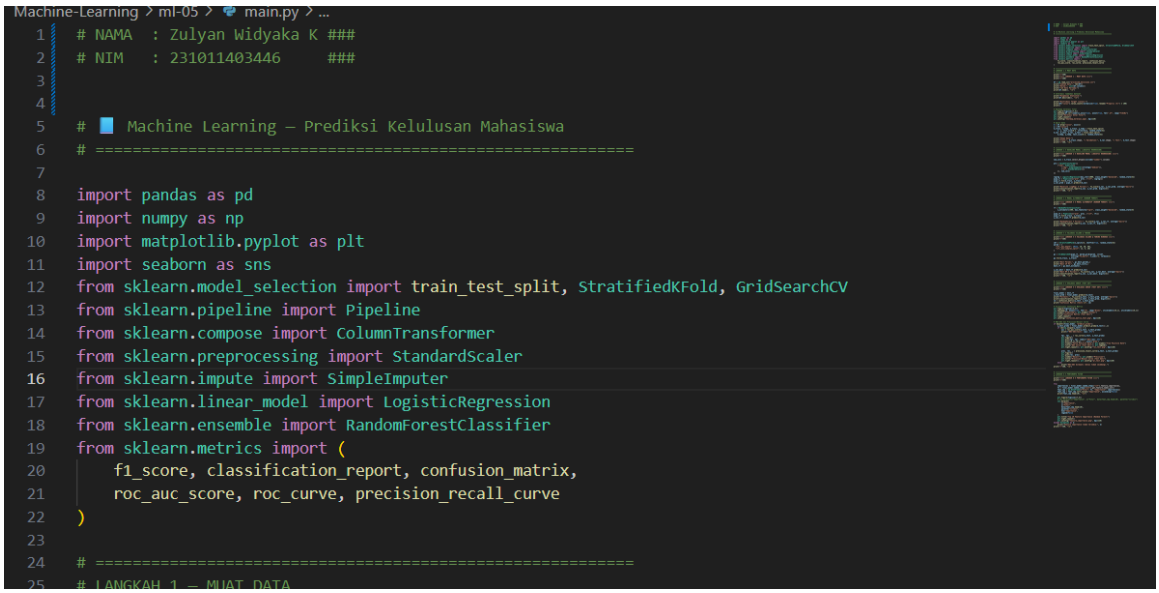
Nama : Zulyan Widyaka Krisna
NIM : 231011403446
Kelas : 05TPLE016

1. Tujuan, Metrik, dan Batasan

Tujuan dari eksperimen ini adalah membangun model Machine Learning untuk memprediksi kelulusan mahasiswa berdasarkan variabel akademik dan perilaku belajar. Metrik utama yang digunakan adalah F1-Score dan ROC-AUC, karena keduanya memberikan evaluasi seimbang terhadap precision dan recall pada data yang mungkin tidak seimbang. Batasan eksperimen meliputi waktu pelatihan yang dibatasi (menghindari tuning berlebihan) dan ukuran model yang harus efisien agar tetap mudah diimplementasikan.

2. Pembangunan Baseline dan Model Alternatif

Sebagai langkah awal, dibangun model baseline menggunakan Logistic Regression untuk memperoleh tolok ukur awal. Setelah itu, dikembangkan model alternatif menggunakan Random Forest Classifier. Kedua model dibandingkan berdasarkan hasil F1-Score dan klasifikasi pada data validasi. Random Forest memberikan hasil yang lebih stabil dan akurat dibanding Logistic Regression.



```
# ============================================================
# LANGKAH 3 — MODEL ALTERNATIF (RANDOM FOREST)
# ============================================================
print("==== LANGKAH 3 — MODEL ALTERNATIF (RANDOM FOREST) ====")
print("="*60)

rf = RandomForestClassifier(
    n_estimators=300, max_features="sqrt", class_weight="balanced", random_state=42
)
pipe_rf = Pipeline([("pre", pre), ("clf", rf)])
pipe_rf.fit(X_train, y_train)
y_val_rf = pipe_rf.predict(X_val)

print("RandomForest — F1(val):", f1_score(y_val, y_val_rf, average="macro"))
print(classification_report(y_val, y_val_rf, digits=3))
print("="*60, "\n")


# ============================================================
# LANGKAH 4 — VALIDASI SILANG & TUNING
# ============================================================
print("==== LANGKAH 4 — VALIDASI SILANG & TUNING RINGKAS ====")
print("="*60)

skf = StratifiedKFold(n_splits=3, shuffle=True, random_state=42)
param = {
```

