

!! 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: Case Study Hyperparameter Tuning

Data Science dan Data Analyst Bootcamp

Periode Pembelajaran

Hyperparameter Tuning in Python

Objectives

1. Student mampu melakukan data cleaning
2. Student mampu memisahkan dataset untuk keperluan training, validasi, dan testing machine learning model
3. Student mampu melakukan seleksi fitur yang redundan berdasarkan plot korelasi, untuk mencegah multicollinearity
4. Student mampu mentransformasi kategorikal fitur sehingga dapat digunakan untuk model training
5. Student mampu memilih metrik pemodelan dengan tepat, bergantung pada konteks permasalahan
6. Student mampu memilih hyperparameter terbaik dengan menggunakan konsep cross validation
7. Student mampu melakukan evaluasi model dengan berbagai metrik, serta menginterpretasikannya
8. Student mampu melakukan eksplorasi dari proses membuat model supervised dan menjawab pertanyaan yang sudah disediakan

Deskripsi Assignment

Assignment ini bertujuan untuk membantu student dalam melakukan Hyperparameter Tuning menggunakan teknik cross-validation untuk meningkatkan performa model Machine Learning.

Student akan melakukan data preparation, feature selection, model training, serta tuning hyperparameter untuk mendapatkan kombinasi parameter terbaik yang meningkatkan performa model prediksi churn pelanggan.

Detail Assignment

Link Dataset : [Churn](#)

Untuk menyelesaikan assignment ini, lakukan langkah-langkah berikut:

1. Persiapan Data

- Buka dataset *Churn Prediction* dan eksplorasi data yang tersedia.
- Identifikasi aspek penting seperti jumlah fitur, tipe data, serta distribusi target.
- Tangani potensi masalah seperti nilai yang hilang atau data yang tidak sesuai.

Goals: Memahami struktur dataset dan memastikan data siap digunakan untuk pemodelan.

2. Pembagian Dataset

- Tentukan strategi pembagian dataset yang tepat untuk training, validasi, dan testing.
- Pastikan dataset mencerminkan distribusi yang seimbang untuk menghasilkan model yang andal.

Goals: Membagi dataset dengan efektif untuk memastikan evaluasi model yang akurat.

3. Analisis Korelasi dan Seleksi Fitur

- Analisis hubungan antar fitur dalam dataset untuk memahami potensi multicollinearity.
- Identifikasi dan pilih fitur yang paling relevan untuk model berdasarkan analisis yang dilakukan.

Goals: Menyusun subset fitur yang optimal untuk meningkatkan performa model.

4. Pengolahan Data Kategorikal

- Teliti fitur kategorikal dalam dataset dan pilih metode transformasi yang sesuai.
- Uji berbagai pendekatan encoding untuk meningkatkan performa model.

Goals: Mengonversi fitur kategorikal agar dapat digunakan dalam model tanpa kehilangan informasi penting.

5. Pemilihan Metrik Evaluasi

- Tentukan metrik evaluasi berdasarkan jenis masalah dan karakteristik dataset.
- Jelaskan alasan pemilihan metrik tertentu dan bagaimana metrik tersebut digunakan untuk menilai kinerja model.

Goals: Menggunakan metrik evaluasi yang tepat untuk menilai performa model dengan baik.

6. Pelatihan Model dan Hyperparameter Tuning

- Pilih model machine learning yang sesuai dengan dataset.
- Lakukan eksperimen dengan berbagai kombinasi hyperparameter.

- Gunakan teknik validasi untuk membandingkan performa model dengan berbagai konfigurasi.

Goals: Menemukan kombinasi hyperparameter terbaik untuk meningkatkan performa model.

7. Evaluasi Model

- Bandingkan performa model sebelum dan sesudah tuning hyperparameter.
- Gunakan metrik yang telah dipilih untuk mengevaluasi efektivitas model.
- Interpretasikan hasil evaluasi untuk memahami bagaimana model dapat dioptimalkan lebih lanjut.

Goals: Memastikan model bekerja dengan optimal dan memahami faktor yang mempengaruhi hasil prediksi.

8. Dokumentasi

- Catat seluruh langkah yang dilakukan dalam proses analisis.
- Buat laporan yang ringkas dan jelas untuk menjelaskan temuan utama.
- Simpan file dengan format: **HW_HPTUNING_<YOUR COMPLETE NAME>** dan pastikan dapat diakses oleh mentor.

Goals: Menyusun dokumentasi yang rapi dan mudah dipahami agar dapat dievaluasi dengan baik.

Tools

Google Collaboratory

Pengumpulan Assignment

Deadline :

Maksimal H+7 Kelas (Pukul 23.30 WIB)

Details :

Dikumpulkan dalam bentuk link Google Colab, secara INDIVIDU, di LMS

Indikator Penilaian

No.	Aspek Penilaian	Parameter	Bobot Maksimal
1	Data Cleaning	Student mampu melakukan pembersihan data sebelum pemodelan	10
2	Split Data	Student mampu membagi dataset untuk training, validasi, dan testing	10
3	Analisis Korelasi & Seleksi Fitur	Student mampu mengidentifikasi fitur yang redundant dan mencegah multicollinearity	10

4	Pengolahan Fitur Kategorikal	Student mampu mentransformasi fitur kategorikal agar dapat digunakan dalam model	15
5	Pemilihan Metrik Evaluasi	Student mampu memilih metrik yang sesuai untuk evaluasi model	10
6	Hyperparameter Tuning	Student mampu memilih hyperparameter terbaik menggunakan cross-validation	30
7	Evaluasi Model	Student mampu melakukan evaluasi model dan menginterpretasikan hasilnya	15

Referensi/Template

1. <https://medium.com/swlh/the-hyperparameter-cheat-sheet-770f1fed32ff>
2. https://scikit-learn.org/stable/modules/grid_search.html

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