

Implementasi Algoritma K-Means Dalam Pengelompokan Data Penjualan CV. Widuri Menggunakan Orange

Susi Oktaviani¹, Agus Bahtiar²

Email: Susioktaviani73@gmail.com¹

Email: agusbahtiar038@gmail.com²

¹ STMIK IKMI Cirebon

Abstract

Sales is one of the company's activities to get profit or what is commonly called profit. To achieve the desired profit, companies make every effort to produce the best products for their consumers to enjoy. As is the case with CV Widuri, one of the companies engaged in making snacks or snacks such as pilus, macaroni, uril macaroni, basreng, round tofu, and seblak. Sales data can be utilized by grouping products that are easy to sell and difficult to sell, so that companies can minimize the increase in the number of leftover goods (BS). By using clustering techniques in data mining, companies can identify potential products by grouping the products produced. Basically, the K-Means clustering algorithm can be applied to the problem of understanding consumer behavior to identify opportunities to bring new products to market. K-Means algorithm can also be used to collect objects from many objects to make it easier to describe the properties and characteristics of each group. From the data obtained from calculations using the orange application, namely cluster 1 (C1) 104 transactions, cluster 2 (C2) 1 transaction, cluster 3 (C3) 139 transactions, cluster 4 (C4) 95 transactions, cluster 5 (C5) 102 transactions, cluster 6 (C6) 76 transactions, and cluster 7 (C7) 97 transactions. With the provisions of products that are easy to sell, namely pilus and products that are difficult to sell, namely pilus bantat.

Keywords; Clustering, Data Mining, K-Means, Orange, Sales.

Abstrak

Penjualan merupakan salah satu kegiatan perusahaan untuk mendapatkan keuntungan atau yang biasa disebut laba. Untuk mencapai keuntungan yang diinginkan, perusahaan berusaha semaksimal mungkin untuk menghasilkan produk yang terbaik untuk dinikmati oleh konsumennya. Seperti halnya yang terjadi pada CV. Widuri, salah satu perusahaan yang bergerak dalam pembuatan makanan ringan atau snack seperti pilus, makaroni, makaroni uril, basreng, tahu bulat, dan seblak. Dari data penjualan dapat dimanfaatkan dengan mengelompokkan produk yang mudah terjual dan sulit terjual, sehingga perusahaan dapat meminimalisir terjadinya peningkatan jumlah barang sisa (BS). Dengan menggunakan teknik clustering dalam data mining, perusahaan dapat mengidentifikasi produk potensial dengan mengelompokkan produk yang dihasilkan. Pada dasarnya, algoritma pengelompokan K-Means dapat diterapkan pada masalah pemahaman perilaku konsumen untuk mengidentifikasi peluang membawa produk baru ke pasar. Algoritma K-Means juga dapat digunakan untuk mengumpulkan objek dari banyak objek agar lebih mudah dalam mendeskripsikan properti dan karakteristik masing-masing kelompok. Dari data yang didapatkan dari perhitungan menggunakan aplikasi orange yaitu

cluster 1 (C1) 104 transaksi, cluster 2 (C2) 1 transaksi, cluster 3 (C3) 139 transaksi, cluster 4 (C4) 95 transaksi, cluster 5 (C5) 102 transaksi, cluster 6 (C6) 76 transaksi, dan cluster 7 (C7) 97 transaksi. Dengan ketentuan produk yang mudah laku yaitu pilus dan produk yang sulit laku yaitu pilus bantat.

Kata Kunci; Clustering, Data Mining, K-Means, Orange, Penjualan.

