

Pengolahan Data Pelanggan Dengan Teknik Clustering K-Means Di Aplikasi Weka

Alfin Falakhi

Universitas Mercu Buana

*Corresponding author email:
alfinply@gmail.com

Manuscript history:

Diterima
Direvisi
Diterima untuk terbit

ABSTRAK

Setiap pelanggan memiliki intensitas yang berbeda-beda dalam berbelanja. Beberapa ada yang berbelanja dengan jumlah banyak dan juga ada juga yang sedikit baik dari pelanggan yang berpendapatan tinggi maupun rendah. Kita perlu melakukan analisis data agar bisa mengetahui kebiasaan dari para pelanggan dalam berbelanja. Oleh karena itu penulis melakukan analisis data mining menggunakan data pelanggan. Dalam melakukan analisis penulis menggunakan alat bantu tools WEKA. Metode yang digunakan adalah metode k-means clustering dengan 5 cluster. Dengan Jumlah Sebagai berikut, C1 dengan 44 data, C2 dengan 43 data, C3 dengan 34 data, C4 dengan 35 data, C5 dengan 44 data.

Kata Kunci: Pelanggan, Weka, Kmeans, Clustering, Data Mining

I. Pendahuluan

Karakter Pelanggan dalam berbelanja di sebuah toko atau pusat perbelanjaan tentu akan berbeda-beda. Hal ini dipengaruhi oleh faktor ekonomi. (Studi Pendidikan Ekonomi Koperasi & Muhson, n.d.) Dimana orang berpenghasilan rendah akan membelanjakan Sebagian besar atau keseluruhan uangnya di pasar tradisional dibanding di mall. dari fakta tersebut pihak toko harus mengetahui karakteristik pelanggannya (Budi, 2015) hal tersebut sangatlah penting karena banyak toko yang berujung bangkrut karena penyediaan barang yang dijual tidak sesuai dengan kebutuhan Sebagian besar pelanggan toko (Jusia et al., 2019).

Berdasarkan hasil wawancara dengan salah satu pemilik toko, pak Ko Lim mengatakan, terkadang barang yang dijual tidak tepat sasaran karena kurangnya pemahaman data pelanggan, dan kebiasaan pelanggan yang tidak terprediksi. Maka penjualan menjadi tidak maksimal (Mardi, 2017). Harapannya pihak toko bisa melihat persentase pelanggan untuk penentuan karakter pelanggan, jika berhasil dan berdampak terhadap pengambilan keputusan toko dalam penyediaan barang, maka pengolahan data pelanggan dengan data miner akan direkomendasikan ke toko lain (Susanto & Sudiyatno, 2014).

Data mining merupakan proses dengan teknik statistik, kecerdasan buatan, matematika, dan machine learning untuk Ekstrak dan akses informasi yang berguna dan pengetahuan yang relevan dari database besar (Zulfa et al., 2020). Tujuan utama dari data mining adalah untuk mencari, mengeksplorasi atau mengekstraksi pengetahuan dari data atau informasi yang tersedia bagi kita. Clustering adalah proses yang mengidentifikasi kesamaan sifat dalam suatu kelompok untuk menghasilkan informasi yang berguna. Metode pengelompokan data telah banyak digunakan di berbagai bidang, misalnya dalam pemodelan, pengambilan keputusan dalam penambangan data, pengenalan pola, dan dalam bidang kesehatan. Ada banyak algoritma untuk melakukan proses integrasi pada kumpulan data yang besar. Pada penelitian ini, peneliti akan menggunakan algoritma K-Means untuk memilih jumlah cluster yang optimal. KMeans adalah algoritma yang banyak digunakan karena efisien dan efektif. Hal ini dikarenakan K-means mudah dipelajari dan waktu komputasinya relatif singkat (Setyo & Wardhana, 2019).

II. Landasan Teori

A. Clustering

Clustering adalah suatu pemrosesan data dimana data dikelompokkan menjadi beberapa kelompok sehingga objek dalam satu kelompok memiliki banyak kesamaan dan memiliki banyak perbedaan dengan objek dalam kelompok lain. Perbedaan dan persamaan seringkali didasarkan pada nilai properti objek dan dapat juga berupa perkiraan jarak. Clustering sendiri juga disebut klasifikasi tanpa pengawasan, karena clustering sebagian besar dipelajari dengan minat. Analisis cluster adalah proses membagi satu set objek data menjadi subset. Setiap subgrup merupakan satu grup, sehingga item-item dalam grup tersebut serupa, dan memiliki perbedaan dengan item-item dari grup lainnya. Klasifikasi tidak dilakukan oleh algoritma pengelompokan manual. Oleh karena itu, pengelompokan sangat berguna dan dapat mengidentifikasi kelompok yang tidak diketahui dalam data. Analisis cluster banyak digunakan dalam berbagai aplikasi seperti Business Intelligence, Pengenalan Pola Gambar,

