

# IMPLEMENTASI ALGORITMA *INTUITIONISTIC FUZZY DECISION TREE* UNTUK IDENTIFIKASI FAKTOR BERPENGARUH PENYAKIT DIABETES PADA PEREMPUAN

<sup>1</sup>Anis Fitri Nur Masruriyah

<sup>2</sup>Hilda Yulia Novita

Teknik Informatika, Fakultas Teknik dan Ilmu Komputer,  
Universitas Buana Perjuangan Karawang  
anis.masruriyah@ubpkarawang.ac.id<sup>1</sup>

## ABSTRAK

Diabetes menjadi salah satu penyakit yang mematikan. Penyakit ini tidak menular, namun dapat diturunkan pada anak dari penderita diabetes. Secara umum, diabetes dapat dideteksi pada pasien usia produktif. Jika tidak ditangani dengan baik, maka akan terjadi peningkatan risiko penyakit penyerta. Di antara berbagai penyakit, kerusakan ginjal, penyakit jantung hingga kerusakan syaraf adalah penyakit penyerta dari diabetes. Naskah ilmiah ini, dibuat untuk membangun model klasifikasi dengan *metode Intuitionistic Fuzzy Decision Tree* sehingga mampu membantu untuk melakukan diagnosa lebih awal. Hasil adari penelitian ini terdapat lima variabel utama yang mempengaruhi penyakit diabetes. Variabel tersebut dapat digunakan sebagai acuan untuk melakukan diagnosa dini pasien diabetes.

Kata kunci: *Decision Tree*, Diabetes, *Intuitionistic Fuzzy*,

## PENDAHULUAN

Diabetes diketahui sebagai penyakit tidak menular, namun menjadi salah satu penyakit mematikan karena jumlah pasien diabetes terus meningkat setiap tahun (World Health Organization, 2021). Secara umum, diabetes dibagi menjadi dua jenis, yaitu diabetes tipe 1 (autoimun) dan diabetes tipe 2 (melitus) (Maharani et al., 2019; Rahatara Ferdousi et al., 2019; Sulistijo, 2019). Penelitian juga menunjukkan bahwa ada jenis diabetes yang bisa sembuh total dan hanya ditemukan pada perempuan hamil disebut dengan diabetes gestasional (Mutmainna, 2019; Sulistijo, 2019). Diabetes gestasional ini pada beberapa kasus dapat sembuh setelah proses persalinan. Beberapa ciri umum diabetes adalah sering merasa haus, kemudian pada malam hari intensitas buang air kecil meningkat dan sering merasa sangat lapar (Walker, 2020). Pada beberapa pasien terjadi penurunan berat badan yang signifikan, merasa lemah bahkan luka yang cenderung sembuh lebih lama. Namun, pada wanita hamil gejala berbeda dan ada beberapa factor risiko yang menyertai.

Sebelumnya, beberapa penelitian melakukan klasifikasi dari penyakit diabetes (Dutta et al., 2019; Geman et al., 2017; Lakhwani et al., 2020; Sisodia & Sisodia, 2018; Zou et al., 2018). Penelitian yang dilakukan oleh Geman *et al.* (2017) menggunakan 768 data pasien diabetes untuk melakukan klasifikasi penyakit diabetes. Metode yang digunakan untuk melakukan klasifikasi

adalah *Adaptive Neuro-Fuzzy Inference System* (ANFIS). Proses klasifikasi dimulai dengan mengatur fitur vektor menggunakan fungsi diabetes pidedegree untuk menentukan basis aturan dari fuzzy dengan beberapa premis. Selanjutnya, model ANFIS dievaluasi menggunakan spesifitas, presisi dan sensitifitas sehingga hasil klasifikasi mencapai 84.27%. Kemudian, Sisodia dan Sisodia (2018) juga membangun model klasifikasi pada data dari *Pima Indians Diabetes Database* (PIDD) dengan total delapan variabel dan dua kelas. Model klasifikasi yang dibangun membandingkan algoritma Pohon Keputusan, *Support Vector Machine* (SVM) dan Naïve Bayes. Model dari hasil penelitian tersebut menunjukkan bahwa Naïve Bayes memiliki akurasi tertinggi dibanding algoritma yang lain.

