

ANALISIS KLASIFIKASI STATUS PINJAMAN NASABAH DENGAN METODE DECISION TREE

I Kadek Abdi Adinugraha¹, I Wayan Sudiarsa², I Nyoman Yuda Kusuma³, I Gusti Putu Agung Putra Wangaya⁴, I Komang Arisandi⁵

kadekabdi76@gmail.com¹, sudiarsa@instiki.ac.id², yudakusuma1306@gmail.com³,
putrawangaya@gmail.com⁴, mangari891@gmail.com⁵

Institut Bisnis Dan Teknologi Indonesia

ABSTRAK

Penilaian kelayakan pinjaman nasabah merupakan langkah penting dalam sektor keuangan untuk mengurangi risiko terjadinya kredit macet (Non-Performing Loan). Kesalahan dalam menentukan status pinjaman dapat menyebabkan peningkatan risiko keuangan dan kerugian bagi institusi keuangan. Oleh karena itu, pendekatan yang didasarkan pada data sangat dibutuhkan untuk mendukung pengambilan keputusan yang objektif dan tepat. Penelitian ini bertujuan untuk mengevaluasi klasifikasi status pinjaman nasabah dengan menggunakan metode Decision Tree. Dataset yang digunakan diperoleh dari platform Kaggle dan mencakup karakteristik demografi serta keuangan nasabah, seperti usia, jenis pekerjaan, tingkat pendidikan, jumlah pinjaman, durasi pinjaman, tujuan pinjaman, dan status pinjaman sebagai variabel target. Proses penelitian terdiri dari pemahaman data, pra-pemrosesan data, pembangunan model klasifikasi, dan evaluasi kinerja model. Pra-pemrosesan data mencakup pemeriksaan struktur data, penanganan nilai hilang, transformasi data kategorikal, dan pembagian data menjadi set data latih dan uji. Model Decision Tree dibangun dengan menggunakan algoritma Decision Tree Classifier dengan parameter tertentu untuk menghindari overfitting. Evaluasi model dilakukan dengan menggunakan confusion matrix, akurasi, presisi, recall, dan skor f1. Hasil penelitian menunjukkan bahwa model Decision Tree berhasil mengklasifikasikan status pinjaman nasabah dengan akurasi mencapai 83,75%. Model ini juga menghasilkan aturan keputusan yang mudah dipahami dan diinterpretasikan. Oleh karena itu, metode Decision Tree dianggap cukup efektif sebagai sistem pendukung keputusan awal dalam proses penyaluran pinjaman di lembaga keuangan.

Kata Kunci: Klasifikasi, Status Pinjaman, Decision Tree, Data Mining, Machine Learning.

ABSTRACT

Customer loan eligibility assessment is an important process in the financial sector to reduce the risk of non-performing loans (NPL). Errors in determining loan status can lead to increased financial risk and losses for financial institutions. Therefore, a data-driven approach is highly required to support objective and accurate decision-making. This study aims to evaluate the classification of customer loan status using the Decision Tree method. The dataset used in this research was obtained from the Kaggle platform and includes customers' demographic and financial characteristics, such as age, job type, education level, loan amount, loan duration, loan purpose, and loan status as the target variable. The research process consists of data understanding, data preprocessing, classification model development, and model performance evaluation. Data preprocessing includes data structure inspection, handling missing values, transformation of categorical variables, and splitting the dataset into training and testing sets. The Decision Tree model was developed using the Decision Tree Classifier algorithm with specific parameter settings to prevent overfitting. Model evaluation was conducted using a confusion matrix, accuracy, precision, recall, and f1-score. The results show that the Decision Tree model successfully classified customer loan status with an accuracy of 83.75%. In addition, the model generates interpretable decision rules that can be easily understood. Therefore, the Decision Tree method is considered effective as an initial decision-support system for loan approval processes in financial institutions.

