



EKSPERIMEN TUGAS AKHIR DATA MINING

Nama : Naufal Arsyaputra Pradana

NIM : A11.2022.14606

Kelompok Kelas : A11.4504

JUDUL

Analisis Klasterisasi Pada Transaksi Penjualan Menggunakan Algoritma K-Means Clustering

DESKRIPSI SINGKAT

Penelitian ini berfokus pada penerapan teknik Data Mining, khususnya algoritma K-means Clustering (Model dari Unsupervised Learning), untuk menganalisis data transaksi penjualan (Data Tabular yang juga mengandung Data Kuantitatif) di sebuah mini market. Dengan memanfaatkan data dari dataset publik kaggle yang berisi informasi tentang pembelian pelanggan (Dataset terdiri dari 5 kolom, yaitu kode_barang, nama_barang, jumlah_transaksi, total_penjualan, rata_rata). Hasil analisis ini diharapkan agar dapat memberikan wawasan dalam mengoptimalkan strategi pemasaran dan pengelolaan stok barang.

MASALAH DALAM EKSPERIMEN

Data transaksi sering kali tidak terorganisir, dan tanpa analisis yang mendalam, maka pengelolaan stok, dan strategi pemasaran tidak dapat dilakukan dengan efektif. Hal ini dapat mengakibatkan kerugian pada minimarket, sehingga perlu adanya identifikasi segmen pelanggan untuk memahami preferensi mereka.

TUJUAN MELAKUKAN EKSPERIMEN

1. Menerapkan Algoritma K-means clustering untuk pengelompokan data transaksi penjualan.
2. Mengidentifikasi pola pembelian pelanggan.
3. Memberikan visualisasi yang informatif dari hasil klasterisasi.

Goals : Dengan melakukan eksperimen ini diharapkan dapat dengan mudah memahami dan membuat keputusan yang tepat dalam strategi pemasaran berdasarkan pola pembelian pelanggan.

TAHAPAN EKSPERIMEN

1. Pengumpulan Data
2. Pra Proses Data (Cleansing, Selection, Transform)
3. Penerapan Algoritma K-means clustering (Modelling)
4. Analisis Hasil (Evaluasi)
5. Visualisasi Data

PENGUMPULAN DATA

Data Transaksi Penjualan diambil dari dataset publik kaggle. Data tersebut mencakup informasi mengenai kode_barang, nama_barang, jumlah_transaksi, total_penjualan, dan rata_rata penjualan. Data ini juga mengandung elemen Data Kuantitatif, karena terdapat nilai numerik seperti "jumlah_transaksi," "total_penjualan," dan "rata_rata".

PRA PROSES DATA

Pra Proses ini meliputi pembersihan data dari duplikasi (Cleansing), kesalahan penulisan, dan data yang hilang, data juga akan dinormalisasi agar variabel-variabel yang berbeda dapat dibandingkan satu sama lain, dan data tidak semuanya dipakai, jadi hanya data yang sesuai yang akan dianalisis (Selection), lalu data akan diubah atau digabung ke dalam format yang sesuai untuk diproses (Transform).

PENERAPAN ALGORITMA K-MEANS CLUSTERING

Setelah data di proses, algoritma K-means clustering akan diterapkan (Modelling). Proses ini dimulai dengan menentukan jumlah cluster yang optimal, Metode Klasterisasi yang akan digunakan adalah Metode Elbow, karena Metode Elbow digunakan untuk membantu menentukan jumlah cluster yang paling sesuai, dimana grafik akan menunjukkan hubungan antara jumlah cluster dan variasi dalam cluster. Setelah jumlah cluster ditentukan, algoritma K-means akan dijalankan untuk mengelompokkan data transaksi ke dalam cluster yang sesuai berdasarkan kesamaan karakteristik. Model klasterisasi ini merupakan bentuk data Descriptive, bukan Predictive, jadi hasilnya merupakan informasi pola pembelian pelanggan untuk strategi pemasaran dan pengelolaan stok barang. (Sumber article [k-means-clustering](#)).

ANALISIS HASIL

Setelah klasterisasi, hasil cluster tersebut akan digunakan untuk menentukan pola pembelian di setiap segmen pelanggan (Evaluasi). Cluster yang akan terbentuk ada tiga yaitu, Data Penjualan Rendah, Data Penjualan Sedang, dan Data Penjualan Tinggi. Setiap cluster yang terbentuk akan dievaluasi berdasarkan atribut-atribut yang relevan, seperti total penjualan dan frekuensi pembelian.

VISUALISASI DATA

Visualisasi Data digunakan untuk menggambarkan hasil analisis. Grafik dan Diagram akan dibuat untuk menunjukkan distribusi setiap cluster dan hubungan antar atribut. Alat visualisasi yang digunakan adalah Matplotlib dan Seaborn.

PENJELASAN DATASET

Data yang digunakan dalam eksperimen ini adalah dataset publik yang berasal dari kaggle, berikut adalah link datasetnya [📄 barang_keluar](#), dimana dalam dataset tersebut berisi setiap transaksi penjualan yang telah dilakukan oleh pelanggan. Dataset transaksi penjualan tersebut terdiri dari sekitar 7400 baris yang masih menjadi data entry awal (Data mentah).

Berikut adalah struktur serta penjelasan dari masing-masing kolom:

1. **kode_barang**: Kode unik yang mewakili setiap jenis barang yang dijual.
2. **nama_barang**: Nama dari barang yang dijual, yang mencakup deskripsi produk seperti ukuran atau jumlah dalam kemasan.
3. **jumlah_transaksi**: Total transaksi yang dilakukan untuk barang tersebut .
4. **total_penjualan**: Jumlah total unit barang yang terjual selama periode tersebut.
5. **rata_rata**: Perhitungan rerata jumlah unit per transaksi.

TIMELINE EKSPERIMEN

Timeline seluruh proses eksperimen tugas akhir dalam menganalisis klasterisasi pada transaksi penjualan menggunakan algoritma K-means Clustering. Proses eksperimen melibatkan beberapa tahapan, yaitu dari pengumpulan data hingga evaluasi hasil klasterisasi.

No.	Tahapan Eksperimen	Tanggal Mulai	Tanggal Selesai	Durasi
1.	Pembuatan Laporan Eksperimen	7 November 2024	8 November 2024	2 hari
2.	Pengumpulan Data	9 November 2024	10 November 2024	2 hari
3.	Pra Proses Data (Cleansing, Selection, Transform)	11 November 2024	14 November 2024	4 hari
4.	Penerapan Algoritma K-Means Clustering	15 November 2024	21 November 2024	7 hari
5.	Analisis Hasil	22 November 2024	28 November 2024	7 hari
6.	Visualisasi Data	29 November 2024	5 Desember 2024	7 hari
7.	Laporan atau Deployment awal	6 Desember 2024	12 Desember 2024	7 hari
8.	Revisi dan Perbaikan	13 Desember 2024	17 Desember 2024	5 hari
9.	Laporan atau Deployment akhir (Penyelesaian)	18 Desember 2024	26 Desember 2024	9 hari

Total waktu yang dibutuhkan untuk menyelesaikan eksperimen ini adalah **50 hari** dari tanggal 7 November hingga 26 Desember 2024. Diharapkan hasil dari eksperimen ini dapat memberikan wawasan dalam pengambilan keputusan, terutama terkait pengelolaan stok dan strategi pemasaran sesuai dengan pola pembelian pelanggan.