I. Pendahuluan

Penjualan merupakan salah satu kegiatan perusahaan untuk mendapatkan keuntungan atau yang biasa disebut profit. Untuk mencapai laba yang diinginkan, perusahaan berusaha semaksimal mungkin untuk menghasilkan produk yang terbaik untuk dinikmati oleh konsumennya. Perusahaan berusaha untuk memberikan produk yang berkualitas tinggi dan higienis, mulai dari pembuatan bahan baku sampai dengan peralatan produksi dan pengemasannya, sehingga produk yang dihasilkan merupakan produk yang terbaik. (Triyandana et al., 2022)

Data mining adalah sebuah proses yang dapat dilakukan untuk menemukan hubungan dari data yang belum diketahui oleh pengguna dengan menyajikannya dengan cara yang lebih mudah untuk dipahami agar dapat menjadi dasar dalam sebuah pengambilan keputusan. (Dharma et al., 2021). Dalam kegiatan jual beli ini sering terjadi yang namanya produk sisa, karna tidak semua customer butuh atau tertarik dengan produk yang dibuat oleh suatu perusahaan, oleh karna itu sering terjadi pengelompokan terhadap produk yang mudah laku dan sulit laku. Produk industri makanan memiliki elastisitas harga penawaran dan permintaan yang relatif tinggi dan cenderung tidak stagnan, sehingga industri makanan berpotensi menjadi mesin penggerak perekonomian. Berbeda dengan komoditas dan komoditas pertanian mentah yang memiliki elastisitas harga yang rendah, sehingga harga jatuh saat panen raya.

peneliti melakukan literatur terhadap penelitian-penelitian terdahulu yang masih relevan terhadap objek permasalahan dalam penelitian ini. Adapun penelitian terdahulu diantaranya yaitu (Penerapan K-Means pada Pengelompokan Penjualan Produk Smartphone) karya Fatimah Putri Arfani Hasibuan, Sumarno Sumarno, dan Iin Parlina, (Jurnal Sains Teknologi dan Sistem Informasi) Vol. 1 No. 1 (2021) 15-20, Lisensi: Creative Commons Attribution 4.0 Internasional (CC BY 4.0). Penelitian ini dilakukan oleh PT. Vivo yang mana dilakukan dengan menggunakan penganalisisan terhadap sejumlah lembaga riset pasar smartphone dunia, salah satunya Canalys. Canalys menjabarkan lima besar vendor smartphone global berturut-turut pada kuartal 2 tahun 2021 antara lain: Samsung (19%), Xiaomi (17%), Apple (14%), Oppo (10%), dan Vivo (10%). Penelitian ini menggunakan data transaksi yang dikelompokkan untuk memeriksa produk yang mudah dijual dan sulit dijual. Metode yang digunakan adalah metode clustering K-Means. Dengan hasil yang telah diperoleh akan dibagi menjadi 3 kelompok yaitu penjualan rendah, penjualan sedang, dan penjualan tinggi. Berdasarkan hasil Pengujian menggunakan metode K-Means terhadap data penjualan smartphone Vivo, diperoleh bahwa penjualan kelompok tertinggi hanya ada 1 data, yakni Vivo Y12 3+32GB.

Ada beberapa data yang menjadi sumber data dalam penelitian. Data tersebut menjelaskan bahwa penjualan sangat berpengaruh terhadap pendapatan atau laba perusahaan. Perusahaan harus dapat memastikan bahwa produk yang diproduksi akan laku dipasaran sesuai dengan kebutuhan dan minat customer. Penganalisisan ini dilakukan guna mengurangi angka kenaikan terhadap barang sisa (BS).

