

ANALISIS POLA PENJUALAN BBM DI PT PERTAMINA RETAIL MENGUNAKAN METODE CLUSTERING K-MEANS DENGAN RAPIDMINER

Adam Nurhidayat, Agus Sulistyanto, Anton Sianipar Zulkarnain

Sistem Informasi, STMIK Jayakarta
Jalan Salemba Raya no 24 Jakarta, Indonesia
adamhiraga@gmail.com

ABSTRAK

Manajemen produk BBM subsidi dan non-subsidi di SPBU Pertamina Retail sangat krusial untuk menjaga efisiensi operasional dan kepuasan pelanggan. Permasalahan utama yang dihadapi adalah kesulitan dalam mengidentifikasi pola penjualan yang bervariasi antar wilayah, yang berisiko pada keakuratan stok karena tingginya penjualan. Penelitian ini bertujuan untuk mengelompokkan produk berdasarkan tingkat minat konsumen guna mengoptimalkan distribusi dan strategi stok. Metode yang digunakan adalah clustering K-Means dengan mengolah data historis berupa kode produk, jumlah transaksi, dan rata-rata penjualan menggunakan aplikasi RapidMiner. Hasil penelitian berhasil memetakan produk ke dalam tiga kelompok: stok banyak (paling diminati), sedang, dan sedikit. K-Means terbukti efektif mengidentifikasi karakteristik penjualan unik di setiap wilayah SPBU, baik yang didominasi BBM subsidi maupun non-subsidi. Temuan ini menjadi dasar bagi PT. Pertamina Retail dalam menyusun strategi distribusi kedepan agar lebih tepat sasaran; wilayah subsidi memerlukan pengawasan ketat, sementara wilayah non-subsidi potensial untuk pengembangan promosi. Dengan demikian, pengelolaan stok menjadi lebih efisien dan efektif.

Kata kunci : Penjualan BBM, K-Means clustering, Data Mining, Pola Analisis, Pt Pertamina Retail.

1. PENDAHULUAN

Penjualan bahan bakar bersubsidi (BBM) dan non-subsidi (BBK) harus terdistribusi secara tepat pada SPBU yang dikelola oleh PT Pertamina Retail. Produk BBM subsidi meliputi Peralite, Biosolar, dan Dexlite, sedangkan BBK non-subsidi terdiri dari Pertamina, Pertamina Green, Pertamina Turbo, dan Pertamina Dex. PT Pertamina Retail sebagai anak perusahaan PT Pertamina Patra Niaga berperan dalam pendistribusian BBM melalui SPBU COCO (Company Owned, Company Operated). Seiring meningkatnya kebutuhan energi, analisis pola penjualan BBM berdasarkan wilayah menjadi penting untuk mendukung strategi distribusi, manajemen stok, dan pengambilan keputusan manajerial.

Belum adanya analisis mendalam dengan metode Clustering K-Means dengan Aplikasi RapidMiner terhadap data penjualan BBM pada PT Pertamina Retail, Belum adanya identifikasi pola konsumsi BBM berdasarkan kategori subsidi dan non-subsidi pada produk produk yang dikeluarkan oleh Pertamina berdasarkan daerah mana saja yang paling banyak menggunakan salah satu produk tertentu untuk pengawasan distribusi dan pengembangan promosi, Distribusi dan pengelolaan stok BBM menuju SPBU masih menghadapi kendala akibat keterbatasan analisis data historis yang belum di clustering, Belum optimalnya pemanfaatan metode data mining dalam mendukung pengambilan keputusan terkait distribusi dan stok BBM Subsidi dan Non Subsidi.

Untuk mengatasi permasalahan tersebut, penulis merencanakan langkah-langkah penyelesaian sebagai berikut: Pengumpulan Data: Melakukan observasi dan wawancara langsung di PT. Pertamina Retail Graha Pertamina untuk mendapatkan data penjualan yang

akurat (periode tahun 2025). Preprocessing Data: Mengolah data historis yang mencakup kode produk, jumlah transaksi, dan rata-rata penjualan agar siap dianalisis. Analisis Data Mining: Mengimplementasikan algoritma K-Means Clustering menggunakan Altair AI Studio (RapidMiner). Proses ini melibatkan penentuan nilai k yang optimal serta penggunaan Euclidean Distance untuk mengukur kedekatan data. Evaluasi Model: Menguji hasil klasterisasi menggunakan operator Cluster Performance dengan parameter Davies-Bouldin Index (DBI) untuk menjamin validitas hasil pengelompokan. Interpretasi Hasil: Menganalisis karakteristik tiap klaster (stok banyak, sedang, sedikit) untuk menghasilkan wawasan manajerial mengenai distribusi yang tepat sasaran.

Berdasarkan latar belakang tersebut, penelitian ini berjudul “Analisis Pola Penjualan BBM di PT Pertamina Retail Menggunakan Metode Clustering K-Means dengan RapidMiner”, dengan tujuan memberikan gambaran pola penjualan BBM subsidi dan non-subsidi untuk mendukung optimalisasi strategi distribusi perusahaan.