Selanjutnya, Zou et al. (2018) menerapkan metode *Principal Component Analysis* (PCA) dan *minimum redundancy maximum relevance* (mRMR) untuk mengurangi dimensi dari data sebelum proses klasifikasi. Data yang digunakan pada penelitian ini diperoleh dari Rumah Sakit Lhou di Cina dengan total 164431 data pasien dan 14 variabel. Hasil klasifikasi mencapai 80.84% dengan metode klasifikasi yang digunakan adalah *Random Forest*. Kemudian Dutta et al. (2019) pada penelitiannya melakukan fitur ekstraksi untuk menemukan variabel yang mempengaruhi penyakit diabetes. Dalam penelitian tersebut, lima variabel yang mempengaruhi penyakit diabetes adalah kadar gula dalam darah, insulin, usia, indeks masa tubuh dan riwayat diabetes. Setelah mendapatkan variabel yang berpengaruh, kemudian dilakukan klasifikasi menggunakan *random forest* dengan hasil klasifikasi sebesar 84%. Setelah itu, Lakhwani et al. (2020) menggunakan algoritma Jaringan Syaraf Tiruan (JST) untuk melakukan prediksi penyakit diabetes. Penelitian tersebut membangun model prediksi dengan menggunakan fungsi *logistic-activation* sebagai fungsi aktivasi, selanjutnya menggunakan metode Quasi Newton sebagai algoritma pelatihan. Model yang diusulkan mencapai keberhasilan sebesar 55%.

Berdasarkan penelitian yang telah dilakukan, terbukti bahwa berbagai teknik klasifikasi mampu membangun model untuk membantu diagnosis dan prediksi penyakit diabetes. Namun, penelitian tersebut belum ada yang menghitung derajat keragu-raguan pada setiap variabel yang mempengaruhi penyakit diabetes. Sehingga, penelitian ini membangun model dengan *Intuitionistic Fuzzy Decision Tree*. Detail selanjutnya pada bagian Metode Penelitian akan membahas cara kerja Intuitionistic fuzzy Decision Tree, selanjutnya hasil akan dibahas pada bagian Hasil. Kemudian, kesimpulan akan dibahas dibagian akhir.

## **METODE PENELITIAN**

### ***C50 Algorithm***

Secara umum pohon keputusan terdiri dari akar, simpul awal dan simpul akhir. Sama seperti algoritma pohon keputusan yang lain, pada *Intuitionistic Fuzzy Decision Tree* juga mencari nilai entropi dan *information gain*. Namun, perhitungan entropi pada *Intuitionistic Fuzzy Decision Tree* menggunakan *Intuitionistic Fuzzy Entropy* yang ditunjukkan pada persamaan 1. Di mana M dan N merupakan elemen dari *Intuitionistic Fuzzy* ( $\langle \mu, \nu, \pi \rangle$ ) yang menunjukkan fungsi keanggotaan (M) dan bukan fungsi keanggotaan (N). Kemudian,  $\mu$  adalah fungsi keanggotaan, lalu  $\nu$  merepresentasikan non fungsi keanggotaan dan  $\pi$  menunjukkan derajat keragu-raguan. Selanjutnya, dengan menggunakan pengembangan dari Hamming *Distance*, maka  $H_{IFS}$  merupakan nilai tertinggi dari fungsi keanggotaan, non-fungsi keanggotaan dan derajat keragu-raguan. Persamaan 2 dan 3 adalah cara untuk mendapatkan nilai  $H_{IFS}$ . Setelah mendapatkan nilai entropi, langkah selanjutnya adalah menghitung *Information Gain* sesuai persamaan 4 di mana  $E_{IFS(x)}$  adalah nilai entropi dari variabel yang memiliki pengaruh.

$$E_{IFS} = \frac{1}{n} \sum_{n=1}^k \left( \frac{\min\{H_{IFS}(x, M), H_{IFS}(x, N)\}}{\max\{H_{IFS}(x, M), H_{IFS}(x, N)\}} \right) \quad (1)$$

$$H_{IFS}(x, M) = \frac{1}{2} (|\mu_x - 1| + |\nu_x - 0| + |\pi_x - 0|) \quad (2)$$

$$H_{IFS}(x, N) = \frac{1}{2} (|\mu_x - 0| + |\nu_x - 1| + |\pi_x - 0|) \quad (3)$$

