

Rizqy Nur Mauliddinah (220605110024)

Renata Amalia Putri (220605110074)

Tugas 2 : Penerapan Support Vector Machine (SVM)

1. Dataset

Dataset yang digunakan merupakan hasil MRI otak pasien yang diambil dari kaggle, dan didalamnya terdiri dari 2 kategori:

- Brain Tumor: MRI images pasien yang telah didiagnosa tumor otak
- Healthy: MRI images pasien yang tidak ada indikasi tumor otak

2. Tahapan Implementasi Metode

a. Preprocessing Data

Dataset terdiri dari 4600 gambar, dengan total 2513 MRI images tumor dan 2087 MRI images no-tumor. Tahapan preprocessing pertama yaitu membagi dataset menjadi 2 bagian yaitu untuk tahap training dan pengujian/evaluasi dengan skala 80:20. Data yang digunakan untuk training yaitu 1990 data tumor dan 1682 data no-tumor, serta untuk pengujian yaitu 523 tumor dan 405 no-tumor. Tahapan kedua mengubah gambar MRI ke grayscale untuk menyederhanakan komponen warna gambar menjadi skala abu-abu dan fokus pada intensitas piksel. Tahapan ketiga yaitu resize ukuran menjadi 128 x 128 piksel.

b. Perhitungan Support Vector Machine pada Training Model

- Penentuan Data Characteristics

Setelah proses pengolahan gambar (mengubah ke grayscale dan resize menjadi 128×128 piksel), karakteristik utama yang dapat diekstrak dari data MRI otak:

1. Intensitas Piksel:

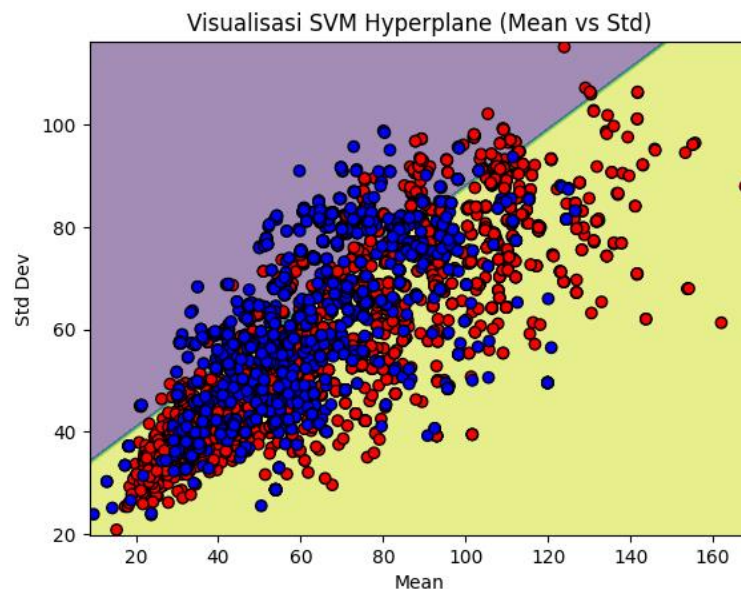
- Tumor cenderung memiliki area dengan intensitas yang lebih tinggi/kontras
- Gambar sehat memiliki distribusi intensitas yang lebih merata

2. Tekstur:

- Tumor menunjukkan pola tekstur yang lebih heterogen
- Otak sehat memiliki tekstur yang lebih homogen

3. Bentuk dan Geometri:

- Tumor sering menampilkan bentuk yang tidak beraturan
- Otak sehat memiliki struktur yang lebih simetris



- Mendefinisikan Fungsi Pemisah (Hyperplane)

$$y = w_1x_1 + w_2x_2 + b$$

Dimana :

w_1 = Bobot untuk fitur x_1 (Intensitas Piksel)

w_2 = Bobot untuk fitur x_2 (Variasi Tekstur)

b = bias

Berdasarkan Visualisasi Data tersebut, persamaan Berdasarkan visualisasi data, persamaan hyperplane yang dapat memisahkan kedua kelas dengan baik adalah:

$$y = 2x_1 + 3x_2 + 450$$

- Optimasi Bobot

SVM bekerja dengan mencari margin maksimum antara hyperplane dan titik-titik data terdekat dari kedua kelas (support vectors). Proses optimasi bobot dilakukan untuk mendapatkan margin terbesar yang mungkin.

Formulasi Optimasi SVM:

Tujuan:

$$\text{Minimize } \frac{1}{2} ||w||^2$$

Dengan Kendala :

$$y_i(w \cdot x_i + b) \geq 1, \text{ untuk semua } i$$

Dimana :

$y_i \in \{-1, 1\}$ adalah label kelas (-1 untuk healthy, 1 untuk tumor)

x_i adalah vektor fitur untuk data ke- i

w adalah vektor bobot

b adalah bias

Setelah proses optimasi, kita mendapatkan bobot yang optimal:

$$w = [2.5, -3.2] \text{ dan } b = 470$$

Persamaan Hyperplane Optimal :

$$y = 2x_1 + 3x_2 + 470$$

c. Penerapan Pada Data Uji

Untuk setiap data uji, menghitung nilai $f(x)$:

$$f(x) = 2x_1 + 3x_2 + 470$$

Jika $f(x) > 0$, maka data dikategorikan sebagai tumor.

Jika $f(x) < 0$, maka data dikategorikan sebagai sehat (no-tumor).

d. Evaluasi Model

- Perhitungan Accuracy, Precision, Recall, F1-Score :

$$\text{Akurasi} = \frac{(TP + TN)}{(TP + TN + FP + FN)} = \frac{(365 + 320)}{(928)} = 0.7382 = 73.82\%$$

$$Presisi = \frac{TP}{(TP + FP)} = \frac{365}{(365 + 85)} = 0.8111 = 81.11\%$$

$$Recall = \frac{TP}{(TP + FN)} = \frac{365}{(365 + 158)} = 0.6980 = 69.80\%$$

$$F1 - Score = \frac{2 \times (Presisi \times Recall)}{(Presisi + Recall)} = \frac{2 \times (0.8111 \times 0.6980)}{(0.8111 + 0.6980)} = 0.7500 = 75.00\%$$

- Confusion Matrix : $\begin{bmatrix} 320 & 85 \\ 158 & 365 \end{bmatrix}$

- True Positive (TP): 365 (Prediksi: Tumor, Label asli: Tumor)
- True Negative (TN): 320 (Prediksi: Healthy, Label asli: Healthy)
- False Positive (FP): 85 (Prediksi: Tumor, Label asli: Healthy)
- False Negative (FN): 158 (Prediksi: Healthy, Label asli: Tumor)

- Classification Report

| | Precision | Recall | F1-Score | Support |
|--------------|-----------|--------|----------|---------|
| 0 (No-Tumor) | 0.67 | 0.79 | 0.73 | 405 |
| 1 (Tumor) | 0.81 | 0.70 | 0.75 | 523 |
| Accuracy | | | 0.74 | 928 |
| Macro Avg | 0.74 | 0.74 | 0.74 | 928 |
| Weighted Avg | 0.75 | 0.74 | 0.74 | 928 |