

ANALISIS DATA TRANSAKSI KUE MENGGUNAKAN METODE K-MEANS CLUSTERING PADA TOKO RAFA CAKE

Nur Kholifa, Nana Suarna, Willy Prihartono

Teknik Informatika, STMIK IKMI Cirebon

Jl. Perjuangan No.10B, Karyamulya, Kec. Kesambi, Kota Cirebon, Jawa Barat 45135

nkholifa6@gmail.com

ABSTRAK

Penggunaan teknologi dalam bisnis makanan, khususnya di industri toko kue, sangat penting untuk mendapatkan informasi detail tentang produk yang dijual. Toko Rafa Cake menghadapi masalah pengelolaan data transaksi belum dimanfaatkan secara efektif dan penataan produk yang kurang optimal. Banyaknya transaksi mengakibatkan penumpukan data dan stok produk, menghambat pengisian ulang etalase dan pemahaman produk terlaris. Studi ini menerapkan metode algoritma *K-Means Clustering* pada data transaksi toko, dengan tujuan mengidentifikasi produk yang diminati, cukup diminati, dan kurang diminati. Data transaksi yang dianalisis meliputi no, nama produk, kategori, variasi, harga, merek, stok terjual, dan total penjualan melalui *RapidMiner*. Hasilnya dapat meningkatkan efisiensi penjualan produk kue dan manajemen stok, serta dapat menjadi acuan bagi toko kue lainnya dengan masalah yang sama, memanfaatkan data transaksi untuk meningkatkan kinerja bisnis. Hasil *Clustering* diperoleh 3 *Cluster* yang membedakan produk kue, bolu, dan roti berdasarkan kategori produk, stok terjual, dan harga. *Cluster 1* memiliki 1 item menunjukkan penjualan paling diminati dengan nilai transaksi tinggi. Sementara *Cluster 2* memiliki 40 item, menandakan penjualan cukup diminati dengan nilai transaksi sedang. *Cluster 0*, dengan 1272 item, menunjukkan penjualan kurang diminati dengan nilai transaksi rendah. Dengan nilai *DBI* sebesar 0,004, dan jarak rata-rata *Centroid* terbaik di *Cluster 1* sebesar 0,000.

Kata kunci: Analisis, K-Means, Clustering, Toko Kue, Data Transaksi

1. PENDAHULUAN

Industri makanan telah mengalami perkembangan pesat dari tahun ke tahun, termasuk dalam industri makanan kue dan roti. Salah satu contohnya adalah usaha Toko Rafa Cake, yang berlokasi di Jl. Kandang Perahu Gang Jati No.42, Karyamulya, Kec. Kesambi Kota Cirebon. Toko ini telah menjadi favorit para pencinta makanan manis, dengan berbagai produk kue, bolu, dan roti yang lezat dan menggugah selera. Pada dunia bisnis, kemajuan teknologi telah membuka pintu bagi inovasi seperti platform pemesanan online dan pengiriman makanan, yang juga relevan bagi usaha seperti Toko Rafa Cake. Toko Rafa Cake dapat memanfaatkan teknologi informasi untuk memperluas jangkauan pasar, berinteraksi dengan pelanggan melalui media sosial, dan mengoptimalkan rantai pasokan.

Permasalahan yang akan dikaji yaitu, pemahaman tentang preferensi pelanggan dan identifikasi produk yang diminati, cukup diminati atau kurang diminati menjadi kunci kesuksesan dalam industri yang kompetitif. Toko Rafa Cake, sebagai bagian dari industri makanan, juga menghadapi tantangan ini. Dengan mengetahui produk yang diminati oleh pelanggan, Toko Rafa Cake dapat lebih fokus pada pengembangan dan peningkatan kualitas produk tersebut. Sebaliknya, produk yang kurang diminati dapat diidentifikasi dengan tujuan mencari solusi untuk meningkatkan minat pelanggan.

Pada penelitian sebelumnya yang diterapkan oleh Fachriansyah & Bu, dengan judul "*Penerapan Algoritma K-Means Untuk Clustering Bakery Dan Cake Yang Laris*" yang dimuat di Jurnal Bulletin of

Information Technology (BIT). Menghadapi tantangan kurangnya penggunaan analisis statistik dalam memilih strategi pemasaran untuk toko, pendekatan yang digunakan metode *K-Means* untuk mengelompokkan produk roti dan kue yang paling laris di CV Syahfira Bakery and Cake. Berdasarkan temuan Fachriansyah & Bu, dua produk yang paling laris adalah Roti Bungkus Keju dan Roti Bungkus Keju Cokelat setelah menerapkan metode *K-Means* untuk menghitung posisi mereka dalam *Cluster 1* [1].

Selanjutnya, penelitian yang dilakukan oleh Ariestiany & Santoso dengan judul "*Analisis Pengelolaan Stok Kue Menggunakan Algoritma Apriori Pada Toko Kue Sponji*" telah mengajukan suatu pendekatan dengan memanfaatkan Algoritma Apriori dan teknik asosiasi untuk meramalkan jumlah varian kue yang perlu dipersiapkan di toko Sponji pada bulan berikutnya. Temuan yang dihasilkan dalam penelitian ini teridentifikasi bahwa kombinasi rasa dengan varian topping keju/ceres dan topping keju/kacang menonjol dengan tingkat keyakinan tertinggi, mencapai 60%[2]. Berdasarkan temuan penelitian sebelumnya, Dengan menggunakan data transaksi yang mencakup periode waktu yang lebih lama, penelitian ini berupaya untuk menganalisis tren jangka panjang dalam perilaku pembelian pelanggan dan mendukung toko dalam penyesuaian strategi bisnis dengan lebih efektif.

