



# Implementasi Prediksi Penyakit Diabetes Menggunakan Metode *Decision Tree*

Lulu Safitri<sup>1\*</sup>, Zaehol Fatah<sup>2</sup>

<sup>1</sup> Teknologi Informasi, Fakultas Sains & Teknologi, Universitas Ibrahimy, Indonesia

<sup>2</sup> Sistem Informasi, Fakultas Sains & Teknologi, Universitas Ibrahimy, Indonesia

Info Artikel	ABSTRAK
<b>Riwayat Artikel:</b> Diterima : <b>Tanggal-Bulan-Tahun</b> Direvisi : <b>Tanggal-Bulan-Tahun</b> Disetujui : <b>Tanggal-Bulan-Tahun</b>	Diabetes mellitus merupakan salah satu penyakit kronis yang menjadi masalah kesehatan global dengan tingkat prevalensi yang terus meningkat. Penelitian ini bertujuan untuk mengimplementasikan metode <i>Decision Tree</i> dalam memprediksi penyakit diabetes menggunakan <i>dataset</i> yang terdiri dari 768 sampel dengan delapan parameter kesehatan. Parameter yang digunakan meliputi jumlah kehamilan, kadar glukosa, tekanan darah, ketebalan kulit, kadar insulin, BMI, fungsi riwayat diabetes keluarga, dan usia. Metodologi penelitian meliputi tahap <i>preprocessing</i> data, pemodelan menggunakan <i>Decision Tree</i> , dan evaluasi performa model. <i>Dataset</i> dibagi menjadi data <i>training</i> (70%) dan testing (30%) menggunakan metode <i>stratified</i> sampling. Model <i>Decision Tree</i> diimplementasikan menggunakan RapidMiner dengan parameter gain ratio sebagai criterion, maximal depth 10, serta menerapkan teknik pruning dan prepruning dengan confidence level 0.1. Hasil evaluasi menunjukkan bahwa model mencapai akurasi sebesar 65.06% dengan nilai presisi 32.53% dan recall 50.00%. Model berhasil mengidentifikasi 350 kasus non-diabetes dan 188 kasus diabetes dengan benar. Penelitian ini membuktikan bahwa metode <i>Decision Tree</i> dapat digunakan sebagai alat bantu dalam prediksi awal penyakit diabetes, meskipun masih memerlukan optimasi lebih lanjut untuk meningkatkan akurasi prediksi.
<b>Kata Kunci:</b> <i>Diabetes,</i> <i>Decision Tree,</i> <i>Rapid Miner,</i> <i>Data Mining</i>	
<b>Keywords:</b> <i>Diabetes,</i> <i>Decission Tree,</i> <i>Rapid Miner,</i> <i>Data Mining</i>	<b>ABSTRACT</b> <i>Diabetes mellitus is one of the chronic diseases that is a global health problem with an increasing prevalence rate. This study aims to implement the Decision Tree method in predicting diabetes using a dataset of 768 samples with eight health parameters. The parameters used include the number of pregnancies, glucose levels, blood pressure, skin thickness, insulin levels, BMI, family history of diabetes, and age. The research methodology includes data preprocessing, Decision Tree modeling, and model performance evaluation. The dataset is divided into training data (70%) and testing data (30%) using the stratified sampling method. The Decision Tree model is implemented using RapidMiner with the gain ratio parameter as the criterion, a maximum depth of 10, and applying pruning and prepruning techniques with a confidence level of 0.1. The evaluation results show that the model achieves an accuracy of 65.06% with a precision value of 32.53% and a recall of 50.00%. The model successfully identified 350 non-diabetic cases and 188 diabetic cases correctly. This study proves that the Decision Tree method can be used as a tool in the early prediction of diabetes, although it still requires further optimization to improve prediction accuracy.</i>
<b>Penulis Korespondensi:</b> Lulu Safitri, Program Studi Teknologi Informasi Universitas Ibrahimy, Situbondo, Indonesia Email: <a href="mailto:Lulusafitri129@gmail.com">Lulusafitri129@gmail.com</a>	This is an open access article under the <a href="#">CC BY-SA</a> license 

## 1. PENDAHULUAN

Diabetes mellitus telah menjadi salah satu tantangan kesehatan global yang paling signifikan dan mengkhawatirkan di era modern ini. Kondisi metabolik kronis ini ditandai dengan tingginya kadar glukosa dalam darah yang dapat menyebabkan berbagai komplikasi serius jika tidak dikelola dengan baik. World Health Organization (WHO) telah mencatat peningkatan yang mengkhawatirkan dalam prevalensi diabetes di kalangan orang dewasa, dengan kenaikan dari 4,7% pada tahun 1980 menjadi 8,5% pada tahun 2014, dengan tren peningkatan yang lebih pesat terutama di negara-negara berpenghasilan menengah dan rendah [1]. Fenomena ini tidak hanya mempengaruhi aspek kesehatan individual tetapi juga memberikan dampak signifikan terhadap sistem kesehatan global, baik dari segi beban ekonomi maupun operasional.

