

Homework Solution - Unsupervised Learning and Clustering

I Gusti Ngurah Agung Hari Vijaya Kusuma Batch 57

August 16, 2025

Submission Links

- Repository: github.com/AgungHari

1 Pendahuluan

1.1 Latar Belakang

Diberikan tugas untuk membuat model unsupervised learning dengan menggunakan dataset yang berisi data customer sebuah perusahaan penerbangan. Dataset ini mencakup berbagai fitur yang dapat menggambarkan nilai dari setiap customer, seperti ID Member, tanggal bergabung dalam program Frequent Flyer, jenis kelamin, tier program, kota asal, provinsi asal, negara asal, umur, jumlah penerbangan yang telah dilakukan, dan informasi terkait jarak penerbangan serta poin yang diperoleh.

Tujuan dari tugas ini adalah untuk menjawab Soal soal yang diberikan oleh Rakamin Academy. Dimana diharapkan outputnya dapat memberikan wawasan yang lebih dalam mengenai pola perilaku customer, segmentasi pasar, dan rekomendasi bisnis yang relevan berdasarkan hasil clustering. Dengan demikian, perusahaan penerbangan dapat mengoptimalkan strategi pemasaran dan meningkatkan pengalaman pelanggan.

1.2 Homework Unsupervised Learning

Adapun beberapa soal yang diberikan dalam tugas ini adalah sebagai berikut:

1. Lakukan EDA pada dataset untuk mendapatkan pemahaman umum mengenai data dan memandu proses feature engineering (20 poin)
 - Pastikan setiap kolom dataset memiliki tipe data yang tepat, tidak ada data kosong, bebas dari duplikat, dan berada di range value yang tepat.
 - Keluarkan statistik kolom baik numerik maupun kategorikal, cari bentuk distribusi setiap kolom (numerik), dan jumlah baris untuk setiap unique value (kategorikal).

- Cari tahu apakah ada kolom-kolom yang berkorelasi kuat satu sama lain.
2. Pilih fitur-fitur yang menurut teman-teman masuk akal secara bisnis untuk digunakan sebagai fitur clustering. Lakukan feature engineering! (20 poin)
 - Dari sekian banyak kolom yang ada, tentukan 3-6 fitur untuk digunakan sebagai fitur clustering. Tulis alasan teman-teman memilih fitur tersebut.
 - Lakukan preprocessing dan feature engineering (apabila fitur yang teman-teman pilih merupakan fitur baru yang dihasilkan dari fitur-fitur yang sudah ada).
 3. Lakukan clustering K-means! Temukan jumlah cluster yang menurut teman-teman optimal dan evaluasi cluster yang dihasilkan dengan visualisasi dan silhouette score (30 poin)
 - Temukan jumlah cluster yang optimal dengan menggunakan elbow method.
 - Lakukan clustering menggunakan K-means.
 - Evaluasi cluster yang dihasilkan dengan menggunakan visualisasi, gunakan PCA apabila diperlukan.
 4. Interpretasi cluster yang dihasilkan secara bisnis dan berikan rekomendasi yang sesuai dengan cluster yang dihasilkan (30 poin)
 - Tempelkan kembali label yang dihasilkan ke dataframe asal, dan keluarkan statistik fitur dari setiap cluster.
 - Deskripsikan secara kontekstual customer seperti apa yang ada di masing-masing cluster.
 - Berdasarkan cluster tersebut, berikan 1-2 rekomendasi bisnis.

2 Tinjauan Pustaka

3 Desain dan Implementasi

3.1 Deskripsi sistem

4 Pengujian dan analisis

4.1 Pengujian Sistem

5 Kesimpulan dan Saran