2. TINJAUAN PUSTAKA

Penulis membuat penelitian terdahulu dengan tujuan untuk menjadi referensi agar mempermudah penulis dalam melakukan penelitian sehingga penelitian ini dapat diselesaikan dengan akurat dan efektif. Berikut merupakan hasil review jurnal yang telah penulis lakukan sebagai berikut:

Implementasi RapidMiner untuk clustering data penjualan pakaian menggunakan metode k-means. Oleh Idham, Herliana Rosika, Yuliadi Yuliadi, Permasalahan Sulit mengetahui pola penjualan

perusahaan tidak bisa langsung lihat produk mana yang paling diminati, mana yang kurang laku, atau musim tertentu Dengan Analisa menggunakan metode K-Means dan hasil Clustering berhasil membagi produk ke dalam beberapa kelompok, Misalnya Cluster 1 = produk.

Analisa cluster dengan k-means clustering untuk pengelompokan data cybercrime. Oleh Wulan Permata Sari¹, Tata Sutabri, Permasalahan Data cybercrime sulit dianalisis karena jumlahnya besar, jenisnya banyak, dan sifatnya kompleks. Metode K-Means clustering berhasil mengelompokkan data cybercrime menjadi beberapa cluster (contohnya: cluster serangan tinggi, cluster serangan sedang, cluster serangan rendah).

Implementasi algoritma k-means clustering pada sistem informasi geografis fasilitas kesehatan BPJS kesehatan kota Semarang. Oleh Bima Aditya Hendriansyah, Aris Tri Jaka Harjanta, Khoiriya Latifah, Permasalahan Tidak meratanya distribusi fasilitas kesehatan terhadap jumlah peserta BPJS, Metode K-Means dapat mengklasterisasi yang dilakukan berdasarkan kedekatan lokasi fasilitas dan jumlah pengguna

3. METODE PENELITIAN

Penelitian ini dilakukan di PT Pertamina Retail selama periode September–November 2025 dengan menggunakan pendekatan kuantitatif. Data yang digunakan merupakan data penjualan BBM subsidi dan non-subsidi yang diperoleh dari dashboard penjualan perusahaan serta didukung oleh observasi dan wawancara singkat dengan staf terkait.

Metode analisis yang digunakan adalah data mining dengan algoritma K-Means Clustering untuk mengelompokkan pola penjualan BBM berdasarkan wilayah dan jenis produk. Proses analisis mengikuti tahapan *Knowledge Discovery in Database* (KDD) yang meliputi pemilihan data, pembersihan data, transformasi data, proses clustering, serta interpretasi hasil. Dataset yang dianalisis berasal dari 15 SPBU dengan cakupan 11 wilayah/provinsi, menggunakan atribut utama berupa wilayah (region), jenis produk BBM, volume penjualan, volume stok, total penjualan, dan jumlah transaksi. Data nominal dikonversi ke bentuk numerik agar dapat diproses secara matematis. Proses clustering dilakukan menggunakan Altair AI Studio (RapidMiner) dengan menentukan jumlah cluster (k) dan menghitung jarak antar data menggunakan Euclidean Distance. Evaluasi kualitas hasil clustering dilakukan menggunakan **Davies-Bouldin Index** (DBI), di mana nilai DBI yang lebih kecil menunjukkan hasil pengelompokan yang lebih baik. Hasil clustering digunakan untuk mengidentifikasi pola konsumsi BBM di setiap wilayah sebagai dasar pendukung pengambilan keputusan distribusi dan pengelolaan stok BBM.

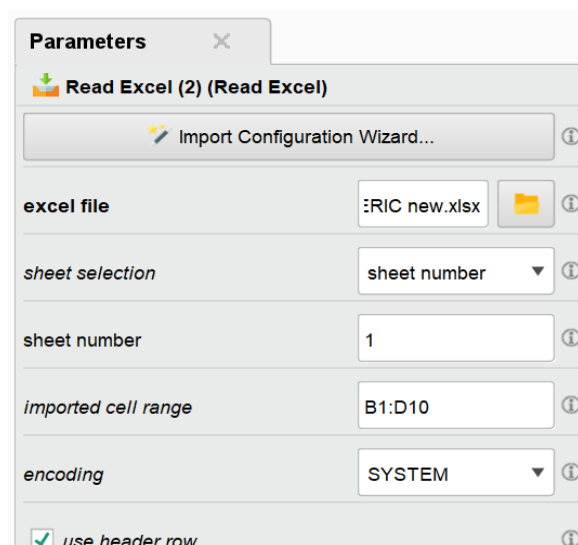
Proses pengumpulan data dilakukan dengan cara mencari referensi dari jurnal, buku, artikel maupun

informasi data yang berkaitan dengan penelitian yang dilakukan.

4. HASIL DAN PEMBAHASAN

Memuat Setelah data siap, pengguna dapat memulai proses klasterisasi dengan menyeret operator K-Means ke dalam process panel di RapidMiner. Dalam penerapan metode ini, pengguna perlu menentukan jumlah klaster (k) yang diinginkan berdasarkan tujuan analisis. Misalnya, apabila tujuan klasterisasi adalah untuk mengelompokkan Produk Subsidi atau Non Subsidi ke dalam 11 kelompok besar berdasarkan penjualan terbanyak pada 11 provinsi pada 15 SPBU, maka nilai k yang digunakan adalah 11. Proses ini dilakukan dengan menyeret operator seperti Read Excel untuk memuat data, Replace Missing Values untuk membersihkan data, Normalize untuk menyesuaikan skala data, dan Select Attributes untuk memilih atribut yang relevan. Setelah itu, operator K-Means digunakan untuk melakukan klasterisasi dan dilanjutkan dengan operator Cluster Model serta Scatter Plot untuk mengevaluasi dan memvisualisasikan hasil klasterisasi.

