

PENERAPAN DATA MINING ALGORITMA DECISION TREE UNTUK MEMPREDIKSI KETERLAMBATAN PEMBAYARAN PIUTANG USAHA PADA PT XYZ

Muhammad Tantowi Jauhari^{1*}, Windarsyah², Finki Dona Marleny³

^{1,2,3}Program Studi Informatika, Fakultas Teknik, Universitas Muhammadiyah Banjarmasin; Jl. Gubernur Syarkawi, Semangat Dalam, Kec. Alalak, Kab. Barito Kuala, Kalimantan Selatan 70581, Telp. +62 821-9675-9664

Keywords:

Keterlambatan Pembayaran;
Prediksi;
Piutang Usaha;
Decision Tree.

Correspondent Email:

m_tantowi_jauhari@umbjm.a
c.id



JITET is licensed under
a Creative Commons
Attribution-NonCommercial
4.0 International License.

Abstrak. Tingginya frekuensi dan volume keterlambatan pembayaran piutang usaha menjadi permasalahan utama yang dihadapi PT XYZ, sebuah agen BBM industri, yang secara langsung mengganggu arus kas perusahaan. Proses pemantauan piutang yang masih berjalan manual dinilai tidak efisien dan rentan terhadap kesalahan. Penelitian ini bertujuan membangun model prediktif untuk status pembayaran bulan berikutnya dengan menerapkan metode *Knowledge Discovery in Databases* (KDD) dan algoritma *Decision Tree*. Data yang diolah merupakan data internal perusahaan dari Januari 2021 hingga Juni 2022, yang mencakup laporan penjualan dan piutang beredar. Hasil penelitian menunjukkan bahwa model klasifikasi *Decision Tree* yang dibangun berhasil mencapai akurasi keseluruhan sebesar 92.70%. Analisis atribut mengidentifikasi "Status Pembayaran" bulan berjalan sebagai faktor paling dominan dalam memprediksi keterlambatan. Model ini dapat menjadi landasan bagi perusahaan untuk beralih ke manajemen piutang yang lebih proaktif dan berbasis data untuk memitigasi risiko keterlambatan pembayaran secara lebih efektif.

Abstract. The high frequency and volume of late payments on accounts receivable is the main problem faced by PT XYZ, an industrial fuel agent, which directly disrupts the company's cash flow. The existing manual accounts receivable monitoring process is considered inefficient and prone to errors. This study aims to build a predictive model for the following month's payment status by applying the *Knowledge Discovery in Databases* (KDD) method and the *Decision Tree* algorithm. The processed data is the company's internal data from January 2021 to June 2022, which includes sales and outstanding receivables reports. The results show that the developed *Decision Tree* classification model successfully achieved an overall accuracy of 92.70%. Feature analysis identified the current month's 'Payment Status' as the most dominant factor in predicting delinquency. The resulting model can serve as a foundation for the company to transition towards more proactive, data-driven accounts receivable management to more effectively mitigate the risk of late payments.

1. PENDAHULUAN

Manajemen arus kas yang efektif merupakan syarat yang harus dilakukan bagi kelangsungan dan pertumbuhan perusahaan [1]. Salah satu komponen krusial yang secara langsung memengaruhinya adalah piutang usaha, yaitu tagihan kepada pelanggan atas

barang atau jasa yang telah diberikan [2]. Meskipun penjualan kredit menjadi strategi umum untuk meningkatkan volume penjualan, praktik ini membawa risiko keterlambatan pembayaran atau bahkan gagal bayar (*bad debt*) [3], yang dapat mengganggu likuiditas dan

kemampuan perusahaan memenuhi kewajiban finansialnya [4].