Tujuan utama dari penelitian ini adalah mengidentifikasi produk paling diminati, cukup diminati dan kurang diminati oleh konsumen melalui teknik algoritma *K-Means Clustering* pada data transaksi toko. Penelitian ini bertujuan untuk

mengoptimalkan strategi pemasaran dan mengurangi penumpukan stok produk yang kurang diminati di toko.

2. TINJAUAN PUSTAKA

2.1. Analisis

Analisis merujuk pada proses pemecahan sistem informasi secara menyeluruh ke dalam komponen-komponen individu. Tujuannya adalah untuk mengidentifikasi, mengevaluasi, serta memahami permasalahan, peluang, kendala, dan kebutuhan yang terkait guna memperoleh saran-saran perbaikan yang diperlukan[3].

2.2. Proses Pemasaran

Untuk bersaing efektif dengan produk dari pesaing, perusahaan perlu menerapkan strategi pemasaran yang efisien. Dengan menerapkan strategi bersaing yang kuat, perusahaan dapat mempertahankan posisi terdepan di pasar yang mereka tuju. Hal ini bertujuan untuk memastikan keunggulan posisi perusahaan dalam kompetisi pasar yang ketat. Untuk meningkatkan volume penjualan, diperlukan penerapan strategi pemasaran yang diawali dengan analisis menyeluruh terhadap strategi pemasaran yang baru, seperti produk, harga, distribusi, promosi, personel, proses, dan lingkungan fisik. Analisis atas strategi pemasaran baru tersebut dilakukan untuk mengidentifikasi variabel yang memiliki dampak paling signifikan terhadap peningkatan volume penjualan. Dengan demikian, dapat diidentifikasi kekuatan, kelemahan, peluang, dan ancaman yang dapat dijadikan strategi kompetitif dalam upaya meningkatkan volume penjualan[4].

2.3. Kue, Bolu dan Roti

Bolu merupakan hasil olahan dari campuran tepung terigu atau variasi tepung lainnya yang diperkaya dengan tambahan berbagai bahan untuk meningkatkan nilai gizi. Proses pembuatan bolu dapat dilakukan dengan metode pengukusan atau pemanggangan[5].

Kue atau bolu merupakan produk kuliner yang terbuat dari campuran bahan dasar seperti tepung, yang biasanya adalah tepung terigu, gula, serta telur. Proses pembuatan kue atau bolu melibatkan teknik memasak dengan menggunakan oven, alat pemanggangan khusus, atau dalam beberapa kasus, metode pengukusan. Keberagaman bentuk dan rasa yang manis menjadikannya sebagai sajian yang diminati secara luas dalam masyarakat pada masa kini. Selain dihidangkan sebagai hidangan penutup, kue atau bolu juga sering disajikan sebagai tambahan saat jamuan teh atau acara santai lainnya[6].

Roti adalah produk pangan yang dihasilkan dari adonan tepung terigu yang telah difermentasi dengan ragi *Saccharomyces cerevisiae*, lalu dipanggang menggunakan oven. Variasi komposisi bahan dan metode pembuatan mempengaruhi variasi jenis dan bentuk roti yang[7].

2.4. KDD (Knowledge Discovery in Database)

Knowledge Discovery and Database adalah proses yang bertujuan untuk mengungkap informasi yang tersembunyi dalam data dan mengidentifikasi pola-pola karakteristik yang ada dalam dataset tersebut[3].

2.5. Data Mining

Penambangan data merupakan proses yang bertujuan untuk memperoleh informasi baru dari sekumpulan data dengan menerapkan algoritma dan teknik yang terkait dengan disiplin ilmu statistik, machine learning, serta manajemen basis data. Praktik *Data Mining* bertujuan untuk mengekstraksi informasi yang signifikan yang tersembunyi di dalam kumpulan data yang besar[8].

2.6. Clustering

Clustering atau pengelompokan merupakan metode yang dimanfaatkan untuk mengelompokkan rangkaian data menjadi beberapa kelompok berdasarkan karakteristik yang telah ditetapkan sebelumnya. Kelompok (*Cluster*) adalah kumpulan objek data yang memiliki kesamaan di dalam satu kelompok dan perbedaan dengan objek-objek dalam kelompok lainnya. Objek data dapat tergabung dalam satu atau lebih kelompok sehingga objek-objek yang berada dalam satu kelompok akan memiliki kesamaan yang tinggi di antara satu dengan yang lainnya[9].

2.7. Algoritma K-Means

K-Means merupakan salah satu algoritma *Clustering* yang terkenal dan sederhana. Algoritma ini bertujuan mengelompokkan data menjadi sejumlah klaster yang telah ditentukan sebelumnya. Tahapan awal *K-Means* dimulai dengan menetapkan jumlah klaster yang diinginkan (k) dan memilih titik awal yang berfungsi sebagai *Centroid* dari setiap klaster. Selanjutnya, setiap data akan dikelompokkan ke dalam klaster yang memiliki *Centroid* terdekat dengan data tersebut. Setelah pengelompokan seluruh data, posisi *Centroid* di setiap klaster akan diperbaharui berdasarkan nilai rata-rata dari semua data yang termasuk ke dalam klaster tersebut. Proses ini akan diiterasi beberapa kali hingga tidak ada perubahan dalam pengelompokan data yang terjadi[10].

