

LAPORAN TUGAS

Unsupervised Learning: Segmentasi Pelanggan Menggunakan K-Means

1. Pendahuluan

Perkembangan bisnis ritel berbasis daring (online retail) menghasilkan data transaksi dalam jumlah besar yang dapat dimanfaatkan untuk pengambilan keputusan. Salah satu pemanfaatan data tersebut adalah melakukan segmentasi pelanggan berdasarkan perilaku pembelian. Segmentasi pelanggan penting untuk membantu perusahaan memahami karakteristik pelanggan sehingga strategi pemasaran dapat dilakukan secara lebih tepat sasaran.

Pada tugas ini digunakan pendekatan *unsupervised learning*, yaitu metode pembelajaran mesin tanpa label, untuk mengelompokkan pelanggan berdasarkan kesamaan pola transaksi. Algoritma yang digunakan adalah **K-Means Clustering** dengan pendekatan **RFM (Recency, Frequency, Monetary)**.

2. Dataset

Dataset yang digunakan adalah **Online Retail Dataset**, yang berisi data transaksi penjualan dari sebuah perusahaan ritel online.

2.1 Deskripsi Dataset

Dataset memiliki atribut sebagai berikut:

- **InvoiceNo** : Nomor transaksi
- **StockCode** : Kode produk
- **Description** : Deskripsi produk

- **Quantity** : Jumlah produk yang dibeli
- **InvoiceDate** : Tanggal dan waktu transaksi
- **UnitPrice** : Harga satuan produk
- **CustomerID** : Identitas pelanggan
- **Country** : Negara pelanggan

Dataset ini bersifat transaksional dan tidak memiliki label kelas, sehingga sesuai untuk penerapan *unsupervised learning*.

3. Metodologi Penelitian

Tahapan penelitian yang dilakukan terdiri dari beberapa langkah utama, yaitu pengumpulan data, pembersihan data, pembentukan fitur, pemodelan clustering, serta analisis hasil.

3.1 Pembersihan Data (Data Cleaning)

Tahap pembersihan data dilakukan untuk meningkatkan kualitas data, dengan langkah-langkah sebagai berikut: 1. Menghapus data yang memiliki nilai kosong (*missing value*). 2. Menghapus transaksi dengan nilai Quantity dan UnitPrice kurang dari atau sama dengan nol, karena menunjukkan retur atau kesalahan pencatatan. 3. Mengonversi kolom InvoiceDate ke dalam format tanggal (*datetime*).

3.2 Feature Engineering (RFM Analysis)

Untuk merepresentasikan perilaku pelanggan, digunakan pendekatan RFM yang terdiri dari:

- **Recency (R)** : Jarak waktu (dalam hari) sejak transaksi terakhir pelanggan.
- **Frequency (F)** : Jumlah transaksi yang dilakukan pelanggan.
- **Monetary (M)** : Total nilai pembelian pelanggan.

Nilai RFM dihitung berdasarkan data transaksi dan dikelompokkan berdasarkan CustomerID.

3.3 Normalisasi Data

Karena algoritma K-Means sensitif terhadap skala data, maka dilakukan normalisasi menggunakan **StandardScaler** agar setiap fitur memiliki skala yang sebanding.

3.4 Clustering dengan K-Means

Algoritma **K-Means** digunakan untuk mengelompokkan pelanggan. Jumlah cluster ditentukan menggunakan metode **Elbow Method**, dengan memilih titik siku pada grafik WCSS.

Berdasarkan hasil pengujian, jumlah cluster optimal adalah **3 cluster**.

4. Hasil dan Pembahasan

4.1 Hasil Clustering

Proses clustering menghasilkan tiga kelompok pelanggan dengan karakteristik yang berbeda berdasarkan nilai rata-rata RFM pada masing-masing cluster.

4.2 Interpretasi Cluster

a) Cluster 0 – Pelanggan Loyal Bernilai Tinggi

Cluster ini memiliki nilai Recency rendah, Frequency tinggi, dan Monetary tinggi. Pelanggan dalam cluster ini merupakan pelanggan yang aktif bertransaksi dan memberikan kontribusi pendapatan terbesar bagi perusahaan.

Implikasi: Perusahaan perlu mempertahankan pelanggan pada cluster ini melalui program loyalitas dan penawaran eksklusif.

b) Cluster 1 – Pelanggan Potensial

Cluster ini memiliki Frequency dan Monetary yang relatif rendah, namun Recency yang cukup baik. Pelanggan dalam cluster ini tergolong pelanggan baru atau pelanggan yang masih jarang bertransaksi.

Implikasi: Pelanggan dalam cluster ini memiliki potensi untuk dikembangkan melalui strategi promosi dan peningkatan keterlibatan pelanggan.

c) Cluster 2 – Pelanggan Tidak Aktif

Cluster ini ditandai dengan nilai Recency tinggi serta Frequency dan Monetary yang rendah. Hal ini menunjukkan bahwa pelanggan jarang melakukan transaksi dan berpotensi berhenti menjadi pelanggan aktif.

Implikasi: Perusahaan dapat melakukan strategi re-engagement atau menurunkan prioritas pemasaran pada kelompok ini.

5. Visualisasi

Visualisasi dilakukan dengan memetakan hasil clustering ke dalam grafik dua dimensi menggunakan atribut Recency dan Monetary. Visualisasi ini membantu dalam memahami perbedaan karakteristik antar cluster secara lebih jelas.

6. Kesimpulan

Berdasarkan hasil penelitian, dapat disimpulkan bahwa penerapan *unsupervised learning* menggunakan algoritma K-Means dengan pendekatan RFM mampu mengelompokkan pelanggan online retail ke dalam beberapa segmen yang bermakna. Segmentasi ini dapat digunakan sebagai dasar dalam pengambilan keputusan bisnis, khususnya dalam penyusunan strategi pemasaran dan pengelolaan pelanggan.

7. Saran

Untuk pengembangan selanjutnya, penelitian dapat dilakukan dengan:

1. Menambahkan algoritma clustering lain seperti DBSCAN atau Hierarchical.
ClusteringSaran ini diberikan untuk menguji apakah pola segmentasi yang diperoleh konsisten ketika menggunakan algoritma clustering yang berbeda
2. Menggunakan variabel tambahan seperti kategori produk atau negara pelanggan.
Variabel tambahan disarankan untuk penelitian lanjutan guna memperkaya analisis tanpa mengganggu fokus penelitian saat ini.
3. Membandingkan hasil clustering dengan periode waktu yang berbeda.
Perbandingan antar periode waktu diperlukan untuk mengetahui stabilitas dan dinamika segmentasi pelanggan