

!! Catatan: Diharapkan seluruh pengerjaan Assignment tidak sepenuhnya mengandalkan penggunaan AI **!!**

"Proses belajar ibarat menanam pohon. Jika hanya mengandalkan AI tanpa memahami esensinya, yang berkembang bukan kompetensimu, melainkan ketergantungan yang melemahkan."

- Learning Design Dibimbing

Assignment Guidance: Create Supervised Machine Learning with Python

Data Science dan Data Analyst Bootcamp

Periode Pembelajaran

Exploring Machine Learning Models with Python: Classification Techniques

Exploring Machine Learning Models with Python: Advanced Classification Techniques

Objectives

1. Student mampu melakukan data preparation.
2. Student mampu melakukan Exploratory Data Analysis (EDA).
3. Student mampu membuat model Supervised Learning.
4. Student mampu melakukan interpretasi model dan hasil.
5. Student mampu melakukan evaluasi model.
6. Student mampu melakukan data preprocessing (fitur engineering).
7. Student mampu mengukur hasil model menggunakan metrik seperti AUC, Precision, dan Recall Score.
8. Student mampu melakukan eksplorasi dari proses membuat model supervised.


Deskripsi Assignment

Assignment ini bertujuan untuk membantu student dalam menganalisis dan memprediksi churn pelanggan menggunakan model supervised learning berbasis algoritma klasifikasi.

Student akan melakukan data preparation, eksplorasi dataset (EDA), feature engineering, membangun model klasifikasi, serta mengevaluasi model menggunakan metrik yang sesuai.

Detail Assignment

Link Google Colab

 Copy of HW Machine Learning Models with Python (Supervised).ipynb

Link Dataset :

<https://drive.google.com/file/d/1YEw9yPDv4p5jw475P8S-haq9zycW6qWf/view?usp=sharing>

1. Pra-pemrosesan Data

- Import library yang dibutuhkan (*numpy, pandas, seaborn, matplotlib, sklearn, imblearn*).
- Muat dataset *Telco Customer Churn* dari sumber yang diberikan.
- Lakukan pembersihan data, termasuk menangani *missing values* dan melakukan encoding data kategorikal.
- Pisahkan dataset menjadi fitur (X) dan target (y).

Goals: Mempersiapkan dataset agar siap digunakan untuk analisis dan pemodelan.

2. Exploratory Data Analysis (EDA)

- Visualisasikan distribusi data untuk memahami pola dan karakteristiknya.
- Identifikasi korelasi antar fitur dan fitur dengan target.
- Lakukan analisis terhadap distribusi data numerik dan kategorikal.

Goals: Memahami hubungan antara variabel dan menentukan fitur yang memiliki potensi signifikan dalam prediksi.

3. Pemrosesan Data dan Feature Engineering

- Lakukan pembersihan data lanjutan seperti transformasi variabel skewed, normalisasi, atau scaling jika diperlukan.
- Pisahkan dataset menjadi *training set* dan *test set*.

Goals: Menghasilkan dataset yang optimal untuk proses pelatihan model.

4. Pemodelan

- Gunakan minimal dua model *Supervised Learning* untuk memprediksi churn (misalnya, *Decision Tree, Random Forest, XGBoost*).
- Lakukan pembagian data menjadi *training set* dan *test set*.

Goals: Membangun dan mengoptimalkan model untuk menghasilkan prediksi yang lebih akurat.

5. Evaluasi Model

- Gunakan metrik evaluasi seperti *Accuracy*, *Precision*, *Recall*, *F1-score*, dan *AUC-ROC*.
- Bandingkan performa berbagai model untuk menentukan model terbaik.
- Berikan penjelasan mengapa model tertentu dipilih sebagai yang terbaik.

Goals: Mengevaluasi performa model berdasarkan metrik yang relevan dan memilih model yang paling sesuai untuk kasus bisnis.

Tools

Google Colaboratory

Pengumpulan Assignment

Deadline :

Maksimal H+7 Kelas (Pukul 23.30 WIB)

Details :

Dikumpulkan dalam bentuk Link Google Colaboratory, secara INDIVIDU, di LMS

Indikator Penilaian

No.	Aspek Penilaian	Parameter	Bobot Maksimal
1	Data Preparation	Student mampu melakukan pembersihan dan pemrosesan data	15
2	Exploratory Data Analysis	Student mampu melakukan EDA dan visualisasi data	15
3	Data Preprocessing	Student mampu menangani data kategorikal, normalisasi, dan split dataset	20
4	Pemodelan	Student mampu membangun model dengan minimal 2 algoritma	20
5	Evaluasi Model	Student mampu menggunakan metrik yang tepat dan menjelaskan pilihannya	15
6	Dokumentasi	Student mampu mendokumentasikan seluruh proses dengan baik	15

Referensi/Template

1. https://scikit-learn.org/stable/supervised_learning.html
2. <https://www.datacamp.com/tutorial/exploratory-data-analysis-python>

Sanksi Penggunaan AI:

Apabila student terdeteksi 100% menggunakan AI, maka hasil assignment akan diberikan skor 0

Ketentuan Pencapaian Nilai:

Nilai minimum Lulus Penyaluran Kerja: 75

Nilai minimum Lulus Bootcamp: 65

Ketentuan Penilaian:

Mengumpulkan Assignment tepat waktu: Sesuai dengan nilai yang diberikan mentor

Mengumpulkan Assignment 12 jam setelah deadline: - 3 dari nilai yang diberikan mentor

Mengumpulkan Assignment 1 x 24 Jam setelah deadline: - 6 dari nilai yang diberikan mentor

Mengumpulkan Assignment 2 x 24 Jam setelah deadline: - 12 dari nilai yang diberikan mentor

Mengumpulkan Assignment 3 x 24 Jam setelah deadline: - 18 dari nilai yang diberikan mentor

Mengumpulkan Assignment 4 x 24 Jam setelah deadline: - 24 dari nilai yang diberikan mentor

Mengumpulkan Assignment 5 x 24 Jam setelah deadline: - 30 dari nilai yang diberikan mentor

Mengumpulkan Assignment 6 x 24 Jam setelah deadline: - 36 dari nilai yang diberikan mentor

Mengumpulkan Assignment 7 x 24 Jam setelah deadline: - 42 dari nilai yang diberikan mentor