

Project Brief

Prediksi Risiko Penyakit Kronis Berbasis Data Sosial-Ekonomi dan Gaya Hidup

ID Grup : **LAI25-SM082**

Tema yang di pilih : Keberlanjutan dan Kesejahteraan ▾

Nama Advisor : **Ahmad Zein Al Wafi**, [Sesi mentoring dilakukan pada] **5 Juni 2025**

Anggota Grup :

1. A001XBM122 – Dewi Kunthi Siswati Suryo – IPB University – [Aktif]
2. A001XBM196 – Halida Fiadnin – IPB University – [Aktif]
3. A001XBM223 – Illinia Malika Putri – IPB University – [Aktif]
4. A001YBM349 – Muhammad Rizky Fajar – IPB University – [Aktif]

LATAR BELAKANG:

Penyakit kronis seperti diabetes, stroke, dan penyakit jantung semakin banyak menyerang kelompok usia produktif hingga lanjut usia akibat pola hidup yang tidak sehat dan tekanan sosial-ekonomi. Padahal, risiko ini dapat ditekan melalui deteksi dini dan intervensi gaya hidup yang tepat. Saat ini, belum banyak sistem prediksi yang secara komprehensif menysasar berbagai kelompok usia dengan mempertimbangkan faktor sosial-ekonomi dan gaya hidup sebagai indikator risiko. Proyek ini bertujuan untuk mengembangkan model *machine learning* berbasis data sosial-ekonomi dan kebiasaan hidup untuk memprediksi risiko penyakit kronis pada masyarakat umum, sehingga mereka dapat mengambil langkah pencegahan lebih awal. Proyek ini dimulai dengan mempersiapkan data melalui pra-pemrosesan data termasuk penanganan nilai hilang, penanganan data *imbalance* menggunakan teknik SMOTEN, *encoding* kolom kategorikal, dan standarisasi fitur numerik. Data kemudian dibagi menjadi set latih dan uji. Model prediksi dibangun menggunakan TensorFlow untuk *Multi-output Neural Network* (MNN5) dengan *hyperparameter* yang digunakan yaitu jumlah unit pada setiap *hidden layer* terdiri dari 192 unit pada *layer* pertama, 256 unit pada *layer* kedua, dan 224 unit pada *layer* ketiga. Untuk mencegah *overfitting*, diterapkan *dropout* dengan nilai 0,2 pada *layer* pertama, 0,3 pada *layer* kedua, dan 0,4 pada *layer* ketiga. *Learning rate* diatur pada 0,001. Proses pelatihan dijalankan selama 100 *epoch* dengan pelatihan dimulai pada *epoch* ke-10. Evaluasi model dilakukan dengan akurasi dan *loss* dihitung untuk masing-masing *output*. Prediksi dilakukan dengan menggunakan probabilitas (*output sigmoid*), yang kemudian dikonversi ke nilai 0 atau 1 dengan *threshold* 0,5. Selain itu, menggunakan tiga metrik utama yaitu *precision* yang mengukur akurasi prediksi positif, *recall* yang menunjukkan seberapa banyak kasus positif yang berhasil ditemukan, dan *F1-score* yang merupakan *harmonic mean* antara *precision* dan *recall* untuk memberikan gambaran keseimbangan antara keduanya, serta *confusion matrix*. Setelah model diuji, model disimpan ke format [Tensorflow.js](https://www.tensorflow.js.org/). Website berbasis Streamlit dapat digunakan untuk memprediksi risiko penyakit kronis secara langsung, membantu masyarakat dalam deteksi dini dan pencegahan penyakit. Hasil yang diperoleh menunjukkan akurasi di atas 85% yang menunjukkan bahwa model memiliki kemampuan yang sangat baik dalam memprediksi risiko penyakit kronis dengan tingkat kesalahan yang relatif rendah. Selain itu, metrik lain menunjukkan bahwa model ini mampu mendeteksi penyakit dengan tingkat keandalan yang tinggi, baik dalam hal identifikasi kasus positif (*recall*) maupun keseimbangan antara hasil positif dan negatif yang dihasilkan (*precision*).