Salah satu teknik pengelompokan data tidak beraturan (*border*) yang dapat membagi data menjadi dua kumpulan atau lebih adalah *K-Means*. Dengan strategi ini, data akan dibagi menjadi beberapa kategori, dengan data yang berbagi kategori serupa ditempatkan dalam satu grup dan data yang memiliki fitur berbeda ditempatkan di grup yang berbeda. Pengelompokan bertujuan untuk mengurangi fungsi tujuan yang telah ditetapkan, umumnya ini bertujuan untuk meningkatkan keragaman antar kelompok tanpa meminimalkan variasi dalam pengelompokan[11].

Metode *K-Means* mampu mengidentifikasi jumlah klaster (k) yang optimal. Klaster (k) dihasilkan dengan menggabungkan sejumlah besar objek yang

memiliki karakteristik serupa melalui penggunaan algoritma *K-Means* yang merupakan teknik yang relatif sederhana[12].

2.8. DBI (Davies Bouldin Index)

Davies Bouldin Index (DBI) adalah metode validasi kluster yang dikembangkan oleh D.L. Davies. *DBI* adalah suatu fungsi yang mengukur rasio distribusi di dalam kluster terhadap pemisahan antar kluster. Penggunaan *DBI* bertujuan untuk mengoptimalkan jarak antar kluster. *DBI* digunakan untuk mengevaluasi kevalidan data pada masing-masing kluster. Sebuah kluster dianggap memiliki pengelompokan yang optimal apabila memperoleh nilai Indeks Davies Bouldin yang minimal[3].

Nilai kluster yang optimal bisa diperoleh dengan menemukan nilai *Davies Bouldin Index (DBI)* yang mendekati nol[13].

2.9. RapidMiner

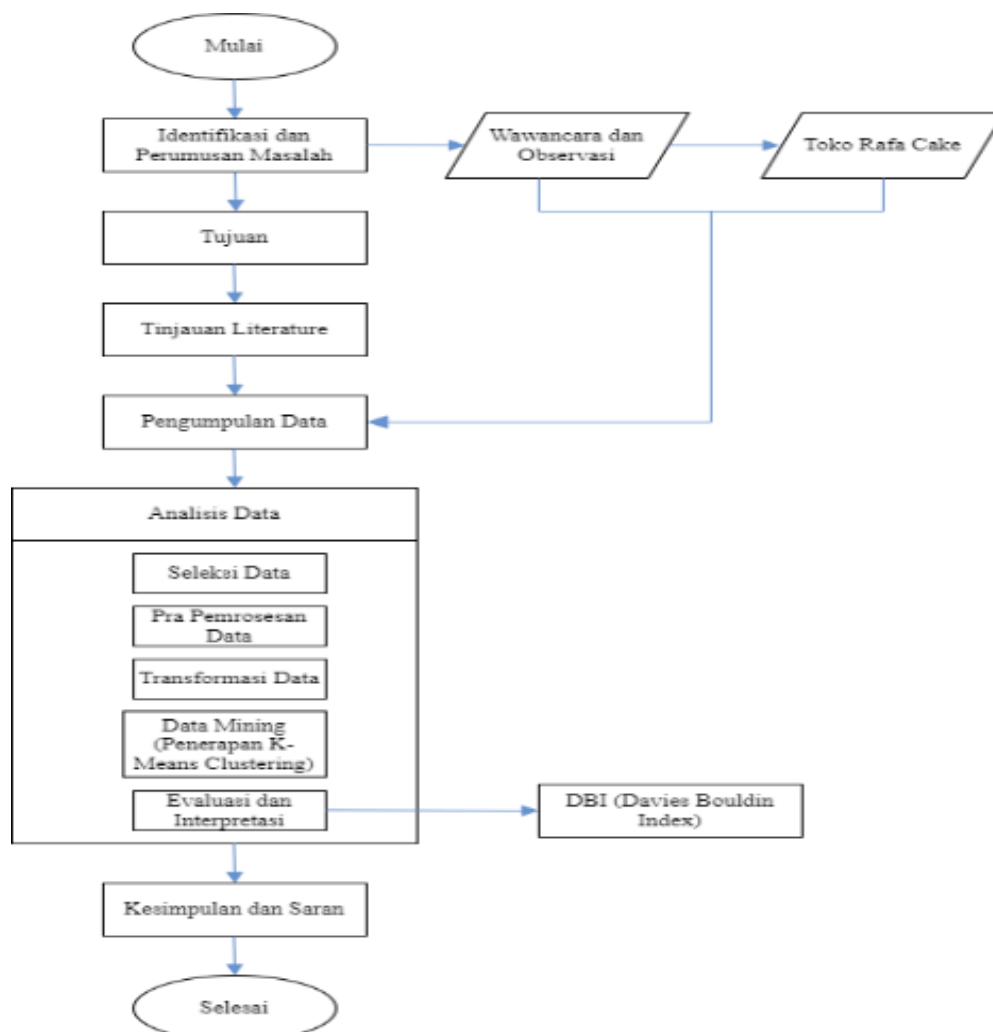
RapidMiner ialah sebuah perangkat lunak yang digunakan untuk menerapkan berbagai teknik dan metode dalam pengolahan *Data Mining*. Tujuan utamanya adalah mengubah data menjadi informasi yang bermanfaat dalam konteks pengambilan

keputusan. Fitur-fitur *RapidMiner* mencakup prosedur dalam bidang *Data Mining* dan pembelajaran mesin, termasuk ekstraksi, transformasi, pengisian data, pra-pemrosesan data, visualisasi, pemodelan, serta evaluasi data. Kemudahan yang ditawarkan *RapidMiner* terletak pada kemampuannya untuk melakukan perhitungan data dalam skala besardengan memanfaatkan operator-operator yang ada di dalamnya. Operator-operator ini berperan dalam proses ekstraksi data untuk menghasilkan informasi yang diinginkan. *RapidMiner* juga memungkinkan tampilan visual dari hasil pengujian data melalui grafik-grafik yang disajikan[3].

