PT Jurnal Cendekia Indonesia
Journal of Information Technology and Informatics Engineering
Homepage: https://journal.jci.co.id/jitie
Vol. 1 No. 1 (2025) pp: 30-36
P-ISSN: XXXX-XXXX, e-ISSN: XXXX-XXXX

PREDIKSI DIABETES BERDASARKAN FAKTOR MEDIS PASIEN MENGGUNAKAN ALGORITMA *DECISION TREE*

[1]Ahmad Reza Farrasy, [2]Amar Naufal Al-kharits, [3]Muhamad Bustomi, [4]Nazar Maulana, [5]Taupik Abdul Rahman

[1]Sistem Informasi, Ilmu Komputer, Universitas Pamulang, Kota Tangerang Selatan, Indonesia
[2]Sistem Informasi, Ilmu Komputer, Universitas Pamulang, Kota Tangerang Selatan, Indonesia
[3]Sistem Informasi, Ilmu Komputer, Universitas Pamulang, Kota Tangerang Selatan, Indonesia
[4]Sistem Informasi, Ilmu Komputer, Universitas Pamulang, Kota Tangerang Selatan, Indonesia
[5]Sistem Informasi, Ilmu Komputer, Universitas Pamulang, Kota Tangerang Selatan, Indonesia

[1]Farrasyreza1@gmail.com, [2]Naufalalkharits11@gmail.com, [3]Bustomimuhammad03@gmail.com,
[4]Nazarmaulana916@gmail.com, [5]Taupikabdulrahman96271@gmail.com

Abstract

Early detection of diabetes risk is crucial to prevent severe complications. This study develops a predictive model for diabetes using the Decision Tree algorithm based on patient medical data. The dataset consists of 768 records with eight health-related attributes, of which 99 labeled instances are used to train the model. The process includes data cleaning, target attribute assignment, and model construction using RapidMiner. Results indicate that variables such as age and glucose levels significantly influence diabetes classification. Although the initial findings show promising potential, further validation with larger and more balanced datasets is needed to improve the model's accuracy.

Keywords: *diabetes, decision tree, data mining, classification, RapidMiner*

Abstrak

Pendeteksian dini terhadap risiko diabetes sangat penting untuk mencegah komplikasi serius. Penelitian ini mengembangkan model prediksi diabetes menggunakan algoritma *Decision Tree* berdasarkan data medis pasien. Dataset yang digunakan mencakup 768 data pasien dengan delapan atribut kesehatan, di mana 99 data berlabel digunakan untuk pelatihan model. Proses dilakukan melalui tahapan pembersihan data, penentuan atribut target, dan pembangunan model menggunakan RapidMiner. Hasil menunjukkan bahwa atribut seperti usia dan kadar glukosa memiliki pengaruh kuat terhadap klasifikasi diabetes. Meskipun hasil awal menunjukkan potensi besar, model masih memerlukan validasi lebih lanjut dengan data yang lebih besar dan distribusi label yang seimbang untuk meningkatkan keakuratannya.

Kata Kunci: diabetes, *decision tree*, data mining, klasifikasi, RapidMiner

A. PENDAHULUAN

Diabetes merupakan salah satu penyakit kronis yang prevalensinya terus meningkat di seluruh dunia. Penyakit ini dapat menyebabkan berbagai komplikasi serius seperti penyakit jantung, gagal ginjal, kebutaan, dan amputasi. Sayangnya, banyak penderita diabetes tidak menyadari kondisinya hingga memasuki tahap yang berbahaya. Oleh karena itu, pendeteksian dini terhadap risiko diabetes sangat penting untuk dilakukan guna mencegah dampak buruk yang lebih lanjut.

Kemajuan teknologi di bidang informasi dan kesehatan membuka peluang besar dalam mendukung deteksi dini penyakit secara otomatis. Salah satu pendekatan yang kini banyak digunakan adalah teknik information mining, yaitu metode penggalian informasi dari information besar untuk menemukan pola-pola yang tersembunyi. Dengan memanfaatkan information medis pasien, algoritma information mining dapat membantu memprediksi risiko penyakit berdasarkan karakteristik dan riwayat kesehatan pasien tersebut.

Salah satu algoritma yang populer dalam klasifikasi information adalah *Decision Tree*. Algoritma ini bekerja dengan membuat cabang-cabang keputusan berdasarkan nilai dari atribut information, hingga akhirnya mengarah ke kesimpulan atau prediksi tertentu. Kelebihan *Decision Tree* terletak pada kemampuannya menghasilkan demonstrate yang mudah dipahami oleh manusia karena bentuknya menyerupai logika keputusan.

Penerapan *Decision Tree* dalam kasus prediksi diabetes memungkinkan sistem untuk menganalisis berbagai faktor

Diterima Redaksi: 10-06-2025 | Selesai Revisi: 20-06-2025 | Diterbitkan Online: 30-06-2025

medis seperti kadar glukosa darah, tekanan darah, indeks massa tubuh (BMI), usia, dan parameter lainnya untuk menentukan apakah seorang pasien memiliki potensi menderita diabetes atau tidak. Dengan begitu, dokter atau tenaga kesehatan dapat memperoleh gambaran awal secara otomatis sebelum pemeriksaan lebih lanjut dilakukan.

