



EVALUASI TENGAH SEMESTER - SEMESTER GASAL 2023/2024
DEPARTEMEN STATISTIKA BISNIS – FAKULTAS VOKASI ITS



Mata kuliah / Kelas : Pembelajaran Mesin/ Kelas A
 Hari , Tanggal : Jum'at, 20 Oktober 2023
 Sifat / waktu : Project/20-25 Oktober 2023
 Dosen : Mukti Ratna Dewi, S.Si., M.Sc.

Sub CPMK	Capaian Pembelajaran	Soal nomor
CPMK 1	Konsep dasar Pembelajaran Mesin dan Algoritma	1
CPMK 2	Pemodelan Regresi : Linear, Logistik, SVR/SVM, Decision Tree, Random Forest	1

❖ Deskripsi Kasus

Amazon adalah sebuah perusahaan berbasis teknologi multinasional Amerika yang berbasis di Seattle, Amerika Serikat. Perusahaan ini mulanya berjualan buku secara online, namun saat ini berfokus pada e-Commerce, *cloud computing*, streaming digital, dan kecerdasan buatan. Perusahaan ini menampung sekitar 1.000.000.000 gigabyte data di lebih dari 1.400.000 server. Melalui inovasi terus-menerus dalam *data science* dan *big data*, Amazon selalu terdepan dalam memahami pelanggannya.

Sebagai salah satu perusahaan dengan bisnis e-Commerce yang signifikan, Amazon memiliki risiko tinggi terhadap penipuan ritel. Sebagai tindakan pencegahan, perusahaan mengumpulkan data historis dan *real-time* untuk setiap pesanan. Perusahaan ini menggunakan algoritma pembelajaran mesin untuk menemukan transaksi dengan kemungkinan penipuan (*fraud*) yang lebih tinggi. Tindakan proaktif ini telah membantu perusahaan membatasi klien dengan jumlah pengembalian produk yang berlebihan.

Pada *final project* ini, diberikan sebuah dataset yang berisi transaksi yang dilakukan dengan kartu kredit pada bulan September 2013. Dataset ini menyajikan transaksi yang terjadi dalam dua hari di mana terdapat 492 penipuan (*fraud*) dari 284.807 transaksi. Dataset hanya berisi variabel numerik yang merupakan hasil transformasi PCA. Sayangnya, fitur asli dan informasi terkait latar belakang mengenai variabel ini tidak dapat diberikan lebih lanjut karena masalah kerahasiaan data. Fitur V1, V2, ... V28 adalah komponen utama yang diperoleh dengan PCA. Satu-satunya fitur yang belum diubah dengan PCA adalah 'Waktu (*Time*)' dan 'Jumlah (*Amount*)'. Fitur 'Waktu (*Time*)' berisi detik-detik yang berlalu antara setiap transaksi dan transaksi pertama dalam dataset. Fitur 'Jumlah (*Amount*)' adalah Jumlah transaksi. Fitur 'Kelas (*Class*)' adalah variabel kategorik yang bernilai 1 jika terjadi penipuan (*fraud*) dan 0 jika terjadi sebaliknya.

❖ Tugas Anda adalah melakukan sebuah analisis berikut:

1. Buat sebuah model prediksi jumlah transaksi (*amount*) menggunakan metode regresi, *Support Vector Regression*, dan satu metode prediksi lainnya (bebas). Tentukan model terbaik di antara ketiganya.
2. Buat sebuah model klasifikasi mengenai jenis transaksi menggunakan metode regresi logistik, *decision tree*, *Support Vector Machine*, dan satu metode klasifikasi lainnya (bebas). Tentukan model terbaik di antara keempatnya.

❖ Petunjuk tambahan:

- *Project* dikerjakan secara individu.
- Terdapat dua dataset yang diberikan, yaitu data training dan testing.
- Penilaian berdasarkan kelengkapan dan kedalaman analisis serta seberapa baik model prediksi terpilih (poin 1) dan klasifikasi terpilih (poin 2) yang dihasilkan. Kebaikan model akan dilihat dari evaluasi model pada data *testing*. Semakin tinggi kebaikan model prediksi maupun

Evaluasi Tengah Semester – DSB - 1

Soal Evaluasi sudah sesuai dengan CPMK yang ditetapkan	Tanggal Validasi :	Tanda tangan Validator
--	-----------------------------	------------------------



EVALUASI TENGAH SEMESTER - SEMESTER GASAL 2023/2024
DEPARTEMEN STATISTIKA BISNIS – FAKULTAS VOKASI ITS



Mata kuliah / Kelas : Pembelajaran Mesin/ Kelas A
Hari , Tanggal : Jum'at, 20 Oktober 2023
Sifat / waktu : Project/20-25 Oktober 2023
Dosen : Mukti Ratna Dewi, S.Si., M.Sc.

klasifikasinya maka akan semakin tinggi nilai yang didapatkan. Bila ada dua mahasiswa yang mendapatkan kebaikan model yang sama maka mahasiswa yang terlebih dahulu *submit* yang akan mendapatkan nilai lebih tinggi.

- Yang dibatasi pada *final project* ini adalah metode yang wajib dipakai (metode regresi, SVR, regresi logistik, *decision tree*, dan SVM). Selebihnya mahasiswa dapat berkreaitivitas saat mengolah data untuk mendapatkan hasil prediksi dan klasifikasi terbaik, termasuk pemilihan satu metode prediksi dan klasifikasi di luar metode wajib.
- Wajib menggunakan R untuk analisis.
- Dosen melakukan pengecekan analisis berdasarkan syntax R yang disubmit sehingga pastikan bahwa syntax R bisa dirunning ulang dengan hasil yang sama persis seperti yang tertulis di laporan (*Tips: Gunakan set.seed() apabila terdapat proses randomisasi data*). Bila syntax R yang disubmit **tidak bisa dirunning ulang atau hasilnya berbeda** maka mahasiswa otomatis mendapatkan **nilai nol**.
- Mahasiswa yang terdeteksi saling **copy-paste** dan **modifikasi** syntax serta **menggunakan jasa joki tugas** maka otomatis akan mendapatkan **nilai nol**.
- Hasil analisis dibuat dalam bentuk laporan dengan rincian bab:
 - i. Latar Belakang
 - ii. Metodologi (berisi penjelasan metode lain di luar metode wajib serta tahapan analisis secara rinci).
 - iii. Hasil Analisis
 - iv. Kesimpulan dan SaranLaporan ditulis maksimal 10 lembar halaman A4, TNR 12pt, spasi 1.15, margin kanan-kiri-atas-bawah 1" (1 inch). Tidak termasuk sampul laporan dan daftar pustaka (bila ada).
- File yang perlu dikumpulkan:
 - Laporan akhir dalam format *.pdf
 - Syntax R
 - Hasil prediksi dan klasifikasi dalam format *.csv dengan format penamaan **NRP lengkap_prediksi** dan **NRP lengkap_klasifikasi**. Perlu diingat bahwa hasil prediksi dan klasifikasi ini berjumlah 56.962 sesuai dengan jumlah data testing.
- Deadline: **Rabu, 25 Oktober 2023 pukul 23.59 WIB** dan semua berkas wajib disubmit di myITS Classroom.

Soal Evaluasi sudah sesuai dengan CPMK yang ditetapkan	Tanggal Validasi :	Tanda tangan Validator
--	-----------------------------	------------------------