

Penerapan Metode K-Means Clustering Untuk Analisa Penjualan Komoditas Toko Tani Indonesia

Dzulfan Abid¹, Rahmatullah Wirya Adikusuma², Allif Mufti Al Fikri³, dan Rinci Kembang Hapsari^{4*}

^{1,2,3,4}Teknik Informatika, Institut Teknologi Adhi Tama Surabaya

*Penulis Korespondensi: rincikembang@itats.ac.id

ABSTRACT

Agricultural Commodities are the results of farming activities that can be traded, stored, and exchanged. Toko Tani Indonesia Center is a shop that buys agricultural products from farmers and sells agricultural materials and products to other farmers and customers. In its operations, it experienced problems related to the limited capacity of the shop warehouse. So, in this study grouping of goods was carried out at the Tani Indonesia Center shop. The grouping carried out in this study uses the K-Means algorithm. Algorithm testing uses goods sales data at the Tani Indonesia Center shop. Three types of data groups are produced based on the sample data taken at the Tani Indonesia Center shop. The data groups are insufficient sales data, moderate sales data, and high sales data. So that with this data grouping, the Toko Tani Indonesia Center can find out the types of goods that are selling best and which are not so that the goods in the warehouse do not accumulate.

Article History

Received 30-01-2023
Revised 08-02-2023
Accepted 08-02-2023

Key words

Analisis sentiment
Support vector machine
Sequential minimal
optimization

ABSTRAK

Komoditas Tani adalah hasil dari usaha tani yang dapat diperdagangkan, disimpan, dan dipertukarkan. Toko Tani Indonesia Center merupakan toko yang membeli hasil pertanian dari petani, serta menjual bahan peryanian dan hasil pertanian ke petani dan customer yang lain. Dalam operasionalnya mengalami kendala terkait dengan keterbatasan kapasitas gudang toko. Sehingga dalam penelitian ini dilakukan pengelompokan barang di toko Tani Indonesia Center. Pengelompokan yang dilakukan dalam penelitian ini menggunakan algoritma K-Means. Pengujian algoritma yang dilakukan dengan menggunakan data penjualan barang di toko Tani Indonesia Center. Berdasarkan sampel data yang diambil di toko Tani Indonesia Center, maka menghasilkan tiga jenis kelompok data. Kelompok data tersebut, yaitu data penjualan rendah, data penjualan sedang, dan data penjualan tinggi. Sehingga dengan adanya pengelompokan data ini pihak tokoTani Indonesia Center dapat mengetahui jenis barang yang laris terjual dan tidak. Sehingga barang yang ada di gudang tidak menumpuk..

PENDAHULUAN

Komoditas adalah sebuah barang atau produk yang dapat diperdagangkan. Tentu saja tujuan dari jual beli komoditas adalah untuk memperoleh keuntungan, sedangkan Komoditas Tani adalah hasil usaha tani yang dapat diperdagangkan, disimpan, dan dipertukarkan[1]. Komoditas tani termasuk ke dalam jenis komoditas lunak yang didapatkan dari hasil perhutanan, pertanian, dan peternakan[2]. Contohnya seperti gula, beras, kedelai, susu sapi, jagung, biji kopi, sawit dan lain sebagainya. Dalam era globalisasi, perkembangan kecanggihan teknologi yang semakin pesat merupakan aspek yang dapat dimanfaatkan untuk mencapai kemudahan-kemudahan, tidak terkecuali dalam arus informasi. Kemajuan teknologi tersebut terlihat semakin marak dengan penggunaan komputer yang memang sudah sangat luas diberbagai bidang kehidupan misalnya di bidang pendidikan, kesehatan, hiburan, terlebih pada bidang bisnis yang semuanya itu menuntut penggunaan dari komputer.

Sayangnya dalam toko, mini market dan swalayan masih ada proses-proses yang dilakukan secara manual sehingga sering terjadi kesalahan dalam pecatatan data-data yang ada dan juga kurangnya efisiensi waktu yang diperlukan. Seperti halnya pada Toko Tani Indonesia Center. Pada saat ini Toko Tani Indonesia Center masih memproses data penjualannya secara manual. Disamping itu Toko Tani Indonesia Center tidak dapat mengelompokkan produk yang laris dan yang tidak laris terjual. Sehingga kesulitan yang dialami yaitu seringnya kekurangan stok produk yang laku

karena penjualannya tinggi, dan menumpuknya produk yang tidak laku di gudang karena penjualannya rendah.

