



MODUL PRAKTIKUM PEMBAJARAN MESIN

Mohammad Bayu Anggara, S.Kom., M.Kom.



Nama Mata Praktikum	: Pembelajaran Mesin
Dosen Pengampu	: Mohammad Bayu Anggara, S.Kom., M.Kom.
Semester	: 5
Modul 15	: REVIEW & PENUGASAN TUGAS BESAR

MODUL 14: REVIEW & PENUGASAN TUGAS BESAR

Setelah menyelesaikan modul ini, mahasiswa mampu:

1. Memetakan Algoritma: Mengidentifikasi dan memilih algoritma yang tepat (dari 14 topik yang telah dipelajari) sesuai dengan jenis masalah dan data.
2. Merancang Solusi: Menyusun alur kerja proyek Machine Learning secara lengkap (*end-to-end*), mulai dari akuisisi data, preprocessing, hingga evaluasi.
3. Justifikasi Teknis: Memberikan alasan logis di balik pemilihan metode dan strategi komparasi model untuk Tugas Besar.
4. Evaluasi Objektif: Menentukan metrik keberhasilan proyek yang relevan (seperti F1-Score atau RMSE) untuk mengukur performa model secara akurat.

1. Pendahuluan

Selamat! Anda telah menuntaskan kurikulum yang sangat komprehensif. Dimulai dari pengenalan ekosistem Python, memahami model linear sederhana, mendalami algoritma berbasis pohon yang canggih, hingga menyentuh ranah Deep Learning dengan LSTM dan ANN.

Sekarang, saatnya mengubah pengetahuan tersebut menjadi solusi nyata. Pertemuan ini difokuskan untuk memetakan kembali "senjata" yang Anda miliki dan merancang strategi untuk Tugas Besar (Capstone Project).

2. Peta Algoritma: Apa yang Harus Saya Pakai?

Berikut adalah ringkasan dari **14 Topik** yang telah kita pelajari, dikelompokkan berdasarkan karakteristiknya agar Anda mudah memilih untuk Tugas Besar:

A. Supervised Learning: Klasifikasi & Regresi (Tabular Data)

Kelompok ini membutuhkan data berlabel (*Labelled Data*).

Algoritma	Karakteristik Utama	Kapan Sebaiknya Digunakan?
Logistic Regression	Linear, Probabilistik, Sederhana.	Sebagai Baseline Model (pembanding). Cocok untuk klasifikasi biner sederhana (Ya/Tidak).
KNN	Berbasis Jarak (<i>Distance</i>), <i>Lazy Learning</i> .	Data dengan pola lokal yang jelas, dataset ukuran kecil-menengah. Wajib <i>Scaling</i> !

Naive Bayes	Berbasis Teorema Bayes, Asumsi Independen.	Sangat cepat. Jawara untuk Klasifikasi Teks (NLP), sentimen analisis, filter spam.
SVM	Berbasis Margin (<i>Hyperplane</i>), Geometris.	Data dimensi tinggi, butuh akurasi presisi. Bagus jika data terpisah jelas.

B. Tree-Based Models (Ensemble Learning)

Kelompok "Kekuatan Hutan". Sangat kuat untuk data tabular yang kompleks.

Algoritma	Karakteristik Utama	Kapan Sebaiknya Digunakan?
Decision Trees	Aturan <i>If-Then</i> , Pohon Tunggal.	Butuh model yang mudah dijelaskan (<i>interpretable</i>) visualisasinya. Rentan <i>overfitting</i> .
Random Forest	<i>Bagging</i> (Paralel), Banyak Pohon.	Mengurangi variansi Decision Tree. Sangat stabil dan jarang <i>overfitting</i> .
XGBoost	<i>Boosting</i> (Sekuensial), Optimasi Ekstrem.	Mencari Akurasi Tertinggi . Sangat disarankan untuk kompetisi atau data tabular rumit.

C. Deep Learning & Sequence Models

Kelompok Jaringan Saraf Tiruan.

Algoritma	Karakteristik Utama	Kapan Sebaiknya Digunakan?
ANNs	Multi-layer Perceptron, Non-linear.	Data yang polanya sangat rumit dan tidak bisa diselesaikan model ML biasa.
LSTM	<i>Memory cell</i> , Menangani data urutan (<i>Sequence</i>).	Wajib untuk data Time-Series (Prediksi Saham/Cuaca) atau Teks Panjang (NLP).

