

Penentuan Menu Potensial Menggunakan Algoritma K-Means Pada Titiksetara Coffee Bekasi

¹Mohammad Alviansyah, dan ²Prima Dina Atika

Informatika, Ilmu Komputer, Universitas Bhayangkara Jakarta Raya

Email : ¹mohammad.alviansyah19@mhs.ubharajaya.ac.id, ²prima.dina@dsn.ubharajaya.ac.id

ABSTRAKSI

Penelitian ini bertujuan untuk mengidentifikasi pelanggan potensial, membantu pelaku usaha kopi lokal bersaing dengan menyediakan produk yang diinginkan pelanggan dan mengatasi masalah dalam menentukan menu potensial menggunakan Algoritma K-Means. Data yang digunakan merupakan data pendapatan TitikSetara Coffee dari Oktober 2023 hingga Januari 2024. Dalam penelitian ini digunakan software RapidMiner untuk melakukan pengelompokan menggunakan algoritma K-Means Clustering. Data penjualan dikelompokkan ke dalam tiga cluster, dimana cluster 0 untuk menu dengan potensi tinggi, cluster 1 untuk menu dengan potensi menengah, dan cluster 2 untuk menu dengan potensi rendah. Dari 26 data uji yang digunakan hasil menunjukkan cluster pertama terdiri dari 4 menu, cluster kedua terdiri dari 14 menu, dan cluster ketiga terdiri dari 8 menu dengan nilai Davies-Bouldin Index sebesar 0,374. Menu potensial yang dihasilkan dapat dijadikan referensi oleh pihak TitikSetara Coffee dan diharapkan menu potensial dari hasil penelitian ini dapat benar benar tepat sasaran.

Kata Kunci: data mining, k-means clustering, menu potensial, rapidminer.

ABSTRACT

This study aims to identify potential customers and help local coffee entrepreneurs compete by offering products that customers desire, as well as address challenges in determining high-potential menu items using the K-Means algorithm. The data consist of TitikSetara Coffee's revenue records from October 2023 through January 2024. RapidMiner software was used to perform clustering via the K-Means algorithm. The sales data were grouped into three clusters: Cluster 0 for high-potential menu items, Cluster 1 for medium-potential menu items, and Cluster 2 for low-potential menu items. Out of the 26 records analyzed, Cluster 0 contains 4 menu items, Cluster 1 contains 14 menu items, and Cluster 2 contains 8 menu items, with a Davies-Bouldin Index of 0.374. The identified high-potential menu items can serve as a reference for TitikSetara Coffee, and it is expected that these recommendations will be well-targeted.

Keywords: data mining, k-means clustering, menu potensial, RapidMiner.

Penulis Korespondensi

Muhammad Hasbi Abdillah

Tanggal Submit : 09/07/2024

Tanggal Diterima : 12/12/2024

Tanggal Terbit : 28/02/2025

This is an open access article under the [CC-BY-NC-SA](https://creativecommons.org/licenses/by-nc-sa/4.0/) license



Copyright: © 2023 by the authors. Submitted for possible open access publication under the terms and conditions of the Creative Commons Attribution-NonCommercial-ShareAlike 4.0 (CC BY-NC-SA 4.0) International License (<https://creativecommons.org/licenses/by-nc-sa/4.0/>).

Publisher's Note: JPPM stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.

I. PENDAHULUAN

Metode text mining adalah evolusi dari data mining yang berfokus pada analisis teks. Ini adalah proses otomatis yang dilakukan oleh komputer untuk mengekstraksi informasi berharga dari kumpulan teks dalam sebuah dokumen. Langkah utama dalam metode ini melibatkan identifikasi kata-kata yang mencerminkan isi dokumen, kemudian menganalisis hubungan antar dokumen menggunakan metode statistik seperti teknik clustering.

Terdapat berbagai algoritma yang diajukan untuk melakukan proses clustering pada dataset berukuran besar. Dalam penelitian ini, peneliti akan menerapkan algoritma K-Means untuk menentukan jumlah cluster terbaik. Algoritma K-Means sering digunakan karena keefektifan dan efisiensinya. Kelebihan K-Means adalah mudah dipelajari dan memiliki waktu proses komputasi yang relatif singkat. K-Means adalah salah satu metode

clustering yang digunakan dalam data mining untuk mengelompokkan data menjadi beberapa kelompok berdasarkan titik pusat (centroid) [1].

