



Klasifikasi Barang Menggunakan Metode Clustering K-Means Dalam Penentuan Prediksi Stok Barang

Francis Matheos Sarimole^{1*}, Lukamanul Hakim²

^{1,2,3} Program Studi Teknik Informatika, Sekolah Tinggi Ilmu Komputer Cipta Karya Informatika, Jakarta Timur, Indonesia

¹matheosfrancis.s@gmail.com, ²hakimlukmanul224@gmail.com

Abstrak— TB Bambu Kuning merupakan toko yang bergerak dalam penjualan bahan – bahan bangunan yang tingkat penjualannya cukup tinggi dan berlokasi di desa Muara Bakti Bekasi. Saat ini TB Bambu Kuning masih melakukan pemenuhan stok barang atau produk dan melakukan pencatatan transaksi secara manual sehingga sering terjadi kesalahan dalam pencatatan data data yang ada dan juga kurangnya efisiensi waktu yang diperlukan. Maka diperlukan suatu proses pengolahan data besar dengan menggunakan suatu teknik data mining. Teknik data mining yang akan digunakan pada penelitian ini adalah metode K-Means Clustering. Metodologi yang diterapkan dalam penelitian ini adalah CRISP-DM, dengan data penjualan di toko TB Bambu kuning bulan Januari sampai Maret 2023. Tujuannya Mempermudah pengendalian barang yang dikelola, Mendapatkan proses yang lebih akurat dan efektif dalam menentukan prediksi stok barang dan Mengukur tingkat akurasi dan efektifitas penerapan algoritma K-Means clustering untuk menentukan prediksi stok barang menggunakan Algoritma K-means diterapkan dalam pembentukan cluster berdasarkan model Recency, Frequency dan Monetary (RFM). Dengan bantuan tools jupiter notebook. Pada penentuan jumlah cluster (k) terbaik digunakan metode Elbow. Hasil yang didapat dengan 20 data transaksi terbagi menjadi tiga cluster penjualan terlaris, sedang dan kurang laris. Dari akurasi yang dihasilkan dengan nilai SSE yang rendah yaitu 0,0041 dapat disimpulkan bahwa metode Clustering K-Means dapat digunakan untuk proses clustering dengan hasil cluster yang baik.

Kata Kunci: K-means Clustering, CRISP-DM, RFM, Stok barang

Abstract—TB Bambu Kuning is a shop that sells building materials with quite high sales levels and is located in Muara Bakti village, Bekasi. Currently TB Bambu Kuning is still fulfilling stock of goods or products and recording transactions manually so that errors often occur in recording existing data and there is also a lack of time efficiency required. So a big data processing process is needed using a data mining technique. The data mining technique that will be used in this research is the K-Means Clustering method. The methodology applied in this research is CRISP-DM, with sales data at the TB Bambu Kuning store from January to March 2023. The aim is to make it easier to control managed goods, obtain a more accurate and effective process in determining stock predictions and measure the level of accuracy and effectiveness application of the K-Means clustering algorithm to determine stock predictions using the K-means algorithm applied in forming clusters based on the Recency, Frequency and Monetary (RFM) model. With the help of Jupiter notebook tools. To determine the best number of clusters (k), the Elbow method is used. The results obtained with 20 transaction data were divided into three clusters of best-selling, medium and least-selling sales. From the resulting accuracy with a low SSE value of 0.0041, it can be concluded that the K-Means Clustering method can be used for the clustering process with good cluster results.

Keywords: K-means Clustering, CRISP-DM, RFM, Inventory

I. PENDAHULUAN

Sebagai sebuah toko TB Bambu Kuning yang bertempat di Bekasi merupakan sebuah toko yang cukup laris dalam penjualannya dengan model penjualan secara grosir dan eceran, TB Bambu Kuning sebuah toko yang menyediakan dan menjual bahan – bahan bangunan untuk kebutuhan masyarakat guna memperlancar proses pembangunan yang dikerjakan masyarakat tersebut. Untuk mengatasi masalah yang terjadi maka TB Bambu Kuning membutuhkan suatu metode dan sistem perencanaan stok barang yang lebih baik sehingga dapat menentukan barang mana yang harus di stok banyak, sedang atau sedikit, agar tidak mengalami kekurangan atau bahkan kelebihan dalam pemenuhan. Nilai akurasi yang didapatkan sebesar 72%. Manfaat yang dihasilkan dari prediksi HbA1c adalah dapat digunakan sebagai alternatif solusi untuk mengatasi keterbatasan pada laboratorium dalam hal pelayanan pemeriksaan HbA1c dan hasil prediksi HbA1c dapat juga digunakan sebagai rekomendasi oleh dokter dalam menentukan keputusan medis pada penderita diabetes [1]. Komplikasi mikrovaskular yakni komplikasi yang dapat

menyebabkan kerusakan pada mata, ginjal, indra perasa, dan lain-lain yang terjadi pada pembuluh darah kecil. Diprediksi ada 439 juta orang yang pada tahun 2030 akan menderita penyakit diabetes [2]. Hal tersebut terjadi secara berulang dan menimbulkan penumpukan terhadap data mahasiswa baru [3]. Oleh sebab itu para penjual diuntut untuk dapat berfikir cerdas dalam membuat strategi penjualan yang pastinya akan berguna untuk mengetahui dengan jelas produk mana yang paling diminati oleh para customer, dan juga untuk meningkatkan penjualan pada toko mereka. [4]. Penyakit ini berkaitan dengan kondisi lingkungan dan perilaku masyarakat. Penelitian ini membahas tentang pengelompokan jumlah daerah yang terjangkit demam berdarah dengue (DBD) berdasarkan provinsi. Metode yang digunakan adalah Data mining KMeans Clustering. [5]. Meski program ini sangat bisa membantu keluarga miskin dalam menyekolahkan anak mereka, terkadang ada beberapa keluarga yang tergolong miskin namun malah tidak mendapatkannya keuntungan dari program ini, melainkan keluarga yang mampu. [6]. Perusahaan ini bertujuan menentukan tiga kelompok cluster

