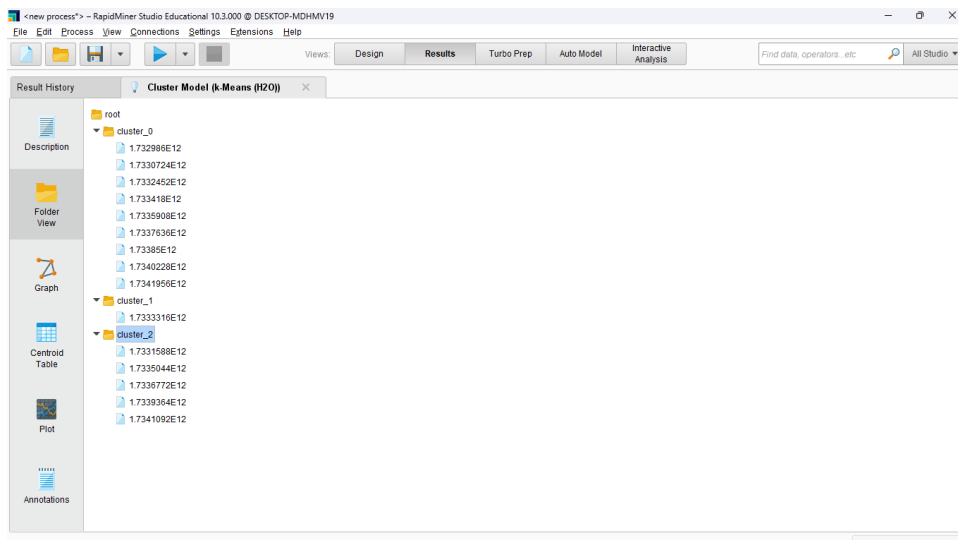
Tampilan hasil seperti gambar 6 berupa hasil pengujian dataset yang berjumlah 15 data menggunakan software rapidminer 10.3, sehingga berbentuk 3 cluster. Pada cluster 0 (cluster pertama) terdapat 9 data, cluster 1 (cluster kedua) terdapat 1 data dan cluster 2 (cluster ketiga) terdapat 5 data



**Gambar 1.** Cluster Model Hasil Data Mining

### 3.1.7. Folder View

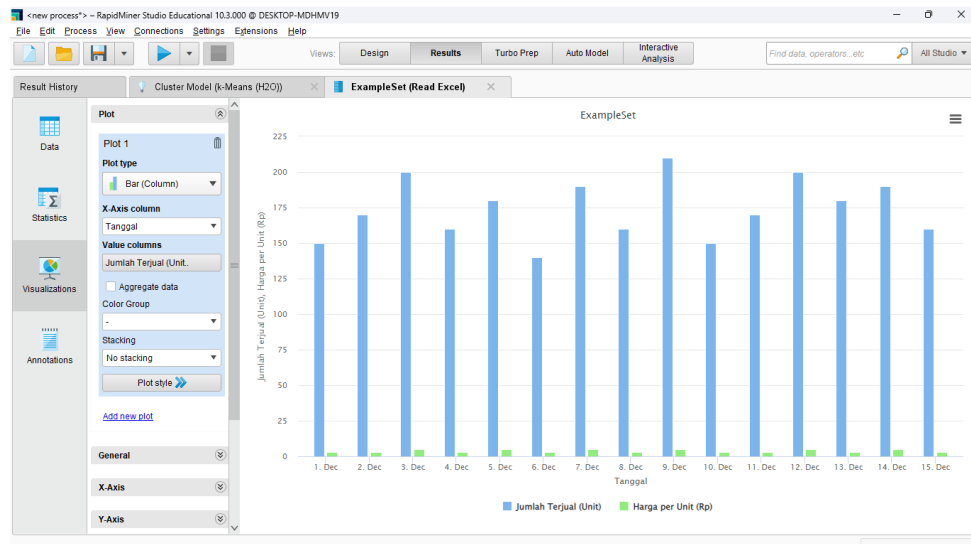
Gambar 7 di bawah ini merupakan ilustrasi tampilan folder view yang menampilkan analisis pola penjualan di cv. havas p2s2. Tampilan ini memvisualisasikan data penjualan berdasarkan kategori produk tertentu, yang bertujuan untuk membantu identifikasi pola penjualan produk terbaik, sebagaimana terlihat pada gambar 7 berikut:



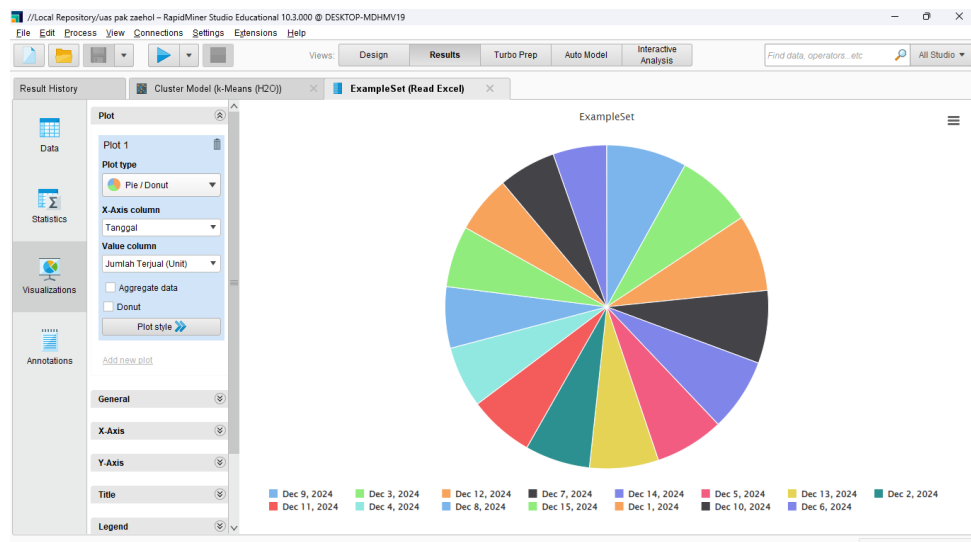
**Gambar 7.** Folder View

### 3.2 Visualization Hasil Clustering

Pada gambar 8 terlihat hasil visualialisasi data pola penjualan berdasarkan hasil total penjualan. Dengan atribut jumlah awal, terjual, harga jual dan total penjualan. Pemilihan fitur, pada tahap ini, dilakukan pemilihan fitur yang relevan untuk mendapatkan hasil *clustering* yang bermakna. Adapun fitur yang digunakan adalah parameter *multiplay*, metode *k-means* untuk menentukan berapa jumlah kluster yang akan digunakan untuk mendapatkan hasil yang paling tepat. Pada gambar 8. Model proses *clustering*, menggunakan parameter *k-means*



**Gambar 8.** Visualization data penjualan



**Gambar 2.** Tampilan Pie pada Chart

#### 4. KESIMPULAN DAN SARAN

Berdasarkan hasil implementasi algoritma K-means untuk menganalisis pola penjualan pada CV. Hafas P2S2, dapat disimpulkan bahwa metode ini efektif dalam menggali informasi terkait hubungan asosiatif antar produk yang sering terjual bersamaan. Analisis yang dilakukan berhasil mengidentifikasi kombinasi produk yang memiliki tingkat pembelian yang tinggi secara bersamaan, yang memberikan wawasan penting bagi perusahaan dalam hal pengelolaan stok dan strategi pemasaran.

Temuan utama dari penelitian ini adalah beberapa item produk yang memiliki pola pembelian yang kuat dan sering dibeli bersama oleh pelanggan. Informasi ini dapat dimanfaatkan oleh perusahaan untuk menyusun strategi penataan produk, seperti penempatan barang yang lebih strategis di rak atau menyusun paket promo berdasarkan kombinasi produk yang sering dibeli bersama. Selain itu, perusahaan juga dapat mengoptimalkan pengelolaan persediaan dengan lebih fokus pada produk yang memiliki tingkat permintaan tinggi dan sering dibeli bersamaan.

Implementasi algoritma K-means pada data penjualan CV. Hafas P2S2 memberikan kontribusi yang signifikan dalam mempercepat proses pengambilan keputusan berbasis data, terutama dalam merancang strategi pemasaran yang lebih tepat sasaran dan meningkatkan efisiensi operasional. Dengan memahami pola-pola penjualan yang terjadi, perusahaan dapat merespons lebih cepat terhadap kebutuhan pasar dan memperkuat daya saingnya.

**DAFTAR PUSTAKA**

- [1] Z. Fatah, J. S. Informasi, and U. Ibrahimy, “sistem informasi geografis pemetaan kantor dinas pemerintah kabupaten bondowoso berbasis web,” vol. 1, no. 2, pp. 342–349, 2024.
- [2] Nurul arimayanti, *APLIKASI PENJUALAN AIR MINERAL SYIFA P2S2 DI UD.ASSYARIF PONDOK PESANTREN SALAFIYAH SYAFI’IYAH SUKOREJO SITUBONDO BERBASIS WEB*. 2016.
- [3] Y. Syahra and E. F. Ginting, “Implementasi Data Mining Untuk Menganalisa Pola Penjualan Menggunakan Metode Apriori Pada CV. Putra Mineral Water,” *J. Cyber Tech*, 2019.
- [4] A. Yogiarto, A. Homaidi, and Z. Fatah, “Implementasi Metode K-Nearest Neighbors (KNN) untuk Klasifikasi Penyakit Jantung,” *G-Tech J. Teknol. Terap.*, vol. 8, no. 3, pp. 1720–1728, 2024, doi: 10.33379/gtech.v8i3.4495.
- [5] SYAFRIDA HAFNI SAHIR, “E-Book Metodologi Penelitian Syafrida”.
- [6] M. K. Amril Mutoi Siregar, S.Kom., M.Kom. DAN Adam Puspabhuana, S.Kom., “DATA MINING Pengolahan Data Menjadi Informasi dengan RapidMiner”.
- [7] M. A. dan M. Nasir, *Data Mining Algoritma dan Implementasi*.
- [8] M. S. Deny Jollyta, Prihandoko, Alyauma Hajjah, Elin Haerani, *Algoritma Klasifikasi untuk Pemula Solusi Python dan RapidMiner*. 2023.
- [9] M. Rafi Nahjan, Nono Heryana, and Apriade Voutama, “Implementasi Rapidminer Dengan Metode Clustering K-Means Untuk Analisa Penjualan Pada Toko Oj Cell,” *JATI (Jurnal Mhs. Tek. Inform.*, vol. 7, no. 1, pp. 101–104, 2023, doi: 10.36040/jati.v7i1.6094.