Kondisi ini menjadi tantangan utama bagi PT XYZ, sebuah perusahaan agen bahan bakar minyak (BBM) solar industri di Kalimantan Selatan. Model bisnis menuntut perusahaan untuk membeli BBM secara tunai dari pemasok utama, PT. Pertamina Patra Niaga. Namun, untuk beberapa pertimbangan, seperti menjaga daya saing perusahaan, mempertahankan volume penjualan, menjaga loyalitas dan mendukung kelancaran proyek-proyek yang dikerjakan oleh pelanggan, PT XYZ menjual produknya dengan sistem pembayaran kredit berjangka waktu 30 hingga 60 hari (Net 30 – Net 60). Kesenjangan antara arus kas keluar yang bersifat tunai dan arus kas masuk yang bersifat kredit menempatkan perusahaan pada posisi yang sangat rentan.

Permasalahan utama yang melatarbelakangi penelitian ini adalah tingginya frekuensi dan volume keterlambatan pembayaran yang dialami oleh PT XYZ. Pengamatan terhadap proses internal mengungkap dua faktor penyebab utama. Pertama, perusahaan belum memiliki sistem pemantauan umur piutang (*aging schedule*) yang otomatis dan rinci per faktur. Kedua, proses penagihan yang masih dilakukan dengan penelusuran faktur secara manual menggunakan Microsoft Excel. Ketergantungan pada metode manual ini tidak hanya memakan waktu dan sumber daya, tetapi juga rentan terhadap kesalahan manusia dan kurangnya visibilitas secara keseluruhan [5].

Menghadapi tantangan tersebut, penelitian ini mengusulkan pemanfaatan kemajuan teknologi informasi, khususnya teknik data mining, sebagai solusi. Data mining adalah proses penggalan pola untuk menemukan informasi berharga dari kumpulan data besar yang dapat digunakan untuk pengambilan keputusan bisnis [6].

Secara spesifik, algoritma *decision tree* dipilih sebagai teknik utama untuk melakukan data mining dalam penelitian ini. Pendekatan *decision tree* populer karena kemampuannya memodelkan hubungan yang kompleks [7], dapat menangani baik atribut data kategorikal maupun numerik serta nilai yang hilang [8], dan yang terpenting, mampu menghasilkan aturan “*IF-THEN*” yang relatif mudah diinterpretasikan oleh manusia [9].

Tujuan utama dari penelitian ini adalah untuk mengidentifikasi faktor-faktor paling berpengaruh terhadap potensi keterlambatan pembayaran dan membangun model prediktif untuk status pembayaran bulan selanjutnya. Harapannya, model ini dapat menjadi landasan bagi PT XYZ untuk bertransformasi dari pendekatan manajemen piutang yang reaktif menjadi lebih proaktif dan berbasis data, sehingga dapat memitigasi risiko arus kas secara lebih efektif.

Untuk menjaga etika penelitian, nama perusahaan dan pelanggan dalam penelitian ini penulis samarkan. Hal ini dilakukan untuk melindungi kerahasiaan informasi yang bersifat sensitif serta menghormati privasi entitas yang terlibat. Penyebutan nama secara eksplisit dapat menimbulkan risiko terhadap reputasi perusahaan maupun pelanggan, terutama apabila temuan penelitian berkaitan dengan kinerja pembayaran atau potensi risiko kredit.

2. TINJAUAN PUSTAKA

Penelitian yang dilakukan oleh Salsabila, Missi, dan Ruuhwan (2024) bertujuan untuk memprediksi kelulusan mahasiswa tepat waktu di Universitas Perjuangan Tasikmalaya. Dalam penelitian ini, metode data mining dengan algoritma C4.5 digunakan dengan bantuan bahasa pemrograman Python untuk menganalisis data wisudawan sebanyak 1,932 sampel. Hasil penelitian menunjukkan bahwa algoritma C4.5 memiliki akurasi yang tinggi, yaitu mencapai 92,233% saat menggunakan seluruh dataset sebagai data latih dan 90,24% dengan menggunakan *cross validation* 80% [10].

