

# Klasifikasi Risiko Stroke Menggunakan Metode Decision Tree pada Healthcare Dataset Stroke Data

Zaehol Fatah<sup>1</sup>, Mahattir Rahman Muhammad<sup>2</sup>

<sup>1,2</sup> Universitas Ibrahimy, Situbondo

Email: Mahattirahmad@gmail.com

## Abstrak

Artikel ini menguraikan tahapan penggunaan algoritma Decision Tree guna membangun model prediktif untuk klasifikasi status pekerjaan menggunakan atribut usia dan level glukosa rata-rata sebagai variabel prediktor. Fokus utama adalah pada tantangan ketidakseimbangan kelas dalam data sosial dan kesehatan, yang mempengaruhi performa model dalam mengidentifikasi kelas minoritas. Dalam era Big Data, data mining dan *machine learning* memungkinkan ekstraksi pola dari data kompleks, tetapi sering kali bias terhadap kelas mayoritas terjadi. Oleh karena itu, proses pra-pemrosesan data yang meliputi teknik penanganan ketidakseimbangan seperti SMOTE penting dilakukan sebelum pembangunan model. Model yang dibangun menggunakan perangkat lunak RapidMiner Studio divalidasi melalui *Cross Validation* 10-Fold guna memastikan generalisasi dan mengurangi risiko overfitting. Hasil evaluasi menunjukkan bahwa meskipun Decision Tree mampu menghasilkan model interpretatif dan cukup akurat, performa pada kelas minoritas masih rendah akibat ketidakseimbangan data, tercermin dari rendahnya nilai *Recall*. Penggunaan teknik *balancing* data seperti SMOTE terbukti meningkatkan *Recall* dan keadilan model, menghasilkan prediksi yang lebih seimbang dan adil. Temuan ini menegaskan pentingnya penanganan ketidakseimbangan data dalam membangun model klasifikasi yang handal serta sebagai panduan pengembangan metode klasifikasi yang lebih efektif dalam bidang sosial dan kesehatan.

**Kata Kunci:** Decision Tree, Ketidakseimbangan, Data Mining, Klasifikasi, Kesehatan

## Abstrak

*This article describes the stages of using the Decision Tree algorithm to build a predictive model for occupational status classification using age and average glucose level attributes as predictor variables. The main focus is on the challenge of class imbalance in social and health data, which affects the model's performance in identifying minority classes. In the Big Data era, data mining and machine learning enable pattern extraction from complex data, but often bias towards the majority class occurs. Therefore, data preprocessing processes that include imbalance handling techniques such as SMOTE are important before model development. The model built using RapidMiner Studio software was validated through 10-Fold Cross Validation to ensure generalization and reduce the risk of overfitting. The evaluation results show that although Decision Tree is capable of producing interpretive and fairly accurate models, its performance on minority classes is still low due to data imbalance, reflected in the low Recall value. The use of data balancing techniques such as SMOTE has been shown to improve model Recall and fairness, resulting in more balanced and fair predictions. These findings emphasize the importance of handling data imbalance in building reliable classification models and serve as a guide for developing more effective classification methods in the social and health fields.*

**Keyword:** Decision Tree, Imbalance, Data Mining, Classification, Health

## PENDAHULUAN

Pemanfaatan teknik Data Mining dan pembelajaran mesin telah menjadi landasan esensial dalam era Big Data, memungkinkan peneliti untuk mengekstrak pola, tren, dan pengetahuan tersembunyi dari himpunan data yang sangat besar dan kompleks di berbagai domain terapan (Saputra, 2025; Putri & Widiarti, 2024).

Dalam konteks analisis data demografi dan sosial, tugas Klasifikasi status pekerjaan individu yang merupakan variabel penting dalam pemodelan stabilitas sosial dan perencanaan kebijakan publik memiliki signifikansi yang tinggi. Prediksi status pekerjaan didukung oleh berbagai atribut pendukung, termasuk faktor *Age* (Usia) dan potensi indikator kesehatan seperti *Average Glucose Level* (Tingkat Glukosa Rata-rata). Usia secara universal diakui sebagai faktor demografi penentu dalam struktur angkatan kerja dan pengambilan keputusan, sementara variabel terkait kesehatan dapat memberikan wawasan mengenai faktor risiko atau kondisi yang memengaruhi kemampuan individu untuk berpartisipasi dalam pasar kerja (Simarmata, 2024).

