



BPJS Visit Type Classification for Prediction Healthy-Sick using XGBoost Approach

Presented by Group 6



Introduction

Kesehatan adalah kebutuhan dasar yang penting untuk menunjang kualitas hidup masyarakat. Di Indonesia, akses layanan kesehatan yang merata dan terjangkau dijamin melalui BPJS Kesehatan, lembaga yang didirikan pemerintah untuk mengimplementasikan Jaminan Kesehatan Nasional (JKN). Namun, pengelolaan BPJS Kesehatan menghadapi tantangan, seperti efisiensi pelayanan dan pengelolaan data untuk memprediksi jenis kunjungan peserta di Fasilitas Kesehatan Tingkat Pertama (FKTP).

Proyek ini menggunakan algoritma XGBoost, yang dikenal mampu menangani data besar dengan akurasi tinggi, untuk mengklasifikasikan jenis kunjungan peserta BPJS Kesehatan, seperti "Berobat Jalan", "Rujuk Lanjut", dan lainnya. Dengan pendekatan berbasis data ini, BPJS Kesehatan diharapkan dapat mengoptimalkan alokasi sumber daya dan meningkatkan efisiensi serta efektivitas layanan kesehatan di FKTP.

Project Objectives

01.

Memenuhi syarat penyelesaian mata kuliah Data Mining T.A. 2024/2025

02.

Mendalami konsep, metode, dan aplikasi algoritma XGBoost untuk memecahkan masalah nyata, khususnya prediksi status jenis kunjungan peserta FKTP dalam konteks BPJS Kesehatan

03.

Mengembangkan model prediktif berbasis data untuk meningkatkan efisiensi dan efektivitas layanan kesehatan, dengan fokus pada analisis status kunjungan peserta (kunjungan sakit dan sehat) guna mendukung perencanaan sumber daya yang lebih optimal di FKTP.

Project Benefits

1. Memberikan pengalaman langsung kepada mahasiswa dalam membangun model prediktif menggunakan algoritma XGBoost, keterampilan penting untuk pengembangan karir di bidang data science dan machine learning.
2. Menunjukkan penerapan nyata data mining dalam menyelesaikan tantangan kompleks di sektor kesehatan.
3. Membantu BPJS Kesehatan mengoptimalkan perencanaan sumber daya di FKTP melalui model prediktif yang meningkatkan efisiensi dan kualitas layanan kesehatan.





■ 01

Dataset yang digunakan adalah Data Kunjungan yaitu berisi catatan kunjungan peserta ke FKTP.

■ 02

Metodologi menggunakan pendekatan CRISP-DM yang meliputi business understanding, data understanding, data preparation, modeling, evaluation, dan deployment.

■ 03

Pendekatan Algoritma menggunakan XGBoost, algoritma machine learning unggul untuk klasifikasi jenis kunjungan peserta di FKTP (sehat/sakit) dengan mempertimbangkan status kunjungan peserta.

Project Scope

Methodology

Cross Industry Standard Process for Data Mining (CRISP-DM) adalah model proses yang dirancang untuk memandu pelaksanaan proyek data mining secara efisien dan dapat diandalkan. Dikembangkan oleh perusahaan seperti DaimlerChrysler, SPSS, NCR, dan OHRA, CRISP-DM bertujuan menciptakan kerangka kerja yang fleksibel untuk berbagai kebutuhan industri tanpa bergantung pada sektor tertentu atau teknologi yang digunakan. Model ini terdiri dari enam fase utama yang terstruktur namun tetap adaptif, sehingga mendukung efisiensi, reproduktifitas, dan kolaborasi yang lebih baik antar tim dalam proyek data mining.

DAMI

-

Project



01

Business Understanding

Analisis ini bertujuan membantu BPJS Kesehatan mengklasifikasikan jenis kunjungan peserta di FKTP untuk mengoptimalkan alokasi sumber daya seperti tenaga medis dan fasilitas.