Mubarokah, Nugraha, & Yunus (2024) menerapkan algoritma C4.5 yang merupakan bagian dari *decision tree* guna membentuk pohon keputusan dan mengklasifikasikan keterlambatan pembayaran SPP di MTs Nurul Huda. Metode yang digunakan adalah eksperimen dan pengujian dengan data peserta didik. Hasil penelitian menunjukkan bahwa akurasi tertinggi sebesar 90% dicapai dengan pembagian *data training* 80% dan *data testing* 20%. Dengan hasil tersebut, disimpulkan bahwa algoritma C4.5 cocok untuk mengklasifikasikan data pembayaran SPP [11].

Penelitian Anief & Rufiyanto (2019) menggunakan data mining dengan algoritma *Decision Tree* C4.5 untuk memprediksi

kelulusan mahasiswa di Universitas Pandanaran, yang memiliki mahasiswa kelas reguler dan karyawan. Penelitian ini mengolah data mahasiswa dengan parameter seperti jurusan, usia, jenis kelamin, status pekerjaan, dan IPK semester 1-4. Hasilnya berupa 10 aturan kelulusan dengan akurasi 65,98% dan nilai AUC 0,874, yang dikategorikan sebagai klasifikasi data yang baik [12].

Serupa dengan penelitian sebelumnya, Gaol (2020) melakukan penelitian untuk memprediksi mahasiswa yang berpotensi non-aktif di STMIK Triguna Dharma Medan menggunakan algoritma C4.5. Penelitian ini bertujuan untuk mengurangi angka drop out dengan mengidentifikasi faktor-faktor seperti jadwal kuliah, nilai absensi, nilai gagal, dan status pembayaran uang kuliah. Hasil penelitian ini adalah rule prediksi dengan tingkat keakuratan yang dinilai sangat baik [13].

Batubara, Windarto, & Irawan (2022) menganalisis prediksi keterlambatan pembayaran listrik, sebuah masalah yang menyebabkan kerugian bagi PT. PLN dan loket pembayaran seperti PT Jaya Nuhgra Pratama. Penelitian ini membandingkan metode klasifikasi *Decision Tree* dan *Support Vector Machine* (SVM) menggunakan data kuisioner dengan atribut seperti penghasilan, jumlah tanggungan, dan tanggal penerimaan gaji. Dengan menggunakan perangkat lunak WEKA, ditemukan bahwa algoritma C4.5 lebih tepat dengan akurasi 70% dibandingkan SVM (67%), dan mengidentifikasi tanggal penerimaan gaji sebagai faktor utama penyebab keterlambatan [14].

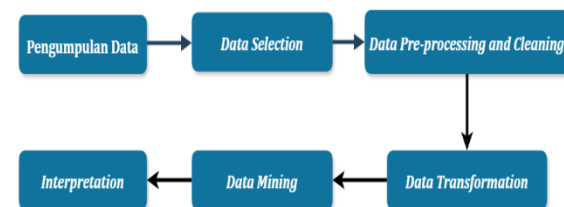
Dalam penelitian ini, model prediksi dibangun menggunakan algoritma *decision tree* yang diimplementasikan melalui kelas *DecisionTreeClassifier* dari library *scikit-learn* di Python. Parameter *splitting criterion* berbasis *entropy* dipilih agar proses pemisahan simpul (*node*) pada pohon keputusan didasarkan pada metrik *Information Gain*. Pendekatan ini secara konsep sejalan dengan prinsip yang digunakan oleh algoritma seperti C4.5 [15].

Berbeda dengan penelitian-penelitian sebelumnya yang mengaplikasikan algoritma *decision tree* dan turunannya seperti algoritma C4.5 di berbagai sektor, penelitian ini menawarkan pendekatan yang lebih spesifik pada konteks perusahaan agen BBM solar industri dengan model bisnis yang melibatkan

pembelian secara tunai dan penjualan secara kredit. Kebaharuan penelitian ini terletak pada penerapan algoritma *decision tree* berbasis *entropy* dalam kerangka kerja *Knowledge Discovery in Databases* (KDD) menggunakan dataset yang dihasilkan dari transformasi data internal yang mencakup laporan penjualan dan laporan piutang beredar (*outstanding*) dengan atribut-atribut operasional perusahaan seperti jumlah total kiloliter, jumlah total tunggakan, kategori masa tunggakan, dan status pembayaran bulan lalu untuk memprediksi status pembayaran bulan berikutnya.

