

Nama : Nabilah Salwa

NIM : 1103204060

## Introduction

Halo semua, perkenalkan nama saya Nabilah Salwa dengan NIM 1103204060, pada kesempatan kali ini saya akan membahas tentang machine learning menggunakan Classification dengan metode XGBoots, Source yang saya gunakan adalah diabetes dataset. Disini kita akan melakukan EDA, Visualisasi Data, Training Data, Evaluating Data, dan Inputing Data

Langkah pertama adalah mengimpor library pandas, library numpy, library matplotlib.pyplot, library seaborn, library scikit-learn, dan library XGBoost

Setelah kita masukan dataset ke google drive selanjutnya kita menghubungkan google colab dengan google drive, kode ini digunakan untuk membacaca dataset menggunakan Pandas dengan path sesuai dataset di Google Drive

Berikutnya kita lakukan EDA (Exploratory Data Analysis) adalah suatu pendekatan dalam analisis data yang bertujuan untuk meringkas dan memahami karakteristik utama dari suatu dataset. EDA dilakukan dengan menggunakan teknik statistik dan visualisasi data untuk mengidentifikasi pola, hubungan, anomali, dan tren dalam data. (Tampilkan dan jelaskan kode tentang EDA dan berbagai macam caranya yang ada di colab). Dengan EDA, kita bisa melihat struktur dataset, seimbang atau tidaknya kelas, dan pola-pola menarik dalam datanya.

Tahap selanjutnya adalah Visualisasi EDA (Exploratory Data Analysis) adalah representasi grafis dari data yang digunakan untuk membantu pemahaman karakteristik dan pola dalam dataset. Visualisasi EDA bertujuan untuk memberikan wawasan visual yang dapat membantu peneliti atau analis dalam mengeksplorasi data, mengidentifikasi pola, hubungan, dan tren, serta memahami distribusi variabel. (Tampilkan dan jelaskan kode tentang Visualisasi EDA dan berbagai macam caranya yang ada di colab).

Selanjutnya kita akan melakukan Training data adalah data yang digunakan untuk melatih (mengajar) model dalam proses pembelajaran mesin (machine learning). Pada umumnya, dataset dibagi menjadi dua bagian utama: training set (data latih) dan test set (data uji). Training set digunakan untuk melatih model, sedangkan test set digunakan untuk menguji performa model pada data yang belum pernah dilihat sebelumnya. (Tampilkan dan jelaskan kode tentang Training data dan berbagai macam caranya yang ada di colab).

Setelah training data dilakukan selanjutnya kita akan melakukan Evaluating data, dimana evaluating data itu adalah proses penilaian dan analisis terhadap dataset untuk memahami karakteristik, kualitas, dan struktur datanya. Evaluasi data bertujuan untuk mendapatkan wawasan

yang mendalam tentang informasi yang terkandung dalam dataset dan mempersiapkan data tersebut untuk tahap selanjutnya, seperti pemodelan atau analisis. (Tampilkan dan jelaskan kode tentang Evaluating data dan berbagai macam caranya yang ada di colab).

Langkah selanjutnya kita akan melakukan classification report untuk menampilkan laporan dari classification

Langkah terakhir yaitu Inputing New Data adalah proses memasukkan atau memberikan data ke dalam suatu sistem atau program. Ini adalah langkah awal yang penting dalam pemrosesan data, karena data yang baik dan benar akan mempengaruhi kualitas analisis atau model yang dibangun. (Tampilkan dan jelaskan kode tentang Inputing Data dan berbagai macam caranya yang ada di colab).