Pencarian Web, Biologi, dan Keamanan. Dalam intelijen bisnis, pengelompokan dapat mengatur banyak pelanggan menjadi banyak kelompok. Misalnya, mengelompokkan pelanggan ke dalam beberapa kelompok dengan karakteristik kuat yang serupa. Clustering juga dikenal sebagai partisi data, karena clustering mengelompokkan banyak kumpulan data berdasarkan kesamaan. Clustering juga dapat digunakan sebagai deteksi outlier, dimana outlier mungkin lebih menarik daripada kasus normal. Sistem tersebut adalah Outlier Detection, mendeteksi penipuan kartu dan melacak aktivitas kriminal di e-commerce. Contohnya adalah pengecualian dalam transaksi kartu kredit.

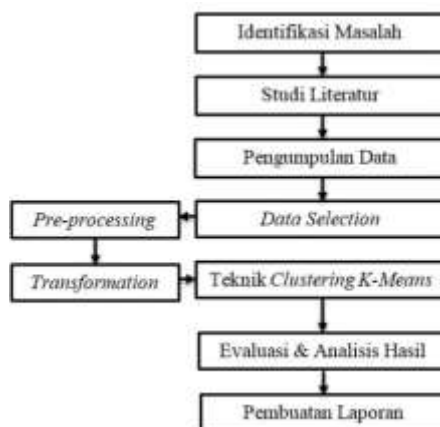
B. K-means

Teknik Pengelompokan K-Means adalah teknik pengelompokan sederhana tanpa pengawasan. Misalkan D adalah dataset dari n objek, dan k adalah jumlah grup yang akan dibuat, algoritma klasifikasi mengatur objek menjadi k bagian ($k \leq n$), dengan setiap bagian mewakili grup. Setiap kluster dirancang untuk mengoptimalkan kriteria klasifikasi, seperti fungsi keragaman versus jarak, sehingga objek dalam kluster serupa, dan objek dalam kluster berbeda berbeda dalam hal fitur dataset. Persamaan untuk menghitung jarak antar data pada K-Means menggunakan rumus Euclidean Distance (D) yang tertera pada rumus tersebut.

III. Metode Penelitian

A. Tahapan Penelitian

Proses penelitian adalah langkah-langkah yang dilakukan selama penelitian (Irmayani, 2021). Proses survei dirancang untuk memfasilitasi akses ke hasil survei, untuk menyelesaikan survei secara tepat waktu dan untuk melaksanakan survei seperti yang diharapkan. Algoritma pencarian yang digunakan dapat dilihat pada gambar.



Gambar 1 Kerangka kerja penelitian

1) Identifikasi Masalah

- Mengetahui data pelanggan dari berbagai kelas ekonomi
- Pengolahan data yang buruk menjadi faktor utama penyebab terjadinya pengambilan keputusan yang kurang tepat oleh perusahaan
- Data pelanggan menjadi hal yang penting untuk menjadi pertimbangan perusahaan dalam pengambilan keputusan
- Mengolah data memerlukan metode inovatif supaya memudahkan dalam pemrosesan data dan data yang di hasilkan akurat.

2) Pengumpulan data

- Metode Pengumpulan Data

Metode pengumpulan adalah cara atau proses yang dilakukan untuk mengumpulkan data. Dalam penelitian pengumpulan data diperlukan untuk mencapai tujuan penelitian. Pada Jurnal ini penulis melakukan penelitian pada data pelanggan mall. 3 cara berikut merupakan cara yang digunakan :

1. Observasi

Observasi atau pengamatan adalah Suatu metode pengumpulan data yang dilakukan dengan pengamatan secara langsung, serta melihat dan mengambil data dari tempat penelitian tersebut

2. Wawancara

Teknik pengumpulan data yang dilakukan dengan cara tanya jawab.

3. Dokumentasi

Dokumentasi adalah teknik pengumpulan data dengan melihat langsung sumber-sumber dokumen yang terkait baik tertulis maupun elektronik.

