

SILABUS TRAINING **DATA SCIENCE**

Machine Learning using Python

Machine Learning (ML) merupakan salah satu bidang yang termasuk ke dalam sub bidang dari Artificial Intelligence (AI). ML menggunakan data yang sudah dibersihkan, divalidasi, dan ditransformasikan agar dapat dimodelkan untuk membuat keputusan tanpa menerima instruksi secara eksplisit dari manusia, hal ini yang membedakannya dengan aplikasi tradisional. Keputusan dihasilkan sebagai hasil dari pembelajaran data, sehingga dapat mengenali pola data dan memberikan prediksi.

Materi training ini membahas tentang implementasi bahasa Python untuk penggunaan package-package yang berhubungan dengan siklus pengembangan ML. Beberapa pemodelan yang dibahas di training ini yaitu: Simple dan Multiple Regresi Linier, Regresi Logistic, Supervised Learning, Unsupervised Learning, beserta teknik-teknik untuk evaluasi dan validasi. Kemampuan ML ini merupakan tahap prasyarat untuk melangkah lebih lanjut ke sub topik ML lainnya seperti Deep Learning.

Setelah mengikuti training ini peserta diharapkan dapat memahami konsep dan mengimplementasikan bahasa Python untuk kebutuhan pengembangan ML yang dapat membantu dalam pengambilan keputusan bisnis organisasi atau perusahaan.

Materi training disusun dari berbagai sumber dan media pembelajaran. Training disampaikan dalam bentuk pengenalan konsep, teori, dan praktek dalam porsi yang seimbang, serta dipandu oleh trainer atau instruktur yang telah berpengalaman di bidangnya.

Durasi dan Waktu

- Durasi Pelaksanaan: 3 (Tiga) Hari
- Waktu Pelaksanaan: 09.00 - 15.00 WIB

Prasyarat

- Peserta harus sudah memiliki pengetahuan dasar pemrograman Python seperti: penggunaan library, variabel, tipe data, struktur data (list, tuple, dictionary, set, array), struktur keputusan, struktur perulangan, pembuatan fungsi, dan pengetahuan tentang Data Analysis.

Outline Materi

■ Instalasi dan Konfigurasi Tools

Awal pertemuan training ini akan diawali dengan pembahasan tentang instalasi dan konfigurasi tools yang akan digunakan selama training. Tools yang akan digunakan yaitu JupyterLab yang dapat diinstal melalui [Pip](#) (packager installer for Python) atau distribusi [Anaconda](#). Ekstensi JupyterLab dan Google Colab juga dapat diinstal di [Visual Studio Code](#) sehingga dapat dijadikan sebagai pilihan utama tools lainnya yang dapat digunakan.

■ Definisi dan Alur Kerja Pengembangan ML

Sebelum masuk ke pembahasan materi inti, pada sesi ini akan dibahas tentang apa itu ML, istilah-istilah, dan siklus pengembangan ML.

■ Simple dan Multiple Regresi Linier

Regresi Linier baik itu untuk jenis Simple dan Multiple digunakan untuk membuat model yang dapat memprediksikan data numerik yang kontinyu, misalnya memprediksi harga mobil atau harga rumah dari beberapa fitur atau variabel-variabel yang mempengaruhinya seperti warna, ukuran mesin, konsumsi bahan bakar (untuk prediksi harga mobil). Visualisasi dapat membantu secara grafis untuk melihat perbandingan data nyata dengan data hasil prediksi seberapa akurat model yang dibuat, tentunya juga dianalisa dari perhitungan metrik yang dihasilkan.

■ Regresi Logistik

Regresi Logistik digunakan untuk memprediksi nilai yang hanya akan memberikan dua nilai keputusan yang dihasilkan seperti ya atau tidak, nol atau satu. Model ini memberikan probabilitas statistik suatu observasi ke dalam salah satu dari dua kelas (misalnya: ya atau tidak). Regresi ini juga disebut sebagai pengklasifikasi biner. Contoh nyata implementasinya yaitu memprediksi apakah seorang pelanggan akan berhenti berlangganan internet atau tidak berdasarkan data historis pelanggan, atau peluang seseorang beresiko terkena sebuah penyakit berdasarkan informasi demografis seperti rentang usia dan jenis kelamin.

■ Supervised Learning

Pemodelan yang digunakan dalam kelompok ini digunakan untuk klasifikasi (clustering) yang telah dilatih menggunakan data pelatihan yang berlabel untuk memprediksi label pada data baru. Label dalam klasifikasi sifatnya kategorikal dan diskrit. Pemodelan ini diperlukan apabila data kompleks dan memerlukan interpretasi yang tidak sederhana regresi logistik. Teknik pemodelan yang dibahas pada sesi ini yaitu: Multi-class Classification, Decision Trees, Regression Trees, SVM, K-NN, dan Random Forest.

■ Unsupervised Learning

Model-model yang dihasilkan dalam kelompok ini menggunakan data yang tidak berlabel untuk kebutuhan clustering data. Proses clustering dilakukan secara mandiri untuk menemukan polanya berdasarkan karakteristik atau kesamaan tertentu. Contoh implementasi nyata seperti identifikasi genre musik, segmentasi pelanggan berdasarkan kategori tertentu, dan deteksi anomali. Selain itu juga dibahas teknik reduksi dimensi untuk menyederhanakan struktur data, dan untuk improve hasil clustering. Pemodelan yang dibahas di sesi ini yaitu: K-Means, DBSCAN, dan HDBSCAN, sedangkan untuk reduksi dimensi yaitu: PCA, t-SNE, dan UMAP.

■ Evaluasi dan Validasi Model

Evaluasi dan validasi model dibahas pada sesi terakhir ini. Beberapa teknik yang dibahas disini yaitu: metrik-metrik klasifikasi dan evaluasinya seperti: accuracy confusion matrix, precision, recall, dan F1 score, metrik-metrik model regresi seperti: MAE, MSE, RMSE, dan R-squared. Untuk validasi model dilakukan proses membagi data ke dalam data training, validasi, dan test, termasuk di dalamnya yaitu cross validation method untuk mencegah overfitting terhadap data test. Pada akhirnya semua alur kerja atau proses telah dilakukan sehingga kita akan memiliki model-model yang terbaik untuk mendapatkan wawasan yang dijadikan sebagai argumen untuk memberikan keputusan bisnis organisasi atau perusahaan.

Note

- Reguler Training dan Online Training tidak ada minimum kuota peserta.
- In-House Training dan Off-Site Training minimum kuota peserta 8 orang.
- Online Training menggunakan software Zoom Meeting.
- Hubungi kami untuk biaya training serta pembuatan Surat Penawaran.