Pelaku usaha harus bisa melakukan inovasi atau menyediakan produk sesuai dengan apa yang diinginkan oleh pelanggannya seperti meningkatkan pelayanannya, promosi, citra perusahaan, variasi menu dan segala yang dibutuhkan oleh pelanggan. Namun dalam menjalankan sebuah usaha pastinya akan selalu ada yang namanya hambatan atau masalah. Masalah yang kerap terjadi pada bidang usaha kedai kopi adalah cara mengetahui pelanggan terkait menu mana yang paling memiliki potensi untuk para pelanggan agar membeli kembali di sebuah kedai kopi. Untuk itu, teknik yang akan diterapkan dalam karya ilmiah ini adalah data mining untuk pengelompokan dengan algoritma K-Means Clustering. Metode K-Means Clustering digunakan untuk mengelompokkan cabang yang produktif berdasarkan kriteria yang telah ditentukan. Clustering dalam data mining seperti yang diketahui berfungsi untuk menganalisis objek berdasarkan kesamaan sifat antar objek tersebut. Penelitian ini akan mencakup perancangan perangkat lunak menggunakan Microsoft Excel dan RapidMiner yang diharapkan dapat menyediakan solusi untuk masalah yang diidentifikasi.

Berdasarkan permasalahan di atas, maka peneliti tertarik untuk meneliti lebih lanjut dalam penelitian yang berjudul Penentuan Menu Potensial Menggunakan Algoritma K-Means Pada Titiksetara Coffee Bekasi.

II. LANDASAN TEOR

1. Penelitian Terkait

Penelitian ini dilakukan untuk mengklasifikasikan tingkat penjualan menu (Food and Beverage) menggunakan algoritma K-Means. Hasil penelitian menunjukkan bahwa berdasarkan proses clustering menu kelompok menu dengan penjualan tertinggi ditemukan di cluster 1 dengan nilai rata-rata penjualan 351,29 dan mencakup 5 jenis menu. Menu dengan penjualan sedang terletak di cluster 0 dengan nilai rata-rata penjualan 141,89 dan meliputi 11 jenis menu. Menu dengan penjualan rendah berada di cluster 2 dengan nilai rata-rata penjualan 53,52 dan terdiri dari 25 jenis menu. Sedangkan cluster 3 yang mencerminkan menu dengan penjualan sangat rendah, memiliki nilai rata-rata penjualan 8,71 dan mencakup 131 jenis menu. Hasil penelitian ini menunjukkan bahwa metode clustering dengan algoritma K-Means memberikan hasil yang signifikan dalam mengelompokkan data penjualan [2].

2. K-Means Clustering

Mencantumkan langkah-langkah algoritma pengelompokan K-Means. Algoritma ini dimulai dengan nilai K tertentu dan awalnya menetapkan pusat massa dari

cluster K. Ini dilanjutkan dengan meminta masing-masing n titik data dalam kumpulan data bergabung dengan cluster terdekatnya, memperbarui pusat massa hingga tidak lagi berubah. Untuk menentukan cluster terdekat ke titik data, hitung jarak ke cluster data. Vektor rata-rata titik data dalam suatu cluster sering digunakan sebagai pusat massa. Gunakan ukuran kesamaan atau ketidaksamaan untuk menghitung jarak dari titik data ke cluster. Salah satu metode untuk menetapkan pusat massa awal untuk k cluster adalah dengan memilih secara acak K titik data dari kumpulan data. Meskipun metode ini menggunakan titik data spesifik untuk menyiapkan pusat massa awal, cluster K pada awalnya tidak memiliki titik data. Metode lain termasuk menggunakan hasil pengelompokan hierarki atau pusat cluster yang diperoleh untuk menginisialisasi algoritma pengelompokan K-means [3].

Algoritma K-Means clustering adalah metode yang mudah dan efektif untuk mengidentifikasi cluster dalam data. Algoritmanya berkembang sebagai berikut:

- 1) Pengguna menentukan jumlah cluster (k)
- 2) Melibatkan penetapan k record secara acak sebagai lokasi pusat cluster awal yang diinginkan
- 3) Menugaskan setiap record ke pusat cluster terdekat, membuat cluster C_1, C_2, \dots, C_k . Lokasi centroid setiap kelompok diambil dari rata-rata (mean) semua nilai data pada setiap fiturnya.
- 4) Menemukan pusat cluster untuk setiap cluster dan memperbarui lokasi pusatnya. Ada beberapa cara yang dapat dilakukan untuk mengukur jarak data ke pusat kelompok, diantaranya adalah Euclidean.
- 5) Langkah 3-5 diulangi hingga konvergensi atau terminasi. Kembali ke langkah 3, apabila masih ada data yang berpindah kelompok atau apabila ada perubahan nilai centroid di atas nilai ambang yang ditentukan, atau apabila perubahan nilai pada fungsi objektif yang digunakan masih di atas nilai ambang yang ditentukan [4].