3. METODE PENELITIAN

3.1. Metode Penelitian

Dalam konteks ini, metode yang digunakan adalah pendekatan kuantitatif dengan penerapan algoritma *K-Means*. Algoritma ini diterapkan melalui serangkaian langkah-langkah yang mencakup tahapan mulai, identifikasi dan perumusan masalah, wawancara dan observasi, tujuan, tinjauan literature, pengumpulan data, analisis data, evaluasi dan interpretasi, kesimpulan dan saran, selesai. Tahapan model penelitian ditunjukkan pada gambar 1



Gambar 1. Tahapan metode penelitian

3.2. Sumber Data

Dalam penelitian ini, peneliti menggunakan sumber data primer yang berupa data transaksi toko Rafa Cake yang terdiri dari 8 (delapan) atribut seperti no, nama produk, kategori produk, variasi, harga, merek, stok yang terjual, dan total transaksi penjualan. Data primer ini diperoleh dari pemilik toko Rafa Cake pada saat observasi dan wawancara di Toko Rafa Cake, yang berlokasi di Jl. Kandang Perahu Gang Jati No.42, Karyamulya, Kec. Kesambi Kota Cirebon.

3.3. Populasi dan Sampel

Populasi dalam studi ini mencakup pada keseluruhan data transaksi penjualan kue di toko Rafa Cake yang menjadi fokus penelitian. Dalam dataset tersebut populasi nya sebanyak 1313 record data transaksi penjualan kue selama periode satu tahun terakhir, yaitu dari bulan Juni 2022 hingga Juli 2023 dengan 8 (delapan) parameter diantaranya no, nama produk, kategori produk, variasi, harga, merek, stok yang terjual, dan total transaksi penjualan. Dengan mengambil seluruh total data transaksi untuk dijadikan sampel, analisis dalam penelitian ini bertujuan untuk mengambil kesimpulan yang lebih pasti tentang pola pembelian dan preferensi pelanggan selama periode waktu tersebut.

Sampel yang digunakan dalam penelitian ini adalah sampel jenuh (census). Dengan menggunakan sampel jenuh, peneliti akan memeriksa dan menganalisis seluruh transaksi yang ada dalam populasi tertentu tanpa melakukan pengambilan sampel yang lebih kecil.

3.4. Teknik Pengumpulan Data

Berikut teknik pengumpulan data yang dilakukan dalam penelitian:

3.4.1. Observasi

Observasi dilakukan secara langsung di tempat penelitian, bertujuan untuk memperoleh wawasan mendalam mengenai sejumlah proses penting terkait. Pengamatan ini mencakup pemantauan rinci terhadap tahapan-tahapan yang terlibat dalam pemrosesan data produk, termasuk tahapan pengumpulan, pengolahan, serta penyimpanan data tersebut. Observasi juga merangkum pengamatan terkait ketersediaan stok produk, mencakup aspek pengelolaan dan penataan stok, dengan fokus pada keberlanjutan dan ketersediaan produk secara keseluruhan.

3.4.2. Wawancara

Proses pengumpulan data dengan tahapan wawancara mengusung metode pengajuan pertanyaan sebagai cara utama dalam memperoleh informasi yang relevan. Wawancara difokuskan pada pemilik Rafa Cake sebagai narasumber utama, Proses ini melibatkan pertanyaan yang dirancang untuk mengeksplorasi pengetahuan dan pengalaman langsung dari pemilik toko Rafa Cake. Dari interaksi ini, peneliti bisa mendapatkan pemahaman mendalam tentang praktik

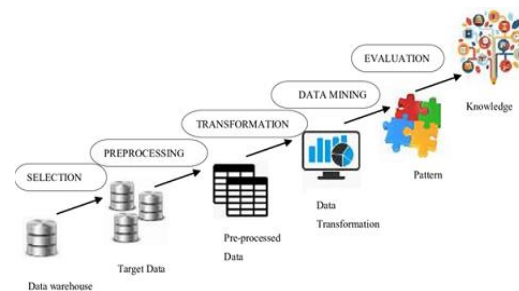
bisnis, strategi penjualan, tantangan yang dihadapi, dan cara pengelolaan stok produk oleh pemilik toko.

3.4.3. Studi Literature

Pengumpulan data melalui studi literatur mencakup eksplorasi dan analisis menyeluruh terhadap sumber-sumber tertulis seperti buku, artikel jurnal, dan teori-teori terkait dengan topik penelitian.

3.5. Teknik Analisis Data

Teknik analisis data dalam penelitian ini dengan menerapkan proses *KDD* (*Knowledge Discovery in Database*) seperti yang terlampir pada gambar 2.



Gambar 2. Tahapan proses KDD[12]

Dari gambar tahapan proses KDD pada gambar 2 dapat diuraikan sebagai berikut:

1. Data Selection

Langkah awal dilakukan dengan *Data Selection*, yaitu pemilihan data yang relevan. Data yang dipilih adalah data transaksi penjualan dari toko Rafa Cake, tercakup dari bulan Juni 2022 hingga Juli 2023, dengan total 1313 data transaksi. Data ini terdiri dari 8 atribut yang meliputi nomor, nama produk, kategori produk, variasi, harga, merek, stok terjual, dan total nilai transaksi penjualan.

2. Preprocessing

Tahap ini peneliti melakukan pra-pemrosesan data yang komprehensif. Ini termasuk pembersihan data dengan memeriksa data dari *Missing Value* dan pemilihan atribut yang relevan (select atribut). Data dinormalisasi untuk memastikan skala yang konsisten dan format data diubah jika diperlukan agar siap digunakan dalam analisis.

