

Clustering Data Persediaan Barang dengan Menggunakan Metode K-Means

¹ **Dadan Ramdhan**

Program Studi Teknik Informatika STMIK IKMI Cirebon
Jl. Perjuangan no.10B Cirebon Jawa Barat
E-Mail: ramdhand55@gmail.com

² **Gifthera Dwilestari**

Program Studi Sistem Informasi STMIK IKMI Cirebon
Jl. Perjuangan no.10B Cirebon Jawa Barat
E-Mail: ggdwilestari@gmail.com

³ **Raditya Danar Dana**

Program Studi Manajemen Informatika STMIK IKMI Cirebon
Jl. Perjuangan no.10B Cirebon Jawa Barat
E-Mail: radith_danar@yahoo.com

⁴ **Abdul Ajiz**

Program Studi Komputerisasi Akuntansi STMIK IKMI Cirebon
Jl. Perjuangan no.10B Cirebon Jawa Barat
E-Mail: abdulajiz00@gmail.com

⁵ **Kaslani**

Program Studi Komputerisasi Akuntansi STMIK IKMI Cirebon
Jl. Perjuangan no.10B Cirebon Jawa Barat
E-Mail: kaslani123@gmail.com

ABSTRAK

UD. Anugerah Sukses Mandiri merupakan perusahaan yang bergerak dibidang distribusi food dan non food. Transaksi barang yang berjalan terus meningkat, sehingga perusahaan mengalami permasalahan dalam menentukan jumlah persediaan barang, dikarenakan jumlah permintaan barang yang dibutuhkan selalu berubah setiap waktu. Persediaan barang merupakan suatu aktivitas lancar yang meliputi barang-barang milik perusahaan dengan maksud dijual kembali pada suatu periode usaha normal. Data mining merupakan proses yang menggunakan teknik statistik, matematika, kecerdasan buatan dan machine learning untuk mengekstraksi dan mengidentifikasi informasi yang terkait dari berbagai warehouse. Tujuan penelitian ini dengan memanfaatkan data mining yaitu untuk melakukan pengelompokan barang dan meningkatkan akurasi klasterisasi data persediaan barang dengan menggunakan metode K-Means Clustering. Dengan metode K-Means ini dapat mempartisi data ke dalam kelompok sehingga data berkarakteristik sama akan dimasukkan ke dalam satu kelompok yang sama dan data yang berkarakteristik berbeda dikelompokkan kedalam kelompok yang lain, karena metode ini menggunakan centroid (rata-rata) sebagai model dari cluster. Hasil penelitian yang didapat berupa pengelompokan data menjadi 2 kluster yaitu data dengan kluster terendah/sedikit dan kluster tertinggi/terbanyak. Sehingga mendapatkan kesimpulan bahwa clustering persediaan barang dengan menggunakan metode K-Means ini cukup baik dari sisi nilai average within distance dan kompleksitas waktu.

Keyword : Data Mining, K-Means Clustering, Persediaan barang

PENDAHULUAN

Di era perkembangan teknologi yang cukup pesat ini membantu manusia memudahkan setiap kegiatan diberbagai bidang sehingga menjadi lebih efektif dan efisien. Salah satu teknologi yang dapat

digunakan pada bidang industri yaitu dalam melakukan persediaan barang. Ketersediaan barang dan kelengkapan pada suatu perusahaan merupakan suatu elemen yang sangat penting, sehingga sebagai manajemen yang baik dalam proses mengatur ketersediaan stok barang sangat

diperlukan untuk menghindari penumpukan barang yang sama dan barang yang kurang diminati.

Persediaan barang merupakan suatu aktiva lancar yang meliputi barang-barang yang merupakan milik perusahaan dengan sebuah maksud supaya dijual dalam suatu periode usaha normal ataupun persediaan barang yang masih dalam pekerjaan sebuah proses produksi maupun persediaan bahan baku yang juga menunggu penggunaannya di dalam suatu proses (Assauri, 2005). Data mining merupakan proses yang menggunakan teknik statistik, matematika, kecerdasan buatan, dan machine learning untuk mengekstraksi dan mengidentifikasi informasi yang bermanfaat dan pengetahuan yang terkait dari berbagai database besar/data Warehouse (Turban, 2013), sedangkan metode K-Means merupakan salah satu metode pengelompokan (clustering) dan non hirarki (sekatan) yang berusaha mempartisi data yang ada kedalam bentuk dua atau lebih kelompok (Prasetyo, 2012).

Dalam melakukan penelitian ini, peneliti mengacu pada penelitian-penelitian terdahulu yang pernah dilakukan. Seperti penelitian oleh najia salsabila, (2019), topik pembahasan yang dilakukan yaitu klasifikasi barang menggunakan metode Clustering K-Means dalam penentuan prediksi barang. Tujuan penelitian ini untuk membantu pengolahan prediksi stok barang sehingga menjadi lebih baik dan dapat memenuhi stok yang paling dibutuhkan pada UKM Mar'ah.

Dalam metode yang digunakan untuk menyelesaikan permasalahan tersebut yaitu dengan pemanfaatan algoritma K-Means Clustering, dikarenakan algoritma ini mempunyai kelebihan mudah diterapkan dan dijalankan, relatif cepat dan mudah untuk dipraktekan untuk data mining. Dan dari hasil yang telah didapatkan dengan menggunakan metode K-Means terhadap pengelompokan barang menggunakan RFM memiliki tingkat akurasi sebesar 70%, sedangkan untuk hasil clustering tanpa menggunakan RFM adalah sebesar 76,7%. dengan detail akurasi kategori penjualan tertinggi dan pemberi keuntungan terbesar.