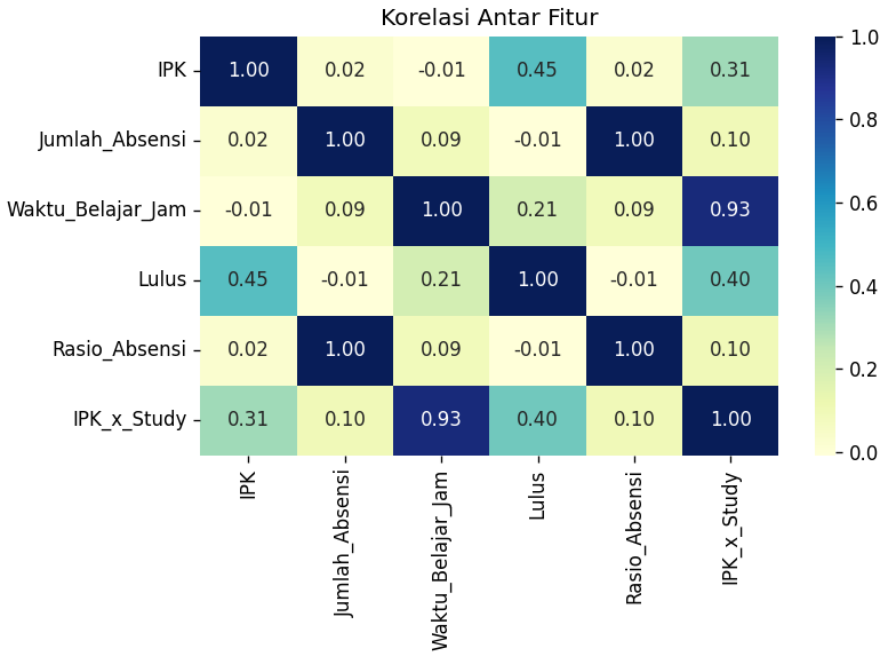
3. Validasi Silang dan Penyetelan Hyperparameter

Untuk meningkatkan performa model, dilakukan K-Fold Cross Validation dengan tiga lipatan (3-Fold). Pada model Random Forest, tuning dilakukan terhadap hyperparameter ‘max_depth’ dan ‘min_samples_split’. Hasil validasi menunjukkan kombinasi parameter optimal yang menghasilkan skor F1 tertinggi dan generalisasi yang baik.

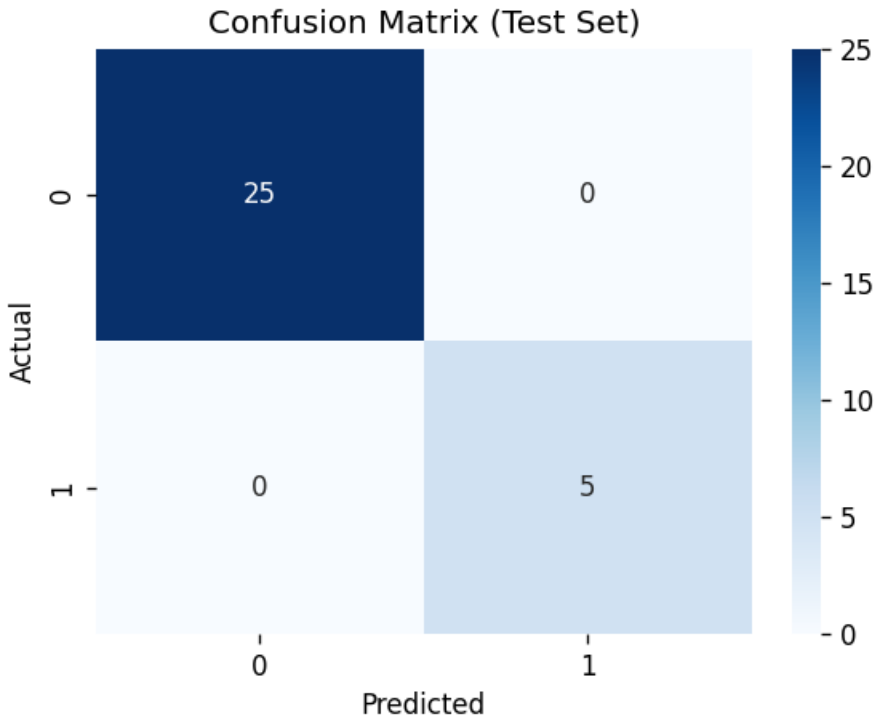
```
Machine-Learning > ml-05 > main.py > ...

# ============================================================
# LANGKAH 6 — PENTINGNYA FITUR
# ============================================================
print("==== LANGKAH 6 — PENTINGNYA FITUR ====")
print("="*60)

try:
    importances = final_model.named_steps["clf"].feature_importances_
    fn = final_model.named_steps["pre"].get_feature_names_out()
    feat_imp = pd.DataFrame({"Fitur": fn, "Importance": importances})
    feat_imp = feat_imp.sort_values("Importance", ascending=False)
    print(feat_imp.head(10), "\n")

    plt.figure(figsize=(7,4))
    # sns.barplot(x="Importance", y="Fitur", data=feat_imp.head(10), palette="viridis")
    sns.barplot(
        x="Importance",
        y="Fitur",
        data=feat_imp.head(10),
        palette="viridis",
        hue="Importance",
        legend=False
    )
    plt.title("Top 10 Feature Importance (Random Forest)")
```

