

1. Jelaskan yang dimaksud dengan *supervised learning* dan cakupannya!

Supervised learning adalah pendekatan *machine learning* yang menggunakan dataset berlabel (punya nilai y). Terdapat dua kategori untuk *supervised learning*, yaitu regresi dan klasifikasi.

2. Jelaskan cara kerja algoritma yang telah diimplementasikan!

Secara garis besar, algoritma-algoritma yang telah diimplementasikan adalah sebagai berikut.

- **K-Nearest Neighbours (KNN)**: Secara singkat, KNN **memprediksi label** dari sebuah rangkaian titik (atau vektor) baru dengan **melihat sebanyak K buah data terdekat yang sudah dilabeli**. Jarak dari titik baru dengan titik yang sudah dilabeli **diukur dengan jarak Euclidean**. Jika sudah mendapatkan **K data terdekat**, dihitung **mean** (untuk kategori **regresi**) atau **modus** (untuk kategori **klasifikasi**) dari label yang sudah ada untuk memberi label kepada data baru.
- **Logistic Regression (LR)**: Jika dianalogikan dalam *deep learning*, algoritma ini **mirip dengan pemelajaran neural network**. Jika *neural network* memiliki banyak layer dan neuron, algoritma ini hanya punya satu neuron, atau biasa disebut **perceptron**. Hanya saja, perceptron ini **harus menggunakan aktivasi sigmoid**. Optimizer, loss, dan metrics lain sebetulnya dapat disesuaikan sesuai dengan pembuat program. Maka dari itu, dalam algoritma ini **terdapat juga learning rate dan epochs** seperti dalam *deep learning*.
- **Decision Tree ID3**: Dalam algoritma ini terdapat pengukuran bernama *information gain* (IG) yang dihitung dari *entropy*. Langkah untuk mengekspan node ID3 ini adalah sebagai berikut.
 1. Hitung seluruh IG dari setiap fitur (kolom).
 2. Cari kolom dengan IG maksimal, pasang kolom tersebut sebagai “pembagi” pada node tersebut. Eliminasi kolom tersebut dari pemilihan pembagi pada simpul anak.
 3. Lakukan split data dengan label 0 dan 1 (untuk kasus binary tree) dari kolom yang dipilih pada langkah nomor 2.
 4. Untuk setiap data dan kolom yang tersisa, dibuat simpul-simpul anak. Simpul tersebut akan terus diekspan sampai kolom habis atau sampai IG maksimum di bawah toleransi (jika dispesifikasi) atau sampai kedalaman maksimum sudah tercapai (jika dispesifikasi).
 5. Pendefinisian batas IG maksimum atau kedalaman maksimum merupakan teknik pruning yang opsional untuk dilakukan.

3. Bandingkan ketiga algoritma tersebut, lalu tuliskan kelebihan dan kelemahannya!
 - LR dan ID3 memprediksi sebuah dataset dengan melakukan learning terlebih dahulu. Ada parameter yang bisa disimpan untuk penggunaan selanjutnya. Namun, KNN tidak bekerja demikian, maka KNN juga disebut juga *lazy learner*. KNN memprediksi sebuah data secara langsung (*on the spot learning*). Dengan demikian, penggunaan Logistic Regression atau ID3 secara komputasional lebih hemat daripada KNN (tidak ada parameter yang bisa disimpan).
 - LR dapat membobot signifikansi suatu fitur/kolom lebih baik daripada ID3 atau KNN.
 - KNN dan ID3 dapat mendeteksi pola non-linear lebih baik daripada LR.
4. Jelaskan penerapan dari algoritma *supervised* di berbagai bidang (misalnya industri atau kesehatan)!

Dalam bidang kesehatan, *supervised learning* dapat digunakan dalam pengklasifikasian gambar hasil radiologi untuk mendeteksi kanker atau penyakit paru-paru.