



---

## UJIAN TENGAH SEMESTER GASAL T.A. 2024/2025

MATA KULIAH : Pembelajaran Mesin dan Pembelajaran Mendalam  
PRODI : Informatika  
HARI, TANGGAL : Selasa, 22 Oktober 2024  
WAKTU : -  
SIFAT UJIAN : Kumpul Tugas Daring  
DOSEN PENGUJI : Aloysius Gonzaga Pradnya Sidhawara, S.T., M.Eng.  
Yohanes Sigit Purnomo, S.T., M.Kom., Ph.D.

---

### Projek UTS PMDPM – Klasifikasi dan Regresi Data Numerik

#### CP Mata Kuliah

(CPMK1) Mahasiswa mampu menerapkan dan mengkombinasikan ilmu pengetahuan domain, informatika, dan/atau prinsip rekayasa untuk menyelesaikan masalah di bidang pembelajaran mesin dengan benar dan terukur (CPL1).

#### Sub-CPMK

(Sub-CPMK1) Mahasiswa mampu mengkombinasikan ilmu pengetahuan domain, informatika, dan/atau prinsip rekayasa untuk merancang solusi permasalahan klasifikasi atau regresi di bidang pembelajaran mesin dengan benar dan terukur.

#### Penjelasan Tugas

1. Tugas dikerjakan secara berkelompok. Satu kelompok terdiri atas 4-5 orang.
2. Soal studi kasus dan dataset sudah disediakan.
3. Mahasiswa mendiskusikan soal studi kasus dan membuat solusi dalam bentuk code dengan format Notebook (.ipynb), model (.pkl) dan antarmuka Streamlit Python (.py). Tools menggunakan VS Code/Jupyter Notebook/Google Collaboratory.

#### Studi Kasus

Perusahaan PT. Ayodya Property yang sudah lama bergerak di bidang properti melihat pemanfaatan teknologi informasi menjadi potensi besar dalam upaya mengembangkan bisnis. PT. Ayodya Property meminta bantuan tim anda untuk mengembangkan sebuah sistem berbasis model pembelajaran mesin yang dapat memprediksi harga properti berdasarkan data-data yang diperoleh di lapangan.

Untuk mengembangkan sistem tersebut, tim anda mengumpulkan data dari klien yang ingin menjual rumah kepada PT. Ayodya Property. Data yang dikumpulkan meliputi informasi luas tanah dalam meter persegi (squaremeters), jumlah kamar (numberofrooms), adanya halaman (hasyard), adanya kolam renang (haspool), jumlah lantai (floors), kode lokasi (citycode), eksklusivitas kawasan (citypartrange), jumlah pemilik sebelumnya (numprevowners), tahun pembuatan (made), gedung baru atau tidak (isnewbuilt), apakah memiliki pelindung badai (hasstormprotector), luas basement (basement), luas loteng (attic), luas garasi (garage), memiliki gudang atau tidak (hasstorageroom), memiliki ruang tamu atau tidak (hasguestroom), harga (price), dan kategori property (category). Secara umum kategori properti dibagi menjadi Basic, Middle, dan Luxury. Setelah mengumpulkan data yang diperlukan, tim anda menetapkan SOP untuk melatih dan menguji model pembelajaran mesin tersebut.

### Instruksi

Terdapat dua proyek yang harus dikerjakan yaitu klasifikasi kategori properti dan regresi harga properti. Instruksi dari pimpinan tim yang kalian terima adalah sebagai berikut:

1. Buatlah perbandingan kinerja model pembelajaran mesin untuk **kasus Klasifikasi Kategori Properti dan Regresi Harga Properti**.
2. Sebelum memulai pembuatan model, pastikan kalian memahami dataset yang akan digunakan. **Periksa apakah terdapat missing value, data duplikat, dan ubah data kategorik string menjadi numerik**. Jika jumlah kelas pada data latih tidak seimbang, pastikan metode cross-validation yang digunakan adalah **Stratified k-fold**.
3. Tentukan fitur dan target: tentukan fitur dan target yang akan digunakan untuk melatih model. Untuk **klasifikasi**, pastikan **Kategori menjadi target** dan **kolom Harga dihapus**. Untuk **regresi**, pastikan **Harga menjadi target** dan **kolom Kategori dihapus**.
4. Untuk nilai parameter **random\_state**, sesuaikan dengan **dua (2) digit terakhir nomor pegawai terbesar (red: dua digit terakhir NPM TERBESAR)**. Silahkan memodifikasi persentase train-test split terbaik antara 80:20, 75:25, atau 70:30.
5. **Tugas Klasifikasi**: buatlah **DUA (2) NOTEBOOK** untuk **perbandingan SATU (1) algoritme berbasis tree dan SATU (1) algoritme berbasis linear**. Notebook pertama membandingkan **Random Forest dan Logistic Regression**, Notebook kedua membandingkan **Gradient Boosting Classifier dan Support Vector Machine**. Tahap pemodelan dimulai dari **data scaling, feature selection**, hingga **algoritme classifier**. Bandingkan **DUA (2) metode penskalaan** yaitu **StandardScaler dan MinMaxScaler** untuk **model berbasis linear**. Kalian perlu bereksperimen dengan membandingkan **DUA (2) metode feature selection (SelectKBest & SelectPercentile)** dan jumlah feature yang dipilih menggunakan parameter grid. Kalian wajib menyesuaikan parameter dari algoritme classifier agar model dapat bekerja dengan baik pada dataset.
6. **Tugas Regresi**: buatlah **DUA (2) NOTEBOOK** untuk **perbandingan SATU (1) algoritme berbasis linear regression dan SATU (1) algoritme regressor berbasis model**. Notebook pertama

