

SOAL SELEKSI DSA



Submission Guidelines

Berikut adalah ketentuan pengumpulan *submission* problem set dari penyisihan Data Science Academy Compfest 13:

1. Setiap tim mengumpulkan *submission* problem set dalam **satu buah file** dengan urutan:
 - Jawaban soal non-*dataset* dan
 - Laporan analisis *dataset*.
2. *Submission* problem set dibuat dengan ketentuan sebagai berikut:
 - *Submission* diketik secara terurut dan rapi menggunakan ukuran kertas A4 dengan margin kiri, atas, kanan, bawah **1, 1, 1, 1** (dalam inci).
 - Font **Times New Roman** ukuran **12** dengan *line spacing* **1,5**.
 - Penulisan jawaban menggunakan **Bahasa Indonesia** dan harap sesuai dengan Pedoman Umum Ejaan Bahasa Indonesia (PUEBI).
3. Khusus laporan analisis *dataset* dibuat dengan ketentuan sebagai berikut:
 - Laporan disusun dengan kriteria berikut:
 - a. Halaman Judul,
 - b. Latar Belakang,
 - c. Jawaban Soal,
 - d. Hasil Analisis Tambahan,
 - e. Kesimpulan,
 - f. Daftar Pustaka (menggunakan **APA style**), dan
 - g. Lampiran (opsional).
 - Laporan **maksimal 20 halaman** tidak termasuk halaman judul dan lampiran.
4. Jawaban dikumpulkan dalam format file **.pdf** dan format penamaan **DSA-[Nama Tim]**.

Problem Set

Ketentuan pengumpulan jawaban dapat dilihat pada Submission Guidelines.

Bagian 1: Analisis *Dataset*

Berikut link untuk mengakses *dataset* dan referensi tambahan yang digunakan dalam problem set penyisihan Data Science Academy Compfest 13:

Dataset : tiny.cc/Datacovidjakarta

Referensi tambahan : corona.jakarta.go.id

1. Dari *dataset* yang disediakan, temukan nilai *mean*, median, dan modus dari positif COVID-19 harian Jakarta.
2. Dari *dataset* yang disediakan, temukan nilai minimal dan maksimal dari positif COVID-19 harian Jakarta.
3. Dari *dataset* yang disediakan, temukan nilai-nilai *outlier* yang ada (menggunakan variabel yang kalian tentukan).
4. Dari *dataset* yang disediakan, usulkan dua buah variabel dan berikan analisis korelasi antara kedua variabel tersebut. Jelaskan apa kesimpulan yang dapat diambil berdasarkan analisis kalian.
5. Dari *dataset* yang disediakan, buatlah analisis dengan runtunan berikut:
 - a. Problem Statement,
 - b. Hypothesis,
 - c. Exploratory Data Analysis,
 - d. Initial Findings,
 - e. Deep Dive Analysis, dan
 - f. Conclusion and Recommendation.

6. Terkait pertanyaan sebelumnya, jelaskan dengan lebih detail usaha *exploratory data analysis* (EDA) yang kalian lakukan dan mengapa kalian melakukan teknik EDA tersebut.
-

Bagian 2: Data Science Fundamentals (non-*dataset*)

1. Jelaskan secara teori statistik mengenai *outlier* (pencilan), implikasinya dalam analisis data, serta bagaimana melakukan manajemen data terhadap kasus *outlier*.
2. Jelaskan konsep dan prinsip korelasi, lalu sebisa mungkin kaitkan dengan dasar-dasar statistik serta implikasinya terhadap konsep/teori statistik lain.
3. Sebutkan teori dasar *machine learning* yang kalian ketahui, lalu jelaskan dalam bahasa sederhana mengenai teori tersebut dan implikasinya.
4. Menggunakan bahasa kalian sendiri, jelaskan kaitan antara *artificial intelligence*, *machine learning*, dan *deep learning*.
5. Apakah yang kalian ketahui mengenai interpretasi data? Bagaimana signifikansi dan tantangannya? Bagaimana kaitan interpretasi data dengan *data story telling* dan *decision making*?