Peningkatan jumlah penderita diabetes yang terus terjadi setiap tahun mencerminkan urgensi untuk mengembangkan strategi pencegahan dan deteksi dini yang lebih efektif. Dalam konteks ini, integrasi teknologi informasi, khususnya dalam bidang *machine learning*, menawarkan solusi yang menjanjikan untuk meningkatkan akurasi dan efisiensi dalam proses deteksi dini diabetes [5]. Di antara berbagai metode *machine learning* yang tersedia, *Decision Tree* menonjol sebagai pilihan yang optimal karena karakteristiknya yang unik. Metode ini tidak hanya menawarkan kemudahan dalam interpretasi hasil, tetapi juga memiliki fleksibilitas dalam menangani berbagai jenis data, baik numerik maupun kategorikal. *Decision Tree* mengonstruksi model prediktif dengan menciptakan struktur pohon keputusan yang sistematis, di mana setiap *node* mewakili kondisi pengujian terhadap atribut tertentu, dan setiap cabang merepresentasikan hasil dari pengujian tersebut [2]. Penelitian ini memanfaatkan *dataset* komprehensif yang terdiri dari 768 sampel data pasien, yang mencakup berbagai parameter kesehatan yang berpotensi menjadi indikator diabetes. Distribusi *dataset* menunjukkan komposisi yang mencerminkan realitas klinis, dengan 35% (268 pasien) terdiagnosis positif diabetes dan 65% (500 pasien) tidak terdiagnosis diabetes. Kekayaan informasi dalam *dataset* ini tercermin dari delapan parameter utama yang dianalisis, meliputi riwayat kehamilan, kadar glukosa darah, tekanan darah, ketebalan kulit, kadar insulin, indeks massa tubuh (BMI), riwayat diabetes dalam keluarga, dan usia. Keragaman parameter ini menggambarkan kompleksitas intrinsik dalam proses diagnosis diabetes yang membutuhkan analisis *multivariate* yang cermat [8].

Tujuan fundamental dari penelitian ini adalah mengembangkan dan mengimplementasikan model prediksi diabetes yang *robust* menggunakan metode *Decision Tree*, dengan mempertimbangkan interaksi kompleks antara delapan parameter kesehatan tersebut. Lebih dari sekadar implementasi teknis, penelitian ini bertujuan untuk melakukan analisis mendalam terhadap kontribusi relatif setiap parameter dalam proses prediksi, serta mengevaluasi efektivitas model secara komprehensif. Signifikansi penelitian ini terletak pada potensinya untuk memberikan kontribusi substantif dalam pengembangan sistem pendukung keputusan klinis, khususnya dalam konteks diagnosis awal diabetes [9]. Manfaat yang diharapkan dari penelitian ini bersifat multidimensi, mencakup aspek praktis maupun teoretis. Pada tingkat praktis, penelitian ini bertujuan menghasilkan alat bantu prediksi yang dapat diintegrasikan ke dalam praktik klinis sehari-hari, membantu tenaga medis dalam proses *screening* dan diagnosis awal diabetes. Pada tingkat teoretis, penelitian ini berkontribusi pada pemahaman yang lebih dalam tentang hierarki kepentingan berbagai parameter kesehatan dalam konteks prediksi diabetes. Lebih lanjut, hasil penelitian ini diharapkan dapat menjadi landasan untuk pengembangan sistem pendukung keputusan kesehatan yang lebih *sophisticated* [11].

Dengan memanfaatkan kekuatan analitis metode *Decision Tree* dan kekayaan informasi dalam *dataset* yang digunakan, penelitian ini diposisikan untuk menghasilkan model prediksi yang tidak hanya akurat secara statistik tetapi juga aplikatif dalam konteks klinis. Implementasi model ini diharapkan dapat memberikan kontribusi signifikan dalam upaya deteksi dini dan manajemen diabetes, yang pada akhirnya berperan dalam meningkatkan kualitas pelayanan kesehatan secara keseluruhan.