Keywords: *Klasifikasi, Status Pinjaman, Decision Tree, Data Mining, Machine Learning.*

PENDAHULUAN

Perkembangan teknologi informasi dan penerapan data mining telah memberikan dampak signifikan dalam mendukung pengambilan keputusan di berbagai bidang, khususnya pada sektor keuangan. Data mining merupakan proses untuk mengekstraksi pola dan informasi yang bermakna dari kumpulan data dalam jumlah besar guna mendukung pengambilan keputusan (Han, Kamber, dan Pei 2012). Salah satu permasalahan utama yang dihadapi oleh lembaga keuangan adalah risiko kredit bermasalah yang timbul akibat kesalahan dalam menilai kelayakan pinjaman nasabah. Risiko kredit merupakan salah satu tantangan utama dalam sektor keuangan yang memerlukan pendekatan analitis untuk meminimalkan potensi kerugian (Shmueli et al. 2020). Proses evaluasi pinjaman yang kurang akurat dapat menyebabkan peningkatan kredit macet dan kerugian finansial.

Klasifikasi status pinjaman nasabah menjadi salah satu solusi untuk mengatasi permasalahan tersebut. Dengan memanfaatkan data historis nasabah, data historis nasabah dapat dimanfaatkan untuk membangun model prediktif yang mampu mengidentifikasi pola perilaku dan risiko di masa mendatang (James et al. 2013). Teknik machine learning dapat digunakan untuk memprediksi status pinjaman secara lebih objektif dan sistematis. Dalam data mining, teknik klasifikasi digunakan untuk memetakan data ke dalam kelas tertentu berdasarkan karakteristik yang dimiliki sehingga memudahkan proses analisis dan pengambilan keputusan (Han, Kamber, dan Pei 2012).

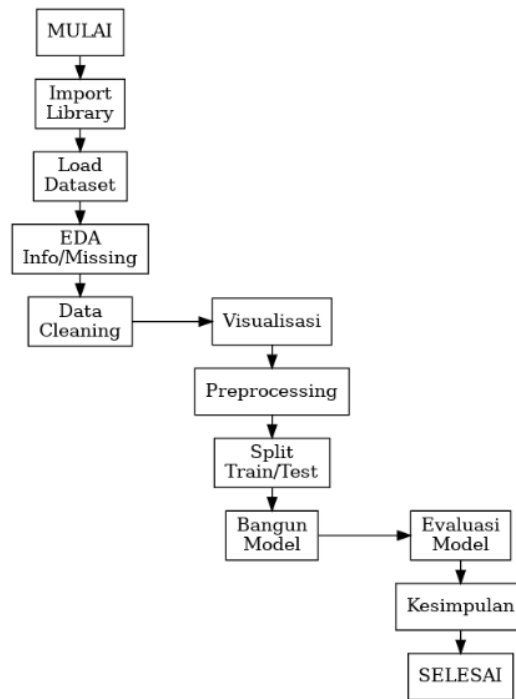
Salah satu algoritma klasifikasi yang banyak digunakan adalah Decision Tree. Decision Tree merupakan salah satu metode klasifikasi yang banyak digunakan karena kemampuannya menangani data numerik dan kategorikal serta menghasilkan aturan keputusan yang mudah dipahami (Han, Kamber, dan Pei 2012). Metode ini memiliki keunggulan dalam menghasilkan aturan keputusan yang mudah dipahami dan diinterpretasikan. Selain itu, Decision Tree mampu menangani data numerik maupun kategorikal dengan baik serta tidak memerlukan asumsi statistik yang kompleks.

Berdasarkan latar belakang tersebut, penelitian ini dilakukan untuk menganalisis klasifikasi status pinjaman nasabah menggunakan metode Decision Tree. Penelitian ini diharapkan dapat memberikan gambaran mengenai kinerja metode Decision Tree dalam mengklasifikasikan status pinjaman serta membantu meningkatkan efektivitas pengambilan keputusan pada lembaga keuangan.

