Jurnal Informatika: Jurnal pengembangan IT

Vol. 9, No. 3, 2024
ISSN : 2477-5126, DOI:10.30591/jpit.v9ix.xxx

Segmentasi Pembelian Produk Menggunakan Algoritma K-Means Berdasarkan Clusterisasi pada pemilihan menu yang ada diUMKM Kuliner

Lolanda Hamim Annisa [1], Dian Rusvinasari [2]
[1, 2]Universitas Putra Bangsa, Ronggowarsito 18, Kebumen, 54361, Indonesia

Info Artikel

Riwayat Artikel:

Received 2024-02-10
Revised 2024-08-10
Accepted 2024-08-18

Corresponding Author:

Lolanda Hamim Annisa

Email: lolandaannisa@gmail.com

Abstract *– Marketing strategy can be seen as one of the bases used in preparing comprehensive SME planning. One of the SMEs that will be highlighted in this research is restaurants. Customer loyalty is an important thing that must be maintained by companies for the sustainability of the company and can improve good relationships between service provider companies and their customers. K-Means Cluster Analysis is a non-hierarchical cluster analysis method that attempts to partition existing objects into one or more clusters or groups of objects based on their characteristics, so that objects that have the same characteristics are grouped in the same cluster and objects that have similar characteristics. different groups are grouped into other clusters. The purpose of this research is to find out how to group menus that have high selling power and also the relationship between one menu variable and another menu when a transaction or purchase occurs by a customer. The results obtained in this research were to create a segmentation of products purchased by customers in the period May-November 2023 in SMEs operating in the culinary sector in the Central Java area. The results showed that the types of products most frequently purchased were Chicken Rice, Tea, Chicken, & White Rice which is at the highest purchase order. Where Chicken Rice was purchased 5409 times, Tea 1867 times, White Rice 1452 times, Chicken 1110 times. In the K-Means Algorithm to determine which products sell frequently and require more inventory and which do not*

Keywords: *K-Means; Clustering; SMEs; Culinary.*

Abstrak *– Strategi pemasaran dapat di pandang sebagai salah satu dasar yang di pakai dalam menyusun perencanaan UMKM secara menyeluruh. Salah satu UMKM yang akan diangkat pada penelitian ini adalah rumah makan. Loyalitas pelanggan adalah hal penting yang harus dijaga oleh perusahaan demi keberlangsungan perusahaan serta dapat meningkatkan hubungan yang baik antara perusahaan penyedia jasa dengan para pelanggannya. K-Means Cluster Analysis merupakan salah satu metode cluster analysis non hirarki yang berusaha untuk mempartisi objek yang ada kedalam satu atau lebih cluster atau kelompok objek berdasarkan karakteristiknya, sehingga objek yang mempunyai karakteristik yang sama dikelompokan dalam satu cluster yang sama dan objek yang mempunyai karakteristik yang berbeda dikelompokan kedalam cluster yang lain. Tujuan dari penelitian ini adalah untuk mengetahui bagaimana bisa melakukan pengelompokkan terhadap menu yang memiliki daya jual tinggi dan juga keterkaitan satu variabel menu dengan menu yang lain saat terjadi transaksi atau pembelian oleh pelanggan. Hasil yang telah didapatkan pada penelitian ini adalah membuat segmentasi produk yang dibeli oleh pelanggan dalam kurun waktu Mei-November 2023 di UMKM yang bergerak di bidang kuliner di daerah Jawa Tengah didapatkan hasil berupa jenis produk yang paling sering dibeli adalah Nasi Ayam, Teh, Ayam, & Nasi Putih yang berada pada urutan pembelian tertinggi. Dimana Nasi Ayam dibeli sebanyak 5409 kali, Teh 1867 kali, Nasi Putih 1452 kali, Ayam 1110 kali. Dalam Algoritma K-Means untuk menentukan produk mana yang sering laku dan memerlukan lebih banyak persediaan dan mana yang tidak*

Kata Kunci*: K-Means, Clustering, UMKM, Kuliner.*

I. PENDAHULUAN

Strategi pemasaran adalah salah satu cara untuk memenangkan keunggulan bersaing yang berkesinambungan baik itu untuk perusahaan yang memproduksi barang atau jasa. Strategi pemasaran dapat di pandang sebagai salah satu dasar yang di pakai dalam menyusun perencanaan UMKM secara menyeluruh. Dipandang dari luasnya permasalahan yang ada dalam suatu usaha, maka di perlukan adanya perencanaan yang menyeluruh untuk dijadikan pedoman bagi segmen suatu usaha dalam menjalankan kegiatanya, alasan lain yang menunjukan pentingnya strategi pemasaran adalah semakin kerasnya persaingan perusahaan pada umumnya [1]. Loyalitas pelanggan adalah hal penting yang harus dijaga oleh perusahaan demi keberlangsungan perusahaan serta dapat meningkatkan hubungan yang baik antara perusahaan penyedia jasa dengan para pelanggannya. Pelanggan yang loyal akan memberi keuntungan bagi perusahaan karena pelanggan yang loyal secara tidak langsung dapat berkontribusi dalam memperkenalkan produk atau jasa yang telah

mereka rasakan kepada keluarga atau rekannya [2]. Segmentasi konsumen mengelompokkan data konsumen menjadi kelompok-kelompok konsumen tertentu dimana masing-masing kelompok memiliki karakteristik yang serupa yang terkait dengan pemasaran produk, seperti umur, jenis kelamin, kecenderungan memilih produk tertentu ataupun seberapa banyak uang yang dikeluarkan konsumen dalam berbelanja. Segmentasi konsumen menjadi cara yang handal dalam mengidentifikasi pangsa pasar suatu perusahaan. Dengan menggunakan data konsumen perusahaan kemudian dapat merencanakan strategi seperti meningkatkan layanan, memberikan promosi yang menarik atau mengembangkan produk baru. Mampu mengelompokkan konsumen memungkinkan perusahaan untuk lebih memahami konsumen lebih jauh lagi mengetahui siapa saja yang menjadi target konsumen dengan cara yang lebih efisien [3]. Sistem Pendukung Keputusan adalah sebuah sistem berbasis model yang terdiri dari prosedur - prosedur dalam pemrosesan data dan pertimbanganya untuk membantu dalam mengambil sebuah keputusan. Untuk itu sebuah perusahaan/usaha memerlukan sistem pendukung keputusan untuk menentukan customer segment dan segment pasar yang lebih signifikan agar bisa lebih mengetahui tentang prospek peminat dari produk perusahaan/usaha tersebut dan terdata rapih untuk mengetahui bahwa rata rata customer segment dan segment pasar yang biasa didapatkan itu lebih banyak kemana dan pendapatan omsetnya lebih berpengaruh besar kemana agar bisa ditetapkan menjadi strategi / target utama untuk kedepanya [4].