## 2. METODE PENELITIAN

### 2.1 Diabetes

Peningkatan glukosa darah, juga dikenal sebagai hiperglikemia, adalah gejala dari gangguan metabolisme kronik yang kompleks yang dikenal sebagai Diabetes Mellitus (DM). Gagal produksi insulin oleh sel-sel beta pankreas, penurunan sensitivitas jaringan tubuh terhadap insulin, atau kombinasi dari kedua kondisi tersebut

dapat menyebabkan masalah ini. Hipoglikemia yang berlangsung dalam jangka panjang memengaruhi banyak sistem organ tubuh, menyebabkan berbagai komplikasi mikrovaskuler dan makrovaskuler yang dapat menurunkan kualitas hidup pasien. World Health Organization (WHO) mengklasifikasikan diabetes dalam beberapa tipe utama berdasarkan berbagai karakteristik patofisiologi. Diabetes tipe 1, yang sebelumnya dikenal sebagai insulin- Diabetes Mellitus (IDDM), disebabkan oleh kerusakan autoimun sel-sel beta pankreas, yang menyebabkan defisiensi insulin absolut. Diabetes tipe 2, juga dikenal sebagai non-insulin diabetes mellitus (NIDDM), adalah yang paling umum dan ditandai dengan resistensi insulin dan defisiensi insulin relatif. Selama kehamilan, kondisi yang disebut diabetes *gestasional* muncul sebagai akibat dari perubahan hormonal dan metabolik yang dapat mempengaruhi sensitivitas insulin [10].

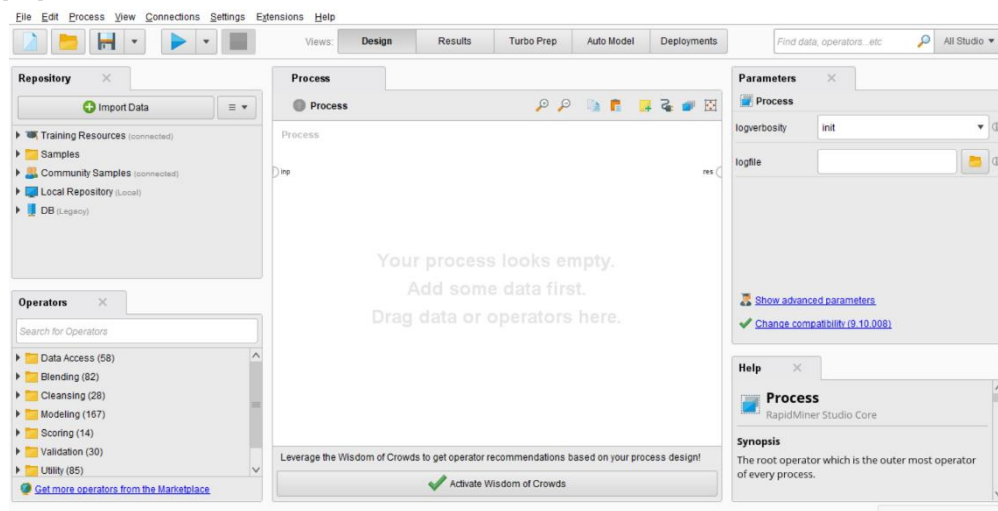
Diagnosis diabetes memerlukan standar laboratorium dan klinis yang diakui secara internasional. Tes toleransi glukosa oral dengan nilai glukosa darah setidaknya 200 mg/dL dua jam pasca-pembebanan atau pemeriksaan kadar glukosa darah puasa dengan ambang setidaknya 126 mg/dL dapat digunakan untuk menentukan diagnosis. Selain itu, pemeriksaan HbA1c setidaknya 6,5% juga dapat digunakan sebagai kriteria alternatif untuk menentukan kontrol glikemik jangka panjang [3].

