
Penerapan Data Mining Untuk Evaluasi Data Penjualan Menggunakan Metode Clustering Dan Algoritma Hirarki Divisive Studi Kasus Toko Sembako Puj

Ade Eka Febriyanti¹, Syaiful Zuhri Harahap², Masrizal³

Sistem Informasi, Fakultas Sains dan Teknologi, Universitas Labuhanbatu^{1,2,3}

Email: seka98193@gmail.com¹, syaifulzuhriharahap@gmail.com²,
masrizal120405@gmail.com³

Corresponding Author : seka98193@gmail.com

Abstract

The larger a company, the longer the company stands, the more companies have branches, of course, the greater the data owned. These data can be consumer data, purchase data, sales data, payroll data, and many other data. All data will usually be stored in a database. But many companies, even the Information Technology (IT) division, do not realize how valuable the pile of old data generated by the company in transactions and activities. Data mining is the study of methods for generating knowledge or finding patterns for processing data. So it's not just information, it's knowledge. Data Mining has several methods including clustering. Clustering is a well-known and widely used method in data mining. The main purpose of this clustering method is to Group a number of data/objects into clusters (groups) so that the cluster will contain the same data as each group. In this study, Divisive hierarchy algorithm is used to form clusters. From the pattern obtained is expected to provide knowledge for the company Media World Pekanbaru as a supporting tool to take policy.

Keywords: Divisive Hierarchy Algorithm, Clustering, Data Mining.

I. Pendahuluan

Teknologi internet berkembang menjadi alat bantu yang mampu menyediakan informasi dan juga mampu mengolah informasi. Internet menjadi media yang diandalkan untuk penyebaran informasi dari semua aktifitas di segala bidang. Saat ini sistem penjualan secara tradisional atau manual mulai ditinggalkan dan tergantikan dengan sistem penjualan yang berbasis internet. Pemanfaatan data yang ada di dalam sistem informasi untuk menunjang kegiatan pengambilan keputusan, tidak cukup hanya mengandalkan data operasional saja, diperlukan suatu

analisis data untuk menggali potensi-potensi informasi yang ada. Para pengambil keputusan berusaha untuk memanfaatkan gudang data yang sudah dimiliki untuk menggali informasi yang berguna membantu mengambil keputusan, hal ini mendorong munculnya cabang ilmu baru untuk mengatasi masalah penggalian informasi atau pola yang penting atau menarik dari data dalam jumlah besar, yang disebut dengan data mining. Penggunaan teknik data mining diharapkan dapat memberikan pengetahuan-pengetahuan yang sebelumnya tersembunyi didalam gudang data sehingga menjadi informasi yang

berharga.

Data merupakan sekumpulan fakta mentah berupa simbol, angka atau dan kata-kata. Data terbagi menjadi dua tipe, yaitu data primer dan data sekunder. Data primer adalah data yang belum pernah dikumpulkan sebelumnya dan di kumpulkan untuk kepentingan penelitian saja. Data primer mengacu pada data yang berasal dari peneliti untuk pertama kali. Data sekunder adalah data yang sudah ada sebelumnya, dikumpulkan oleh lembaga dan organisasi penelitian sebelumnya[1]. Pentingnya data sebagai komponen utama dalam sistem informasi menjadi hal yang perlu diperhatikan dan dikaji karena data merupakan dasar dalam menyediakan informasi dan menentukan kualitas informasi dengan cepat, akurat, dan relevan, sehingga informasi yang disajikan berupa informasi yang absah.

Toko sembako pujo merupakan toko supplier yang bergerak dibidang kebutuhan pokok masyarakat. Pelanggan merupakan target utama toko dalam meningkatkan penjualan bahan pokok untuk mencari keuntungan perusahaan. Untuk mengolah data penjualan bahan pokok selama ini di toko sembako menggunakan aplikasi Microsoft Excell. Dengan menggunakan sistem ini toko bisa mengolah data penjualan mulai dari barang-masuk dan keluar. Data yang diinput tersimpan di dalam database. Oleh karena itu, data yang ada di dalam database berjumlah ribuan data dan toko memandangnya hanya sebagai arsip atau bahkan yang sudah menjadi berkas-berkas tidak terpakai dan bisa dihancurkan kapan saja. Hal tersebut tentu saja merupakan pandangan yang salah, sebab dengan penanganan data yang tepat, data-data tersebut

dapat diberdayakan dan nantinya bisa digunakan untuk memprediksi strategi bisnis yang dijalankan di masa yang akan datang dan berguna untuk took tersebut.