Untuk membangun model prediktif dalam tugas klasifikasi ini, algoritma Decision Tree dipilih. Keputusan ini didasarkan pada keunggulan metodologis Decision Tree, yaitu kemampuannya dalam menangani data yang tidak linier dan yang paling penting, menghasilkan model yang sangat interpretable (mudah dipahami) dalam bentuk struktur pohon keputusan. Keunggulan interpretasi ini memungkinkan peneliti untuk mengidentifikasi secara hierarkis atribut mana yang paling menentukan pemecahan kelas, memberikan kejelasan yang sering hilang dalam model black box yang lebih kompleks (Qisthiano, Prayesy, & Ruswita, 2023; Muharram & Kusuma, 2024; Nurdiansyah, 2024).

Penerapan algoritma Decision Tree telah terbukti efektif dalam beragam studi klasifikasi di Indonesia, mulai dari identifikasi risiko drop out mahasiswa hingga prediksi pelanggan berisiko tinggi

dalam penjualan, menegaskan keandalan metodologinya (Sari, 2023; Sapitri & Costaner, 2025). Guna menjamin validitas dan generalisasi hasil, seluruh proses data mining dilakukan menggunakan perangkat lunak RapidMiner Studio, dan model divalidasi dengan teknik Cross Validation (CV) K-Fold, sebuah prosedur standar yang meminimalkan risiko overfitting dan memberikan estimasi kinerja model yang robust pada data yang tidak terlihat (Syahputra & Andryani, 2024; Wijaya, 2023).

Meskipun penerapan metodologi yang robust telah dilakukan, implementasi model klasifikasi pada data status pekerjaan seringkali menghadapi hambatan empiris yang serius yang dikenal sebagai masalah ketidakseimbangan kelas (Imbalanced Dataset). Masalah ini muncul ketika distribusi sampel antara kelas target tidak merata, di mana kelas mayoritas (kelas yang dominan) jauh melebihi jumlah kelas minoritas (kelas yang kurang terwakili).

Dalam konteks penelitian riil, kondisi ini lumrah terjadi pada data sosial-demografi, yang menciptakan kesenjangan penelitian metodologis yang signifikan. Secara struktural, algoritma klasifikasi default, termasuk Decision Tree, dirancang untuk mengoptimalkan Akurasi (Accuracy) keseluruhan, yang menyebabkan bias alami; model cenderung terfokus pada kinerja prediksi kelas mayoritas karena kesalahan prediksi pada kelas minoritas hanya sedikit memengaruhi akurasi total (Kaope & Pristyanto, 2023; Hidayat & Yulia, 2024).

Konsekuensi dari bias ini sangat merugikan: meskipun akurasi keseluruhan yang dicapai model terlihat memadai, metrik kinerja pada kelas minoritas, khususnya Daya Panggil (*Recall*), mengalami penurunan drastis, seringkali mendekati angka nol (Anggraini & Sujana, 2024). Hal ini berarti bahwa model gagal mengidentifikasi atau mengenali pola pada sampel minoritas, sehingga model menjadi tidak berguna untuk tujuan prediksi kategori yang kurang terwakili, meskipun

secara teknis model telah "berhasil" pada mayoritas data. Berbagai studi terbaru secara konsisten menyoroti pentingnya mengatasi isu ini, menegaskan bahwa tanpa intervensi penanganan *imbalance* data seperti teknik penyeimbangan data *oversampling* yang spesifik, yaitu algoritma SMOTE dan ADASYN model klasifikasi tidak akan mampu mencapai kinerja yang seimbang dan adil di semua kelas target (Nurhayati & Rahardi, 2025; Putra & Azhari, 2025). Oleh karena itu, akurasi total tidak boleh menjadi satu-satunya indikator kinerja, melainkan harus didukung oleh analisis Precision dan Recall per kelas.