3. Davies Bouldin Index

Davies-Bouldin Index (DBI) merupakan metode untuk menilai validitas internal dalam evaluasi clustering. DBI mengukur seberapa efektif clustering dengan menghitung kuantitas dan fitur turunan dari dataset. Tahapan perhitungan Davies-Bouldin Index meliputi langkah-langkah berikut [5]:

- 1) Sum of Square Within cluster (SSW)

Sum of Squares Within cluster (SSW) adalah rumus yang digunakan untuk mengukur matriks

kohesi dalam sebuah cluster ke-i. Persamaan untuk menghitung SSW adalah sebagai berikut:

$$SSWi = \frac{1}{mi} \sum_{j=i}^{mi} d(xj, ci)$$

2) Sum of Square Between cluster (SSB)

Sementara itu, metrik untuk mengukur separasi antara dua cluster menggunakan formula Sum of Squares Between Clusters (SSB). SSB mengukur jarak antara centroid c_i dan c_j dari dua cluster yang berbeda, dan dirumuskan dengan persamaan sebagai berikut:

$$SSBi, j = d(ci, cj)$$

3) Ratio (Rasio)

Langkah berikutnya adalah menghitung nilai R_{ij} yang merupakan ukuran rasio untuk menilai perbandingan antara cluster ke-i dan cluster ke-j. Nilai

R_{ij} dihitung berdasarkan komponen kohesi dan separasi. Cluster yang ideal adalah yang memiliki kohesi sekecil mungkin dan separasi sebesar mungkin. Rumus untuk menghitung R_{ij} adalah sebagai berikut:

$$R_{ij} = \frac{SSWi + SSWj}{SSBij}$$

4) Davies-Bouldin Index (DBI)

Nilai rasio yang diperoleh tersebut digunakan untuk mencari nilai Davies- Bouldin Index (DBI) dengan menggunakan persamaan berikut:

$$DBI = \frac{1}{k} \sum_{i=1}^k \max_{i \neq j} (R_{i,j})$$

Seperti yang terlihat dari persamaan di atas, semakin rendah nilai DBI maka cluster yang dihasilkan akan semakin baik, yang berarti kemiripan antar data dalam suatu cluster akan semakin mirip [6].

4. Menu Potensial

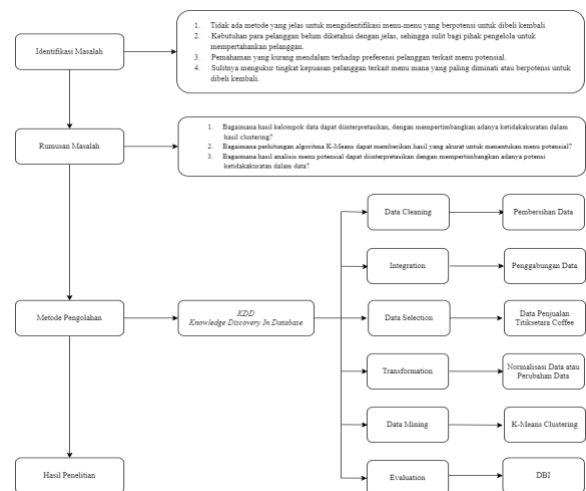
Menu potensial dapat merujuk pada daftar hidangan atau variasi hidangan yang memiliki kemampuan, peluang, atau harapan untuk dikembangkan lebih lanjut. Dalam konteks ini, perencanaan menu dijelaskan sebagai serangkaian kegiatan untuk menyusun menu dengan variasi yang serasi untuk manajemen penyelenggaraan makanan di institusi. Proses perencanaan menu dapat dianggap sebagai suatu potensi karena melibatkan pengembangan dan peluncuran menu baru, menunjukkan bahwa menu tersebut

memiliki kemungkinan atau harapan untuk berkembang lebih lanjut. Menu potensial dapat diartikan sebagai daftar hidangan atau variasi hidangan yang memiliki potensi untuk ditingkatkan, dikembangkan, atau diadaptasi sesuai dengan kebutuhan atau preferensi yang mungkin muncul di masa depan [7].