METODOLOGI

Penelitian ini menggunakan metode deskriptif kuantitatif dengan pendekatan data mining untuk menganalisis dan mengklasifikasikan status pinjaman nasabah. Data mining adalah proses analitis yang dirancang untuk mengeksplorasi sejumlah besar data untuk menemukan pola yang berharga, konsisten, dan tersembunyi (Nugroho dkk., 2022). Metode ini dipilih karena mampu mengolah data numerik dan kategorikal secara sistematis serta menghasilkan model prediktif yang dapat digunakan sebagai sistem pendukung pengambilan keputusan. Proses penelitian dilakukan melalui beberapa tahapan utama, yaitu pemahaman data, pra-pemrosesan data, pembangunan model klasifikasi, dan evaluasi kinerja model. Menggunakan Decision Tree sebagai alat klasifikasi. Decision Tree merupakan salah satu metode klasifikasi yang banyak digunakan karena mudah untuk diinterpretasikan (Surahmat & Sutrisno, t.t.).

Flowchart Metode Penelitian



Gambar diatas Adalah flowchart dari tahapan Metode Penelitian yang kami lakukan.

- Import Library
Mengambil semua Pustaka dari python yang dibutuhkan dalam project ini.
- Load Dataset
Membaca file Dataset dari sumber. (dapat berbentuk .CSV, .Excel)
- EDA (Explore Data Analysis)
Dalam tahap ini kami dapat melihat bentuk data, menganalisis tipe data, dan mengetahui jumlah nilai data.
- Data Cleaning
Membersihkan data yang hilang, agar data ini siap untuk digunakan.
- Visualisasi Data
Fase pembuatan grafik untuk memahami pola.
- Pre-Processing
Menyiapkan data untuk model seperti, encoding data menjadi angka, dan normalisasi atau standarisasi.
- Split Train
Pemisahan dataset menjadi dua.
- Bangun Model
Menggunakan algoritma seperti descision tree untuk mempelajari pola data.
- Evaluasi Model
Mengukur performa model menggunakan accuracy, precision, recall.

HASIL DAN PEMBAHASAN

Tahap pra-pemrosesan data merupakan tahapan awal yang sangat penting dalam penelitian ini karena kualitas data yang digunakan memengaruhi hasil penelitian. Dataset yang digunakan dalam penelitian ini diperoleh dari platform Kaggle dan berisi data karakteristik nasabah yang berkaitan dengan status pinjaman. Sebelum dilakukan

pemodelan menggunakan metode Decision Tree, data terlebih dahulu dianalisis untuk memastikan kelengkapan dan konsistensinya.

Tabel 1 Ringkasan Atribut

NO	Atribut	Tipe Data	Nilai Sample	Keterangan
1	Age	Numerik	35	Usia Nasabah
2	Job	Kategorikal	admin	Jenis Pekerjaan
3	Marital	Kategorikal	married	Status Pernikahan
4	Education	Kategorikal	Secondary	Tingkat Pendidikan
5	Credit Amount	Numerik	1500	Jumlah Pinjaman
6	Housing	Kategorikal	yes	Kepemilikan rumah
7	Duration	Numerik	24	Lama Pinjaman
8	Purpose	Kategorikal	car	Tujuan Pinjaman
9	Status Pinjaman	Kategorikal	lancar	Label Target

Tabel 1 Menampilkan ringkasan tentang atribut yang digunakan dalam penelitian ini. Informasi yang diberikan mencakup nama atribut, jenis data, serta nilai sample sebagai ilustrasi karakteristik data tersebut. Tujuan dari penyajian ini adalah memberikan gambaran awal mengenai struktur dataset tanpa menampilkan data mentah secara berlebihan.