Dalam penelitian yang berjudul “Klasterisasi Pola Penjualan Pestisida Menggunakan Metode K-Means Clustering” tahun 2020 oleh Sabrina Aulia Rahmah yang membahas tentang cara mengelompokkan peringkat pestisida dengan penjualan tertinggi. Hasilnya adalah dengan adanya metode K-means Clustering membantu dalam pengelompokan pola penjualan selama satu musim. Persamaan dari penelitian ini adalah memakai metode K-Means Clustering dan klasifikasi hasilnya. Untuk perbedaannya yaitu pada proses analisa yang menggunakan *software* RapidMiner sebagai implementasinya.

Dalam Penelitian yang berjudul “Klasterisasi Penjualan Alat-Alat Bangunan Menggunakan Metode K-Means (Studi Kasus di Toko Adi Bangun)” tahun 2018 oleh M. Hasyim Siregar yang membahas tentang penerapan data mining untuk mengolah data penjualan khususnya penjualan bahan bangunan. Hasilnya adalah grafik untuk mengklasifikasi barang bangunan yang laris yaitu dan tidak laris. Persamaan dari penelitian ini adalah menggunakan klasifikasi stok yang mirip seperti laris dan tidak laris. Untuk perbedaannya yaitu pada bantuan *software* rapid miner.

II. Landasan Teori

Clustering

Clustering (pengelompokan data) mempertimbangkan sebuah pendekatan penting untuk mencari kesamaan dalam data dan menempatkan data yang sama kedalam kelompok-kelompok. Clustering membagi kumpulan data kedalam bebe-

rapa kelompok dimana kesamaan dalam sebuah kelompok adalah lebih besar dari pada diantara kelompok-kelompok.

Hirarki Divisive

Langkah awal yang dilakukan dalam algoritma hirarki divisive adalah membentuk satu klaster yang besar yang dapat ditempati oleh semua obyek data. Pada Langkah berikutnya, satu klaster besar tersebut dipisah-pisah menjadi beberapa klaster yang lebih kecil dengan karakteristik data yang mempunyai lebih besar kesamaan satu dengan yang lainnya, sehingga data yang tidak memiliki kemiripan yang cukup besar berada pada klaster yang terpisah.

III. Metode Penelitian

Metode penelitian ini menggunakan Algoritma Clustering. Tujuan dari data yang akan di cluster adalah meminimalisasi objective function yang diset dalam clustering. Pada umum meminimalisasi variasi di dalam suatu cluster dan memaksimalkan variasi antar cluster. Tiap cluster dihubungkan dengan sebuah centroid. Tiap centroid ditempatkan ke dalam cluster dengan centroid terdekat. Jumlah cluster harus ditentukan.

IV. Hasil Dan Pembahasan Data

Dalam proses *Knowledge Discovery in Database* (KDD) seleksi data merupakan proses himpunan data dan menciptakan data target pada sampel data, dimana penemuan akan dilakukan selanjutnya hasil seleksi disimpan dalam suatu berkas yang terpisah dari data basis data operasional. Setelah data didapatkan dari pihak Secom Infotech seperti data awal diatas maka dilakukan proses seleksi data, proses ini dilakukan agar memudahkan proses perhitungan algoritma K-

Means Clustering . Adapun data yang digunakan adalah data pada bulan Januari-Juni 2018 yang dapat dilihat pada tabel 1.

