

Clustering Data Penjualan Produk pada Toko Yudha dengan Algoritma K-Means

Arnold Ropen Sinaga, Gema Dodi Pranata

Fakultas Teknologi Informatika Universitas Informatika dan Bisnis Indonesia

Email: arnoldropen@unibi.ac.id, gema.dp@student.unibi.ac.id

Abstrak

Pesatnya perkembangan teknologi informasi yang menjadikan semua informasi dapat disimpan dalam jaringan komputer, membuat munculnya sistem basis data yang sangat besar. *Data Mining* salah satunya, merupakan teknologi yang sangat berguna untuk membantu perusahaan menemukan informasi yang sangat penting dari gudang data (*Data Warehouse*). *Clustering* merupakan salah satu metode dalam *Data Mining* yang bersifat tanpa arahan (*unsupervised*). Penggunaan algoritma *K-Means* adalah dapat membantu dalam mengelompokkan data, informasi yang ditampilkan berupa nilai centeroid dari tiap-tiap *cluster*, untuk mengelompokkan barang berdasarkan data penjualan produk pada Toko Yudha.

Kata Kunci: *Data Mining, Naive Bayes, Prestasi.*

Abstract

The rapid development of information technology which makes all information stored in a computer network has created a very large database system. Data Mining, one of them, is a very useful technology to help companies find very important information from data warehouses (Data Warehouse). Clustering is a method in Data Mining that is unsupervised. The use of the K-Means algorithm is able to assist in grouping data, the information displayed is in the form of a centeroid value from each cluster, to group goods based on product sales data at the Yudha Store.

Keywords: *Data Mining, K-Means, Rapidminer, Clustering.*

1. PENDAHULUAN

Pada saat sekarang ini proses penjualan mengalami perkembangan yang sangat pesat seiring perkembangan teknologi. Berbagai cara dan metode yang dilakukan untuk meningkatkan penjualan barang termasuk pada penjualan bahan-bahan pokok maupun kebutuhan sehari-hari. Dengan berkembangnya teknologi, para penjual harus turut ikut serta dalam proses perkembangan tersebut. Tetapi tidak menutup kemungkinan pada Toko Yudha maupun Toko kelontong sejenis yang mengalami kerugian di akibatkan banyaknya produk yang tidak laku terjual. Untuk itu diperlukan suatu analisa yang

digunakan menggunakan salah satu metode yang ada pada *Data Mining* yaitu *clustering*.

Data Mining atau *Knowledge Discovery in Database* (KDD) merupakan penggalian data yang tersembunyi dari database. *Data Mining* dapat melakukan pengklasifikasian, memprediksi, memperkirakan dan mendapatkan informasi lain yang bermanfaat dari kumpulan data dalam jumlah yang besar.

Salah satu metode *clustering* adalah Algoritma *K-Means*. Algoritma *K-Means* merupakan metode non-hierarki dimana pada langkah awalnya mengambil sebagian dari banyaknya komponen populasi untuk dijadikan pusat awal *cluster*. Proses *clustering* merupakan suatu proses pengelompokkan

berdasarkan atas prinsip kesamaan kelas serta mengurangi kesamaan antar kelas. Keakuratan perkiraan penjualan memiliki dampak yang besar pada penjualan. Hasil perkiraan penjualan menggunakan metode *clustering* dengan Algoritma *K-Means* yang tepat dan akurat dapat menjadi jembatan antara banyak penawaran dan permintaan sehingga mampu mengurangi biaya dan mempertahankan jumlah stok barang.

Menerapkan algoritma pengelompokan *K-Means* tradisional dan jarak pengukuran Euclidean dari kesamaan dipilih untuk digunakan dalam analisis nilai. Algoritma *K-Means* bekerja dengan baik untuk dokumen tertentu. Dengan menggunakan metode ini, data-data yang telah didapatkan dapat dikelompokkan kedalam beberapa *cluster* berdasarkan kemiripan dari data-data tersebut, sehingga data-data yang memiliki karakteristik yang sama dikelompokkan dalam satu *cluster* dan yang memiliki karakteristik yang berbeda dikelompokkan dalam *cluster* yang lain yang memiliki karakteristik yang sama.

