

## Penerapan *Recursive Feature Elimination* (RFE) pada *Tree-Based Classifier* untuk Identifikasi Risiko Diabetes

Nadia Annisa Maori<sup>1\*</sup>, Noor Azizah<sup>2</sup>

<sup>1</sup> Jurusan Teknik Informatika, Fakultas Sains Dan Teknologi, Universitas Islam Nahdlatul Ulama Jepara

<sup>2</sup> Jurusan Sistem Informasi, Fakultas Sains Dan Teknologi, Universitas Islam Nahdlatul Ulama Jepara

\*Email: nadia@unisnu.ac.id

### Abstrak

Diabetes mellitus adalah penyakit kronis yang umum dengan dampak global yang signifikan. Identifikasi dini individu berisiko tinggi terkena diabetes sangat penting untuk pencegahan dan manajemen penyakit ini. Penelitian ini mengeksplorasi penggunaan *Recursive Feature Elimination* (RFE) pada classifier berbasis pohon keputusan (*tree-based classifiers*) untuk meningkatkan akurasi prediksi risiko diabetes. Dataset Pima Indians Diabetes Database (PIDD) digunakan sebagai basis data, dan algoritma seperti *Decision Tree*, *Random Forest*, *Gradient Boosting*, dan *Xtreme Gradient Boosting* diuji. Hasil penelitian menunjukkan bahwa penerapan RFE meningkatkan akurasi model, dengan *Random Forest* dan *Gradient Boosting* mencapai akurasi tertinggi sebesar 77.27%. RFE juga berhasil mengidentifikasi fitur-fitur yang paling relevan, mengurangi risiko overfitting, dan meningkatkan interpretasi model. Penelitian ini memberikan dasar kuat untuk pengembangan alat prediksi yang lebih efektif dalam manajemen dan pencegahan diabetes. Penelitian selanjutnya disarankan untuk menguji generalisasi pendekatan ini pada dataset yang lebih luas dan dalam berbagai konteks klinis.

**Kata kunci:** diabetes, recursive feature selection, tree-based classifier.

### Abstract

*Diabetes mellitus is a common chronic disease with significant global impact. Early identification of individuals at high risk of developing diabetes is critical for the prevention and management of the disease. This study explores the use of Recursive Feature Elimination (RFE) in decision tree-based classifiers to improve the accuracy of diabetes risk prediction. The Pima Indians Diabetes Database (PIDD) dataset was used as the database, and algorithms such as Decision Tree, Random Forest, Gradient Boosting, and Xtreme Gradient Boosting were tested. The results showed that the application of RFE improved the model accuracy, with Random Forest and Gradient Boosting achieving the highest accuracy of 77.27%. RFE also successfully identified the most relevant features, reduced the risk of overfitting, and improved model interpretability. This study provides a strong foundation for the development of more effective predictive tools in diabetes management and prevention. Future studies are recommended to test the generalizability of this approach to a wider dataset and in various clinical contexts.*

**Keywords:** diabetes, recursive feature selection, tree-based classifier.

## PENDAHULUAN

Diabetes mellitus adalah salah satu penyakit kronis yang paling umum di dunia, yang memengaruhi jutaan orang dari berbagai latar belakang dan usia. Kondisi ini ditandai oleh tingginya kadar gula dalam darah, yang dapat menyebabkan berbagai komplikasi serius seperti penyakit jantung, kerusakan saraf, dan kerusakan ginjal (Budianto et al., 2022). Oleh karena itu, identifikasi dini individu yang berisiko tinggi terkena diabetes sangat penting untuk pencegahan dan manajemen penyakit ini.

Dalam beberapa tahun terakhir, kemajuan dalam teknologi informasi dan ketersediaan data medis yang besar telah membuka peluang baru dalam analisis risiko kesehatan menggunakan metode pembelajaran mesin (*machine learning*). Pembelajaran mesin memungkinkan identifikasi pola-pola kompleks dalam data yang mungkin tidak terlihat oleh metode analisis tradisional. Salah satu pendekatan yang efektif untuk analisis data kesehatan adalah penggunaan *classifier* berbasis pohon keputusan (*tree-based classifiers*) seperti *Decision Tree*, *Random Forest*, dan *Gradient Boosting*.

