

**Topik:** Analisis Prediksi Risiko Diabetes Menggunakan Algoritma C4.5 pada Dataset Kaggle

**Datasets:** <https://www.kaggle.com/datasets/salihacur/diabetes>

## Pendahuluan

Diabetes merupakan penyakit kronis yang prevalensinya terus meningkat secara global. Deteksi dini sangat penting untuk mencegah komplikasi serius dan membantu pengambilan keputusan medis secara lebih cepat dan efisien. Seiring perkembangan teknologi, metode *data mining* seperti algoritma C4.5 mulai banyak digunakan untuk menganalisis data klinis dan memprediksi kemungkinan pasien menderita diabetes. C4.5 menjadi populer karena kemampuannya membentuk pohon keputusan yang mudah diinterpretasikan oleh praktisi medis.

## Tinjauan Literatur

Penelitian oleh **Safitri dan Praba (2024)** menggunakan algoritma C4.5 untuk memprediksi penyakit diabetes berdasarkan dataset dari Sylhet Diabetes Hospital. Dengan 520 data pasien, mereka membangun model klasifikasi menggunakan RapidMiner dan validasi silang 10-fold. Hasilnya menunjukkan **akurasi sebesar 96,16%**, dengan precision dan recall yang tinggi. Mereka juga mengidentifikasi atribut *gender* sebagai faktor paling berpengaruh dalam klasifikasi, sedangkan *usia* tidak terlalu signifikan.

Studi lain oleh **Fida Maisa Hana (2020)** juga menerapkan algoritma C4.5 pada dataset yang sama, dan menunjukkan akurasi lebih tinggi yaitu **97,12%**. Penelitian ini menambahkan analisis Area Under Curve (AUC) yang mencapai 0,994, termasuk dalam kategori “excellent classification”. Data dibagi menjadi 80% untuk pelatihan dan 20% untuk pengujian. Hasil recall class positive mencapai 100%, menunjukkan kemampuan model dalam mendeteksi kasus diabetes dengan sangat baik.

Sementara itu, **Noviandi (2018)** menggunakan dataset PIMA Indian Diabetes sebanyak 768 data. Peneliti menerapkan algoritma C4.5 dengan validasi silang, dan memperoleh akurasi **70,32%**. Penelitian ini menghasilkan sembilan rule klasifikasi, tetapi hasil akurasinya tergolong rendah. Tidak ada visualisasi decision tree atau pembahasan preprocessing data secara mendalam.

## **Analisis Kritis**

Dari ketiga studi di atas, terlihat bahwa algoritma C4.5 memberikan hasil yang sangat baik ketika digunakan pada data yang bersih dan representatif seperti data dari Sylhet Diabetes Hospital. Penelitian oleh Safitri dan Hana menunjukkan akurasi yang tinggi ( $>96\%$ ) dan evaluasi yang lengkap. Namun, kedua penelitian tersebut tidak melakukan perbandingan dengan algoritma lain seperti Naïve Bayes atau SVM.

Sebaliknya, studi oleh Noviandi yang menggunakan dataset PIMA Indian memiliki akurasi yang lebih rendah karena kemungkinan besar tidak dilakukan proses preprocessing secara optimal. Selain itu, visualisasi hasil dan evaluasi tambahan seperti AUC atau f-measure tidak disediakan.

Secara umum, metode validasi silang sering digunakan pada ketiga penelitian ini, dan RapidMiner menjadi tools utama dalam eksperimen. Meskipun hasil dari dua penelitian sangat tinggi, celah penelitian tetap ada dalam bentuk perbandingan algoritma serta eksplorasi atribut lain yang lebih mendalam.

## **Kesimpulan**

Algoritma C4.5 terbukti efektif dan stabil dalam melakukan klasifikasi penyakit diabetes, terutama ketika digunakan dengan dataset yang bersih dan memiliki atribut yang relevan. Untuk hasil optimal, penting untuk melakukan preprocessing data, memilih atribut yang signifikan, serta menggunakan teknik validasi yang tepat. Penelitian selanjutnya disarankan untuk membandingkan C4.5 dengan algoritma lain dan menguji hasilnya pada data nyata dari institusi kesehatan lokal. Selain itu, penggabungan dengan teknik ensemble juga dapat menjadi peluang pengembangan.

## **REFERENSI**

- Hana, F. M. (2020). Klasifikasi Penderita Penyakit Diabetes Menggunakan Algoritma Decision Tree C4.5. *Jurnal Teknologi Informasi dan Komputer*.
- Noviandi, F. (2018). IMPLEMENTASI ALGORITMA DECISION TREE C4.5 UNTUK PREDIKSI PENYAKIT DIABETES. *Jurnal Sistem dan Teknologi Informasi*.

Safitri, M. &. (2024). PREDIKSI PENYAKIT DIABETES DENGAN MENGGUNAKAN ALGORITMA C4.5. *JIKA (Jurnal of Informatics)*, Universitas Muhammadiyah Tangerang.