Data Mining dengan teknik *clustering* bertujuan untuk meminimalisir kerugian yang diakibatkan karena banyaknya produk yang tidak laku terjual dengan menggunakan Algoritma *K-Means* untuk mengelompokkan produk kedalam beberapa *cluster* berdasarkan kemiripan data dan mendapatkan hasil yang lebih akurat terhadap produk penjualan baik yang sedang dijual maupun yang akan dijual.

Penerapan metode *K-Means* dalam pengelompokan data penjualan pada Toko Yudha dapat menghasilkan rekomendasi barang yang laris, Kurang laris dan cukup laris (Indriyani & Irfiani, 2019). Siregar, M. H. (2018) menyatakan bahwa Metode *K-Means* dapat memberi gambaran bahwa *cluster* 1 adalah *cluster* yang penjualan kurang laris pada toko adi bangunan. Sedangkan *cluster* 2 adalah *cluster* yang dikategorikan barang yang laris di pasaran pada kelompok ini hanya dua barang yang menunjukkan tingkat penjualan yang tinggi. Menurut Hadi, F., & Diana, Y (2020) menyatakan bahwa pada iterasi yang dilakukan mulai iterasi 1 sampai iterasi 5 didapatkan bahwa pada iterasi 1 mempunyai pola *cluster* centroid yang berbeda dengan pola *cluster*

iterasi 2. Sedangkan pada iterasi selanjutnya seperti iterasi 3 juga mempunyai pola *cluster* centroid yang berbeda dengan iterasi 1 dan 2.

Diharapkan artikel ini dapat memberikan manfaat bagi Toko Yudha maupun Toko kelontong yang lain untuk meningkatkan penjualan serta efisiensi, dan bagi penulis ataupun pembaca dapat menambah pengetahuan mengenai *Data Mining*, teknik *Data Mining* serta algoritma *K-Means*.

2. KAJIAN PUSTAKA

2.1 *Data Mining*

Data Mining merupakan proses penggalian informasi dan pola yang bermanfaat dari data yang sangat besar (Arhami dan Nasir, 2020). *Data Mining* mencakup pengumpulan data, ekstraksi data, analisis data, dan statistik data. *Data Mining* juga dikenal sebagai *Knowledge discovery*, *Knowledge extraction*, *data/pattern analysis*, *information harvesting*, dan lain-lain.

Data Mining atau *Knowledge Discovery in Database* (KDD) merupakan proses pengambilan informasi yang tersembunyi, di mana informasi tersebut sebelumnya tidak dikenal. Proses KDD meliputi proses pendekatan secara teknis seperti *clustering*, klasifikasi dengan meliputi metode yang merupakan irisan *Artificial Intelligence* (AI), *Machine Learning* (ML), dan statistik.

Secara garis besar kegunaan *Data Mining* dibagi menjadi 2 (dua) yaitu: deskriptif dan prediktif. Secara deskriptif artinya *Data Mining* berarti menemukan pola-pola yang digunakan untuk menjelaskan karakteristik data. Sedangkan secara prediktif *Data Mining* berarti dapat digunakan untuk menemukan model-model pengetahuan yang digunakan untuk melakukan prediksi.

Data Mining berdasarkan fungsionalitasnya dapat dikelompokkan menjadi 6 (enam) bagian yaitu (Suyanto, 2017):

1. Klasifikasi (*classification*)
2. Klastering (*clustering*)
3. Regresi (*Regresion*)
4. *Association Rule*
5. *Anomaly detection*

2.2 Clustering

Jika diberikan sekumpulan data titik, masing-masing diantaranya memiliki atribut, dan kesamaan ukuran, sehingga data ditemukan gugus (*cluster*) sehingga data titik dalam satu *cluster* (*intracluster*) lebih serupa kepada yang lain (Mufflikhah, Ratnawati, dan Putri, 2018). Namun data titik dalam *cluster* berbeda (*intercluster*) lebih nampak kurang serupa kepada yang lain.

2.3 K-Means Clustering

Menurut Suyanto (2017:262) “*k-Means* merupakan algoritma klasterisasi yang memiliki ide dasar sederhana dengan cara meminimalkan Sum of Squared Error (SSE) antara objek-objek data dengan sejumlah *k* centroid”.

Salah satu metode *Data Mining* yang sering digunakan yaitu algoritma *K-Means Clustering*, untuk mengidentifikasi kelompok alami dari sebuah kasus yang didasarkan pada pengelompokan data yang memiliki kemiripan sehingga dengan pengelompokan tersebut dapat dilakukan analisis (Vulandari, 2017). Data-data dipilih menjadi beberapa kelompok dengan kriteria yang telah ditentukan lalu dikumpulkan menjadi satu dalam sebuah *cluster*, di mana setiap *cluster* ini memiliki centroid atau titik pusat (Suntoro, 2019).

