

Nama: Andi Sri Mulyani Ananda Halilintar

No. registrasi: 1939538840-225

Kelas: A

Soal Pre-Test Associate Data Scientist (30 Soal)

1. Data pencilan (*outlier*) adalah ?
 - a) Data yang nilainya paling kecil.
 - b) Data yang nilainya paling besar.
 - c) **Daya yang berbeda dari data mayoritas.**
 - d) Data testing.
2. Dibawah ini, manakah yang bukan ciri dari data terstruktur ?
 - a) Data memiliki struktur yang jelas.
 - b) **Data tidak memiliki struktur yang jelas.**
 - c) Sebagian data memiliki struktur yang jelas.
 - d) Data yang strukturnya bisa berubah.
3. Masalah Data Science yang diatasi dengan supervised learning adalah ?
 - a) **Klasifikasi dan regresi.**
 - b) Regresi dan Clustering.
 - c) Kalsifikasi dan Dimensionality Reductionn.
 - d) Regresi
4. Pilihlah pernyataan yang benar mengenai jenis data ?
 - a) Data nominal memiliki interval yang tetap.
 - b) Data ordinal tidak memiliki urutan.
 - c) Data interval memiliki nilai terkecil (nilai 0).
 - d) **Data rasio memiliki interval yang tetap antar dua data.**
5. Asumsi model regresi linier antara lain ?
 - a) Antar residual model tidak saling berkorelasi.
 - b) Varians dari residual tidak tergantung kepada nilai fitur prediktor.
 - c) Fitur prediktor memiliki korelasi tinggi dengan fitur target.
 - d) **Semua benar**
6. Membersihkan data meliputi proses untuk ?
 - a) **Menangani data yang hilang, menangani data berulang, dan menangani data pencilan (outlier).**
 - b) Memilih fitur prediktor yang dominan,
 - c) Merepresentasikan fitur nominal dengan teknik one-hot-encoding.
 - d) Semua jawaban salah.

7. Karakteristik dari data outlier sebuah fitur secara statistik adalah ?
 - a) Nilai data diluar dari selang $[\mu - 3\sigma, \mu + 3\sigma]$ dimana μ adalah rata-rata data dan σ adalah deviasi standar data.
 - b) Nilai data diluar dari selang $[\mu - 2\sigma, \mu + 2\sigma]$ dimana μ adalah rata-rata data dan σ adalah deviasi standar data.
 - c) Nilai data diluar dari selang $[\mu - 1.5\sigma, \mu + 1.5\sigma]$ dimana μ adalah rata-rata data dan σ adalah deviasi standar data.
 - d) Tidak ada yang benar.
8. Apa yang dimaksud imbalanced data ?
 - a) Data yang nilai-nilainya terhadap rata-rata tidak seimbang.
 - b) Data setiap kategori jumlahnya tidak seimbang.
 - c) Data yang jumlah kategorinya merupakan angka ganjil.
 - d) Tidak ada yang benar.
9. Suatu besaran yang dapat diubah atau berubah sehingga mempengaruhi peristiwa atau hasil penelitian dinamakan ...
 - a) Variabel
 - b) Tipe Data
 - c) Operator
 - d) Record
10. Berapa rentang tipe data untuk integer byte ...
 - a) 0 .. 255
 - b) -128 .. 127
 - c) 0 .. 65535
 - d) 0 .. 552366
11. Dibawah ini yang merupakan ruang lingkup statistik deskriptif ialah sebagai berikut, kecuali
 - a) Penyajian data biasanya dalam bentuk tabel, grafik, diagram
 - b) Melakukan pengujian hipotesis
 - c) Hasil analisis berupa proporsi dan varians
 - d) Hasil analisis berupa modus, median, mean
12. Dibawah ini type chart yang paling tepat untuk dapat digunakan pada data data inflasi suatu negara dari tahun ke tahun adalah
 - a) Scatter Plot dan Line Chart
 - b) Pie Chart dan Bar Chart
 - c) Pie Chart dan Scatter Plot
 - d) Heatmap dan Bar Chart
13. Pilihlah type chart yang dapat digunakan untuk dapat memvisualisasikan variable-variable berikut secara bersamaan : Waktu, jumlah positif covid, jumlah kematian covid, jumlah kesembuhan , kecuali
 - a) Line Chart
 - b) Pie Chart
 - c) Heatmap
 - d) Bar chart