K-Means *Cluster Analysis* merupakan salah satu metode cluster analysis non hirarki yang berusaha untuk mempartisi objek yang ada kedalam satu atau lebih cluster atau kelompok objek berdasarkan karakteristiknya, sehingga objek yang mempunyai karakteristik yang sama dikelompokan dalam satu cluster yang sama dan objek yang mempunyai karakteristik yang berbeda dikelompokan kedalam cluster yang lain [5]. Perubahan kebutuhan pelanggan mendorong terjadinya perubahan dalam bidang pemasaran. Bagian pemasaran perusahaan memiliki peran penting dalam menghadapi persaingan yang semakin ketat dengan perusahaan lain. Perusahaan yang berorientasi pasar atau dengan kata lain menjual produk ke pelanggan, umumnya akan menghadapi masalah di bidang pemasaran. Diperlukan survei pasar untuk mendapatkan informasi tentang permintaan dan kebutuhan spesifik pelanggan. Perusahaan harus mempertimbangkan karakteristik setiap pelanggan. Informasi itu digunakan untuk mengembangkan produk yang mengandung kombinasi dan atribut optimal yang diinginkan oleh pelanggan [6].

Segmentasi terus menjadi konsep pemasaran yang penting juga dalam konteks relationship marketing. Meningkatkan hubungan dengan pelanggan menjadi lebih menarik dan akan menghasilkan pemahaman yang lebih baik tentang kebutuhan pelanggan. Segmentasi adalah proses membagi pelanggan menjadi beberapa cluster dengan kategori loyalitas pelanggan untuk membangun strategi pemasaran. Segmentasi pelanggan adalah salah satu langkah awal dalam membuat model bisnis [7]. Menganalisis segmentasi pasar terhadap rumah makan menjadi sangat penting untuk memahami siapa yang akan menjadi pangsa pasar. Penting bagi pemilik rumah makan untuk mengetahui bagaimana sebuah kebiasaan dilokasi usaha untuk dapat mengetahui segmentasi [8].Pembuatan sistem informasi segmentasi pelanggan menggunakan metode clustering K-Means ini bertujuan untuk mengelola data pesanan dan data pelanggan yang ada pada suatu perusahaan perdagangan. Hal tersebut dilakukan untuk membantu perusahaan dalam melakukan strategi pemasaran guna mempertahankan pelanggan, baik pelanggan lama maupun baru. Pelanggan tersebut perlu dipertahankan dikarenakan sangat berpengaruh terhadap perusahaan, untuk itu perlu dilakukan proses segmentasi dengan mengelompokkan pelanggan yang memiliki kesamaan tertentu menjadi satu [9]. Menerapkan konsep CRM (*Customer Relationship Management*), perusahaan dapat melakukan identifikasi konsumen dengan melakukan segmentasi konsumen. Tujuan dari proses segmentasi konsumen adalah untuk mengetahui perilaku konsumen dan menerapkan strategi pemasaran yang tepat sehingga mendatangkan keuntungan bagi pihak perusahaan [10].

Data arsitektur dapat didefinisikan sebagai "rangkaian struktur, komponen, serta interaksi antara mereka yang membentuk kerangka kerja untuk pengelolaan, pengolahan, dan penggunaan data dalam suatu organisasi." Data arsitektur berperan penting dalam menyediakan landasan yang kokoh untuk sistem informasi, pemrosesan transaksi, data warehousing, analisis bisnis, dan inisiatif terkait data lainnya [8] Cluster yakni kumpulan data yang mirip, dengan perbedaan dibanding data yang berasal dari cluster yang lain. Diferensiasi utama diantara algoritma clustering dan klasifikasi yakni clustering tak punya target/kelas/label, sehingga merupakan pembelajaran tanpa pengawasan. Clustering dipakai menjadi fase awal di proses data mining dan cluster yang terbentuk bakal jadi input untuk algoritma selanjutnya dipakai. Identifikasi dan pengelompokkan gaya belajar siswa memerlukan pendekatan clustering, yaitu pengelompokkan data. Pengelompokkan ini mempunyai tujuan mengelompokkan objek menurut karakteristik antar objek. Dari analisis cluster, kita dapat menemukan grup yang memiliki karakteristik masing-masing grup [11] Clustering hirarki, menghubungkan atau memisahkan subset dari point-point dan bukan pointpoint individual, jarak antara point-point individu harus digeneralkan terhadap jarak antara subset. Ukuran kedekatan yang diperoleh disebut metrik berhubungan. Tipe mertik hubungan yang digunakan secara signifikan memperoleh algoritma hisrarki, karena merefleksikan konsep tertentu dari

kedekatan dan konektivitas [10]. Ada dua jenis data clustering yang sering dipergunakan dalam proses pengelompokan data yaitu hierarchical (hirarki) data clustering dan non-hierarchical (non hirarki) data clustering. Dalam cluster hirarki dimulai dengan membuat K cluster dimana setiap cluster beranggotakan satu objek dan berakhir dengan satu cluster dimana anggotanya K objek, pada setiap tahap prosedurnya, satu cluster digabung dengan satu cluster lain, lalu dapat dipilih klaster yang diinginkan dengan menentukan cut-off pada tingkat tertentu. Dalam non-hierarchical pengelompokan objek dimasukan ke dalam K cluster, dapat dilakukan dengan menentukan pusat klaster awal lalu dilakukan realokasi objek berdasarkan kriteria tertentu sampai dicapai pengelompokan yang optimum [2].