Tabel 1. Daftar Seleksi Data Penjualan

Jumlah Barang	Jumlah Terjual
20	15
12	8
14	5
15	10
6	2
10	3
10	7
20	13
13	6
12	8
5	4
4	3
11	6
19	13
16	14
17	12
20	14
24	20
25	19
10	8
15	12
14	7
16	12
17	11
18	16
20	13
10	8
5	3
8	2
12	7

Pada tahap ini dilakukan proses utama yaitu pengelompokan data penjualan barang yang diakses dari database, yaitu metode clustering algoritma K-Means. Dari banyak data penjualan yang diperoleh, percobaan dilakukan dengan menggunakan parameter-parameter berikut :

- Jumlah cluster : 2
- Jumlah data : 30
- Jumlah atribut : 2

Menentukan Titik Pusat Centroid

Secara Random Setelah menentukan cluster sebanyak 2 cluster, selanjutnya ditentukan titik pusat secara random.

C_0 diambil dari nomor 1, C_1 diambil dari nomor 7 dapat dilihat pada tabel 2.

Tabel 2. Nilai Centroid Random

Centroid	Jumlah Barang	Jumlah Terjual
C_0	20	15
C_1	10	7

Pengujian

Data yang dikumpulkan terdiri dari 204 data penjualan setiap pelanggan dan dimulai dilakukan pengujian secara manual dengan menggunakan algoritma hirarki divisive yang terdiri dari tiga langkah:

Langkah 1

Perhitungan pada langkah pertama yaitu perhitungan nilai rata-rata (*mean*) dari semua nilai yang ada pada tabel frekuensi.

Perhitungan titik pusat 1

$$= (\text{nilai data terkecil} + \text{nilai rata-rata}) / 2$$

$$= (1 + 34,67647059) / 2$$

$$= 17,83823529$$

Perhitungan titik pusat 2

$$= \text{diambil dari nilai rata-rata}$$

$$= 34,67647059$$

Perhitungan titik pusat 3

$$= (\text{nilai rata-rata} + \text{nilai data terbesar}) / 2$$

$$= (34,67647059 + 358) / 2$$

$$= 196,8382353$$

Maka hasil titik pusat dari langkah pertama adalah:

- a. Titik Pusat1 = 17,83823529
- b. Titik Pusat2 = 34,67647059
- c. Titik Pusat3 = 196,8382353

Perhitungan rentang tertinggi

$$= (\text{Titik Pusat 3} - \text{Titik Pusat 2}) / 2$$

$$= (196,8382353 - 34,67647059) / 2$$

$$= 81,08088235$$

Batas rentang tertinggi dari titik pusat 2

$$= (\text{Titik Pusat 3} - \text{jml rentang tertinggi})$$

$$= 196,8382353 - 81,08088235$$

$$= 115,7573529$$

Sehingga anggota klaster dari titik pusat 2 diambil dari rentang titik frekuensi data $26,25735294 \leq \text{Anggota klaster titik pusat 2} < 115,7573529$

Langkah 2

Perhitungan langkah kedua berdasarkan data dari langkah pertama yaitu : Titik Pusat 1

Jumlah semua item data pada klaster 1 = 135 item.

Kemudian menentukan nilai median pada rentang data di atas untuk menentukan titik pusatnya.

$$(N + 1) / 2 = 135 + 1 / 2 = 68$$

Jumlah semua item data pada klaster 1 = 51 item

Kemudian menentukan nilai median pada rentang data di atas untuk menentukan titik pusatnya.

$$(N + 1) / 2 = 51 + 1 / 2 = 26$$

Nilai yang terletak pada item 26 adalah nilai tengah dari nilai 44

Jadi nilai titik pusat 2 pada langkah kedua adalah 44

Hasil perhitungan dari titik pusat 2 pada langkah kedua kemudian dibandingkan dengan nilai titik pusat 2 pada langkah pertama, yaitu $44 \neq 34,67647059$, dari keterangan ini dapat disimpulkan bahwa titik pusat masih berubah.

