



Ujian Akhir Semester Gasal Tahun Akademik 2022/2023

MATA UJIAN	: Pembelajaran Mesin dan Pembelajaran Mendalam
PRODI	: Informatika
HARI, TANGGAL	: Jumat, 16 Desember 2022
WAKTU	: -
SIFAT UJIAN	: Kumpul Tugas Daring
DOSEN PENGUJI	: Aloysius Gonzaga Pradnya Sidhawara, S.T., M.Eng.

Proyek UAS PMDPM

1. Proyek dibuat secara berkelompok **dua orang** dan menggunakan file notebook Jupyter yang sudah disediakan di LMS.
2. Gunakan dataset untuk kasus klasifikasi yang sudah disediakan di LMS.
3. Buatlah proyek modelling algoritme pembelajaran mesin untuk kasus klasifikasi dengan ketentuan proses sebagai berikut:
 - a. Inisialisasi: import library yang dibutuhkan
 - b. Data loading: proses memuat data ke notebook
 - c. Data cleansing: proses pembersihan data (handling missing values, jika ada)
 - d. Data visualization: proses visualisasi data
 - e. Feature engineering: proses rekayasa fitur seperti categorical data encoding (jika ada data kategorik string), label encoding (jika label masih dalam bentuk string).
 - f. Modelling: proses pembuatan model algoritme pembelajaran mesin menggunakan Pipeline dan grid search cross-validation. Tahap di dalam Pipeline termasuk data scaling dan feature selection.
 - g. Evaluasi model: proses evaluasi model terbaik menggunakan classification report dan skor AUC.
4. Ketentuan detail proyek:
 - a. Random_state untuk semua objek yang menggunakan random_state adalah **0** untuk kelompok yang anggota pertamanya NPM ganjil dan **42** untuk kelompok yang anggota pertamanya NPM genap.
 - b. Dalam proses visualisasi data, silahkan pilih 5 feature yang menurut Anda perlu divisualisasikan agar mendapat insight dari data. Berikan keterangan atau deskripsi singkat mengenai informasi yang diperoleh dari visualisasi data tersebut.
 - c. Gunakan Stratified K-fold Cross Validation dengan nilai k=5.
 - d. Pipeline terdiri atas data scaling, feature selection, dan classifier.
 - e. Step data scaling untuk membandingkan perlunya data scaling atau tidak. Gunakan tiga nilai pada parameter grid yaitu: [None,Scaler1,Scaler2] Scaler yang digunakan antara lain MinMax, Standard, atau Robust.
 - f. Step feature selection untuk membandingkan metode feature selection yang digunakan. Gunakan tiga metode pada parameter grid yaitu: [None,Selector1,Selector2]. Metode yang digunakan harus statistik univariat (antara SelectKBest atau SelectPercentile) dan SelectFromModel (boleh menggunakan tree-based model atau linear model). Kalian juga perlu melakukan setting parameter grid nilai k atau percentile-nya dan setting model untuk SelectFromModel.
 - g. Step classifier untuk membandingkan algoritme classifier. Gunakan 4 algoritme pembelajaran mesin yaitu Random Forest, Gradient Boosting Classifier, Logistic Regression, dan SVM. Eksplorasi pengaturan parameter yang ada pada algoritme yang akan Anda gunakan, tidak terbatas pada 1-2 parameter saja supaya optimal (selain



Ujian Akhir Semester Gasal Tahun Akademik 2022/2023

MATA UJIAN	: Pembelajaran Mesin dan Pembelajaran Mendalam
PRODI	: Informatika
HARI, TANGGAL	: Jumat, 16 Desember 2022
WAKTU	: -
SIFAT UJIAN	: Kumpul Tugas Daring
DOSEN PENGUJI	: Aloysius Gonzaga Pradnya Sidhawara, S.T., M.Eng.

- random_state). Silahkan bereksperimen untuk parameter seperti `n_estimators`, `max_depth` untuk tree-based classifier; `C` untuk Logistic Regression; dan `C`, `gamma`, serta kernel untuk SVM agar mendapat akurasi terbaik.
- h. Tampilkan best estimator dan nilai best score dari grid-search cross-validation. Fit best estimator pada train set kemudian ujian pada test set dengan fungsi `score`. Format score 3 digit di belakang koma (.3f).
- i. Tampilkan classification report dan AUC dari best estimator yang diperoleh!
- j. Contoh ketentuan penulisan variabel (`aaaaa=nickname mahasiswa 1`, `bbbbbb=nickname mahasiswa 2`):
- Nama dataframe (sesuai dataset): `df_housing_aaaaa_bbbbb`
 - Nama feature yang sudah dilengkapi/dibersihkan (sesuai nama feature yang dibersihkan): `namafeature_aaaaa_bbbbb`
 - Nama grid search cross-validation yang digunakan: `GSCV_aaaaa_bbbbb`
5. Buatlah laporan hasil pengerjaan proyek individu pada cell markdown di notebook yang sudah disediakan. Jelaskan secara singkat tentang:
- dataset yang digunakan beserta deskripsi feature dan targetnya
 - proses data cleansing yang dilakukan
 - proses data visualization yang dilakukan
 - proses feature engineering yang dilakukan
 - proses modelling yang dilakukan
 - model terbaik
 - hasil evaluasi model
6. Jika ada yang terindikasi **mencontek/copy paste** dari internet ataupun orang lain, **nilai UAS=0**.