



Project-Based Internship

Prediction Model

Membuat model untuk memprediksi credit risk



Table of Content

- 1 Challenge
- 2 Prerequisite
- 3 Hints
- 4 Submission

Challenge

Sebagai Data Scientist di **ID/X Partners**, Anda akan terlibat dalam sebuah proyek dari perusahaan pemberi pinjaman (*multifinance*), dimana *client* Anda ingin meningkatkan keakuratan dalam menilai dan mengelola risiko kredit, sehingga dapat mengoptimalkan keputusan bisnis mereka dan mengurangi potensi kerugian.

Tugas Anda adalah mengembangkan *model machine learning* yang dapat memprediksi risiko kredit (*credit risk*) berdasarkan dataset yang disediakan, yang mencakup data pinjaman yang disetujui dan ditolak. Dalam pengembangan modelnya Anda juga perlu melakukan beberapa tahap dimulai dengan Data Understanding, Exploratory Data Analysis (EDA), Data Preparation, Data Modelling, dan Evaluation.

➤ Data Understanding

- Analisis dataset yang diberikan dan buatlah ringkasan tentang struktur dataset, termasuk jumlah baris dan kolom.
- Mengidentifikasi dan memahami setiap atribut atau kolom data.
- Melakukan eksplorasi awal terhadap distribusi variabel, statistik deskriptif, dan pola umum dalam data.

Challenge

- Exploratory Data Analysis (EDA)
 - Melakukan visualisasi data menggunakan grafik dan plot untuk memahami hubungan antar variabel.
 - Menganalisis korelasi antara fitur-fitur dalam dataset.
 - Melakukan analisis univariat dan bivariat untuk mendapatkan wawasan yang lebih dalam tentang data.
- Data Preparation
 - Menangani nilai yang hilang, baik dengan menghapusnya atau mengisi dengan teknik imputasi yang sesuai.
 - Mengatasi outlier jika diperlukan, entah dengan menghapusnya atau menyesuaikan nilainya.
 - Melakukan encoding pada variabel kategorikal jika diperlukan.
 - Melakukan scaling atau normalisasi pada fitur-fitur numerik jika diperlukan.
 - Membagi data menjadi train set dan test set.

Challenge

➤ Data Modelling

- Memilih model machine learning yang sesuai dengan jenis masalah dan tujuan yang ingin dicapai.
- Melakukan pelatihan model pada set pelatihan.
- Menyesuaikan parameter model menggunakan teknik seperti cross-validation atau grid search.
- Mengevaluasi kinerja model menggunakan metrik yang relevan seperti akurasi, presisi, recall, atau ROC-AUC.
- Memeriksa overfitting atau underfitting model dan mengambil tindakan yang sesuai.

➤ Evaluation

- Melakukan evaluasi kinerja model terhadap set data pengujian yang terpisah.
- Menganalisis metrik evaluasi yang diperoleh untuk mengevaluasi seberapa baik model bekerja.
- Menyusun ringkasan dari hasil evaluasi dan menyajikannya dengan jelas.

Challenge

Dalam pengerjaan final task ini anda dapat menggunakan bahasa pemrograman **Python** dan untuk algoritma machine learning yang digunakan anda wajib menggunakan minimal 2 jenis algoritma, dimana **Logistic Regression** adalah algoritma yang wajib Anda buat dalam project ini.

Dataset yang digunakan: [Loan Dataset](#) ([Data Dictionary](#))

Prerequisite

Untuk persiapan pengerjaan tugas akhir dalam pengembangan model machine learning yang mencakup Data Understanding, Exploratory Data Analysis (EDA), Data Preparation, Data Modelling, dan Evaluation, berikut adalah langkah-langkah yang perlu dipersiapkan:

1. Pahami konsep dasar machine learning dan terminologi yang terkait seperti supervised learning. Karena data yang kamu akan olah adalah data yang berlabel.
2. Penting untuk memahami domain knowledge pada sektor multifinance untuk dapat menentukan pinjaman yang disetujui dan tidak disetujui.
3. Siapkan tools yang akan kamu gunakan untuk membantu pengerjaan soal. Untuk python, kamu dapat menggunakan [jupyter notebook](#) atau [google colab](#).
4. Pelajari teknik-teknik untuk memahami data, termasuk eksplorasi data, visualisasi, dan pre-processing.
5. Pahami teknik-teknik pemodelan machine learning seperti pemilihan model, tuning parameter, dan validasi model.
6. Pengumpulan tugas dikumpulkan dalam format:
 - a. File Code dengan format .py
 - b. File Python Notebook dengan format .ipynb
 - c. File infografis sebagai media presentasi end-to-end solution yang telah kamu buat

Silahkan paparkan dan sematkan hasil pekerjaanmu dalam bentuk portofolio dengan template yang telah ditentukan.