Metode yang digunakan penelitian ini adalah K-Means clustering, yang merupakan salah satu teknik data mining yang dapat digunakan untuk mengelompokkan berdasarkan kemiripan karakteristik tertentu dimana data-data yang memiliki kemiripan akan berada pada cluster yang sama [12] K-means merupakan metode clustering secara partitioning yang memisahkan data ke dalam kelompok yang berbeda. Dengan partitioning secara iteratif, K-Means mampu meminimalkan rata-rata jarak setiap data ke cluster-nya Terdapat beberapa ukuran jarak yang digunakan sebagai ukuran kemiripan suatu instance data [7] K-Means merupakan algoritma clustering yang berulang-ulang dengan menetapkan jumlah dari cluster (k) secara acak, dimana nilai tersebut menjadi pusat dari cluster untuk sementara. Pusat cluster biasa disebut dengan centroid, mean atau means [8] Model RFM dapat digunakan sebagai variabel/atribut data yang digunakan untuk proses clustering dapat menemukan jumlah cluster yang paling optimal. Selain itu, dalam penelitian ini algoritma k-means digunakan karena komputasi yang ringan, kemudian sesuai dengan kebutuhan data yang akan di cluster, dan sesuai dengan tujuan untuk menentukan jumlah cluster di awal untuk memperoleh jumlah kelompok pelanggan yang optimal. Hasil dari clustering dengan RFM model ini, menghasilkan hasil pengkasteran dengan berbagai karakteristik pelanggannya. Hal ini dapat dijadikan referensi dalam menentukan pemasaran sesuai dengan karakteristik pelanggannya. Algoritma K-Means menggunakan suatu persamaan untuk mengukur jarak yang paling minimum pada objek yaitu *Distance Space* atau *Euclidean Distance Space*. *Distance space* adalah suatu komponen dari K-Means Clustering yang digunakan untuk menghitung jarak antar objek [13]

Ilmu Data telah mengubah banyak hal dalam bidang ilmu yang salah satunya adalah bidang bioinformatika dari pengurangan dimensi kumpulan data yang sangat besar menjadi visualisasi data. Pasar bioinformatika dan kebutuhan profesi di bidang bioinformatika semakin berkembang setiap tahunnya. Dalam beberapa tahun mendatang, terdapat prediksi bahwa akan ada kebutuhan yang sangat besar bagi para profesional yang memiliki kemampuan bioinformatika [14] Dalam analisis data, algoritma pengelompokan sering digunakan untuk mengelompokkan titik data dengan fitur yang sama ke dalam cluster yang sama, sambil sebisa mungkin menghindari fitur serupa di cluster yang berbeda. Analisis klaster adalah membagi sekelompok data ke dalam beberapa kategori, sehingga semakin besar kemiripan antar kelompok data yang sama maka semakin baik, sedangkan kemiripan antar kelompok yang lain semakin kecil. Penyempurnaan algoritma clustering merupakan penyempurnaan dari algoritma clustering tradisional, antara lain metode pengukuran jarak clustering, optimalisasi algoritma clustering, dan analisis *multiobjective clustering*, guna meningkatkan kualitas dan akurasi hasil clustering [15] K-Means digunakan karena merupakan teknik pengelompokan yang dapat diterapkan secara luas yang berupaya membangun kluster yang berdekatan dan bekerja dengan baik dengan data berdimensi rendah. Ketika jumlah variabel sangat banyak, algoritma K-means mengungguli pengelompokan hierarki [16] Dalam implementasi K-means ini adalah digunakan sebagai pengelompok data dokumen pada saat mengembangkan sebuah perangkat lunak, pada tahap pertama diperlukan pengelompokan melalui penemuan potensi kesamaan antar dokumen. Kelompok dokumen yang dihasilkan (cluster dokumen) dapat digabungkan. Pendekatan seperti itu memungkinkan aplikasi untuk mengalokasikan kumpulan dokumen yang lebih menjanjikan demi kepentingan pengguna. Perangkat lunak pengguna dapat menjalankan penggabungan aglomeratif berdasarkan metode yang dipilih [17]

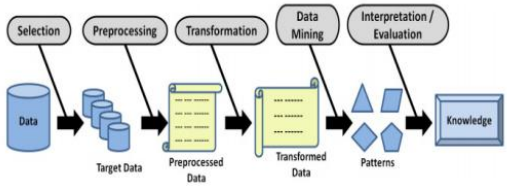
Penelitian yang dilakukan menggunakan algoritma K-Means dengan klasterisasi dalam mengolah data meghasilkan kelompok data yang akan dibagi pada beberapa kelompok. Pada penelitian ini, peneliti mencoba untuk *state of the art* yang menggunakan model sebagai focus utama. Model yang dimaksudkan pada penelitian ini adalah bagaimana para peneliti bisa memodelkan atau mempersiapkan data, untuk bisa menghasilkan data yang sesuai dengan kebutuhan dan menjawab rumusan masalah pada penelitian ini. Namun dalam penelitian ini berfokus pada pemodelan data/persiapan data yang telah disesuaikan dengan kebutuhan output yang diinginkan oleh peneliti. Dengan demikian diharapkan gap tersebut dapat menjadi *state of the art* pada penelitian ini.

II. METODE

Objek penelitian ini menggunakan data dari UMKM yang bergerak pada bidang kuliner di daerah Jawa Tengah dalam kurun waktu Mei 2023-November 2023 Laporan transaksi pembelian.

A. Knowledge Discovery in Databases (KDD)

Knowledge Discovery in Databases (KDD) adalah suatu proses yang melibatkan identifikasi pola atau pengetahuan yang berharga dari data yang besar, kompleks, dan mungkin terstruktur. Proses ini memiliki beberapa tahapan yang membantu dalam memahami dan mengekstraksi informasi yang dapat digunakan untuk pengambilan keputusan atau pemahaman yang lebih baik tentang suatu fenomena. Berikut adalah tahapan umum dalam proses *Knowledge Discovery in Databases* (KDD) [18]:



Gambar 1. Proses Knowledge Discovery in Databases (KDD)
Sumber: [18]

1. Selection

Selection digunakan untuk menentukan variabel yang akan diambil agar tidak ada kesamaan dan terjadi perulangan yang tidak diperlukan dalam pengolahan data mining.

2. Preprocessing

Pada prepocessing terdapat dua tahap, yaitu sebagi berikut :

- Data Cleaning
  Menghilangkan data yang tidak diperlukan seperti menangani missing value, noise data serta menangani data – data yang tidak konsisten dan relevan.
- Data Integration
  Dilakukan terhadap atribut yang mengidentifikasikan entitas yang unik.

