

Tugas Praktikum: Classification dengan Ensemble Method & Class Imbalance

Deskripsi Tugas

Pada praktikum ini, mahasiswa diminta untuk membangun model klasifikasi menggunakan ensemble method pada dataset yang memiliki distribusi kelas tidak seimbang. Praktikum ini bertujuan untuk menerapkan teknik ensemble method seperti Random Forest atau Boosting, serta pendekatan untuk menangani class imbalance seperti oversampling atau cost-sensitive learning.

Tujuan Praktikum

1. Memahami penerapan ensemble method dalam klasifikasi data.
2. Mengatasi masalah class imbalance pada dataset.
3. Mengevaluasi performa model menggunakan metrik yang sesuai seperti F1-score, recall, dan AUC-ROC.

Langkah-Langkah Praktikum

1. **Unduh Dataset**

Gunakan dataset Credit Card Fraud Detection dari Kaggle. Dataset ini berisi data transaksi kartu kredit dengan label `Class`, di mana 0 menunjukkan transaksi normal dan 1 menunjukkan transaksi penipuan (fraud).

Link dataset: <https://www.kaggle.com/mlg-ulb/creditcardfraud>

2. **Praproses Data**

- Identifikasi distribusi kelas pada dataset (`Class`).
- Terapkan metode oversampling (seperti SMOTE) atau undersampling untuk menangani class imbalance.
- Bagi dataset menjadi data latih dan data uji dengan perbandingan 80:20.

3. **Membangun Model Klasifikasi**

- Gunakan ensemble method seperti Random Forest atau Gradient Boosting.
- Bandingkan model dengan atau tanpa penanganan class imbalance.

4. **Evaluasi Model**

- Hitung metrik evaluasi seperti precision, recall, F1-score, dan AUC-ROC untuk menilai performa model pada kelas minoritas (fraud).
- Gunakan confusion matrix untuk memahami distribusi prediksi.

5. **Laporan Praktikum**

- Buat laporan singkat (2-3 halaman) yang mencakup:
- Hasil distribusi kelas awal.
- Pendekatan yang digunakan untuk menangani class imbalance.
- Performansi model dengan metrik evaluasi.

Hasil yang Diharapkan

1. Model klasifikasi yang dapat mengatasi masalah class imbalance.
2. Laporan yang berisi analisis distribusi kelas, pendekatan yang digunakan, dan hasil evaluasi model.

Penilaian

1. ****Praproses Data (20%)****: Kemampuan mengidentifikasi dan menangani class imbalance.
2. ****Penerapan Ensemble Method (40%)****: Implementasi model Random Forest atau Boosting dengan tepat.
3. ****Evaluasi Model (30%)****: Penggunaan metrik evaluasi yang sesuai dan interpretasi hasil.
4. ****Laporan Praktikum (10%)****: Kerapihan, kelengkapan, dan analisis hasil.