

Nama : Felicia Amanda Cahyadewi
NIM : A11.2022.14090
Mata Kuliah : Sistem Temu Kembali Informasi - A11.4701
UTS

1. Topik / Judul

Prediksi Diabetes Menggunakan Algoritma Neural Network

2. Deskripsi Singkat (Business Understanding dan Data Understanding)

Pada sistem prediksi diabetes ini menggunakan data pasien, yang mencakup faktor seperti usia, jenis kelamin, hipertensi, penyakit jantung, riwayat merokok, BMI, tingkat HbA1c, dan kadar glukosa darah. Data ini diolah agar model dapat memprediksi risiko diabetes dengan akurat, sehingga dapat membantu tenaga medis dalam identifikasi dini pasien yang berisiko tinggi.

3. Masalah dan Tujuan yang Akan Diselesaikan

Masalah: Diabetes merupakan penyakit kronis yang perlu dideteksi secara dini untuk mengurangi komplikasi. Namun, proses identifikasi diabetes secara manual membutuhkan waktu dan sumber daya yang besar.

Tujuan: Mengembangkan sistem prediksi otomatis untuk diabetes dengan menggunakan algoritma neural network (MLPClassifier) yang dapat membantu dalam menganalisis data pasien dengan cepat dan akurat. Dengan sistem ini, diharapkan dapat memberikan indikasi awal tentang kemungkinan diabetes berdasarkan fitur-fitur medis yang ada.

4. Penjelasan Dataset

Dataset yang digunakan memiliki beberapa fitur utama yang berhubungan dengan diabetes:

- **gender**: Kategori jenis kelamin pasien (Female, Male, Other).
- **age**: Usia pasien.
- **hypertension**: Apakah pasien memiliki hipertensi (0 = Tidak, 1 = Ya).
- **heart_disease**: Apakah pasien memiliki riwayat penyakit jantung (0 = Tidak, 1 = Ya).
- **smoking_history**: Riwayat merokok pasien (No Info, never, former, current, not current, ever).
- **bmi**: Body Mass Index (BMI) pasien, sebuah indikator umum untuk melihat berat badan yang sehat.
- **HbA1c_level**: Tingkat HbA1c, indikator kontrol gula darah jangka panjang.
- **blood_glucose_level**: Tingkat glukosa darah saat ini.

5. Alur / Tahapan / Kerangka Eksperimen

1. Pengumpulan Data: Memastikan dataset yang berkualitas dan relevan tersedia untuk eksperimen ini.
2. Pra-pemrosesan Data: Mengonversi data kategori seperti gender dan riwayat merokok menjadi bentuk numerik agar bisa digunakan oleh model neural network.
3. Pelatihan Model: Menggunakan MLPClassifier untuk melatih model prediksi diabetes berdasarkan data yang telah disiapkan.
4. Evaluasi Model: Menggunakan data uji untuk mengevaluasi akurasi model dan memastikan bahwa model dapat memprediksi diabetes dengan akurat.
5. Implementasi Web App: Membangun aplikasi berbasis Streamlit untuk memudahkan penggunaan model oleh pengguna umum.
6. Pengujian dan Pengoptimalan: Menguji aplikasi secara keseluruhan, dan melakukan pengoptimalan jika diperlukan.

6. Timeline Eksperimen (Sampai dengan Minggu ke-7)

No	Minggu	Kegiatan Eksperimen
1	Minggu ke-1	Pengumpulan dan eksplorasi dataset.
2	Minggu ke-2	Pra-pemrosesan data (konversi fitur kategori, penanganan missing values jika ada).
3	Minggu ke-3	Pembagian data untuk pelatihan dan pengujian, serta pelatihan awal model neural network.
4	Minggu ke-4	Evaluasi awal model dan tuning hyperparameter untuk meningkatkan akurasi.
5	Minggu ke-5	Implementasi model ke dalam aplikasi Streamlit.
6	Minggu ke-6	Pengujian aplikasi dan optimasi akhir untuk memastikan kinerja yang stabil.
7	Minggu ke-7	Dokumentasi akhir dan pengumpulan.