UD. Anugrah Sukses Mandiri merupakan salah satu distributor penyedia barang food dan non-food yang berdomisili di Cirebon. Distribusi UD. Anugrah Sukses Mandiri meliputi seluruh wilayah 3 Cirebon. Pada saat ini untuk melakukan pemenuhan stok barang memang sudah menggunakan software tetapi masih ada hal yang membuat persediaan barang tidak sesuai dengan data penjualan yang telah berjalan. Jumlah permintaan barang dari konsumen yang fluktuatif mengakibatkan stok yang harus disiapkan menjadi tidak stabil. Selain itu produk yang cukup beragam jenisnya menjadikan manajemen stok yang dilakukan kurang akurat sehingga tidak efektif dan seringkali mengecewakan konsumen karena kekosongan suatu produk barang. Dan disamping

itu dalam hal pengelompokan produk atau barang yang cukup laris dan tidak, masih mengalami kesulitan seperti barang yang seharusnya laris tetapi kurangnya stok karena tingginya permintaan konsumen sedangkan barang yang kurang laris menjadi terus menumpuk di karenakan penjualan yang cukup rendah.

Permasalahan yang terjadi tersebut disebabkan karena mengalami kesulitan dalam menentukan batas stok minimum pada tipe barang yang harus dipenuhi berdasarkan minat konsumen. Untuk menangani masalah tersebut maka dibutuhkan suatu manajemen data persediaan barang yang baik untuk pengolahan data historis transaksi dengan menggunakan teknik data mining dengan metode algoritma K-Means Clustering. Maka dengan data yang sudah dikelompokkan (clustering) menggunakan metode K-means diharapkan dapat mempermudah perusahaan dalam menentukan persediaan barang. Sehingga perusahaan dapat melakukan efisiensi secara cermat dalam hal pengadaan produk/barang.

LANDASAN TEORI

1. Persediaan Barang

Pada umumnya, persediaan (inventory) merupakan barang dagangan yang utama dalam perusahaan dagang. Persediaan termasuk dalam golongan aset lancar perusahaan yang berperan penting dalam menghasilkan laba perusahaan. secara umum istilah persediaan dipakai untuk menunjukan barang-barang yang dimiliki untuk dijual kembali atau digunakan untuk memproduksi barang-barang yang akan dijual. Dalam perusahaan dagang, persediaan merupakan barang-barang yang diperoleh atau dibeli dengan tujuan untuk dijual kembali tanpa mengubah barang itu sendiri.

Persediaan adalah barang-barang yang akan disimpan untuk digunakan atau dijual pada masa yang akan datang (Ristono, 2009). Sedangkan menurut ikatan akuntan Indonesia pengertian persediaan barang adalah aset tersedia untuk dijual dalam kegiatan usaha normal, dan dalam proses produksi atau dalam perjalanan maupun dalam bentuk bahan perlengkapan (supplies) untuk digunakan dalam proses produksi atau pemberi jasa.

Dan dari pendapat lain persediaan merupakan suatu aktiva yang meliputi barang-barang milik perusahaan dengan maksud untuk dijual dalam suatu periode usaha tertentu atau persediaan barang-barang yang masih dalam pengerjaan atau proses produksi ataupun persediaan bahan baku yang menunggu penggunaannya dalam proses produksi (Gumelar & alexandr, 2019).

Jenis persediaan biasanya mencakup beberapa jenis persediaan seperti persediaan bahan mentah, persediaan bahan setengah jadi, dan persediaan barang jadi (barang dagangan) (Hanafi, 2010). Bahan mentah adalah

bahan yang akan digunakan untuk memproduksi barang dagangan. Barang setengah jadi adalah barang yang belum selesai sepenuhnya menjadi barang dagangan. Barang jadi ialah barang yang sudah selesai dikerjakan dan siap untuk dijual.

Pada dasarnya jenis jenis persediaan barang adalah persediaan barang mentah, barang dalam proses dan barang persediaan jadi. Perusahaan menggunakan jenis persediaan barang jadi yang didapat dengan cara dibeli dengan tujuan dijual kembali tanpa mengubah bentuk fisik barang dagangan tersebut.

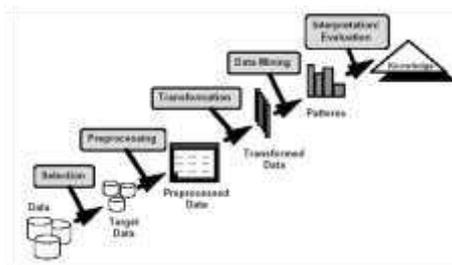
2. Data Mining

Data mining adalah proses yang mempekerjakan suatu atau lebih teknik pembelajaran komputer (machine learning) untuk menganalisa dan mengekstraksi pengetahuan (Knowledge) secara otomatis. Atau dapat disimpulkan data mining adalah suatu proses menganalisa dan mengekstrak dari sekumpulan data yang sangat besar serta digunakan untuk membuat suatu keputusan bisnis yang sangat penting secara otomatis.

1. Tahapan proses data mining

Istilah data mining dan knowledge discovery in database (KDD) seringkali digunakan secara bergantian untuk menjelaskan proses penggalian informasi tersembunyi dalam suatu basis data yang besar. Dan tahapan dalam data mining ini adalah seperti proses KDD yang secara garis besar dapat dijelaskan sebagai berikut (Luthfi, 2009).

- a. Data selection
Pemilihan (seleksi) data baru sekumpulan data operasional perlu dilakukan sebelum tahap penggalian informasi dalam data mining dimulai.
- b. Pre-processing/Cleaning
Sebelum proses data mining dapat dilaksanakan, perlu dilakukan proses cleaning pada data yang menjadi fokus KDD. Proses cleaning mencakup antara lain membuang duplikasi data, memeriksa data yang inkonsisten, dan memperbaiki kesalahan pada data, seperti kesalahan cetak (tipografi).
- c. Transformation
Coding adalah proses transformasi pada data yang telah dipilih, sehingga data tersebut sesuai untuk proses data mining.
- d. Data mining
Data mining adalah proses mencari pola atau informasi menarik dalam data terpilih dengan menggunakan teknik atau metode tertentu.
- e. Interpretation/Evaluation
Pola informasi yang dihasilkan dari proses data mining perlu ditampilkan dalam bentuk yang mudah dimengerti oleh pihak yang berkepentingan. Tahap ini merupakan bagian dari proses KDD yang disebut interpretation (Luthfi, 2009).