Penelitian pengelompokan data telah banyak dilakukan, diantaranya adalah pengelompokan barang yang laku dipasaran dengan menggunakan algoritma Apriori. Pengelompokan ini digunakan untuk menyelesaikan permasalahan pedagang baru yang terbatas modalnya. Sehingga para pedagang bisa mengetahui barang apa saja yang laris di pasaran[3]. Penelitian pengelompokan lain adalah pengelompokan terhadap data penjualan mesin dan bahan-bahan digital printing dengan menggunakan algoritma Hirarki Divisive. Dari hasil pengelompokan bisa digunakan sebagai alat pendukung untuk mengambil kebijakan di perusahaan[4].

Penelitian ini dilakukan untuk memberikan informasi tentang bagaimana mengelompokkan data penjualan barang. Oleh karena itu tujuan dari penelitian ini adalah untuk mengelompokkan data komoditas tani menjadi 3 kelompok yakni data penjualan rendah, data penjualan sedang, dan data penjualan tinggi.

TINJAUAN PUSTAKA

Data Mining

Data mining dapat didefinisikan sebagai prosedur untuk menentukan pola atau pola data untuk tujuan memperoleh dan menemukan informasi yang berguna dari sekelompok data yang sangat besar. Yang pada akhirnya memungkinkan keputusan atau penilaian tentang suatu hal tertentu. Beberapa alat dan teknik data mining telah dikembangkan dan telah digunakan dalam proyek untuk tujuan data mining spesifik yang terkait dengan jaringan saraf, pohon keputusan, prediksi, pengelompokan, dan klasifikasi. Jenis teknologi eksplorasi ini memiliki aturan dan metode khusus yang membantu menentukan masalah apa yang perlu dipecahkan[5].

Clustering

Clustering juga dikenal sebagai segmentasi. Metode ini digunakan untuk mengidentifikasi grup natural dari suatu kasus berdasarkan grup atribut, mengelompokkan data yang memiliki kemiripan atribut[6].

Clustering merupakan bagian data mining yang bersifat *unsupervised*. Dimana merupakan proses pembagian data kedalam *cluster* berdasarkan tingkat kemiripan. Dalam *clustering*, data yang mempunyai kemiripan digabungkan dalam *cluster* yang sama, sedangkan data yang tidak memiliki kemiripan dimasukkan dalam *cluster* yang berbeda[7].

K-Means

Algoritma K-Means merupakan salah satu metode data *clustering non-hierarki* yang berusaha mempartisi data yang ada ke dalam bentuk satu atau lebih *cluster* atau kelompok sehingga data yang memiliki karakteristik yang sama dikelompokkan ke dalam satu *cluster* yang sama dan data karakteristik yang berbeda dikelompokkan ke dalam kelompok lainnya[8].

Algoritma K-Means mengatur nilai cluster(k) secara acak, sementara nilai yang menjadi pusat cluster biasa disebut centroid, mean atau "means". Kemudian jarak tiap data yang ada terhadap tiap centroid dihitung menggunakan rumus Euclidean sehingga dicari jarak terdekat pada tiap data dengan centroid. Kemudian dilakukan klasifikasi berdasarkan kedekatannya dengan centroid[9].

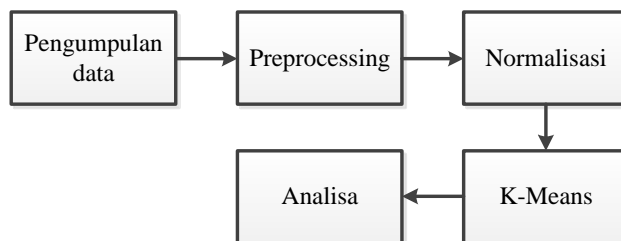
Algoritma K-Means pada dasarnya melakukan dua proses, yakni proses pendeteksian lokasi pusat tiap *cluster* dan proses pencarian anggota dari tiap-tiap *cluster*. Cara kerja algoritma K-Means[10] :

1. Tentukan K sebagai jumlah cluster yang ingin dibentuk.
2. Bangkitkan k centroid (titik pusat cluster) awal secara random.
3. Hitung jarak setiap data ke masing masing centroid
4. Setiap data memilih centroid yang terdekat

5. Tentukan posisi centroid yang baru dengan cara menghitung nilai rata- rata dari data-data yang terletak pada centroid yang sama
6. jika posisi centroid baru dengan centroid yang lama tidak sama ulangi langkah 3

METODE

Penelitian ini dimulai dengan melakukan pengumpulan data dilanjutkan dengan pengolahan data atau preprocessing seperti pengecekan data yang tidak valid, transformasi, dan seleksi fitur. Berikut alur proses penelitian dijelaskan pada Gambar 1.