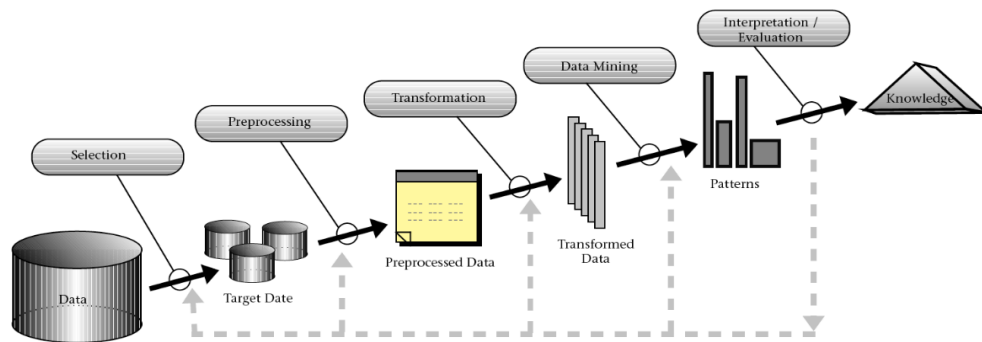
Dataset yang digunakan dalam penelitian ini merupakan hasil wawancara langsung terhadap pemilik dari CV. Widuri, data yang di peroleh berupa data penjualan produk selama 1 tahun yaitu dari bulan Januari-Desember 2022, yang terdiri dari 614 data. Dari data tersebut terdapat 6 atribut yang meliputi nomor, tanggal transaksi, nama barang, jumlah barang, harga satuan barang, dan total penjualan.

Ada beberapa masalah yang timbul akibat adanya pengolahan data penjualan secara manual. Sulitnya melakukan pengelompokan terhadap produk yang mudah laku dan produk yang sulit laku, yang di sebabkan oleh banyak nya produk makanan yang di produksi oleh CV. Widuri dan banyak nya buku-buku yang harus di catat serta panjang nya proses pengolahan data secara manual. Selain itu, adanya kemungkinan data-data yang hilang akibat kertas yang hilang, sobek,adan lainnya yang mengakibatkan penghambatan dalam pengolahan data.

Tujuan dilakukannya penelitian ini yaitu untuk mengatasi permasalahan-permasalahan yang muncul di CV. Widuri dengan melakukan pengolahan data penjualan secara otomatis menggunakan aplikasi dengan sistem data mining sehingga mempermudah pihak perusahaan dalam pengelompokan penjualan produk dan pengolahan data yang lebih efisien.

II. Metode Penelitian

Metode berisi tahapan atau prosedur penelitian dan algoritma yang digunakan dalam penelitian, formula permasalahan yang diteliti dengan lebih rinci, serta perancangan sistem juga dibutuhkan. Metode penelitian yang digunakan dalam penelitian ini menggunakan Knowledge Discovery in Database (KDD) adalah proses pengumpulan, penggunaan data untuk menentukan aturan atau pola tertentu dalam jumlah data yang sangat besar.



Gambar 1 Proses KDD

Fase penelitian ini mencakup desain dari awal hingga akhir. Fase ini adalah fase analisis dan desain model yang digunakan untuk pengelompokan, detail pada tahap ini adalah:

1. Seleksi Data

Tahap pertama dari proses data mining ini adalah penentuan database yang akan digunakan sesuai dengan permasalahan yang berkaitan dengan proses clustering. Data yang kami gunakan adalah data penjualan dari CV. Widuri dari bulan Januari-Desember 2022. Pemilihan data bertujuan untuk menyeleksi data, karena tidak semua data digunakan, maka hanya data yang sesuai untuk dianalisis saja yang diambil dari database.

2. Pre-Processing

Data pada tahap pre-processing merupakan tahap pembersihan data dari data yang mengandung outlier, inkonsistensi, duplikat, nilai yang hilang, atau data yang tidak lengkap. Kemudian data penjualan barang diperiksa kembali. Data yang duplikat atau tidak lengkap akan dihapus. (Ramadhan et al., 2021)

3. Data Transformasi

Pada tahap ini, dilakukan transformasi terhadap data distribusi komoditas yang diperoleh pada tahap pembersihan. Transformasi ini menginisialisasi variabel data distribusi menjadi kode awal yang unik dan bertujuan untuk mengimplementasikannya dalam perhitungan menggunakan metode k-means. Karena algoritma k-means hanya dapat menangani data numerik, maka variabel-variabel tersebut diinisialisasi dalam bentuk numerik ketika alat bantu digunakan untuk mengimplementasikannya.