III. METODE PENELITIAN

1. Kerangka Penelitian

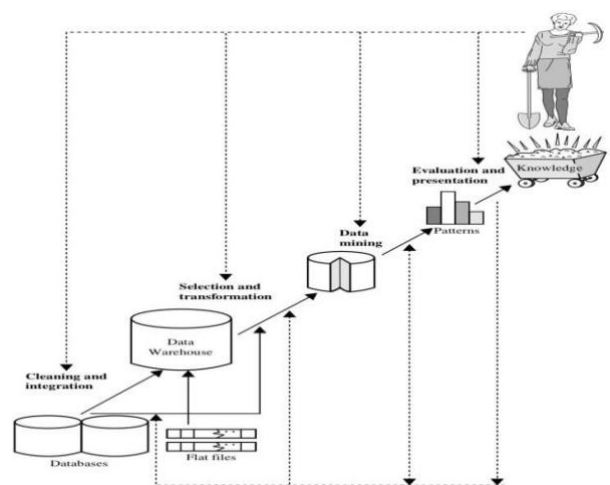
Pada penelitian ini akan dilaksanakan sesuai dengan tahap – tahapan yang ada pada gambar merupakan rincian dari rancangan penelitian yang akan dilaksanakan:



Gambar 1. Kerangka Penelitian

2. Tahapan Penelitian

Pada penelitian ini akan dilaksanakan sesuai dengan tahap – tahap yang ada pada metode KDD gambar merupakan rincian dari rancangan penelitian yang akan dilaksanakan. Berikut gambar KDD:



Gambar 2. Skema Dari Proses KDD

1) Data Cleaning

Data yang digunakan dalam penelitian ini mencakup data penjualan dari bulan Oktober 2023 hingga Januari 2024 pada Titiksetara Coffee. Proses pembersihan data (data cleaning) dilakukan untuk menangani nilai yang hilang, gangguan, dan ketidakonsistenan dalam data. Proses ini dilakukan secara manual untuk memastikan data yang diperoleh memenuhi standar kualitas yang diinginkan. Setelah tahap pembersihan, data yang telah diperbaiki dimasukkan ke dalam file dataset Excel agar siap untuk diproses pada tahap berikutnya.

2) Integration

Setelah proses pembersihan data selesai tahap berikutnya adalah integrasi data yaitu proses menggabungkan data dari berbagai sumber. Dalam hal ini, file dataset Excel yang telah dibersihkan digunakan untuk membuat lembar kerja baru yang menggabungkan data pemasukan harian menjadi pemasukan mingguan. Proses ini membantu mengurangi redundansi data dan ketidakonsistenan yang mungkin timbul dari pengambilan data dari berbagai sumber.

3) Selection

Setelah tahap penggabungan data selesai, langkah berikutnya adalah seleksi data. Proses ini bertujuan untuk memilih data yang relevan untuk analisis mining sambil memastikan data tetap mewakili kondisi aslinya. Data yang telah diseleksi kemudian diubah dari pemasukan mingguan menjadi total pemasukan untuk bulan Oktober. Tahap ini melibatkan pengumpulan data penjualan dari Titiksetara Coffee untuk periode dari Oktober 2023 hingga Januari 2024. Data ini selanjutnya digunakan dalam penelitian untuk analisis lebih lanjut.

4) Transformation

Setelah tahap pembersihan data, integrasi data, dan seleksi data selesai, proses selanjutnya adalah transformasi data. Proses dilakukan dengan menerapkan algoritma K-Means yang cukup populer, dikarenakan kesederhanaan dan kecepatan dalam prosesnya. Tahap ini bertujuan untuk mengetahui informasi yang relevan dari kumpulan data pada komputer. Pada tahap ini melakukan perubahan data set yaitu dengan memberikan kode menu.

5) Data mining

Tahapan ini melibatkan penerapan metode data mining untuk memproses data yang ada. Metode yang digunakan adalah clustering K-Means yang telah dijelaskan sebelumnya dengan proses menghitung jarak setiap data ke pusat cluster atau centroid menggunakan jarak Euclidean sesuai rumus yang telah ditetapkan. Output dari proses ini adalah identifikasi menu yang memiliki potensi dalam distribusi produk serta mengetahui produk yang paling diminati oleh konsumen. Berdasarkan pengolahan data yang melibatkan atribut seperti kode

produk, nama produk, harga produk, total penjualan produk dan total pendapatan, akan dihasilkan tiga cluster yakni potensi tinggi, potensi sedang, dan potensi rendah.