3. METODE PENELITIAN

Knowledge Discovery in Databases (KDD) adalah metode yang digunakan untuk menggali pengetahuan dari data yang tersedia untuk mendukung pengambilan keputusan [16]. Proses KDD dalam penelitian ini melibatkan beberapa langkah, mulai dari pengumpulan data yang kemudian dilanjutkan ke *data selection*, dilanjutkan dengan *data pre-processing and cleaning*, *data transformation*, *data mining*, hingga akhirnya *interpretation*.



Gambar 1. Alur KDD

3.1. Pengumpulan Data

Pada tahap ini dilakukan proses pengumpulan data yang menjadi dasar untuk melakukan penelitian. Data yang digunakan yaitu data internal perusahaan dari PT XYZ yang terdiri dari laporan penjualan perusahaan dan laporan piutang beredar (*outstanding*) yang dirancang tiap bulannya dalam rentang waktu Januari 2021 hingga Juni 2022. Data ini diperoleh secara langsung dari divisi keuangan PT XYZ dalam format berkas digital *Microsoft Excel*.

Tabel 1. Data Penjualan Bulan Januari

Tanggal	Nama Pelanggan	Jumlah Kiloliter	INV. No
01-Jan-21	C001	20	0001

01-Jan-21	C002	50	0002
01-Jan-21	C003	5	0003
01-Jan-21	C004	5	0004
...	...	...	...
31-Jan-21	C251	10	0251

Tabel 2. Laporan Outstanding Bulan Januari

Pelanggan	Total Tagihan	Keterangan
C010	Rp5,000,000	Belum Lunas
C014	Rp13,000,000	Belum Lunas
C017	Rp 0	Lunas
...	...	...
C030	Rp7,800,000,000	Piutang tak tertagih, pengadilan

3.2. Data Selection

Tahap seleksi data bertujuan untuk mengidentifikasi dan memilih atribut-atribut yang paling relevan dari kumpulan data yang telah diperoleh untuk membangun model.

3.3. Data Preprocessing and Cleaning

Pre-processing data berfokus pada peningkatan kualitas dataset untuk memastikan akurasi dalam proses pelatihan model. Dalam tahap ini dilakukan penanganan nilai yang hilang (*missing values*) yang dilakukan untuk menjaga integritas data dan akurasi model, setiap baris data yang mengandung nilai kosong pada atribut kunci dihapus dari dataset. Pendekatan ini dipilih untuk mencegah *bias* atau kesalahan interpretasi oleh model selama proses *training*.

3.4. Data Transformation

Dalam *data transformation* melibatkan beberapa tahapan yang harus dilakukan untuk merancang sebuah dataset agar siap digunakan dalam proses *data mining*. Dalam konteks penelitian ini, yang pertama kali dilakukan yaitu melakukan agregasi terhadap data penjualan bulanan untuk mengakumulasi total volume pembelian per pelanggan. Setelah proses agregasi data penjualan selesai, dataset tersebut kemudian digabung dengan data piutang beredar. Langkah selanjutnya adalah rekayasa fitur (*feature engineering*) di mana

beberapa atribut baru dibentuk. Atribut-atribut baru ini mencakup status pembayaran bulan sebelumnya, kategori masa tunggakan, dan status pembayaran bulan depan yang ditetapkan sebagai variabel target prediksi.

3.5. Data Mining

Pada tahap *data mining*, dilakukan proses inti untuk menemukan pola dan pengetahuan tersembunyi dari dataset yang telah dipersiapkan. Penelitian ini akan menggunakan kelas *DecisionTreeClassifier* dari *library scikit-learn* dengan kriteria pemisahan (*splitting criterion*) berbasis *entropy* untuk membangun model klasifikasi.