membandingkan **Ridge Regression** dan **Support Vector Regressor**, Notebook kedua membandingkan **Lasso Regression** dan **Random Forest Regressor**. Tahap pemodelan terdiri dari **data scaling, feature selection, dan algoritme regressor**. Bandingkan **DUA (2) metode penskalaan** yaitu **StandardScaler** dan **MinMaxScaler** untuk **semua model**. Kalian perlu bereksperimen dengan membandingkan **DUA (2) metode feature selection (SelectKBest & SelectPercentile)** dan jumlah feature yang dipilih menggunakan parameter grid. Kalian wajib menyesuaikan parameter dari algoritme regressor agar model dapat bekerja dengan baik pada dataset.

7. Setelah membuat Pipeline (**DUA (2) Pipeline untuk masing-masing algoritme dalam satu Notebook**) untuk tugas klasifikasi dan regresi, lakukan evaluasi untuk mengetahui performa masing-masing model. Evaluasi dilakukan dengan membuat masing-masing **DUA (2) Grid Search Cross Validation** dengan metode Stratified KFold Cross Validation. Untuk **klasifikasi**, tampilkan **feature yang relevan** dipilih oleh model, hasil pengukuran kinerja model klasifikasi dengan **confusion matrix**, serta metrik **accuracy, precision, recall, dan F1-score menggunakan Classification Report**. Untuk memudahkan pimpinan divisi membaca hasil, buatlah **confusion matrix dalam bentuk representasi visual** menggunakan fungsi ConfusionMatrixDisplay dari library scikit-learn.
8. Untuk **regresi**, tampilkan hasil pengukuran kinerja model regresi dengan **Mean Absolute Error, Mean Squared Error, dan Root Mean Squared Error**. Untuk memudahkan pimpinan divisi membaca hasil, buatlah **dataframe yang menampilkan perbandingan harga asli properti, harga hasil prediksi dua model regresi, selisih prediksi dengan harga asli, beserta grafik visualnya**.
9. Berdasarkan nilai metrik evaluasi terbaik pada klasifikasi dan regresi dan jumlah fitur terbaik, tentukan **SATU (1) model terbaik untuk KLASIFIKASI dan SATU (1) model terbaik untuk REGRESI**. Ekspor model menjadi file Pickle untuk digunakan pada aplikasi Streamlit.
10. Buatlah **SATU (1) aplikasi Streamlit** dengan pilihan menu **Prediksi Kategori Properti dan Prediksi Harga Properti** menggunakan **model terbaik yang sudah kalian peroleh**. Template aplikasi akan diberikan. **Pastikan jenis input pada tampilan aplikasi sesuai dengna jenis fitur**.
11. **Deploy aplikasi menggunakan Streamlit Cloud dan akun GitHub**.
12. Semoga instruksi ini dapat membantu tim baru dalam membuat perbandingan kinerja model pembelajaran mesin untuk kasus Klasifikasi dan Regresi. Jika ada pertanyaan jangan ragu menghubungi saya.

**Aspek Penilaian:**

- Projek code: Kualitas code proyek Klasifikasi dan Regresi data numerik yang dibuat mahasiswa.