- Sumber Data

Sumber data yang digunakan antara lain :

1. Data Primer, Data yang diperoleh peneliti dari tangan pertama

2. Data Sekunder, Data yang diperoleh peneliti dari sumber yang sudah ada.

3) *Pemrosesan data*

Data yang sudah diperoleh akan diolah dengan menggunakan tools aplikasi WEKA dengan proses sebagai berikut

- Data selection
- Pre-processing
- Transformation
- Clustering k-means

B. Data

Metode pengumpulan data yang dipakai adalah metode dokumentasi. pengumpulan data dilakukan dengan pengambilan data yang sudah ada dan tercatat dalam database pelanggan mall (Iriadi et al., 2021). Selanjutnya data dirubah menjadi numerik supaya bisa diolah menggunakan aplikasi WEKA .

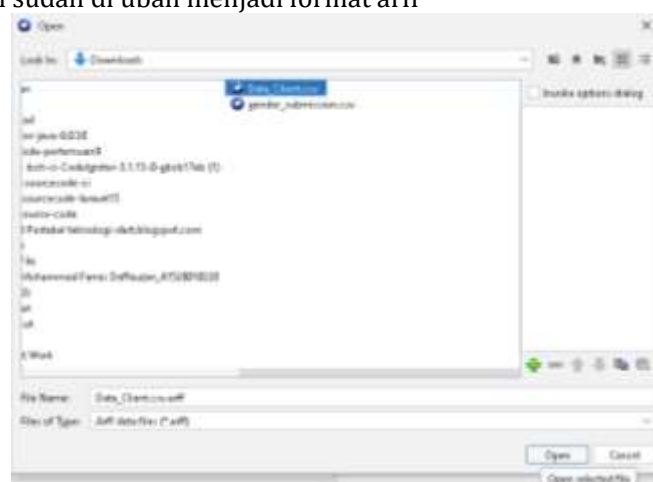
IV. Hasil Dan Pembahasan

A. Hasil

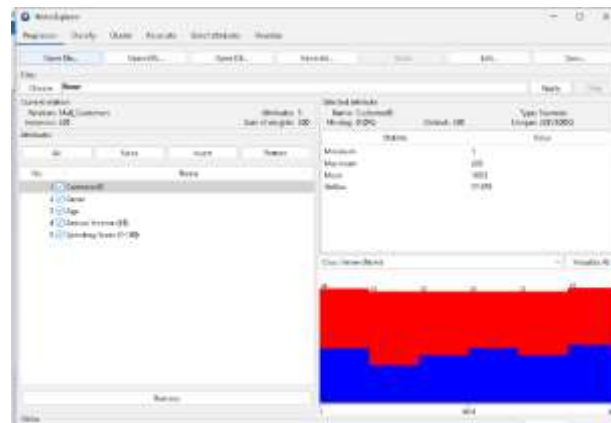
WEKA merupakan aplikasi pengolahan data yang sangat berguna dan memudahkan dengan berbagai tools, untuk melihat hasilnya kita perhatikan Langkah langkah berikut.

1) *Persiapkan data set*

Dataset yang digunakan bernama *Data_Client.csv.arff*. dataset adalah data yang telah dikumpulkan dengan format csv tapi sudah di ubah menjadi format arff

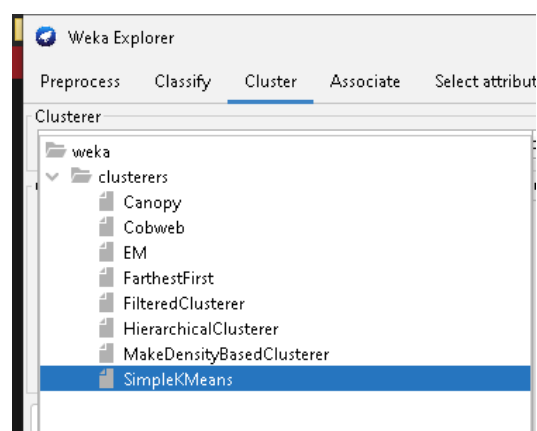


Gambar 2 Buka dataset di weka



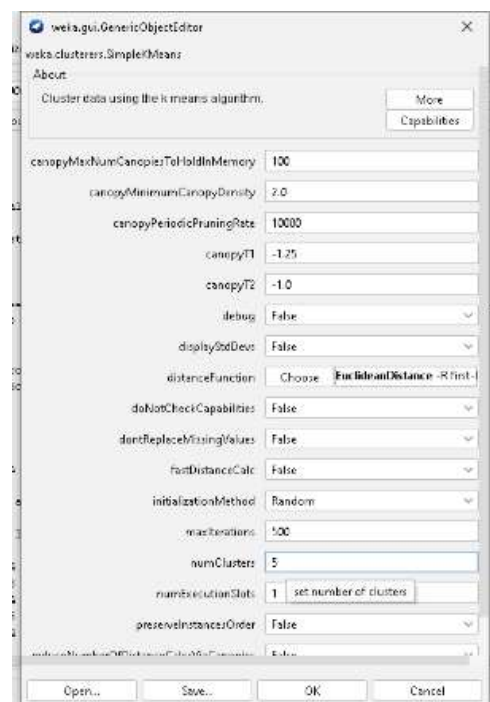
Gambar 3 Tampilan data set di weka

Untuk melakukan klustering Klik tab cluster lalu pilih opsi simplekmeans.