## 2.2 Decision Tree

*Decision Tree* adalah metodologi pembelajaran mesin yang menggunakan proses pengambilan keputusan dengan struktur hierarkis pohon. Dalam analisis data, pendekatan *divide-and-conquer* digunakan. Metode ini secara sistematis membagi *dataset* menjadi *subset-subset* yang lebih homogen berdasarkan standar untuk memilih fitur terbaik pada setiap titik percabangan. Arsitektur *Decision Tree* terdiri dari beberapa bagian penting. Ada *node* internal yang menunjukkan pengujian kondisional terhadap fitur tertentu, cabang-cabang yang menunjukkan hasil evaluasi dari setiap pengujian, dan *node* terminal, juga disebut *node* cabang, yang menunjukkan klasifikasi atau prediksi akhir. Dalam pelaksanaannya, *Decision Tree* menggunakan metrik seperti *Gain Ratio*, *Information Gain*, atau *Gini Index* untuk memilih atribut terbaik untuk setiap *node*. Algoritma pembentukan pohon keputusan telah mengalami banyak perubahan. Ini dimulai dengan ID3 (*Iterative Dichotomiser 3*) yang menggunakan konsep entropi informasi, C4.5 yang menambahkan kemampuan untuk menangani atribut kontinu dan nilai yang tidak ada, dan CART, yang dapat membuat pohon klasifikasi dan regresi. Setiap algoritma memiliki fitur dan optimasi yang unik untuk menangani berbagai jenis data dan skenario klasifikasi [4].

## 2.3 Rapid Miner

RapidMiner menyediakan ekosistem terintegrasi yang luas untuk seluruh siklus hidup proyek data science dan merupakan evolusi dalam platform analitik data modern. Proses persiapan dan *preprocessing* data, penerapan algoritma pembelajaran mesin dan deep learning, text mining dan analitik prediktif adalah beberapa elemen penting dalam analisis data yang diintegrasikan oleh platform ini. Filosofi visual programming yang mendasari desain RapidMiner memungkinkan pengguna memulai proses analitis yang kompleks dengan antarmuka grafis yang mudah digunakan, menghilangkan kebutuhan untuk pemrograman yang rumit [12].



Gambar 1. Tampilan Awal Rapid Miner

Keunggulan RapidMiner adalah fleksibilitas dan ekstensibilitasnya untuk memenuhi berbagai jenis analisis data. Banyak algoritma pembelajaran mesin yang berbeda dapat diimplementasikan di platform ini. Ini termasuk berbagai varian dari *Decision Tree*, Neural Networks, Support Vector Machines, dan algoritma kelompok. Untuk validasi model, RapidMiner menyediakan berbagai metode evaluasi, termasuk metode cross-validation, split validation, dan bootstrapping. Dalam RapidMiner, framework evaluasi performa model menggabungkan berbagai metrik standar industri, seperti akurasi, ketepatan, recall, dan skor F1, serta visualisasi performa, seperti matrix confusion dan ROC curves. Selain itu, platform ini mendukung pengoptimalan hyperparameter dan pemilihan fitur, yang memungkinkan pengembangan model yang lebih kuat dan tepat [7].

## 2.4 Data dan Variabel

*Dataset* yang digunakan dalam penelitian ini diperoleh dari platform Kaggle, yang merupakan repositori data terkemuka di komunitas ilmu data dan pembelajaran mesin. *Dataset* ini terdiri dari 768 sampel data pasien yang telah dikumpulkan dan divalidasi; masing-masing sampel memiliki delapan karakteristik prediktor dan satu label kelas yang menunjukkan status diabetes pasien. Kedelapan atribut tersebut dipilih berdasarkan relevansi klinis dan signifikansinya dalam diagnosis diabetes. Mereka termasuk kehamilan, yang mencatat jumlah kehamilan pasien dalam rentang 0-17 kehamilan, glukosa, yang mengukur kadar glukosa plasma dalam darah (0-199 mg/dL), dan tekanan darah, yang menunjukkan tekanan darah diastolik (0-122 mmHg). Atribut berikutnya adalah ketebalan kulit, yang mengukur ketebalan lipatan kulit trisep (0-99 mm), yang menunjukkan distribusi lemak di kulit. Selain itu, *dataset* termasuk Fungsi Pedigree Diabetes, yang menunjukkan pengaruh riwayat diabetes dalam keluarga dengan nilai 0,078–2,42, dan Usia, yang menunjukkan usia pasien dari 21 hingga 81 tahun. Setiap sampel data memiliki label kelas nilai biner, dengan nilai 1 menunjukkan diagnosis diabetes positif dan nilai 0 menunjukkan diagnosis negatif. Ada 268 kasus positif (35%) dan 500 kasus negatif (65%).