D. Lain-Lain (Unsupervised & RL)

Algoritma	Karakteristik Utama	Kapan Sebaiknya Digunakan?
K-Means	Clustering, Berbasis Centroid.	Tidak ada label target. Segmentasi pelanggan, pengelompokan wilayah.
Reinforcement Learning	Agent, Reward, Environment.	Game, Navigasi Robot, Optimasi Logistik, Pemecahan Maze.

3. Strategi Mengerjakan Tugas Besar

Untuk Tugas Besar, Anda diharapkan mengikuti alur kerja standar industri (Pipeline):

Tahap 1: Pemilihan Masalah & Data

- Pilih dataset minimal 1000 baris (kecuali Time-Series atau RL).
- Hindari dataset "Hello World" (Iris, Titanic). Cari yang unik di Kaggle/UCI.

Tahap 2: Preprocessing (Wajib Disesuaikan Algoritma)

Ingat aturan main setiap algoritma yang sudah dipelajari:

- Pakai KNN, SVM, ANN, K-Means → Wajib StandardScaler/MinMaxScaler.
- Pakai Naive Bayes/XGBoost → Wajib Encoding (Ubah teks jadi angka).
- Pakai LSTM → Wajib ubah data jadi format Sequence/Sliding Window.
- Pakai Logistic Regression → Cek multikolinearitas.

Tahap 3: Modeling & Komparasi

Anda wajib membandingkan minimal 2 algoritma.

- *Skenario 1 (Klasifikasi)*: Bandingkan Logistic Regression (Baseline) vs XGBoost (Advanced).
- *Skenario 2 (Time Series)*: Bandingkan Linear Regression (Basic) vs LSTM (Advanced).
- *Skenario 3 (Teks)*: Bandingkan Naive Bayes vs LSTM atau SVM.

Tahap 4: Evaluasi

- Jangan cuma pamer Akurasi!
- Jika data *imbalance*, gunakan F1-Score.
- Jika kasus regresi/forecasting, gunakan RMSE/MAE.

4. Ide Topik Berdasarkan Materi

Berikut inspirasi judul berdasarkan materi yang sudah dipelajari:

1. Prediksi Harga Saham / Cryptocurrency
 - *Metode*: LSTM (Long Short-Term Memory).
2. Deteksi Penyakit (Jantung/Diabetes)
 - *Metode*: Komparasi KNN vs ANN vs Random Forest.
3. Segmentasi Pelanggan Mall
 - *Metode*: K-Means Clustering (dilanjutkan analisis profil).

4. Prediksi Churn (Pelanggan Kabur)
 - *Metode*: XGBoost (karena biasanya datanya *imbalance*).
5. Agen Cerdas Bermain Game Sederhana
 - *Metode*: Reinforcement Learning (Q-Learning).

5. Output Tugas Besar

1. **Laporan Proyek (PDF):**
 - https://docs.google.com/document/d/15Dt9nHmyrKdlR9bRS4V1ZMIng_FwvT DU/edit?usp=share_link&ouid=100945484765281148605&rtpof=true&sd=true
2. **Source Code (.ipynb):**
 - Notebook yang rapi dan bisa dijalankan (*runnable*).
3. **Deployment Berbasis Web**
4. **Video Demo:**
 - Penjelasan solusi, kode, hasil, dan deployment aplikasi yang dibuat.

TUGAS BESAR (CAPSTONE PROJECT)

Mata Kuliah: Machine Learning / Data Science **Tema:** Penerapan End-to-End Machine Learning untuk Penyelesaian Masalah Dunia Nyata **Sifat:** INDIVIDU (Perorangan)

A. Latar Belakang & Tujuan

Tugas Besar ini adalah evaluasi akhir semester untuk mengukur kompetensi setiap mahasiswa dalam membangun solusi *Machine Learning* secara mandiri. Anda ditantang untuk menyelesaikan masalah spesifik menggunakan data riil, mulai dari preprocessing hingga *deployment* aplikasi.

Batasan Topik: Untuk mendorong variasi dan kedalaman analisis teknis, topik **Analisis Sentimen (Sentiment Analysis)** **DILARANG** digunakan dalam tugas ini. Mahasiswa diminta fokus pada masalah prediktif lain seperti Klasifikasi, Regresi, Forecasting, Clustering, atau Computer Vision sederhana.

Capaian Pembelajaran:

1. Mampu merancang *pipeline* ML lengkap secara mandiri.
2. Mampu membandingkan performa antar-algoritma secara objektif.
3. Mampu men-deploy model menjadi aplikasi web interaktif.