Hasil klasterisasi yang ditampilkan dalam bentuk scatter plot akan menunjukkan bagaimana setiap SPBU dikelompokkan berdasarkan kesamaan penjualan produk terbanyak. Setiap titik dalam visualisasi tersebut merepresentasikan satu SPBU, dan setiap warna atau simbol menunjukkan klaster tempat SPBU tersebut tergabung. Visualisasi ini sangat membantu dalam mengidentifikasi pola distribusi BBM dan penjualan diantara SPBU, sehingga PT. Pertamina Retail dapat dengan cepat mengenali kelompok-kelompok yang memiliki potensi atau kecenderungan serupa. Data tersebut diolah menggunakan Algoritma K-Means clustering dengan operator yang sudah peneliti rancang sebagai berikut,



Gambar 1. Operator Read Excel
(Sumber : Peneliti)

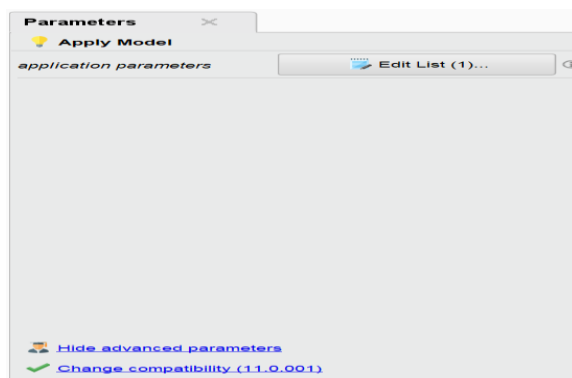
Menunjukkan pengaturan parameter pada mode Read Excel K-MEANS Analytics Platform. File yang

digunakan bernama "DATASET_SPBU_NUMERIC_NEW.XLX" dengan metode pemilihan sheet berdasarkan nomor, yaitu sheet nomor 1. Rentang sel yang diimpor adalah dari B1 hingga D10.



Gambar 2. Operator Clustering K-Means (Sumber : Peneliti)

Konfigurasi pada gambar K:11 menunjukkan bahwa pengguna ingin membagi data ke dalam 11 kelompok yaitu berdasarkan jumlah Provinsi dimasing-masing SPBU menggunakan algoritma K-Means.



Gambar 3. Operator Apply Model (Sumber : Peneliti)

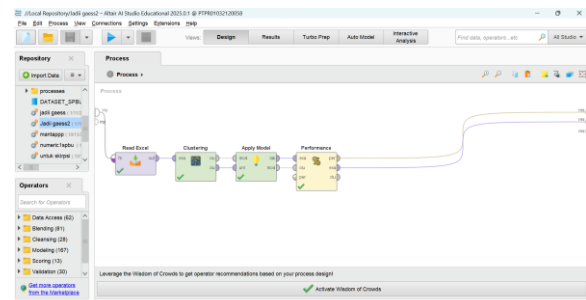
Operator Apply Model dalam RapidMiner digunakan untuk Menggunakan model hasil training algoritma K-Means menerapkan model yang sudah dilatih (trained model) ke data baru.



Gambar 4. Konfigurasi Parameter Pada Operator (Sumber : Peneliti)

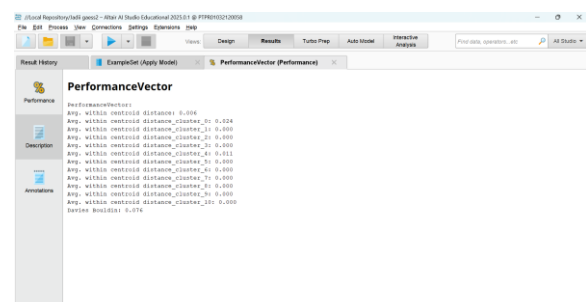
Performance (Cluster Distance Performance) diRapidMiner. Pada bagian main criterion, pengguna memilih metrik Avg. within centroid distance, yang

berfungsi untuk menghitung rata-rata jarak data terhadap pusat klasternya. Desain Clustering yang sudah peneliti analisa dengan data Penjualan BBM Penerapan proses algoritma K-Means menggunakan aplikasi RapidMiner studio yang berguna dalam membantu melakukan kegiatan penelitian ini berikut Desain data mining, Klustering yang sudah peneliti buat.