4. Interpretasi/Evaluasi

Interpretasi dan evaluasi adalah tahap mengevaluasi hasil penelitian setelah data mining. Evaluasi dilakukan untuk mentransformasikan pola-pola yang dihasilkan oleh metode clustering ke dalam bentuk yang lebih mudah dipahami oleh pihak-pihak yang berkepentingan.

5. Algoritma K-Means

Algoritma K-Means adalah salah satu metode clustering non hirarki dengan cara mempartisi data yang ada ke dalam satu atau lebih cluster. Metode ini mempartisi data ke dalam cluster, sehingga daya yang memiliki karakteristik yang

sama akan berada dalam 1 cluster. Sedangkan Segmentasi gambar berarti mempartisi gambar ke dalam berbagai kelompok atau kelompok dengan fitur yang sama atau memiliki beberapa kesamaan. K-means merupakan metode pengelompokan partisi gambar kedalam kelompok sedemikian rupa sehingga setidaknya satu bagian dari cluster mengandung gambar dengan area utama dari bagian yang digunakan. (Suroyo, 2019)

6. Proses Algoritma K-Means

Tahapan dalam proses algoritma clustering menggunakan metode K-Means adalah: (Bakery, 2020)

- a. Memilih jumlah cluster k.
- b. Ada banyak cara untuk menginisialisasi pusat-pusat kluster dalam proses ini, tetapi pilihan utama pada tingkat kluster adalah acak.
- c. Alokasikan semua data/objek ke dalam kluster terdekat. Kami memproses data untuk setiap titik pusat kluster menggunakan teori jarak Euclidean, yang dirumuskan sebagai berikut:

$$d(x_i, \mu_j) = \sqrt{\sum (x_i - \mu_j)^2}$$

Dimana :

x_i : data kriteria,

μ_j : centroid pada cluster ke-j

- d. Proses selanjutnya adalah menghitung ulang pusat-pusat cluster menggunakan keanggotaan cluster terbaru. Rata-rata dari semua data/objek di dalam cluster adalah pusat cluster. Oleh karena itu, rata-rata bukanlah parameter yang lebih disukai untuk digunakan.
- e. Terlepas dari apakah pusat cluster berubah, petakan kembali setiap objek menggunakan pusat cluster yang baru sampai proses pengelompokan selesai, sampai Anda menemukan bahwa nilai pusat cluster tidak berubah. Ulangi proses 'c'.