6) Evaluation

Dalam penelitian ini, pengukuran menu potensial dilakukan dengan menggunakan model Davies Bouldin Index. Davies Bouldin Index adalah metode evaluasi untuk mengukur kualitas cluster dengan mempertimbangkan kesamaan ukuran cluster dan perbedaan antar-cluster dalam distribusi data. Teknik ini bertujuan untuk menilai kinerja model dalam menentukan jumlah cluster yang optimal. Semakin kecil nilai Davies Bouldin Index, semakin baik hasil clustering yang dicapai.

IV. HASIL DAN PEMBAHASAN

1. Hasil Penelitian

1) Proses Clustering

Proses clustering yang akan dilakukan adalah dengan mengelompokkan data penjualan TitikSetara Coffee dalam waktu 4 bulan yaitu dari bulan Oktober 2023 hingga Januari 2024. Tahapan proses clustering akan menghasilkan kelompok data ke dalam cluster Berdasarkan karakteristik yang dimiliki oleh data. Berikut tahapan proses clustering:

- a. Memilih jumlah Cluster
- b. Pengelompokkan Data
- c. Cluster baru dengan pengambilan rata-rata dari semua data
- d. Pengulangan Langkah b dan c
- e. Konvergensi

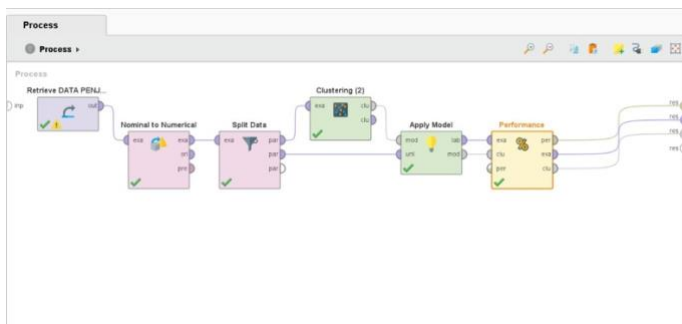
Dalam penelitian ini cluster yang dipilih adalah cluster dengan 3 kelompok dengan operator split data dan apply model untuk data testing dan data training. Setelah split data diberikan ratio data testing dan data training selanjutnya adalah memilih jumlah clustering pada operator cluster dengan di operatorkan pada apply model. Setelah program dijalankan dengan tahapan diatas maka akan keluar hasil performance vector dari hasil pengolahan tersebut. Data yang diambil untuk penelitian berjumlah 132 data penjualan dengan atribut kode barang, nama menu, harga menu, jumlah transaksi dan total pendapatan. Setelah dilakukan split data, data berjumlah 26 data yang akan digunakan dalam penerapan algoritma K-Means. Percobaan ini melibatkan penggunaan parameter-parameter berikut:

Jumlah cluster	: 3
Jumlah data	: 26
Jumlah atribut	: 6

2) Hasil Uji Rapid Miner

a. Design Process

Design process dalam RapidMiner adalah pendekatan sistematis untuk merancang alur kerja dalam analisis data. Ini mencakup pemilihan operator yang tepat, penyusunan alur kerja secara visual dengan menghubungkan operator, konfigurasi parameter operator, eksekusi alur kerja, dan evaluasi hasil untuk memastikan relevansi dan interpretasi yang akurat dari informasi yang dihasilkan. Dalam penelitian ini, penulis menggunakan perangkat RapidMiner sebagai alat utama untuk menguji dataset yang digunakan. Adapun tahapan-tahapan yang dilakukan dalam proses pengujian mencakup langkah-langkah berikut secara berurutan:



Gambar 3. Design Process Clustering

Gambar 3 menggambarkan proses desain yang terdiri dari empat tahapan yang dilakukan sebagai berikut:

a) Read excel

Tahapan ini melibatkan operasi untuk mengimpor dataset yang berupa file dengan ekstensi .xls, yang berisi data penjualan TitikSetara Coffee di Bekasi selama 4 bulan.

b) Nominal to Numerical

Setelah dilakukan import data pada design proses, Langkah selanjutnya adalah mengganti data nominal dengan numerikal.

c) Split Data

Split data digunakan untuk membagi dataset menjadi beberapa subset yang dapat digunakan untuk berbagai tujuan seperti pelatihan (training) dan pengujian (testing). Pembagian data ini penting untuk memastikan bahwa model yang dibangun dapat diuji dan divalidasi dengan data yang belum pernah dilihat sebelumnya, yang membantu menghindari overfitting dan memastikan bahwa model dapat menggeneralisasi dengan baik ke data baru.

d) Clustering

Tahapan ini melibatkan operasi clustering menggunakan algoritma tertentu yang digunakan dalam penelitian ini. Dalam RapidMiner terdapat tools operator yang digunakan untuk memilih metode yang ingin digunakan yang mana dalam penelitian ini menggunakan