Gambar 5. Input proses Clustering Rapidminer Pola Penjualan BBM (Sumber : Peneliti)

Setelah data diinput, maka hasil yang didapat dari software Rapidminer dapat dilihat sebagai berikut,



Gambar 6. Output Performance Davies Bouldin Pola Penjualan BBM (Sumber : Peneliti)

Untuk mengukur performa pengelompokan, penelitian ini menggunakan Davies-Bouldin Index (DBI) melalui perangkat lunak RapidMiner. Prinsip utama DBI adalah semakin kecil nilai indeks yang dihasilkan, maka semakin baik kualitas kluster tersebut (memiliki kepadatan yang tinggi dan pemisahan antar kluster yang jelas). Data dan Parameter Penelitian dengan sampel data penjualan dari 15 SPBU dengan performa tertinggi di 11 provinsi (Periode: Januari – Oktober 2025). Sumber Data berasal dari Dashboard Sales Operation PT. Pertamina Retail yang diekstrak ke dalam format Excel. Atribut Analisis berfokus pada komponen Subsidi dan Non-Subsidi dengan rasio jumlah transaksi (*Trans Count*) terhadap rata-rata volume penjualan. Penggunaan DBI memastikan bahwa hasil pemetaan stok BBM didasarkan pada perhitungan matematis yang objektif, sehingga strategi distribusi yang dihasilkan lebih akurat. Untuk dianalisa data penjualannya dengan rasio Trans Count dengan rata rata penjualan yang tertera pada perbandingan nilai transaksi/Trans Count yang berada didashboard penjualan Pertamina Retail yang

sudah dijadikan summary dengan diekstrak menjadi file Excel dengan capture Sebagai berikut :

Period	No	ME	Product	Grp	Vol	Ord	Own	Total	Subsidi	Non-Subsidi	Total	Subsidi	Non-Subsidi	Total
2025-01-01	6	5	BIO SOLAR	BBM	1524.64	0	0	1524.64	23833	10387460.00	.00	.00	10387460.00	44
2025-01-01	6	5	DEULITE	BBM	192.62	0	0	192.62	27833	2619600.00	.00	.00	2619600.00	8
2025-01-01	6	5	PERTALITE	BBM	3963.18	0	0	3963.18	22537	39631710.00	.00	.00	39631710.00	972
2025-01-01	6	5	PERTAMAX	BBM	3403.15	0	0	3403.15	11875	42539530.00	.00	.00	42539530.00	212
2025-01-01	6	5	PERTAMAX TURBO	BBM	231.11	0	0	231.11	18293	3186140.00	.00	.00	3186140.00	10
2025-01-01	6	5	PERTAMAX DES	BBM	273.64	0	0	273.64	635	3789660.00	.00	.00	3789660.00	8
2025-01-02	6	5	BIO SOLAR	BBM	1826.66	0	0	1826.66	28141	12421280.00	.00	.00	12421280.00	57
2025-01-02	6	5	DEULITE	BBM	246.92	0	0	246.92	27951	3357990.00	.00	.00	3357990.00	9
2025-01-02	6	5	PERTALITE	BBM	5417.39	0	0	5417.39	17479	54174010.00	.00	.00	54174010.00	936
2025-01-02	6	5	PERTAMAX	BBM	3014.43	0	0	3014.43	19873	37869020.00	.00	.00	37869020.00	231
2025-01-02	6	5	PERTAMAX TURBO	BBM	461.44	0	0	461.44	17679	6135750.00	.00	.00	6135750.00	22
2025-01-02	6	5	PERTAMAX DES	BBM	225.67	0	0	225.67	420	3138840.00	.00	.00	3138840.00	7
2025-01-03	6	5	BIO SOLAR	BBM	2018.06	0	0	2018.06	26086	13734630.00	.00	.00	13734630.00	68
2025-01-03	6	5	DEULITE	BBM	388.08	0	0	388.08	27187	5277890.00	.00	.00	5277890.00	12
2025-01-03	6	5	PERTALITE	BBM	6115.74	0	0	6115.74	19348	61157330.00	.00	.00	61157330.00	995
2025-01-03	6	5	PERTAMAX	BBM	3359.04	0	0	3359.04	13044	44887970.00	.00	.00	44887970.00	231
2025-01-03	6	5	PERTAMAX TURBO	BBM	613.49	0	0	613.49	17208	8404880.00	.00	.00	8404880.00	23
2025-01-03	6	5	PERTAMAX DES	BBM	295.89	0	0	295.89	148	4121440.00	.00	.00	4121440.00	10
2025-01-04	6	5	BIO SOLAR	BBM	1655.95	0	0	1655.95	24422	11260380.00	.00	.00	11260380.00	55

Gambar 7. Capture Excel Penjualan SPBU Coco Pertamina Retail (Sumber : Peneliti)

Melalui Nilai transaksi peneliti dapat membandingkan nilai transaksi berdasarkan indikasi pola konsumsi dengan transaction Count, bukan volume liter, Peneliti menghitung Persentase konsumsi produk subsidi dan non-subsidi dihitung berdasarkan hasil kompilasi dari Transaction Count pada masing-masing produk,

Trans count menggambarkan nilai pembelian rata-rata pertransaksi, sehingga dapat digunakan untuk mengukur proporsi konsumsi antara produk subsidi dan non-subsidi. Pemilihan Transaction Count dilakukan karena metrik ini menggambarkan jumlah transaksi aktual yang terjadi, sehingga dapat menampilkan frekuensi konsumsi setiap jenis produk. Berbeda dengan liter penjualan yang dapat dipengaruhi oleh kapasitas tangki kendaraan, Transaction Count dinilai lebih mewakili perilaku dan preferensi konsumen dalam memilih produk subsidi maupun non-subsidi. Metode Perhitungan Peneliti melakukan kompilasi seluruh Trans Count berdasarkan kelompok produk : Kelompok Subsidi, Pertalite, Biosolar. Kelompok Non-Subsidi, Pertamina, Pertamax Turbo, Dexlite. Setelah seluruh Trans Count dijumlahkan berdasarkan kelompok produknya, langkah berikutnya adalah menghitung persentase proporsi konsumsi menggunakan rumus sederhana : Persentase Subsidi $\approx$ (Total Transaksi Subsidi / Total Seluruh Transaksi) $\times$ 100% $\approx$ Persentase NonSubsidi $\approx$ (Total Transaksi Non-Subsidi / Total Seluruh Transaksi) $\times$ 100% $\approx$ SPBU yang akan diteliti dengan