3. Transformasi

Tahap selanjutnya adalah transformasi data. Pada tahap ini data yang telah diproses disesuaikan dengan format yang tepat untuk penerapan Algoritma *K-Means*. Peneliti memilih atribut yang akan digunakan dalam analisis dan mengubahnya ke dalam type data yang sesuai dengan ketentuan algoritma.

4. Data Mining

Selanjutnya, implementasi teknik *Data Mining K-Means Clustering* guna mengelompokkan data berdasarkan kategori yang serupa. Algoritma ini diterapkan pada data transaksi toko untuk mengidentifikasi produk mana yang paling diminati, cukup diminati dan kurang diminati oleh

pelanggan menggunakan *RapidMiner*. Berikut diagram alir algoritma *K-Means* ditampilkan pada gambar 3.



Gambar 3. Diagram alir algoritma *K-Means*

Gambar 3 diagram alir di atas menggambarkan serangkaian langkah dalam algoritma *K-Means*, yaitu sebagai berikut:

1. Tentukan jumlah *Cluster*
Dari 1313 data transaksi kue akan dikelompokkan ke dalam 3 cluster yaitu, *Cluster* paling diminati, *Cluster* cukup diminati dan *Cluster* kurang diminati.
2. Tentukan nilai *Centroid* atau pusat *Cluster*
Dari 1313 data transaksi kue akan dipilih 3 pusat *Cluster* “*Centroid*” Menghitung jarak antara objek data dengan pusat *Cluster* Jumlah data transaksi kue akan dihitung ke pusat *Cluster* yang sudah ditentukan menggunakan model Euclidean yang dirumuskan sebagai berikut: [1]

$$d_{ij} = \sqrt{(x_{1i} - x_{1j})^2 + (x_{2i} - x_{2j})^2 + \dots + (x_{ki} - x_{kj})^2} \quad (1)$$

Dengan keterangan:

d_{ij} = Jarak data ke - i ke titik pusat (*Cluster j*).

x_{ki} = Data ke - i atribut data ke - k.

x_{kj} = Data ke - j atribut data ke - k.

3. Evaluation

Evaluasi adalah tahapan akhir yang membicarakan tentang hasil. Hasil tersebut kemudian akan di evaluasi dan di validasi menggunakan *DBI* (Davies Bouldin Index) untuk mengukur kualitas klasterisasi, Semakin rendah nilai *DBI*, semakin baik klasterisasi tersebut. Hasil ini akan memberikan informasi tentang produk apa saja yang paling diminati, cukup diminati dan kurang diminati oleh pelanggan berdasarkan kategori produk, jumlah penjualan dan harga produk pada data transaksi toko. *DBI* dapat dihitung dengan rumus berikut : [3]

$$R_{ij} = \frac{Var(C_i) - Var(C_j)}{||C_i - C_j||}$$

$$R_i = \max_{j=1, \dots, k, i \neq j} R_{ij}$$

$$DB = \frac{1}{K} \sum_{i=1}^K R_i$$

$$Var(x) = \frac{1}{N-1} \sum_{i=1}^N (x_i - \bar{x})^2$$

(2)

Keterangan:

R : jarak antar Cluster

Var : variance dari data

K : data ke-i

k : rata-rata dari tiap Cluster

DB : validasi Davies Bouldin

4. HASIL DAN PEMBAHASAN

4.1. Data

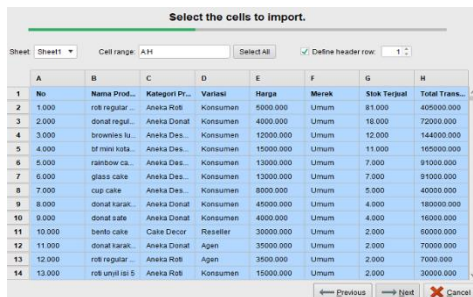
Dataset yang digunakan dalam penelitian ini berupa data primer yang bersumber dari Toko Rafa Cake yang berlokasi di Jl. Kandang Perahu Gang Jati No.42, Karyamulya, Kec. Kesambi Kota Cirebon. Dataset tersebut dihasilkan dalam kurun waktu satu tahun, terhitung dari Juni 2022 sampai dengan Juli 2023 yang memuat sebanyak 1313 record dan 8 atribut diantaranya no, nama produk, kategori produk, variasi, harga, merek, stok yang terjual, dan total transaksi penjualan. Data tersebut dalam bentuk dokumen soft file format Microsoft Excel seperti pada tampilan tabel 1 berikut ini:

Tabel 1. Data transaksi kue

No	Nama Produk	kategori produk	Variasi	Harga	Merek	Stok Terjual	Total Transaksi Penjualan
1	roti regular all varian	Aneka Roti	Konsumen	5000	Umum	81	405000
2	donat regular all varian	Aneka Donat	Konsumen	4000	Umum	18	72000
3	brownies lumer 300ml	Aneka Dessert	Konsumen	12000	Umum	12	144000
4	bf mini kotak/bulet	Aneka Dessert	Konsumen	15000	Umum	11	165000
5	rainbow Cake segitiga	Aneka Dessert	Konsumen	13000	Umum	7	91000
6	glass Cake	Aneka Dessert	Konsumen	13000	Umum	7	91000
7	cup Cake	Aneka Dessert	Konsumen	8000	Umum	5	40000
8	donat karakter/topping isi 12	Aneka Donat	Konsumen	45000	Umum	4	180000
9	donat sate	Aneka Donat	Konsumen	4000	Umum	4	16000
10	bento Cake	Cake Decor	Reseller	30000	Umum	2	60000
...	...	...	...	...	...	...	...
1313	Cake segitiga rainbow	Aneka Dessert	Agen	11000	Umum	5	55000

4.2. Data Selection

Langkah pertama yang dilakukan adalah seleksi data yang akan dilakukan dalam analisis yaitu dengan memilih operator "Read Excel" dalam RapidMiner untuk melakukan Import data yang dipilih untuk analisis. Pemilihan ini disebabkan oleh format dataset yang digunakan, yaitu dalam bentuk file Excel. Proses langkah tersebut ditunjukkan pada gambar 4.



