

1. Gaussian Naive Bayes adalah algoritma probabilistik yang bekerja berdasarkan teorema bayes dengan asumsi independensi antar fitur. Model ini disebut naive karena membuat asumsi sederhana bahwa semua fitur saling independen given the class label, meskipun dalam praktiknya di dunia nyata asumsi ini jarang sepenuhnya benar. Cara kerja
  - a. Training ( Fit method)
    - i. Mengidentifikasi kelas unik dari data target
    - ii. Untuk setiap kelas, menghitung:
      1. Mean setiap fitur
      2. Variance setiap fitur
      3. Prior probability setiap kelas
  - b. Probability Density Function
    - i. Menggunakan distribusi Gaussian untuk menghitung likelihood
    - ii. Menambahkan epsilon kecil untuk menghindari division by zero
  - c. Prediction (Predict method)
    - i. Untuk setiap instance baru, menghitung posterior probability setiap kelas
    - ii. Menggunakan log probabilities untuk menghindari underflow numerik
    - iii. Memilih kelas dengan posterior probability tertinggi
2. Berdasarkan hasil evaluasi, model dari Sklearn memiliki nilai Recall 0.56 dan F1 score 0.57, sementara untuk model yang saya buat sendiri juga mendapatkan nilai Recall 0.56 dan F1 score 0.57. Hal ini menunjukkan performa model yang sama persis antar kedua model.
3. Beberapa improvement yang dapat dilakukan
  - a. Menambahkan Laplace smoothing untuk mengatasi masalah probabilitas nol. Laplace smoothing mengatasi ini dengan menambahkan nilai kecil (biasanya 1) ke setiap hitungan frekuensi fitur, sehingga tidak ada fitur yang memiliki probabilitas nol, dan model menjadi lebih robust.
  - b. Vectorized operation pada prediction sehingga perhitungan menjadi lebih cepat