3.6. Interpretation

Tahap interpretasi berfokus pada evaluasi dan pemahaman terhadap model prediktif yang telah dibangun. Pola-pola keputusan yang dihasilkan oleh model disajikan dalam bentuk yang dapat diinterpretasi, yaitu melalui visualisasi struktur *decision tree* dan aturan atau *rules "IF-THEN"*. Selain itu, dilakukan analisis terhadap kinerja model seperti pengukuran akurasi untuk menilai keandalan model prediksi dalam mengklasifikasikan status pembayaran pelanggan di bulan depan.

4. HASIL DAN PEMBAHASAN

Bagian ini menyajikan hasil dari proses pengolahan data, pemodelan, hingga evaluasi model klasifikasi status pembayaran pelanggan.

4.1. Data Selection

Dalam penelitian ini, atribut yang dinilai relevan dari kedua macam data yang telah diperoleh meliputi nama pelanggan dan jumlah kiloliter yang bersumber dari data penjualan, serta nama pelanggan, total tagihan, dan status pembayaran dari laporan piutang beredar.

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 251 entries, 0 to 250
Data columns (total 4 columns):
#   Column              Non-Null Count  Dtype
---  -
0   No                   251 non-null    int64
1   Tanggal              251 non-null    object
2   Nama Pelanggan       251 non-null    object
3   Jumlah Kiloliter     251 non-null    int64
dtypes: int64(2), object(2)
memory usage: 8.0+ KB
```

Gambar 2. Hasil Selection Data Penjualan

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 211 entries, 0 to 210
Data columns (total 3 columns):
#   Column              Non-Null Count  Dtype
---  -
0   Nama Pelanggan      211 non-null   object
1   Total Tunggalan     211 non-null   float64
2   Status Pembayaran   211 non-null   object
dtypes: float64(1), object(2)
memory usage: 5.1+ KB
```

Gambar 3. Hasil Selection Piutang Beredar

4.2. Data Preprocessing and Cleaning

Tahap awal melibatkan pembersihan data dengan menghapus kolom-kolom kosong menggunakan metode *dropna* dari library Pandas dari Python. Setelah dilakukan pembersihan dari nilai kosong, dilakukan pengecekan dengan fungsi *isnull().sum()*.

```
import pandas as pd
df_penjualan = pd.read_csv('Final/Januari 2021 - Standardized.csv', sep=',')
df_penjualan.isnull().sum()

✓ 0.0s
No      0
Date    0
Customer Name  0
Kiloliter  0
dtype: int64

df_outstanding = pd.read_csv('Final/Januari 2021 agg.csv', sep=',')
df_outstanding.isnull().sum()

✓ 0.0s
Customer Name      0
Total Outstanding  0
Status Pembayaran  0
dtype: int64
```

Gambar 4. Hasil Cleaning Nilai Kosong

4.3. Data Transformation

Setelah proses *cleaning* selesai, dilakukan transformasi data yang mencakup pemetaan nama pelanggan dan penggabungan data pelanggan dengan nama yang sama. Proses ini secara otomatis mengakumulasi total kiloliter pembelian untuk setiap pelanggan. Kolom tanggal dari data penjualan kemudian disesuaikan sehingga hanya menampilkan bulan saja.

Selanjutnya, dilakukan rekayasa fitur (*feature engineering*) untuk membuat variabel-variabel berbasis waktu yang krusial untuk pemodelan. Atribut “Status Pembayaran Bulan Lalu” dibentuk untuk menangkap riwayat pembayaran pelanggan. Hal ini dicapai dengan melakukan iterasi data secara bulanan dan menggunakan sebuah struktur *data dictionary* untuk menyimpan status pembayaran terakhir dari setiap pelanggan. Saat memproses data bulan berjalan, nilai atribut ini diambil dari catatan yang tersimpan dari bulan sebelumnya.

Bagi pelanggan yang baru pertama kali muncul, nilai default “no data” akan diberikan. Setelah itu, status pembayaran bulan berjalan dicatat untuk digunakan pada iterasi bulan berikutnya.

