

Kuis 1 Kecerdasan Artifisial

Muhammad Khalid Atthoriq
2341720114
TI/2C

Studi Kasus

Sebuah perusahaan layanan streaming (seperti Netflix atau Spotify) ingin memprediksi pelanggan yang mungkin berhenti berlangganan

Soal & Jawaban

1. Apa definisi masalahnya?
 - Mencari tahu pelanggan dari perusahaan layanan streaming yang kemungkinan besar akan berhenti berlangganan(Churn), berdasarkan data yang sudah ada. Tujuannya adalah agar perusahaan dapat mengambil langkah preventif untuk membuat mereka tetap bertahan. (Churn and not churn)
2. Apa saja langkah untuk menyiapkan data?
 - Mengumpulkan beberapa variabel data seperti berapa lama seseorang berlangganan, seberapa sering mereka memakai layanan, keluhan feedback dari pelanggan, dan catatan pembayaran.
 - Membersihkan data dengan membuang data ganda (redundant data), data yang tidak relevan, mengisi data yang hilang, dan memperbaiki data outlier.
 - Ubah data kategori menjadi angka dengan cara seperti one-hot encoding, dan sesuaikan angka besar-kecil supaya rapi.
3. Bagaimana data akan dibagi?
 - Data akan dibagi menjadi tiga:
 - 70% untuk melatih model
 - 15% untuk validasi guna menyempurnakan parameter model
 - 15% untuk pengujian performa akhir
 - Menggunakan stratified split agar distribusi pelanggan yang churn dan tidak churn tetap proporsional di setiap subset, memastikan representasi yang adil.
4. Bagaimana model akan dilatih?
 - Model akan dilatih menggunakan LightGBM, algoritma berbasis gradient boosting yang efisien dan cepat untuk dataset besar. Pelatihan dilakukan pada data training dengan optimasi hyper-parameter melalui bayesian optimization, dan cross-validation lima lipatan diterapkan untuk memverifikasi kemampuan generalisasi model pada data yang belum dilihat.

5. Bagaimana model akan dievaluasi?
 - Model dievaluasi pada data pengujian dengan metrik seperti AUC-ROC untuk menilai kemampuan menangani ketidakseimbangan kelas, serta precision dan recall untuk mengukur akurasi prediksi churn. Performa dibandingkan dengan baseline sederhana seperti prediksi acak, memberikan gambaran jelas tentang keefektifan model.
6. Bagaimana model akan digunakan?
 - Model diterapkan pada data pelanggan aktif untuk menghasilkan probabilitas churn individu. Perusahaan kemudian menggunakan hasil ini untuk mengidentifikasi pelanggan berisiko tinggi dan menawarkan insentif seperti diskon atau fitur tambahan, bertujuan meningkatkan retensi dan mengurangi tingkat churn.
7. Bagaimana cara mengukur pemantauan kinerja?
 - Pemantauan kinerja dilakukan dengan meacak AUC-ROC dan recall secara rutin pada data baru, serta mengevaluasi dampak bisnis seperti tingkat keberhasilan retensi pelanggan. Jika performa menurun akibat perubahan perilaku pengguna, model dilatih ulang dengan data terbaru untuk menjaga akurasi.

Catatan Tambahan:

1. Churn: Pelanggan yang berhenti langgan
2. One-hot encoding: Mengubah kategori menjadi angka biner
3. Stratified split: Cara bagi data supaya proporsi kelompok tertentu tetap sama di tiap bagian.
4. LIGHTGBM: Algoritma berbasis gradient boosting yang efisien dan cepat untuk dataset besar.
5. Bayesian optimization: Cara cerdas untuk mengoptimalkan hyper-parameter model tanpa mencoba semua kemungkinan.
6. AUC-ROC: Ukuran seberapa baik model membedakan yang pergi dan yang tetap, makin tinggi makin bagus.