Berdasarkan latar belakang masalah dan kesenjangan penelitian yang teridentifikasi, penelitian ini memiliki dua tujuan utama: (1) Mengimplementasikan dan mengevaluasi kinerja algoritma Decision Tree dalam tugas klasifikasi status pekerjaan dengan menggunakan atribut Age dan Average Glucose Level, dan (2) Menganalisis secara kritis dan mendemonstrasikan secara empiris dampak langsung dari kondisi ketidakseimbangan data terhadap metrik evaluasi model, dengan fokus pada perbedaan kinerja Recall antara kelas mayoritas dan kelas minoritas. Kontribusi utama dari penelitian ini adalah menyajikan bukti empiris yang terperinci mengenai keterbatasan model klasifikasi dalam menghadapi imbalanced dataset dan menyediakan landasan argumentasi yang kuat bagi peneliti di masa depan. Rekomendasi yang akan dihasilkan akan mengarahkan pada kebutuhan untuk mengintegrasikan teknik data balancing pada tahap pra-pemrosesan data untuk mengurangi bias dan menghasilkan model prediktif yang adil, valid, dan efektif di seluruh kategori status pekerjaan.

## METODE

Bagian ini menguraikan secara rinci prosedur dan teknik yang digunakan dalam penelitian, mulai dari pengumpulan data, tahapan preprocessing, model/algoritma hingga pengujian. Keseluruhan proses data

mining dilakukan menggunakan perangkat lunak RapidMiner Studio.

### Jenis Penelitian

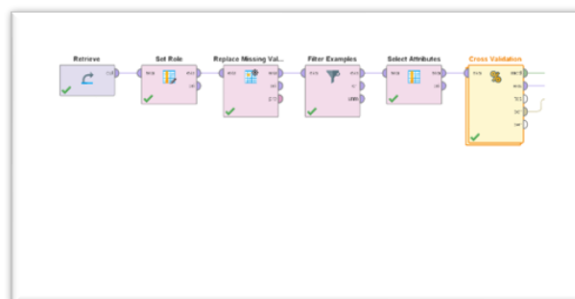
Studi ini menerapkan pendekatan kuantitatif, menggunakan teknik Data Mining dengan fokus utama tertuju pada tugas Klasifikasi untuk memprediksi status pekerjaan.

1. Algoritma : Algoritma utama yang diimplementasikan adalah Decision Tree (Pohon Keputusan), dipilih karena kemampuannya dalam memodelkan hubungan non-linear dan menghasilkan aturan yang mudah diinterpretasikan.
2. Perangkat Lunak : Seluruh tahapan proses data mining dan validasi model dilakukan menggunakan perangkat lunak RapidMiner Studio, yang memungkinkan implementasi alur kerja visual dan evaluasi kinerja model secara komprehensif.

### 1. Pengumpulan Data

Dataset yang digunakan dalam penelitian ini merupakan dataset sekunder yang berasal dari kaggle (<https://www.kaggle.com/datasets/godfath/erfigure/healthcare-dataset-stroke-data>). Secara khusus membatasi ruang fokus hanya pada dua atribut prediktor utama, yaitu Age (usia) dan Average Glucose Level (Tingkat Glukosa Rata-Rata) Atribut Status Pekerjaan ditetapkan sebagai variabel target yang memiliki lima kelas kategori, yaitu *Private*, *Self-employed*, *Govt\_job*, *children*, dan *Never\_worked*.

### 2. Pra-pemrosesan (Preprocessing)



Gambar 1. PraProses pada RapidMiner Studio

Pra-pemrosesan dilakukan untuk memastikan bahwa data dalam kondisi

optimal sebelum digunakan untuk pemodelan. Tahapan ini dikerjakan menggunakan RapidMiner Studio dan meliputi beberapa proses utama:

**a. Set Role**

- Pada tahap ini, dataset yang diambil dari *kaggle* dimasukkan ke operator *Set Role* untuk menetapkan peran setiap atribut.
- Atribut Status Pekerjaan ditetapkan sebagai Label, yakni output yang ingin diprediksi.
- Atribut *Age* dan *Average Glucose Level* dipertahankan sebagai *Regular Attribute*.