Dengan mekanisme serupa, atribut “Kategori Masa Tunggakan” dibuat untuk melacak durasi tunggakan pelanggan. Atribut ini berfungsi sebagai penanda umur piutang (*receivable aging*). Jika seorang pelanggan memiliki status “belum lunas” pada bulan berjalan, kategori tunggakannya akan ditingkatkan ke tingkat selanjutnya (misalnya, dari “31-60 hari” menjadi “61-90 hari”). Sebaliknya, jika pelanggan melakukan pelunasan, kategori tunggakannya akan diatur ulang kembali ke kondisi awal, yang menandakan piutang kembali sehat.

Dua atribut penting, yaitu Total_KL (total BBM yang dibeli pelanggan selama sebulan) dan Total_Tunggakan (total nominal tagihan yang belum dibayar), dikelompokkan ke dalam kategori-kategori deskriptif. Pengelompokan ini dilakukan berdasarkan statistik deskriptif yang diperoleh dari fungsi *describe()* pada library Pandas, di mana nilai dari kuartil pertama (25%) hingga nilai maksimum dikonversi menjadi kategori masing-masing.

Puncak dari transformasi data adalah pembentukan variabel target, yaitu “Status Pembayaran Bulan Depan”. Atribut ini dibuat setelah semua data bulanan digabungkan dan diurutkan berdasarkan pelanggan serta bulan. Menggunakan fungsi *shift(-1)* yang dikelompokkan per pelanggan, nilai status pembayaran dari bulan berikutnya diambil dan ditetapkan sebagai target untuk bulan berjalan.

Dari penggabungan dan transformasi ini, dihasilkan dataset akhir dengan beberapa kolom kunci untuk analisis. Dataset yang digunakan untuk pemodelan terdiri dari 1,368 baris data.

```
<class 'pandas.core.frame.DataFrame'>
Index: 1368 entries, 0 to 3727
Data columns (total 8 columns):
#   Column              Non-Null Count  Dtype
---  -
0   Nama Pelanggan      1368 non-null   object
1   Bulan               1368 non-null   object
2   Total_KL            1368 non-null   float64
3   Total_Tunggakan     1368 non-null   float64
4   Status_Pembayaran_Bulan_Lalu  1368 non-null   object
5   Kategori_Masa_Tunggakan  1368 non-null   object
6   Status_Pembayaran   1368 non-null   object
7   Status_Pembayaran_Bulan_Depan  1368 non-null   object
dtypes: float64(2), object(6)
memory usage: 96.2+ KB
```

Gambar 5. Dataset Akhir

4.4. Data Mining

Untuk memprediksi status pembayaran pelanggan pada bulan berikutnya, data yang tidak relevan seperti nama pelanggan dan bulan dikeluarkan sebagai atribut penentu agar tidak mempengaruhi hasil prediksi. Pada tahap ini, digunakan model klasifikasi dari kelas *DecisionTreeClassifier* dari library *scikit-learn* dengan parameter *criterion* diatur ke *entropy*. Setelah atribut yang tidak relevan dikeluarkan, dilakukan pengaturan parameter *training*. Dengan menggunakan *test_size=0.2* untuk melakukan pembagian data, proporsi data diatur agar menggunakan data *training* sebesar 80%, dan data *testing* sebesar 20%. Kemudian, parameter *max_depth* ditambahkan dalam proses training agar model tidak *overfitting*.

Penentuan *max_depth* dilakukan dengan melatih model pada beberapa nilai kedalaman yang berbeda, yaitu dari 1 hingga 10. Dari serangkaian pengujian tersebut, dipilih nilai *max_depth* yang memberikan keseimbangan terbaik antara performa pada data *training* dan data *testing*. Tujuannya adalah untuk mendapatkan akurasi *testing* yang setinggi mungkin tanpa mengorbankan kemampuan model untuk generalisasi pada data baru, yang ditandai dengan selisih kecil antara kedua nilai akurasi tersebut. Untuk penelitian ini, penulis akan menentukan nilai *max_depth* di nilai 2.