penjualan terbesar terdiri dari beberapa area yaitu area Banda Aceh 1 SPBU, area Riau 1 SPBU, area Sumatera Selatan 1 SPBU, area Bandar lampung 1 SPBU, area Banten 1 SPBU, area Jakarta 1 SPBU, area Jawa Barat 3 SPBU, area Jawa tengah 1 SPBU, area Jawa Timur 1 SPBU, area Kalimantan Timur 2 SPBU, area Sulawesi Selatan 2 SPBU total 15 SPBU untuk dianalisa dari total keseluruhan 205 SPBU.

Data yang digunakan pada penelitian ini adalah data penjualan dari export dashboard penjualan tahun 2025 yang didapat dari tim ICT Pertamina Retail sebagai pengelola dashboard. Data digunakan untuk mengetahui dalam mengidentifikasi pola konsumsi BBM berdasarkan kategori subsidi dan non-subsidi pada produk produk yang dikeluarkan oleh Pertamina berdasarkan daerah mana saja yang paling banyak menggunakan salah satu produk tertentu untuk ketahanan stok. Terdapat data mentah yang digunakan berjumlah : 14.571.200 data penjualan selama dari bulan Januari sampai bulan Oktober tahun 2025 dari jumlah 15 SPBU dengan penjualan terbesar pada 11 area provinsi yang diklasterisasi dengan 3 variabel yang ada didalam jenis yang digunakanya BBM Subsidi dan Non Subsidi. Pada tahap ini, Peneliti melakukan proses clustering terhadap data penjualan BBM Subsidi dan Non Subsidi dari 15 SPBU pada 11 Provinsi yang menjadi sampel penelitian. Tujuan dari clustering adalah untuk mengelompokkan SPBU berdasarkan kesamaan pola penjualan sehingga dapat memberikan informasi yang lebih terarah terkait strategi distribusi, monitoring penjualan, serta perencanaan stok BBM. Proses clustering dilakukan menggunakan metode K-Means dengan jumlah cluster yang ditetapkan sebanyak sebelas kelompok (Cluster 1, Cluster 2, dan Cluster 3, Cluster 4, Cluster 5, Cluster 6, Cluster 7, Cluster 8, Cluster 9, Cluster 10, Cluster 11). Penentuan jumlah cluster ini diperoleh berdasarkan perhitungan metode Davies-Bouldin Index yang menunjukkan Cluster optimal adalah yang menghasilkan nilai Davies Bouldin terendah pada nilai $k = 11$.

Tabel 1. Centroid Cluster BBM Subsidi dan Non Subsidi

Attribute	Clust er_0	Clust er_1	Clust er_2	Clust er_3	Clust er_4	Clust er_5	Clust er_6	Clust er_7	Clust er_8	Clust er_9	Cluste r_10
REGION_SPBU_ENCODED	7	11	1	4	10	9	5	6	8	2	3
SUBSIDI	0.760	0.755	0.670	0.740	0.415	0.760	0	0	0.740	0.690	0.640
NON_SUBSIDI	0.240	0.245	0.330	0.260	0.585	0.230	1	1	0.350	0.310	0.360

Pada tabel ini menunjukkan nilai rata-rata (centroid) dari tiap variabel dalam 11 cluster. (1) Setiap kolom (cluster_0, cluster_1, dst.) mewakili kelompok SPBU yang punya karakteristik mirip. (2) Baris REGION_SPBU_ENCODED menunjukkan wilayah

dominan dicluster itu (hasil rata-rata atau nilai kode area). (3) Baris SUBSIDI dan NON_SUBSIDI menunjukkan proporsi atau dominasi BBM jenis tersebut didalam cluster.

Tabel 2. Status Percluster

Cluster	Wilayah dominan	BBM Subsidi	BBM Non-Subsidi	Interprestasi
cluster_0	7	0.760	0.240	Wilayah 7, sangat dominan subsidi tapi masih ada penjualan non-subsidi
cluster_1	3	0.755	0.245	Wilayah 11, fokus pada subsidi
cluster_2	11	0.670	0.330	Wilayah 1, masih dominan subsidi
cluster_3	9	0.740	0.260	Wilayah 4, dominan subsidi tapi masih ada non-subsidi
cluster_4	5	0.415	0.585	Wilayah 10, penjualan terbesar non-subsidi
cluster_5	6	0.760	0.230	Wilayah 9, penjualan terbesar subsidi
cluster_6	8	0	1	Wilayah 5, lebih banyak non subsidi
cluster_7	2	0	1	Wilayah 6, dominan subsidi
cluster_8	1	0.740	0.350	Wilayah 8, dominan subsidi
cluster_9	4	0.690	0.310	Wilayah 2, lebih banyak subsidi
cluster_10	10	0.640	0.360	Wilayah 3, dominan subsidi

Pada rincian status per cluster pada tabel diatas (cluster 0 hingga cluster 10), terlihat pola penjualan BBM Subsidi dan Non-Subsidi diberbagai wilayah. Dari 11 wilayah (1–11), mayoritas SPBU menunjukkan penjualan BBM subsidi yang lebih dominan, yaitu pada 8 dari 11 cluster. Sementara itu, wilayah pada cluster 4 serta cluster 6 dan 7 memperlihatkan porsi penjualan non-subsidi yang cukup besar. Adapun wilayah pada cluster 2, cluster 8, cluster 9, dan cluster 10 memiliki tingkat penjualan yang relatif seimbang antara subsidi dan non-subsidi, namun tetap menunjukkan kecenderungan lebih tinggi pada penjualan BBM subsidi.