3. Transformation

Merubah data sesuai format ekstention yang sesuai dalam pengolahan data mining karena beberapa metode pada data mining memerlukan format khusus sebelum dapat diproses pada data mining.

4. Data mining

Proses utama pada metode yang diterapkan untuk mendapatkan pengetahuan baru dari data yang diproses. Pada penelitian ini diterapkan teknik clustering yaitu metode KMeans Clustering.

5. Evaluation/ Interpretation

Mengidentifikasi pola – polayang menarik kedalam knowdge base yang diidentifikasi. Pada tahap ini, menghasilkan pola – pola khas maupun model prediksi yang dievaluasi untuk menilai kajian yang ada sudah memenuhi target yang diinginkan.

6. Knowledge

Pola-pola yang dihasilkan akan dipresentasikan kepada pengguna. Pada tahapan ini pengetahuan baru yang dihasilkan bisa dipahami semua orang yang akan dijadikan acuan pengambilan keputusan.

B. Proses Algoritma K-Means

Tentukan k sebagai jumlah cluster yang akan dibentuk, menggunakan metode elbow criterion dengan rumus sebagai berikut:

1. Tentukan k titik pusat cluster (centroid) awal yang dilakukan secara random. Penentuan pada centroid awal dilakukan secara acak atau random dari objek yang etrsedia sebanyak k-cluster,untuk menghitung centroid cluster ke-i berikutnya, digunakan rumus sebagai berikut:
2. Menghitung jarak dari setiap objek ke masing – masing centroid dari masing – masing cluster menggunakan Euclidean Distance, dengan rumus sebagai berikut:
3. Perhitungan tersebut dapat diseleseikan dengan perhitungan manual, namun dalam penelitian ini akan dilakukan menggunakan aplikasi.
4. Aplikasi yang digunakan terbagi menjadi 2, yaitu aplikasi yang pertama tentang pengolahan & analisis data, dan aplikasi yang selanjutnya tentang visualisasi data.

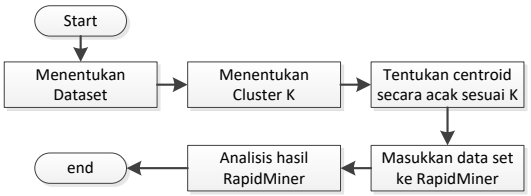
C. Analisis & Visualisasi Data

1. RapidMiner

Rapid miner adalah sebuah aplikasi atau perangkat lunak untuk mengolah sebuah dataset, rapid miner juga merupakan aplikasi yang bersifat *opensource* yang banyak diminati karna mudah untuk dipelajari dan rapidminer juga memiliki lisensi AGPL (GNU *Affero General Public License*) yang mampuh untuk mengolah data mining yang dikembangkan oleh Ralf Klinkenrberg, Ingo Mierswa, dan Simon Fischer di *Artificial Intelligence* Unit dari University of Dortmund.

2. Tableau

Ada banyak tools yang bisa kamu manfaatkan untuk data visualization, salah satunya yaitu Tableau. Tableau adalah perangkat lunak visualisasi data dan analisis bisnis yang memungkinkan pengguna untuk menghubungkan, memvisualisasikan, dan berbagi data dengan cara yang memfasilitasi pemahaman dan pengambilan keputusan. Tableau secara luas digunakan di berbagai industri untuk analisis data dan pelaporan, dan telah menjadi pilihan populer bagi organisasi yang ingin mengambil keputusan berbasis data.



Gambar 2. Alur Penelitian

III. HASIL DAN PEMBAHASAN

Penelitian ini menggunakan data yang berasal dari Laporan Transaksi Jual Beli pada salah satu UMKM yang bergerak di bidang kuliner di wilayah Jawa Tengah. Data yang digunakan memiliki rentang waktu antara Mei 2023 - November 2023, waktu ini ditentukan dari pertama kali UMKM ini meluncurkan produk hingga waktu pengambilan data pada penelitian ini. Parameter yang digunakan adalah nama produk, jumlah item produk yang dibeli setiap bulan dari Mei 2023 - November 2023. Proses cluster dilakukan untuk mengetahui pola banyaknya produk yang dibeli pada toko tersebut. Hal ini dilakukan bertujuan untuk mengetahui produk apa saja yang akan diproduksi lebih banyak sesuai dengan cluster terbesar pembelian produk yang ada pada toko tersebut. Data yang digunakan memilki transaksi sebnayak 8782 transaksi dan memiliki jumlah atribut sebnayak 13 atribut yang menggambarkan data transksi yang dibutuhkan oleh pengguna UMKM dalah melihat data transaksi yang ada pada tokonya selama kurun waktu buan Mei-Nov 2023 seperti terlihat dalam gambar 3 dibawah ini:

No. Struk	Tanggal	Bulan	Jam	Kode Produk	Produk	Jumlah Produk	Jumlah Dibatalkan	Harga Per Produk	Subtotal	Tipe Harga	Total	Status
321195L9	01-05-2023	05	19:18:08	23	Nasi Putih	3	0	3000	9000	Normal	9000	Transaksi
321195L9	01-05-2023	05	19:18:08	2	Ayam	2	0	7000	14000	Normal	14000	Transaksi
321195L9	01-05-2023	05	19:18:08	25	Rendang	1	0	9000	9000	Normal	9000	Transaksi
321183I7	01-05-2023	05	18:54:44	23	Nasi Putih	2	0	3000	6000	Normal	6000	Transaksi
321183I7	01-05-2023	05	18:54:44	25	Rendang	2	0	9000	18000	Normal	18000	Transaksi
321183I7	01-05-2023	05	18:54:44	12	Lele Goreng	2	0	8000	16000	Normal	16000	Transaksi
32118AVO	01-05-2023	05	18:53:42	23	Nasi Putih	1	1	3000	0	Normal	0	Dibatalkan
32118AVO	01-05-2023	05	18:53:42	25	Rendang	1	1	9000	0	Normal	0	Dibatalkan
32118AVO	01-05-2023	05	18:53:42	23	Nasi Putih	1	1	3000	0	Normal	0	Dibatalkan
32118AVO	01-05-2023	05	18:53:42	25	Rendang	1	1	9000	0	Normal	0	Dibatalkan
32118AVO	01-05-2023	05	18:53:42	12	Lele Goreng	1	1	8000	0	Normal	0	Dibatalkan
32118PBU	01-05-2023	05	18:31:36	23	Nasi Putih	1	0	3000	3000	Normal	3000	Transaksi
32118PBU	01-05-2023	05	18:31:36	2	Ayam	1	0	7000	7000	Normal	7000	Transaksi
32118SAS	01-05-2023	05	18:22:54	23	Nasi Putih	4	0	3000	12000	Normal	12000	Transaksi
32118SAS	01-05-2023	05	18:22:54	25	Rendang	4	0	9000	36000	Normal	36000	Transaksi
32118SAS	01-05-2023	05	18:22:54	24	Perkedel	4	0	4000	16000	Normal	16000	Transaksi
32117C5T	01-05-2023	05	17:59:10	23	Nasi Putih	1	0	3000	3000	Normal	3000	Transaksi
32117C5T	01-05-2023	05	17:59:10	25	Rendang	1	0	9000	9000	Normal	9000	Transaksi
321173R3	01-05-2023	05	17:39:56	2	Ayam	2	0	7000	14000	Normal	14000	Transaksi
321173R3	01-05-2023	05	17:39:56	25	Rendang	1	0	9000	9000	Normal	9000	Transaksi
321173R3	01-05-2023	05	17:39:56	26	Sambal Ijo	3	0	1000	3000	Normal	3000	Transaksi
321173R3	01-05-2023	05	17:39:56	6	Daun Singkong Rebus	3	0	1000	3000	Normal	3000	Transaksi
32117NCY	01-05-2023	05	17:22:16	2	Ayam	1	0	7000	7000	Normal	7000	Transaksi
32117NCY	01-05-2023	05	17:22:16	12	Lele Goreng	1	0	8000	8000	Normal	8000	Transaksi
32117Z8Z	01-05-2023	05	17:21:55	2	Ayam	1	0	7000	7000	Normal	7000	Transaksi
32117Z8Z	01-05-2023	05	17:21:55	12	Lele Goreng	1	0	8000	8000	Normal	8000	Transaksi

Gambar 3. Potongan Dataset

A. Data Cleaning

Pada tahap data cleaning ini yang dilakukan adalah membuang atribut yang tidak relevan atau tidak konsisten. Atribut yang dibuang dari dataset yang sudah ada yaitu no.struk, jam pembelian, harga satuan, total pembelian. Atribut yang digunakan yaitu kode produk, nama produk, jumlah pembelian, dan tanggal pembelian dataset yang telah di cleaning dapat dilihat pada tabel 1.

TABEL 1
DATA CLEANING

Kode	Nama Produk	Total Pembelian
1	Ati Ampela	130
2	Ayam	1110
3	Bawal	87
4	Ceker	25
5	Cincang	9
6	Daun Singkong Rebus	329
…..	…..	…..
…..	…..	…..
41	Susu Coklat	10
42	Susu Putih	13
43	Teh	1867

B. Data Integration

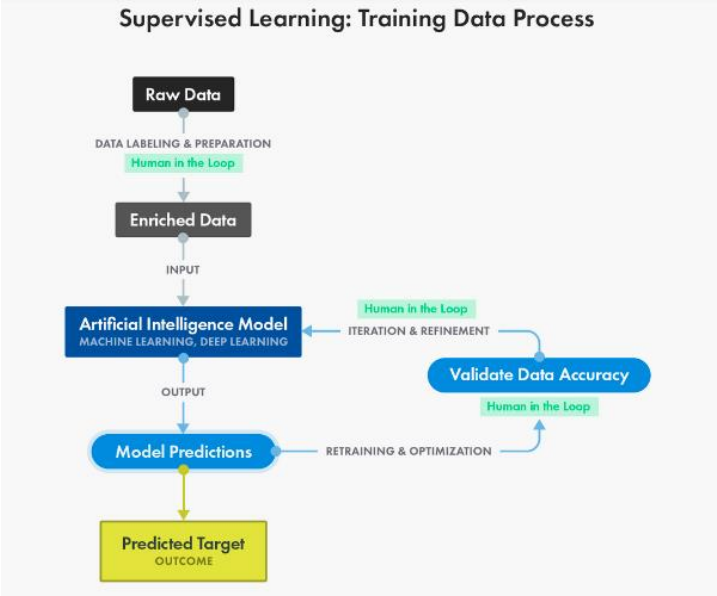
Pada tahap data integration ini dilakukan penggabungan data-data produk yang terdapat pada laporan transaksi bulanan selama pembukaan toko pada UMKM studi kasus penelitian. Pada atribut data-data produk akan berintegrasi atribut banyaknya total pemakaian. Untuk pembelian produk yang tidak ada jumlah pembeliannya akan diberi nilai 0, seperti yang terlihat di tabel 2

TABEL 2
DATA INTEGRATION

Kode	Nama Produk	Total Penjualan
1	Ati Ampela	130
2	Ayam	1110
3	Bawal	87
4	Ceker	25
5	Cincang	9
6	Daun Singkong Rebus	329
….	….	….
….	….	….
41	Susu Coklat	10
42	Susu Putih	13
43	Teh	1867

C. Data Training

Data Training merujuk pada kumpulan data yang digunakan untuk melatih suatu model dalam konteks pembelajaran mesin. Dalam proses pembelajaran mesin, model dipelajari dari data historis atau data training untuk dapat membuat prediksi atau mengidentifikasi pola pada data yang baru. Ini adalah data yang telah dikumpulkan atau dihasilkan sebelumnya dan digunakan sebagai input untuk melatih model. Data ini umumnya memiliki label atau hasil yang sudah diketahui yang digunakan untuk mengajar model bagaimana melakukan tugas tertentu. Dalam data training pada penelitian ini adalah pencarian cluster dari data transaksi yang telah ada, data dapat dilihat pada gambar 4.