Hints

- Gunakan Data Dictionary sebagai Panduan Utama:
 - Manfaatkan data dictionary sebagai sumber informasi utama untuk memahami setiap atribut dalam dataset dengan baik.
 - Gunakan definisi dan keterangan yang terdapat dalam data dictionary untuk mengartikan variabel dan mengidentifikasi nilai yang mungkin.
- Lakukan Pemeriksaan Kualitas Data Awal:
 - Sebelum memulai analisis, lakukan pemeriksaan kualitas data awal untuk mendeteksi masalah umum seperti nilai yang hilang, outlier, atau duplikat.
 - Gunakan teknik visualisasi sederhana seperti histogram atau scatter plot untuk mengidentifikasi pola yang mencurigakan.
- Berfokus pada Pemahaman Masalah yang Mendalam:
 - Jelajahi latar belakang masalah dengan baik untuk memahami konteksnya secara menyeluruh.
 - Gunakan wawasan dari pemangku kepentingan dan pemahaman domain untuk membantu memandu analisis data.

Hints

- Gunakan EDA secara Kreatif:
 - Jangan terbatas pada visualisasi standar; coba eksplorasi berbagai jenis grafik dan plot untuk menemukan pola yang menarik.
 - Gunakan teknik seperti clustering atau analisis faktor untuk mendapatkan wawasan yang lebih dalam tentang struktur data.
- Lakukan Eksperimen dengan Berbagai Model:
 - Cobalah berbagai jenis model dan algoritma untuk menemukan yang paling cocok dengan data Anda.
 - Lakukan tuning parameter secara cermat untuk meningkatkan kinerja model.
- Buat Dokumentasi yang Komprehensif:
 - Selalu dokumentasikan setiap langkah yang Anda ambil dalam proses analisis data, termasuk pemilihan model, parameter yang digunakan, dan hasil evaluasi.
 - Pastikan untuk menjelaskan alasan di balik setiap keputusan yang Anda buat dalam dokumen Anda.

Hints

- **Gunakan Alat dan Sumber Daya yang Tersedia:**
 - Manfaatkan alat dan sumber daya online seperti tutorial, dokumentasi, dan forum diskusi untuk memperdalam pemahaman Anda tentang teknik analisis data dan pemodelan.
 - Jangan ragu untuk mencari bantuan dari instruktur atau rekan sekelas jika Anda mengalami kesulitan atau pertanyaan.
- **Selalu Evaluasi dan Tinjau Kembali:**
 - Lakukan evaluasi berkala terhadap kemajuan Anda dan tinjau kembali tujuan proyek untuk memastikan Anda tetap berada di jalur yang benar.
 - Gunakan umpan balik dari pemangku kepentingan dan rekan tim untuk membantu Anda memperbaiki dan mengembangkan solusi Anda.
- **Identifikasi Kolom Performa Kredit:**
 - Mulailah dengan mengidentifikasi kolom dalam dataset yang menggambarkan performa kredit dari peminjam.

Hints

- Periksa Kolom dengan Teliti:
 - Periksa isi dari kolom tersebut secara cermat untuk melihat apakah sudah jelas menunjukkan status kredit yang baik atau buruk. Jika ya, label sudah dapat ditetapkan secara langsung.
- Penentuan Label:
 - Jika kolom tersebut sudah jelas menunjukkan status kredit, tetapkan label sesuai dengan kategori yang sudah ditetapkan (misalnya, GOOD atau BAD).
- Teliti Kembali Isi Kolom:
 - Jika isi kolom belum secara langsung menunjukkan status kredit, teliti lagi isi kolom tersebut dengan lebih detail.
- Pemahaman Sejarah Pembayaran:
 - Cari tahu arti dari data dalam kolom tersebut dan perhatikan sejarah pembayaran kredit yang dimiliki oleh masing-masing peminjam.
- Pertimbangkan Keterlambatan Pembayaran:
 - Berdasarkan sejarah pembayaran kredit, tentukan sendiri apakah peminjam termasuk dalam kategori yang baik atau buruk. Biasanya, hal ini dapat ditentukan berdasarkan keterlambatan pembayaran historis yang dimiliki oleh peminjam.

Submission

Setelah Anda menyelesaikan Challenge yang telah diberikan, Anda **DIWAJIBKAN** untuk membuat **File Final Submission** dalam format PPT dengan template [berikut](#) yang berisi:

- Biodata Diri
- Hasil Pengerjaan
- Link Folder/File Hasil Pengerjaan
- Link Github Hasil Pengerjaan
- Link Video Presentasi Hasil Pekerjaan / Penjelasan Project yang Telah Anda Kerjaan

Video presentasi yang telah dibuat, harus Anda upload pada **YouTube** atau **Google Drive** untuk mendapatkan link yang akan dicantumkan pada file final submission nantinya.

Kemudian, **kumpulkan** *File Final Submission* tersebut dengan format **Final Task_ID/X Partners_Data Scientist_<Nama Lengkap>**

File tersebut yang akan **diberikan kepada company** sebagai hasil akhir dari program VIX ini. Diharapkan Anda dapat membuat *File Final Submission* ini dengan sungguh-sungguh. Terima kasih.



Thank You!