Dilakukan pemeriksaan terhadap struktur dataset untuk mengidentifikasi jumlah atribut, tipe data, serta kemungkinan adanya nilai kosong atau data yang tidak relevan. Berdasarkan hasil pemeriksaan, dataset terdiri dari beberapa atribut numerik dan kategorikal yang merepresentasikan karakteristik nasabah, seperti usia, jenis pekerjaan, tingkat pendidikan, jumlah pinjaman, serta tujuan pinjaman. Variabel status pinjaman digunakan sebagai variabel target dalam proses klasifikasi.

Setelah proses pemeriksaan selesai, dataset dibagi menjadi dua bagian, yaitu data latih dan data uji. Data latih digunakan untuk membangun model Decision Tree, sedangkan data uji digunakan untuk mengevaluasi kinerja model yang dihasilkan. Pembagian data dilakukan dengan proporsi tertentu agar model dapat belajar dari sebagian besar data dan tetap diuji pada data yang belum pernah dilihat sebelumnya. Pembagian ini bertujuan untuk menghindari bias dan memberikan gambaran yang lebih objektif terhadap kemampuan generalisasi model.

Setelah melalui seluruh langkah pra-pemrosesan, dataset yang digunakan dalam penelitian ini disimpulkan sudah siap digunakan untuk membangun model klasifikasi dengan metode Decision Tree. Hasil dari proses pra-pemrosesan ini menjadi dasar yang sangat penting dalam menghasilkan model yang akurat dan dapat dipercaya di tahap berikutnya.

Hasil Pembangunan Model Decision Tree

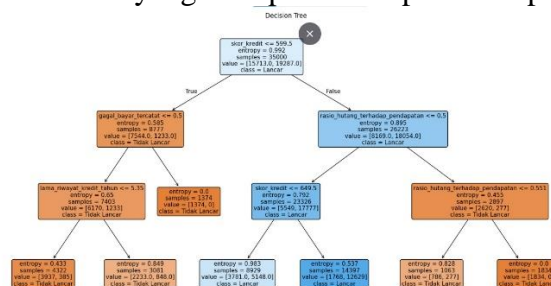
Tahap pembangunan model adalah tahap berikutnya setelah proses pra-pemrosesan data selesai. Pada tahap ini, dataset yang sudah dibersihkan dan diubah bentuknya digunakan untuk membuat model klasifikasi status pinjaman nasabah dengan metode Decision Tree. Proses pembuatan model dilakukan dengan menggunakan data latih yang telah ditentukan sebelumnya.

```
dt = DecisionTreeClassifier(  
    criterion='entropy',  
    max_depth=5,  
    random_state=42  
)  
  
dt.fit(X_train, y_train)
```

Gambar 1 Source code untuk pengembangan model Decision Tree

Model Decision Tree dibuat menggunakan algoritma Decision Tree Classifier dengan parameter yang disesuaikan sesuai kebutuhan penelitian. Parameter tersebut dipilih agar model dapat belajar pola hubungan antara atribut nasabah dan status pinjaman secara efektif, sekaligus mencegah terjadinya overfitting. Proses pelatihan model dilakukan dengan membangun struktur pohon keputusan berdasarkan atribut yang memiliki pengaruh terbesar terhadap variabel target.

Model yang sudah dibuat kemudian dianalisis untuk mengetahui struktur dan tingkat kompleksitas pohon keputusan yang dihasilkan. Struktur ini menunjukkan atribut-atribut terpenting dalam proses klasifikasi. Selain itu, model Decision Tree memiliki keunggulan dalam hal kemudahan pemahaman, karena setiap keputusan yang dihasilkan bisa dijelaskan melalui aturan-aturan yang terdapat dalam pohon Keputusan (Decision Tree).