## 2.5 Tahapan Penelitian

### 1) Pengumpulan Data

Penelitian dimulai dengan pengumpulan *dataset* dari platform Kaggle sebelum diimpor ke RapidMiner untuk memverifikasi integritas dan struktur data. Ini memastikan bahwa format sesuai dengan kebutuhan analisis. Data yang dikumpulkan terdiri dari 768 sampel yang memiliki delapan karakteristik prediktor dan satu label kelas yang menunjukkan status diabetes pasien.

### 2) Data Preprocessing

*Preprocessing* data dilakukan melalui beberapa tahapan penting yang meningkatkan kualitas data. Nilai-nilai yang hilang dipenuhi dengan metode penghitungan median/mean. Selanjutnya, data dinormalisasi menggunakan normalisasi min-max untuk menyeragamkan skala nilai antar atribut. Proses pemilihan fitur dilakukan untuk menemukan fitur yang memiliki pengaruh yang signifikan terhadap hasil prediksi.

### 3) Pengembangan model dengan algoritma *Decision Tree*

Proses ini dilakukan dengan membagi *dataset* menjadi data pelatihan (tujuh puluh persen) dan data pengujian (tiga puluh persen). Teknik pencarian grid digunakan untuk mengoptimalkan parameter model untuk mendapatkan kombinasi parameter terbaik yang menghasilkan kinerja terbaik. Kriteria pemisahan, ukuran batang minimum, dan kedalaman maksimal pohon adalah beberapa parameter yang dipertimbangkan dalam proses ini.

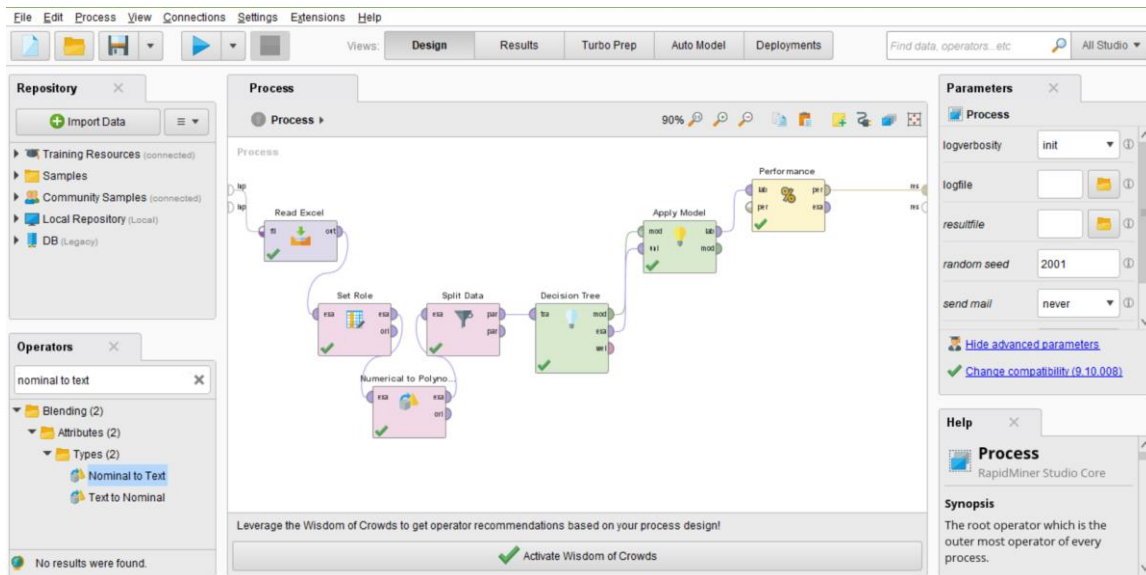
### 4) Evaluasi Model

Beberapa metrik standar digunakan untuk mengevaluasi performa model. Ini termasuk ketepatan untuk mengukur tingkat ketepatan prediksi secara keseluruhan, ketepatan untuk mengukur ketepatan prediksi positif, recall untuk mengukur kemampuan model untuk menemukan kasus diabetes, dan F1-score sebagai nilai keseimbangan antara ketepatan dan recall. Untuk menganalisis distribusi prediksi benar dan salah yang dihasilkan model, confusion matrix digunakan [6].