Dalam penelitian ini, digunakan information sebanyak 768 baris pasien yang memuat delapan atribut medis. Dari jumlah tersebut, 99 information telah diberi *name* klasifikasi sebagai "*Yes*" (positif diabetes) dan "*No*" (negatif diabetes). Dengan *information* tersebut, dibangun demonstrate prediksi menggunakan algoritma *Decision Tree* di dalam perangkat lunak RapidMiner. Proses meliputi tahap *preprocessing information*, pelatihan *demonstrate*, dan prediksi terhadap *information* baru.

Masalah utama yang diangkat dalam penelitian ini adalah bagaimana mengembangkan sistem prediksi berbasis *information* medis yang dapat digunakan untuk membantu mendeteksi risiko diabetes secara otomatis dan cepat. Penelitian ini tidak hanya memfokuskan pada akurasi hasil, tetapi juga pada kemudahan interpretasi demonstrate dan efektivitasnya dalam konteks jumlah *information* yang terbatas.

Penelitian ini diharapkan dapat memberikan kontribusi dalam pengembangan sistem pendukung keputusan (*Decision support system*) di bidang kesehatan, khususnya dalam hal deteksi dini penyakit diabetes. Selain itu, penelitian ini juga menjadi studi awal untuk penerapan algoritma pembelajaran mesin (*machine learning*) dalam kasus nyata menggunakan computer program yang mudah diakses seperti RapidMiner.

Dengan adanya show prediksi ini, diharapkan proses skrining awal terhadap risiko diabetes bisa dilakukan secara lebih efisien, sehingga memungkinkan pencegahan dan penanganan lebih dini. Ke depan, pendekatan serupa dapat diterapkan pula pada penyakit-penyakit kronis lainnya yang memerlukan deteksi awal berdasarkan analisis *information* medis.

B. PELAKSANAAN DAN METODE

Algoritma *Decision Tree* merupakan salah satu metode pembelajaran mesin yang digunakan untuk menyelesaikan masalah klasifikasi maupun regresi. Pada dasarnya, *Decision Tree* bekerja dengan membuat struktur pohon yang terdiri dari simpul keputusan (*decision node*) dan simpul akhir (*leaf node*). Setiap simpul keputusan merepresentasikan kondisi logika atas atribut tertentu, yang akan membagi data ke dalam cabang-cabang yang lebih spesifik.

Proses pembentukan pohon dilakukan secara rekursif, di mana algoritma memilih atribut yang paling memberikan nilai informasi tinggi (*information gain*) untuk dijadikan akar pohon (*root node*), lalu membagi dataset sesuai kondisi atribut tersebut. Proses ini diulang hingga data habis atau mencapai kriteria tertentu, misalnya kedalaman pohon maksimal atau jumlah minimum data pada tiap cabang.

Keunggulan metode ini adalah kemudahan interpretasi karena hasilnya menyerupai logika "if-then", serta tidak memerlukan asumsi statistik tertentu. Namun, kelemahan utamanya adalah risiko overfitting jika pohon terlalu dalam atau data terlalu kecil.

Penelitian ini menggunakan dataset diabetes dari kaggle berisi information medis pasien, yang masing-masing memuat informasi terkait beberapa faktor kesehatan yang berkaitan erat dengan risiko diabetes. Setiap entri dalam dataset memiliki delapan atribut utama, yaitu: jumlah kehamilan (*Pregnancies*), kadar glukosa dalam darah (*Glucose*), tekanan darah diastolik (*BloodPressure*), ketebalan lipatan kulit (*SkinThickness*), kadar affront (*Affront*), indeks massa tubuh atau *Body Mass Record* (BMI), fungsi silsilah diabetes (*DiabetesPedigreeFunction*), serta usia pasien (*Age*).

Atribut-atribut tersebut dipilih karena secara klinis dianggap memiliki pengaruh signifikan terhadap kemungkinan seseorang menderita diabetes. Selain itu, terdapat satu atribut target atau name yang disebut dengan "*Outcome*", yang berfungsi sebagai indikator hasil klasifikasi. Name ini memiliki dua nilai kategorikal, yakni "*Yes*" yang menunjukkan pasien positif diabetes, dan "*No*" untuk pasien yang negatif diabetes. Dari seluruh *information* yang tersedia, hanya sebanyak 99 *information* yang telah memiliki name Result dan digunakan sebagai information pelatihan (*preparing set*), sedangkan sisanya digunakan untuk prediksi (*unlabelled information*).