Meskipun *tree-based classifiers* memiliki keunggulan dalam hal interpretabilitas dan kemampuan menangani data dengan fitur non-linear, mereka sering kali menghadapi tantangan dalam hal pemilihan fitur yang relevan. Fitur yang tidak relevan atau berlebihan dapat memperburuk performa model dan meningkatkan risiko *overfitting* (Wahyuni, 2016). Oleh karena itu, pemilihan fitur yang efektif merupakan langkah penting dalam pengembangan model prediksi yang handal (Gunawan et al., 2023).

*Recursive Feature Elimination* (RFE) adalah salah satu metode pemilihan fitur yang dapat membantu mengatasi tantangan ini (Edwar et al., 2023). RFE bekerja dengan cara menghapus fitur secara rekursif dan membangun model pada setiap iterasi untuk mengidentifikasi fitur-fitur yang paling penting (Pratama et al., 2022). Dengan demikian, RFE dapat meningkatkan akurasi model sekaligus mengurangi kompleksitasnya.

Penelitian sebelumnya yang berjudul “Klasifikasi Penyakit Diabetes Menggunakan Algoritma *Decision Tree*” menghasilkan akurasi tinggi. Namun, nilai *recall* yang rendah menunjukkan adanya keterbatasan kinerja dalam proses klasifikasi. Menentukan atribut relevan menjadi alternatif lanjutan untuk meningkatkan

kinerja model dalam proses klasifikasi (Nurussakinah & Faisal, 2023).

Dalam konteks identifikasi risiko diabetes, penerapan RFE pada *tree-based classifier* dapat memberikan manfaat signifikan. Dengan menggunakan RFE, kita dapat menyaring fitur-fitur yang paling relevan dari dataset medis yang besar, sehingga menghasilkan model prediksi yang lebih akurat dan efisien. Penelitian ini bertujuan untuk mengeksplorasi dan menguji efektivitas penggunaan RFE pada *tree-based classifier* dalam mengidentifikasi individu yang berisiko tinggi terkena diabetes.

Penelitian ini diharapkan dapat memberikan kontribusi dalam bidang kesehatan dengan menawarkan pendekatan analitis yang dapat diandalkan untuk deteksi dini risiko diabetes. Selain itu, hasil dari penelitian ini dapat digunakan oleh praktisi kesehatan dan peneliti lain untuk mengembangkan strategi pencegahan dan intervensi yang lebih efektif, sehingga membantu mengurangi beban penyakit diabetes di masyarakat.

## TINJAUAN PUSTAKA

Penelitian terdahulu yang relevan dengan klasifikasi dalam mengidentifikasi penyakit resiko diabetes dilakukan (Nurussakinah & Faisal, 2023) menggunakan algoritma *decision tree* menunjukkan nilai akurasi yang tinggi, yaitu 0.78 dan rendahnya nilai *recall*, yaitu 0.45. Hal ini disebabkan adanya proses pengklasifikasian yang terbatas. Terdapat kesalahan dalam proses klasifikasi karena terdapat beragamnya variabel yang mempengaruhi tingkat gula darah seseorang. Selanjutnya, pada penelitian yang dilakukan (Pratama et al., 2022) menerapkan RFE untuk seleksi fitur pada algoritma SVM. Penelitian ini menghasilkan peningkatan nilai akurasi sebesar 2% jika dibandingkan tanpa RFE.

Dari hasil penelitian terdahulu menjadi acuan peneliti untuk menerapkan seleksi fitur RFE pada algoritma *tree-based classifier* pada identifikasi resiko penyakit diabetes untuk mendapatkan peningkatan akurasi yang optimal, sehingga deteksi dini resiko diabetes dapat diandalkan dalam dunia medis.

### 2.1 Tree-Based Clasifiers

#### 2.1.1 Decision Tree

*Decision Tree* adalah metode yang sering digunakan dalam pembelajaran mesin untuk

klasifikasi dan regresi karena kesederhanaan dan kemudahan interpretasinya. Metode ini memisahkan data menjadi cabang-cabang berdasarkan nilai fitur tertentu, mulai dari akar sebagai titik awal. Setiap cabang mengarah ke node yang melakukan pengambilan keputusan hingga mencapai daun yang memberikan hasil akhir prediksi (Nasrullah, 2021). Proses ini memungkinkan *Decision Tree* untuk membuat keputusan yang jelas dan mudah dipahami, sehingga memudahkan interpretasi hasil model (Antika et al., 2023).

