

!! Catatan: Diharapkan seluruh pengerjaan Assignment tidak sepenuhnya mengandalkan penggunaan AI **!!**

"Proses belajar ibarat menanam pohon. Jika hanya mengandalkan AI tanpa memahami esensinya, yang berkembang bukan kompetensimu, melainkan ketergantungan yang melemahkan."

- Learning Design Dibimbing

Assignment Guidance: Exploring Clustering Techniques with Python

Data Science dan Data Analyst Bootcamp

Periode Pembelajaran

Introduction to Unsupervised Machine Learning Models with Python

Exploring Clustering Techniques with Python

Objectives

1. Student mampu melakukan data preprocessing dengan menggunakan Python untuk mempersiapkan dataset.
2. Student mampu melakukan Exploratory Data Analysis (EDA) untuk memahami pola dan distribusi data.
3. Student mampu membuat model K-Means Clustering untuk mengelompokkan pelanggan berdasarkan data perjalanan mereka.
4. Student mampu melakukan visualisasi hasil clustering menggunakan grafik yang sesuai untuk mempermudah interpretasi.
5. Student mampu melakukan interpretasi model dan hasil clustering untuk menemukan insight yang berguna.
6. Student mampu melakukan evaluasi model clustering menggunakan metrik seperti Silhouette Score.
7. Student mampu melakukan analisis setiap cluster dan memberikan rekomendasi bisnis berdasarkan hasil analisis tersebut.

Deskripsi Assignment

Dalam assignment ini, student akan melakukan segmentasi pelanggan menggunakan teknik K-Means Clustering. Sebagai Data Scientist di sebuah perusahaan penerbangan, student diminta untuk mengelompokkan pelanggan berdasarkan pola perjalanan mereka, serta memberikan rekomendasi bisnis berdasarkan hasil segmentasi tersebut.

Tujuan utama dari tugas ini adalah untuk membantu student memahami cara membuat model clustering menggunakan K-Means, mengoptimalkan jumlah cluster yang tepat, serta memberikan wawasan bisnis yang berguna dari hasil clustering tersebut.

Detail Assignment

Link Dataset: [Flight](#)

Code	Description
MEMBER_NO-b	: ID Member
FFP_DATE	: Frequent Flyer Program Join Date
FIRST_FLIGHT_DATE	: Tanggal Penerbangan pertama
GENDER	: Jenis Kelamin
FFP_TIER	: Tier dari Frequent Flyer Program
WORK_CITY	: Kota Asal
WORK_PROVINCE	: Provinsi Asal
WORK_COUNTRY	: Negara Asal
AGE	: Umur Customer
LOAD_TIME	: Tanggal data diambil
FLIGHT_COUNT	: Jumlah penerbangan Customer
BP_SUM	: Rencana Perjalanan
SUM_YR_1	: Total credit/point di tahun pertama
SUM_YR_2	: Total credit/point di tahun kedua
SEG_KM_SUM	: Total jarak(km) penerbangan yg sudah dilakukan
LAST_FLIGHT_DATE	: Tanggal penerbangan terakhir
LAST_TO_END	: Jarak waktu penerbangan terakhir ke pesanan penerbangan paling akhir
AVG_INTERVAL	: Rata-rata jarak waktu
MAX_INTERVAL	: Maksimal jarak waktu
EXCHANGE_COUNT	: Jumlah penukaran
avg_discount	: Rata rata discount yang didapat customer
Points_Sum	: Jumlah poin yang didapat customer
Point_NotFlight	: point yang tidak digunakan oleh members

Untuk menyelesaikan assignment ini, **student harus mengikuti langkah-langkah berikut:**

1. Data Preprocessing

- Import library yang dibutuhkan: **numpy, pandas, seaborn, matplotlib, sklearn.**
- Muat dataset yang telah disediakan dan lakukan **pembersihan data**,

- termasuk menangani **missing values** dan **feature selection**.
- Pastikan dataset siap digunakan untuk analisis clustering.

2. Exploratory Data Analysis (EDA)

- Lakukan **analisis statistik deskriptif** terhadap dataset.
- Visualisasikan distribusi data numerik dan hubungan antar fitur menggunakan **scatter plot** atau **heatmap**.
- Identifikasi pola dalam data yang mungkin mempengaruhi hasil clustering.