- Tahapan Processing

Sebelum *information* digunakan untuk membangun show prediksi, dilakukan beberapa tahapan preprocessing guna memastikan bahwa *information* dalam kondisi bersih, relevan, dan siap diproses secara algoritmik. Tahap awal dimulai dengan menghapus *information* kosong atau tidak substantial, khususnya pada atribut yang bersifat numerik dan kritis seperti *Glucose* atau BMI, karena nilai-nilai kosong dapat memengaruhi hasil analisis dan akurasi show.

Langkah selanjutnya adalah menentukan peran dari setiap atribut. Dalam konteks RapidMiner, proses ini dilakukan menggunakan administrator "*Set Role*" yang menetapkan kolom "*Outcome*" sebagai atribut *name* atau target klasifikasi. Penetapan name ini sangat penting karena menjadi dasar bagi algoritma *Decision Tree* untuk mempelajari pola dari information berlabel.

Setelah itu, dilakukan penyaringan *information*, yaitu memisahkan *information* yang sudah memiliki *name Result* sebanyak 99 baris untuk digunakan dalam proses pelatihan *demonstrate*. Sementara itu, *information* lainnya yang tidak memiliki *name Result* digunakan sebagai *information*

prediksi, yang hasil klasifikasinya akan ditentukan oleh *demonstrate* yang sudah dilatih sebelumnya.

Proses ini membantu membedakan antara *information* yang digunakan untuk membentuk *demonstrate* dan *information* yang nantinya akan diprediksi, sehingga tidak terjadi kebocoran *information* (*information spillage*) yang dapat menurunkan kualitas hasil analisis.

- Algoritma dan Alat

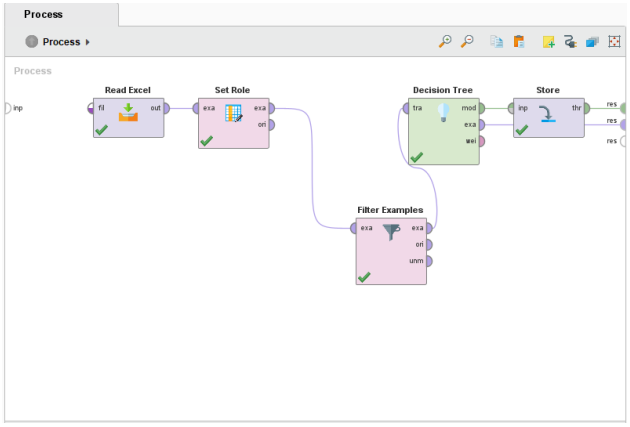
Algoritma utama yang digunakan dalam penelitian ini adalah *Decision Tree*, yang merupakan salah satu metode klasifikasi yang *withering* mudah dipahami dan banyak digunakan dalam analisis *information*. *Decision Tree* bekerja dengan cara membagi *information* berdasarkan atribut yang *withering* memberikan informasi (*data pick up*) untuk memisahkan kelas-kelas *information*. Proses pembentukan pohon keputusan dilakukan secara rekursif hingga terbentuk aturan-aturan klasifikasi yang dapat digunakan untuk menentukan hasil prediksi terhadap *information* baru.

Untuk mendukung seluruh proses analisis, digunakan perangkat lunak RapidMiner, sebuah *stage open-source* yang menyediakan antarmuka grafis (GUI) berbasis *drag-and-drop* sehingga sangat memudahkan dalam membangun dan memvisualisasikan alur proses *information mining*. RapidMiner juga menyediakan berbagai administrator dan algoritma klasifikasi, termasuk *Decision Tree*, yang dapat langsung digunakan tanpa perlu pemrograman manual.

Dengan kombinasi algoritma *Decision Tree* dan kemudahan visualisasi dari RapidMiner, penelitian ini dapat merancang sistem prediksi diabetes yang sederhana namun fungsional, serta mudah dipahami oleh pengguna dari latar belakang non-teknis sekalipun.

C. HASIL DAN PEMBAHASAN

Rancangan Alur Proses Pemodelan
alur proses yang digunakan dalam pembangunan show prediksi diabetes di Altair AI Studio. Proses dimulai dari pengambilan information dalam *arrange Exceed expectations*, dilanjutkan dengan penetapan atribut target (*Result*) sebagai *name*, penyaringan *information* berdasarkan kondisi tertentu, pelatihan *demonstrate* menggunakan algoritma *Decision Tree*, dan penyimpanan hasil prediksi.

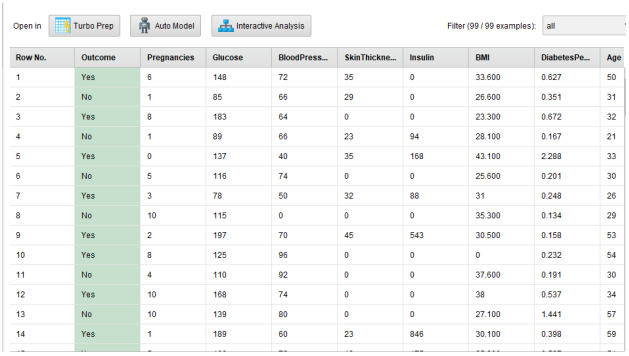


Pada Gambar di atas menunjukkan alur proses pembangunan model prediksi diabetes menggunakan perangkat lunak RapidMiner. Proses dimulai dari penggunaan operator Read Excel yang berfungsi untuk membaca dataset pasien dalam format Excel. Dataset ini berisi informasi medis dari pasien, seperti jumlah kehamilan, kadar glukosa, tekanan darah, dan atribut kesehatan lainnya. Data yang dibaca dari Excel kemudian diteruskan ke operator berikutnya untuk diproses lebih lanjut sesuai kebutuhan klasifikasi.