Gambar 1. Proses KDD

Tahapan yang dilakukan pada proses data mining di atas diawali dari seleksi data dari data sumber ke data target, tahap preprocessing untuk memperbaiki kualitas data, transformasi, data mining serta tahap interpretasi dan evaluasi yang menghasilkan output berupa pengetahuan baru yang diharapkan memberikan kontribusi yang lebih baik.

Data mining dibagi menjadi beberapa kelompok berdasarkan tugas yang dapat dilakukan (Luthfi, 2009), yaitu:

- a. Deskripsi
Terkadang peneliti dan analisis secara sederhana ingin mencoba mencari cara untuk menggambarkan pola dan kecenderungan yang terdapat dalam data.
- b. Estimasi
Estimasi hampir sama dengan klasifikasi, kecuali variabel target estimasi lebih ke arah numerik daripada ke arah kategori. Model dibangun menggunakan record lengkap yang menyediakan nilai dari variabel target sebagai nilai prediksi. Selanjutnya, pada peninjauan berikutnya estimasi nilai dari variabel target dibuat berdasarkan nilai variabel prediksi.
- c. Prediksi
Prediksi hampir sama dengan klasifikasi dan estimasi, kecuali bahwa dalam prediksi nilai dari hasil akan ada di masa mendatang. Beberapa metode dan teknik yang digunakan dalam klasifikasi dan estimasi dapat pula digunakan (untuk keadaan yang tepat) untuk prediksi.
- d. Klasifikasi
Dalam klasifikasi, terdapat target variabel kategori. Sebagai contoh, penggolongan pendapatan dapat dipisahkan dalam tiga kategori, yaitu pendapatan tinggi, sedang, dan rendah.
- e. Pengklasteran
Pengklasteran merupakan pengelompokan record, pengamatan, atau memperhatikan dan membentuk kelas objek-objek yang memiliki kemiripan. Klaster adalah kumpulan record yang memiliki kemiripan satu dengan yang lainnya dan memiliki ketidakmiripan dengan record-record dalam klaster lain. Pengklasteran berbeda dengan klasifikasi yaitu tidak adanya variabel target dalam pengklasteran. Pengklasteran tidak mencoba

untuk melakukan klasifikasi, mengestimasi, atau memprediksi nilai dari variabel target. Akan tetapi, algoritma pengklasteran mencoba untuk melakukan pembagian terhadap keseluruhan data menjadi kelompok-kelompok yang memiliki kemiripan (homogen), yang mana kemiripan record dalam satu kelompok akan bernilai maksimal, sedangkan kemiripan dengan record dalam kelompok lain akan bernilai minimal.

f. Asosiasi

Tugas asosiasi dalam data mining adalah menemukan atribut yang muncul dalam satu waktu. Dalam dunia bisnis lebih umum disebut analisis keranjang belanja.

3. Clustering

Analisis klaster atau clustering dipilih dalam menyelesaikan masalah ini dikarenakan clustering bertujuan menemukan kelompok (cluster) objek yang berguna, dimana digunakannya tergantung dari tujuan analisa data (Hermawati, 2005). Analisis cluster dapat dibagi menjadi teknik pengelompokan hierarchial (hirarki) dan teknik pengelompokan non-hierarchial (non hirarki)

Teknik klaster termasuk teknik yang sudah cukup banyak dikenal dan banyak dipakai dalam data mining. Sampai sekarang para ilmuwan dalam bidang data mining masih melakukan berbagai usaha melakukan perbaikan model cluster, karena metode yang dikembangkan sekarang masih bersifat heuristik (Budi, 2007).

4. K - Means

K-Means merupakan salah satu metode pengelompokan dan non hierarki (sekatan) yang berusaha mempartisi data yang ada ke dalam bentuk dua atau lebih kelompok (Prasetyo, 2012). Metode ini mempartisi data ke dalam kelompok sehingga data berkarakteristik sama di masukan ke dalam satu kelompok yang sama dan data yang berkarakteristik berbeda dikelompokkan ke dalam kelompok yang lain.

Algoritma K-Means memiliki performa yang baik dari algoritma lainnya, baik dari sisi nilai average within centroid distance dan kompleksitas waktu. Algoritma K-Means menggunakan centroid (rata-rata) sebagai model dari cluster.

Algoritma K-Means merupakan algoritma clustering sederhana yang bersifat tanpa arahan (unsupervised). Misalkan D adalah sebuah dataset dari n objek, dan k adalah jumlah cluster yang akan dibentuk, algoritma partisi mengatur objek-objek tersebut kedalam partisi k ($k \leq n$), dimana setiap partisi menggambarkan sebuah cluster. Setiap cluster di bentuk untuk mengoptimalkan kriteria partisi, seperti fungsi perbedaan berdasarkan jarak, sehingga objek-objek di dalam sebuah cluster adalah mirip, sedangkan objek-objek pada cluster yang berbeda adalah tidak mirip dalam hal atribut dataset. Persamaan untuk

menghitung jarak antar data pada K-Means menggunakan rumus Euclidean Distance (D) yang ditunjukkan pada persamaan

$$D(x_2, x_1) = \sqrt{\sum_{j=1}^p (x_{2j} - x_{1j})^2} \dots \dots \dots$$

Keterangan

P = dimensi data

X₁ = posisi titik 1

X₂ = posisi titik 2

BCV (Between Cluster Variation)

WCV (Within Cluster Variation)

M = Mean (Rata-rata)

D = data

Algoritma standar dari K-Means adalah sebagai berikut

1. langkah 1, tentukan jumlah clustering yang diinginkan (misalkan k3).