### 2.1.2 Random Forest

*Random Forest* merupakan *ensemble* dari beberapa *decision tree* untuk meningkatkan akurasi dan stabilitas prediksi. *Random Forest* mengurangi risiko *overfitting* yang sering terjadi pada *decision tree* individual dengan melakukan averaging prediksi dari beberapa *tree* yang dibangun pada subset data yang berbeda (Khoeruddin et al., 2023).

### 2.1.3 Gradient Boosting

*Gradient Boosting* adalah teknik yang menggabungkan beberapa *decision tree* secara iteratif, di mana setiap *tree* baru diperbaiki berdasarkan kesalahan dari *tree* sebelumnya (Suryana, Warsito, et al., 2021). Metode ini dikenal sangat efektif namun memerlukan pengaturan parameter yang hati-hati untuk menghindari *overfitting*.

### 2.1.4 Extreme Gradient Boosting

XGBoost (*Extreme Gradient Boosting*) merupakan implementasi dari algoritma *gradient boosting* yang dirancang untuk efisiensi dan kinerja tinggi. XGBoost telah menjadi populer karena kemampuannya dalam menangani data dengan volume besar dan kompleksitas tinggi (Rizky et al., 2022). XGBoost menggunakan berbagai teknik optimasi, seperti *tree pruning*, regularisasi, dan *parallel processing*, untuk meningkatkan akurasi prediksi dan mengurangi risiko *overfitting*.

## 2.2 Recursive Feature Elimination (RFE)

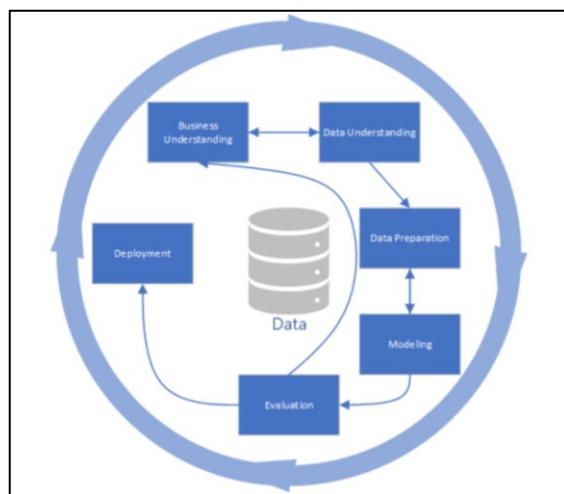
RFE adalah teknik pemilihan fitur yang populer dalam pembelajaran mesin. Metode ini bekerja dengan cara menghapus fitur secara bertahap dan membangun model pada setiap iterasi untuk mengidentifikasi fitur-fitur yang

paling penting (Siswa & Wibowo, 2023). Proses ini diulangi hingga tersisa fitur-fitur yang paling relevan untuk model. RFE efektif dalam meningkatkan performa model dengan mengurangi dimensi data dan menghilangkan fitur yang tidak relevan atau redundan.

Dengan demikian, kombinasi antara RFE dan *tree-based classifiers* menawarkan pendekatan yang kuat untuk identifikasi risiko diabetes. RFE membantu mengidentifikasi fitur-fitur yang paling relevan, sementara *tree-based classifiers* memberikan model prediksi yang akurat dan interpretatif. Penelitian ini akan mengeksplorasi efektivitas pendekatan ini dalam mengidentifikasi individu yang berisiko tinggi terhadap diabetes, serta potensi penerapannya dalam praktik klinis.

## METODE

Penelitian ini bertujuan untuk mengembangkan dan mengevaluasi model prediksi risiko diabetes dengan menggunakan kombinasi *Recursive Feature Elimination* (RFE) dan *tree-based classifiers*. Metode yang diterapkan adalah Crisp-DM, Gambar 1. menggambarkan metode yang digunakan dalam penelitian.