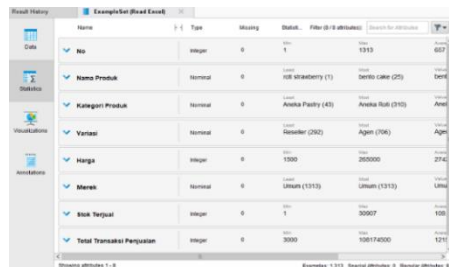
Gambar 4. Data Selection

4.3. Data Preprocessing

Tahap selanjutnya adalah Data Preprocessing atau pembersihan data dengan memeriksa data dari Missing Value, pemilihan atribut yang relevan (Select Attributes) dan menetapkan atribut nomor produk sebagai identifikasi (ID) utama menggunakan operator (Set Role).

1) Missing Value

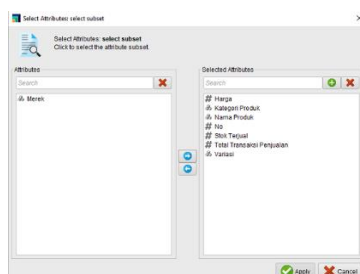
Proses Missing Value dilihat pada gambar 5.



Gambar 5. Missing Value

2) Select Attribute

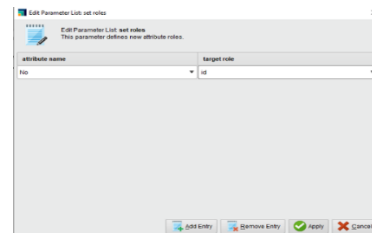
Proses Select Attribute ditampilkan pada gambar 6.



Gambar 6. Select Attribute

3) Set Role

Proses Set Role direpresentasikan pada gambar 7.



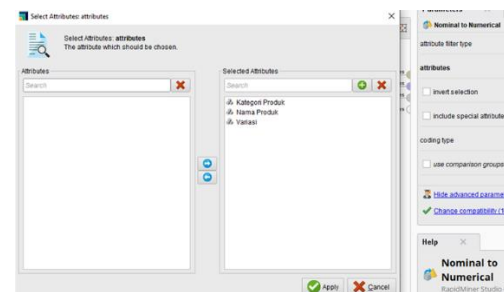
Gambar 7. Set Role

4.4. Data Transformasi

Tahap transformasi adalah kegiatan mengubah atribut ke dalam type data yang sesuai dengan ketentuan algoritma K-Means untuk analisis lebih lanjut menggunakan operator Nominal To Numerical, Kemudian data dinormalisasi untuk memastikan skala yang konsisten menggunakan operator normalize.

1) Nominal To Numerical

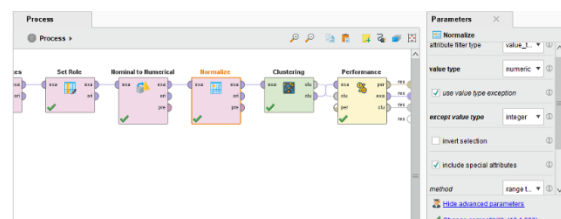
Proses Nominal To Numerical dilihat pada gambar 8.



Gambar 8. Nominal To Numerical

2) Normalize

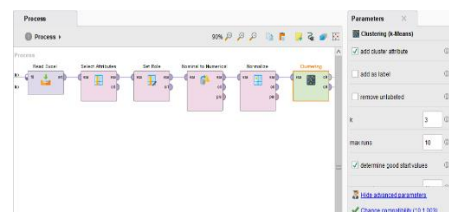
Proses Normalize dilihat pada gambar 9.



Gambar 9. Normalize

4.5. Data Mining

Tahap berikutnya setelah proses transformasi data adalah menerapkan model algoritma K-Means Clustering dalam platform RapidMiner yang ditampilkan pada gambar 10.



Gambar 10. K-Means Clustering

Pada pengaturan parameter Clustering K-Means, parameter nilai K menandakan jumlah Cluster yang

ditentukan dalam proses *Clustering*. Dalam kasus ini, peneliti menentukan 3 *Cluster* ($k=3$) sebagai acuan yang ditentukan dikarenakan penelitian ini akan mengelompokkan data ke dalam 3 kelompok berdasarkan karakteristik yang serupa guna mengidentifikasi produk dengan kategori paling diminati, cukup diminati dan kurang diminati. Pengelompokan ini dilakukan dengan proses max run sebanyak 10 kali percobaan, dengan pengukurannya menggunakan tipe Numerical Measures.

Setelah menjalankan proses tersebut, pengelompokan dataset kue berdasarkan hasil penjualan dari Toko Rafa Cake di *RapidMiner* menghasilkan 3 kelompok data atau model *Cluster* dapat dilihat pada gambar 11.