Langkah selanjutnya adalah penggunaan operator Set Role, yang bertugas menetapkan peran masing-masing atribut dalam dataset. Pada tahap ini, atribut "Result" diatur sebagai *label* atau *target class*, yang berarti atribut ini menjadi dasar bagi algoritma *Decision Tree* dalam belajar dan membentuk pola klasifikasi. Setelah peran ditentukan, data dilanjutkan ke operator Filter Examples, yang memisahkan data berlabel (sebanyak 99 baris) untuk keperluan pelatihan model, dari data yang tidak berlabel yang nantinya akan diprediksi. Proses ini penting agar tidak terjadi *data leakage* yang dapat menurunkan validitas model.

Operator *Decision Tree* digunakan untuk membangun struktur pohon klasifikasi berdasarkan data yang telah difilter. Model yang dihasilkan kemudian disimpan menggunakan operator Store agar dapat digunakan kembali untuk prediksi data baru di masa mendatang. Secara keseluruhan, proses ini mencerminkan pipeline yang sederhana namun efektif dalam melakukan klasifikasi risiko diabetes secara otomatis, dengan memanfaatkan antarmuka grafis dari RapidMiner yang intuitif dan mudah digunakan tanpa perlu menulis kode pemrograman secara manual.

Information yang Digunakan
Dataset yang digunakan terdiri dari 99 baris *information* pasien yang memuat 8 atribut medis serta 1 atribut *name* (*Result*). Setiap atribut merepresentasikan faktor-faktor medis yang mempengaruhi risiko diabetes, seperti jumlah kehamilan (*Pregnancies*), kadar glukosa (*Glucose*), tekanan darah (*BloodPressure*), dan lainnya.



Open in | Turbo Prep | Auto Model | Interactive Analysis | Filter (99 / 99 examples): all

Row No.	Outcome	Pregnancies	Glucose	BloodPress...	SkinThickne...	Insulin	BMI	DiabetesPe...	Age
1	Yes	6	148	72	35	0	33.600	0.627	50
2	No	1	85	66	29	0	26.600	0.351	31
3	Yes	8	183	64	0	0	23.300	0.672	32
4	No	1	89	66	23	94	28.100	0.167	21
5	Yes	0	137	40	35	168	43.100	2.288	33
6	No	5	116	74	0	0	25.600	0.201	30
7	Yes	3	78	50	32	88	31	0.248	26
8	No	10	115	0	0	0	35.300	0.134	29
9	Yes	2	197	70	45	543	30.500	0.158	53
10	Yes	8	125	96	0	0	0	0.232	54
11	No	4	110	92	0	0	37.600	0.191	30
12	Yes	10	168	74	0	0	38	0.537	34
13	No	10	139	80	0	0	27.100	1.441	57
14	Yes	1	189	60	23	846	30.100	0.398	59

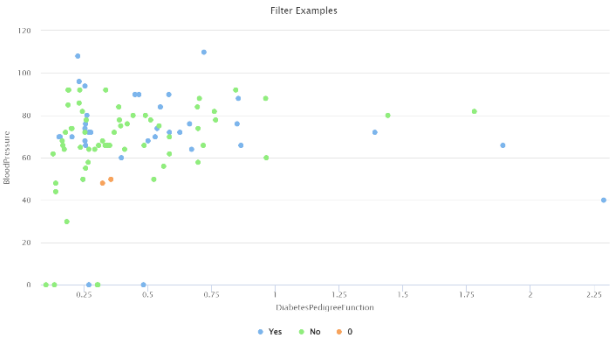
Gambar di atas memperlihatkan tampilan dataset hasil impor ke dalam RapidMiner yang terdiri dari 99 baris data medis pasien dengan atribut-atribut yang relevan untuk proses prediksi risiko diabetes. Setiap baris pada tabel merepresentasikan satu pasien, sedangkan kolom-kolomnya memuat informasi medis seperti *Pregnancies* (jumlah kehamilan), *Glucose* (kadar glukosa darah), *BloodPressure* (tekanan darah), *SkinThickness* (ketebalan lipatan kulit), *Insulin*, *BMI* (indeks massa tubuh), *DiabetesPedigreeFunction* (fungsi silsilah diabetes), serta *Age* (usia). Kolom terakhir, yaitu "Outcome", merupakan label target klasifikasi yang menunjukkan apakah pasien tersebut terdiagnosis diabetes ("*Yes*") atau tidak ("*No*").