2. Langkah 2, pilih centroid awal secara acak. Pada langkah ini secara acak akan dipilih 3 buah data sebagai centroid.

3. Langkah 3, hitung jarak dengan centroid. Pada langkah ini setiap data akan ditentukan centroid terdekatnya, dan data tersebut akan ditetapkan sebagai anggota kelompok yang terdekat dengan centroid.

Untuk menghitung jarak ke centroid masing-masing cluster

Misalkan data (x,y), centroid M1 : (a1,b1), centroid M2 : (a2,b2), centroid M3 : (a3,b3).

$$DM1 = \sqrt{(x - a1)^2 + (y - b1)^2} = ?$$

$$DM2 = \sqrt{(x - a2)^2 + (y - b2)^2} = ?$$

$$DM3 = \sqrt{(x - a3)^2 + (y - b3)^2} = ?$$

Buat tabel hasil perhitungan jarak selengkapnya antara masing masing data dengan centroid, maka didapatkan keanggotaan dari masing masing cluster.

Pada langkah ini dihitung pula rasio antara BCV (Between Cluster Variation) dengan WCV (Within Cluster Variation), Karena centroid M1 = (a1,b1), M2 = (a2,b2), M3 = (a3,b3),

$$d(m1, m2) = \sqrt{(a1 - a2)^2 + (b1 - b2)^2} = ?$$

$$d(m1, m3) = \sqrt{(a1 - a3)^2 + (b1 - b3)^2} = ?$$

$$d(m2, m3) = \sqrt{(a2 - a3)^2 + (b2 - b3)^2} = ?$$

$$BCV = d(m1, m2) + d(m1, m3) + d(m2, m3) = ?$$

Dalam hal ini d(mi,mj) menyatakan jarak euclidean dari m ke mj menghitung WCV yaitu dengan memilih jarak terkecil yang terdapat pada tabel keanggotaan.

$$WCV = c1^2 + c2^2 + c3^2 + N = ?$$

$$\text{Sehingga beasr rasio} = BCV/WCV = ?$$

Karena langkah ini merupakan iterasi 1 maka lanjutkan ke langkah berikutnya.

4. Langkah 4, pembaruan centroid dengan menghitung rata rata nilai pada masing-masing cluster.

Setelah menghitung rata-rata nilai pada masing-masing cluster didapatkan centroid baru yaitu : $M1 = (a1,b1)$, $M2 = (a2,b2)$, $M3 = (a3,b3)$.

- Langkah 5, (uterasi ke 2) kembali kelangkah ke 3, jika masih ada data yang berpindah cluster atau jika nilai centroid di atas ambang. Atau jika pada fungsi objektif yang digunakan masih diatas ambang. Jika tidak maka iterasi dihentikan.

Maka dapat disimpulkan bahwa K-Means adalah salah satu algoritma dalam data mining yang berfungsi untuk menentukan divisi terbaik entitas n kedalam K (disebut cluster), sehingga total antara jarak anggota kelompok dan centroid sesuai, terlepas dari kelompok diminimalkan. Setiap entitas milik cluster dengan mean terdekat.

5. Tools RapidMiner

Ada banyak tools yang tersedia untuk data mining. Tujuan utama data mining tools adalah untuk menemukan data, mengekstrak data, menyaring data, mendistribusikan informasi dan memonetisasinya. Dalam penelitian ini akan menggunakan tools yang bernama RapidMiner.

RapidMiner adalah sebuah lingkungan machine learning data mining, text mining dan predictive analytics (Han and Kamber dalam artikel Rima Dias Ramadhani, 2013 : 3). RapidMiner adalah salah satu software untuk pengolahan data mining. Pekerjaan yang dilakukan oleh RapidMiner text mining adalah berkisar dengan analisis teks, mengekstrak pola-pola dari dataset yang besar dan mengkombinasikannya dengan metode statistika, kecerdasan buatan, dan database. Tujuan dari analisis teks ini adalah untuk mendapatkan informasi bermutu tertinggi dari teks yang diolah. RapidMiner menyediakan prosedur data mining dan machine learning, di dalamnya termasuk: ETL (extraction, transformation, loading), data preprocessing, visualisasi, modelling dan evaluasi. Proses data mining tersusun atas operator-operator yang nestable, dideskripsikan dengan XML, dan dibuat dengan GUI. Penyajiannya dituliskan dalam bahasa pemrograman Java.

METODE PENELITIAN

Berdasarkan pada landasan teori yang telah dijabarkan pada Bab sebelumnya, pada Bab ini menjelaskan cara kerja algoritma yang akan digunakan dan proses yang akan dibangun untuk melakukan pengelompokan.

1. Pendekatan Penelitian

Pendekatan penelitian adalah keseluruhan cara atau kegiatan dalam suatu penelitian yang dimulai dari perumusan masalah sampai suatu kesimpulan. Pendekatan penelitian ada dua macam yaitu pendekatan kuantitatif dan pendekatan kualitatif. Pendekatan kuantitatif artinya informasi atau data yang disajikan berupa

angka sedangkan pendekatan kualitatif yaitu informasi atau data yang disajikan berupa pernyataan.

Berdasarkan uraian yang telah dikemukakan di atas bahwa dapat disimpulkan penelitian ini menggunakan pendekatan kuantitatif karena data yang diperoleh berupa angka-angka mulai dari pengumpulan data, penafsiran, serta penampilan hasilnya.

Jenis penelitian yang digunakan dalam penelitian ini adalah jenis penelitian terapan. Penelitian terapan adalah sebuah penelitian untuk menghasilkan sesuatu yang langsung bisa diterapkan untuk memecahkan suatu masalah. Pada analisis ini menerapkan metode K-Means clustering untuk menentukan persediaan barang berdasarkan banyaknya barang yang keluar, barang yang tidak terlalu banyak dan tidak terlalu sedikit, dan barang yang keluar cukup sedikit. Data penelitian ini diambil dari objek penelitian pada UD.ASM Cirebon.