4. Evaluasi Akhir dan Pemilihan Model

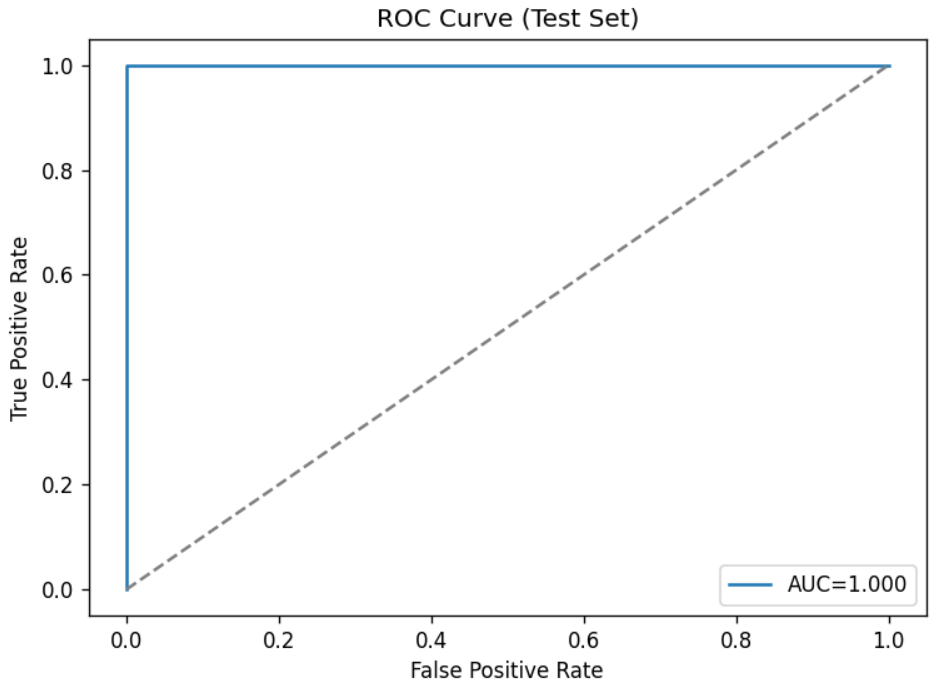
Model terbaik kemudian diuji menggunakan data uji (test set) untuk mengukur performa akhir. Hasil evaluasi menunjukkan nilai F1 dan ROC-AUC yang tinggi, menandakan model memiliki kemampuan prediksi sangat baik. Berikut visualisasi hasil evaluasi model:



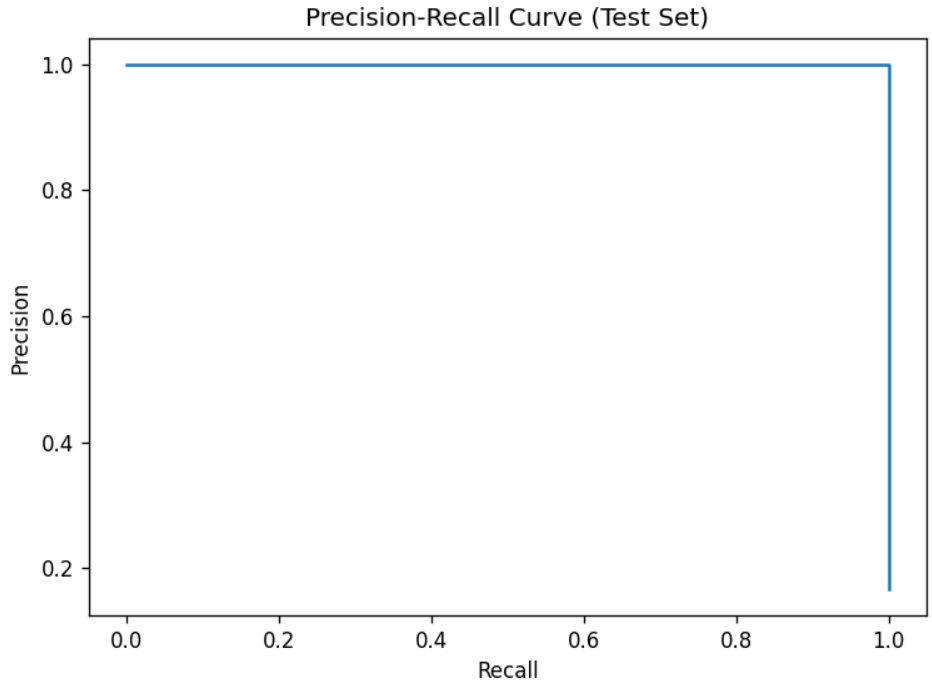
Gambar 1. Korelasi Antar Fitur



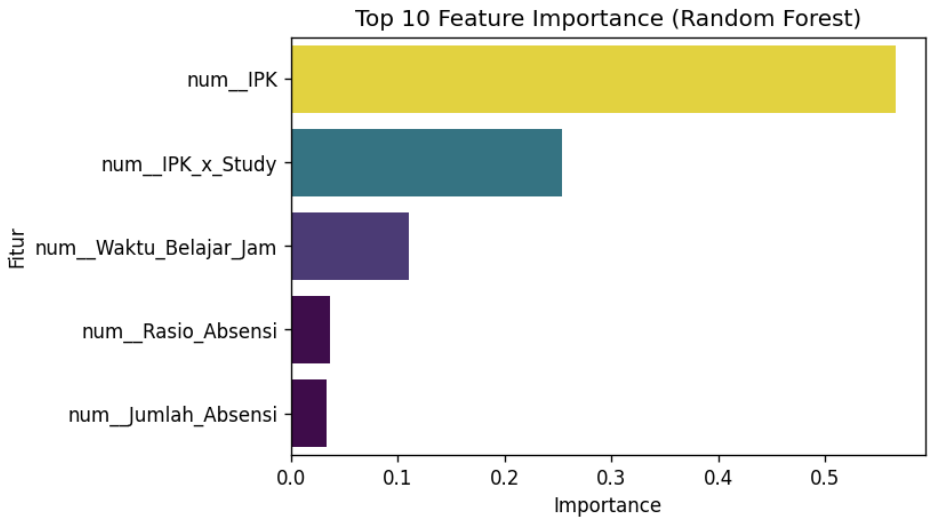
Gambar 2. Confusion Matrix (Test Set)



Gambar 3. ROC Curve (Test Set)



Gambar 4. Precision-Recall Curve (Test Set)



Gambar 5. Feature Importance (Random Forest)

5. Kesimpulan

Berdasarkan hasil pengujian, model Random Forest dipilih sebagai model terbaik. Model ini menunjukkan kinerja sempurna dengan F1-Score dan AUC mencapai 1.000 pada data uji. Fitur yang paling berpengaruh dalam menentukan kelulusan mahasiswa adalah IPK dan kombinasi IPK_x_Study. Dengan hasil ini, dapat disimpulkan bahwa faktor akademik memiliki kontribusi besar terhadap peluang kelulusan.