Dataset ini berasal dari sumber publik yang telah banyak digunakan dalam penelitian serupa, dan telah difilter untuk hanya menampilkan data yang telah memiliki label (berjumlah 99 data) guna digunakan sebagai *training set*. Keberadaan label ini sangat penting karena berfungsi sebagai dasar bagi algoritma *Decision Tree* untuk mempelajari pola hubungan antara atribut medis dengan status diagnosis diabetes. Proses ini merupakan bagian dari tahap preprocessing awal yang dilakukan sebelum model pelatihan dibangun.

Atribut-atribut yang ditampilkan pada tabel memiliki pengaruh yang signifikan terhadap diagnosis diabetes berdasarkan literatur medis maupun hasil penelitian sebelumnya. Sebagai contoh, kadar glukosa yang tinggi dan BMI di atas normal diketahui menjadi indikator utama dalam menentukan risiko diabetes. Namun, data ini juga memperlihatkan beberapa nilai yang tidak wajar, seperti angka nol pada kolom *Insulin* dan *SkinThickness*, yang secara klinis tidak mungkin bernilai nol. Oleh karena itu, proses pembersihan data (*data cleaning*) menjadi sangat penting untuk memastikan bahwa informasi yang digunakan dalam pelatihan model adalah valid dan tidak menyesatkan.

Dengan melihat data secara tabular, pengguna atau peneliti dapat melakukan inspeksi awal terhadap distribusi nilai, mendeteksi adanya anomali, serta menentukan strategi preprocessing yang tepat, seperti pengisian nilai kosong atau normalisasi. Visualisasi dalam bentuk tabel seperti ini juga memudahkan dalam pemilihan fitur yang akan digunakan dalam model, serta dalam menetapkan peran atribut menggunakan fitur *Set Role* pada RapidMiner. Dalam penelitian ini, ke-99 data berlabel tersebut dijadikan sebagai data latih untuk membangun model *Decision Tree* yang mampu mengklasifikasikan data pasien lain yang belum memiliki label berdasarkan atribut-atribut medis yang sama.

Visualisasi Scatter Plot



Gambar di atas menampilkan visualisasi scatter plot yang menggambarkan distribusi data pasien berdasarkan dua atribut medis penting, yaitu *Diabetes Pedigree Function* pada sumbu horizontal (X) dan *Blood Pressure* pada sumbu vertikal (Y). Masing-masing titik pada grafik merepresentasikan satu entri pasien dari dataset yang digunakan dalam pelatihan model *Decision Tree*. Warna titik mengindikasikan label klasifikasi pada atribut "Outcome": warna biru menunjukkan pasien dengan diagnosis diabetes ("*Yes*"), warna hijau untuk pasien non-diabetes ("*No*"), dan oranye untuk data yang belum memiliki label (jika ada).

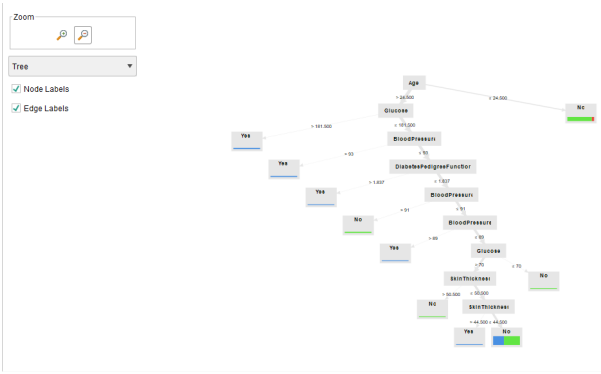
Tujuan utama dari visualisasi ini adalah untuk memahami sebaran nilai-nilai atribut medis yang digunakan dalam pemodelan, serta melihat bagaimana kelas data terdistribusi secara alami sebelum maupun sesudah proses klasifikasi. Berdasarkan pola sebaran pada grafik, terlihat bahwa pasien dengan nilai *Diabetes Pedigree Function* yang rendah dan tekanan darah sedang hingga tinggi lebih sering diklasifikasikan sebagai "No". Sementara itu, data pasien yang berada pada nilai *Diabetes Pedigree Function* yang tinggi menunjukkan kecenderungan diklasifikasikan sebagai "*Yes*", terutama jika tekanan darahnya rendah atau sangat tinggi.

Temuan ini sejalan dengan struktur model *Decision Tree* yang dibangun dalam penelitian, di mana atribut *Blood Pressure* dan *Diabetes Pedigree Function* digunakan sebagai pemisah lanjutan setelah atribut utama seperti *Age* dan *Glucose*. Hal ini memperkuat kesimpulan bahwa kombinasi faktor keturunan dan tekanan darah memiliki peran penting dalam menentukan risiko diabetes. Visualisasi ini juga membantu mendeteksi kemungkinan *outlier* yang dapat memengaruhi kinerja model, seperti titik data dengan tekanan darah sangat rendah namun memiliki nilai keturunan tinggi.