3. Clustering dengan K-Means

- Gunakan metode **K-Means Clustering** untuk membagi pelanggan ke dalam beberapa kelompok (cluster).
- Tentukan jumlah cluster yang optimal dengan menggunakan metode seperti **Elbow Method** atau **Silhouette Score**.

4. Visualisasi Hasil Clustering

- Gunakan **scatter plot** atau metode visualisasi lainnya untuk menampilkan hasil clustering.
- Pastikan visualisasi dapat menunjukkan perbedaan antar cluster secara jelas.

5. Evaluasi Model

- Gunakan metrik seperti **Silhouette Score** untuk mengevaluasi hasil clustering.
- Bandingkan hasil clustering dengan jumlah cluster yang berbeda untuk menemukan model yang paling optimal.

6. Insight dan Rekomendasi

- Analisis setiap cluster dan identifikasi **karakteristik unik** dari masing-masing kelompok pelanggan.
- Berikan rekomendasi bisnis berdasarkan hasil analisis cluster, misalnya strategi pemasaran atau retensi pelanggan.

Ekspektasi Output

Student harus mengumpulkan link **Google Colaboratory** yang berisi:

- **Script Python** untuk seluruh langkah-langkah di atas.
- **Visualisasi grafik** yang menunjukkan hasil clustering dan karakteristik setiap cluster.
- **Dokumentasi dalam bentuk markdown atau teks** yang menjelaskan:
 - Langkah-langkah analisis
 - Model K-Means yang digunakan
 - Hasil evaluasi dan interpretasi dari setiap cluster.

Tools

Google Colaboratory

Pengumpulan Assignment

Deadline :

Maksimal H+7 Kelas (Pukul 23.30 WIB)

Details :

Dikumpulkan dalam bentuk Link Google Colaboratory, secara INDIVIDU, di LMS

Indikator Penilaian

No.	Aspek Penilaian	Parameter	Bobot Maksimal
1	Data Preprocessing	Student mampu melakukan pembersihan data, feature selection, dan persiapan dataset	25
2	Clustering	Student mampu menerapkan metode K-Means dengan jumlah cluster yang optimal	25
3	Visualisasi Data	Student mampu menampilkan hasil clustering dalam bentuk grafik	25
4	Insight dan Rekomendasi	Student mampu menganalisis hasil clustering dan memberikan rekomendasi bisnis	25

Referensi/Template

1. [Data Science: 6th International Conference of Pioneering Computer Scientists](#)
2. <https://drive.google.com/drive/folders/1R2zzE8cTf-JwNzP-orxaGDWoQGh2qmAy>
3. <https://ieeexplore.ieee.org/document/9736735>
4. <https://ieeexplore.ieee.org/document/9967896>

Sanksi Penggunaan AI:

Apabila student terdeteksi 100% menggunakan AI, maka hasil assignment akan diberikan skor 0

Ketentuan Pencapaian Nilai:

Nilai minimum Lulus Penyaluran Kerja: 75

Nilai minimum Lulus Bootcamp: 65

Ketentuan Penilaian:

Mengumpulkan Assignment tepat waktu: Sesuai dengan nilai yang diberikan mentor

Mengumpulkan Assignment 12 jam setelah deadline: - 3 dari nilai yang diberikan mentor

Mengumpulkan Assignment 1 x 24 Jam setelah deadline: - 6 dari nilai yang diberikan mentor

Mengumpulkan Assignment 2 x 24 Jam setelah deadline: - 12 dari nilai yang diberikan mentor

Mengumpulkan Assignment 3 x 24 Jam setelah deadline: - 18 dari nilai yang diberikan mentor

Mengumpulkan Assignment 4 x 24 Jam setelah deadline: - 24 dari nilai yang diberikan mentor

Mengumpulkan Assignment 5 x 24 Jam setelah deadline: - 30 dari nilai yang diberikan mentor

Mengumpulkan Assignment 6 x 24 Jam setelah deadline: - 36 dari nilai yang diberikan mentor

Mengumpulkan Assignment 7 x 24 Jam setelah deadline: - 42 dari nilai yang diberikan mentor