dengan kemiripan produknya agar dijadikan rekomendasi bagi manajemen perusahaan dalam merencanakan stok barang. Pada penelitian ditentukan 3 cluster dengan cluster 1 merupakan produk yang paling laris, cluster 2 produk yang laris dan yang cluster 3 produk yang kurang laris[7].Oleh karena itu perlu diterapkan data mining menggunakan metode K-Means pada Toko Helai. Metode K-Means dapat diterapkan pada Toko Helai untuk menentukan penjualan baju mana yang sangat laris, laris dan kurang laris [8]. Metode K-Means clustering digunakan agar mempermudah dalam menganalisis serta mengelompokkan data untuk mengetahui beberapa daerah rawan bencana di wilayah Purbalingga. Dalam penelitian ini pengolahan data menggunakan tools rapidminer [9]. K-Means dengan teknik clustering untuk data mining menghasilkan informasi produk asuransi yang lebih diminati para calon nasabah. [10]. Teknik Clustering menggunakan metode K-Means[11]. Metode K-Means ini diharapkan mampu membantu pemerintah Desa Sukamulia Timur dalam pengambilan keputusan dan menemukan informasi yang dibutuhkan untuk memecahkan masalah dalam mendata penduduk miskin dengan tepat[12]. K-means merupakan salah satu metode pengelompokkan data non-hierarki yang mempartisi data yang ada ke dalam bentuk dua atau lebih kelompok[13].Untuk mengatasi permasalahan tersebut dengan menerapkan metode Algoritma K-Means. [14]. Algoritma K-means Data akan diolah dengan melakukan clushtering dalam 3 clushter yaitu clushter tingkat keluhan kesehatan tinggi, clushter tingkat keluhan kesehatan sedang dan rendah. Centroid data untuk cluster tingkat populasi tinggi 37.48, Centroid data untuk cluster tingkat populasi sedang 27.08, dan Centroid data untuk cluster tingkat populasi rendah 14.89[15].

II. METODE PENELITIAN

Pada bagian ini akan dijelaskan mengenai perancangan sistem dalam penelitian ini. Metode dalam penelitian ini dilakukan melalui beberapa tahap. Secara garis besar, alur penelitian dapat dijelaskan dibawah ini:

- CRISP-DM memiliki kepanjangan Cross-Industry Standard Process for Data Mining adalah sebuah metode data mining yang dikembangkan bersama antara Daimler-Chrysler, SPSS, dan NCR dimana dari Namanya merupakan sebuah metode netral dan dapat digunakan dalam segala lini bisnis dan berbagai tool.
- Sebagai sebuah metodologi, CRISP-DM menggambarkan fase dari tahapan-tahapan dalam sebuah proyek, pekerjaan yang terkait dalam tiap fase dan penjabaran terkait hubungan antar pekerjaan tersebut serta memberikan sebuah gambaran siklus hidup (life-cycle) dari Data Mining bila dilihat sebagai Model Proses.
- Dari penggambaran tersebut metode ini memberikan sebuah proses standar yang bersifat umum atau tidak eksklusif dalam strategi pemecahan masalah dalam sebuah unit bisnis atau penelitian dengan menggunakan Data Mining yang sesuai atau tepat.

A. Sistem

Sistem adalah Sistem berasal dari bahasa Latin (Systema) dan bahasa Yunani (Sustema) yang berarti suatu kesatuan yang terdiri dari komponen atau elemen yang dihubungkan bersama untuk memudahkan aliran informasi, materi atau energi untuk mencapai suatu tujuan. Istilah ini sering dipergunakan untuk menggambarkan suatu entitas yang berinteraksi. Sistem juga merupakan kesatuan bagian-bagian yang saling berhubungan yang berada dalam suatu wilayah serta memiliki item-item penggerak, contoh umum misalnya seperti negara[16].

B. Bahan Bangunan

Bahan bangunan adalah setiap bahan yang digunakan untuk tujuan konstruksi Banyak bahan alami seperti tanah liat, pasir, kayu, batu bahan daun dan ranting telah digunakan untuk membangun bangunan. Selain dari bahan alami produk buatan banyak digunakan menjadi bahan bangunan[7].

C. Clustering

Clustering merupakan salah satu bagian dari teknik data mining yaitu sekumpulan objek yang mempunyai kesamaan diantara anggotanya dan memiliki ketidaksamaan dengan objek lain pada cluster lainnya, dengan kata lain sebuah cluster adalah sekumpulan objek yang digabung bersama karena persamaan kedekatannya[15].

D. Algoritma K-Mens

K-means menginisialisasi cluster means dengan menghasilkan titik k secara acak di ruang data. Ini biasanya dilakukan dengan menghasilkan nilai secara seragam secara acak dalam rentang untuk setiap dimensi[11].

Berikut adalah langkah-langkah umum dalam algoritma k-means:

- Inisialisasi: Tentukan jumlah kluster yang diinginkan (k). Pilih secara acak k titik pusat awal (centroid) dari data.
- Alokasi: Hitung jarak antara setiap data dengan centroid dan alokasikan data ke kluster dengan centroid terdekat.
- Pembaruan centroid: Hitung ulang centroid baru untuk setiap kluster berdasarkan rata-rata dari data yang termasuk dalam kluster tersebut.
- Iterasi: Ulangi langkah kedua dan ketiga sampai tidak ada perubahan alokasi kluster atau mencapai kriteria penghentian yang ditentukan (misalnya, jumlah iterasi maksimum).
- Hasil: Setelah iterasi selesai, data akan terkluster menjadi k kelompok berdasarkan centroid yang dihasilkan.

E. RFM (Recency, Frequency, Monetary)

RFM merupakan suatu model perhitungan yang terdiri dari tiga atribut domain yang memperhatikan transaksi pelanggan berdasarkan transaksi terakhir (Recency), jumlah transaksi (Frequency) serta nominal transaksi (Monetary). Metodologi ini sangat bermanfaat dalam segmentasi pelanggan dengan membagi pelanggan kedalam beberapa kelompok. RFM terdiri dari tiga dimensi, yaitu Recency,Frequency,Monetary[9].