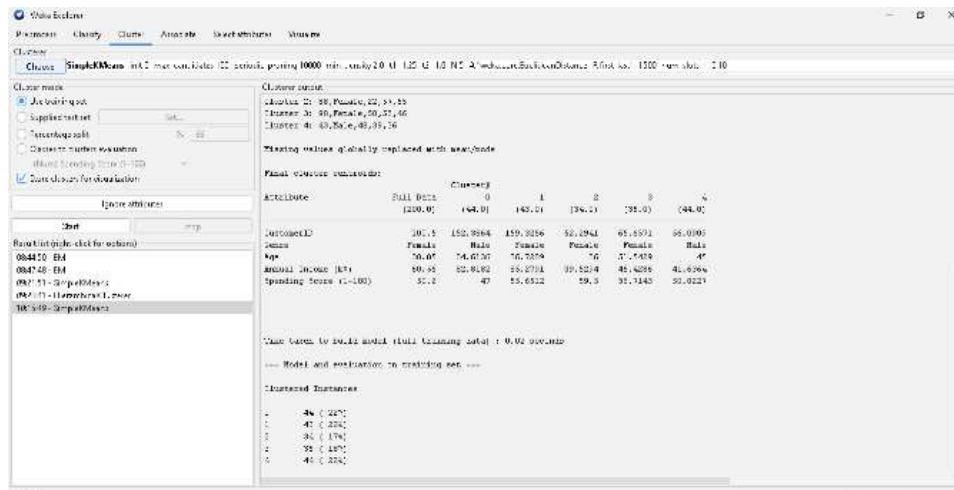


Gambar 4 .tab cluster

Klik 2 kali pada tulisan simplekmeans untuk mengatur jumlah cluster dengan mengganti pada kolom cluster num

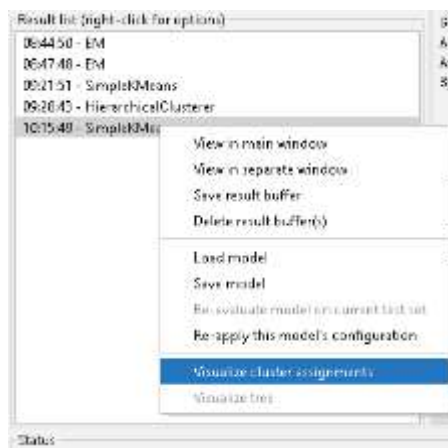


Gambar 5 mengatur jumlah cluster



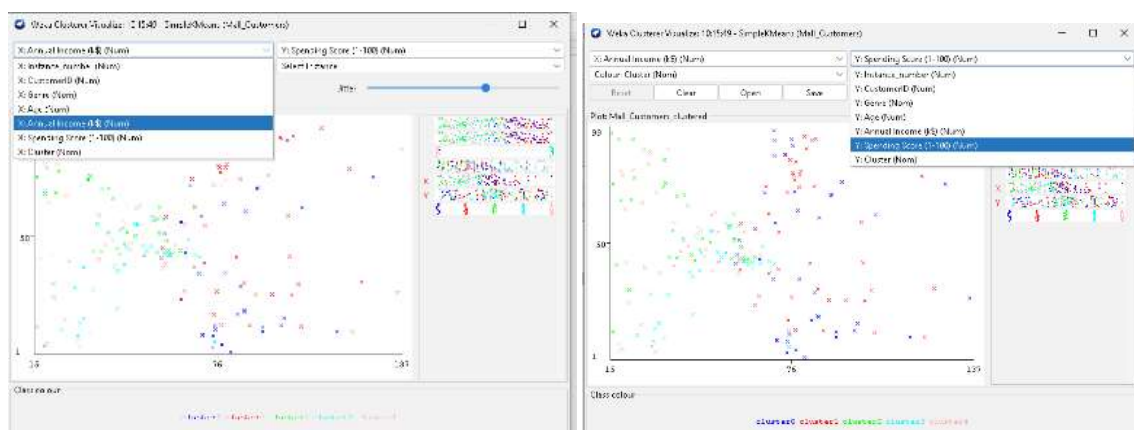
Gambar 6 Hasil clustering

Untuk visualisasi data kita klik bagian algoritma yang kita pakai yaitu simplekmeans dibagian kiri bawah lalu klik visualize cluster assignments