**Nilai code dihitung dengan rubrik:**

<b>Kriteria Penilaian</b>	<b>Deskripsi Kriteria Penilaian</b>	<b>Nilai</b>
Data Loading	Data loading meliputi proses (1) membaca data dari file eksternal, (2) menampilkan contoh isi data, (3) menampilkan informasi mengenai masing-masing kolom (info), dan (4) menampilkan deskripsi terkait gambaran isi data (describe)	95
	Data loading meliputi 3 dari 4 aspek yang disebutkan di kriteria sebelumnya	75
	Data loading meliputi 2 dari 4 aspek yang disebutkan di kriteria sebelumnya	55
	Data loading meliputi 1 dari 4 aspek yang disebutkan di kriteria sebelumnya	40
Data Checking	Data checking meliputi proses pengecekan apakah ada data yang (1) null, (2) kosong, (3) NaN, maupun (4) outlier.	95
	Data checking meliputi 3 dari 4 aspek yang disebutkan di kriteria sebelumnya.	75
	Data checking meliputi 2 dari 4 aspek yang disebutkan di kriteria sebelumnya.	55
	Data checking meliputi 1 dari 4 aspek yang disebutkan di kriteria sebelumnya.	40
Data Preparation	Data preparation meliputi proses penanganan data yang null (1), kosong (2), NaN (3), maupun outlier (4).	95
	Data preparation meliputi proses penanganan 3 dari 4 aspek yang disebutkan di kriteria sebelumnya.	75
	Data preparation meliputi proses penanganan 2 dari 4 aspek yang disebutkan di kriteria sebelumnya.	55
	Data preparation meliputi proses penanganan 1 dari 4 aspek yang disebutkan di kriteria sebelumnya.	40
Data Encoding	Data encoding meliputi proses (1) pemilihan data yang akan di-encoding, (2) pemilihan metode encoding, (3) proses encoding, dan (4) menampilkan contoh data hasil encoding.	95

	Data encoding meliputi 3 dari 4 proses yang disebutkan di kriteria sebelumnya.	75
	Data encoding meliputi 2 dari 4 proses yang disebutkan di kriteria sebelumnya.	55
	Data encoding meliputi 1 dari 4 proses yang disebutkan di kriteria sebelumnya.	40
Model Pipeline	Model pipeline meliputi proses pembuatan pipeline untuk fungsi penskalaan data (1), pemilihan fitur (2), pemilihan metode regresi/klasifikasi (3), dan penentuan parameter untuk setiap tahapan pipeline (4).	95
	Model pipeline meliputi 3 dari 4 proses yang disebutkan di kriteria sebelumnya.	75
	Model pipeline meliputi 2 dari 4 proses yang disebutkan di kriteria sebelumnya.	55
	Model pipeline meliputi 1 dari 4 proses yang disebutkan di kriteria sebelumnya.	40
Model Training and Evaluation	Model training and evaluation meliputi proses training model menggunakan Grid Search dan Cross validation (1), pengecekan model terbaik (2), pengecekan fitur yang signifikan (3), dan visualisasi hasil evaluasi (confusion matrix untuk klasifikasi dan grafik prediksi untuk regresi) (4)	95
	Model training and evaluation meliputi meliputi 3 dari 4 proses yang disebutkan di kriteria sebelumnya.	75
	Model training and evaluation meliputi meliputi 2 dari 4 proses yang disebutkan di kriteria sebelumnya.	55
	Model training and evaluation meliputi meliputi 1 dari 4 proses yang disebutkan di kriteria sebelumnya.	40
Model Deployment	Model deployment meliputi proses dump model terbaik (1), pembuatan antarmuka aplikasi dengan streamlit dan berjalan di lokal (2), publikasi ke github (3), dan deployment aplikasi di streamlit cloud (4)	95
	Model deployment meliputi meliputi 3 dari 4 proses yang disebutkan di kriteria sebelumnya.	75
	Model deployment meliputi meliputi 2 dari 4 proses yang disebutkan di kriteria sebelumnya.	55

	Model deployment meliputi meliputi 1 dari 4 proses yang disebutkan di kriteria sebelumnya.	40
<b>Aspek Penilaian</b>	<b>Bobot (%)</b>	<b>Skala Penilaian</b>
Data Loading	5	Mengikuti rubrik code
Data Checking	5	
Data Preparation	10	
Data Encoding	10	
Model Pipeline	25	
Model Training and Evaluation	30	
Model Deployment	15	
TOTAL	100	

