

1. EDA (Exploratory Data Analysis) merupakan proses awal analisis data untuk memahami pola, mengidentifikasi struktur, menguji hipotesis serta menemukan komponen penting dalam suatu dataset.

EDA penting dalam proses analisis data karena dengan melakukan EDA, kita dapat menemukan kesalahan data, outlier serta mengetahui hubungan antardata. Dengan begitu, kita dapat lebih memahami dataset yang kita gunakan.

2. Perbedaan Supervised Learning, Unsupervised Learning dan Reinforcement Learning adalah pada cara kerjanya. Pada Supervised Learning, dataset nya sudah mempunyai label dan outputnya sudah diketahui, lalu pengguna memetakan input ke output tersebut. Pada Unsupervised Learning, dataset nya belum mempunyai label, pola dalam data diidentifikasi serta dikelompokkan berdasarkan kesamaan. Sedangkan pada Reinforcement Learning, dataset yang digunakan ada umpan balik setelah melakukan tindakan tertentu. (bisa berupa reward atau penalty)
Problemset pada penugasan ini termasuk tipe Supervised Learning.
3. Overfitting adalah suatu hasil dari penentuan akurasi data yang memiliki skor akurasi tinggi pada data training, namun skor rendah pada data testing. Sedangkan underfitting adalah skor akurasi rendah pada data training dan data testing.
4. Saya tidak menggunakan data eksternal.
5. Menurut saya, problemset dalam penugasan ini sudah cukup baik, dan cukup membutuhkan ketelitian dan usaha yang tinggi untuk menganalisis serta memperbaiki data kosongnya.