Secara keseluruhan, scatter plot ini menjadi alat bantu eksploratif yang penting dalam proses *data mining* karena memungkinkan analisis visual terhadap kecenderungan

distribusi label dan hubungan antar atribut. Dalam konteks implementasi sistem prediksi berbasis *Decision Tree* di RapidMiner, visualisasi semacam ini tidak hanya memudahkan interpretasi hasil, tetapi juga dapat membantu dalam proses validasi manual terhadap prediksi sistem yang dihasilkan. Dengan demikian, grafik ini berkontribusi dalam mendukung keakuratan dan keandalan sistem pendeteksi dini diabetes yang sedang dikembangkan.

Visualisasi Struktur Model *Decision Tree*



Setelah proses pelatihan menggunakan 99 *information* medis pasien yang telah diberi *name* pada atribut *Result*, algoritma *Decision Tree* berhasil membentuk sebuah pohon keputusan yang merepresentasikan alur klasifikasi berdasarkan atribut-atribut medis yang relevan. Hasil visualisasi menunjukkan bahwa atribut *Age* menjadi pemisah awal (*root hub*), dengan nilai ambang 24,5. Hal ini mengindikasikan bahwa usia pasien memiliki pengaruh yang sangat signifikan terhadap kemungkinan terdiagnosis diabetes. Jika usia pasien lebih kecil atau sama dengan 24,5 tahun, sistem secara langsung mengklasifikasikan pasien sebagai tidak menderita diabetes.

Namun, jika usia > 24,5 tahun, maka algoritma melanjutkan pemisahan berdasarkan nilai *Glucose*. Misalnya, apabila *Glucose* > 181,5 maka klasifikasi langsung menjadi "*Yes*", menunjukkan kemungkinan besar menderita diabetes. Jika nilai *Glucose* ≤ 181,5, maka pohon akan terus mengevaluasi atribut lainnya seperti *Blood Weight*, *Diabetes Family Work*, *Skin Thickness*, dan BMI hingga mencapai simpul akhir (*leaf hub*). Setiap jalur dalam pohon merepresentasikan kombinasi kondisi medis yang menghasilkan klasifikasi tertentu.

Interpretasi Aturan-aturan yang Dihasilkan
Berikut adalah beberapa aturan klasifikasi (*if-then rules*) yang dapat disimpulkan dari model pohon keputusan:

Tree

```
Age > 24.500
|   Glucose > 181.500: Yes {Yes=5, No=0, 0=0}
|   Glucose ≤ 181.500
|   |   BloodPressure > 93: Yes {Yes=4, No=0, 0=0}
|   |   BloodPressure ≤ 93
|   |   |   DiabetesPedigreeFunction > 1.837: Yes {Yes=2, No=0, 0=0}
|   |   |   DiabetesPedigreeFunction ≤ 1.837
|   |   |   |   BloodPressure > 91: No {Yes=0, No=5, 0=0}
|   |   |   |   BloodPressure ≤ 91
|   |   |   |   |   BloodPressure > 89: Yes {Yes=2, No=0, 0=0}
|   |   |   |   |   BloodPressure ≤ 89
|   |   |   |   |   |   Glucose > 70
|   |   |   |   |   |   |   SkinThickness > 50.500: No {Yes=0, No=2, 0=0}
|   |   |   |   |   |   |   SkinThickness ≤ 50.500
|   |   |   |   |   |   |   |   SkinThickness > 44.500: Yes {Yes=2, No=0, 0=0}
|   |   |   |   |   |   |   |   SkinThickness ≤ 44.500: No {Yes=21, No=31, 0=0}
|   |   |   |   |   |   Glucose ≤ 70: No {Yes=0, No=2, 0=0}
Age ≤ 24.500: No {Yes=0, No=21, 0=2}
```

Rule 1:
In case Age > 24.5 AND Glucose > 181.5
At that point Result = Yes
Run the show 2:
In case Age > 24.5 AND Glucose ≤ 181.5 AND BloodPressure ≤ 93 AND DiabetesPedigreeFunction ≤ 1.837 AND BloodPressure ≤ 91 AND BloodPressure > 89
At that point Result = Yes

Rule 3:
In case Age > 24.5 AND Glucose ≤ 181.5 AND BloodPressure ≤ 93 AND DiabetesPedigreeFunction ≤ 1.837 AND BloodPressure ≤ 91 AND BloodPressure ≤ 89 AND Glucose > 70 AND SkinThickness > 44.5
At that point Result = Yes

Aturan-aturan ini sangat berguna dalam praktik klinis untuk membantu dokter atau tenaga medis dalam melakukan asesmen awal terhadap pasien.

Visualisasi dan Statistik Information
Information hasil pemrosesan menunjukkan bahwa:
• Rata-rata *Glucose* pasien: 117,89
• Rata-rata *Blood Pressure*: 67,85
• Nilai maksimum Insulin mencapai 846
• Rata-rata BMI: 30,81
Berdasarkan *scramble plot* antara *Pregnancies* dan *Glucose* yang diberi warna berdasarkan kelas *Result*, terdapat kecenderungan bahwa pasien dengan kadar *Glucose* tinggi dan riwayat kehamilan lebih banyak cenderung diklasifikasikan sebagai "*Yes*".