Gambar 1. Metode Crisp-DM  
(Ajitomo & Pratama, 2024)

### 1. Business Understanding

Pada fase ini, tujuan utama adalah memahami tujuan bisnis dan mengubahnya menjadi tujuan data mining yang spesifik. Tujuan bisnisnya untuk mengidentifikasi

individu yang berisiko tinggi terkena diabetes untuk meningkatkan pencegahan dan intervensi dini dan tujuan data mining untuk mengembangkan model prediksi risiko diabetes dengan menggunakan kombinasi RFE dan *tree-based classifiers*.

## 2. Pemahaman Data

Data yang digunakan dalam penelitian ini diambil dari dataset kesehatan yang berisi informasi tentang individu, termasuk data demografi, hasil tes laboratorium, dan riwayat kesehatan. Dataset yang digunakan adalah dataset diabetes dari [Pima Indians Diabetes Database \(PIDD\)](#) yang tersedia secara publik. Dataset ini terdiri dari beberapa fitur seperti kehamilan, usia, berat badan, tekanan darah, kadar glukosa, ketebalan kulit, Insulin, dan fungsi pedigree diabetes yang berjumlah sebanyak 768 records. Tabel 1 menunjukkan dataset yang diterapkan dalam penelitian

Tabel 1. Dataset PIDD

Pregnancies	Glucose	....	Outcome
6	148	...	1
1	85	...	0
8	183	...	1
1	89	...	0
0	137	...	1
...	...	...	...
1	93	...	0

## 3. Persiapan Data

Data dipersiapkan untuk menghasilkan data yang berkualitas. Tahapan yang dilakukan dimulai dari pembersihan data untuk menyelesaikan permasalahan data yang kosong. Selanjutnya, pemilihan atribut yang digunakan dalam penelitian. Atribut yang digunakan terdapat 9, yaitu *Pregnancies*, *Glucose*, *BloodPressure*, *SkinThickness*, *Insulin*, *BMI*, *DiabetesPedigreeFunction*, dan *Age*, sementara atribut *Outcome* menjadi label. Tahap terakhir adalah memisahkan data menjadi set pelatihan (80%) dan set pengujian (20%) secara acak.

## 4. Pemodelan

Tahap pemodelan melibatkan pemilihan model, pelatihan model, dan evaluasi model. Beberapa model *tree-based classifier* akan digunakan, yaitu *decision tree*, *random forest*, *gradient boosting*, dan *Xtreme gradient boosting*. Pemilihan fitur menggunakan RFE

untuk meningkatkan akurasi pada algoritma klasifikasi berbasis *tree*.

## 5. Evaluasi

Fase ini melibatkan evaluasi model untuk memastikan bahwa model memenuhi tujuan bisnis yang telah ditetapkan. Evaluasi yang digunakan adalah tingkat akurasi, *precision*, *recall*, dan *F1-score* untuk mengukur keakuratan model dalam mengklasifikasikan data (Mulyo & Maori, 2024).

## 6. Penyajian

Pada fase ini, model yang telah dievaluasi dan dipilih akan diimplementasikan dalam lingkungan produksi.

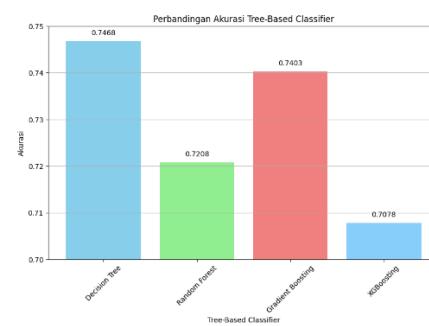
## HASIL DAN PEMBAHASAN

Data yang diperoleh melalui tahap pemrosesan, selanjutnya dievaluasi menggunakan algoritma *tree-based classifier* dan RFE untuk mendapatkan model terbaik dengan pemilihan fitur yang saling relevan.

### 4.1 Hasil Evaluasi Model

#### 4.1.1 Evaluasi Tree-Based Classifier

Hasil evaluasi algoritma *tree-based classifier* ditunjukkan pada Gambar 2.



