**1. Judul**

**Prediksi Penyakit Diabetes dengan Metode K-Nearest Neighbor (KNN) dan Seleksi Fitur Information Gain**

**2. Penulis**

* Devian
* Puspita Nurul Sabrina
* Agus Komarudin

**3. Afiliasi dan Korespondensi**

* Program Studi Teknik Informatika, Universitas Jenderal Achmad Yani  
  Alamat: Jl. Terusan Jend. Sudirman, Cibeber, Cimahi Selatan, Jawa Barat  
  Email: devian20@if.unjani.ac.id

**4. Jurnal / Seminar**

* *JATI (Jurnal Mahasiswa Teknik Informatika)*  
  Volume 8 No. 6, Desember 2024

**5. Abstrak**

* **Latar belakang**: Diabetes melitus mempengaruhi lebih dari 422 juta orang secara global. Deteksi dini sangat penting namun sering terlambat karena keterbatasan sumber daya medis.
* **Tujuan**: Meningkatkan akurasi prediksi KNN melalui seleksi fitur dengan *Information Gain*.
* **Metode dan Data**:
  + Dataset dari Kaggle ("Diabetes Dataset Prediction")
  + Pengujian dilakukan pada empat skenario rasio (90:10, 80:20, 70:30, 60:40).
  + Dibandingkan model KNN dengan dan tanpa seleksi fitur.
* **Hasil**:
  + KNN tanpa seleksi fitur: Akurasi tertinggi 69,11%
  + KNN dengan Information Gain: Akurasi tertinggi 72,93% (K=17)
* **Kata Kunci**: KNN, data mining, diabetes, dataset, Information Gain

**6. Bab 1 - Pendahuluan**

* **Masalah umum**: Diabetes menyebabkan kerusakan jangka panjang pada organ jika tidak terdiagnosis dini.
* **Masalah utama**: Keterbatasan dokter vs pasien → perlunya bantuan teknologi (ML/KNN).
* **Penelitian lain**:
  + Penerapan *Decision Tree* di RSUP Sardjito dengan akurasi 88,42%.
  + Penggunaan *Information Gain* untuk seleksi fitur pada dataset kesehatan.
* **Kontribusi**: Meningkatkan performa KNN dengan menerapkan metode seleksi fitur Information Gain.

**7. Sitasi (Referensi)**

* Menggunakan **penomoran** [1] hingga [11] → **harus disusun berdasarkan urutan kemunculan dalam teks**.

**8. Bab 2 - Tinjauan Pustaka**

* **Teori kasus**: Diabetes Melitus, penyebab, dampak, dan faktor risiko.
* **Teori metode**:
  + Data Mining dan proses KDD.
  + Algoritma KNN: cara kerja, perhitungan jarak (Euclidean).
  + Seleksi fitur dengan Information Gain: digunakan untuk mengurangi atribut tidak relevan.
  + Evaluasi menggunakan confusion matrix (akurasi, presisi, recall).

**9. Bab 3 - Metode**

* **Data**: Dataset dari Kaggle (Pima Indian Diabetes).
* **Alur sistem**:
  1. Preprocessing data
  2. Seleksi fitur menggunakan Information Gain
  3. Pengujian dengan algoritma KNN pada 4 rasio data
  4. Evaluasi hasil menggunakan confusion matrix
* **Diskusi**:
  1. Penggunaan Information Gain efektif dalam menyaring atribut yang relevan.
  2. Peningkatan akurasi menunjukkan seleksi fitur membantu klasifikasi.

**10. Bab 4 - Kesimpulan dan Saran**

* **Kesimpulan**:  
  Penerapan Information Gain meningkatkan akurasi KNN. Kombinasi seleksi fitur dan algoritma klasifikasi memberikan hasil lebih baik dibanding hanya KNN.
* **Saran**:  
  Dapat diperluas dengan algoritma lain atau model ensembel untuk perbandingan lebih lanjut.