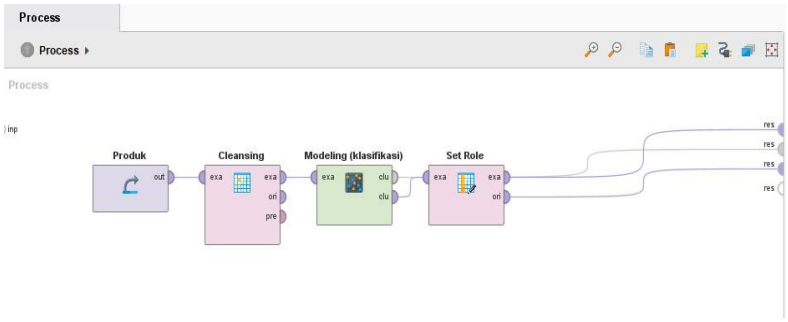


Gambar 4. Data Trining
(Sumber: https://www.cloudfactory.com/training-data-guide)

D. Data Modeling

Model data adalah representasi struktural dari suatu sistem informasi atau konsep bisnis dalam bentuk yang dapat dimengerti dan diolah oleh komputer. Data modeling (pemodelan data) adalah proses menciptakan model data ini untuk menggambarkan hubungan antara data, informasi, bisnis, dan aplikasi. Tujuannya adalah untuk memahami, menggambarkan, dan mengorganisir data agar dapat diakses, dikelola, dan dimanfaatkan secara efektif. Data modeling dapat dilakukan menggunakan berbagai teknik dan notasi, seperti *Entity-Relationship Diagrams* (ERD), UML (*Unified Modeling Language*), atau bahkan menggunakan bahasa formal seperti SQL (Structured Query Language) untuk mendefinisikan struktur basis data. Pemodelan data sangat penting dalam pengembangan perangkat lunak dan desain basis data karena membantu mengklarifikasi persyaratan bisnis, memudahkan pemahaman struktur data, dan membantu tim pengembang dalam membangun sistem yang efisien dan mudah dielola.

Dalam pengujian ini menggunakan rapid miner dengan operator cleansing data, k-means, dan operasi classification. Dimana 3 operator ini digunakan untuk menganalisa data dan melakukan data integration. Seperti yang telah dilakukan pada langkah sebelumnya data integration adalah melakukan analisa data yang memiliki kolerasi atau analisa data yang akan di temukan keterhubungannya. Berikut adalah proses dari modelling data dimana modeling ini akan memberikan gambaran pengolahan data yang akan dilakukan. Data modelling ini, memerlukan beberapa operasi yang digunakan untuk mendukung analisis yang akan dilakukan pada dataset. Gambar Data Modelling akan disajikan pada gambar 5 dibawah ini:



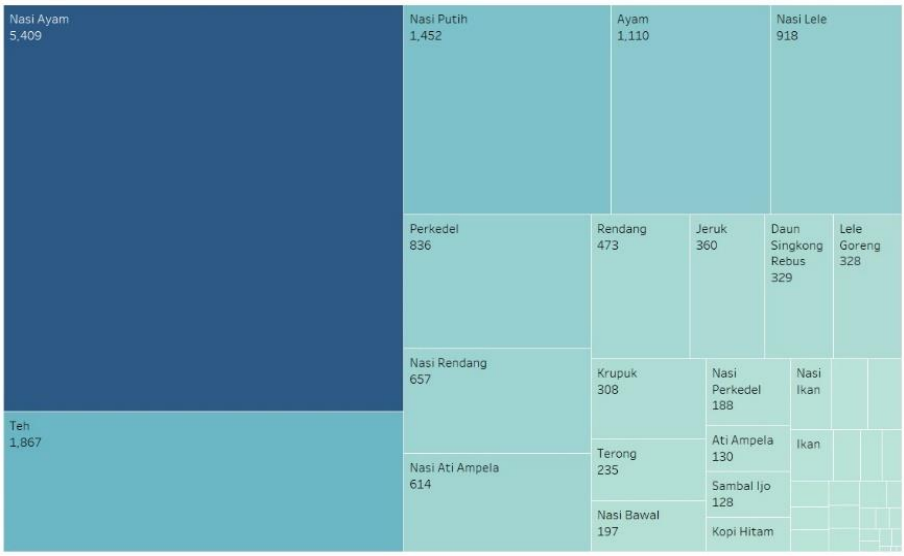
Gambar 5. Data Modelling pada RapidMiner

Data Modelling pada penelitian ini menempuh 3 langkah penggunaan operation pada aplikasi RapidMiner. Pada Langkah yang pertama setelah pemilihan atribut pada dataset maka data harus memiliki kesesuaian dalam tipe data, data yang dapat diproses salah satunya harus memiliki tipe data yang sama. Pada

aplikasi Rapidminer ldapat menggunakan beberapa jenis tipe data, salah satunya pada penelitian ini menggunakan binominal, dimana data yang dapat diproses adalah dengan tipe data integer & karakter.

Langkah yang selanjutnya adalah melakukan modeling, dimana modeling ini adalah memilih algoritma yang akan digunakan. Pada penelitian ini menggunaka algoritma k-means yang akan berujung pada klasifikasi data atau clusterisasi data. Pada tahapan ini, peneliti dapat memilih operator untuk algoritma yang akan diterapkan. Setelah memilih operator k-means maka peneliti dapat menghubungkan antara dataset, data cleaning, dan algoritma yang akan digunakan. Penelitian ini berfokus pada bagaimana memodelkan sebuah implementasi algoritma dari sebuah dataset yang nantunya akan disesuaikan dengan tujuan penelitian. Pada penelitian ini, peneliti menguraikan bagaimana persiapan data untuk dapat digunakan sebagai data modeling yang sesuai dengan tujuan penelitian.