Titik Pusat 3

Anggota klaster titik pusat ketiga pada langkah pertama adalah :

$$[117,121,123,141,143,150,183,194,200,201,205,206,213,242,252,287,317,358].$$

Jumlah semua item pada klaster 3 = 18 item

Median pada rentang data klaster 2 $(N + 1) / 2 = 18 + 1 / 2 = 9,5$

Nilai yang terletak pada item 9,5 adalah nilai tengah dari nilai 200 dan 201 = $(200 + 201) / 2 = 200,5$

Perhitungan jarak rentang tersebut adalah :

Titik Pusat 1

Rentang titik pusat 1

$$\begin{aligned} &= (\text{titik pusat 2} - \text{titik pusat 1}) / 2 \\ &= (44 - 3) / 2 \\ &= 20,5 \end{aligned}$$

Maka batas rentang anggota dari titik pusat 1 pada langkah 2 adalah:

$$\begin{aligned} &= (\text{titik pusat 1} + \text{jml rentang titik pusat 1}) \\ &= 3 + 20,5 \\ &= 23,5 \end{aligned}$$

Anggota klaster 1 pada langkah 2 < data frekuensi bernilai 23,5

Titik pusat 2

Rentang titik pusat 2

$$\begin{aligned} &= (\text{titik pusat 2} - \text{titik pusat 1}) / 2 \\ &= (44 - 3) / 2 \\ &= 20,5 \end{aligned}$$

Perhitungan rentang terendah

$$\begin{aligned} &= (\text{titik pusat 2} - \text{titik pusat 1}) / 2 \\ &= (44 - 3) / 2 \\ &= 20,5 \end{aligned}$$

Maka batas rentang terendah titik pusat 2 adalah

$$\begin{aligned} &= (\text{titik pusat 1} + \text{jml rentang titik pusat 2}) \\ &= 3 + 20,5 \\ &= 23,5 \end{aligned}$$

Batas rentang tertinggi dari titik pusat 2 adalah

$$\begin{aligned} &= (\text{titik pusat 3} - \text{jml rentang tertinggi}) \\ &= 200,5 - 78,25 \\ &= 122,25 \end{aligned}$$

Titik Pusat 3

Rentang titik pusat 3

$$\begin{aligned} &= (\text{titik pusat 3} - \text{titik pusat 2}) / 2 \\ &= (200,5 - 44) / 2 \\ &= 78,25 \end{aligned}$$

Maka batas rentang anggota dari titik pusat 3 pada langkah 2 adalah :

$$\begin{aligned} &= (\text{titik pusat 2} + \text{jml rentang titik pusat 3}) \\ &= 44 + 78,25 \end{aligned}$$

$$= 122,25$$

Anggota klaster 3 pada langkah 2 < data frekuensi bernilai 122,25.

Langkah 3

Langkah ketiga merupakan langkah yang digunakan jika hasil titik pusat pada perhitungan kedua masih berubah dari perhitungan pertama. Fungsi perhitungan yang digunakan sama dengan fungsi pada perhitungan kedua, yaitu dengan cara mencari nilai tengah (median).

Titik Pusat 1

Jumlah semua item data pada klaster 1 = 135 item

Kemudian menentukan nilai median pada rentang data di atas untuk menentukan titik pusatnya.

$$(N + 1) / 2 = 135 + 1 / 2 = 68$$

Titik Pusat 2

Anggota klaster titik pusat kedua pada langkah pertama adalah:

Jumlah semua item data pada klaster 1 = 51 item.

Kemudian menentukan nilai median pada rentang data di atas untuk menentukan titik pusatnya.

$$(N + 1) / 2 = 51 + 1 / 2 = 26$$

Jadi nilai titik pusat 2 pada langkah ketiga adalah 44

Hasil perhitungan dari titik pusat pada langkah ketiga kemudian dibandingkan dengan nilai titik pusat pada langkah kedua, yaitu 44 = 44, dari keterangan ini dapat disimpulkan bahwa titik pusat tidak berubah.

Titik Pusat 3

Anggota klaster titik pusat ketiga pada langkah pertama adalah:

[117,121,123,141,143,150,183,194,200,201,205,206,213,242,252,287,317,358].

Jumlah semua item pada klaster 3 = 18 item.

Median pada rentang data klaster 2 $(N + 1) / 2 = 18 + 1 / 2 = 9,5$

Nilai yang terletak pada item 9,5 adalah nilai tengah dari nilai 200 dan $201 = (200 + 201) / 2 = 200,5$.

Sehingga diperoleh titik pusat 3 pada langkah ketiga adalah 200,5

Hasil perhitungan dari titik pusat pada langkah ketiga kemudian dibandingkan dengan nilai titik pusat pada langkah kedua, yaitu $200,5 = 200,5$, dari keterangan ini dapat disimpulkan bahwa titik pusat tidak berubah.