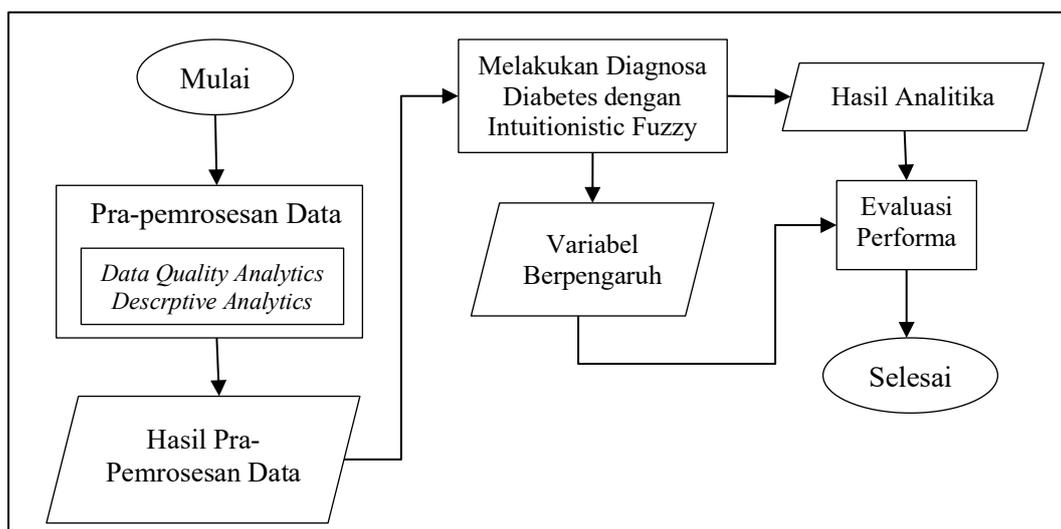
$$Information\ Gain = E_{IFS} - E_{IFS(x)} \quad (4)$$

### Waktu dan Tempat Penelitian

Penelitian ini dimulai sejak Januari 2021 hingga Oktober 2021 di laboratorium Riset pada Universitas Buana Perjuangan.

### Prosedur Penelitian

Proses penelitian ditunjukkan pada Gambar 1 yang dimulai dengan melakukan pra-pemrosesan data. Tahap pra-pemrosesan data dilakukan untuk menghilangkan *outlier* dan menyelesaikan masalah *missing value*. Setelah data siap untuk digunakan, selanjutnya melakukan diagnosa penyakit diabetes menggunakan *Intuitionistic Fuzzy Decision Tree*. Hasil dari proses *Intuitionistic Fuzzy Decision Tree* berupa variabel berpengaruh dan juga model klasifikasi untuk diagnosa penyakit diabetes.



Gambar 1 Alur Prosedur Penelitian

### Data, Instrumen, dan Teknik Pengumpulan Data

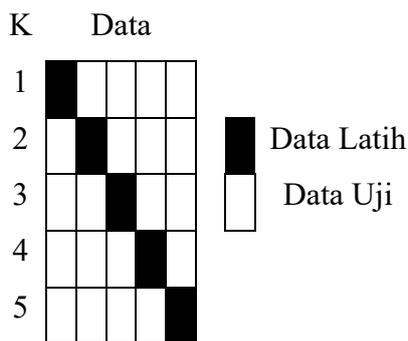
Data yang digunakan merupakan data kuantitatif berdasarkan responden 768 wanita dengan 8 variabel dan kelas diabetes positif dan negatif. Variabel terdiri dari usia dengan rentang usia 21 tahun hingga 81 tahun, kemudian riwayat hamil, glukosa, tekanan darah, pengukuran lemak tubuh, insulin, indeks masa tubuh dan riwayat diabetes pada keluarga.