- a) Recency adalah mengukur nilai pelanggan berdasarkan pembelian paling akhir yang dilakukan pelanggan. Data terpenting yang diperlukan untuk menghitung nilai recency tanggal pembelian terakhir. Nilai recency berkaitan dengan jarak antara tanggal terakhir transaksi dengan periode analisis sehingga semakin dekat proses pembelian terakhir maka pelanggan tersebut semakin royal. Nilai recency ini biasanya dinyatakan dalam hari.
- b) Fequency adalah mengukur nilai pelanggan berdasarkan seberapa sering pelanggan tersebut melakukan transaksi. Semakin sering melakukan transaksi ini memungkinkan bahwa pelanggan tersebut merupakan pelanggan potensial. Transaksi yang dimaksud tidak tergantung pada berapa banyak pelanggan yang beli tetapi satu transaksi pun bisa mewakili bahwa pelanggan tersebut potensial.
- c) Monetary adalah mengukur nilai pelanggan dengan melihat nilai pembelian yang dilakukan pelanggan dalam periode tertentu. Semakin tinggi nilai pembelian maka semakin tinggi pula potensial pelanggan tersebut.

F. Alur Penelitian

Dalam proses klasifikasi barang dalam penentuan prediksi stok barang, peneliti menggunakan tool jupiter notebook dan excel sebagai alat bantu untuk pengolahan data, dan menggunakan elbow metode untuk menentukan nilai K berapa cluster yang paling bagus untuk data yang akan di olah serta menggunakan metode CRISP-DM dalam metodologi prosesnya

G. Implementasi Metodologi

Dalam melakukan clustering, data yang diperoleh akan dihitung terlebih dahulu dengan menggunakan metodologi proses CRISP-DM dan melalui tahapan – tahapan metodologi K Means Clustering.

H. Business Understanding

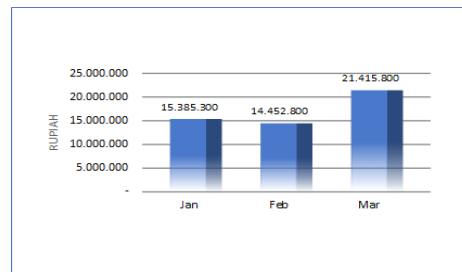
Kegiatan pendalaman ke toko TB Bambu kuning secara terbuka agar dapat menentukan tujuan dan menilai kondisi yang terjadi. Wawancara juga dilakukan untuk mengetahui secara mendalam tentang informasi dan kondisi yang terjadi di toko TB Bambu kuning dan menghasilkan data yang dapat dimanfaatkan sebagai pengetahuan atau informasi, salah satunya adalah data transaksi pada toko TB Bambu kuning yang berisi history transaksi penjualan setiap harinya. Data tersebut dijadikan sebagai bahan untuk menyusun strategi prediksi pengadaan stok pada toko TB Bambu kuning.

I. Data Understanding

Proses yang dilakukan dalam pemahaman data pengumpulan dan pemahaman data apa saja yang diperlukan, yang berupa data penjualan di toko TB Bambu Kuning selama bulan januari – maret 2023. Potongan data penjualan dapat dilihat pada gambar 1 dan merupakan gambaran masing- masing informasi transaksi penjualan di toko TB Bambu Kuning, yaitu kode barang,nama barang, jumlah transaksi dan nominal transaksi.

No	kode	Nama Barang	Jumlah	Total Harga
1	BK1	Slang Benang Fuso	4	380.000
2	BK2	Pipa PVC 2 1/2"	4	970.000
3	BK3	Kawat Patri	6	377.400
4	BK4	Pipa PVC 1"	6	777.000
5	BK5	Kran Cuci Piring	8	1.280.000
6	BK6	Kawat Las RB 2,6	8	784.000
7	BK7	Baut Reng Baja Ringan 250 pcs	8	940.000
8	BK8	Pipa PVC 1/2"	8	460.000
9	BK9	Bata 100 Pcs	36	12.600.000
10	BK10	Catylac ext Grey Sand 5 kg	16	2.320.000
11	BK11	Catylac ext putih 5 kg	16	2.160.000
12	BK12	Besi K58	90	3.690.000
13	BK13	Klem Slang	60	72.000
14	BK14	Lem Isarplas Tube	70	486.500
15	BK15	Thiner 1 Ltr	80	1.760.000
16	BK16	Kuas Crocodile 3"	80	1.920.000
17	BK17	Paku Bintang	80	3.280.000
18	BK18	Besi K56	85	3.281.000
19	BK19	Semen TR 50 Kg	180	12.546.000
20	BK20	Kawat Beton	90	1.170.000

Gambar 1. Data Penjualan



Gambar 2. Grafik Penjualan Jan sd Mar 2023

Pada Gambar 2 dapat dilihat pendalaman data pada gambar 1 yaitu total penjualan pada tiap bulannya. Toko TB Bambu kuning mempunyai total transaksi yang tinggi pada bulan maret 2023 yaitu sebesar 366 data transaksi dan total transaksi sebesar Rp.21.415.800. Di bulan januari data transaksi sebanyak 284 dan total transaksi sebesar Rp.15.385.300. Di bulan februari data transaksi sebanyak 285 dan total transaksi sebesar Rp.14.452.500 dan sebagai penjualan nominal terendah.

J. Data Pre- Preparation

Pengolahan data terdiri dari kegiatan untuk menyusun data set yang nantinya akan di inputkan di tool pemodelan data mining. Data transaksi penjualan yang didapat dari tahap pengumpulan di ubah menjadi model RFM, sehingga data set terdiri dari sejumlah penjualan dengan tiga atribut yaitu Recency, Frequency dan Monetary. Adapun atribut dari model RFM data penjualan pada toko TB Bambu kuning, beserta contohnya di jelaskan pada tabel 1.

Tabel 1. Atribut Model RFM



No	Atribut	Ulasan
1	Recency	Selisih hari antara tanggal dilakukannya analisis dan tanggal terakhir penjualan di toko TB Bambu kuning Recency = Tanggal penelitian dilakukan - tanggal terakhir transaksi penjualan Contoh : 31 Maret 2023 - 29 Maret 2023 = 2
2	Frequency	Total penjualan yang dilakukan oleh toko TB Bambu kuning selama rentan waktu analisis Contoh : Penjualan BK10 yang dilakukan toko TB Bambu kuning di bulan Januari sampai Maret sebanyak 30 transaksi jadi poin <i>frequency</i> nya adalah 30
3	Monetary	Total nominal yang didapatkan toko TB Bambu kuning selama rentan waktu analisis Contoh : Total nominal yang didapatkan toko TB Bambu kuning saat menjual BK10 sebesar Rp.200.000 maka poin <i>monetary</i> nya adalah 200 000

Berdasarkan tabel 1 maka semua atribut pada model RFM menggunakan bentuk data numerik, data penjualan di toko TB Bambu kuning. Model RFM digunakan sebagai pemodelan data penjualan dalam penelitian. Potongan hasil perubahan bentuk dari data penjualan kedalam bentuk RFM dapat dilihat pada tabel 2.