Gambar 2. Alur Penelitian

3. Tahapan Penelitian

Ada beberapa tahapan dalam melakukan suatu penelitian. Pada penelitian ini menggunakan metode KDD, yaitu

- Selected data atau teknik pengumpulan data Untuk meningkatkan akurasi dan kualitas hasil data mining. Data diambil dengan rentang waktu satu tahun terhitung dari 1 november 2020 – 30 november 2021 dengan kisaran 12 bulan, data barang yang diambil merupakan produk non food dan food dari UD. Anugerah Sukses Mandiri berupa kamper pakaian, pewangi ruangan, handbody, sabun, coklat, tepung kue dsb, dengan banyaknya produk sebanyak 361 item.
- Preprocessed data atau Pembersihan data pada tahap ini yaitu melakukan pembersihan data karena data yang diperoleh masih berupa data mentah seperti data yang kurang lengkap dan atribut yang tidak relevan, data yang masih mentah ini dapat menyebabkan hasil yang kurang akurat dan bisa menimbulkan error pada saat pemrosesan data sehingga pada proses ini perlu dilakukan pengolahan data atau pembersihan data.

Pada proses ini dapat menghapus data yang tidak lengkap dan yang tidak bersinambung dikarenakan produk tersebut sudah tidak diproduksi dan variabel serta kolom yang tidak memiliki nilai. Pada proses preprocessed atau pembersihan data menggunakan file yang berformat Excel. Dari total data yang di gunakan untuk penelitian ini sebanyak 361 data dan setelah dilakukan pembersihan maka data yang dapat digunakan untuk penelitian sebanyak 249 data.

3. Data mining

Data mining mengolah data yang tersimpan sehingga menghasilkan informasi yang berguna. Pengolahan data dapat dimanfaatkan untuk pembuatan laporan, mengontrol kebutuhan, menemukan informasi dan pengetahuan baru. Pada penelitian ini data mining yang digunakan yaitu menggunakan metode K-Means Clustering dengan menggunakan software Rapidminer. Metode ini mempartisi data menjadi berkelompok yang sama dimasukkan pada karakteristik yang sama dan data berkelompok yang berbeda dikelompokkan ke dalam kelompok yang lain. Pada proses data mining menggunakan software RapidMiner ini kita dapat membagi data menjadi beberapa cluster atau kelompok, disini peneliti akan membagi data menjadi 2 cluster yaitu cluster terendah dan tertinggi. Setelah proses data mining berhasil teroutput kita dapat mengetahui data mana saja yang menjadi kelompok pada tiap cluster, dan hasil centroid atau titik pusat pada setiap cluster.

4. Interpretation atau Evaluation

Tahap ini mengetahui dari hasil penggalian dan pengolahan data terhadap sampel yang telah dihasilkan mengenai penentuan stok produk untuk jumlah terendah, sedang dan terbanyak yang telah diproses oleh RapidMiner. Hasil tersebut dapat berupa gambar atau tabel dan digunakan sebagai sumber pengetahuan yang digunakan untuk tujuan penelitian lebih lanjut dan mempresentasikan pengetahuan dalam bentuk yang mudah dipahami. Selanjutnya proses data akan dihitung tingkat akurasi dengan menggunakan metode confusion matrix sehingga akan diketahui jumlah akurasi dari proses yang didapat. Pada tahap ini juga kita dapat melihat apakah data yang dimasukkan sesuai untuk menyelesaikan masalah pada penelitian ini

HASIL DAN PEMBAHASAN

Proses data mining pada penelitian ini menggunakan program RapidMiner 9.8. program tersebut digunakan untuk membantu penelitian dalam proses mengelompokkan data ke dalam klaster dengan menggunakan salah satu metode algoritma klasterisasi K-Means. Data yang akan

digunakan untuk diolah pada penelitian ini meliputi data penjualan barang, dan data persediaan barang yang ada pada UD. ASM. Sehingga hasil dari olahan data-data tersebut akan menghasilkan output yang mana bisa dijadikan sebuah informasi yang cukup berguna. Dan informasi tersebut berupa kelompok data yang tercluster menjadi 2 bagian yaitu dapat diketahui item barang apa saja yang terjual dengan jumlah banyak dan terjual dengan jumlah sedikit. Sehingga dengan penelitian ini dapat melakukan sebuah proses penyetokan barang dengan lebih tepat, dan dapat terhindar dari proses penyetokan barang yang sia-sia.

Hasil Penelitian

Berikut adalah hasil penelitian yang dilakukan :

1. Pengumpulan data

a. Data Penjualan

Nama Barang	Jumlah	Unit	Harga Satuan	Mata Uang
Dahlia Banded K240g Buy 2 Get 1	516	Pcs	18000	Idr
Dahlia Bc Disf Spray Refill 1000ml	72	Pcs	51000	Idr
Dahlia Bc Disf Spray Refill 500ml	431	Pcs	25909.1	Idr
Dahlia Bc Disf Spray W/Noz 500ml	1006	Pcs	28685.93	Idr
Dahlia Bc-001 Blue Clean	452	Pcs	7500	Idr
Dahlia Bc-002 Blue Clean 2pcs	5999	Pcs	15443.63	Idr
Dahlia D-112 Serap Lembab	225	Pcs	17273.33	Idr
Dahlia D-113 Banded Buy 2 Get 1	3	Pcs	18000	Idr
Dahlia D-113 Serap Lembab Paket Refill	1532	Pcs	9502.95	Idr
Dahlia D-115 Serap Lembab Wangi	37	Pcs	17725	Idr
Dahlia D-116 Refill Serap Lembab Wangi	114	Pcs	9550	Idr
Dahlia D-117 Serap Bau Kulkas	126	Pcs	14100.07	Idr
Dahlia D-118 Serap Bau Lemari	3	Pcs	14100	Idr
Dahlia D-119 Serap Bau Sepatu	10	Pcs	14100	Idr
Dahlia D-120 Shoe Deodorizer	28	Pcs	13362.14	Idr
Dahlia Dis Aerosol Eucalyptus Mint	687	Pcs	18700	Idr
Dahlia Disinfectant Aerosol Classic	1741	Pcs	18704.3	Idr
Dahlia F-601 Air Freshener Apple	18227	Pcs	8302.07	Idr

