

1. Algoritma KNN adalah salah satu jenis algoritma machine learning sederhana yang bekerja dengan prinsip similarity based learning. Algoritma ini termasuk dalam lazy learner karena pada tahap training, algoritma tidak melakukan proses kompleks apapun, melainkan hanya menyimpan data training tersebut dan melakukan komputasi pada saat prediksi. Ketika ada data baru yang akan diprediksi, algoritma ini akan mencari sejumlah tetangga terdekat dari data training dan melakukan voting mayoritas untuk menentukan labelnya. Cara kerjanya adalah:
 - a. Inisialisasi parameter
 - i. k: Menentukan jumlah tetangga terdekat yang akan diperhitungkan (default: 5)
 - ii. metric: Metode penghitungan jarak (euclidean, manhattan, minkowski)
 - iii. p: Parameter khusus untuk metric minkowski
 - b. Fase training (Fit method)
 - i. Menyimpan seluruh data training (X_train dan y_train)
 - ii. Tidak ada proses learning atau optimization yang kompleks
 - iii. Hanya melakukan storage data untuk digunakan pada fase prediksi
 - c. Fase prediksi (Predict method), untuk setiap data training
 - i. Hitung jarak ke semua data training menggunakan metric yang ditentukan
 - ii. Urutkan jarak dari yang terkecil hingga terbesar
 - iii. Ambil k tetangga terdekat berdasarkan jarak terkecil
 - iv. Lakukan voting mayoritas pada label k tetangga terdekat
 - v. Tentukan prediksi berdasarkan label yang paling banyak muncul
 - d. Metode penghitungan jarak
 - i. Euclidean: Jarak garis lurus antara dua titik dalam space multidimensi
 - ii. Manhattan: Jarak yang dihitung seperti jalur grid kota (horizontal dan vertikal)
 - iii. Minkowski: Generalisasi dari kedua metode di atas dengan parameter p
2. Berdasarkan hasil evaluasi, model dari Sklearn memiliki nilai Recall 0.55 dan F1 score 0.57, sementara untuk model yang saya buat sendiri mendapatkan nilai Recall 0.57 dan F1 score 0.58. Hal ini menunjukkan performa model yang sudah sangat mirip.
3. Beberapa improvement yang dapat dilakukan
 - a. Hyperparameter tuning dengan menggunakan tools seperti Grid Search, Optuna, dll untuk mendapatkan kombinasi parameter terbaik untuk model pada dataset ini.

- b. Ensemble model yaitu menggabungkan beberapa model (termasuk KNN) untuk melakukan prediksi sehingga kita dapat memperoleh model yang lebih robust dan mendapat akurasi yang lebih baik.