

Nama: Mikail Achmad
NIM: 24/542370/PA/23026

Soal

1. Dengan menggunakan kalimat Anda sendiri, jelaskan yang dimaksud dengan EDA (Exploratory Data Analysis)! Mengapa hal tersebut penting dalam proses analisis data?
2. Jelaskan perbedaan antara supervised learning, unsupervised learning, dan reinforcement learning! Termasuk kategori yang manakah problemset pada penugasan open recruitment ini?
3. Apa yang dimaksud dengan overfitting dan underfitting dalam konteks machine learning? Apakah dalam pengerjaan penugasan praktek Anda mengalami salah satu atau kedua masalah tersebut? Bagaimana Anda menanganinya?
4. Seandainya dalam proses prediksi penugasan problemset diperbolehkan menambahkan data eksternal, apakah Anda akan menggunakan data eksternal? Jika iya, data apa yang akan Anda gunakan dan jelaskan alasannya! (NB: selain data primer harga laptop dengan spesifikasi yang sama, contoh: data harga laptop di marketplace)
5. Bagaimana tanggapan dan evaluasi Anda terhadap problem set pada penugasan praktek dan soal teori pada proses open recruitment ini?

Jawaban

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. Supervised Learning: Pada metode supervised learning kita memberikan pelatihan kepada model dengan memberi tahu soal beserta kunci jawabannya. Ini membuat model machine learning ketika nanti diberi soal lain yang serupa akan dapat menjawabnya dengan benar karena sudah belajar dari pola soal dan jawaban yang sudah diberikan sebelumnya.

Unsupervised Learning: Bedanya dengan supervised learning adalah metode ini tidak memberikan kunci jawaban kepada model. Jadi, model akan mengelompokkan data sesuai pola yang ia temukan, jadi model hanya fokus pada pengenalan pola tanpa target atau output yang spesifik.

Reinforcement Learning: Pada metode ini sebuah agen akan berulang kali melakukan tindakan sesuai apa yang benar menurutnya. Ketika ia menjalankan tindakan yang benar dia akan mendapatkan reward (+ point) dan ketika ia melakukan tindakan yang salah ia akan mendapatkan punishment (- point). Dengan mencoba berulang-ulang dan mengumpulkan pengalaman, agen belajar memilih tindakan terbaik agar mendapat hadiah paling banyak dan mencapai tujuannya. Sehingga model machine learning dengan metode ini memiliki pengalaman yang sangat tinggi dan dapat mencapai tujuannya dengan benar.

3. *Overfitting* adalah suatu kondisi yang terjadi ketika model mempelajari train dataset terlalu baik sehingga model tidak bisa memprediksi dataset selain train dataset dengan baik. Sementara itu, *underfitting* adalah kondisi yang terjadi ketika model tidak bisa mempelajari train dataset dengan baik sehingga model tidak bisa memprediksi dataset manapun dengan baik.

Dalam pengerjaan penugasan praktik ini tidak terjadi *underfitting* karena jumlah data yang dimiliki cukup besar, kemudian terjadi sedikit *overfitting* dimana hasil prediksi test dataset tidak sebagus ketika memprediksi train dataset, tetapi *overfitting* tidak terjadi dalam margin yang begitu besar.

4. Saya akan menggunakan data eksternal yang relevan dengan studi kasus. Dalam studi kasus dataset ini, bisa ditambahkan data tingkat edukasi, besaran umr daerah tempat bekerja, dsb.

Nama: Mikail Achmad
NIM: 24/542370/PA/23026

5. Menurut saya, problem set dalam penugasan ini sudah cukup baik, dan cukup membutuhkan ketelitian dan usaha yang tinggi untuk menganalisis serta memperbaiki data kosongnya.