Gambar 3. data penjualan

Diatas ini merupakan tampilan data penjualan pada UD. ASM per tanggal 1 November 2020 – 30 November 2021 dengan jumlah sebanyak 361 data, yang sedang digunakan dalam bahan penelitian.

b. Data Persediaan

Bc002b	Dahlia Bc-002 Blue Clean 2pcs Buy 2=1	1.00	Pcs
D112	Dahlia D-112 Serap Lembab	32.00	Pcs
D112b	Dahlia D-112 Buy 1 Get 1 Serap Lembab Banded (D113)	0.00	Pcs
D113	Dahlia D-113 Serap Lembab Paket Refill	176.00	Pcs
D113b	Dahlia D-113 Banded Buy 2 Get 1	0.00	Pcs
D115	Dahlia D-115 Serap Lembab Wangi	37.00	Pcs
D115b	Dahlia D-115 Buy 1 Get 1 Serap Lembab Wangi Banded D116	0.00	Pcs
D116	Dahlia D-116 Refill Serap Lembab Wangi	1.00	Pcs
D116b	Dahlia D-116 Buy 2 Get 1	0.00	Pcs
D117	Dahlia D-117 Serap Bau Kulkas	45.00	Pcs
D118	Dahlia D-118 Serap Bau Lemari	120.00	Pcs
D119	Dahlia D-119 Serap Bau Sepatu	35.00	Pcs
D120	Dahlia D-120 Shoe Deodorizer	46.00	Pcs
Dis	Dahlia Disinfectant Aerosol Classic	472.00	Pcs
Dis110	Dahlia Bc Disf Spray Refill 1000ml	115.00	Pcs
Dis500	Dahlia Bc Disf Spray W/Noz 500ml	9.00	Pcs
Dis501	Dahlia Bc Disf Spray Refill 500ml	196.00	Pcs
Disa	Dahlia Dis Aerosol Eucalyptus Mint	323.00	Pcs
F6010m	Dahlia F-601 Air Freshener Mix	364.00	Pcs

Gambar 4. Data persediaan

Diatas ini merupakan tampilan data Persediaan pada UD. ASM dalam bentuk format tabel, yang sedang digunakan dalam bahan penelitian.

2. Data Cleaning / pemrosesan data

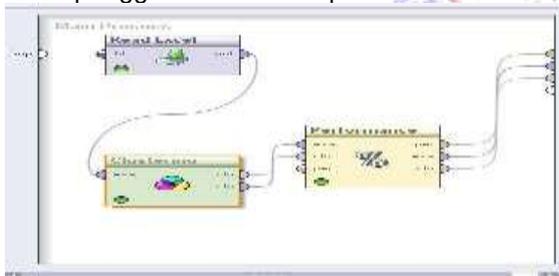
No	Nama Barang	Stok Awal	Terjual	Stok Terakhir	Vol Penjualan	Rata-Rata Penjualan
1	Dahlia Bc-001 Bisc Clean	825.00	675.00	150.00	487.50	1.38
2	Dahlia Bc-002 Bisc Clean 2pac	7374.00	7155.00	219.00	5796.50	1.88
3	Dahlia D-112 Serap Lembab	311.00	279.00	32.00	171.50	1.63
4	Dahlia D-113 Serap Lembab Paket Refill	1989.00	1813.00	176.00	1082.50	1.67
5	Dahlia D-113 Banded Buy 2 Get 1	6.00	6.00	0.00	3.00	2.00
6	Dahlia D-113 Serap Lembab Wangi	114.00	77.00	37.00	75.50	1.02
7	Dahlia D-116 Refill Serap Lembab Wangi	163.00	162.00	1.00	82.00	1.98
8	Dahlia D-117 Serap Bau Kulkas	187.00	142.00	45.00	116.00	1.22
9	Dahlia D-118 Serap Bau Lemari	123.00	3.00	120.00	121.50	0.02
10	Dahlia D-119 Serap Bau Sepatu	65.00	10.00	35.00	40.00	0.25

Gambar 5. Data yang sudah proses Cleaning

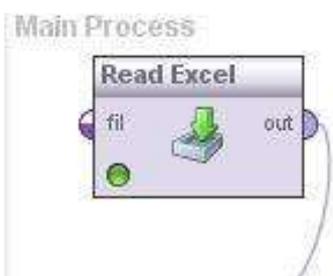
Pada gambar diatas tahap awal yang dilakukan adalah melakukan pembersihan data, data yang kurang relevan dan menambahkan atribut yang dibutuhkan. Data yang diproses harus dalam keadaan normal, bebas dari missing value dan redundansi data.

3. Data Mining

Pada tahap ini melakukan penggalian data digunakan untuk menemukan pola-pola yang tidak diketahui sebelumnya menjadi informasi. Untuk membantu melakukan proses penggalian data. Penelitian ini menggunakan perangkat lunak RapidMiner, untuk dapat melakukan proses data yang sesuai dengan teknik data mining maka digunakan operator-operator yang telah tersedia pada tools RapidMiner. Dibawah ini adalah model proses penggunaan tools RapidMiner



Gambar 6. Model Proses



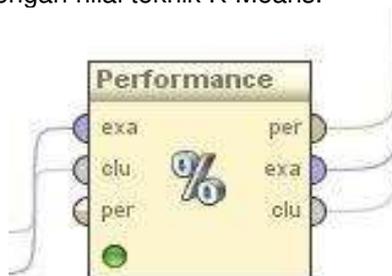
Gambar 7. Read excel

Operator Read Excel ini digunakan untuk memanggil data atau membuka file yang berekstensi xls (data dalam bentuk excel)



Gambar 8. operator K-Means

Operator ini digunakan untuk proses pengolahan data dengan nilai teknik K-Means.