02

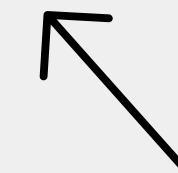
Data Understanding

Tahapan ini mencakup pengumpulan data diagnosis (ICD-10), fasilitas kesehatan, pola rujukan, dan riwayat kunjungan. Data dianalisis untuk memahami distribusi, pola, dan variabilitas, serta divalidasi untuk memastikan kualitas melalui pemeriksaan nilai hilang, outlier, dan inkonsistensi.

03

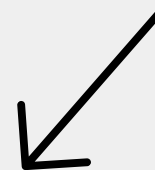
Data Preparation

Pada tahap ini, data dipilih, dibersihkan, dan dikonstruksi untuk meningkatkan kualitas dan relevansi. Fitur seperti diagnosis dan pola rujukan dipilih, data duplikat dan outlier dihapus, label target seperti "Berobat Jalan" atau "Kunjungan Sehat" ditentukan.



CRISP-DM

for project



04

Modeling

Model dibangun menggunakan algoritma XGBoost setelah menyusun skenario pengujian dengan data uji dan metrik seperti accuracy, precision, recall, dan F1 score. Model dilatih dengan data latih untuk mengidentifikasi pola pada jenis kunjungan peserta.

05

Evaluation

Evaluasi dilakukan untuk memastikan model memenuhi tujuan bisnis dengan menggunakan metrik relevan. Jika hasilnya belum optimal, model ditinjau ulang untuk meningkatkan performa, seperti dengan tuning hyperparameter atau mencoba pendekatan lain.

06

Deployment

Model diterapkan dalam sistem untuk klasifikasi otomatis, dengan rencana pemeliharaan guna menjaga performa optimal.

65%

accuracy

Model menghasilkan akurasi 64.98%, dengan performa baik pada kelas mayoritas (0 dan 1) dengan rata-rata f1-score 0.66. Namun, performa pada kelas minoritas (2 dan 4) rendah, masing-masing f1-score hanya 0.36 dan 0.08, menunjukkan kesulitan dalam mengenali pola pada kelas ini. Kelas 3 memiliki f1-score sempurna 1 karena fitur yang mendukung pelatihan. Rata-rata macro avg precision, recall, dan f1-score sebesar 0.55, mencerminkan ketidakseimbangan performa antar kelas, sehingga diperlukan perbaikan pada data minoritas.

0.08

09

DAMI

■ Akurasi Keseluruhan

Lorem ipsum dolor sit amet, consectetur adipiscing elit, sed do eiusmod tempor incididunt ut labore et dolore magna aliqua. Ut enim ad minim veniam.

■ Kendala pada Kelas Minoritas

Model kesulitan dalam mengklasifikasikan kelas minoritas, yang ditunjukkan oleh nilai **f1-score** rendah pada kelas tertentu. Hal ini mengindikasikan perlunya strategi tambahan untuk meningkatkan performa pada kelas ini.

■ Potensi Peningkatan Model

Teknik seperti **oversampling**, **cost-sensitive learning**, atau pengembangan model khusus untuk data imbalan dapat diterapkan untuk mengatasi ketidakseimbangan kelas.

■ Keberhasilan Proyek

Proyek ini berhasil membangun pipeline klasifikasi multikelas yang lengkap dan memberikan landasan yang kuat untuk pengembangan model lebih lanjut dengan pendekatan yang lebih baik pada data tidak seimbang.



Final reflections and future steps

Sistem ini berhasil melakukan klasifikasi jenis kunjungan pasien berdasarkan input seperti provinsi, poliklinik, diagnosis, dan jenis kunjungan. Dengan antarmuka interaktif berbasis Flask, sistem menampilkan prediksi dalam tabel, membantu fasilitas kesehatan merencanakan sumber daya, serta memiliki potensi untuk pengembangan lebih lanjut dalam pengelolaan data BPJS.

Disarankan menambahkan pesan kesalahan yang informatif, mengoptimalkan kinerja model, dan meningkatkan desain antarmuka. Untuk penggunaan skala besar, perlu ditambahkan fitur keamanan seperti autentikasi, enkripsi data, dan proteksi serangan guna memastikan aplikasi tetap aman dan efisien.

Thank You

by Group 6