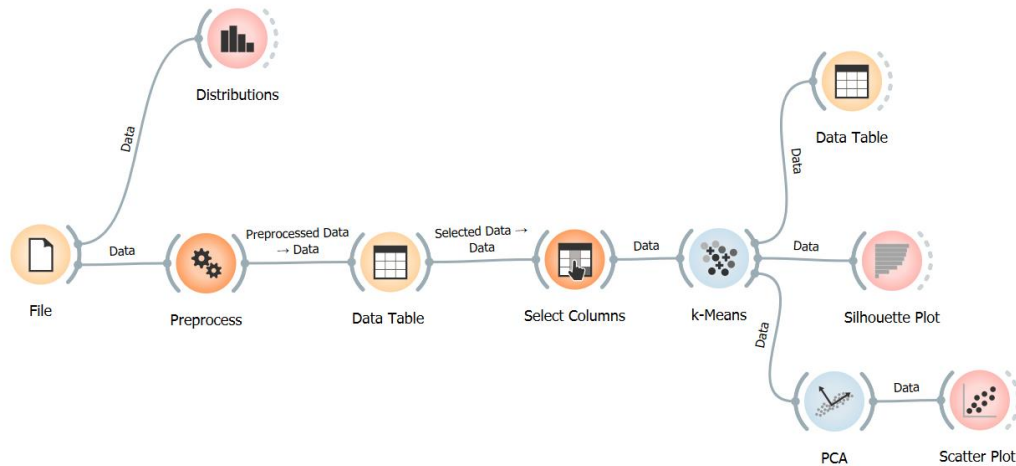
7. Orange

Orange Data Mining adalah paket perangkat lunak pemrograman visual berbasis komponen untuk visualisasi data pembelajaran mesin, penggalian data, dan analisis data. Pemrograman visual diimplementasikan melalui antarmuka yang menghubungkan widget yang telah ditentukan atau yang dirancang pengguna untuk membuat alur kerja. Orange menyediakan lingkungan serbaguna bagi para pengembang, peneliti, dan profesional data mining. Berkat Python, bahasa skrip dan lingkungan pemrograman generasi baru, skrip data mining menjadi sederhana, namun hasilnya berskala perusahaan. (Nawawi et al., 2021)

III. Hasil dan Pembahasan

1. Skenario Penelitian

Tampilan desain widget clustering K-Means dengan menggunakan Orange Data Mining seperti gambar berikut ini.



Gambar 2 Widget Clustering K-Means

Data yang diperoleh dari hasil wawancara akan diinput dan dianalisis berdasarkan objek. Kemudian akan dihubungkan menggunakan widget yang sesuai dengan apa yang akan digunakan dan dibutuhkan dalam penelitian ini .

2. Data Selection

Pada tahap selection ini akan menyeleksi atau memilih data yang diperlukan data tersebut kemudian disesuaikan dengan kebutuhan. Data dalam penelitian ini terdiri dari 6 atribut dengan jumlah data 614 record.

3. Preprosesing

Tahapan preprocessing data dalam penelitian ini menggunakan *widget* preprocessing data dengan melakukan *Normalize features* pada tools orange kemudian pilih *Normalize to interval [-1,1]*.

4. Data Table

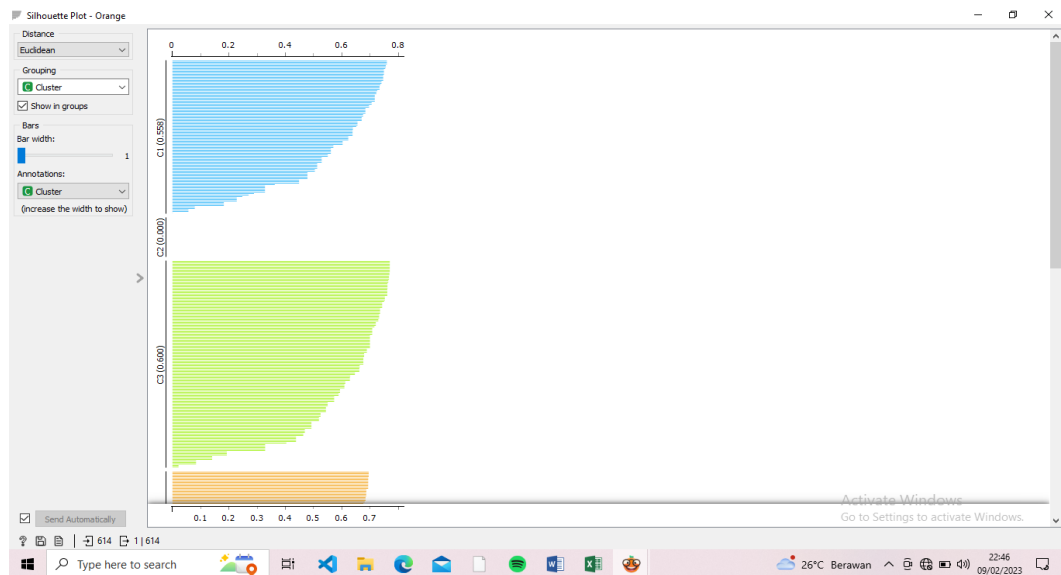
Data tabel merupakan hasil untuk menampilkan atribut mana saja yang dipilih untuk ditampilkan sebagai output dalam widget select column.