Gambar 9. Cluster distance performance

Operator ini digunakan untuk evaluasi kinerja metode Clustering berbasis centroid. Operator ini memberikan daftar nilai kriteria kinerja berdasarkan centroid cluster.

4. Evaluation hasil implementasi tools RapidMiner

Setelah model proses dibuat, selanjutnya adalah hasil dari mengeksekusi data dengan menggunakan RapidMiner. Berikut hasil dari pemrosesan K-Means Clustering.

Cluster Model

```
Cluster 0: 42 items
Cluster 1: 207 items
Total number of items: 249
```

Gambar 10. Cluster Model

Dari gambar di atas menunjukan bahwa data yang tertera sebanyak 249 record data. terdiri dari cluster 0 sebanyak 42 item dan cluster 1 sebanyak 207 item. Dengan total keseluruhan sebanyak 249 item.

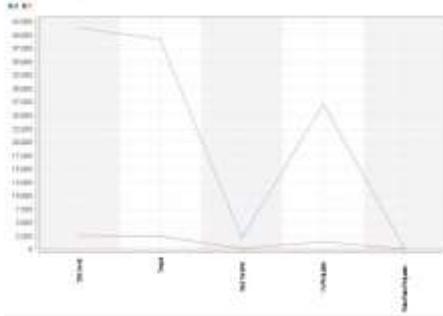
Penentuan cluster dalam penelitian ini dilakukan pengelompokan berdasarkan karakter barang yang ada pada data persediaan UD. Anugerah Sukses Mandiri yang dibagi menjadi 2 cluster yaitu cluster tertinggi dan cluster terendah, untuk cluster tertinggi menjadikan stok yang stabil. Sedangkan untuk cluster terendah menjadikan persediaan yang harus di restock yaitu apakah

data barang tersebut harus ditambahkan kembali atau discontinue/dihentikan.

Atribut	cluster_0	cluster_1
Stok awal	41350,024	2604,976
Terjual	39260,071	2394,546
Stok Terakhir	2089,952	210,430
Vol Penjualan	27164,036	1407,793
Rata-Rata Penjualan	1,888	1,583

Gambar 11. Hasil Centroid

Dari gambar menunjukkan hasil setiap nilai centroid pada tiap cluster, untuk mengetahui bagaimana pola dari tiap clusternya. Untuk dapat melihat hasil plot klastering dapat dilihat pada gambar berikut



Gambar 12. Plot Klastering

Berdasarkan hasil dari gambar plot klastering maka dapat disimpulkan bahwasannya klaster 0 merupakan kelompok yang memiliki tingkat tertinggi dengan rata-rata nilai lebih di atas 10.000. sedangkan untuk klaster 1 merupakan kelompok yang memiliki tingkat terendah yang memiliki nilai rata-rata yang bervariasi dengan rata-rata nilai pada tiap data dibawah 10.000. Agar hasil klaster lebih mudah dimengerti, maka nama klaster diubah sesuai dengan rata-ratanya. Sebagai berikut

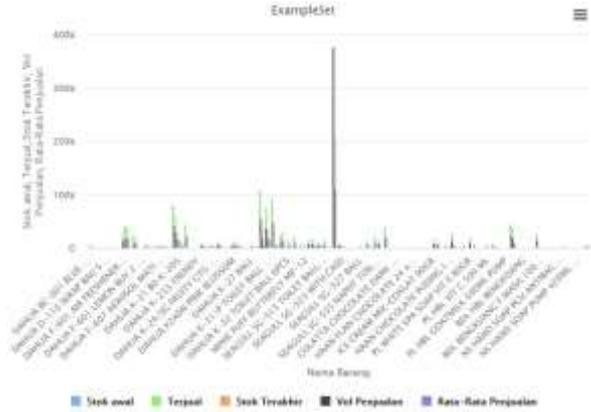
Table 1. Pembagian Cluster

Klaster	Nama Klaster
Klaster 0	Tinggi
Klaster 1	Rendah

cluster	Nama Barang	Stok awal	Terjual	Stok Terakhir	Vol Penjualan	Rata-Rata Penjualan
cluster_0	DAHJUA F-601 AIR FRESHER APPLE	29184,0	28453,0	731,0	14957,5	1,9
cluster_0	DAHJUA F-601 AIR FRESHER LEMON	42464,0	42181,0	283,0	21373,5	2,0
cluster_0	DAHJUA F-601 AIR FRESHER ORANGE	40157,0	39893,0	264,0	20210,5	2,0
cluster_0	DAHJUA F601 AIR FRESHER CHERRY BLOSSOM	25403,0	25232,0	171,0	12787,0	2,0
cluster_0	DAHJUA F601 AIR FRESHER JUIT PUNCH	21098,0	20900,0	998,0	10048,0	1,9
cluster_0	DAHJUA K-19/17 NAPH 25GR	79286,0	72076,0	7210,0	43248,0	1,7
cluster_0	DAHJUA K-19/W NAPH 25GR	99553,0	96242,0	3311,0	31432,0	1,8
cluster_0	DAHJUA K-205 NAPH 300GR	31553,0	30387,0	3166,0	16399,5	1,9
cluster_0	DAHJUA K-21 NAPH 300GR	13379,0	13155,0	224,0	6601,5	1,9
cluster_0	DAHJUA K-22 NAPH 350GR	44451,0	44348,0	103,0	22277,0	2,0
cluster_0	DAHJUA K24/20N WHITE DAISY	11183,0	10984,0	199,0	5691,0	1,9

Gambar 13. Data pembagian Cluster

Untuk mengetahui daftar dari tiap data yang sudah dikelompokkan pada tiap tiap cluster, gambar diatas adalah daftar data pembagian cluster



Gambar 14 Data visualisasi

Pada gambar 14 di atas merupakan data visualisasi diagram setiap data.