Gambar 10. Proses Klasterisasi (Sumber : Peneliti)

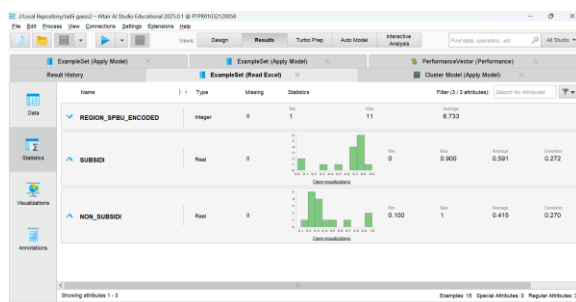
Menampilkan Hasil Dari Proses Klasterisasi menggunakan algoritma K Means di RapidMiner, yang ditampilkan pada jendela Cluster Model (Clustering). Hasil tersebut menunjukkan bahwa data berhasil dikelompokkan menjadi 11 klaster, yaitu Cluster 0 dengan 3 item, Cluster 1 dengan 2 item, dan Cluster 2 dengan 1 item, Cluster 3 dengan 1 item, Cluster 4 dengan 2 item, dan Cluster 5 dengan 1 item, Cluster 6 dengan 1 item, Cluster 7 dengan 1 item, dan Cluster 8 dengan 1 item, Cluster 9 dengan 1 item, Cluster 10 dengan 1 item. Secara keseluruhan terdapat 15 item yang dikelompokkan. Distribusi ini menunjukkan bahwa mayoritas data terdapat dalam Cluster 0, Sedangkan klaster lainnya memiliki jumlah item lebih sedikit. Informasi ini berguna untuk memahami penyebaran data dalam masing-masing kelompok dan karakteristik dominan pada tiap klaster.

Gambar 11. ExampleSet (1-13) (Sumber : Peneliti)

Gambar 12. ExampleSet (3-15) (Sumber : Peneliti)

ExampleSet (15 examples, 2 special attributes, 3 regular attributes) menunjukkan hasil klasterisasi pada tab ExampleSet di RapidMiner. Data terdiri dari 15 examples berarti ada 15 baris data (15 SPBU), 2 special attributes atribut spesial seperti id dan cluster (Menunjukkan tiap data masuk ke cluster mana bukan variabel yang digunakan untuk perhitungan, tapi sebagai identitas dan hasil pengelompokan (output clustering), 3 regular attributes atribut numerik yang digunakan untuk proses clustering, yaitu, REGION_SPBU_ENCODED - Kode wilayah SPBU (numerik hasil encoding) 1,2, 7, 5, 6, 11, dst, SUBSIDI - Proporsi BBM subsidi, Contoh: 0.9 berarti 90% penjualan subsidi, NON_SUBSIDI - Proporsi BBM non-subsidi, Contoh: 0.2 berarti 20% penjualan non-subsidi. Yang masing-masing memiliki atribut id, cluster, REGION_SPBU_ENCODED, SUBSIDI, dan NON_SUBSIDI tiga diantaranya merupakan atribut numerik. Kolom cluster menampilkan hasil pengelompokan data ke dalam 11 klaster, yaitu cluster_0 hingga cluster_10. Distribusi Data Tiap Cluster (berdasarkan tampilan): cluster_0 : 3 data (id

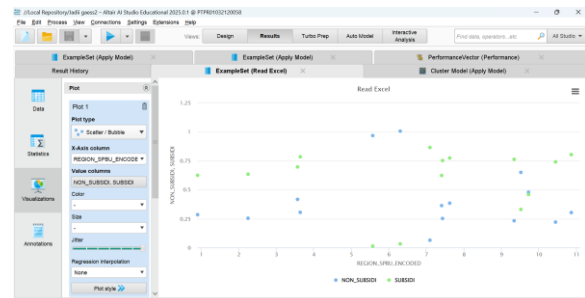
1, 4, 5), cluster_1 : 2 data (id 8,15), cluster_2 : 1 data (id 10), cluster_3 : 1 data (id 11), cluster_4 : 2 data (id 13,14), cluster_5 : 1 data (id 7), cluster_6 : 1 data (id 2), cluster_7 : 1 data (id 3), cluster_8 : 1 data (id 6), cluster_9 : 1 data (id 9), cluster_10 : 1 data (id 12). Lalu berdasarkan hasilnya, mayoritas data masuk ke cluster_0 dan cluster_4 sedangkan Cluster lain umumnya hanya berisi 1 data (SPBU) setelah diklaster. Tampilan ini memberikan gambaran yang jelas tentang bagaimana setiap SPBU dikelompokkan berdasarkan kemiripan nilai atribut, sehingga mempermudah analisis karakteristik tiap klaster secara visual (cluster grouping) dan numerik (nilai variable SPBU).