3. METODE PENELITIAN

3.1 Metode Penelitian

Metode penelitian yang dilakukan ialah menggunakan metode penelitian kualitatif karena dengan metode penelitian ini lebih cenderung berfokus kepada pembahasan mengapa suatu peristiwa atau fenomena terjadi dimana peristiwa yang dimaksudkan disini adalah objek penelitian. Adapun metode ini dapat dilakukan melalui serangkaian kegiatan berupa teknik pengumpulan data melalui observasi dan wawancara.

3.2 Sumber Data

Sumber data dalam penelitian ini diperoleh dari data primer, yang mana data

langsung dikumpulkan oleh orang yang berkepentingan atau yang memakai data tersebut serta data yang diperoleh melalui wawancara atau kuesioner.

3.3 Teknik Pengumpulan Data

Peneliti menggunakan teknik pengumpulan data sebagai berikut:

1. Observasi, Observasi merupakan teknik pengumpulan data dengan jalan pengamatan dan pencatatan secara sistematis, logis, objektif dan rasional mengenai fenomena.

Dalam penelitian ini, peneliti mengacu pada proses observasi participant (pengamatan berperan serta) yaitu dengan cara peneliti melibatkan secara langsung dan berinteraksi pada kegiatan yang dilakukan oleh subyek penelitian dalam lingkungan, selain itu juga mengumpulkan data secara sistematis dalam bentuk catatan lapangan. Teknik ini dilakukan dengan jalan mengadakan pengamatan sistematis terhadap objek, baru kemudian dilakukan pencatatan setelah penelitian itu.

2. Wawancara, Wawancara merupakan teknik pengumpulan data yang dilakukan melalui percakapan dan tanya jawab, baik secara langsung maupun tidak langsung dengan responden untuk mencapai tujuan tertentu. Adapun tujuan dari wawancara adalah untuk memperoleh informasi secara langsung, menyelami dunia fikiran dan perasaan seseorang, membuat suatu konstruksi kejadian dan pengalaman yang telah lalu dan memproyeksikan suatu kemungkinan yang diharapkan akan terjadi di masa yang akan datang.

3.4 Skema Penelitian

Skema penelitian yang digunakan:

- a. Mempelajari Studi Literatur
- b. Mengumpulkan Data
- c. Merumuskan Masalah
- d. Menganalisa dan Merancang Secara Umum
- e. Mentransformasikan Data

- f. Menguji Data Dengan Algoritma *K-Means*
- g. Menarik Kesimpulan.

4. HASIL DAN PEMBAHASAN

4.1 Pengumpulan Data

Data barang sebanyak 15 jenis barang berdasarkan dari kebutuhan yang kemudian diolah dengan menggunakan Algoritma *K-Means* untuk menentukan mana yang lebih laris maupun tidak laris menggunakan Rapid Miner.

4.2 Perhitungan Algoritma *K-Means* dengan Rapid Miner

Langkah pertama dilakukan dengan menentukan menentukan banyaknya *cluster* dan nilai centroid awal, jumlah *cluster* ditentukan berdasarkan variabel pengelompokan data penjualan laris, cukup laris dan kurang laris. langkah berikutnya menentukan centroid awal (k), dapat diambil data secara random atau acak

Tabel 1. Data barang terjual

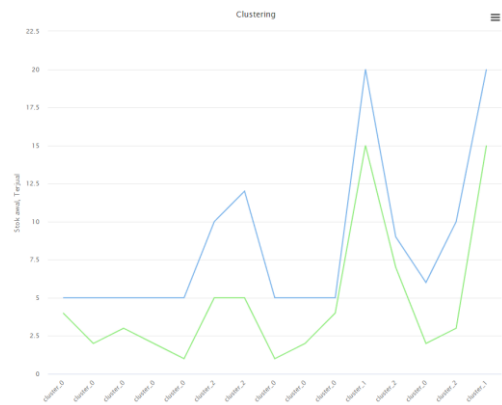
Type	Kode Barang	Stok Awal	Terjual
Kebutuhan 1	B101	5	4
	B102	5	2
	B103	5	3
	B104	5	2
	B105	5	1
Kebutuhan 2	B201	10	5
	B202	12	5
	B203	5	1
	B204	5	2
	B205	5	4
Kebutuhan 3	B301	20	15
	B302	9	7
	B303	6	2
	B304	10	3
	B305	20	15