	TOTAL PENJ (2)	NO	TANGGAL	NAMA BARANG	TOTAL PENJ (1)	HARGA/BAL
1	Rp. 57.200.000	0.000	2022-01-03 00:00:00	makaroni 5kg	0.476481	0.34897
2	Rp. 39.769.000	0.00163	2022-01-03 00:00:00	pilus	0.2500	0.933066
3	Rp. 1.475.000	0.00326	2022-01-03 00:00:00	Kerupuk Seblak	0.019164	0.0915332
4	Rp. 612.000	0.00489	2022-01-03 00:00:00	Tahu Bulat	0.006098	0.104119
5	Rp. 57.200.000	0.00653	2022-01-04 00:00:00	makaroni 5kg	0.476481	0.34897
6	Rp. 30.915.000	0.00816	2022-01-04 00:00:00	pilus	0.128049	0.933066
7	Rp. 80.808.000	0.00979	2022-01-05 00:00:00	makaroni 5kg	0.674216	0.34897
8	Rp. 44.311.500	0.01142	2022-01-05 00:00:00	pilus	0.184669	0.933066
9	Rp. 32.151.600	0.01305	2022-01-06 00:00:00	pilus	0.133275	0.933066
10	Rp. 39.208.000	0.01468	2022-01-06 00:00:00	makaroni 5kg	0.325784	0.34897
11	Rp. 15.250.000	0.01631	2022-01-06 00:00:00	Makaroni Unil	0.106272	0.451945
12	Rp. 7.213.500	0.01794	2022-01-07 00:00:00	pilus	0.027875	0.933066
13	Rp. 62.400.000	0.01958	2022-01-07 00:00:00	makaroni 5kg	0.520035	0.34897
14	Rp. 39.000.000	0.02121	2022-01-08 00:00:00	Basreng 2,5kg	0.650697	0.0514874
15	Rp. 62.400.000	0.02284	2022-01-08 00:00:00	makaroni 5kg	0.520035	0.34897
16	Rp. 48.190.000	0.02447	2022-01-10 00:00:00	Makaroni Unil	0.341463	0.451945
17	Rp. 24.960.000	0.02610	2022-01-10 00:00:00	makaroni 5kg	0.206446	0.34897
18	Rp. 918.000	0.02773	2022-01-10 00:00:00	Tahu Bulat	0.010453	0.104119
19	Rp. 990.000	0.02936	2022-01-10 00:00:00	Makaroni Kerang	0.006098	0.320366
20	Rp. 39.769.000	0.03100	2022-01-11 00:00:00	pilus	0.2500	0.933066
21	Rp. 60.320.000	0.03263	2022-01-11 00:00:00	makaroni 5kg	0.502613	0.34897
22	Rp. 60.552.000	0.03426	2022-01-12 00:00:00	pilus	0.2500	0.948513
23	Rp. 63.800.000	0.03589	2022-01-12 00:00:00	makaroni 5kg	0.502613	0.383295
24	Rp. 590.000	0.03752	2022-01-12 00:00:00	kerupuk seblak	0.006098	0.0915332
25	Rp. 39.000.000	0.03915	2022-01-13 00:00:00	basreng 2,5kg	0.650697	0.0514874
26	Rp. 66.000.000	0.04078	2022-01-13 00:00:00	Makaroni 5kg	0.520035	0.383295
27	Rp. 116.092.800	0.04241	2022-01-14 00:00:00	Pilus 9kg	0.481707	0.948513
28	Rp. 55.000.000	0.04405	2022-01-14 00:00:00	Makaroni 5kg	0.432927	0.383295
29	Rp. 70.365.600	0.04568	2022-01-15 00:00:00	Pilus 9kg	0.290941	0.948513
30	Rp. 53.900.000	0.04731	2022-01-15 00:00:00	Makaroni 5kg	0.424216	0.383295
31	Rp. 40.000.000	0.04894	2022-01-17 00:00:00	Makaroni 5kg	0.468153	0.383295