Tabel 2. Hasil perubahan data penjualan ke bentuk RFM

No	Code	Recency	Frequency	Monetary
1	BK1	35	4	380.000
2	BK2	34	4	970.000
3	BK3	35	6	377.400
4	BK4	34	6	777.000
5	BK5	34	8	1.280.000

Sebelum di klusterisasi menggunakan *K Means* nilai RFM dinormalisasikan menggunakan normalisasi standar. Adapun untuk normalisasi dilakukan dengan membagikan nilai setiap data RFM dengan nilai RFM yang terbesar pada kolom RFM tersebut. Nilai RFM sebelum di normalisasi terdapat pada tabel berikut :

Tabel 3. Nilai RFM sebelum di normalisasi

No	Code	Recency	Frequency	Monetary
1	BK1	35	4	380.000
2	BK2	34	4	970.000
3	BK3	35	6	377.400
4	BK4	34	6	777.000
5	BK5	34	8	1.280.000
6	BK6	35	8	784.000
7	BK7	33	8	940.000
8	BK8	35	8	460.000
9	BK9	31	36	12.600.000
10	BK10	30	16	2.320.000
11	BK11	30	16	2.160.000
12	BK12	34	90	3.690.000
13	BK13	30	60	72.000
14	BK14	30	70	486.500
15	BK15	30	80	1.760.000
16	BK16	30	80	1.920.000
17	BK17	30	80	3.280.000
18	BK18	30	85	3.281.000
19	BK19	30	180	12.546.000
20	BK20	30	90	1.170.000

Dari data nilai RFM ditemukan nilai maksimum koefisien normalisasi tertera pada tabel:

Tabel 4 Nilai Maksimum RFM

R	F	M
35	180	12.600.000

Menggunakan rumus simple feature scaling untuk menormalisasi data RFM. Rumus simple feature scaling :

$$X_{new} = \frac{X_{old}}{X_{max}} \tag{1}$$

Hasil perhitungan pada tabel 3

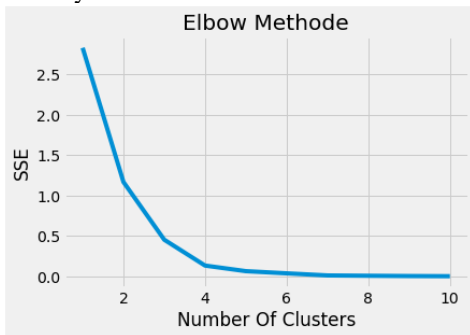
Tabel 5. Nilai RFM setelah Normalisasi

Code	Recency	Frequency	Monetary
BK1	1,00	0,02	0,03
BK2	0,97	0,02	0,08
BK3	1,00	0,03	0,03
BK4	0,97	0,03	0,06
BK5	0,97	0,04	0,10
BK6	1,00	0,04	0,06
BK7	0,94	0,04	0,07
BK8	1,00	0,04	0,04
BK9	0,89	0,20	1,00
BK10	0,86	0,09	0,18
BK11	0,86	0,09	0,17
BK12	0,97	0,50	0,29
BK13	0,86	0,33	0,01
BK14	0,86	0,39	0,04
BK15	0,86	0,44	0,14
BK16	0,86	0,44	0,15
BK17	0,86	0,44	0,26
BK18	0,86	0,47	0,26
BK19	0,86	1,00	1,00
BK20	0,86	0,50	0,09

K. Modeling

Clustering menjadi pada tahap pemodelan untuk mengelompokkan penjualan dengan karakter yang serupa. Langkah – langkah dalam pemodelan menggunakan algoritma Kmeans sebagai berikut :

- a) Data yang digunakan sebagai input pada algoritma Kmeans adalah data RFM ternormalisasi seperti yang terlihat pada tabel 4.4, dengan jumlah data sebanyak 20 data di toko TB Bmbu kuning dengan 3 atribut yaitu atribut Recency, Frequency dan Monetary.
- b) Data yang digunakan sebagai input pada algoritma Kmeans adalah data RFM ternormalisasi seperti yang terlihat pada tabel 4.4, dengan jumlah data sebanyak 20 data di toko TB Bmbu kuning dengan 3 atribut yaitu atribut Recency, Frequency dan Monetary.



Gambar 3. Grafik elbow

- c) Pada tahap ini akan dilakukan proses utama yaitu segmentasi atau pengelompokan data penjualan barang menggunakan K Means Clustering. Berikut ini merupakan diagram flowchart dari algoritma Kmeans Clustering, dengan parameter input jumlah dataset sebanyak n data dan jumlah inisialisasi centroid K =3 sesuai penelitian. Dari banyak data penjualan yang diperoleh , diambil 20 jenis barang yang dijadikan sampel untuk penerapan algoritma KMeans . Percobaan dilakukan dengan menggunakan parameter berikut :

Jumlah cluster = 3

Jumlah data = 20

Tabel 6 Sampel Data

No	Code	Frequency	Monetary
1	BK1	0,02	0,03
2	BK2	0,02	0,08
3	BK3	0,03	0,03
4	BK4	0,03	0,06
5	BK5	0,04	0,10
6	BK6	0,04	0,06
7	BK7	0,04	0,07
8	BK8	0,04	0,04
9	BK9	0,20	1,00
10	BK10	0,09	0,18
11	BK11	0,09	0,17
12	BK12	0,50	0,29
13	BK13	0,33	0,01
14	BK14	0,39	0,04
15	BK15	0,44	0,14
16	BK16	0,44	0,15
17	BK17	0,44	0,26
18	BK18	0,47	0,26
19	BK19	1,00	1,00
20	BK20	0,50	0,09