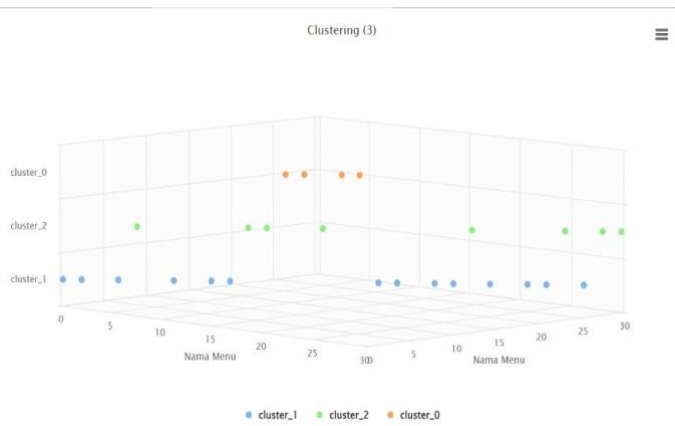
k-means clustering. Namun setelah dicari pada kolom operator “k-means” ditarik dan di drop pada design process.

e) Performance

Tahapan performa dalam metode clustering K-Means melibatkan evaluasi hasil clustering setelah data dibagi, model diterapkan, dan operator performa digunakan. Setelah operator Split Data membagi dataset menjadi subset pelatihan dan pengujian. Selanjutnya, operator Apply Model diterapkan pada subset pengujian untuk memprediksi cluster anggota data. Operator Performance kemudian digunakan untuk mengevaluasi hasil clustering dengan berbagai metrik. Dalam konteks clustering metrik evaluasi Davies-Bouldin Index digunakan untuk mengukur seberapa baik data dikelompokkan dalam cluster. Operator Performance di RapidMiner menghitung dan menampilkan metrik-metrik ini, memungkinkan peneliti untuk menilai kualitas dan ketepatan pengelompokan yang dilakukan oleh algoritma K-Means. Evaluasi ini penting untuk memastikan bahwa cluster yang dihasilkan memiliki interpretasi yang bermakna dan dapat digunakan untuk keputusan yang lebih baik.

b. Scatter 3D Result

Scatter 3D Result adalah fitur yang memungkinkan pengguna untuk memvisualisasikan hubungan antara tiga variabel numerik secara simultan dalam ruang tiga dimensi. Dengan menggunakan Scatter 3D Result, pengguna dapat melihat pola dan hubungan kompleks di antara ketiga variabel tersebut dengan lebih jelas dan intuitif. Setiap titik pada plot mewakili satu instansi data, di mana sumbu x, y, dan z masing-masing mewakili nilai dari tiga variabel yang dipilih. Plot ini sangat bermanfaat dalam eksplorasi data untuk menemukan korelasi yang kompleks atau pola-pola yang mungkin sulit dideteksi melalui analisis dua dimensi atau hanya dengan memeriksa data mentah dalam bentuk tabel.



Gambar 4. Scatter 3D Result

Gambar 4. merupakan hasil Scatter 3D result yang diolah menggunakan RapidMiner menunjukkan bahwa cluster_0 atau cluster 1 merupakan cluster dengan data penjualan yang paling tinggi, lalu cluster yang merupakan data penjualan sedang yaitu cluster_2 atau cluster 3 dan yang terakhir cluster yang merupakan data penjualan paling kecil yaitu cluster_1 atau cluster 2

c. Evaluasi

Pada tahap ini dilakukan pengujian untuk mengevaluasi serta mengetahui performa dari model yang digunakan dalam memprediksi data. Proses evaluasi pada penelitian ini menggunakan Davies Bouldin Index atau DBI,

yaitu dengan menentukan cluster terbaik pada rapid miner. dimana metode ini merupakan salah satu evaluasi internal cluster dilihat dari nilai terendah, dimana hasil jumlah cluster dapat ditinjau dari kohesi dan separasi.

Davies Bouldin Index (DBI) merupakan salah satu teknik validitas internal yang digunakan untuk menilai atau mengevaluasi kualitas sebuah cluster. Seperti yang diilustrasikan oleh persamaan di atas nilai DBI yang lebih rendah menunjukkan bahwa kualitas cluster yang dihasilkan lebih baik karena menunjukkan kemiripan antar data dalam satu cluster yang lebih tinggi. Berdasarkan hasil dari ketiga performance vector tersebut dapat disimpulkan bahwa nilai DBI yang paling mendekati nol diperoleh dengan penggunaan 3 cluster. Perhitungan dengan algoritma K-Means menghasilkan nilai DBI sebesar 0,374, yang menunjukkan bahwa objek-objek dalam cluster tersebut memiliki tingkat kemiripan yang relatif tinggi, mendekati angka 0.