Tabel 2. Centroid

Attribute	cluster_0	cluster_1	cluster_2
Stok awal	5,111	20	10,250
Terjual	2,333	15	5

Row No.	Kode Barang	cluster	Stok awal	Terjual
1	B101	cluster_0	5	4
2	B102	cluster_0	5	2
3	B103	cluster_0	5	3
4	B104	cluster_0	5	2
5	B105	cluster_0	5	1
6	B201	cluster_2	10	5
7	B202	cluster_2	12	5
8	B203	cluster_0	5	1
9	B204	cluster_0	5	2
10	B205	cluster_0	5	4
11	B301	cluster_1	20	15
12	B302	cluster_2	9	7
13	B303	cluster_0	6	2
14	B304	cluster_2	10	3
15	B305	cluster_1	20	15

Gambar 1. Tampilan Hasil Clustering



Gambar 2. Hasil Pengolahan Data pada Rapidminer

5. KESIMPULAN

Untuk melakukan pengelompokan terhadap data penjualan pada Toko Yudha dapat dilakukan dengan menerapkan *K-Means Clustering*. Data-data yang diperoleh, diproses dengan menggunakan software Microsoft Excel ataupun dengan software lainnya seperti SPSS ataupun Rapidminer. Pengelompokan akan sangat dibutuhkan untuk melihat barang-barang yang sering beli konsumen berdasarkan jumlah terjual.

DAFTAR PUSTAKA

- Hadi, F., & Diana, Y. (2020). Pengklusteran Penjualan Bahan Bangunan Menggunakan Algoritma K-Means. *JOISIE (Journal Inf. Syst. Informatics Eng.)*, 4(1), 22.
- Indriyani, F., & Irfiani, E. (2019). Clustering Data Penjualan pada Toko Perlengkapan Outdoor Menggunakan Metode K-Means. *JUITA: Jurnal Informatika*, 7(2), 109-113.
- Mardalius, M. (2018). Pemanfaatan Rapid Miner Studio 8.2 Untuk Pengelompokan Data Penjualan Aksesoris Menggunakan Algoritma K-Means. *JURTEKSI (Jurnal Teknologi dan Sistem Informasi)*, 4(2), 123-132.
- Mudzakir, B. D. (2018). Pengelompokan Data Penjualan Produk pada PT Advanta Seeds Indonesia Menggunakan Metode K-Means. *JATI (Jurnal Mahasiswa Teknik Informatika)*, 2(2), 34-40.
- Normah, N., Nurajizah, S., & Salbinda, A. (2021). Penerapan data mining metode k-means clustering untuk analisa penjualan pada toko fashion hijab Banten. *Jurnal Khatulistiwa Informatika*, 7(2), 158-163.
- Pambudi, W. T., & Witanti, A. (2022). Penerapan Algoritma K-Means Untuk Menganalisis Data Penjualan Pada Toko Ayu Collection Berbasis Web. *Jurnal Sistem Informasi dan Bisnis Cerdas*, 15(1), 1-9.
- Aulia, S. (2020). Klasterisasi Pola Penjualan Pestisida Menggunakan Metode K-Means Clustering (Studi Kasus Di Toko Juanda Tani Kecamatan Hutabayu Raja). *Djtechno: Jurnal Teknologi Informasi*, 1(1), 1-5.
- Silalahi, M. (2018). Analisis Clustering Menggunakan Algoritma K-Means Terhadap Penjualan Produk Padapt Batamas Niaga Jaya. *Computer Based Information System Journal*, 6(2), 20-35.
- Siregar, M. H. (2018). Data Mining Klasterisasi Penjualan Alat-Alat Bangunan Menggunakan Metode K-Means (Studi Kasus Di Toko Adi Bangunan). *Jurnal Teknologi Dan Open Source*, 1(2), 83-91.
- Tamba, S. P., & Kesuma, F. T. (2019). Penerapan Data Mining Untuk Menentukan Penjualan Sparepart Toyota Dengan

Metode K-Means Clustering: data mining; k-means-clustering. *Jurnal Sistem Informasi dan Ilmu Komputer Prima (JUSIKOM PRIMA)*, 2(2), 67-72.