Iterasi ke-1

Penentuan pusat awal cluster

Pusat awal cluster atau centroid didapatkan secara random, untuk penentuan awal cluster adalah :

Pusat cluster 1 : (0.03 , 0.06)

Pusat cluster 2 : (0.04 , 0.10)

Pusat cluster 3 : (0.09 , 0.18)

Perhitungan jarak pusat cluster

Untuk mengukur antara data dengan pusat cluster yang digunakan Euclidean distance kemudian akan didapatkan matrik jarak sebagai berikut :

Rumus Euclidean Distance :

$$d = |x - y| = \sqrt{\sum_{i=1}^n (x_i - y_i)^2} \tag{2}$$

X = Pusat Cluster

Y= Data

Dari 20 data yang dijadikan sampel telah dipilih pusat awal cluster yaitu C1(0.30, 0.06) C2(0.04,0.10) C3(0.09,0.18). Lalu dilakukan perhitungan jarak dari sisa sampel data dengan pusat cluster. Hasil iterasi ke-1 dapat dilihat pada tabel 7.

Tabel 7 Hasil Iterasi ke-1

Code	C1	C2	C3
BK1	0,01	0,26	0,04
BK2	0,82	0,91	0,87
BK3	0,83	0,89	0,88
BK4	0,15	0,13	0,21
BK5	1,68	0,98	1,82
BK6	0,00	0,34	0,03
BK7	0,34	0,00	0,44
BK8	0,03	0,44	0,00
BK9	0,59	0,97	0,60
BK10	0,00	0,31	0,03
BK11	1,01	0,83	1,09
BK12	0,25	0,42	0,29
BK13	0,91	1,10	0,95
BK14	0,96	0,78	1,04
BK15	0,05	0,46	0,02
BK16	0,14	0,15	0,19
BK17	0,69	0,99	0,71
BK18	0,00	0,40	0,02
BK19	0,02	0,36	0,00
BK20	0,05	0,35	0,03

AnggotaC1=

{BK1,BK2,BK3,BK6,BK9,BK10,BK12,BK13, BK16,BK17,BK18}

Anggota C2 = {BK4,BK5,BK7,BK11,BK14}

Anggota C3 = {BK8,BK15,BK19,BK20}

a) Hasil pusat baru

Tentukan posisi centroid baru (Ck)dengan cara menghitung nilai rata rata dari data – data yang ada pada centroid yang sama.

$$Ck = \left(\frac{1}{nk}\right)\sum d1$$

Dimana nk adalah jumlah dokumen dalam cluster k dan d1 dokumen dalam cluster k, sehingga di dapatkan titik pusat atau centroid yang baru yaitu :

C1 : (0.24,0.21)

C2 : (0.12 , 0.09)

C3 : (0.50 , 0.32)

b) Perhitungan pusat jarak cluster

Hitung euclidean distance dari semua data ke titik pusat yang baru (C1,C2,C3) seperti yang telah dilakukan pada tahap 1 dan setelah hitungan kita dapatkan, kemudian bandingkan hasil tersebut. Jika hasil posisi cluster pada iterasi ke 2 sama dengan posisi dengan iterasi yang pertama, maka proses dihentikan, namun jika tidak maka proses dilanjutkan ke iterasi ke 3.

Tabel 8 Hasil itersi ke 2

Code	C1	C2	C3
BK1	0,25	0,10	0,56
BK2	0,24	0,10	0,54
BK3	0,24	0,09	0,55
BK4	0,23	0,09	0,53
BK5	0,21	0,08	0,50
BK6	0,22	0,08	0,52
BK7	0,21	0,08	0,52
BK8	0,23	0,08	0,54
BK9	0,66	0,91	0,76
BK10	0,15	0,04	0,43
BK11	0,15	0,04	0,43
BK12	0,27	0,42	0,00
BK13	0,14	0,22	0,27
BK14	0,18	0,27	0,19
BK15	0,21	0,33	0,09
BK16	0,21	0,33	0,08
BK17	0,21	0,35	0,06
BK18	0,23	0,38	0,03
BK19	1,38	1,70	0,96
BK20	0,27	0,38	0,05

C1 : (0.31,0.35)

C2 : (0.04 , 0.06)

C3 : (0.44 , 0.27)

Anggota C1 = {BK9,BK13,BK14}

AnggotaC2=

{BK1,BK2,BK3,BK4,BK5,BK6,BK7,BK8,BK10,BK11}

AnggotaC3=

{BK12,BK15,BK16,BK17,BK18,BK19,BK20}

Tabel 9. Hasil iterasi ke 3

Code	C1	C2	C3
BK1	0,39	0,02	0,48
BK2	0,36	0,02	0,46
BK3	0,38	0,01	0,46
BK4	0,36	0,01	0,45
BK5	0,33	0,01	0,42
BK6	0,35	0,00	0,44
BK7	0,34	0,00	0,43
BK8	0,36	0,00	0,45
BK9	0,53	1,04	0,77
BK10	0,25	0,06	0,36
BK11	0,25	0,06	0,36
BK12	0,19	0,51	0,06
BK13	0,14	0,30	0,18
BK14	0,18	0,35	0,10
BK15	0,18	0,41	0,02
BK16	0,17	0,41	0,02
BK17	0,14	0,44	0,00
BK18	0,17	0,47	0,03
BK19	1,11	1,84	1,09
BK20	0,26	0,46	0,09

Anggota C1 = {BK9,BK13,}

AnggotaC2=

{BK1,BK2,BK3,BK4,BK5,BK6,BK7,BK8,BK10,BK11}

AnggotaC3=

{BK12,BK14,BK15,BK16,BK17,BK18,BK19,BK20}

Dikarenakan hasil posisi cluster pada iterrasi ke 3 tidak sama dengan posisi cluster pada itersi kedua , maka proses di lanjutkan pada iterasi ke 4. Maka harus ditentukan centroid yang baru yaitu :

C1 : (0.60, 0.59)

C2 : (0.05, 0.08)

C3 : (0.52 , 0.28)