Gambar 1. Blok Diagram Penelitian

Pengumpulan Data

Data yang dipakai dalam penelitian ini bersumber dari situs Dinas Pangan, Peternakan dan Kesehatan Hewan Provinsi Kalimantan Barat per 13 Mei 2020 yang berisi data penjualan dan stok komoditas di Toko Tani Indonesia Center (TTIC). Dimana dataset yang digunakan ditunjukkan pada Tabel 1.

Tabel 1. Dataset (data penjualan)

Nama Barang	Volume Terjual	Harga Rerata	Total Omzet
Beras	810	8800	7128000
Bawang Merah	12	43000	516000
Bawang Putih	12	23000	276000
Gula Pasir	250	12500	3125000
Garam	15	6000	90000
Minyak Goreng	107	11500	1230500
Telur Ayam	185	1450	268250
Daging Ayam	13	28000	364000
Nuget Ayam	25	17000	425000
Daging Kerbau	56	78000	4368000
Cabe Keriting	26	17000	442000

Preprocessing

Preprocessing dan cleaning data merupakan operasi dasar yang dilakukan seperti penghapusan noise. Proses cleaning mencakup antara lain membuang duplikasi data, memeriksa data yang inkonsisten, dan memperbaiki kesalahan pada data, seperti kesalahan cetak. Data bisa diperkaya dengan data atau informasi eksternal yang relevan.

Tabel 2. Hasil normalisasi dataset

Nama Barang	Volume Terjual	Harga Rerata
Beras	1.0000	0.0960
Bawang Merah	-	0.5428
Bawang Putih	-	0.2815
Gula Pasir	0.2982	0.1444
Garam	0.0038	0.0594
Minyak Goreng	0.1190	0.1313
Telur Ayam	0.2168	-
Daging Ayam	0.0013	0.3468
Nuget Ayam	0.0163	0.2031
Daging Kerbau	0.0551	1.0000
Cabe Keriting	0.0175	0.2031

Normalisasi

Proses normalisasi yang dilakukan dalam penelitian ini untuk mentransformasi nilai atribut/parameter data ke dalam *range* 0 sampai 1. Metode normalisasi yang digunakan adalah min-Max normalization, dengan menggunakan persamaan (1)[12]. Hasil normalisasi dataset ditunjukkan pada Table 2.

$$X^* = \frac{X - \min(X)}{\max(X) - \min(X)} \quad (1)$$

HASIL DAN PEMBAHASAN

Analisa Sistem Aktual

Dari hasil penelitian yang dilakukan di toko tani indonesia center Kalimantan Barat, didapatkan sistem yang sudah berjalan dan digunakan masih manual. Disamping itu, pemilik toko tersebut mempunyai kesulitan dalam mengklasifikasikan produk yang laku dengan yang tidak laku. Hal tersebut tentunya akan sulit karena banyaknya produk yang masuk dan keluar setiap harinya.

Analisa Sistem Baru

Dari kekurangan sistem yang sedang berjalan tersebut maka dibuatlah suatu aplikasi pengelompokan penjualan produk sederhana menggunakan excel. Aplikasi ini nantinya dapat mengklasifikasikan penjualan produk yang laku dan kurang laku. Sehingga pemesanan barang yang kurang laku dapat dikurangi.

Perhitungan jarak pusat cluster

Untuk mengukur jarak antara data dengan pusat cluster digunakan *eclidean distance*, dengan persamaan (2).

$$D(x, y) = \sqrt{(x_1 - y_1)^2 + (x_2 - y_2)^2} \quad (2)$$

Dengan,

- D = Jarak
- x = Data
- y = cenroid

Kemudian akan didapatkan matriks jarak yang ditunjukkan pada Tabel 3. Penentuan pusat awal cluster diambil data ke-1 sebagai pusat cluster ke-1 kemudian data ke-3 sebagai pusat cluster ke-2

dan data ke-10 sebagai pusat cluster ke-3. Kemudian dilakukan perhitungan sehingga centroid yang lama dan yang baru sama.