### Aturan pengumpulan proyek

- Batas akhir pengumpulan proyek adalah hari Jumat, 25 Oktober 2024 pukul 23.59 WIB
- File yang dikumpulkan meliputi:
  - Proyek UTS dalam bentuk **EMPAT (4) Notebook Python (.ipynb)**, **SATU (1) model terbaik untuk klasifikasi (.pkl)**, **SATU (1) model terbaik untuk regresi (.pkl)**, **SATU (1) file Python aplikasi Streamlit (.py)**, **SATU (1) file PDF berisi semua code yang dikerjakan (.pdf)**, dan **file requirements (.txt)** yang diunggah ke repository GitHub. **Pastikan GitHub dapat diakses secara PUBLIK.**
  - Format penamaan Notebook: **Notebook\_KLASIFIKASI/REGRESI\_Kelas\_NamaSB\_ALG1\_VS\_ALG2\_NamaPanggilan.ipynb** di mana Notebook dibedakan berdasarkan tugas Klasifikasi atau Regresi, Kelas diisi kode kelas (A/B/C), NamaSB diisi nama kelompok Study Buddy, ALG1 dan ALG2 diisi nama algoritme yang dibandingkan sesuai instruksi soal, dan nama panggilan dari anggota tim yang bertanggungjawab mengerjakan notebook tersebut.
  - Format penamaan model: **BestModel\_CLF/REG\_Alg\_NamaSB.pkl** di mana CLF adalah kode untuk model klasifikasi dan REG untuk model regresi. Alg adalah nama algoritme yang terbaik yang didump, dan NamaSB diisi nama kelompok Study Buddy.

- Format penamaan file Python Streamlit: **MainStreamlit\_Kelas\_NamaSB.py** di mana kelas diisi kode kelas (A/B/C), NamaSB diisi nama kelompok Study Buddy.
- Seluruh Code Projek UTS dalam bentuk **SATU (1) PDF utuh berisi code EMPAT (4) NOTEBOOK DAN SATU (1) FILE APLIKASI STREAMLIT**. Format penamaan: **Projek UTS\_Kelas\_NamaSB.pdf** di mana Kelas diisi kode kelas (A/B/C) dan NamaSB diisi nama kelompok Study Buddy.
- Semua file diunggah ke **SATU (1) repository GitHub**. Format penamaan repository: **Projek UTS PMDPM\_Kelas\_NamaSB** di mana Kelas diisi kode kelas (A/B/C) dan NamaSB diisi nama kelompok Study Buddy.
- Link GitHub dan link Streamlit Cloud diunggah dalam bentuk file notepad (.txt) ke situs kuliah. Format penamaan file notepad: **Projek UTS PMDPM\_Kelas\_NamaSB.txt** di mana Kelas diisi kode kelas (A/B/C) dan NamaSB diisi nama kelompok Study Buddy.
- Semua anggota kelompok **wajib berbagi tugas dan mengunggah semua file proyek**.
- **Jika ada code yang terindikasi mencontek/copy paste full Notebook dari internet ataupun orang lain, nilai UTS=0.**

### Peer Review

1. Peer review dilakukan sebagai acuan penilaian kinerja individu dalam kerja kelompok.
2. Penilaian ini digunakan untuk PT 7 (Projek Tengah Semester) dan PT 15 (Projek Akhir Semester).
3. Mahasiswa **WAJIB** memberikan penilaian terhadap rekan satu kelompok yang lain, **kecuali dirinya sendiri**.
4. **Jika ada mahasiswa yang tidak mengisi maka penilaian terhadap mahasiswa tersebut sama dengan 1 atau 60% nilai kelompok.**

dengan skor penilaian kinerja sebagai berikut:

Grade	Skor	Kriteria Penilaian
Aktif Berkontribusi	3	Rekan kerja <b>aktif terlibat</b> dalam: <ul style="list-style-type: none"> <li>• <b>koordinasi</b> diskusi analisis studi kasus,</li> <li>• dan membuat proyek code</li> </ul> sesuai pembagian yang sudah ditentukan dalam kelompok serta <b>membantu kesulitan rekan kerja yang lain</b>
Cukup Berkontribusi	2	Rekan kerja terlibat dalam <b>koordinasi</b> diskusi membuat proyek code <b>sesuai pembagian yang sudah ditentukan dalam kelompok</b>
Kurang Berkontribusi	1	Rekan kerja <b>tidak memenuhi pembagian</b> yang sudah ditentukan dalam kelompok
Tidak Berkontribusi	0	Rekan kerja <b>sama sekali tidak terlibat</b> dalam <b>semua tugas</b>

5. Skor penilaian per individu dihitung reratanya. Rerata skor sebagai acuan untuk menghitung nilai individu dengan grade sebagai berikut:

<b>Grade</b>	<b>Rerata Skor</b>	<b>Persentase Nilai Individu</b>
Aktif Berkontribusi	$\geq 2,67$	100% nilai kelompok
Cukup Berkontribusi	$\geq 2$	80% nilai kelompok
Kurang Berkontribusi	$\geq 1$	60% nilai kelompok
Tidak Berkontribusi	0	0

Good luck!