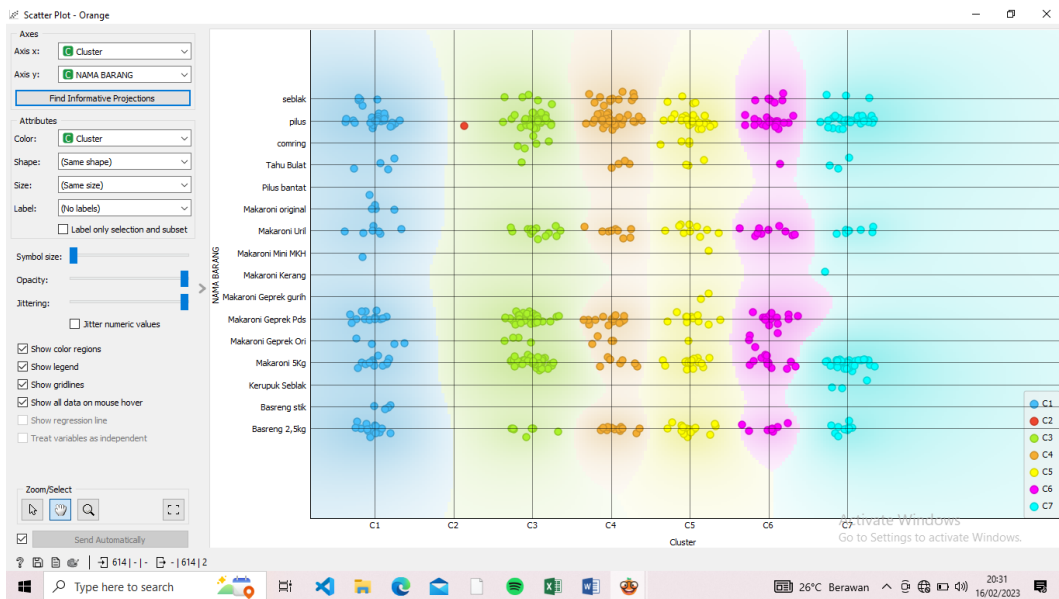
Gambar 3 Data Tabel

Gambar 3 merupakan hasil dari widget Data Tabel-Orange Setelah dilakukan proses preprocessing data menggunakan widget preprocessing terhadap data penjualan produk CV. Widuri.

Dari rangkaian pengolahan data penjualan CV. Widuri menggunakan algoritma K-Means dengan aplikasi Orange. Menunjukkan terdapat 7 cluster.



Gambar 4 Widget Silhouette Plot Orange



Gambar 5 Widget Scatter Plot

Pada Gambar 5 merupakan hasil dari penerapan *widget Scatter Plot* yang digunakan untuk memvisualisasikan hasil kluster $k=7$ dari data penjualan produk CV. Widuri berdasarkan nama-nama produk yang di buat oleh CV. Widuri menggunakan algoritma K-Means pada tools orange. Pada widget scatter plot ini menjelaskan hasil cluster berdasarkan nama-nama produk yang di produksi oleh CV. Widuri yang terdiri dari 7 cluster, dalam setiap cluster ini terdapat beberapa macan produk yang mana antara cluster satu dengan yang lain nya itu berbeda. Dalam *cluster ke-1* terdapat beberapa produk diantaranya yaitu, seblak, pilus, tahu bulat, pilus bantat, makaroni original, makaroni uril, makaroni mini MKH, makaroni geprek pedas, makaroni geprek ori, makaroni 5Kg, basreng stik, basreng 2,5Kg. Kemudian dalam *cluster ke-2* hanya terdapat satu jenis produk yaitu Pilus. Kemudian dalam *cluster ke-3* yaitu seblak, pilus, comring, makaroni uril, makaroni geprek pedas, makaroni geprek ori, makaroni 5Kg, basreng 2,5Kg. Kemudian dalam *cluster ke-4* yaitu seblak, pilus, tahu bulat, makaroni uril, makaroni geprek pedas, makaroni 5Kg, basreng 2,5Kg. Kemudian dalam *cluster ke-5* yaitu seblak, pilus, comring, tahu bulat, makaroni uril, makaroni geprek gurih, makaroni geprek pedas, makaroni geprek ori, makaroni 5Kg, basreng 2,5Kg. Kemudian dalam *cluster ke-6* yaitu seblak, pilus, tahu bulat, makaroni uril, makaroni geprek pedas, makaroni geprek ori, makaroni 5Kg, basreng 2,5Kg. Kemudian dalam *cluster ke-7* yaitu seblak, pilus, tahu bulat, makaroni uril, makaroni kerang, Makaroni 5Kg, kerupuk seblak, basreng 2,5Kg.