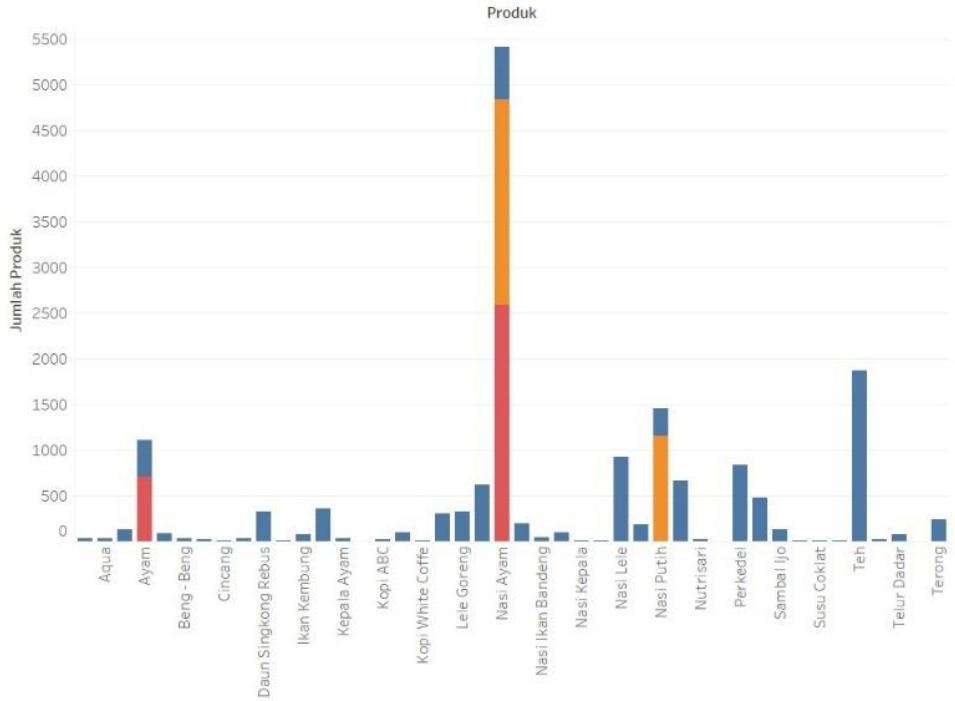
Langkah terkahir pada data modeling di penelitian ini adalah memilih hasil data yang ingin dilihat atau ditunjukkan. Dalam proses tersebut peneliti dapat memilik operation yang diinginkan, pada penelitian ini peneliti memilih untuk menggunakan klasifikasi maka dapat di setting demikian dengan memilih menggunakan set role. Setelah data modelling selesai untuk pemodelan maka peneliti sudah dapat running model tersebut dan menghasilkan clusterisasi sesuai dengan yang diinginkan oleh peneliti. Pengelompokan data penjualan dapat dilihat pada gambar 6 yang menunjukkan treemap untuk data produk penjualan.



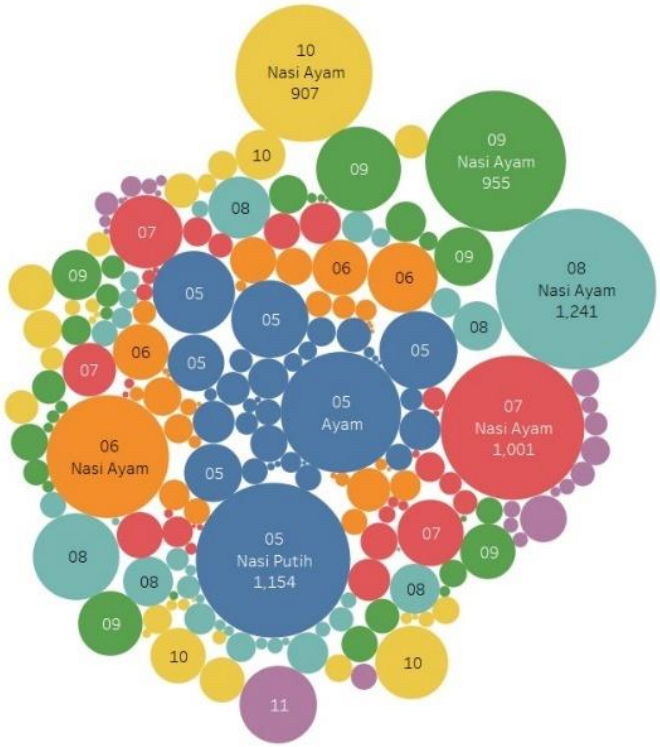
Gambar 6. Treemap data penjualan

E. Data Visualization

Data Visualization (Visualisasi Data) adalah representasi grafis dari informasi dan data. Tujuannya adalah untuk membantu orang memahami konteks, menganalisis pola, dan mengambil wawasan dari data yang kompleks atau besar. Dengan menggunakan elemen visual seperti grafik, diagram, peta, dan tata letak visual lainnya, data visualization mempermudah pengguna untuk menyajikan, membandingkan, dan memahami informasi dengan lebih cepat dan efektif daripada hanya dengan menggunakan teks atau angka. Alat visualisasi data seperti Tableau, Power BI, dan Python libraries seperti Matplotlib dan Seaborn sangat umum digunakan untuk membuat visualisasi data yang interaktif dan informatif. Pada penelitian ini untuk aplikasi visualisasi data menggunakan Tableau untuk menggambarkan visualisasi data cluster. Pada penelitian ini, peneliti menggunakan Tableau untuk memvisualisasikan cluster data. Ada beberapa bentuk visualisasi data dalam data penjualan diatas, dapat dilihat pada gambar 7 & 8.



Gambar 7. Visualisasi Data untuk klusterisasi penjualan produk



Gambar 8. Visualisasi Data untuk gambaran perbandingan penjualan produk

Visualisasi data yang ditunjukkan pada gambar 8 memberikan gambaran banyaknya produk yang terjual. Pada laporan transaksi penjualan akan menampilkan jumlah produk yang dijual sesuai besaran bulatan pada visualisasi data di gambar 8. Semakin besar bulatan pada gambar 8 maka semakin banyak pula angka penjualan untuk produk tersebut. Sedangkan untuk nada sewarna tetapi besaran bulatan tidak sama makahal tersebut menggambarkan bahwa produk yang dibeli adalah sama namun dibeli pada waktu yang berbeda atau pada hari yang berbeda atau pada bulan yang berbeda.

IV. SIMPULAN

Rangakaian penelitian yang telah dilakukan dan memberikan penjelasan pada akhir penelitian maka dapat disimpulkan dari hasil penelitian membuat segmentasi produk yang dibeli oleh pelanggan dalam kurun waktu Mei-November 2023 di UMKM yang bergerak di bidang kuliner di daerah Jawa Tengah didapatkan hasil berupa jenis produk yang paling sering dibeli adalah Nasi Ayam, Teh, Ayam, & Nasi Putih yang berada pada urutan pembelian tertinggi. Dimana Nasi Ayam dibeli sebanyak 5409 kali, Teh 1867 kali, Nasi Putih 1452 kali, Ayam 1110 kali. Dalam Algoritma K-Means untuk menentukan produk mana yang sering laku dan memerlukan lebih banyak persediaan dan mana yang tidak. Pemilihan variabel dan atribut yang akan digunakan memiliki dampak yang signifikan terhadap data yang digunakan. Pengembangan model database yang dipilih akan sangat memberikan dampak signifikan dalam hasil analisis. Sistem ini dari sistem analisis menggunakan aplikasi Ms.Excel menjadi pada aplikasi analisis expert yaitu dengan menggunakan RapidManer dan Tableau dan dilakukan dari perspektif masalah yang timbul dari sistem yang lama. Berikut adalah saran yang dapat dipertimbangkan untuk masa depan; Pada penelitian ini, bereksperimen dengan salah satu metode clustering data mining yang ada yaitu algoritma K-means. Untuk kedepannya perlu ditambahkan terkait dengan pengembangan model pada RapidMiner dan pengembangan model perancangan aplikasi.