Cluster Model

```
Cluster 0: 1272 items
Cluster 1: 1 items
Cluster 2: 40 items
Total number of items: 1313
```

Gambar 11. *Cluster model*

Selanjutnya, dalam proses penentuan jumlah *Cluster* optimal, dilakukan sepuluh kali percobaan untuk menentukan nilai K yang optimal. Percobaan berhenti pada iterasi keenam, alasan percobaan dihentikan setelah iterasi keenam adalah karena pada percobaan keenam nilai K yang optimal berhasil diidentifikasi. Hasilnya menunjukkan bahwa $K=3$ pada percobaan kedua merupakan nilai K optimal untuk jumlah klaster dalam analisis klasterisasi dengan nilai *DBI* paling minimal dibandingkan percobaan lainnya yaitu 0.044. Informasi ini direpresentasikan dalam tabel hasil percobaan yang menunjukkan nilai-nilai Davies Bouldin Index terkait dengan berbagai nilai K yang diuji coba pada proses klasterisasi dapat direpresentasikan dalam tabel 2.

Tabel 2. Penentuan k optimal

Cluster (K)	Davies Bouldin Index (<i>DBI</i>)
2	0.076
3	0.044
4	0.051
5	0.058
6	0.062

Berikut ditampilkan hasil visualisasi dari K optimal pada gambar 12.

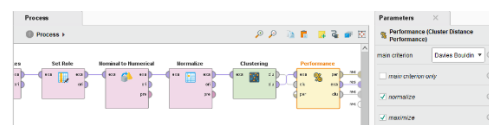


Gambar 12. Visualisasi grafik k optimal

Dalam analisis hasil *Clustering* tersebut, terlihat bahwa *Cluster 1* memiliki 1 item, mencerminkan dataset yang menunjukkan kategori penjualan paling diminati dengan nilai transaksi tinggi. Pada sisi lain, *Cluster 2* terdiri dari 40 item, menggambarkan dataset dengan kategori penjualan cukup diminati dan nilai transaksi sedang. Sementara *Cluster 0* memiliki 1272 item, mewakili dataset dengan kategori penjualan yang kurang diminati dan nilai transaksi rendah.

4.6. Evaluation

Langkah selanjutnya adalah evaluasi hasil dari proses *Clustering* tersebut dengan tujuan mengukur mutu hasil klaster. Untuk melakukan evaluasi ini, Pengujian dilaksanakan dengan menggunakan operator kinerja *Cluster Distance Performance* dapat dilihat pada gambar 13.



Gambar 13. Pengujian performance

Hasil dari evaluasi menampilkan nilai *Davies-Bouldin Index* sebesar 0.044. Nilai ini menunjukkan bahwa hasil dari proses *Clustering* dengan $K=3$ memiliki nilai *Davies-Bouldin Index* paling efisien karena memiliki angka *DBI* yang paling minimal atau mendekati 0 di antara percobaan lainnya. Seperti yang terlihat pada gambar 14 dibawah ini :

Davies Bouldin

Davies Bouldin: 0.044

Gambar 14. Hasil pengujian performance nilai *DBI*

Hasil dari evaluasi performa menggunakan performance vektor pada *RapidMiner* menunjukkan bahwa nilai jarak rata-rata *Centroid* dari tiap klaster berbeda, dengan jarak rata-rata *Centroid* antara ketiga klaster sebesar 598909122519.887. Dari nilai jarak rata-rata *Centroid* untuk masing-masing klaster, dapat disimpulkan bahwa *Cluster 1* memiliki nilai paling optimal yaitu 0.000, yang menunjukkan pemisahan klaster yang sangat baik. Berikut hasil tersebut ditampilkan pada gambar 15.

PerformanceVector

```
PerformanceVector:
Avg. within centroid distance: 598909122519.887
Avg. within centroid distance_cluster_0: 266610943894.785
Avg. within centroid distance_cluster_1: 0.000
Avg. within centroid distance_cluster_2: 11180963930861.139
Davies Bouldin: 0.044
```

Gambar 15. Hasil evaluasi jarak pada *performance vector*

4.7. Knowledge

Klaster 1: Terdapat 1 item dengan kategori produk paling diminati dengan nilai transaksi tinggi, menandakan produk unggulan dengan permintaan tinggi di pasaran.

Tabel 3. Produk *Cluster 1*

Cluster (K)	ID	Nama Produk	Kategori Produk	Stok Terjual	Harga	Kategori
Cluster 1	73	roti regular all varian	Aneka Roti	30907	3500-5000	Paling diminati

Klaster 2: Terdapat 40 item dengan kategori produk cukup diminati dengan nilai transaksi sedang, menunjukkan adanya peluang untuk meningkatkan

daya tarik produk ini melalui strategi pemasaran yang lebih cermat. Berikut produk produk klaster 2 dilihat pada tabel 4.

Tabel 4. Produk *Cluster 2*

Cluster (K)	ID	Nama Produk	Kategori Produk	Stok Terjual	Harga	Kategori
Cluster 2	74	roti regular all varian	Aneka Roti	12002	3500-5000	Cukup diminati
	80	mille crepes	Aneka Dessert	2096	12000-15000	Cukup diminati
	...	...	...	...	...	
	813	Cake potong special	Cake Decor	157	75000-95000	Cukup diminati

Klaster 0: Terdapat 1272 item dengan kategori produk kurang diminati dengan nilai transaksi rendah, memerlukan strategi pemasaran khusus atau evaluasi

lebih lanjut untuk meningkatkan daya tarik di pasar. Berikut produk klaster 0 ditampilkan pada tabel 5.