Gambar 2 Struktur Model Decision Tree

```
[22] ✓ 0d from sklearn.tree import DecisionTreeClassifier, plot_tree
import matplotlib.pyplot as plt

# Model dengan kedalaman terbatas
dt_simple = DecisionTreeClassifier(
    criterion="entropy",
    max_depth=3, # <-- POINT PENTING (atur 2-4)
    min_samples_leaf=100, # agar tidak terlalu detail
    random_state=42
)

dt_simple.fit(X_train, y_train)

# Visualisasi
plt.figure(figsize=(20, 10))
plot_tree(
    dt_simple,
    feature_names=X.columns,
    class_names=["Tidak Lancar", "Lancar"],
    filled=True,
    rounded=True,
    fontsize=11
)
plt.title("Decision Tree")
plt.show()
```

Gambar 3 Source code untuk model Decision Tree

Berdasarkan hasil pembangunan model, pohon keputusan yang dihasilkan mampu memahami pola dari data latih dengan baik dan siap digunakan pada tahap evaluasi. Selanjutnya, model ini diuji menggunakan data uji untuk mengetahui sejauh mana kemampuannya dalam mengklasifikasikan status pinjaman nasabah yang belum pernah dilihat sebelumnya.

Hasil Evaluasi Model

Evaluasi kinerja model dilakukan menggunakan data uji yang sebelumnya tidak digunakan dalam proses pelatihan. Metode evaluasi yang digunakan meliputi confusion matrix, accuracy, precision, recall, dan f1score. Berdasarkan hasil evaluasi, model Decision Tree menunjukkan performa yang baik dalam mengklasifikasikan status pinjaman nasabah. Nilai akurasi yang diperoleh menunjukkan bahwa sebagian besar data uji dapat diklasifikasikan dengan benar. Confusion matrix memperlihatkan bahwa kesalahan klasifikasi relatif kecil dibandingkan dengan jumlah prediksi yang tepat. Model yang dikembangkan berhasil mencapai tingkat akurasi sebesar 83,75%. Angka ini menunjukkan bahwa model memiliki kemampuan generalisasi yang baik dalam memetakan karakteristik demografis dan finansial nasabah (seperti pendapatan tahunan,

skor kredit, dan lama bekerja) terhadap status kelancaran pinjaman mereka. Dalam konteks klasifikasi risiko kredit, akurasi di atas 80% mengindikasikan bahwa metode Decision Tree cukup robust untuk digunakan sebagai alat bantu seleksi awal (pre-screening tool), menggantikan atau mendampingi penilaian manual yang sering kali subjektif.

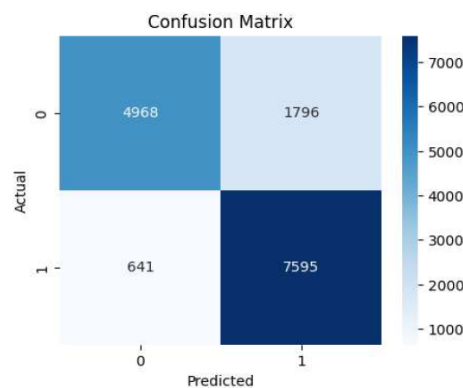
Pembahasan

Evaluasi kinerja model Decision Tree dilakukan menggunakan data pengujian (testing set) sebanyak 15.000 data, yang merepresentasikan 30% dari total dataset. Pengukuran performa dilakukan menggunakan metrik Classification Report untuk melihat nilai statistik, serta Confusion Matrix untuk membedah distribusi prediksi benar dan salah secara lebih rinci. Ringkasan performa model dapat dilihat pada Tabel 1 berikut.

Tabel 1. Classification Report (Evaluasi Performa Model)

Status Pinjaman (Kelas)	Precision	Recall	F1-Score	Support
0 (Tidak Lancar)	0.89	0.73	0.80	6.764
1 (Lancar)	0.81	0.92	0.86	8.236
Akurasi (Accuracy)			0.84	15.000

Berdasarkan Tabel 1, model menghasilkan tingkat akurasi global sebesar 84%. Namun, nilai akurasi saja tidak cukup untuk menggambarkan kualitas model, terutama pada kasus risiko kredit. Oleh karena itu, untuk memvalidasi asal-usul nilai Precision dan F1-Score tersebut, serta memahami letak kesalahan model, disajikan Confusion Matrix pada Gambar 4 di bawah ini.