B. Ketentuan Umum Proyek

1. Kriteria Dataset

- **Sumber:** Kaggle, UCI ML Repository, Data.gov, atau sumber kredibel lain.
- **Ukuran:** Minimal 1.000 baris data (row) dan minimal 5 fitur (column).
- **Larangan Dataset:**
 - DILARANG menggunakan dataset "Hello World" seperti: **Iris, Titanic, Boston Housing**.
 - DILARANG menggunakan dataset teks untuk **Analisis Sentimen** (Ulasan produk, Tweet sentimen, dll).

2. Rekomendasi Topik

- **Bidang Kesehatan:** Prediksi Stroke, Diabetes, Serangan Jantung, atau Klasifikasi X-Ray (Image).
- **Bidang Finansial:** Prediksi Gagal Bayar Kredit (*Credit Scoring*), Deteksi Fraud, Prediksi Harga Rumah/Mobil.
- **Bidang Bisnis:** Prediksi *Churn* (Pelanggan Berhenti), Segmentasi Pelanggan (Clustering), Prediksi Penjualan (Forecasting).
- **Bidang Industri/Saintek:** Klasifikasi Kualitas Air, Prediksi Cuaca, Deteksi Spam Email (Non-sentimen).

3. Metodologi Teknis

- **Komparasi:** Wajib membandingkan minimal **2 Algoritma** (Misal: Random Forest vs XGBoost).
- **Preprocessing:** Wajib ada penanganan *Missing Values*, *Encoding*, dan *Scaling* (jika perlu).
- **Evaluasi:** Gunakan metrik yang tepat. Jika data tidak seimbang (*imbalanced*), dilarang hanya menggunakan Akurasi (Gunakan F1-Score/AUC).

C. Komponen Output (Deliverables)

Tugas dikumpulkan dalam 4 komponen berikut:

1. Laporan Proyek (Format Jurnal)

Laporan adalah dokumen akademis utama.

- **Template Wajib:** Link Google Docs Template
- **Konten Laporan:**
 - **Abstrak:** Ringkasan masalah dan hasil.
 - **Pendahuluan:** Mengapa masalah ini penting untuk diselesaikan?
 - **Metodologi:** Jelaskan teknik preprocessing (SMOTE, Scaling, dll) dan parameter model.
 - **Hasil & Pembahasan:**
 - Tabel perbandingan hasil 2 model.
 - Grafik *Confusion Matrix & ROC Curve*.
 - Analisis *Feature Importance* (Faktor apa yang paling berpengaruh?).
 - **Kesimpulan:** Model mana yang dipilih untuk deployment dan alasannya.
- **Format:** PDF.
- **Hasil Turnitin maksimal 20%**

2. Source Code (.ipynb)

Notebook Python yang rapi dan terdokumentasi.

- Pastikan kode berjalan (*runnable*) dari awal sampai akhir.
- Berikan penjelasan (Markdown) pada setiap tahapan kode: "Mengapa kolom ini di-drop?", "Mengapa pakai median imputation?".
- Simpan model terbaik menggunakan pickle atau joblib untuk tahap deployment.

3. Deployment Aplikasi Web

Model terbaik harus dijadikan aplikasi web agar bisa dicoba oleh pengguna.

- **Tools:** Gunakan library **Streamlit** (Python).
- **Fungsionalitas:**
 - User dapat menginput nilai fitur (melalui slider, text input, atau upload csv).
 - Klik tombol "Prediksi".
 - Muncul hasil prediksi (Kelas/Nilai) dan Probabilitasnya.

- **Pengumpulan:** File app.py disertakan dalam folder tugas, atau link aplikasi jika di-hosting online (Streamlit Cloud).

4. Video Demo

Video presentasi singkat berdurasi **3-5 menit** (Wajib terlihat wajah/suara mahasiswa).

- **Isi Video:**
 1. Perkenalan & Penjelasan Masalah (30 detik).
 2. Penjelasan singkat Preprocessing & Hasil Model Terbaik (2 menit).
 3. **Demo Aplikasi Web:** Tunjukkan cara input data dummy dan jelaskan hasilnya (Sisa waktu).
- **Format:** Link YouTube

D. Timeline Pengerjaan

Tahap	Deadline	Aktivitas
Proposal	Pertemuan 15	Submit Judul & Link Dataset (untuk di-ACC Dosen).
Pengerjaan	Sampai UAS	Preprocessing, Modeling, dan Pembuatan App Streamlit.
Pengumpulan	Pertemuan (UAS)	Submit semua file (Laporan, Kode, App, Video).