

Background

Stroke merupakan penyakit dengan tingkat kematian dan kecacatan yang tinggi. Masalah utama di dunia medis adalah **keterlambatan deteksi dini**, di mana pasien berisiko stroke sering diklasifikasikan sebagai sehat. Selain itu, model machine learning sering hanya berhenti di tahap eksperimen dan tidak siap digunakan dalam sistem nyata. Oleh karena itu, dibutuhkan sistem prediksi risiko stroke yang sensitif, konsisten, dan siap produksi melalui pendekatan MLOps.

Value proposition

Sistem ini memberikan peringatan dini risiko stroke berbasis AI dengan sensitivitas tinggi.

Nilai utama sistem:

1. Mengurangi risiko false negative (pasien stroke dikatakan sehat)
2. Mendukung keputusan medis sebagai second opinion
3. Model dapat dijalankan secara konsisten di berbagai lingkungan (reproducible)
4. Mudah digunakan melalui antarmuka web interaktif

Objectives

1. Membangun sistem prediksi risiko stroke berbasis machine learning
2. Mengimplementasikan workflow MLOps end-to-end
3. Menangani permasalahan imbalanced data pada kasus medis
4. Menyediakan online inference melalui aplikasi web
5. Menjamin reproducibility dengan containerization (Docker)

Solution

Solusi berupa aplikasi web prediksi risiko stroke berbasis **Streamlit**.

Fitur utama:

- Input data pasien (profil, medis, gaya hidup)
- Pilihan model (Random Forest, SVM, Logistic Regression)
- Output probabilitas risiko stroke dengan visual peringatan
- Risk calibration berbasis pengetahuan medis

Out-of-scope: diagnosis klinis final (sistem hanya untuk skrining awal).

Feasibility

Proyek feasible karena:

- Dataset publik tersedia
- Menggunakan library open-source (Scikit-learn, Streamlit)
- Infrastruktur sederhana (Docker, GitHub)
- Model dapat dijalankan di laptop lokal tanpa resource besar

Data

1. Sumber data: healthcare-dataset-stroke-data.csv
2. Data training dan inference berasal dari dataset yang sama
3. Label: stroke (0 = tidak stroke, 1 = stroke)
4. Preprocessing:
 - Imputasi nilai BMI
 - Encoding data kategorikal
 - Standardisasi data numerik
5. Masalah utama: **class imbalance**

Metrics

Metrik utama: **Recall (Sensitivitas)**
Confusion Matrix digunakan sebagai alat evaluasi untuk menganalisis distribusi True Positive, False Negative, False Positive, dan True Negative.

Recall diprioritaskan karena kesalahan false negative sangat berbahaya dalam konteks medis.

Evaluation

Offline evaluation:

- Confusion Matrix
- Recall Score

Online evaluation:

- Observasi hasil prediksi di aplikasi
- Analisis perbedaan output antar model
- Validasi hasil dengan logika medis (risk calibration)

Modeling

Tahap modeling dilakukan dengan pendekatan komparatif, di mana beberapa algoritma machine learning dilatih dan dievaluasi menggunakan pipeline preprocessing yang sama. Model yang digunakan meliputi **Logistic Regression**, **Support Vector Machine (SVM)**, dan **Random Forest**. Evaluasi difokuskan pada metrik recall untuk memastikan kemampuan model dalam mendeteksi pasien berisiko stroke dan meminimalkan kesalahan false negative. Model dengan performa terbaik dipilih sebagai model utama, sementara model lain tetap disediakan sebagai pembanding.

Inferences

Inference dilakukan secara online (**real-time**) melalui aplikasi web berbasis Streamlit. Pengguna memasukkan data pasien, kemudian sistem melakukan preprocessing data, memanggil model terlatih, dan menampilkan probabilitas risiko stroke secara langsung tanpa perlu proses pelatihan ulang.k

Feedback

Feedback diperoleh dari hasil prediksi sistem, perbandingan output antar model, serta analisis kasus dengan risiko tinggi. Feedback ini digunakan sebagai dasar untuk melakukan iterasi model, penyesuaian threshold risiko, dan peningkatan performa sistem di tahap selanjutnya.

Project

Tim: Kelompok 9

1. 122450058 Adil Aulia Rahma Nurhidayah
2. 122450059 Nathanael Daniel Santoso
3. 122450061 Kharisa Harvany
4. 122450066 Cintya Bella
5. 122450102 Daris Samudra

Deliverables:

- Source code (GitHub)
- Model .pkl
- Aplikasi Streamlit
- Dockerfile
- Dokumentasi README

Timeline:

- Data & preprocessing
- Modeling & evaluasi
- Deployment
- Final report & presentasi