Gambar 2. Perbandingan Akurasi Tree-Based Classifier

Dari hasil tersebut menunjukkan bahwa algoritma *decision tree* memiliki nilai akurasi tertinggi dibanding dengan yang lainnya, yaitu sebesar 0.7468, sementara *Gradient Boosting* menghasilkan akurasi sebesar 0.7403, *Random Forest* menghasilkan 0.7208, dan *XGBoosting* sebesar 0.7078.

#### 4.1.2 Evaluasi Tree-Based Classifier dan RFE

Hasil evaluasi dengan menerapkan RFE dapat terlihat pada Tabel 2, dimana akurasi

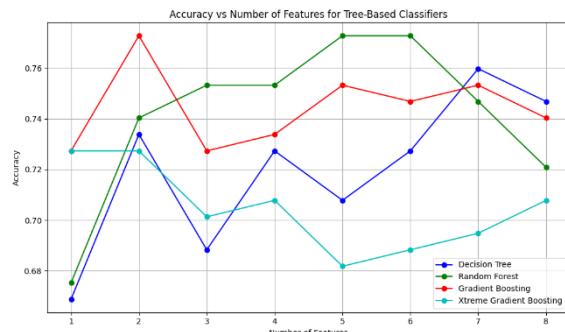
tertinggi dimiliki *Random Forest* dan *Gradient Boosting*, namun memiliki perbedaan pada pemilihan fiturnya. Penggunaan RFE pada *Random Forest* menunjukkan akurasi terbaik sebesar 0.7727 dengan jumlah fitur terbaik sejumlah 5 dan 6 fitur, sementara RFE pada *Gradient Boosting* sebesar 0.7727 sejumlah 2 fitur.

Algoritma lainnya, seperti *Decision Tree* menunjukkan akurasi sebesar 0.7597 dengan fitur terbaik sejumlah 7 dan *XGBoosting* akurasinya mencapai 0.7273 dengan 1 dan 2 fitur terbaiknya.

Tabel 2. Algoritma *Tree-Based Classifier* dan RFE

<i>Tree-Based Classifier + RFE</i>	Akurasi	Jumlah Fitur Terbaik
<i>Decision Tree</i>	0.7597	7
<i>Random Forest</i>	0.7727	5 Dan 6
<i>Gradient Boosting</i>	0.7727	2
<i>Xtreme Gradient Boosting</i>	0.7273	1 Dan 2

Rincian performa RFE pada algoritma klasifikasi berbasis *tree* dapat tergambar pada Gambar 3.



Gambar 3. Perbandingan Akurasi *Tree-Based Classifier* dan RFE

## 4.2 Pembahasan

### 4.2.1 Peningkatan Performa

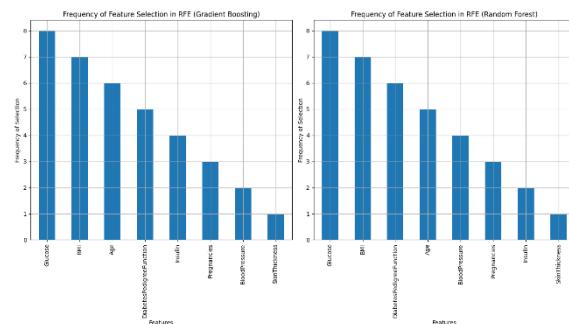
Hasil evaluasi menunjukkan bahwa penerapan *Recursive Feature Elimination* (RFE) berhasil meningkatkan performa sebagian besar model *tree-based classifier* yang diuji. Secara khusus, *Random Forest* dan *Gradient Boosting* menunjukkan peningkatan yang signifikan dalam akurasi setelah penerapan RFE, walaupun dengan pemilihan jumlah fitur yang berbeda. *Random Forest* mengalami peningkatan akurasi

dari 72.08% menjadi 77.27%, sementara *Gradient Boosting* juga mengalami peningkatan dari 74.03% menjadi 77.27%.

### 4.2.2 Analisis Fitur yang Dipilih

RFE membantu mengidentifikasi fitur-fitur yang paling relevan untuk prediksi risiko diabetes. Hasil pengujian menunjukkan bahwa jumlah fitur terbaik yang dipilih bervariasi untuk setiap model.