**b. Replace Missing Values**

Metode imputasi dipilih secara otomatis untuk mempertahankan konsistensi dan integritas dataset.

- Operator *Filter Examples* digunakan apabila terdapat data yang tidak memenuhi syarat untuk dimasukkan dalam proses pemodelan, misalnya outlier tertentu atau baris dengan nilai yang tidak logis.

**c. Select Attributes**

Operator ini memastikan bahwa hanya dua atribut utama (*Age* dan *Average Glucose Level*) yang dipertahankan dalam dataset untuk pemodelan. Langkah ini membantu :

- Membantu mengurangi kompleksitas data.
- Mempercepat waktu komputasi.
- Mengurangi risiko overfitting yang dapat muncul dari atribut yang tidak relevan

**3. Model / Algoritma Decision Tree**

Proses inti pemodelan menggunakan algoritma *Decision Tree*. Proses ini dilakukan di dalam *sub-process Cross Validation*.

**a. Proses Training**

Pada tahap training, data latih diproses untuk membangun model *Decision Tree*.

**1. Data latih masuk ke operator Decision Tree.**

Data (*tra*) digunakan sebagai

dasar pembentukan model.

**2. Perhitungan ukuran pemisahan (split criteria).**

Algoritma mengevaluasi atribut menggunakan:

- Entropy
- Information Gain
- Gini Index

untuk memilih atribut terbaik sebagai node pemisah.

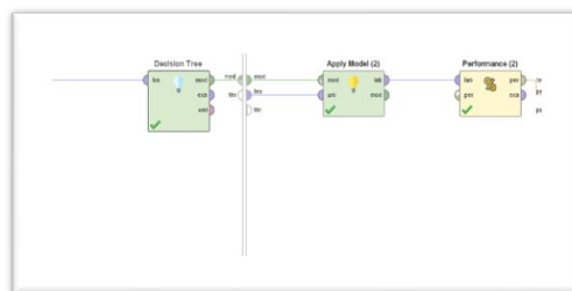
**3. Pembentukan pohon keputusan**

Berdasarkan nilai split terbaik, model membentuk struktur pohon yang berisi:

- Node keputusan
- Cabang aturan
- Daun (kelas akhir)

**4. Output**

Hasil proses ini adalah model terlatih (*mod*) yang berisi aturan klasifikasi berbasis hubungan antara *Age* dan *Average Glucose Level*



Gambar 2. Langkah Training & Testing

**4. Pengujian Model**

**b. Proses Testing**

**1. Apply Model**

- Model terlatih (*mod*) diterapkan pada data uji (*tes*).
- Operator menghasilkan label prediksi (*lab*) untuk setiap baris data uji berdasarkan aturan pohon keputusan.

**2. Performance**

Hasil prediksi dibandingkan dengan label asli menggunakan operator *Performance*.

Metrik evaluasi yang dihasilkan meliputi:

- Confusion Matrix

- Akurasi
- Precision untuk setiap kelas
- Recall untuk setiap kelas

**HASIL DAN PEMBAHASAN**

**1. Hasil Preprocessing dan Analisis Awal**

	true Private	true Self-emplo...	true Govt_job	true children	true Never_wor...	class precision
sd Private	2814	748	638	41	15	66.12%
sd Self-empl...	71	63	15	7	0	40.38%
sd Govt_job	4	1	2	0	0	28.57%
sd children	34	7	2	639	7	92.74%
sd Never_w...	2	0	0	0	0	0.00%
ss recall	96.21%	7.69%	0.30%	93.01%	0.00%	

**Gambar 4 Confusion Matrix**

Dataset yang digunakan berisi atribut *Age*, *Average Glucose Level*, dan *Work Type* sebagai label yang ingin diprediksi. Tahap *preprocessing* dilakukan berupa pembersihan data hilang dan pemilihan fitur yang relevan.

Analisis awal menunjukkan bahwa jumlah data pada setiap kelas tidak seimbang kelas *Private* jauh lebih banyak dibanding kelas *Govt job* dan *Never worked*. Kondisi ini mempengaruhi *performa* model pada tahap evaluasi.