STATUS PROYEK: 100% Proyek selesai sesuai dengan rencana proyek

TANGKAPAN LAYAR/VIDEO DEMO:

Prediksi Penyakit (Heart Disease, Stroke, Diabetes)

Hipertensi
Tidak

Kolesterol tinggi
Tidak

Pemeriksaan kolesterol dalam 5 tahun Terakhir
Tidak

Body Mass Index (BMI)
25

Perokok aktif
Tidak

Melakukan aktivitas fisik dalam 30 hari terakhir
Tidak

Mengonsumsi buah setiap hari
Tidak

Mengonsumsi sayur setiap hari
Tidak

Mengonsumsi lebih dari 7 alkohol dalam seminggu
Tidak

Mempunyai asuransi kesehatan
Tidak

Membayar untuk berobat ke dokter
Tidak

Hari buruk mental, 0-30
0

Hari buruk fisik, 0-30
0

Kesulitan berjalan atau menaiki tangga
Tidak

Gender
Perempuan

Usia
40

Status Kesehatan Umum
Sangat Baik

Tingkat Pendidikan
Tidak Pernah Sekolah/Hanya TK

Pendapatan Bulanan
< Rp 1,000,000

Prediksi

Gambar 1 Tangkapan layar demo

LINK DATASET:

[Heart Disease Health Indicators Dataset](#)

LINK DEPLOYMENT:

Dikarenakan streamlit sedang tidak bisa untuk deploy (masalah dengan versi python), maka kami menggunakan URL ngrok.

<https://c9fc-34-87-17-167.ngrok-free.app/>

LINK REPOSITORI GITHUB:

[Repository Grup Capstone Laskar AI \(ID: LAI25-SM082\)](#)

LINK SLIDE PRESENTASI:

[Slide Presentasi Grup Capstone Laskar AI \(ID: LAI25-SM082\)](#)

ANALISIS PASAR**1. PASAR TARGET**

- a. Rentang usia, Profesi, Hobi, dan lain sebagainya.

Rentang usia	:	Kelompok usia produktif hingga lanjut usia
Profesi	:	Masyarakat umum
Hobi	:	Tidak ada hobi tertentu
Lain-lain	:	Individu yang ingin memantau risiko kesehatan mereka secara pribadi

- b. Mengapa target pasar Anda membutuhkan solusi Anda?

i. **Didorong oleh tujuan:**

Solusi ini dibutuhkan untuk mendorong kesadaran kesehatan preventif di masyarakat luas dan membantu mengurangi beban ekonomi jangka panjang akibat penyakit kronis. Penyakit kronis semakin meningkatkan angka kematian manusia, dan pengobatan untuk penyakit ini menghabiskan lebih dari 70% pendapatan pasien, sehingga sangat penting untuk meminimalkan faktor risiko pasien yang dapat menyebabkan kematian.

ii. **Didorong oleh data:**

Proyek ini bertujuan untuk mengembangkan model *machine learning* berbasis data sosial-ekonomi dan kebiasaan hidup untuk memprediksi risiko penyakit kronis. Hal ini disebabkan tingginya angka penyakit kronis akibat kurangnya alat prediksi yang mudah diakses dan berbasis data nyata.

iii. **Alasan Kebijakan Pemerintah:**

Solusi prediksi risiko penyakit berbasis data mendukung kebijakan pemerintah, seperti pada program GERMAS, karena dapat membantu masyarakat mengetahui risiko penyakit sejak awal, sehingga langkah pencegahan bisa dilakukan lebih cepat. Selain itu, penggunaan teknologi seperti ini juga sesuai dengan upaya pemerintah dalam digitalisasi layanan kesehatan dan pengurangan beban biaya pengobatan di masyarakat.

iv. **Stakeholder yang terkait dan mendapat manfaat dari solusi Anda:**

- **Masyarakat umum:**

Mendapat kemudahan untuk mengetahui dan memantau risiko penyakit kronis sejak dini, sehingga bisa melakukan pencegahan lebih awal.