Gambar 12. Statistic ExampleSet (Sumber : Peneliti)

Menunjukkan tampilan tab Statistics ExampleSet di RapidMiner yang menampilkan deskriptif untuk tiga atribut reguler yang dipakai dalam proses klasterisasi: REGION_SPBU_ENCODED (integer, Min=1, Max=11, Rata-rata (Average / Mean)=6.733, Missing: 0 (tidak ada nilai hilang, tidak ada deviasi untuk integer ini, tetapi nilai Min/Max/Average cukup untuk melihat rentang dan pusat, Interpretasi; ini adalah kode wilayah hasil encoding; nilai 1–11 menunjukkan ada 11 wilayah, dan rata-rata 6.7 menunjukkan distribusi lokasi agar condong ke wilayah berkode menengah/atas. SUBSIDY (real/numerik kontinu, Min=0.000, Max=0.900, Rata-rata (Average / Mean)=0.591, Missing: 0, Deviasi (Standard Deviation): StdDev=0.272). Interpretasi sebagian besar nilai SUBSIDY tinggi (rata-rata 0.591) dari histogram terlihat banyak nilai mendekati 0.9. Ada nilai 0 pada beberapa contoh (artinya beberapa SPBU tidak menjual subsidi sama sekali), dan NON_SUBSIDY (real/numerik kontinu, Min=0.100, Max=1.000, Rata-rata (Average / Mean)=0.415, Missing: 0, Deviasi (Standard Deviation): StdDev=0.270). Interpretasi rata-rata non-subsidi relatif rendah (0.415), banyak nilai terletak dikisaran rendah (lihat histogram). Namun ada beberapa SPBU dengan non-subsidi tinggi (maks 1.0), menandakan variabilitas diantara SPBU. Semua atribut tidak memiliki nilai yang hilang. Informasi ini berguna untuk memahami rentang dan sebaran nilai sebelum menginterpretasikan karakteristik tiap klaster. Terlihat keterangan dibawah bahwa dataset berisi 15 contoh (Examples: 15), 0 special attributes pada tampilan ini namun ditampilkan ExampleSet sebelumnya id dan

cluster muncul sebagai special attributes ; tab Statistics ExampleSet hanya menampilkan regular attributes yang dipakai untuk clustering, dan 3 regular attributes yang dianalisis secara deskriptif. Semua atribut tidak memiliki nilai yang hilang (Missing = 0).



Gambar 14. Scatter Plot (Sumber : Peneliti)

Visualisasi *scatter plot* pada RapidMiner menunjukkan distribusi penjualan berdasarkan atribut REGION_SPBU_ENCODED (sumbu X) serta volume SUBSIDY dan NON_SUBSIDY (sumbu Y). Hasilnya memetakan karakteristik 15 SPBU di 11 wilayah ke dalam dua pola utama: (1) Dominasi BBM Subsidi (Titik Hijau): Terkonsentrasi di bagian atas grafik (nilai Y 0.8–1.2). Pola ini mendominasi wilayah 1, 2, 3, 4, 7, 8, dan 9. Hal ini menunjukkan volume penjualan subsidi yang tinggi dan merata di hampir seluruh region, sehingga memerlukan prioritas stok distribusi yang lebih besar. (2) Dominasi BBM Non-Subsidi (Titik Biru): Terkonsentrasi di area bawah (nilai Y 0.1–0.4). Penjualan tertinggi untuk kategori ini ditemukan secara spesifik pada wilayah 5 dan 6. Meskipun wilayah lain memiliki transaksi non-subsidi, volumenya tetap relatif lebih kecil dibandingkan konsumsi subsidi.

Visualisasi ini membantu memahami sebaran data dan kekompakan masing-masing klaster secara intuitif dalam dua dimensi. Untuk Pendistribusian stok BBM pada masing masing area dengan penjualan Subsidi dan non-Subsidi tertinggi.

5. KESIMPULAN DAN SARAN

Berdasarkan hasil penelitian, dapat disimpulkan bahwa setiap wilayah SPBU PT Pertamina Retail memiliki pola penjualan BBM yang berbeda antara BBM subsidi dan non-subsidi. Sebagian besar wilayah didominasi oleh penjualan BBM subsidi, sementara penjualan BBM non-subsidi hanya menonjol pada wilayah tertentu, khususnya wilayah 5 dan 6. Penerapan metode K-Means Clustering menggunakan RapidMiner berhasil mengelompokkan SPBU berdasarkan karakteristik penjualan, sehingga menunjukkan bahwa strategi distribusi BBM tidak dapat diseragamkan antarwilayah. Hasil klasterisasi memberikan dasar yang kuat bagi PT Pertamina Retail dalam mengoptimalkan alokasi stok BBM. Wilayah dengan ketergantungan tinggi terhadap BBM subsidi memerlukan pengawasan distribusi yang lebih ketat, sedangkan wilayah dengan potensi BBM non-subsidi

dapat dijadikan target pengembangan produk. Secara keseluruhan, penelitian ini membuktikan bahwa analisis data berbasis clustering dapat mendukung pengambilan keputusan manajerial dan meningkatkan efisiensi perencanaan distribusi BBM.

