

TUGAS PENAMBANGAN DATA P.4

Berikut adalah jawaban dari pertanyaan latihan yang terdapat pertemuan

QUIZ HALAMAN 21

Soal:

Anda menjalankan perusahaan, dan Anda ingin mengembangkan algoritma pembelajaran untuk mengatasi masing-masing dari dua masalah berikut:

- **Problem 1:** Perusahaan Anda akan segera meluncurkan produk baru. Anda ingin memprediksi **berapa banyak** dari produk tersebut akan terjual selama 3 bulan ke depan.
- **Problem 2:** Anda mengamati aktivitas media sosial terkait peluncuran produk baru dan mencoba untuk mengidentifikasi **apakah opini orang positif atau negatif** terhadap produk tersebut.

Pilihan jawaban:

- a. Memberlakukan keduanya sebagai masalah Classification
- b. Memberlakukan Problem 1 sebagai masalah Classification; Problem 2 sebagai masalah Regression
- c. Memberlakukan Problem 1 sebagai masalah Regression; Problem 2 sebagai masalah Classification
- d. Memberlakukan keduanya sebagai masalah Regression

Jawaban yang benar adalah c. Memberlakukan Problem 1 sebagai masalah Regression; Problem 2 sebagai masalah Classification

Penjelasan:

- **Problem 1 → Regression**
Karena kita ingin memprediksi angka (berapa banyak unit produk yang akan terjual), maka ini merupakan tugas regresi. Regresi digunakan ketika output atau target berupa nilai numerik/berkelanjutan.
- **Problem 2 → Classification**
Di sini kita ingin mengetahui apakah opini orang itu positif atau negatif, yang merupakan kategori. Oleh karena itu, ini adalah tugas klasifikasi, karena output-nya adalah label diskret seperti "positif" atau "negatif".

QUIZ HALAMAN 31

Soal:

Dari contoh berikut, manakah yang akan Anda atasi dengan menggunakan algoritma **supervised learning** dan mana yang menggunakan **unsupervised learning**?

Pernyataan:

- a. Diberi satu set artikel berita yang ditemukan di web, kelompokkan ke dalam set artikel yang berbicara tentang cerita yang sama.
- b. Diberikan kumpulan data informasi cuaca di Jakarta 3 tahun terakhir seperti cerah, hujan, dan mendung, belajar memprediksi cuaca esok hari.
- c. Diberikan kumpulan data pelanggan, temukan segmen pasar secara otomatis dan mengelompokkan pelanggan ke dalam segmen yang berbeda.
- d. Diberikan kumpulan data pasien yang didiagnosis menderita kanker serviks atau tidak, pelajari cara mengklasifikasikan pasien baru menderita kanker serviks atau tidak.

Jawaban dan Penjelasan:

- a. **Unsupervised Learning**
Karena tidak ada label/klasifikasi yang tersedia sebelumnya, ini adalah proses pengelompokan berdasarkan kemiripan isi berita → ini adalah clustering.
- b. **Supervised Learning**
Karena data sebelumnya memiliki label cuaca (cerah, hujan, dll), dan kita ingin memprediksi label baru, maka ini adalah klasifikasi atau prediksi → supervised learning.
- c. **Unsupervised Learning**
Mengelompokkan pelanggan berdasarkan data yang tidak berlabel untuk menemukan segmen pasar termasuk clustering → unsupervised learning.
- d. **Supervised Learning**
Karena data sudah memiliki label (terdiagnosis atau tidak), dan digunakan untuk memprediksi pasien baru, maka ini adalah klasifikasi → supervised learning.