Tabel 10. Hasil Iterasi ke 4

Code	C1	C2	C3
BK1	0,89	0,03	0,56
BK2	0,84	0,03	0,54
BK3	0,88	0,02	0,55
BK4	0,85	0,02	0,53
BK5	0,79	0,01	0,51
BK6	0,83	0,01	0,52
BK7	0,82	0,01	0,52
BK8	0,86	0,01	0,53
BK9	0,57	1,00	0,84
BK10	0,68	0,05	0,44
BK11	0,69	0,05	0,44
BK12	0,19	0,50	0,02
BK13	0,61	0,29	0,26
BK14	0,52	0,34	0,19
BK15	0,36	0,40	0,10
BK16	0,35	0,40	0,09
BK17	0,26	0,43	0,08
BK18	0,24	0,45	0,05
BK19	0,56	1,79	0,99
BK20	0,35	0,45	0,06

Anggota C1 = {BK9,BK19,}

AnggotaC2=

{BK1,BK2,BK3,BK4,BK5,BK6,BK7,BK8,BK10,BK11}

AnggotaC3=

{BK12,BK13,BK14,BK15,BK16,BK17,BK18,BK20}

Dikarenakan hasil posisi cluster pada iterrasi ke 4 tidak sama dengan posisi cluster pada itersi ketiga , maka proses di lanjutkan pada iterasi ke 5 . Maka harus ditentukan centroid yang baru yaitu :

C1 : (0.60, 1.00)

C2 : (0.05, 0.08)

C3 : (0.44 , 0.16)

Tabel 11. Hasil iterasi ke-5

Code	C1	C2	C3
BK1	1,52	0,03	0,43
BK2	1,43	0,03	0,42
BK3	1,51	0,02	0,42
BK4	1,45	0,02	0,42
BK5	1,36	0,01	0,40
BK6	1,43	0,01	0,41
BK7	1,41	0,01	0,40
BK8	1,48	0,01	0,41
BK9	0,40	1,00	0,95
BK10	1,18	0,05	0,35
BK11	1,20	0,05	0,35
BK12	0,60	0,50	0,08
BK13	1,26	0,29	0,13
BK14	1,14	0,34	0,07
BK15	0,90	0,40	0,00
BK16	0,87	0,40	0,00
BK17	0,70	0,43	0,01
BK18	0,67	0,45	0,04
BK19	0,40	1,79	1,26
BK20	0,92	0,45	0,06

Anggota C1 = {BK9,BK19,}

AnggotaC2=

{BK1,BK2,BK3,BK4,BK5,BK6,BK7,BK8,BK10,BK11}

Anggota C3

{BK12,BK13,BK14,BK15,BK16,BK17,BK18,BK20}

Keterangan

Barang dengan kode BK9 , BK19 merupakan anggota C1 dan termasuk barang dengan tingkat penjualan yang tinggi dan jumlah nilai nominal yang tinggi, sedangkan barang dengan code dibawah ini:

BK1,BK2,BK3,BK4,BK5,BK6,BK7,BK8,BK10,BK11 merupakan anggota dari clstuer C2 dan dikategorikan sebagai barang yang tingkat penjualannya yang rendah dan nilai nominal juga rendah, dan barang dengan kode BK12,BK13,BK14,BK15,BK16,BK17,BK18,BK20 merupakan anggota C3 dengan katagori penjualan sedang dan nilai nominal yang sedang.



Gambar 4 Grafik per iterasi

L. Evaluation

Tahap evaluasi dilakukan untuk mengevaluasi nilai random seed (s) yang digunakan. Setiap nilai random seed menghasilkan SSE yang berbeda. Hal ini dikarenakan algoritma Kmeans menggunakan inisialisasi yang acak dalam prosesnya. Oleh karena itu dalam tahap evaluasi dilakukan clustering dengan mengganti nilai random seed yang di gunakan pada tools jupiter notebook. Dari tahap

pemodelan pada gambar 4.3 terlihat bahwa default dari random seed adalah 10. Tabel 4.11 menunjukkan nilai SSE untuk beberapa nilai random seed, yaitu S = 10 sampai s = 20.

```
In [26]: from sklearn.metrics import silhouette_samples,silhouette_score
sse=[]
for i in range(1, 11):
    kmeans=KMeans(n_clusters=i, init='k-means++',
                 max_iter=300,n_init=10,random_state=0)
    kmeans.fit(features)
    sse.append(kmeans.inertia_)

In [25]: sse

Out[25]: [2.8299730457892416,
1.1700410388042046,
0.4557177483166415,
0.1357085646431721,
0.0666122153187201,
0.03966223696145126,
0.013693772980179729,
0.0097511341962711,
0.006213868512219705,
0.004129558075081887]
```

Gambar 5. Hasil nilai SEE beberapa nilai random

Berdasarkan gambar 5 dapat dilihat ketika S = 20 (cetak tebal), nilai SSE yang didapatkan adalah yang terkecil. Oleh karena itu K = 3 dan S = 20 adalah perpaduan yang menghasilkan cluster terbaik.

III. HASIL DAN PEMBAHASAN

Saat ini TB Bambu Kuning masih melakukan pemenuhan stok barang atau produk dan melakukan pencatatan transaksi secara manual sehingga sering terjadi kesalahan dalam pencatatan data data yang ada dan juga kurangnya efisiensi waktu yang diperlukan. Jumlah permintaan dari konsumen yang fluktuatif mengakibatkan stok yang harus disiapkan menjadi tidak stabil. Disamping itu TB Bambu Kuning tidak dapat mengelompokan produk yang laris dan tidak laris terjual. Sehingga kesulitan yang dialami adalah kurangnya stok produk yang laku karena penjualan tinggi, dan menumpuknya produk yang tidak laku karena penjualannya rendah. Maka diperlukan suatu proses pengolahan data besar dengan menggunakan suatu teknik data mining. Teknik data mining yang akan digunakan pada penelitian ini adalah metode K-Means Clustering. Metodologi yang diterapkan dalam penelitian ini adalah CRISP-DM, dengan data penjualan di toko TB Bambu kuning bulan januari sampai maret 2023. Algoritma Kmeans diterapkan dalam pembentukan cluster berdasarkan model Recency, Frequency dan Monetary(RFM). Dengan bantuan tools jupiter note book. Pada penentuan jumlah cluster (k) terbaik digunakan metode Elbow . Hasil yang didapat dengan 20 data transaksi terbagi menjadi tiga cluster penjualan terlaris, sedang dan kurang laris. Dari akurasi yang dihasilkan dapat disimpulkan bahwa metode Clustering K-Means dapat sistem dengan baik.