### Teknik Analisis Data

Analisis data dimulai dengan melakukan pra-pemrosesan data untuk mengisi *missing value*. Teknik yang digunakan untuk menyelesaikan masalah *missing value* adalah mengisi dengan rata-rata data. Sehingga, masalah *outlier* juga dapat diatasi. Setelah data siap untuk diolah, kemudian mengimplementasikan algoritma *Intuitionistic Fuzzy* pada seluruh data. Jika simpul pada pohon memenuhi kondisi  $D$ , maka simpul dijadikan akar dari pohon keputusan. Kondisi ini memiliki syarat yang harus dipenuhi, yaitu proporsi dari kelas  $C_k$  lebih besar sama dengan  $\theta_r$ ,  $\frac{|D^{ci}|}{|D|} \geq \theta_r$ , kemudian total data harus kurang dari  $\theta_n$ . Selanjutnya, tidak boleh ada sub atribut yang menjadi klasifikasi lain, jika semua kondisi terpenuhi maka pembuatan pohon berhenti, lalu dipangkas. Namun, jika simpul tidak memenuhi kondisi  $D$ , maka semua frekuensi kelas dihitung, lalu menurunkan fungsi *Intuitionistic Fuzzy* dan menghitung nilai entropinya. Selanjutnya, memilih nilai *Information Gain* tertinggi sebagai akar simpul dan menurunkan anak simpul untuk setiap *Information Gain*. Lalu semua proses diulang secara rekursif hingga kondisi  $D$  terpenuhi.

Setelah seluruh proses pembentukan model klasifikasi selesai, kemudian dilakukan evaluasi kinerja. Pada penelitian ini, evaluasi menggunakan teknik *K-Fold Cross Validation*.

Cara kerja dari Teknik *K-Fold Cross Validation* adalah dengan membagi data menjadi data uji dan data latih sebanyak K. Ilustrasi untuk teknik ini ditunjukkan pada Gambar 2.



Gambar 2 Ilustrasi *K-Fold*

**HASIL PENELITIAN DAN PEMBAHASAN**

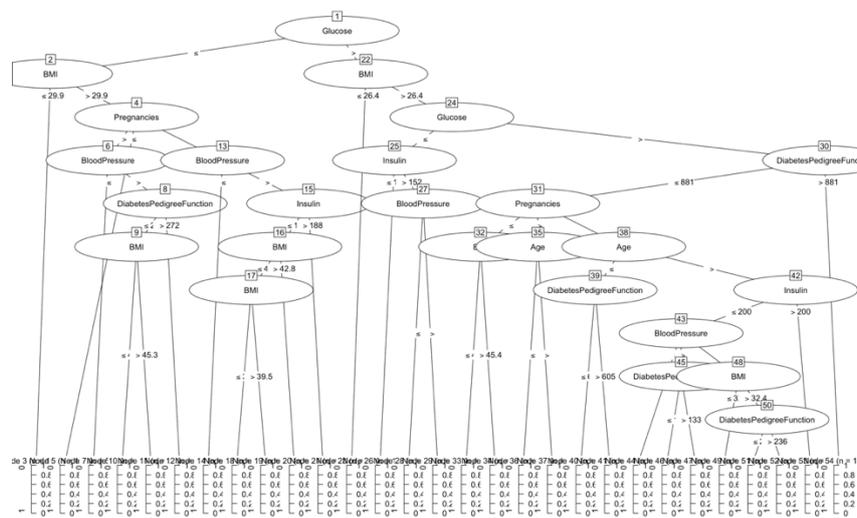
**Hasil Penelitian**

Berdasarkan hasil model *diagnostic analytics* yang dibangun menggunakan algoritme *Intuitionistic Fuzzy based on Decision Tree* menunjukkan nilai akurasi menggunakan Persamaan 7 sebesar 70.12%. Nilai terbaik ini ditunjukkan pada fold ke-10 dengan hasil matriks konfusi ditunjukkan pada Tabel 1.

Tabel 1 Matriks Konfusi

Prediksi	Asli	
	A	B
A	419	35
B	70	167

Kemudian, hasil pohon keputusan ditunjukkan pada Gambar 3. Berdasarkan Gambar 4.3 nampak bahwa kadar glukosa dan indeks masa tubuh berada pada simpul paling atas. Hal ini menandakan bahwa dua variabel tersebut mempengaruhi pasien diabetes secara signifikan.