Tabel 5. Produk *Cluster 0*

Cluster (K)	ID	Nama Produk	Kategori Produk	Stok Terjual	Harga	Kategori
Cluster 0	9	donat sate	Aneka Donat	4	3000– 4000	Kurang diminati
	6	glass Cake	Aneka Dessert	7	11000-15000	Kurang diminati
	...	...	...	...	...	
	1313	Cake segitiga rainbow	Aneka Dessert	5	11000	Kurang diminati

5. KESIMPULAN DAN SARAN

Berdasarkan penelitian, penggunaan algoritma *K-Means* pada data transaksi kue di Toko Rafa Cake membentuk tiga klaster utama yang mengelompokkan produk berdasarkan kategori, tingkat penjualan, dan harga. Klaster 1 menunjukkan produk paling diminati (1 item), Klaster 2 menampilkan produk cukup diminati (40 item), sementara Klaster 0 menunjukkan produk kurang diminati (1272 items). Evaluasi menggunakan *DBI* menegaskan pemisahan klaster yang baik dengan nilai mendekati 0 (0.044) pada tiga klaster ($K=3$), dengan K optimal terdapat pada *Cluster 1* dan jarak rata-rata *Centroid* sebesar 0.000. Saran untuk penelitian selanjutnya adalah melakukan pengembangan dengan menerapkan algoritma lain seperti *K-medoid* atau *K-Prototype* serta menggunakan dataset yang mencakup rentang waktu yang lebih luas dan dengan atribut yang lebih beragam untuk studi yang lebih komprehensif.

DAFTAR PUSTAKA

- [1] A. Fachriansyah and E. Bu, "Penerapan Algoritma K-Means Untuk Clustering Bakery Dan Cake Yang Laris," vol. 4, no. 2, pp. 205–217, 2023, doi: 10.47065/bit.v3i1.675.
- [2] D. Ariestiany and T. Santoso, "Analisis Pengelolaan Stok Kue Menggunakan Algoritma Apriori Pada Toko Kue Sponji," *Zo. J. Sist. Inf.*, vol. 4, no. 2, pp. 167–183, 2022, doi: 10.31849/zn.v4i2.11128.
- [3] A. Muhidin and S. K. Alfandara, "Analisis Data Mining Menggunakan Algoritma K-Means Untuk Clustering Penjualan Studi Kasus Dapur Bu Ipung," *Sigma J. Teknol. Pelita Bangsa*, vol. 13, no. 2, pp. 75–80, 2022.
- [4] A. Padhil, A. Saleh, M. Nusran, H. Ismail, and A. Aidel, "Strategi Pemasaran Kue Bagea Sebagai Kue Khas Kota Palopo Dengan Menggunakan Analisis Swot Pada Ukm Multi Kue Kota Palopo," *J. Ind. Eng. Manag.*, vol. 3, no. 2, pp. 15–23, 2018, doi: 10.33536/jiem.v3i2.230.
- [5] S. A. Wahyono and L. C. Grace, "Substitusi tepung kacang hijau dalam pembuatan bolu pandan kukus," *J. Pariwisata Vokasi*, vol. 4, no. 1, pp. 12–32, 2023.
- [6] S. Firman, Slamet Widodo, and Haerani, "Tanggapan Masyarakat Terhadap Bolu Cukke' Dengan Substitusi Tepung Tempe," *Sinergitas Multidisiplin Ilmu Pengetah. dan Teknol.*, vol. 1, no. 1, pp. 97–103, 2018.
- [7] S. A. Simbolon and K. Sumatupang, "Analisis Proses Produksi Roti Tawar," *J. Ilmu Komputer, Ekon. dan Manaj.*, vol. 3, no. 2, pp. 2195–2204, 2023.
- [8] N. F. Adani *et al.*, "Implementasi Data Mining Untuk Pengelompokan Data Penjualan Berdasarkan Pola Pembelian Menggunakan Algoritma K-Means Clustering Pada Toko Syihan," *J. Cyber Tech*, vol. x, no. x, pp. 1–11, 2019, [Online]. Available: <https://ojs.trigunadharma.ac.id/index.php/jct/article/view/4648%0Ahttps://ojs.trigunadharma.ac.id/index.php/jct/article/download/4648/791>
- [9] S. Aulia, "Klasterisasi Pola Penjualan Pestisida Menggunakan Metode K-Means Clustering (Studi Kasus Di Toko Juanda Tani Kecamatan Hutabayu Raja)," *Djtechno J. Teknol. Inf.*, vol. 1, no. 1, pp. 1–5, 2021, doi: 10.46576/djtechno.v1i1.964.
- [10] M. R. Nahjan, N. Heryana, A. Voutama, F. I. Komputer, U. S. Karawang, and R. Miner,

- “IMPLEMENTASI RAPIDMINER DENGAN METODE CLUSTERING K-MEANS UNTUK ANALISA PENJUALAN PADA TOKO OJ CELL,” vol. 7, no. 1, pp. 101–104, 2023.
- [11] D. A. Manalu and G. Gunadi, “Implementasi Metode Data Mining K-Means Clustering Terhadap Data Pembayaran Transaksi Menggunakan Bahasa Pemrograman Python Pada Cv Digital Dimensi,” *Infotech J. Technol. Inf.*, vol. 8, no. 1, pp. 43–54, 2022, doi: 10.37365/jti.v8i1.131.
- [12] I. Pii, N. Suarna, and N. Rahaningsih, “Penerapan Data Mining Pada Penjualan Produk Pakaian Dameyra Fashion Menggunakan Metode K-Means Clustering,” *JATI (Jurnal Mhs. Tek. Inform.*, vol. 7, no. 1, pp. 423–430, 2023, doi: 10.36040/jati.v7i1.6336.
- [13] N. P. Gantara and I. Ali, “Penerapan Metode K-Means Clustering Pada Penjualan Barang Di Sports Station,” *E-Link J. Tek. Elektro dan Inform.*, vol. 18, no. 1, p. 28, 2023, doi: 10.30587/e-link.v18i1.5339.