A. Tahap Pengujian

Berdasarkan hasil dari tahapan evaluasi maka didapatkan pusat cluster untuk masing masing atribut RFM seperti terlihat pada tabel 12

Tabel 12 Pusat masing masing cluster



Cluster	Frequency	Monetary
0	0,60	1,00
1	0,05	0,80
2	0,44	0,16

B. Nilai Rata-rata RFM

Tabel 12 merupakan nilai hasil pusat segmentasi dari masing masing cluster dengan total Frequency 0.60 dan Monetary 1.00 pada cluster 0. Selanjutnya pada cluster 1 memiliki total Frequency 0.05 dan Monetary 0.80. Nilai segmentasi terakhir pada cluster 2 memiliki total nilai Frequency 0,44 dan Monetary 0,16. Dari clustering dengan algoritma Kmeans dan RFM dalam bentuk modelnya, maka didapatkan hasil segmentasi penjualan di toko TB Bambu kuning. Rata – rata nilai RFM ditunjukkan pada tabel 13.

Tabel 13. Nilai Rata-rata RFM

Cluster	Frequency	Monetary
0	108	12.573.000
1	7	870.700
2	79	1.957.438

Tabel 13 merupakan total rata rata dari hasil segmentasi data penjualan selama analisis dilakukan pada penelitian ini. Langkah berikut dalam tahap penyebaran yakni dengan menentukan cluster mana yang memiliki atau kemampuan rendah , sedang, dan tinggi dari masing masing simbol rank.

Tabel 14. Rank masing – masing Cluster

Cluster	Frequency	Monetary	Potensi
0	108	12.573.000	Tinggi
1	7	870.700	Rendah
2	79	1.957.438	Sedang

Berdasarkan tabel 14 tersebut terdapat pada cluster 0 dengan jumlah rata – rata nilai frequency 108, dan nilai rata-rata pada Monetary transaksinya paling tinggi dari yang lainnya yakni Rp.12.573.000 dibulan januari sampai maret 2023. Jumlah nilai rata-rata tersebut menunjukkan cluster ini memiliki penjualan paling tinggi. Pada cluster 1 terdapat jumlah nilai rata-rata frequency 7, dan jumlah rata-rata nilai monetary yakni Rp.870.700 data transaksi di bulan januari sampai maret 2023, dari data ini menunjukkan memiliki penjualan rendah. Pada cluster 2 terdapat data penjualan dengan nilai rata-rata frequency transaksinya yakni 79 dan nilai rata-rata monetary transaksinya Rp1.957.438, dari data ini menunjukkan penjualan dengan ranking sedang.

IV. KESIMPULAN

Dari hasil pembahasan yang dilakukan dapat ditarik kesimpulan sebagai berikut Diterapkannya metode algoritma Kmeans Clustering berdasarkan analisis dalam bentuk RFM pada data penjualan di toko TB Bambu kuning bulan januari sampai maret 2023. Jumlah cluster paling baik yang didapat yakni menurut hasil penerapan hasil penetapan jumlah cluster dengan memakai metode elbow adalah k = 3. SSE digunakan untuk mengevaluasi

hasil clustering berdasarkan k (jumlah cluster). Nilai random seed yang menghasilkan SSE terkecil yakni random seed S =20. Segmentasi penjualan berdasarkan model RFM yang dihasilkan terbagi menjadi tiga cluster yaitu kategori data C1 dengan frequency dan monetary tinggi dikategorikan penjualan tinggi sehingga dibutuhkan stok banyak untuk memenuhi permintaan konsumen, kategori data C2 dengan frequency dan monetary rendah dikategorikan penjualan rendah sehingga stok yang dibutuhkan sedikit dan kategori data C3 dengan frequency dan monetary sedang dikategori penjualan sedang sehingga dibutuhkan stock yang tidak banyak. Hasil rank dari tiap-tiap cluster dapat diterapkan dalam prediksi pengadaan stok barang pada toko TB Bambu kuning. Dari akurasi yang dihasilkan dapat disimpulkan bahwa sistem dapat mendukung dalam upaya pengelompokan barang menggunakan algoritma Kmeans clustering untuk prediksi pengadaan stok barang.

UCAPAN TERIMA KASIH

Saya ingin mengucapkan terima kasih yang sebesar-besarnya atas bimbingan, dukungan, dan dedikasi Bapak Francis Matheos Sarimole dalam membantu pembuatan jurnal ini. Tanpa panduan dan dorongan yang berharga dari Anda, pencapaian ini tidak akan mungkin terwujud. Saya sangat berterimakasih atas waktu dan pengetahuan yang Anda luangkan untuk membimbing saya dalam menyusun jurnal ini. Dalam setiap pertemuan dan diskusi, Bapak telah memberikan wawasan yang berharga, saran yang berharga, dan arahan yang tepat. Bimbingan Bapak telah memainkan peran penting dalam membantu saya memahami proses penelitian, mengembangkan metodologi yang tepat, dan menganalisis data dengan benar. Selain itu, terima kasih juga karena telah memberikan koreksi dan umpan balik konstruktif dalam setiap tahap penulisan jurnal. Pengamatan dan penilaian Bapak yang mendalam telah membantu saya untuk meningkatkan kualitas tulisan dan memperbaiki kelemahan yang ada. Saya sangat menghargai kesabaran dan ketelitian yang Anda tunjukkan dalam membimbing saya menuju hasil akhir yang memuaskan. Lebih dari sekadar menjadi pembimbing, Bapak juga menjadi inspirasi bagi saya. Kepedulian dan semangat Bapak terhadap penelitian dan pengembangan ilmu pengetahuan telah mendorong saya untuk terus belajar dan berkembang. Saya merasa sangat beruntung dan bersyukur dapat belajar dari Bapak, seorang akademisi yang berkompeten dan berdedikasi. Terima kasih karena telah memberikan kesempatan kepada saya untuk terlibat dalam penelitian ini dan memperluas pengetahuan serta wawasan saya. Saya meyakini bahwa karya tulis ini tidak hanya akan memberikan manfaat bagi saya secara pribadi, tetapi juga akan memberikan kontribusi yang berarti dalam perkembangan bidang ilmu yang kita geluti. Sekali lagi, terima kasih yang tak terhingga atas semua bimbingan, dorongan, dan kepercayaan yang Bapak berikan kepada saya. Saya sangat beruntung dan bangga dapat belajar di bawah

arahan Bapak. Semoga keberhasilan ini menjadi bukti nyata dari dedikasi dan komitmen Bapak dalam membentuk generasi muda yang berkualitas dan berkontribusi positif dalam dunia akademik.