2. Pembahasan

Setelah dilakukan pengujian dengan tool Rapid Miner, jumlah cluster yang digunakan dalam penelitian ini adalah 3 cluster dimana cluster_0 atau cluster 1 memiliki anggota sebanyak 4 menu yang disediakan dalam tabel berikut:

Tabel 1. Anggota Cluster_0

Kode	Nama Menu	Harga Menu	Jumlah Transaksi	Total Pendapatan
A112	Kopi Setara	21000	694	14.574.000
A113	Titik Semangat	23000	543	12.489.000
A115	Cappucino	23000	614	14.122.000
A116	Lemon tea	20000	636	12.720.000
Total			2487	53.905.000

Sumber: Hasil olah peneliti (2024)

Tabel 2. Anggota Cluster_1

Kode	Nama Menu	Harga Menu	Jumlah Transaksi	Total Pendapatan
A100	Mix Platter	25000	186	4.650.000
A101	Indomie Setara	15000	230	3.450.000
A103	Cireng	20000	46	920.000
A106	Kebab Mini	25000	138	3.450.000
A108	Burger	25000	134	3.350.000
A109	Snack Bucket	30000	134	4.020.000

A117	On The Rock	16000	119	1.904.000
A118	Americano	18000	159	2.862.000
A121	Rum	25000	64	1.600.000
A122	Single Origin	20000	154	3.080.000
A123	Japanesse	23000	204	4.692.000
A125	Vanilla	23000	161	3.703.000

Sumber: Hasil olah peneliti (2024)

Tabel 3. Anggota Cluster_2

Kode	Nama Menu	Harga Menu	Jumlah Transaksi	Total Pendapatan
A104	Friench Fries	20000	412	8.240.000
A110	Kocok Setara	23000	355	8.165.000
A111	Setara Aren	23000	310	7.130.000
A114	Coffee Latte	20000	305	6.100.000
A122	Special Beans	30000	260	7.800.000
A127	Huzzelnut	23000	371	8.533.000
A129	Strawberry Milk	25000	299	7.475.000
A130	Vietnam Drip	20000	270	5.400.000
Total			2582	58.843.000

Sumber: Hasil olah peneliti (2024)

Dalam penelitian ini, cluster_1 atau cluster 2 memiliki anggota sebanyak 14 yang disediakan dalam Tabel 3. Dalam penelitian ini, cluster_2 atau cluster 3 memiliki anggota sebanyak 8 yang disediakan dalam tabel berikut:

Tabel yang telah disediakan diatas adalah hasil dari total 33 dataset yang telah diuji menggunakan rapidminer dimana cluster yang memiliki potensi atau penjualan paling diminati adalah cluster_0 atau cluster 1. Menu yang dapat

direkomendasikan kepada pihak TitikSetara Coffee mengingat potensi yang dimiliki menu terdiri dari Kopi Setara, Titik Semangat, Cappucino dan Lemon Tea. Setelah cluster terbentuk seperti penjelasan diatas, cluster diberikan deskripsi mengenai cluster mana yang paling berpotensi, cluster yang memiliki potensi sedang hingga cluster yang memiliki potensi rendah. Maka dapat disimpulkan pada Tabel 4.

Tabel 4. Deskripsi Data Berdasarkan Dengan Cluster

Cluster	Deskripsi
Cluster_0	Anggota cluster merupakan menu dengan potensi tinggi.
Cluster_1	Anggota cluster merupakan menu dengan potensi rendah.
Cluster_2	Anggota cluster merupakan menu dengan potensi sedang.

Sumber: Hasil olah peneliti (2024)

Table 4. memberikan deskripsi data yang dikelompokkan berdasarkan hasil clustering. Setelah proses clustering selesai, langkah berikutnya adalah melakukan evaluasi kinerja untuk menghitung nilai Davies Bouldin Index. Tujuan dari evaluasi ini adalah untuk

memaksimalkan jarak antar cluster dan meminimalkan jarak antara anggota dalam satu cluster, sehingga memastikan kualitas dan kekompakan cluster yang dihasilkan. Penelitian ini telah menguji beberapa data dalam beberapa kelompok atau cluster, namun angka yang paling

mendekati 0 adalah cluster atau kelompok yang berjumlah 3. Dari hasil tersebut diperoleh cluster menu yang paling berpotensi atau sering dibeli oleh pelanggan adalah anggota cluster_0.