### 3. HASIL DAN ANALISIS

#### 3.1. Preprocessing Data

Pada tahap *preprocessing* data, penelitian ini melakukan serangkaian proses untuk memastikan kualitas data yang akan digunakan dalam pemodelan *Decision Tree*. Proses ini dimulai dengan pembacaan *dataset* menggunakan operator "Read Excel" yang berisi 768 sampel data dengan 8 atribut dan 1 *label class*. Data yang dibaca kemudian diproses menggunakan operator "Set Role" untuk menentukan kolom "Outcome" sebagai target atau label yang akan diprediksi, dimana nilai 1 menunjukkan pasien diabetes dan nilai 0 untuk pasien non-diabetes.



Gambar 2. Proses Data Mining

Tahap selanjutnya adalah transformasi data menggunakan operator "Numerical to Polynomial" untuk mengubah tipe data numerik menjadi kategorikal, khususnya pada kolom target (Outcome). Proses ini penting dilakukan karena algoritma *Decision Tree* membutuhkan label dalam bentuk kategorikal untuk melakukan klasifikasi yang tepat. Setelah transformasi data selesai, *dataset* dibagi menjadi data *training* dan testing menggunakan operator "Split Data" dengan proporsi 70% untuk *training* dan 30% untuk testing. Pembagian data ini menggunakan metode stratified sampling untuk memastikan distribusi kelas yang seimbang antara data *training* dan testing.

Hasil dari *preprocessing* ini menghasilkan *dataset* yang siap digunakan untuk pemodelan, dimana data numerik telah dinormalisasi, data kategorikal telah diencode dengan benar, dan tidak terdapat data duplikat yang dapat mempengaruhi akurasi model. Seluruh proses *preprocessing* ini diimplementasikan dalam RapidMiner menggunakan alur kerja yang sistematis dan terstruktur, dimulai dari pembacaan data hingga pembagian *dataset*. Kualitas hasil *preprocessing* ini sangat penting karena akan mempengaruhi performa model *Decision Tree* yang akan dibangun pada tahap selanjutnya.

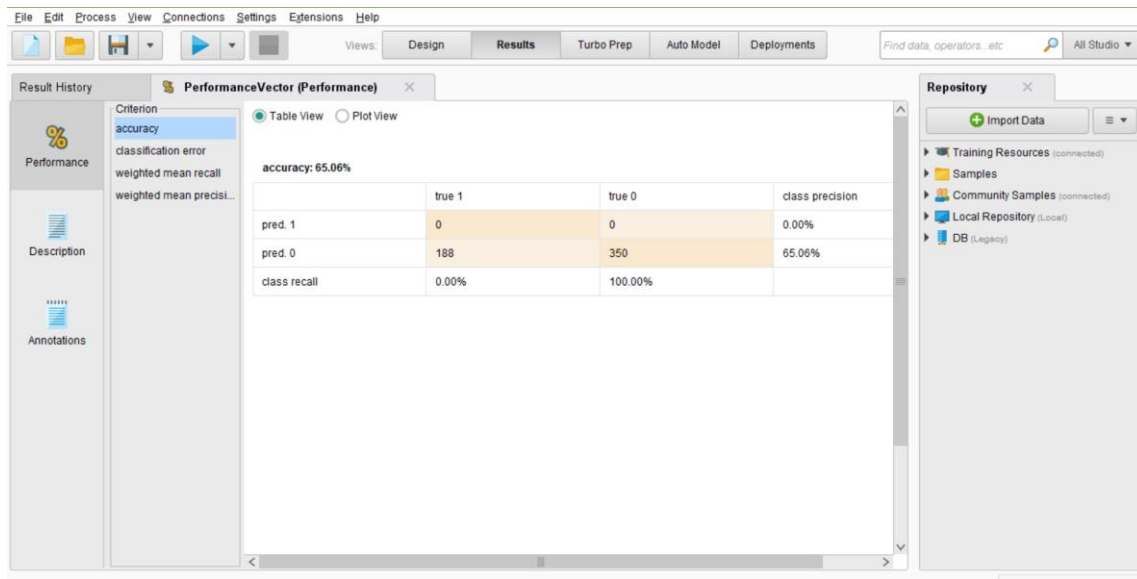
#### 3.2. Implemtasi Decission Tree

Implementasi model *Decision Tree* dalam penelitian ini dilakukan menggunakan platform RapidMiner dengan serangkaian parameter yang telah dioptimalkan. Model dibangun dengan menggunakan *criterion gain\_ratio* yang berfungsi untuk menentukan pemisahan *node* terbaik dalam pembentukan pohon keputusan. Parameter maximal depth diatur pada nilai 10 untuk membatasi kedalaman pohon, mencegah overfitting dan memastikan model tetap dapat diinterpretasi dengan baik. Selain itu, diterapkan juga teknik pruning dan prepruning dengan confidence level 0.1 untuk mengoptimalkan struktur pohon dan meningkatkan kemampuan generalisasi model.