Kesimpulan

Dari hasil penelitian dalam penerapan data mining pengelompokan data penjualan CV. Widuri menggunakan algoritma K-Means di peroleh kesimpulan sebagai berikut :

1. Pengelompokan data penjualan CV. Widuri menggunakan algoritma K-Means dengan hasil pengelompokan penjualan produk yang mudah laku yaitu pilus dan produk yang sulit laku yaitu pilus bantat.
2. Hasil clustering dengan menggunakan aplikasi orange menghasilkan 7 cluster dengan rincian, cluster 1 (C1) 104 transaksi, cluster 2 (C2) 1 transaksi, cluster 3 (C3) 139 transaksi, cluster 4 (C4) 95 transaksi, cluster 5 (C5) 102 transaksi, cluster 6 (C6) 76 transaksi, cluster 7 (C7) 97 transaksi.
3. Pada cluster ke-1 transaksi terendah sebesar Rp. 489.000 dan transaksi tertinggi sebesar Rp. 80.586.000, kemudian cluster ke-2 hanya ada 1 transaksi yaitu sebesar Rp. 64.251.000, cluster ke-3 transaksi terendah sebesar Rp. 1.180.000 dan transaksi tertinggi sebesar Rp. 131.214.000, cluster ke-4 transaksi terendah sebesar Rp. 660.00 dan transaksi tertinggi sebesar Rp. 63.162.000, cluster ke-5 transaksi terendah sebesar Rp. 820.000 dan transaksi tertinggi sebesar Rp. 85.500.000, cluster ke-6 transaksi terendah sebesar Rp. 820.000 dan transaksi tertinggi sebesar Rp. 63.162.000, dan yang terakhir cluster ke-7 transaksi terendah sebesar Rp. 612.000 dan transaksi tertinggi sebesar Rp. 122.210.000.

Dengan demikian diharapkan perusahaan dapat lebih memperhitungkan jumlah produksi agar tidak terjadi penambahan jumlah barang sisa (BS) apalagi terhadap produk yang sulit laku seperti pilus bantat.

Daftar Pustaka

- Bakery, P. J. (2020). *Data Mining Untuk Pengelompokan Data Penjualan Cake Dengan Menggunakan Algoritma K-means Clustering*. x.
- Dharma, Y., Sudarma, M., Bagus, I., & Swamardika, A. (2021). *Cluster ing History Data Penjualan Menggunakan*. 20(2).
- Nawawi, M. S., Sembiring, F., & Erfina, A. (2021). *Implementasi Algoritma K-Means Clustering Menggunakan Orange Untuk Penentuan Produk Busana Muslim Terlaris*. 789–797.
- Ramadhan, D. G., Prihatini, I., & Liantoni, F. (2021). *Analisis Clustering Pengelompokan Penjualan Paket Data Menggunakan Metode K-Means*. 13(1), 33–38.
- Suroyo, H. (2019). *Penerapan Machine Learning dengan Aplikasi Orange Data Mining Untuk Menentukan Jenis Buah Mangga*. 343–347.
- Triyandana, G., Putri, L. A., & Umaidah, Y. (2022). *Penerapan Data Mining Pengelompokan Menu Makanan dan Minuman Berdasarkan Tingkat Penjualan Menggunakan Metode K-Means*. 6(1), 40–46.