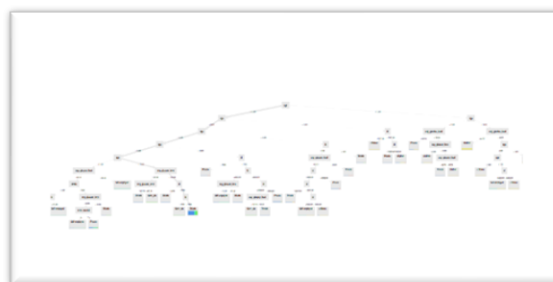
**Tabel 1. Distribusi Data pada Setiap Kategori Work Type**

Work Type	Jumlah Data
Private	4.256
Self-employed	156
Govt_job	7
children	689
Never_worked	2

**2. Hasil Pemodelan Decision Tree**

Model Decision Tree dibangun menggunakan metode *10-Fold Cross Validation*. Algoritma memilih atribut pemecah berdasarkan nilai *Entropy*, *Information Gain*, dan *Gini Index*. Hasil pemodelan menunjukkan bahwa:

- *Average Glucose Level* menjadi pemecah (*split*) pertama dan berfungsi sebagai root node.
- *Age* menjadi atribut pemecah pada node-node berikutnya.



**Gambar 3 Hasil Pohon Keputusan**

**3. Hasil Evaluasi Model**

Model dievaluasi menggunakan *Confusion Matrix* serta perhitungan akurasi, *precision*, dan *recall*. Akurasi yang diperoleh adalah 68,85%.

Hasil prediksi menunjukkan pola:

- Model sangat baik memprediksi *children* dan *Private*.
- Model sangat buruk memprediksi *Govt job* dan *Never worked*.
- Hal ini konsisten dengan ketidakseimbangan data.

**Tabel 2. Confusion Matrix dan Kinerja Klasifikasi Status Pekerjaan**

True \ Predicted	Private	Self-employed	Govt_job	Children	Never_worked	Class Precision
Private	2814	748	638	41	15	66.12%
Self-employed	71	63	15	7	0	40.38%
Govt_job	4	1	2	0	0	28.57%
Children	34	7	2	639	7	92.74%
Never_worked	2	0	0	0	0	0.00%
<b>Class Recall</b>	<b>96.21%</b>	<b>7.69%</b>	<b>0.30%</b>	<b>93.01%</b>	<b>0.00%</b>	<b>Accuracy: 68.85%</b>

**Tabel 3. Nilai Precision dan Recall Tiap Kategori Work Type**

Work Type	Precision	Recall
Private	66.12%	96.21%
Self-employed	40.38%	7.69%
Govt_job	28.57%	0.30%
children	92.74%	93.01%
Never_worked	0.00%	0.00%

**4. Pembahasan Hasil**

**1. Pengaruh Ketidakseimbangan Kelas**

Merujuk pada Tabel 1, ketidakseimbangan jumlah data menyebabkan model lebih dominan

mempelajari pola kelas mayoritas. Hal ini menjelaskan:

- *Recall Private* sangat tinggi (96%)
- *Recall kelas Govt job* dan *Never worked* sangat rendah (0–3%)
- Model “cenderung menebak” kelas mayoritas ketika ragu.

Kosekuensinya, meskipun akurasi cukup baik, performa sebenarnya tidak merata untuk semua kelas.

## 2. Interpretasi Pohon Keputusan

Berdasarkan Gambar 4, struktur pohon memperlihatkan bahwa :

- Glukosa tinggi/normal menjadi faktor utama pemisahan data.
- Usia selanjutnya membedakan antara kategori children dan orang dewasa.
- Hanya dua atribut prediktor menyebabkan pohon keputusan terbatas dan tidak mampu menangkap variasi pekerjaan dengan baik.

Hal ini menjelaskan mengapa model cenderung memetakan banyak data dewasa ke katagori mayoritas (*Private*).

## 3. Evaluasi Performa Model

Dari Tabel 2 (*Confusion Matrix*) dan Tabel 3 (*Precision & Recall*) dapat disimpulkan :

- Performa model tidak merata antar kelas.
- Kelas anak-anak (*children*) mudah teridentifikasi karena ciri usia yang jelas.
- Kelas pekerjaan lain sering tumpang tindih karena kurangnya atribut pembeda.

