

Nama : Zahrani Cahya Priesa

NIM : 1103223074

Mata Kuliah : Machine Learning

Analisis Evaluasi Model Klasifikasi MNIST

I. Persiapan Data dan Klasifikasi Biner (Deteksi Angka '5')

- **Dataset:** Analisis menggunakan dataset **MNIST** yang berisi 70.000 gambar digit tulisan tangan (60.000 data latih, 10.000 data uji), di mana setiap gambar memiliki 784 fitur (piksel).
- **Tujuan Awal:** Mengembangkan pengklasifikasi biner (SGDClassifier) untuk mendeteksi secara spesifik apakah suatu gambar adalah angka '5' atau bukan.

II. Evaluasi Kinerja dan Jebakan Akurasi

Tingkat akurasi tinggi sebesar 95-96% oleh SGDClassifier terlihat menjanjikan, namun hasil ini menyesatkan. Akurasi model yang hanya memprediksi "bukan 5" (model bodoh) sudah mencapai ~91%.

- **Penyebab:** Dataset MNIST memiliki distribusi kelas yang miring (*skewed*) di mana hanya sekitar 10% dari data latih adalah angka '5'. Oleh karena itu, akurasi tinggi hanya mencerminkan kemampuan model untuk memprediksi mayoritas kelas negatif (bukan '5'), bukan kemampuan sebenarnya untuk mendeteksi kelas target '5'.

III. Matriks Kebingungan (*Confusion Matrix*) dan Metrik Esensial

Untuk mendapatkan gambaran kinerja yang sebenarnya, digunakan Matriks Kebingungan dan metrik turunan:

Kelas Aktual/Prediksi	Prediksi Bukan 5 (Negatif)	Prediksi Angka 5 (Positif)
Aktual Bukan 5	True Negative (TN): 53.892	False Positive (FP): 687
Aktual Angka 5	False Negative (FN): 1.891	True Positive (TP): 3.530

Dari Matriks Kebingungan, metrik utama dihitung:

Metrik	Formula	Nilai Kinerja	Arti
Presisi (<i>Precision</i>)	$\$TP / (TP + FP)\$$	0.8371 (83.71%)	Dari semua gambar yang <i>diprediksi</i> sebagai '5', 83.71% benar-benar '5'.
Perolehan (<i>Recall</i>)	$\$TP / (TP + FN)\$$	0.6512 (65.12%)	Dari semua gambar '5' yang sebenarnya ada, model hanya mendeteksi 65.12% darinya.
F1 Score	Rata-rata harmonik Precision dan Recall	0.7325	Keseimbangan antara Presisi dan Recall.

Kesimpulan Kinerja Biner: Model SGDClassifier memiliki Presisi yang cukup baik tetapi Recall yang relatif rendah. Ini mengindikasikan bahwa model cenderung menghasilkan jumlah *False Negatives* (melewatkan angka '5' yang seharusnya) yang lebih tinggi daripada *False Positives* (salah mengklasifikasikan digit lain sebagai '5').

IV. Klasifikasi Multikelas

- **Model yang Diuji:** SGDClassifier, SVC, dan RandomForestClassifier.
- **Strategi:** SVC menggunakan strategi *One-vs-One* (OvO), sedangkan SGDClassifier menggunakan strategi *One-vs-Rest* (OvR) secara *default*.
- **Peran Penskalaan:** Penskalaan data (StandardScaler) sangat penting dan meningkatkan Akurasi SGDClassifier secara signifikan, mencapai rata-rata sekitar **86.5%**.
- **Kinerja Lanjutan:**
 - RandomForestClassifier menunjukkan kinerja yang sangat kuat dengan ROC AUC Score 0.9983 untuk tugas *OvR*.
 - Matriks Kebingungan Multikelas menunjukkan bahwa kesalahan umum terjadi pada digit yang memiliki kemiripan visual, seperti salah mengklasifikasikan angka 5 sebagai 3, atau angka 7 sebagai 9.