Gambar 4. Visualisasi Confusion Matrix Hasil Prediksi

Berdasarkan visualisasi pada Gambar 4, berikut adalah analisis mendalam mengenai korelasi antara Confusion Matrix dengan skor performa yang dihasilkan:

- Analisis Precision dan Tingkat Kesalahan (False Positive) Nilai Precision untuk Kelas 1 (Lancar) tercatat sebesar 0.81. Angka ini dipengaruhi oleh adanya 1.796 data False Positive (prediksi Lancar, aktual Macet). Dalam konteks perbankan, ini merupakan critical error karena model merekomendasikan peminjaman kepada nasabah yang berisiko gagal bayar. Tingginya angka False Positive inilah yang menahan skor Precision Kelas 1 tidak setinggi Kelas 0.
- Analisis Recall dan Sensitivitas Model Sebaliknya, model menunjukkan sensitivitas yang sangat tinggi dalam mendeteksi nasabah Lancar, dibuktikan dengan nilai Recall sebesar 0.92. Hal ini terlihat pada Confusion Matrix di mana jumlah True Positive (7.595) jauh mendominasi dibandingkan False Negative (641). Artinya, model sangat jarang melewati nasabah yang potensial (Lancar).
- Analisis F1-Score (Keseimbangan Model) Nilai F1-Score rata-rata (0.84) menunjukkan bahwa model memiliki keseimbangan yang baik antara ketepatan (precision) dan kelengkapan (recall). Meskipun terdapat ketidakseimbangan pada deteksi kelas minoritas (Macet), performa model secara keseluruhan masih berada dalam batas

toleransi yang baik untuk digunakan sebagai sistem pendukung keputusan awal.

Validitas Metodologis

Model klasifikasi ini berhasil karena proses pra-pemrosesan data yang dilakukan dengan cara yang teratur. Khususnya, dalam mengatasi nilai yang hilang pada atribut usia dan status pekerjaan, digunakan metode imputasi modus seperti yang dijelaskan di Step 4 dalam metode penelitian. Selain itu, untuk menjadikan model lebih mudah dipahami dan menghindari masalah overfitting, struktur pohon keputusan dibatasi dengan menetapkan parameter (`max_depth = 3`). Pembatasan ini memungkinkan model tetap mengidentifikasi pola-pola utama dalam data tanpa membuat struktur pohon yang terlalu rumit. Dengan demikian, model tidak hanya mampu belajar dari data latih secara baik, tetapi juga memiliki kemampuan untuk memprediksi data uji yang belum pernah dilihat sebelumnya dengan baik.

KESIMPULAN

Berdasarkan temuan yang diperoleh dari penelitian yang telah dilakukan, dapat disimpulkan bahwa pendekatan Decision Tree dapat digunakan untuk mengklasifikasikan status pinjaman nasabah secara efisien dengan memanfaatkan informasi tentang karakteristik demografis dan finansial para nasabah. Tahapan pra-pemrosesan data, yang mencakup pembersihan, penanganan data yang hilang, serta transformasi data kategorikal, terbukti sangat berpengaruh dalam meningkatkan kualitas dataset dan performa model yang dihasilkan.

```
accuracy = accuracy_score(y_test, y_pred)
print("Akurasi Model:", accuracy)
```

Akurasi Model: 0.8375333333333334

Gambar 5. Hasil akurasi model

Hasil analisis terhadap model menunjukkan bahwa Decision Tree memiliki kinerja yang cukup baik dengan akurasi mencapai 83,75%. Penilaian menggunakan confusion matrix dan laporan klasifikasi memperlihatkan bahwa model tersebut mampu mengklasifikasikan mayoritas data uji dengan akurat, terutama dalam mendeteksi nasabah yang memiliki status pinjaman lancar. Walaupun masih terdapat sejumlah kesalahan dalam klasifikasi data, secara keseluruhan, model ini memiliki kemampuan generalisasi yang baik dan dapat memberikan gambaran risiko kredit dengan cara yang objektif.