Analisis Model dan Evaluasi
Walaupun tidak dilakukan evaluasi formal seperti *Perplexity Framework* atau pengukuran akurasi menggunakan metode validasi silang (*cross approval*), struktur pohon dapat dikaji secara logis. Show ini memperlihatkan pemisahan yang relevan dengan pengetahuan medis, misalnya:
• Glukosa darah tinggi sering diasosiasikan dengan diabetes.
• Usia dan BMI yang tinggi juga merupakan faktor risiko.
Namun, karena information yang digunakan hanya sebanyak 99 contoh yang berlabel dari add up to 768 information, maka demonstrate ini belum dapat dikatakan generalisasi dengan baik. *Overfitting* masih menjadi risiko tinggi. Distribusi name juga tidak seimbang (*Yes* = 38, *No*

= 61), sehingga berpotensi membuat *show inclination* terhadap kelas mayoritas.

Kelebihan dan Kekurangan Algoritma *Decission Tree*
Kelebihan:
• Transparan dan interpretatif. Struktur pohon dapat divisualisasikan dan dijelaskan dengan mudah.
• Cepat dalam proses pelatihan dan prediksi.
• Tidak memerlukan asumsi statistik yang kompleks.

Kekurangan:
• Rentan terhadap *overfitting*, terutama pada data kecil atau yang memiliki noise.
• Sensitif terhadap ketidakseimbangan data (*class imbalance*).
• Keputusan dipengaruhi oleh pemilihan atribut awal (*greedy algorithm*) yang mungkin melewatkan kombinasi yang lebih optimal.

Implikasi Hasil
Dengan menggunakan show yang telah terbentuk ini, *information* pasien baru yang tidak memiliki *name* dapat langsung diprediksi dengan menjalankan *information* melalui jalur pohon yang telah ada. Ini memungkinkan implementasi sistem pendukung keputusan (*Decision Bolster Framework*) untuk deteksi dini diabetes. Jika *demonstrate* dilatih ulang dengan seluruh *information* berlabel dan disempurnakan menggunakan metode validasi atau *gathering* (misal *Irregular Woodland*), maka akurasi dan generalisasi *demonstrate* akan meningkat.

D. PENUTUP

Simpulan
Berdasarkan hasil penelitian yang dilakukan mengenai prediksi risiko diabetes menggunakan algoritma *Decision Tree*, dapat disimpulkan bahwa metode ini mampu mengidentifikasi pola klasifikasi berdasarkan *information* medis pasien. Dengan menggunakan 99 *information* pasien sebagai information latih yang telah dilabeli, *demonstrate* dapat membentuk struktur pohon keputusan yang mempertimbangkan atribut-atribut penting seperti kadar glukosa, tekanan darah, dan usia. Proses pemodelan dilakukan melalui aplikasi Altair AI Studio dengan tahapan *preprocessing* yang sederhana, seperti menghapus nilai tidak substantial dan menentukan name prediksi.
Hasil visualisasi show pohon keputusan menunjukkan bahwa pasien dengan usia di atas 24,5 tahun dan kadar glukosa tinggi lebih berisiko terkena diabetes. Meskipun demikian, kualitas prediksi *demonstrate* masih dipengaruhi oleh jumlah *information* dan distribusi *name* yang belum merata. Dengan demikian, algoritma *Decision Tree* cukup potensial untuk digunakan dalam pendeteksian awal risiko diabetes, namun hasilnya tetap perlu divalidasi lebih lanjut dengan information yang lebih besar dan beragam.

Saran
1. Jumlah data yang digunakan dalam penelitian ini masih terbatas, sehingga disarankan untuk menggunakan dataset yang lebih besar dan seimbang dalam distribusi label untuk meningkatkan akurasi dan keandalan model.
2. Perlu dilakukan proses preprocessing yang lebih lengkap, seperti normalisasi atau pengisian nilai kosong dengan metode statistik agar data lebih bersih dan hasil lebih valid.
3. Diharapkan penggunaan model ini dapat dikembangkan lebih lanjut dengan menambahkan metode evaluasi lain seperti *confusion matrix* atau ROC *Curve* untuk mengukur performa model secara lebih komprehensif.
4. Penelitian lanjutan dapat membandingkan hasil algoritma *Decision Tree* dengan algoritma lainnya, seperti *Random Forest*, Naive Bayes, atau KNN untuk melihat metode mana yang lebih unggul dalam konteks prediksi penyakit diabetes.

