



PROBLEM:

Big Data Challenge (BDC)

Satria Data 2022

Tim BDC Satria Data 2022



Pendahuluan

- Jaminan sosial berupa asuransi kesehatan banyak diminati masyarakat karena manfaat yang dirasakannya.
- Badan Penyelenggara Jaminan Sosial Kesehatan (BPJS Kesehatan) merupakan penyelenggara program jaminan sosial di bidang kesehatan bagi seluruh rakyat Indonesia.
- Peserta BPJS Kesehatan memperoleh manfaat berupa layanan kesehatan di fasilitas Kesehatan se-Indonesia.

Pelayanan Kesehatan Tingkat Pertama

- Pelayanan Kesehatan Tingkat Pertama adalah pelayanan kesehatan perorangan yang bersifat non spesialisik (primer) meliputi pelayanan rawat jalan dan rawat inap yang diberikan oleh Fasilitas Kesehatan Tingkat Pertama (FKTP) meliputi:
 - Puskesmas atau yang setara
 - Praktik Mandiri Dokter
 - Praktik Mandiri Dokter Gigi
 - Klinik pertama atau yang setara termasuk fasilitas kesehatan tingkat pertama milik TNI/Polri
 - Rumah Sakit Kelas D Pratama atau yang setara
 - Faskes Penunjang: Apotik dan Laboratorim

PROBLEM

- Untuk keperluan mengefektifkan pengelolaan anggaran BPJS Kesehatan, maka pada lomba BDC Satria Data 2022 Anda ditantang untuk melakukan prediksi status pulang peserta berdasarkan data FKTP BPJS Kesehatan.
- Bangun pemodelan untuk klasifikasi status pulang peserta menjadi “Sehat” dan “Belum_Sehat” dengan memanfaatkan data train.
- Lakukan prediksi kelas status pulang peserta pada data test.



Dataset

- Data yang digunakan dalam BDC merupakan catatan kunjungan dari peserta BPJS Kesehatan ke Fasilitas Kesehatan Tingkat Pertama (FKTP) selama tahun 2015 hingga 2020.
- Data kunjungan FKTP BPJS Kesehatan disimpan dalam format *plain text* (.txt), antar *cell data* dipisahkan dengan koma. Urutan *cell* dari kiri ke kanan, sesuai dengan nama variabel yang tersedia dalam file metadata.

Data untuk peserta BDC

- Peserta akan mendapatkan dua jenis data yaitu:
 - Data train FKTP yang disimpan dalam train_fktp.txt
 - Data ini berisi 4,056,898 baris data kunjungan dengan 26 kolom, termasuk kolom target.
 - Data test FKTP yang disimpan dalam test_fktp.txt
 - Data ini berisi 1,014,225 baris data kunjungan yang harus diprediksi dengan model yang diperoleh peserta BDC.

Pengumpulan

- Peserta BDC mengumpulkan hanya hasil prediksi kelas status pulang peserta untuk data test dalam file berformat *comma separated value (csv)*.
- Gunakan file `submission.csv` untuk mengumpulkan jawaban
- Pengukuran yang digunakan oleh Tim Juri BDC pada tahap penyisihan adalah *F1-Score*.
- Gunakan pengkodean: 0 = Belum_Sihat dan 1 = Sehat

BATASAN

- Peserta dilarang menggunakan data lain, selain dari data yang telah ditetapkan oleh tim BDC Satria Data 2022.
- Peserta dilarang mengacak urutan pada file submission.csv ketika melakukan pengumpulan jawaban.

LANJUTAN

- Berdasarkan metrik pengukuran, Tim Juri BDC akan mengambil 20 tim dengan metrik pengukuran tertinggi.
- Selanjutnya, kedua puluh tim akan mendapatkan problem kedua untuk diselesaikan.
- Problem kedua akan diberikan pada tanggal **30 Oktober 2022**.
- Berdasarkan pengerjaan problem tersebut, Tim Juri BDC akan menyeleksi 10 tim terbaik sebagai Finalis BDC Satria Data 2022.
- Peserta memiliki waktu untuk menyelesaikan problem kedua hingga **12 November 2022**.



Diskualifikasi atau sanksi

Diskualifikasi atau sanksi akan diberlakukan terhadap peserta lomba yang karyanya:

1. Tidak sesuai dengan ketentuan dalam panduan, seperti tema/topik, data, *template*, dan sebagainya.
2. Terbukti melakukan plagiarisme.
3. Peserta yang melakukan kecurangan, seperti melakukan prediksi secara manual.