Fitur-fitur yang dipilih termasuk variabel yang konsisten dengan faktor risiko diabetes yang diketahui dari literatur medis. Perbandingan fitur terbaik dapat terlihat pada Gambar 4.



Gambar 4. Perbandingan Pemilihan Fitur Terbaik RFE

Dari gambar tersebut, fitur terbaik dari *Random Forest* menggunakan RFE adalah Kadar gula, BMI, *diabetes Pedigree Function*, *Age*, *Blood Pressure*, dan atau *Pregnancies*.

Sementara, fitur terpilih dari *Gradient Boosting* dan RFE adalah *Glucose* dan *BMI*.

### 4.2.3 Keunggulan Penggunaan RFE

Penggunaan RFE memberikan beberapa keunggulan yang signifikan:

1. Peningkatan Akurasi: Penerapan RFE meningkatkan akurasi prediksi untuk sebagian besar model.
2. Pemilihan Fitur yang Relevan: RFE membantu dalam menghilangkan fitur-fitur yang kurang relevan, yang dapat meningkatkan interpretasi dan performa model.
3. Pengurangan *Overfitting*: RFE dapat membantu mengurangi *overfitting* dengan menghilangkan fitur-fitur yang tidak relevan.

#### 4.2.4 Keterbatasan dan Tantangan

Meskipun hasil penelitian ini menunjukkan peningkatan yang signifikan, terdapat beberapa keterbatasan:

1. *Generalizability*: Hasil ini mungkin tidak dapat langsung diterapkan pada dataset lain tanpa validasi lebih lanjut.
2. Kompleksitas Implementasi: Meskipun RFE meningkatkan performa, proses implementasinya lebih kompleks dan membutuhkan lebih banyak waktu komputasi dibandingkan model yang tidak menggunakan RFE.
3. Ketergantungan pada Kualitas Data: Keberhasilan RFE sangat tergantung pada kualitas dan kelengkapan data yang digunakan.

#### SIMPULAN

Dalam penelitian ini, penerapan *Recursive Feature Elimination* (RFE) terhadap model-model *tree-based classifier* seperti *decision tree*, *random forest*, *gradient boosting*, dan *Xtreme gradient boosting* telah berhasil meningkatkan akurasi prediksi dalam identifikasi risiko diabetes. Hasil evaluasi menunjukkan bahwa hampir semua model mengalami peningkatan performa setelah RFE, dengan beberapa model mencapai akurasi lebih dari 77%. Fitur-fitur yang dipilih oleh RFE memberikan wawasan yang lebih baik tentang faktor risiko yang signifikan dalam pengembangan diabetes. Pemilihan fitur ini tidak hanya meningkatkan akurasi prediksi, tetapi juga membantu mengurangi *overfitting* dan memungkinkan interpretasi yang lebih baik terhadap model-model ini dalam konteks medis. Temuan ini memberikan dasar yang kuat untuk pengembangan alat prediksi yang lebih efektif dalam manajemen dan pencegahan diabetes, serta mendukung keputusan medis yang lebih baik berdasarkan informasi yang lebih akurat.

Penelitian selanjutnya dapat menguji generalisasi dari pendekatan ini pada dataset yang lebih luas dan dalam berbagai konteks klinis. Selain itu, eksplorasi terhadap teknik pemilihan fitur lainnya juga dapat memberikan wawasan tambahan dalam pengembangan sistem prediksi yang lebih efektif untuk risiko diabetes.