Gambar 7 Opsi visualisasi

Pilih sumbu x = annual income pilih sumbu y = spending score .tujuan nya adalah untuk mengvisualisasi ke 5 cluster dengan karakter masing masing cluster berdasarka sumbu x dan y.



Gambar 8 Sumbu X dan Y

- Cluster 0 (biru): penghasilan tinggi tetapi Pembelanjaan lebih sedikit
- Cluster 1 (merah): Rata dalam hal Penghasilan dan Pembelanjaan
- Cluster 2 (hijau): Penghasilan tinggi dan juga Pembelanjaan banyak

Cluster 3 (toska): Penghasilan lebih sedikit tetapi Pembelanjaan lebih banyak
Cluster 4 (ungu): penghasilan lebih sedikit, Pembelanjaan lebih sedikit



Gambar 9 Hasil visualisasi

B. Pembahasan

Data Mining dengan metode k-means mampu dengan baik mengklastering 200 pelanggan yang terdapat pada dataset Data_client yaitu dengan melakukan clustering dibagi menjadi 5 label antara lain, adalah :

- Cluster 0 (biru): penghasilan tinggi tetapi Pembelanjaan lebih sedikit(44%)
- Cluster 1 (merah): Rata dalam hal Penghasilan dan Pembelanjaan(43%)
- Cluster 2 (hijau): Penghasilan tinggi dan juga Pembelanjaan banyak (34%)
- Cluster 3 (toska): Penghasilan lebih sedikit tetapi Pembelanjaan lebih banyak(35%)
- Cluster 4 (ungu): penghasilan lebih sedikit, Pembelanjaan lebih sedikit(44%)

V. Kesimpulan

Dengan menggunakan aplikasi WEKA memudahkan peneliti dalam mengolah data, sehingga data pelanggan dapat diproses dengan baik menggunakan metode clustering K-means dengan baik dan akurat. Dengan ini data yang sudah diolah bisa menjadi pertimbangan bagi perusahaan untuk menentukan kebijakan dan pengambilan keputusan.

Daftar Rujukan

- Budi, A. (2015). Penerapan Data Mining Untuk Prediksi Nilai Tukar Petani Tanaman Pangan Di Indonesia Dengan Metode Linear Regression Dan Support Vector Machine. *Jurnal Informatika Dan Bisnis*, 4(2).
- Iriadi, N., Setioningtias, L., & Priatno, P. (2021). Implementasi Data Mining Pada Klasifikasi Ketidakhadiran Pegawai Menggunakan Metode C4. 5. *Computer Science (CO-SCIENCE)*, 1(1), 53–61.
- Irmayani, W. (2021). Visualisasi Data Pada Data Mining Menggunakan Metode Klasifikasi Naïve Bayes. *Jurnal Khatulistiwa Informatika*, 9(1).
- Jusia, P. A., Irfan, F. M., & Kurniabudi, K. (2019). Clustering data untuk rekomendasi penentuan jurusan perguruan tinggi menggunakan metode K-Means. *IKRA-ITH Informatika: Jurnal Komputer Dan Informatika*, 3(3), 75–84.
- Mardi, Y. (2017). Data Mining: Klasifikasi Menggunakan Algoritma C4. 5. *Jurnal Edik Informatika Penelitian Bidang Komputer Sains Dan Pendidikan Informatika*, 2(2), 213–219.
- Setyo, W. N., & Wardhana, S. (2019). Implementasi Data Mining Pada Penjualan Produk Di Cv Cahaya Setya Menggunakan Algoritma Fp-Growth.
- Susanto, H., & Sudiyatno, S. (2014). Data mining untuk memprediksi prestasi siswa

berdasarkan sosial ekonomi, motivasi, kedisiplinan dan prestasi masa lalu. *Jurnal Pendidikan Vokasi*, 4(2).

Zulfa, I., Rayuwati, R., & Koko, K. (2020). Implementasi data mining untuk menentukan strategi penjualan buku bekas dengan pola pembelian konsumen menggunakan metode apriori. *Teknika: Jurnal Sains Dan Teknologi*, 16(1), 69–82.