

# **Klasifikasi Penyakit Diabetes Menggunakan Metode K-Nearest Neighbor (KNN)**

Disusun dalam rangka memenuhi tugas mata kuliah Penambangan Data  
Dosen pengampu : Abu Salam, M.Kom



Oleh :

❖ Nama : FATU RAHMAT  
❖ NIM : A11.2022.14831  
❖ KELP : A11.4504~Penambangan Data

**PROGRAM STUDI TEKNIK INFORMATIKA**

**FAKULTAS ILMU KOMPUTER**

**UNIVERSITAS DIAN NUSWANTORO**

**SEMARANG**

**2024**

## A. JUDUL

Klasifikasi Penyakit Diabetes Menggunakan Metode K-Nearest Neighbor (KNN)

## B. ABSTRACT

Penelitian ini menggunakan metode K-Nearest Neighbor (KNN) untuk mengklasifikasikan penyakit diabetes berdasarkan beberapa parameter kesehatan. Dataset yang digunakan memuat informasi medis seperti kadar glukosa, tekanan darah, BMI, dan faktor-faktor lainnya. Model KNN dipilih karena kemampuannya dalam melakukan klasifikasi berdasarkan kedekatan karakteristik data, serta kemudahannya dalam implementasi dan interpretasi hasil.

## C. MASALAH DAN TUJUAN

### Masalah:

- Sulitnya mendeteksi diabetes secara dini karena banyaknya parameter yang harus dipertimbangkan
- Perlunya sistem yang dapat membantu tenaga medis dalam melakukan diagnosis awal diabetes
- Pentingnya mengidentifikasi faktor-faktor yang berpengaruh dalam diagnosis diabetes

### Tujuan:

- Membangun model klasifikasi diabetes menggunakan algoritma KNN

## C. TAHAPAN

### • Data Preparation:

- Import library yang dibutuhkan (numpy, pandas, seaborn)
- Load dataset diabetes
- Analisis struktur data (2460 data dengan 9 atribut)
- Pengecekan missing value

### • Preprocessing Data:

- Pemisahan fitur (X) dan target (y)
- Pembagian data training (80%) dan testing (20%)
- Normalisasi data menggunakan RobustScaler

### • Model Development (Decision Tree):

- Implementasi algoritma KNN

- Pencarian nilai k optimal(1-15 neighbor)
- Training model dengan data yang telah diprocessing
- Model Evaluation
  - Perhitungan accuracy score (92.68%)
  - Pembuatan confusion matrix
  - Analisis classification report (precision, recall, f1-score)
  - Perhitungan metrik evaluasi tambahan (TPR, FPR, specificity)
  - Cross validation dengan 10-fold

## D. Dataset

Dataset `dataset-diabetes.csv` memiliki beberapa atribut penting yang dapat membantu dalam melatih dan testing dalam klasifikasi penyakit diabetes menggunakan algoritma KNN. Berikut adalah penjelasan masing-masing atribut dalam dataset ini:

1. Pregnancies: Jumlah kehamilan
2. Glucose: Kadar glukosa
3. BloodPressure: Tekanan darah
4. SkinThickness: Ketebalan kulit
5. Insulin: Kadar insulin
6. BMI: Indeks massa tubuh
7. DiabetesPedigreeFunction: Riwayat diabetes dalam keluarga
8. Age: Usia
9. Outcome: Status diabetes (0: Tidak diabetes, 1: Diabetes)

Atribut-atribut ini akan dianalisis untuk memahami faktor-faktor yang paling memengaruhi keberhasilan mahasiswa MBA berdasarkan metode Decision Tree.

## F. HASIL DAN KESIMPULAN

Hasil evaluasi model menunjukkan:

- Accuracy: 92.68%
- Precision: 94.75%
- Recall: 93.53%
- Specificity: 91.26%
- Cross-validation score rata-rata: 91.51%

Model KNN yang dibangun menunjukkan performa yang baik dalam mengklasifikasikan penyakit diabetes dengan tingkat akurasi di atas 90%. Model ini dapat digunakan sebagai alat bantu diagnosis awal untuk mendeteksi potensi diabetes pada pasien