Di samping itu, keunggulan utama dari metode Decision Tree adalah kemampuannya untuk menghasilkan aturan keputusan yang mudah dipahami, sehingga hal ini mempermudah lembaga keuangan dalam menginterpretasikan hasil klasifikasi. Oleh karena itu, metode Decision Tree dapat berfungsi sebagai alat bantu dalam pengambilan keputusan awal (pre-screening) selama proses pemberian pinjaman, guna membantu mengurangi risiko kredit yang bermasalah dan meningkatkan efektivitas manajemen risiko di lembaga keuangan.

Saran

Berdasarkan penelitian yang telah dilakukan, disarankan untuk melakukan pengembangan lebih lanjut dengan menggunakan dataset yang lebih besar dan beragam agar model dapat bekerja lebih optimal dan generalisasi hasil menjadi lebih baik. Selain itu, penelitian selanjutnya dapat membandingkan metode Decision Tree dengan algoritma klasifikasi lain seperti Random Forest, K-Nearest Neighbor, atau Support Vector Machine guna memperoleh performa yang lebih baik. Penggunaan teknik optimasi parameter dan feature selection juga disarankan untuk meningkatkan akurasi model dalam mengklasifikasikan status pinjaman nasabah.

DAFTAR PUSTAKA

- Han, Jiawei, Micheline Kamber, and Jian Pei. *Data Mining: Concepts and Techniques*. 3rd ed. Burlington: Morgan Kaufmann, 2012.
- James, Gareth, Daniela Witten, Trevor Hastie, and Robert Tibshirani. *An Introduction to Statistical Learning: With Applications in R*. New York: Springer, 2013.
- Knafllic, *Storytelling with Data: A Data Visualization Guide for Business Professionals*. Hoboken: John Wiley & Sons, 2015.
- McKinney, Wes. *Python for Data Analysis: Data Wrangling with Pandas, NumPy, and IPython*. 2nd ed. Sebastopol: O'Reilly Media, 2017.
- Provost, Foster, and Tom Fawcett. *Data Science for Business: What You Need to Know about Data Mining and Data-Analytic Thinking*. Sebastopol: O'Reilly Media, 2013.
- Sari, N. P., & Nugroho, D. A. (2022). Analisis Metode Decision Tree untuk Klasifikasi Data Mahasiswa Baru. *Jurnal Teknologi Informasi dan Sains Terapan (JUST-IT)*, 4(1), 15–25.
- Sari, N. P., & Nugroho, D. A. (2022). Penerapan Algoritma Klasifikasi dalam Data Mining untuk Prediksi Kelulusan Mahasiswa. *Jurnal Teknologi Informasi dan Sains Terapan (JUST-IT)*, 4(1), 26–35.
- Shmueli, Galit, Peter C. Bruce, Peter Gedeck, and Nitin R. Patel. *Data Mining for Business Analytics: Concepts, Techniques, and Applications*. Hoboken: John Wiley & Sons, 2020.
- Tukey, John W. *Exploratory Data Analysis*. Reading, MA: Addison-Wesley, 1977.
- Wickham, Hadley, and Garrett Grolemund. *R for Data Science: Import, Tidy, Transform, Visualize, and Model Data*. Sebastopol: O'Reilly Media, 2017.
- Zaki, Mohammed J., and Wagner Meira Jr. *Data Mining and Analysis: Fundamental Concepts and Algorithms*. Cambridge: Cambridge University Press, 2014.
- Wickham, Hadley. *ggplot2: Elegant Graphics for Data Analysis*. New York: Springer, 2016.