Apabila penelitian ini dilanjutkan, adapun saran yang dibuat oleh peneliti dengan tujuan agar penelitian bersifat menyeluruh diantaranya sebagai berikut, (1) Sebaiknya dapat diteliti semua SPBU Pertamina Retail pada Seluruh Provinsi. (2) Sebaiknya mencoba memakai Segmentation Operator X-Means / Auto Clustering versi otomatis K-Means yang bisa menentukan jumlah kluster terbaik. (3) Sebaiknya memakai fitur Altair AI Studio yang berbayar karena lebih banyak fitur.

DAFTAR PUSTAKA

- [1] I. Idham, H. Rosika, and Y. Yuliadi, "Implementasi RapidMiner untuk clustering transaksi penjualan pakaian menggunakan metode K-Means," *JUTECH: Jurnal Teknologi Informasi*, vol. 5, no. 2, pp. 45–56, 2023, <https://jurnal.stkipppersada.ac.id/jurnal/index.php/jutech/article/view/3642>.
- [2] PT Pertamina Retail, "Profil perusahaan," 2023. [Online]. Available: <https://www.pertaminaretail.com>
- [3] JF Andry, D. Henny Hartono, Honni, Denny and Jeffrey Jo, "Analysis and prediction supermarket sales with data mining using RapidMiner," *AIP Conference Proceedings*, vol. 2693, no. 1, p. 020004, 2023. <https://doi.org/10.1063/5.0118725>.
- [4] J. Han, M. Kamber, and J. Pei, *Data Mining: Concepts and Techniques*, 4th ed. San Francisco, CA, USA: Morgan Kaufmann, 2022.
- [5] IBM, "What is data mining?" *IBM Think*, 2023. [Online]. Available: <https://www.ibm.com/think/topics/data-mining>
- [6] A. Sadri, U. Kose, and O. Deperlioglu, "Characteristics of data mining technology," *ResearchGate*, 2023. [Online]. Available: <https://www.researchgate.net>
- [7] L. Soibelman, H Kim, "Data Preparation Process for Construction Knowledge Generation through Knowledge Discovery in Databases," *Journal of Computing in Civil Engineering*, vol. 16, no. 1, 2022.
- [8] A. Wulandari, Irmayansyah, and L. T. Ningrum, "Penerapan metode K-Means clustering untuk pemetaan perbaikan jalan di Kota Bogor," *JINTEKS: Jurnal Informatika, Teknologi dan Sains*, vol. 7, no. 1, 2025. <https://jurnal.uts.ac.id/index.php/JINTEKS/article/view/5323>.
- [9] W. P. Sari and T. Sutabri, "Analisa cluster dengan K-Means clustering untuk pengelompokan data cybercrime," *JINTEKS: Jurnal Informatika, Teknologi dan Sains*, vol. 5, no. 1, 2023. <https://eprints.binadarma.ac.id/17444/>.
- [10] E. Kurniawan and N. Hendrastuty, "Penerapan algoritma K-Means untuk melakukan klusterisasi pada peringkasan teks," *JINTEKS: Jurnal Informatika, Teknologi dan Sains*, 2023. <https://jurnal.uts.ac.id/index.php/JINTEKS/article/view/4435>.
- [11] B. A. Hendriansyah, A. T. J. Harjanta, and K. Latifah, "Implementasi algoritma K-Means clustering pada sistem informasi geografis fasilitas kesehatan BPJS Kesehatan Kota Semarang," *JINTEKS: Jurnal Informatika, Teknologi dan Sains*, 2023. <https://jurnal.uts.ac.id/index.php/JINTEKS/article/view/5618>.
- [12] Altair, "Launch of Altair RapidMiner 2023 platform delivers next-generation generative AI capabilities," *Altair Newsroom*, Nov. 28, 2023. [Online]. Available: <https://altair.com/news/rapidminer-2023-launch>
- [13] M. F. Mustapha, "The prediction of student's academic performance using RapidMiner," *Indonesian Journal of Electrical Engineering and Computer Science*, vol. 32, no. 1, pp. 363–371, 2023, doi: 10.11591/ijeecs.v32.i1.pp363-371. https://d1wqtxts1xzle7.cloudfront.net/10648168/1/26178_67455_1_PB-libre.pdf.
- [14] RapidMiner Documentation, "What's new in RapidMiner Studio 9.10," 2022. [Online]. Available: <https://docs.rapidminer.com/latest/studio/>
- [15] RapidMiner, "RapidMiner repositories overview," *GitHub*, 2023. [Online]. Available: <https://github.com/rapidminer>
- [16] I. F. Ashari *et al.*, "Application of data mining with the K-Means clustering method and Davies–Bouldin index for grouping IMDB movies," *JAIC*, vol. 6, no. 1, pp. 7–15, Jul. 2022. <https://jurnal.polibatam.ac.id/index.php/JAIC/article/view/3485>.
- [17] D. L. Davies and D. W. Bouldin, "A cluster separation measure," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 1, no. 2, pp. 224–227, 1979, doi: 10.1109/TPAMI.1979.4766909.
- [18] Altair Engineering, "Altair AI Studio," 2024. [Online]. Available: <https://altair.com/altair-ai-studio>