Sehingga akurasi 68,85% tidak sepenuhnya mencerminkan kualitas model secara keseluruhan.

## 4. Kelebihan dan Keterbatasan Model

Kelebihan :

- Mudah diinterpretasi (melalui pohon keputusan).
- Menghasilkan aturan-aturan klasifikasi yang eksplisit.

- Baik sebagai baseline model.

Keterbatasan :

- Sangat sensitif terhadap ketidakseimbangan data
- Kinerja buruk pada kelas minoritas
- Atribut prediktor terlalu sedikit untuk memprediksi kategori pekerjaan kompleks

## SIMPULAN (PENUTUP)

Penelitian ini mengevaluasi kinerja algoritma Decision Tree dalam mengklasifikasikan status pekerjaan berdasarkan fitur *Age* dan *Average Glucose Level*, dengan memanfaatkan skema validasi *Cross Validation* dan diimplementasikan melalui perangkat lunak RapidMiner.

### 1. Kinerja Agregat dan Faktor Dominan

Hasil implementasi model Decision Tree mencapai akurasi agregat sebesar 68.85%. Analisis struktur pohon keputusan mengonfirmasi bahwa Usia (*Age*) merupakan variabel prediktif paling dominan dalam menentukan kategori status pekerjaan.

### 2. Dampak Ketidakseimbangan Kelas

Meskipun akurasi keseluruhan menunjukkan nilai yang memadai, analisis detail pada *Confusion Matrix* secara tegas mengungkapkan adanya bias model yang signifikan, yang disebabkan oleh distribusi kelas yang tidak seimbang (*Imbalanced Dataset*).

### 3. Disparitas Kinerja Recall

Kelas Mayoritas: Kinerja model untuk kelas dominan (*Private dan children*) menunjukkan nilai *Recall* yang sangat tinggi (melampaui 93%), mengonfirmasi kapabilitas model yang efektif dalam mengidentifikasi observasi dari kelas yang terwakili secara berlebih.

Kelas Minoritas: Sebaliknya, kelas minoritas (*Govt\_job, Self-employed, dan Never\_worked*) menunjukkan degradasi kinerja yang parah, dengan nilai *Recall* mendekati nol. Hal ini

mengindikasikan bahwa model hampir gagal total dalam mengidentifikasi individu yang termasuk dalam kategori pekerjaan dengan jumlah sampel kecil.

Simpulan utamanya adalah bahwa model Decision Tree yang dikembangkan tidak memiliki reliabilitas yang memadai untuk prediksi status pekerjaan minoritas. Akurasi 68.85% merupakan artefak dari klasifikasi kelas mayoritas yang efektif dan tidak mencerminkan kemampuan generalisasi model untuk membedakan kategori minoritas.

Sebagai tindak lanjut dari keterbatasan ini, sangat disarankan untuk menerapkan teknik penanganan ketidakseimbangan data (seperti SMOTE) sebelum proses pemodelan. Hal ini bertujuan untuk meningkatkan Recall pada kelas minoritas, sehingga menghasilkan model klasifikasi yang lebih adil dan kuat.

#### UCAPAN TERIMA KASIH

Penulis menyampaikan apresiasi dan terima kasih yang setinggi-tingginya kepada pihak-pihak yang telah berperan penting dalam pelaksanaan dan penyelesaian penelitian ini. Ucapan terima kasih khusus ditujukan pembimbing atas bimbingan, arahan, dan dukungan yang tak ternilai sepanjang pelaksanaan studi ini. Penghargaan tulus juga diberikan kepada seluruh rekan seperjuangan atas motivasi dan kolaborasi yang terjalin, yang turut berkontribusi dalam penyelesaian studi ini. Melalui karya berjudul “Klasifikasi Risiko Stroke Menggunakan Metode Decision Tree pada Healthcare Dataset Stroke Data”, besar harapan penulis agar temuan-temuan di dalamnya mampu menawarkan kontribusi nyata dan dampak yang bermanfaat bagi pembaca, khususnya dalam memajukan penerapan data mining di bidang kesehatan dan prediksi risiko penyakit.