UCAPAN TERIMAKASIH

Ucapan terima kasih kepada Universitas Putra Bangsa yang telah memberikan dana hibah penelitian kepada penulis. Dan ucapan terima kasih kepada UMKM yang telah bersedia menjadi studi kasus untuk terselenggaranya pengembangan penelitian ini.

DAFTAR PUSTAKA

[1] S. R. Arifen, V. D. Purwanty, D. A. Suci, R. H. Agustiawan, And A. R. Sudrajat, “Analisis Strategi Pemasaran Untuk Meningkatkan Daya Saing Umkm.”
[2] A. Pramudiansyah And H. Munte, “Segmentasi Pelanggan Menggunakan Algoritma K-Means Berdasarkan Model Recency Frequency Monetary,” Vol. 7, No. 2, 2021, [Online]. Available: Http://Ejournal.Fikom-Unasman.Ac.Id
[3] K. Auliasari Et Al., “Penerapan Algoritma K-Means Untuk Segmentasi Konsumen Menggunakan R,” 2019.
[4] B. Dewayana, A. Purno, W. Wibowo, And I. Artikel, “Menentukan Customer Segment Dan Segment Pasar Umkm (Foodendez) Dengan Metode Decision Tree,” Jurnal Informatika, Vol. 8, No. 2, 2021, [Online]. Available: Http://Ejournal.Bsi.Ac.Id/Ejurnal/Index.Php/Ji
[5] S. Wira Hadi, M. Fahmi Julianto, S. Rahmatullah, W. Gata, And S. Nusa Mandiri, “Bianglala Informatika Analisa Cluster Aplikasi Pada App Store Dengan Menggunakan Metode K-Means,” Vol. 8, No. 2, P. 2020.
[6] B. E. Adiana, I. Soesanti, And A. E. Permanasari, “Analisis Segmentasi Pelanggan Menggunakan Kombinasi Rfm Model Dan Teknik Clustering,” No. 2, 2018, Doi: 10.21460/Jutei.2017.21.76.
[7] N. H. Harani, C. Prianto, And F. A. Nugraha, “Segmentasi Pelanggan Produk Digital Service Indihome Menggunakan Algoritma K-Means Berbasis Python,” Jurnal Manajemen Informatika (Jamika), Vol. 10, No. 2, Pp. 133–146, 2020, Doi: 10.34010/Jamika.V10i2.
[8] B. Eno Ketherin, A. Anjani Arifiyanti, A. Sodik, J. Sistem Informasi, And I. Teknologi Adhi Tama Surabaya, “Analisa Segmentasi Konsumen Menggunakan Algoritma K-Means Clustering.”
[9] I. Y. Kurniawati, “Segmentasi Pelanggan Menggunakan Clustering K-Means.”
[10] A. H. Lubis And P. Ganesha, “Model Segmentasi Pelanggan Dengan Kernel K-Means Clustering Berbasis Customer Relationship Management,” Jurnal & Penelitian Teknik Informatika, Vol. 1, No. 1, 2016.
[11] F. Handayani, “Aplikasi Data Mining Menggunakan Algoritma K-Means Clustering Untuk Mengelompokkan Mahasiswa Berdasarkan Gaya Belajar,” Jurnal Teknologi Dan Informasi, Doi: 10.34010/Jati.V12i1.
[12] K. Prabhakaran, J. Dridi, M. Amayri, And N. Bouguila, “Explainable K-Means Clustering For Occupancy Estimation,” In Procedia Computer Science, Elsevier B.V., 2022, Pp. 326–333. Doi: 10.1016/J.Procs.2022.07.041.
[13] A. Satriawan, R. Andreswari, And O. N. Pratiwi, “Segmentasi Pelanggan Telkomsel Menggunakan Metode Clustering Dengan Rfm Model Dan Algoritma K-Means Telkomsel Customer Segmentation Using Clustering Method With Rfm Model And K-Means Algorithm.”
[14] N. Iqbal And P. Kumar, “From Data Science To Bioscience: Emerging Era Of Bioinformatics Applications, Tools And Challenges,” In Procedia Computer Science, Elsevier B.V., 2022, Pp. 1516–1528. Doi: 10.1016/J.Procs.2023.01.130.
[15] N. Wang, “Design Of An Intelligent Processing System For Business Data Analysis Based On Improved Clustering Algorithm,” Procedia Comput Sci, Vol. 228, Pp. 1215–1224, 2023, Doi: 10.1016/J.Procs.2023.11.105.
[16] K. E. Setiawan, A. Kurniawan, A. Chowanda, And D. Suhartono, “Clustering Models For Hospitals In Jakarta Using Fuzzy C-Means And K-Means,” In Procedia Computer Science, Elsevier B.V., 2022, Pp. 356–363. Doi: 10.1016/J.Procs.2022.12.146.
[17] K. Sadowski, P. Wolski, And I. Czarnowski, “An Application For Collecting And Mining Reports Referring Vulnerabilities And Exposures In Physical Systems: A Comparative Study Of Selected Clustering Methods,” Procedia Comput Sci, Vol. 225, Pp. 2763–2772, 2023, Doi: 10.1016/J.Procs.2023.10.268.
[18] G. Gustientiedina, M. H. Adiya, And Y. Desnelita, “Penerapan Algoritma K-Means Untuk Clustering Data Obat-Obatan,” Jurnal Nasional Teknologi Dan Sistem Informasi, Vol. 5, No. 1, Pp. 17–24, Apr. 2019, Doi: 10.25077/Teknosi.V5i1.2019.17-24.