Tabel 3. Hasil pengujian SVM dengan optimasi SMO

Nama Barang	Volume terjual	Harga Rata	C1	C2	C3	Cluster
Beras	1.0000	0.0960	-	1.0171	1.3077	C1
Bawang Merah	-	0.5428	1.0953	0.2613	0.4605	C2
Bawang Putih	-	0.2815	1.0171	-	0.7206	C2
Gula Pasir	0.2982	0.1444	0.7034	0.3283	0.8895	C2
Garam	0.0038	0.0594	0.9969	0.2221	0.9420	C2
Minyak Goreng	0.1190	0.1313	0.8817	0.1917	0.8711	C2
Telur Ayam	0.2168	-	0.7891	0.3553	1.0130	C2
Daging Ayam	0.0013	0.3468	1.0298	0.0653	0.6554	C2
Nuget Ayam	0.0163	0.2031	0.9895	0.0801	0.7978	C2
Daging Kerbau	0.0551	1.0000	1.3077	0.7206	-	C3
Cabe Keriting	0.0175	0.2031	0.9883	0.0803	0.7978	C2

KESIMPULAN

Berdasarkan hasil uji coba dan analisis yang telah dilakukan dapat diambil kesimpulan sebagai berikut :

1. Algoritma K-means dapat diterapkan pada toko Tani Indonesia Center untuk menentukan penjualan barang mana yang sangat laris, laris dan kurang laris. Penerapan metode K-Means pada toko Tani Indonesia Center, yaitu dengan cara mengelompokan data stok barang.
2. Metode Data Mining menggunakan algoritma K-Means sangat membantu dan mempermudah toko Tani Indonesia Center untuk mengembangkan strategi persediaan stok barang.

DAFTAR PUSTAKA

- [1] M. I. Riyadh, "ANALISIS NILAI TUKAR PETANI KOMODITAS TANAMAN PANGAN DI SUMATERA UTARA".
- [2] C. Chairunnisa, I. Ernawati, and M. M. Santoni, "Klasifikasi Sentimen Ulasan Pengguna Aplikasi PeduliLindungi di Google Play Menggunakan Algoritma Support Vector Machine dengan Seleksi Fitur Chi-Square," *Inform. J. Ilmu Komput.*, vol. 18, no. 1, p. 69, Aug. 2022, doi: 10.52958/iftk.v17i4.4594.
- [3] A. -, F. Marisa, and D. Purnomo, "Penerapan Algoritma Apriori Terhadap Data Penjualan di Toko Gudang BM," *JOINTECS J. Inf. Technol. Comput. Sci.*, vol. 1, no. 1, Aug. 2016, doi: 10.31328/jointecs.v1i1.408.

- [4] Y. Irawan, "PENERAPAN DATA MINING UNTUK EVALUASI DATA PENJUALAN MENGGUNAKAN METODE CLUSTERING DAN ALGORITMA HIRARKI DIVISIVE," vol. 04.
- [5] Yazeed Al Moaiad, "TECHNOLOGIES USED IN DATA MINING," 2022, doi: 10.13140/RG.2.2.29881.44640.
- [6] W. J. (Thomas) Lee, I. Cheah, I. Phau, M. Teah, and B. A. Elenein, "Conceptualising consumer regiocentrism: Examining consumers' willingness to buy products from their own region," *J. Retail. Consum. Serv.*, vol. 32, pp. 78–85, Sep. 2016, doi: 10.1016/j.jretconser.2016.05.013.
- [7] P. Guruprasad, "OVERVIEW OF DIFFERENT CLUSTERING TECHNIQUES IN DATA MINING".
- [8] N. Bakhshinejad, R. Soltani, U. T. Nguyen, and P. Messina, "A Survey of Machine Learning Based Anti-Money Laundering Solutions".
- [9] M. Imron, U. Hasanah, and B. Humaidi, "Analysis of Data Mining Using K-Means Clustering Algorithm for Product Grouping," *IJIS Int. J. Inform. Inf. Syst.*, vol. 3, no. 1, pp. 12–22, Mar. 2020, doi: 10.47738/ijis.v3i1.3.
- [10] M. Kalra, N. Lal, and S. Qamar, "K-Mean Clustering Algorithm Approach for Data Mining of Heterogeneous Data," in *Information and Communication Technology for Sustainable Development*, vol. 10, D. K. Mishra, M. K. Nayak, and A. Joshi, Eds. Singapore: Springer Singapore, 2018, pp. 61–70. doi: 10.1007/978-981-10-3920-1_7.
- [11] R. K. Hapsari, M. Miswanto, R. Rulaningtyas, and H. Suprajitno, "Identification of Diabetes Mellitus and High Cholesterol Based on Iris Image".
- [12] A. Prawesti, T. Haryanto, and I. Effendi, "Sistem Pakar Identifikasi Varietas Ikan Mas (*Cyprinus carpio*) Berdasarkan Karakteristik Morfologi dan Tingkah Laku," *J. Ilmu Komput. Dan Agri-Inform.*, vol. 4, no. 1, p. 6, Jan. 2017, doi: 10.29244/jika.4.1.6-13.