Berdasarkan hasil olah data dari 3 percobaan jumlah cluster, cluster dengan jumlah 3 kelompok merupakan cluster yang memiliki hasil DBI paling mendekati nol yang dimana semakin data mendekati angka 0 maka kualitas data semakin baik. Dimana percobaan cluster dengan jumlah kelompok 2, percobaan cluster dengan jumlah kelompok 3 dan percobaan cluster dengan jumlah kelompok 4. Berikut ini disediakan tabel hasil dari nilai DBI dari masing-masing percobaan jumlah cluster yaitu:

Hasil dari evaluasi davies bouldin index dengan bantuan cluster distance performance menghasilkan cluster dengan jumlah anggota terbaik yaitu cluster dengan 3 anggota yang mana memiliki nilai value Davies Bouldin Index paling mendekati 0. Maka dapat disimpulkan bahwa dalam penelitian ini cluster dengan 3 anggota merupakan cluster dengan kualitas data yang baik atau data dengan kemiripan paling baik.

V. KESIMPULAN

Penelitian ini berhasil menerapkan model klasifikasi K-Means untuk menentukan menu potensial di TitikSetara Coffee Bekasi berdasarkan 132 data penjualan periode Oktober 2023–Januari 2024. Data dibagi menjadi 80 % untuk pelatihan (106 data) dan 20 % untuk pengujian (26 data). Hasil clustering menunjukkan bahwa cluster 2 (menu potensial) terdiri dari empat item—Kopi Setara, Titik Semangat, Cappuccino, dan Lemon Tea—yang memiliki performa penjualan tertinggi menurut Scatter 3D Plot, sedangkan evaluasi Davies–Bouldin Index sebesar 0,403 mengonfirmasi bahwa pembagian tiga cluster adalah yang paling optimal.

Penggunaan RapidMiner dan Microsoft Excel terbukti memudahkan dan meningkatkan akurasi proses pengolahan data. Seluruh tahapan penelitian—mulai dari pembersihan data, pemeriksaan missing values, hingga penerapan algoritma dan evaluasi DBI—berjalan lancar tanpa kendala data bermasalah, sehingga keandalan hasil dapat dipertanggungjawabkan.

Berdasarkan temuan ini, disarankan agar penelitian selanjutnya memanfaatkan dataset yang lebih besar untuk semakin memperkuat validitas model K-Means, serta terus memadukan RapidMiner dan Excel agar proses analisis tetap efisien. Menu-menu yang teridentifikasi sebagai potensial dapat langsung dijadikan acuan strategis oleh manajemen TitikSetara Coffee, sehingga penawaran produk lebih tepat sasaran dan mampu bersaing di pasar lokal.

DAFTAR PUSTAKA

- [1] N. P. E. Marliana, Ernawati and J. Santoso, "Analisa penentuan jumlah cluster terbaik pada metode k-means clustering," 2015.
- [2] H. Syahputra, L. Mayola and D. Guswandi , "Clustering tingkat penjualan menu (food and beverage)menggunakan algoritma k-means," *Jurnal KomtekInfo*, vol. 9, no. 1, pp. 29-33, 2022.
- [3] N. Ye, *Data mining theories, algorithms and examples*, Boca Raton: CRC Press, 2013.
- [4] W. S. Aziz and D. D. Atmajaya, "Pengelompokan minat baca mahasiswa menggunakan metode k-means," *ILKOM Jurnal Ilmiah*, vol. 8, no. 2, 2016.
- [5] S. Retno, "Peningkatan akurasi algoritma k-means dengan clustering purity sebagai titik pusat cluster awal (centroid)," Universitas Sumatera Utara, 8 Juli 2019.
- [6] D. A. Rahmayanti, R. Juita and C. D. Suhendra, "Penerapan metode k-means untuk clustering data anak berdasarkan kepemilikan akta kelahiran dan KIA," *Informatics Journal*, vol. 7, no. 3, pp. 210-219, 2022.
- [7] Trianasari, Buku ajar : *manajemen makanan dan minuman*, Jakarta: CV. PLC, 2021.
- [8] A. P. Riani, A. Voutama and T. Ridwan, "Penerapan k-means clustering dalam pengelompokan hasil belajar peserta didik dengan metode elbow," *Jurnal Teknologi Sistem Informasi dan Sistem Komputer TGD*, vol. 6, no. 1, pp. 164-172, 2023.