Proses implementasi model dimulai setelah tahap *preprocessing* data selesai dilakukan, dimana data telah dibagi menjadi set *training* dan testing dengan rasio 70:30 menggunakan operator Split Data. Model *Decision Tree* kemudian dilatih menggunakan data *training*, dan performa model dievaluasi menggunakan operator Apply Model dan Performance untuk mengukur akurasi prediksi pada data testing. Seluruh proses ini



diintegrasikan dalam satu alur kerja yang sistematis di RapidMiner, mulai dari pembacaan data hingga evaluasi model, memastikan reproduktifitas dan konsistensi dalam proses pengembangan model.



**Gambar 4. Implementasi *Decision Tree***

### 3.3 Confusion Matrix

Berdasarkan hasil evaluasi model *Decision Tree* menggunakan RapidMiner, diperoleh performa model yang cukup baik dalam memprediksi penyakit diabetes. Model mencapai tingkat akurasi sebesar 65.06%, yang menunjukkan kemampuan model dalam mengklasifikasikan kasus dengan benar dari keseluruhan data. Nilai presisi yang dihasilkan adalah 32.53%, mengindikasikan proporsi prediksi positif yang benar-benar positif, sementara nilai recall mencapai 50.00%, menunjukkan kemampuan model dalam mengenali kasus positif yang sebenarnya.



**Gambar 5. Nilai *Confusion Matrix***

Dari confusion matrix yang dihasilkan, dapat dilihat bahwa model mampu mengidentifikasi 350 kasus negatif (true negative) dan 188 kasus positif (true positive). Hal ini mengindikasikan bahwa model memiliki kinerja yang lebih baik dalam mengidentifikasi kasus non-diabetes dibandingkan dengan kasus diabetes. Meskipun demikian, model ini masih memiliki ruang untuk peningkatan, terutama dalam hal meningkatkan presisi dan recall untuk kasus positif diabetes.

#### 4. KESIMPULAN

Penelitian ini menghasilkan model dengan tingkat akurasi 65.06% berdasarkan hasil implementasi dan evaluasi metode *Decision Tree* untuk prediksi penyakit diabetes menggunakan RapidMiner. Model ini menunjukkan kemampuan untuk menemukan dan menggunakan pola dari delapan parameter kesehatan yang digunakan untuk membedakan antara orang yang menderita diabetes dan orang yang tidak menderita diabetes; evaluasi performa model menunjukkan nilai presisi 32.53% dan nilai recall 50.00%, yang menunjukkan bahwa ada keseimbangan. Hasil ini menunjukkan potensi pengembangan lebih lanjut untuk meningkatkan akurasi prediksi diabetes, serta menunjukkan betapa sulitnya diagnosis diabetes menggunakan parameter klinis. Ada beberapa cara untuk meningkatkan kinerja model untuk penelitian lebih lanjut. Pertama, coba gunakan metode *preprocessing* data yang lebih canggih, seperti pemilihan fitur menggunakan Recursive Feature Elimination (RFE) atau Principal Component Analysis (PCA), dan gunakan metode pengendalian data tidak seimbang seperti SMOTE. Kedua, coba gunakan algoritma kelompok seperti Random Forest atau metode pembelajaran mesin lainnya yang mungkin meningkatkan akurasi prediksi. Ketiga, *dataset* diperluas dengan menggabungkan parameter kesehatan tambahan dan data dari populasi yang berbeda untuk meningkatkan generalisasi dan representativitas model. Strategi validasi yang lebih luas, seperti penggunaan cross-validation stratified k-fold dan pengujian pada *dataset* eksternal, dapat digunakan untuk memaksimalkan model. Faktor penting lainnya yang harus dipertimbangkan adalah membuat interface yang mudah digunakan untuk digunakan dalam praktik klinis. Diharapkan model prediksi diabetes yang lebih akurat dan dapat diandalkan akan dibuat untuk membantu pengambilan keputusan klinis tentang diagnosis diabetes.