#### DAFTAR PUSTAKA

- Ajitomo, W., & Pratama, I. (2024). Jurnal Informatika dan Rekayasa Perangkat Lunak Penerapan Metode Dbscan untuk Identifikasi Kluster Gempa Bumi di Daerah Yogyakarta. *Jurnal Informatika Dan Rekayasa Perangkat Lunak*, 6(1), 40–46.
- Antika, R., Rifa, A., Dikananda, F., Indriya Efendi, D., & Narasati, R. (2023). PENERAPAN ALGORITMA DECISION TREE BERBASIS POHON KEPUTUSAN DALAM KLASIFIKASI PENYAKIT JANTUNG. In *Jurnal Mahasiswa Teknik Informatika* (Vol. 7, Issue 6).
- Budianto, R. E., Linawati, N. M., Arijana, I. G. K. N., Wahyuniari, I. A. I., & Wiryanan, I. G. N. S. (2022). Potensi Senyawa Fitokimia pada Tumbuhan dalam Menurunkan Kadar Glukosa Darah pada Diabetes Melitus. *Jurnal Sains Dan Kesehatan*, 4(5), 548–556. <https://doi.org/10.25026/jsk.v4i5.1259>
- Gunawan, A., Munir, Wibisono, Y., & Furqon, C. (2023). *Sistem Informasi Manajemen Terkini: Meningkatkan Efisiensi dengan Kecerdasan Buatan* (I). PT. Literasi Nusantara Abadi Grup. [www.penerbitlitnus.co.id](http://www.penerbitlitnus.co.id)
- Suryana, S. E., Warsito, B., & Suparti. (2021). PENERAPAN GRADIENT BOOSTING DENGAN HYPEROPT UNTUK MEMPREDIKSI KEBERHASILAN TELEMARKETING BANK. *JURNAL GAUSSIAN*, 10(4), 617–623. <https://ejournal3.undip.ac.id/index.php/gaussian/>
- Edwar, Semadi, I. G. A. N. R., Samsudin, M., & Dharmendra, I. K. (2023). Perbandingan Metode Seleksi Fitur Pada Analisis Sentimen (Studi Kasus Opini PILKADA DKI 2017). *Informatics for Educators And Professionals : Journal of Informatics*, 8(1), 11–18.
- Khoeruddin, A., Andriansyah Sudrajat, F., Purnama, G., Kuwangid, I., & Firmansyah, R. (2023). Optimasi Fitur Seleksi Random Forest Menggunakan GA Dalam Klasifikasi Data Penyakit Gagal Jantung. 1(2), 1–09. <https://doi.org/10.54066/jptis.v1i2.323>

- Mulyo, H., & Maori, N. A. (2024). *PENINGKATAN AKURASI PREDIKSI PEMILIHAN PROGRAM STUDI CALON MAHASISWA BARU MELALUI OPTIMASI ALGORITMA DECISION TREE DENGAN TEKNIK PRUNING DAN ENSEMBLE ENHACING PREDICTION ACCURACY OF NEW STUDENT PROGRAM SELECTION THROUGH DECISION TREE ALGORITHM OPTIMIZATION WITH PRUNING TECHNIQUE AND ENSEMBLE.* 15(1), 15–25. <https://doi.org/10.34001/jdpt>
- Nasrullah, A. H. (2021). *IMPLEMENTASI ALGORITMA DECISION TREE UNTUK KLASIFIKASI PRODUK LARIS.* 7(2). <http://ejournal.fikom-unasman.ac.id>
- Nurussakinah, N., & Faisal, M. (2023). Klasifikasi Penyakit Diabetes Menggunakan Algoritma Decision Tree. *Jurnal Informatika*, 10(2), 143–149. <https://doi.org/10.31294/inf.v10i2.15989>
- Pratama, A. R. I., Latipah, S. A., & Sari, B. N. (2022). *OPTIMASI KLASIFIKASI CURAH HUJAN MENGGUNAKAN SUPPORT VECTOR MACHINE (SVM) DAN RECURSIVE FEATURE ELIMINATION (RFE).*
- Rizky, P. S., Hirzi, R. H., & Hidayaturrohman, U. (2022). Perbandingan Metode LightGBM dan XGBoost dalam Menangani Data dengan Kelas Tidak Seimbang. In *J Statistika* (Vol. 15, Issue 2). [www.unipasby.ac.id](http://www.unipasby.ac.id)
- Siswa, T. A. Y., & Wibowo, R. P. (2023). Komparasi Metode Seleksi Fitur Dalam Prediksi Keterlambatan Pembayaran Biaya Kuliah. *Teknika*, 12(1), 73–82. <https://doi.org/10.34148/teknika.v12i1.601>
- Wahyuni, E. S. (2016). PENERAPAN METODE SELEKSI FITUR UNTUK MENINGKATKAN HASIL DIAGNOSIS KANKER PAYUDARA. *Jurnal SIMETRIS*, 7(1).