V. REFERENSI

- [1] F. Yunita, "PENERAPAN DATA MINING MENGGUNAKAN ALGORITMA K-MEANS CLUSTERING PADA PENERIMAAN MAHASISWA BARU (STUDI KASUS: UNIVERSITAS ISLAM INDRAGIRI)," 2018.
- [2] J. Nasir, "PENERAPAN DATA MINING CLUSTERING DALAM MENGELOMPOKAN BUKU DENGAN METODE K-MEANS," *Jurnal SIMETRIS*, vol. 11, no. 2, 2020.
- [3] K. Fatmawati and A. P. Windarto, "DATA MINING: PENERAPAN RAPIDMINER DENGAN K-MEANS CLUSTER PADA DAERAH TERJANGKIT DEMAM BERDARAH DENGUE (DBD) BERDASARKAN PROVINSI," 2018. [Online]. Available: <https://www.depkes.go.id/>.
- [4] E. Fammaldo and L. Hakim, "PENERAPAN ALGORITMA K-MEANS CLUSTERING UNTUK PENGELOMPOKAN TINGKAT KESEJAHTERAAN KELUARGA UNTUK PROGRAM KARTU INDONESIA PINTAR," 2018.
- [5] S. Nurajizah and A. Salbinda, "Penerapan Data Mining Metode K-Means Clustering Untuk Analisa Penjualan Pada Toko Fashion Hijab Banten," *Jurnal Teknik Komputer AMIK BSI*, vol. 7, no. 2, 2021, doi: 10.31294/jtk.v4i2.
- [6] F. Amin, D. S. Anggraeni, and Q. Aini, "Penerapan Metode K-Means dalam Penjualan Produk Souq.Com," *Applied Information System and Management (AISM)*, vol. 5, no. 1, pp. 7–14, Apr. 2022, doi: 10.15408/aism.v5i1.22534.
- [7] R. Yuliani, "Penerapan Data Mining untuk Mengeluster Data Penduduk Miskin Menggunakan Algoritma K-Means di Dusun Bagik Endep Sukamulia Timur," *Jurnal Informatika dan Teknologi*, vol. 4, no. 1, 2021, doi: 10.29408/jit.v4i1.2986.
- [8] J. Hutagalung and F. Sonata, "Penerapan Metode K-Means Untuk Menganalisis Minat Nasabah," *JURNAL MEDIA INFORMATIKA BUDIDARMA*, vol. 5, no. 3, p. 1187, Jul. 2021, doi: 10.30865/mib.v5i3.3113.
- [9] H. Effendi, A. Syahrial, S. Prayoga, and W. D. Hidayat, "Penerapan Metode K-Means Clustering untuk Pengelompokan Lahan Sawit Produktif pada PT Kasih Agro Mandiri APPLICATION OF K-MEANS CLUSTERING METHOD FOR GROUPING PRODUCTIVE PALM OIL LAND AT PT KASIH AGRO MANDIRI," *TEKNOMATIKA*, vol. 11, no. 02, pp. 1–5, 2021.
- [10] D. I. Ramadhani, O. Damayanti, O. Thaushiyah, and A. R. Kadafi, "Penerapan Metode K-Means Untuk Clustering Desa Rawan Bencana Berdasarkan Data Kejadian Terjadinya Bencana Alam," *JURIKOM (Jurnal Riset Komputer)*, vol. 9, no. 3, p. 749, Jun. 2022, doi: 10.30865/jurikom.v9i3.4326.
- [11] H. Annur, "Penerapan Data Mining Menentukan Strategi Penjualan Variasi Mobil Menggunakan Metode K-Means Clustering (Studi Kasus Toko Luxor Variasi Gorontalo)," 2019.
- [12] A. Asroni, H. Fitri, and E. Prasetyo, "Penerapan Metode Clustering dengan Algoritma K-Means pada Pengelompokan Data Calon Mahasiswa Baru di Universitas Muhammadiyah Yogyakarta (Studi Kasus: Fakultas Kedokteran dan Ilmu Kesehatan, dan Fakultas Ilmu Sosial dan Ilmu Politik)," *Semesta Teknika*, vol. 21, no. 1, 2018, doi: 10.18196/st.211211.
- [13] C. Selvi, D. Sembiring, L. Hanum, and S. Parsaoran Tamba, "PENERAPAN DATA MINING MENGGUNAKAN ALGORITMA K-MEANS UNTUK MENENTUKAN JUDUL SKRIPSI DAN JURNAL PENELITIAN (STUDI KASUS FTIK UNPRI)," *Jurnal Sistem Informasi dan Ilmu Komputer Prima*, vol. 5, no. 2, 2022.
- [14] C. Selvi, D. Sembiring, L. Hanum, and S. Parsaoran Tamba, "PENERAPAN DATA MINING MENGGUNAKAN ALGORITMA K-MEANS UNTUK MENENTUKAN JUDUL SKRIPSI DAN JURNAL PENELITIAN (STUDI KASUS FTIK UNPRI)," *Jurnal Sistem Informasi dan Ilmu Komputer Prima*, vol. 5, no. 2, 2022.
- [15] A. Prasatya, R. R. A. Siregar, and R. Arianto, "Penerapan Metode K-Means Dan C4.5 Untuk Prediksi Penderita Diabetes," *PETIR*, vol. 13, no. 1, pp. 86–100, Mar. 2020, doi: 10.33322/petir.v13i1.925.
- [16] N. Rofiqo, A. P. Windarto, and D. Hartama, "KOMIK (Konferensi Nasional Teknologi Informasi dan Komputer) PENERAPAN CLUSTERING PADA PENDUDUK YANG MEMPUNYAI KELUHAN KESEHATAN DENGAN DATAMINING K-MEANS", [Online]. Available: <http://ejurnal.stmik-budidarma.ac.id/index.php/komik>