14. Mengapa kita membutuhkan transformasi fitur?
- a) Mengubah fitur non-numerik menjadi numerik
 - b) Mengubah ukuran input ke ukuran tetap
 - c) Baik A dan B
 - d) Tidak ada
15. Manakah dari berikut ini yang benar tentang outlier
- a) Titik data yang sangat menyimpang dari pengamatan normal
 - b) Dapat mengurangi akurasi model
 - c) Baik A dan B
 - d) idak ada
16. Algoritma mana yang tidak memerlukan penskalaan fitur?
- a) A. Naive Bayes
 - b) B. Pohon Keputusan
 - c) C. Pohon Keputusan C
 - d) D. Tidak ada
17. Tujuan penskalaan fitur adalah untuk -
- a) Mempercepat waktu pelatihan
 - b) Mendapatkan akurasi yang lebih baik
 - c) Baik A dan B
 - d) Tidak ada
18. Dalam standardisasi, fitur akan diskalakan ulang dengan -
- a) Rata-rata 0 dan Varians 0
 - b) Rata-rata 0 dan Varians 1
 - c) Rata-rata 1 dan Varians 0
 - d) Rata-rata 1 dan Varians 1
19. Apa yang dimaksud dengan Perangkat Variabel Dummy?
- a) Multikolinearitas antar variabel dummy
 - b) Satu variabel memprediksi nilai lainnya
 - c) Baik A dan B
 - d) Tidak ada di atas
20. Label diterapkan sebagai identitas untuk atribut:
- a) Prediktor
 - b) Independent
 - c) Dependent
 - d) Tidak ada yang benar
21. Jika data mempunyai label, maka paling mungkin digunakan untuk pembelajaran model yang bersifat:
- a) Supervised
 - b) Unsupervised
 - c) Semi supervised

- d) Tidak ada yang benar
22. Jika data tidak mempunyai label, maka paling mungkin digunakan untuk pembelajaran model yang bersifat:
- a) Supervised
 - b) **Unsupervised**
 - c) Semi supervised
 - d) Tidak ada yang benar
23. Dalam penyajian dataset berupa matriks, label biasanya merupakan fitur yang posisinya ada di urutan kolom:
- a) Paling awal
 - b) **Paling akhir**
 - c) Bisa di mana saja
 - d) Tidak ada yang benar
24. Pengertian label dan target:
- a) **Label sama dengan target**
 - b) Label tidak sama dengan target
 - c) Label biasanya disebut sebagai prediktor
 - d) Tidak ada yang benar
25. Label regresi adalah nilai yang bersifat
- a) **Kontinyu**
 - b) Kategorikal
 - c) Kontinyu dan kategorikal sekaligus
 - d) Tidak ada yang benar
26. Label klasifikasi adalah nilai yang bersifat
- a) Kontinyu
 - b) **Kategorikal**
 - c) Kontinyu dan kategorikal sekaligus
 - d) Tidak ada yang benar
27. Label klasifikasi adalah nilai yang bersifat
- a) Kontinyu
 - b) **Kategorikal**
 - c) Kontinyu dan kategorikal sekaligus
 - d) Tidak ada yang benar
28. Pelabelan data (atau anotasi data) adalah proses menambahkan atribut target ke data :
- a) pelatihan / training
 - b) testing / pengujian
 - c) **pelatihan maupun pengujian**
 - d) tidak ada yang benar
29. Salah satu teknik untuk meningkatkan efisiensi dan akurasi pelabelan data adalah:
- a) **Konsensus pemberi label**

- b) Konsensus pembuat model
- c) Decision Tree
- d) tidak ada yang benar

30. Salah satu pendekatan pelabelan data adalah :

- a) **Pelabelan internal**
- b) Interpolasi label
- c) Pemrograman
- d) Random

*Jawaban ditandai dengan line kuning