Ucapan Terima Kasih
Penulis mengucapkan terima kasih kepada semua pihak yang telah membantu dalam proses penyusunan penelitian ini, khususnya kepada:
1. Ahmad Reza Farrasy dari Universitas Pamulang yang telah memberikan dukungan fasilitas dan data,
2. Amar Naufal Al-kharits dari Universitas Pamulang yang telah memberikan dukungan fasilitas dan data,
3. Muhamad Bustomi dari Universitas Pamulang yang telah memberikan dukungan fasilitas dan data,
4. Nazar Maulana dari Universitas Pamulang yang telah memberikan dukungan fasilitas dan data,
5. Taupik Abdul Rahman dari Universitas Pamulang yang telah memberikan dukungan fasilitas dan data,
6. Dosen pembimbing Data Mining atas arahan dan masukannya,
7. Teman-teman serta partisipan pengujian sistem yang telah memberikan waktu dan umpan baliknya.
8. Semoga penelitian ini dapat memberikan manfaat dan menjadi referensi untuk pengembangan sistem serupa di masa mendatang.

E. DAFTAR PUSTAKA

Dewi, R., & Hidayat, M. (2021). Perbandingan algoritma C4.5 dan CART dalam prediksi diabetes. Jurnal Sistem Informasi, 17(2), 99–106.

Fitriani, R., & Nugroho, A. (2022). Implementasi data mining untuk memprediksi risiko diabetes pada data rekam medis. Seminar Nasional Aplikasi Teknologi Informasi.

Hossain, M. M., Rawshan, A., & Islam, M. M. (2021). Prediction of diabetes mellitus using decision tree algorithm: A study based on PIMA Indian dataset. Journal of King Saud University - Computer and Information Sciences, 33(1), 137–143. https://doi.org/10.1016/j.jksuci.2018.09.014

Jain, A., & Singh, K. (2020). A machine learning approach to predict the risk of diabetes using decision tree and random forest. International Journal of Advanced Computer Science and Applications, 11(5), 372–378.

Jannah, M., & Ramadhan, A. (2020). Penerapan RapidMiner untuk klasifikasi penyakit diabetes.

Jurnal Teknologi dan Sistem Komputer, 9(1), 112–118.

Kurniawan, R., & Prasetyo, E. (2020). Data mining untuk prediksi diabetes melitus menggunakan algoritma CART. Jurnal Teknologi Informasi dan Ilmu Komputer, 7(1), 11–18.

Li, X., & Wong, L. (2023). Explainable AI for diabetes prediction: A clinical perspective. AI in Healthcare Insights.

Mehta, D., & Joshi, K. (2023). Decision trees vs logistic regression for diabetes classification. International Journal of Data Science.

Novitasari, R. D., & Wahyuni, S. (2021). Prediksi penyakit diabetes menggunakan algoritma decision tree C4.5. Jurnal Informatika dan Komputer, 8(2), 101–107.

Nurmalasari, T., & Fahmi, A. (2023). Prediksi penyakit menggunakan metode decision tree dan klasifikasi data kesehatan di RapidMiner. Jurnal Ilmiah Teknologi Informasi Asia, 17(1), 24–31.

Permana, H., & Rachmawati, D. (2022). Analisis data mining untuk prediksi diabetes menggunakan decision tree dan random forest. Jurnal Sains dan Informatika, 8(1), 23–29.

Putra, A., & Widodo, S. (2020). Pemanfaatan RapidMiner dalam memprediksi risiko diabetes pada pasien di rumah sakit. Jurnal Sistem Informasi, 6(1), 17–23.

Rahmawati, N., & Yuliana, S. (2021). Perbandingan akurasi algoritma decision tree dan naive Bayes dalam prediksi diabetes. Jurnal Informatika Merdeka, 6(1), 55–62.

Rini, D. P., Zarlis, M., & Hartama, D. (2020). Model decision tree untuk prediksi penyakit menggunakan data medis. Jurnal Sistem Cerdas, 3(2), 89–95.

Salehahmadi, Z., Ehsan, A., & Mahmoudi, M. (2022). Application of decision tree algorithms in healthcare data mining. Artificial Intelligence in Medicine, 118, 102101. https://doi.org/10.1016/j.artmed.2021.102101

Saputra, D., & Lestari, S. (2022). Penerapan classification tree untuk mendeteksi sindrom metabolik. Jurnal Ilmu Komputer dan Teknologi Informasi.

Sharma, S., & Kaur, A. (2021). Comparative analysis of decision tree algorithms for disease prediction using machine learning. Materials Today: Proceedings. https://doi.org/10.1016/j.matpr.2021.07.048

Singh, J. P., Kaur, P., & Sharma, R. (2020). Diabetes prediction using machine learning techniques: A comparative study. Procedia Computer Science, 167, 706–716. https://doi.org/10.1016/j.procs.2020.03.327

Susanti, D., & Hartati, S. (2020). Klasifikasi penyakit diabetes menggunakan metode decision tree C4.5 dan naive Bayes. Jurnal Informatika, 14(1), 45–51.

Trisnawati, R., & Syafriandi, S. (2023). Sistem prediksi diabetes dengan algoritma decision tree dan aplikasi RapidMiner. Prosiding Seminar Nasional Teknologi Informasi dan Komputer.