- **Keluarga dan caregiver:**

Mendapat membantu untuk memantau kesehatan anggota keluarga dan mengambil tindakan lebih cepat jika ditemukan risiko.

- **Klinik, puskesmas, dan lembaga kesehatan:**

Menggunakan solusi ini untuk edukasi, skrining awal, dan perencanaan intervensi kesehatan secara lebih efektif.

- **Pemerintah dan instansi kesehatan:**

Memperoleh data pendukung untuk program pencegahan, kebijakan promotif, serta penguatan sistem kesehatan berbasis data.

- **Peneliti dan akademisi:**

Bisa memanfaatkan data dan sistem ini untuk riset lanjutan dan pengembangan inovasi di bidang kesehatan masyarakat.

2. Analisis SWOT dari proyek

a. Kekuatan

- **Pendekatan komprehensif:**

Menggabungkan faktor sosial-ekonomi dan gaya hidup, sehingga prediksi lebih relevan dan menyeluruh.

- **Berbasis data terbuka:**

Memudahkan pengembangan dan kolaborasi, serta memungkinkan penerapan lebih luas.

- **Mendukung upaya preventif:**

Dapat membantu masyarakat dan lembaga kesehatan dalam mengambil langkah pencegahan lebih awal.

b. Kelemahan

- **Keterbatasan data primer:**

Hanya mengandalkan data sekunder/publik, sehingga validasi hasil prediksi secara klinis masih terbatas.

- **Belum diuji di praktik medis langsung:**

Implementasi di dunia nyata membutuhkan uji coba lebih lanjut dan pengawasan dari tenaga medis.

c. Peluang

- **Dukungan kebijakan pemerintah:**

Sejalan dengan program promotif dan preventif kesehatan serta digitalisasi layanan kesehatan nasional.

- **Potensi kolaborasi:**

Bisa dikembangkan bersama lembaga kesehatan, universitas, dan startup teknologi kesehatan.

- **Pengembangan fitur lanjutan:**

Berpotensi dikembangkan lebih lanjut untuk personalisasi rekomendasi atau integrasi dengan layanan kesehatan lain.

d. Ancaman

- **Persaingan dengan solusi serupa:**

Ada kemungkinan muncul aplikasi serupa dengan teknologi lebih maju.

- **Isu keamanan dan privasi data:**

Pengelolaan data kesehatan pribadi harus mengikuti regulasi ketat agar tidak terjadi kebocoran data.

KOMENTAR DARI ADVISOR :

Ka Ahmada Zein selaku advisor memberikan beberapa masukan terkait strategi untuk meningkatkan kualitas model yang dibangun, salah satunya menggunakan teknik balancing data. Selain itu, untuk proyek ini Ka Ahmad Zein juga menyarankan untuk mencoba membangun model multi-output classification tanpa menggunakan variabel kadar gula darah karena hal tersebut merupakan hasil tes untuk pengecekan penyakit diabetes sehingga tidak relevan untuk proyek ini yang berfokus pada pengaruh sosial-ekonomi dan gaya hidup.

Apakah capstone project yang diimplementasikan berbeda dari rencana awal? Jika ya, bagaimana perubahan ini berdampak pada keberhasilan dan hasil proyek?

Iya. *Capstone project* yang diimplementasikan sedikit berbeda dari rencana awal. Awalnya, kami merencanakan untuk menggunakan tiga sumber data, namun setelah melakukan analisis lebih lanjut, kami hanya menggunakan satu sumber data saja untuk mendukung model yang akan dibangun. Oleh karena itu, proses penggabungan data yang sebelumnya direncanakan tidak diperlukan lagi, dan hal ini mempermudah proses pengolahan data serta mengurangi kompleksitas proyek secara keseluruhan.