#### DAFTAR PUSTAKA

Saputra, A. (2025). Analisis Sentimen Pengguna X Terhadap Kebocoran Data Pribadi Menggunakan Algoritma

Naïve Bayes Classifier. *Jurnal XYZ*, 14(01), 32–40.

Putri, D. A. & Widiarti, A. (2024). Peran Data Mining dalam Prediksi Status Sosial dan Ekonomi Masyarakat. *Jurnal Sosial Demografi*, 5(1), 12-25.

Simarmata, R. (2024). Korelasi Antara Indikator Kesehatan (Tingkat Glukosa) dan Status Ketenagakerjaan di Indonesia. *Jurnal Kesehatan dan Bisnis*, 3(2), 55-68.

Qisthiano, M. R., Prayesy, P. A., & Ruswita, I. (2023). Penerapan Algoritma Decision Tree dalam Klasifikasi Data Prediksi Kelulusan Mahasiswa. *G-Tech: Jurnal Teknologi Terapan*, 7(1), 21–28. doi: 10.33379/gtech.v7i1.1850.

Muharram, A. N. & Kusuma, D. (2024). Pengaruh Pemilihan Atribut Root Node terhadap Akurasi Decision Tree dalam Klasifikasi. *Jurnal Ilmu Komputer dan Teknologi Informasi*, 7(3), 201-209.

Nurdiansyah, I. (2024). Studi Komparatif Visualisasi Aturan Decision Tree untuk Interpretasi Model Klasifikasi. *Jurnal Teknika*, 16(1), 88-95.

Sari, A. (2023). Penerapan Data Mining Menggunakan Algoritma Decision Tree C4.5 Untuk Memprediksi Mahasiswa Drop Out Di Universitas Wiraraja. *Jurnal Advance Research Informatika (Jar's)*, 1(2), 1–7. doi: 10.24929/jars.v1i02.2684.

Sapitri, R. & Costaner, L. (2025). Penggunaan Data Mining Untuk Mengidentifikasi Pelanggan Beresiko Tinggi Dalam Penjualan Menggunakan Algoritma Decision Tree C4.5. *Jurnal Ilmiah Multidisiplin Ilmu*, 1(1), 46–51.

Syahputra, R. & Andryani, R. (2024). Validasi Model Klasifikasi dengan Teknik K-Fold Cross Validation dalam Lingkungan RapidMiner. *Jurnal Komputer dan Sains*, 9(1), 77-84.

Wijaya, S. (2023). Pentingnya Kualitas Data dan Pra-pemrosesan dalam Membangun Model Machine

- Learning. *Jurnal Rekayasa Informasi*, 15(4), 301-310.
- Kaope, C. & Pristyanto, Y. (2023). The Effect of Class Imbalance Handling on Datasets Toward Classification Algorithm Performance. *MATRIK Jurnal Manajemen, Teknik Informatika dan Rekayasa Komputer*, 22(2), 227–238. doi: 10.30812/matrik.v22i2.2515.
- Hidayat, R. & Yulia, M. (2024). Evaluasi Metrik Klasifikasi Akurasi, Presisi, dan Recall pada Studi Data Imbalance. *Jurnal Sistem Informasi*, 18(2), 112-120.
- Anggraini, L. & Sujana, D. (2024). Analisis Kinerja Decision Tree pada Data Klasifikasi dengan Skema Cross Validation K-Fold. *Jurnal Informatika Terapan (JIT)*, 10(1), 45-53.
- Nurhayati, L. D. & Rahardi, M. (2025). Impact of SMOTE and ADASYN on Class Imbalance in Metabolic Syndrome Classification Using Random Forest Algorithm. *Journal of Applied Informatics and Computing (JAIC)*, 9(5), 2807–2813.
- Putra, A. B. & Azhari, M. (2025). Implementasi Algoritma Decision Tree Untuk Klasifikasi Pemilihan Sekolah Lanjutan Di MTs Al-Hasanah Medan. *JSI (Jurnal Sistem Informasi) Universitas Suryadarma*, 12(2), 128–137. doi: 10.35968/jsi.v12i2.1520.