

Final Project: Data Mining

Deskripsi Tugas

Mahasiswa diminta untuk mengimplementasikan berbagai metode data mining yang telah dipelajari selama perkuliahan, termasuk eksplorasi data, praproses data, klasifikasi, clustering, dan deteksi anomali. Proyek ini bertujuan untuk memberikan pengalaman menyeluruh dalam menerapkan teknik data mining pada dataset dunia nyata.

Tujuan Proyek

1. Memahami dan menerapkan alur lengkap proses data mining.
2. Menggunakan metode eksplorasi, praproses, klasifikasi, clustering, dan deteksi anomali.
3. Menganalisis hasil implementasi dan menyusun laporan yang terstruktur.

Langkah-Langkah Proyek

1. ****Unduh Dataset****

Gunakan dataset Kaggle - Online Retail Dataset. Dataset ini berisi data transaksi retail yang mencakup informasi tentang produk, harga, jumlah, dan tanggal transaksi.

Link dataset: <https://www.kaggle.com/datasets/mashlyn/online-retail-dataset>

2. ****Eksplorasi Data****

- Analisis struktur dataset dan identifikasi atribut penting.
- Identifikasi nilai hilang, data duplikat, atau anomali.

3. ****Praproses Data****

- Bersihkan dataset dengan menghapus nilai kosong dan duplikasi.
- Normalisasi data numerik jika diperlukan.
- Buat data transaksi berbasis pelanggan untuk analisis lebih lanjut.

4. ****Implementasi Teknik Data Mining****

- ****Eksplorasi Data****: Analisis distribusi data (contoh: histogram, boxplot).
- ****Klasifikasi****: Buat model untuk memprediksi kategori produk berdasarkan atribut lainnya menggunakan algoritma Random Forest atau Gradient Boosting.
- ****Clustering****: Segmentasi pelanggan berdasarkan pola pembelian menggunakan algoritma K-Means atau DBSCAN.
- ****Deteksi Anomali****: Identifikasi transaksi yang mencurigakan menggunakan Isolation Forest atau Local Outlier Factor (LOF).

5. ****Visualisasi dan Analisis Hasil****

- Buat visualisasi hasil untuk setiap teknik yang digunakan.
- Interpretasikan hasil dan berikan wawasan bisnis yang dapat diperoleh dari data.

6. ****Laporan Final Proyek****

- Sertakan langkah-langkah, hasil visualisasi, dan analisis dari setiap metode.
- Buat kesimpulan dan rekomendasi berdasarkan hasil analisis.

Hasil yang Diharapkan

1. Implementasi lengkap metode eksplorasi data, klasifikasi, clustering, dan deteksi anomali.
2. Visualisasi hasil dari setiap metode.
3. Laporan proyek yang mencakup seluruh langkah dan analisis.

Penilaian

1. ****Eksplorasi dan Praproses Data (20%)****: Pemahaman dan kemampuan mempersiapkan dataset.
2. ****Implementasi Teknik Data Mining (50%)****: Kesesuaian dan kelengkapan metode yang diterapkan.
3. ****Analisis Hasil (20%)****: Interpretasi hasil yang sesuai dan relevan.
4. ****Laporan Proyek (10%)****: Kerapihan, kelengkapan, dan analisis dalam laporan.