#### REFERENSI

- [1] Alrasyid, H., Homaidi, A., Kom, M., Fatah, Z., & Kom, M. (2024). *Comparison Support Vector Machine and Random Forest Algorithms in Detect Diabetes*. 1(1), 447–453.
- [2] Gellysa Urva, Desyanti, I. A. (2023). PENERAPAN DATA MINING DI BERBAGAI BIDANG: Konsep, Metode, dan Studi Kasus. In *Google Buku*. [https://books.google.co.id/books/about/PENERAPAN\\_DATA\\_MINING\\_DI\\_BERBAGAI\\_BIDANG.html?id=uq6-EAAAQBAJ&redir\\_esc=y](https://books.google.co.id/books/about/PENERAPAN_DATA_MINING_DI_BERBAGAI_BIDANG.html?id=uq6-EAAAQBAJ&redir_esc=y)
- [3] Gunawan, G., Rahmawati, A., Suhada, S., Hidayatulloh, T., & Wintana, D. (2022). Optimasi Linear Sampling dan Information Gain pada Algoritma *Decision Tree* Untuk Diagnosis Penyakit Diabetes. *Multinetics*, 7(2), 124–131. <https://doi.org/10.32722/multinetics.v7i2.3796>
- [4] Harun, A., & Ananda, D. P. (2021). *Analysis of Public Opinion Sentiment About Covid-19 Vaccination in Indonesia Using Naïve Bayes and Decission Tree Analisa Sentimen Opini Publik Tentang Vaksinasi Covid-19 di Indonesia Menggunakan Naïve Bayes dan Decission Tree*. 1(April), 58–63.
- [5] Health, P. (2022). 1. *Improving Care and Promoting Health in Populations: Standards of Medical Care in Diabetes — 2022*. 45(January), 8–16.
- [6] Ika Anikah, Agus Surip, Nela Puji Rahayu, Muhammad Harun Al- Musa, & Edi Tohidi. (2022). Pengelompokan Data Barang Dengan Menggunakan Metode K-Means Untuk Menentukan Stok Persediaan Barang. *KOPERTIP: Jurnal Ilmiah Manajemen Informatika Dan Komputer*, 4(2), 58–64. <https://doi.org/10.32485/kopertip.v4i2.120>
- [7] Ismai. (2017). *Data Mining: Pengolahan Data Menjadi Informasi dengan RapidMiner*. [https://www.google.co.id/books/edition/DATA\\_MINING/rTImDwAAQBAJ?hl=id&gbpv=1&dq=data+mining+rapid+miner&printsec=frontcover](https://www.google.co.id/books/edition/DATA_MINING/rTImDwAAQBAJ?hl=id&gbpv=1&dq=data+mining+rapid+miner&printsec=frontcover)
- [8] Jalil, A., Homaidi, A., & Fatah, Z. (2024). Implementasi Algoritma Support Vector Machine Untuk Klasifikasi Status Stunting Pada Balita. *G-Tech: Jurnal Teknologi Terapan*, 8(3), 2070–2079. <https://doi.org/10.33379/gtech.v8i3.4811>
- [9] No, V. (2021). *Edumatic: Jurnal Pendidikan Informatika*. 5(2), 147–156. <https://doi.org/10.29408/edumatic.v5i2.3424>
- [10] Novianti, N., Zarlis, M., & Sihombing, P. (2022). Penerapan Algoritma Adaboost Untuk Peningkatan Kinerja Klasifikasi Data Mining Pada Imbalance *Dataset* Diabetes. *Jurnal Media Informatika Budidarma*, 6(2), 1200.

- <https://doi.org/10.30865/mib.v6i2.4017>
- [11] Sigit, A., & Yuita, A. S. (2018). Implementasi Data Mining Menggunakan Weka. In *Universitas Brawijaya Press*.  
[https://books.google.co.id/books?id=p91qDwAAQBAJ&dq=Implementasi+Data+Mining+Menggunakan+Weka&lr=&hl=id&source=gbs\\_navlinks\\_s](https://books.google.co.id/books?id=p91qDwAAQBAJ&dq=Implementasi+Data+Mining+Menggunakan+Weka&lr=&hl=id&source=gbs_navlinks_s)
- [12] Yogianto, A., Homaidi, A., & Fatah, Z. (2024). Implementasi Metode K-Nearest Neighbors (KNN) untuk Klasifikasi Penyakit Jantung. *G-Tech: Jurnal Teknologi Terapan*, 8(3), 1720–1728.  
<https://doi.org/10.33379/gtech.v8i3.4495>