Gambar 3 Model Pohon Keputusan Pasien Diabetes

Pada Gambar 3 lima variabel yang mempengaruhi pasien diabetes terdiri dari kadar glukosa, indeks masa tubuh, riwayat kehamilan, riwayat diabetes dalam keluarga, dan kadar insulin.

## KESIMPULAN DAN IMPLIKASI

Berdasarkan hasil penelitian dapat disimpulkan bahwa algoritma *Intuitionistic Fuzzy Decision Tree* mampu membuat model klasifikasi dengan baik dengan akurasi sebesar 70.12%. Evaluasi model klasifikasi diperoleh hasil *K-Fold Cross Validation*. Selanjutnya, variabel yang mempengaruhi penyakit diabetes pada wanita adalah kadar glukosa, indeks masa tubuh, riwayat kehamilan, riwayat diabetes dalam keluarga, dan kadar insulin.

Selanjutnya, pengembangan model penelitian ini adalah perlu penambahan variabel nutrisi dan informasi pola hidup. Sehingga, tenaga kesehatan dapat memberikan petunjuk agar kondisi pasien menjadi lebih baik.

## DAFTAR PUSTAKA (TIMES NEW ROMAN (TNR-12) BOLD)

- Dutta, D., Paul, D., & Ghosh, P. (2019). Analysing Feature Importances for Diabetes Prediction using Machine Learning. *2018 IEEE 9th Annual Information Technology, Electronics and Mobile Communication Conference, IEMCON 2018, September 2019*, 924–928. <https://doi.org/10.1109/IEMCON.2018.8614871>
- Geman, O., Chiuchisan, I., & Todorean, R. (2017). Application of Adaptive Neuro-Fuzzy Inference System for diabetes classification and prediction. *2017 E-Health and Bioengineering Conference, EHB 2017, Dm*, 639–642. <https://doi.org/10.1109/EHB.2017.7995505>
- Lakhwani, K., Bhargava, S., Hiran, K. K., Bundele, M. M., & Somwanshi, D. (2020). Prediction of the Onset of Diabetes Using Artificial Neural Network and Pima Indians Diabetes Dataset. *2020 5th IEEE International Conference on Recent Advances and Innovations in Engineering, ICRAIE 2020 - Proceeding, March 2021*. <https://doi.org/10.1109/ICRAIE51050.2020.9358308>
- Maharani, M., Latipun, & Djudiyah. (2019). Pelatihan Manajemen Diri untuk Meningkatkan Status Kesehatan Psikologis Penderita Diabetes Melitus. In *Psychology Forum*.
- Mutmainna, A. (2019). Faktor-Faktor Risiko yang Mempengaruhi Manajemen Glukosa pada Pasien dengan Diagnosa Medis Diabetes Mellitus di Makassar, Sulawesi Selatan, Indonesia. *Nursing Inside Community*, 1(2), 61–67. <https://doi.org/10.35892/nic.v1i2.59>
- Rahatara Ferdousi, S. R. M. M. F. I., Bushra, & Yasmin, H. (2019). Early stage diabetes risk

- prediction dataset. Data Set. In *UCI Machine Learning Repository*.  
<https://archive.ics.uci.edu/ml/datasets/Early+stage+diabetes+risk+prediction+dataset>.
- Sisodia, D., & Sisodia, D. S. (2018). Prediction of Diabetes using Classification Algorithms. *Procedia Computer Science*, 132(Iccids), 1578–1585.  
<https://doi.org/10.1016/j.procs.2018.05.122>
- Sulistijo, S. A. (2019). Pedoman Pengelolaan dan Pencegahan Diabetes Melitus Tipe 2 Dewasa di Indonesia. In *PB Perkeni* (Vol. 19, Issue 4).
- Walker, R. (2020). *Take Control of Your Diabetes*.
- World Health Organization. (2021). *Diabetes* (p. 1). [https://www.who.int/health-topics/diabetes#tab=tab\\_1](https://www.who.int/health-topics/diabetes#tab=tab_1)
- Zou, Q., Qu, K., Luo, Y., Yin, D., Ju, Y., & Tang, H. (2018). Predicting Diabetes Mellitus With Machine Learning Techniques. *Frontiers in Genetics*, 9.  
<https://doi.org/10.3389/fgene.2018.00515>