

Nama/NIM : Rafli Limandijaya/1103210243

1. Feature Engineering, EDA, dan Visualisasi Data

Dataset: Dataset yang digunakan berisi 1482 data dengan 20 fitur, termasuk gejala klinis, hasil CT Scan, suhu tubuh, dan sebagainya.

a. Handling Missing Values

- Fitur kategorikal diisi dengan modus.
- Fitur numerik diisi dengan median.
- Fitur dengan missing terlalu banyak atau tidak relevan di-drop.

b. Encoding dan Normalisasi

- Fitur kategorikal diubah menjadi numerik menggunakan Label Encoding.
- Fitur numerik distandardisasi agar sebanding dalam pelatihan model.

c. Visualisasi Data

- Distribusi Diagnosis: Pie chart menunjukkan distribusi kelas H1N1 dan H3N2.
- Heatmap Korelasi: Untuk melihat hubungan antar fitur numerik terhadap diagnosis.
- Boxplot dan Countplot:
 - Usia terhadap diagnosis
 - Gejala (demam, batuk, dsb.) terhadap diagnosis

Hasil EDA menunjukkan beberapa fitur penting dalam membedakan diagnosis seperti:

- Fever (Demam)
- Shortness of Breath (Sesak Napas)
- Temperature (Suhu tubuh)
- Fatigue (Kelelahan)

2. Evaluasi Model Klasifikasi

Model klasifikasi menggunakan Random Forest Classifier.

a. Hasil Evaluasi:

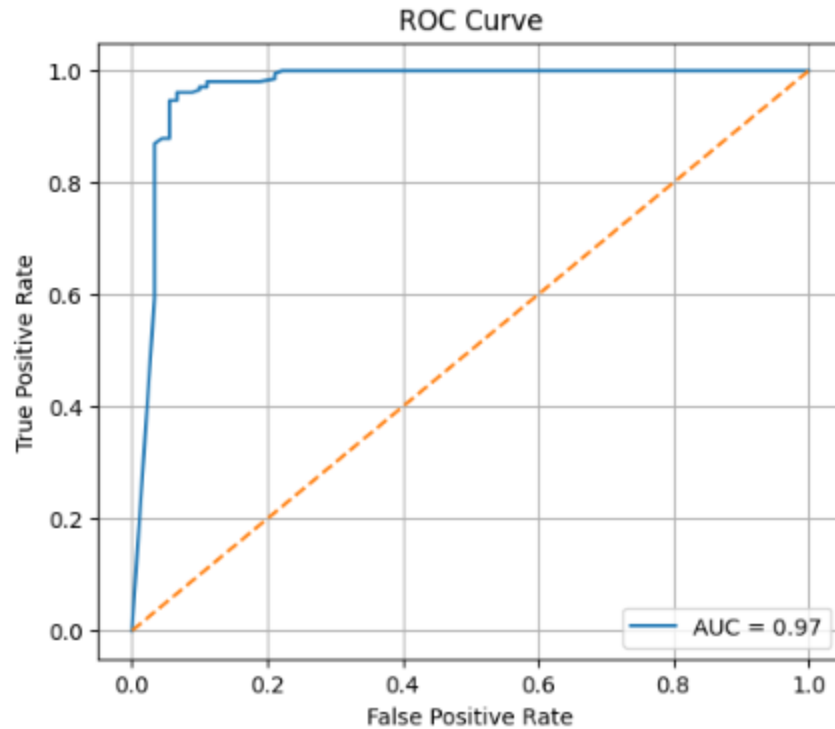
- **Accuracy:** 95%
- **Precision:** 90% (H1N1), 97% (H3N2)
- **Recall:** 93% (H1N1), 96% (H3N2)
- **F1-Score:** 92% (H1N1), 96% (H3N2)
- **AUC Score:** 0.97 (dari ROC Curve)

b. Confusion Matrix:

	Pred H1N1	Pred H3N2
True H1N1	84	6
True H3N2	9	198

c. Interpretasi

Model berhasil mengklasifikasikan dua tipe flu dengan akurasi sangat tinggi. ROC Curve menunjukkan pemisahan kelas yang sangat baik.



3. Penjelasan Matematika Metrik Evaluasi

a.
$$\text{Accuracy} = \frac{TP+TN}{TP+TN+FP+FN}$$

Mengukur seberapa banyak prediksi yang benar dari seluruh data.

B.
$$\text{Precision} = \frac{TP}{TP+FP}$$

Mengukur ketepatan model saat memprediksi kelas positif.

c.
$$\text{Recall} = \frac{TP}{TP+FN}$$

Mengukur seberapa banyak kasus positif yang berhasil terdeteksi.

d.
$$F1 = 2 \cdot \frac{\text{Precision} \cdot \text{Recall}}{\text{Precision} + \text{Recall}}$$

Gabungan harmonik dari Precision dan Recall.

e. AUC (Area Under Curve)

AUC adalah luas di bawah kurva ROC. Semakin mendekati 1, semakin baik performa model dalam membedakan kelas.