Pembahasan

Berdasarkan hasil dari penelitian yang telah dilakukan untuk pembagian persediaan data barang menggunakan algoritma K-Means Clustering dengan aplikasi RapidMiner, didapatkan bahwa dengan data yang digunakan menghasilkan beberapa produk yang sesuai untuk lebih ditingkatkan dalam persediaan barang. Berikut tabel dari hasil proses pada RapidMiner

Table 2. Hasil proses RapidMiner

Keterangan	Proses
	RapidMiner
Jumlah Iterasi	-
Jumlah Data	
Cluster 0	42 item
Cluster 1	207 item
Total	249 item
Terjual Banyak	Cluster 0
Terjual Sedikit	Cluster 1
Centroid Akhir	
Cluster 0	
Stok Awal	41350,024
Terjual	39260,071
Stok Terakhir	2089,952
Volume Penjualan	27164,036
Rata - Rata Penjualan	1,888
Cluster 1	
Stok Awal	2604,976
Terjual	2394,546
Stok Terakhir	210,430

Keterangan	Proses
	RapidMiner
Volume Penjualan	1407,703
Rata - Rata Penjualan	1,503

Dilihat dari tabel diatas maka didapatkan hasil sebagai berikut. Cluster kedua(1) memiliki 207 data dan dapat diartikan sebagai kelompok untuk jumlah terendah sehingga jumlah stok harus di evaluasi agar jumlah produk dapat diproses untuk melakukan peningkatan atau discontinue pada barang yang tidak berkembang sehingga tidak menumpuk pada data stok.

Cluster pertama(0) yaitu cluster terakhir memiliki 42 data dan dapat diartikan sebagai kelompok produk untuk jumlah tinggi atau terbanyak . produk yang termasuk dalam kategori ini harus terus dipertahankan dan terus di sesuaikan antara jumlah stok dan jumlah penjualan.

Dalam melakukan evaluasi terhadap performance algoritma machine learning, kita dapat menggunakan acuan confusion matrix. Confusion matrix mempresentasikan prediksi dan kondisi sebenarnya.

	True cluster_0	True cluster_1	Class precision
Pred cluster_0	42	0	100.00%
Pred cluster_1	0	207	100.00%
Class recall	100.00%	100.00%	

Gambar 15. deskripsi akurasi confusion matrix

Berdasarkan penelitian pada gambar 15 yang telah dilakukan, setelah data yang masuk dan diambil maka di uji tingkat akurasi dengan menggunakan confusion matrix. Perhitungan akurasi dari 249 data, maka hasil akurasi yang di dapat yaitu sebesar 100.00%.

Untuk lebih jelasnya kita lihat pada tabel dibawah sebagai berikut:

Table 3. hasil akurasi data

Accuracy: 100.00 +/- 0.00%(micro average: 100.00%)			
	True cluster_0	True cluster_1	Class precision
Pred. cluster_0	42	0	100.00%
Pred. cluster_1	0	207	100.00%
Class recall	100.00%	100.00%	

KESIMPULAN

Berikut ini adalah kesimpulan yang dapat diambil setelah melakukan penelitian tentang clustering data persediaan barang dengan menggunakan metode K-Means, diantaranya sebagai berikut:

1. Clustering data barang dapat dilakukan dengan menggunakan K-Means Clustering pada UD. Anugerah Sukses mandiri. Pada metode ini barang dikelompokkan menjadi satu dan dicari nilai centroid. Sehingga lebih mudah memilah data barang menjadi barang banyak terjual, terjual sedang dan terjual sedikit.
2. Hasil evaluasi dari clustering data barang menggunakan K-Means dengan dataset sebanyak 249 data menghasilkan 2 kluster yaitu kluster 0 dengan dengan kluster terbanyak dan kluster 1 dengan kluster terendah/sedikit. Dari hasil data tersebut didapatkan pula nilai akurasi sebesar 100.00% sehingga dapat ditarik kesimpulan bahwa algoritma K-means mempunyai tingkat akurasi yang tinggi. Nilai ini juga dapat berubah seiring dengan penambahan data.

Saran

Berikut ini beberapa saran terhadap clustering data barang menggunakan K-Means, agar penelitian ini dapat dikembangkan lagi kedepannya.

1. Agar hasil penelitian ini lebih baik lagi, dapat menggunakan data yang lebih banyak dari penelitian sekarang dan melakukan pengujian dan melakukan pengujian dataset atribut yang lebih lengkap sehingga model yang dihasilkan akan lebih teruji.
2. Pada penelitian berikutnya, diharapkan agar dapat dikomparasi menggunakan metode clustering lain seperti Fuzzy C-Means dan sebagainya.

DAFTAR PUSTAKA

- [2]. Assauri, S. (2005). *marihot dan dearlina sinaga* (p. 50).
- [3]. Budi, S. (2007). *Data Mining Teknik Pemanfaatan Data untuk Keperluan Bisnis*. In *Graha Ilmu*.
- [4]. Gumelar, A., & alexandr. (2019). *Pengertian Persediaan Menurut Para Ahli*. In *Persediaan Menurut Para Ahli*.
- [5]. Hanafi, M. (2010). *Persediaan Barang*.
- [6]. Hermawati, F. A. (2005). *Data Mining - Algoritme dan Implementasi Menggunakan Bahasa Pemrograman PHP. Mining of Massive Datasets*.
- [7]. Luthfi, K. A. E. T. (2009). *Algoritma Data Mining Yogyakarta. February*, 149–176.
- [8]. Prasetyo, E. (2012). *Data Mining: Konsep Dan Aplikasi Menggunakan MATLAB* (Vol. 3, Issue 1, pp. 27–28). Penerbit ANDI. Yogyakarta.
- [9]. Ristono. (2009). *Definisi Persediaan Barang*. In *Persediaan Barang*.
- [10]. Turban, E. (2013). *Implementasi Data Mining Algoritma Apriori pada sistem persediaan alat-alat kesehatan*.