Hasil

Dari analisa dan pengujian yang telah dilakukan, hasil dari pengujian algoritma hirarki divisive

- a. Cluster 1 terdiri dari 135 item yaitu dengan range pembelian 1-25. Customer yang berada di dalam range 1-25.
- b. Cluster 2 terdiri dari 51 item yaitu dengan range pembelian 27-113. Customer yang berada di dalam range 27-113
- c. Cluster 3 terdiri dari 18 item yaitu dengan range pembelian 117-358. Customer yang berada di dalam range 117-358.

Setelah dilakukan pengujian secara manual menggunakan algoritma hirarki divisive dapat diketahui bahwa customer yang loyal berada di cluster 3 dengan range frekuensi pembelian 117-358.

V. Kesimpulan

Setelah dilakukan analisa dan pengujian data maka dapat ditarik kesimpulan bahwa data mining bermanfaat untuk menghasilkan pengetahuan berupa customer loyal yang ada di Toko sembako pujo dan setelah dievaluasi dengan menggunakan algoritma hirarki divisive serta pengolahan data dengan menggunakan

software rapid miner maka ditemukan bahwa customer loyal berada pada cluster 3 dengan range 117-358.

VI. Daftar Pustaka

- A. K. Patnaik, P. K. Bhuyan, and K. V. Krishna Rao, "Divisive Analysis (DIANA) of hierarchical clustering and GPS data for level of service criteria of urban streets," *Alexandria Eng. J.*, vol. 55, no. 1, pp. 407–418, 2016, doi: 10.1016/j.aej.2015.11.003.
- A. R. Jannah, D. Arifianto, and M. Kom, "Penerapan Metode Clustering dengan Algoritma K-Means untuk Prediksi Kelulusan Mahasiswa Jurusan Teknik Informatika di Universitas Muhammadiyah Jember," *J. Manaj. Sist. Inf. dan Teknol.*, vol. 1, no. 1210651237, pp. 1–10, 2015.
- D. Juliawan, F. Amir, and E. Desi, "Penerapan Data Mining Metode Clustering Pada CV. Secom Infotech Menggunakan Algoritma K-Means," *Pros. Semin. Nas. Ris. Inf. Sci.*, vol. 1, no. September, p. 96, 2019, doi: 10.30645/senaris.v1i10.12.
- D. Missa, S. Achmadi, and A. Mahmudi, "Penerapan Metode Clustering Dengan Algoritma K-Means Pada Pengelompokan Data Penghasilan Orang Tua Siswa," *JATI (Jurnal Mhs. Tek. Inform.)*, vol. 5, no. 1, pp. 125–133, 2021, doi: 10.36040/jati.v5i1.3275.
- Dodi Nofri Yoliad, "Data mining Dalam Analisis Tingkat Penjualan Barang Elektronik Menggunakan Algoritma K-means," *Insearch (Information Syst. Res. J.)*, vol. 3, no. 1, 2023.

- E. D. Purnamasari et al., "Critical Success Factors In Agile Software Project : Case Study Astra Graphia Information Technology," vol. 3, pp. 4997–5006, 2023.
- E. K. A. Rizkiyanti, "Penerapan Metode K-Means Clustering Pada Sistem Pendukung Keputusan Penentuan Kualitas Beras Berdasarkan Ciri Fisik," Skripsi. Tidak diterbitkan. Fak. Sains dan Teknol. Univ. Islam Negeri Sultan Syarif Kasim Riau Pekanbaru, 2021.
- I. A. Kasoqi, M. N. Hayati, and R. Goejantoro, "Pengelompokan Desa Atau Kelurahan Di Kutai Kartanegara Menggunakan Algoritma Divisive Analysis," J. Stat. Univ. Muhammadiyah Semarang, vol. 9, no. 2, p. 101, 2021, doi: 10.26714/jsunimus.9.2.2021.101-108.
- I. W. Rahayu, I. Atastina, and A. Herdiani, "Analisis Dan Implementasi Algoritma Agglomerative Hierarchical Untuk Deteksi Komunitas Pada Media Sosial Facebook," e-Proceeding Eng., vol. 5, no. 1, pp. 1460–1468, 2018.
- N. Agustin, "Hierarchical Agglomerative Clustering Pada Klasifikasi Produk Susu Formula Bayi Di Bawah Umur 1 Tahun Di Hypermart Kota Malang," 2022.
- O. Somantri, S. Wiyono, and D. Dairoh, "Metode K-Means untuk Optimalisasi Klasifikasi Tema Tugas Akhir Mahasiswa Menggunakan Support Vector Machine (SVM)," Sci. J. Informatics, vol. 3, no. 1, pp. 34–45, 2016, doi: 10.15294/sji.v3i1.5845.
- R. Kurniawan and R. Dewi, "Penerapan Algoritma K-Means Clustering Dalam Persentase Merokok Pada Penduduk Umur Di Atas 15 Tahun Menurut Provinsi," J. Sist. Komput. dan Inform. Hal, vol. 2, no. 2, pp. 178–186, 2021, doi: 10.30865/json.v2i2.2770.
- R. Takdirillah, "Penerapan Data Mining Menggunakan Algoritma Apriori Terhadap Data Transaksi Sebagai Pendukung Informasi Strategi Penjualan," Edumatic J. Pendidik. Inform., vol. 4, no. 1, pp. 37–46, 2020, doi: 10.29408/edumatic.v4i1.2081.
- Ramadhana, Islamiyah, and A. P. A. Masha, "Penerapan Data Mining Menggunakan Metode K-Means Clustering Pada Data Ekspor Batubara," Adopsi Teknol. dan Sist. Inf., vol. 2, no. 1, pp. 35–42, 2023, doi: 10.30872/atasi.v2i1.595.
- S. W. Hadi, M. F. Julianto, S. Rahmatullah, and W. Gata, "Analisa Cluster Aplikasi Pada App Store Dengan Menggunakan Metode K-Means," Bianglala Inform., vol. 8, no. 2, pp. 86–90, 2020, doi: 10.31294/bi.v8i2.8191.
- T. Tonny, I. R. Munthe, and M. H. Munnandar, "Perancangan Aplikasi Pengenalan Tokoh Penemu Benda-Benda Penting Di Dunia Berbasis Android Menggunakan Metode Sequential Search," MEANS (Media Inf. Anal. dan Sist., vol. 6, no. 1, pp. 90–94, 2021, doi: 10.54367/means.v6i1.1254.
- W. Apriani, "Jurnal Mantik Jurnal Mantik," Sist. Pendukung Keputusan Pemilihan Pimpinan Dengan Metod. Multi Attrib. Util. Theory di PT. Sagami Indones., vol. 3, no. 2, pp. 10–19, 2019.
- W. Lestari, "Clustering Data Mahasiswa Menggunakan Algoritma K-

- Means Untuk Menunjang Strategi Promosi (Studi Kasus : STMIK Bina Bangsa Kendari),” *Simkom*, vol. 4, no. 2, pp. 35–48, 2019, doi: 10.51717/simkom.v4i2.37.
- W. W. Kristianto and C. Rudianto, “Penerapan Data Mining Pada Penjualan Produk Menggunakan Metode K-Means Clustering (Studi Kasus Toko Sepatu Kakikaki),” *J. Pendidik. Teknol. Inf.*, no. 5, pp. 90–98, 2020.
- W. Widyawati, W. L. Y. Saptomo, and Y. R. W. Utami, “Penerapan Agglomerative Hierarchical Clustering Untuk Segmentasi Pelanggan,” *J. Ilm. SINUS*, vol. 18, no. 1, p. 75, 2020, doi: 10.30646/sinus.v18i1.448.
- Z. Muttaqin, “Implementasi Unsupervised Learning Pada Nilai Jasmani Kesamaptaan Sekolah Polisi Negara Dengan Metode Clustering Analysis,” *PROSISKO J. Pengemb. Ris. dan Obs. Sist. Komput.*, vol. 10, no. 1, pp. 18–23, 2023, doi: 10.30656/prosisko.v10i1.6269.