Dilakukan pula konfigurasi pada parameter *class_weight* dan *random_state*. Parameter *class_weight* diatur menjadi *balanced* untuk memberikan bobot penalti yang seimbang, sehingga model dapat belajar secara efektif dari kelas minoritas dan tidak cenderung membuat prediksi yang bias terhadap kelas mayoritas yang mungkin mendominasi data. Sementara itu, *random_state* ditetapkan pada nilai konstan 42 di seluruh proses untuk menghilangkan variabilitas acak.

Pada tahap terakhir, ditambahkan fungsi seperti *accuracy_score*, *classification_report*, dan *feature_importances* untuk menganalisa hasil evaluasi dari model, serta penambahan library *graphviz* untuk visualisasi *decision tree*.

```
# --- 8. Training model DecisionTree dengan entropy ---
model = DecisionTreeClassifier(criterion='entropy',
                              class_weight='balanced',
                              random_state=42,
                              max_depth=2)
model.fit(X_train, y_train)
```

Gambar 6. Training Model

```
# --- 9. Evaluasi ---
y_pred = model.predict(X_test)
print("Akurasi:", accuracy_score(y_test, y_pred))
print("Laporan Klasifikasi:\n",
      classification_report(y_test, y_pred, target_names=['Lunas', 'Belum Lunas']))
print("Fitur Penting:\n",
      pd.Series(model.feature_importances_, index=X.columns).sort_values(ascending=False))
```

Gambar 7. Fungsi Evaluasi

4.5. Interpretation

Tahap *interpretation* dilakukan setelah model dilatih menggunakan algoritma *Decision Tree Classifier* dengan *criterion entropy*. Berdasarkan hasil pengujian, model menunjukkan performa yang sangat baik dengan akurasi keseluruhan mencapai 92.70%.

Akurasi: 0.92700729270073

Laporan Klasifikasi:

	precision	recall	f1-score	support
Lunas	0.94	0.97	0.96	224
Belum Lunas	0.84	0.74	0.79	50
accuracy			0.93	274
macro avg	0.89	0.85	0.87	274
weighted avg	0.92	0.93	0.93	274

Fitur Penting:

Status_Pembayaran	0.899692
Total_Tunggakan	0.088682
Status_Pembayaran_Bulan_Lalu	0.011626
Total_KL	0.000000
Kategori_Masa_Tunggakan	0.000000
dtype:	float64

Gambar 8. Hasil Akurasi Model

4.5.1. Metrik Evaluasi

Secara lebih rinci, metrik evaluasi untuk model sebagai berikut.

Tabel 3. Metrik Evaluasi

	Precision	Recall	F1-Score	Support
Lunas	0.94	0.97	0.96	224
Belum Lunas	0.84	0.74	0.79	50

Berdasarkan metrik evaluasi, model ini menunjukkan performa yang sangat baik dalam memprediksi pelanggan yang akan melakukan pembayaran. Hal ini dibuktikan dengan nilai Precision sebesar 0.94, Recall 0.97, dan F1-Score 0.96. Tingginya nilai Recall (0.97) mengindikasikan bahwa model berhasil mengidentifikasi 97% dari seluruh pelanggan yang sebenarnya “Lunas”.

Sedangkan untuk pelanggan yang “Belum Lunas”, model mencapai Precision 0.84 dan Recall 0.74, yang menghasilkan F1-Score sebesar 0.79. Nilai Precision 0.84 berarti 84% dari prediksi “Belum Lunas” yang dibuat model adalah benar. Namun, nilai Recall 0.74

menunjukkan bahwa model masih belum dapat menangkap sekitar 26% dari total pelanggan yang sebenarnya “Belum Lunas”.

4.5.2. Atribut Penting

Analisis atribut dilakukan untuk mengetahui kontribusi masing-masing atribut yang menjadi faktor penyebab terjadinya keterlambatan pembayaran. Hasil analisis sebagai berikut.

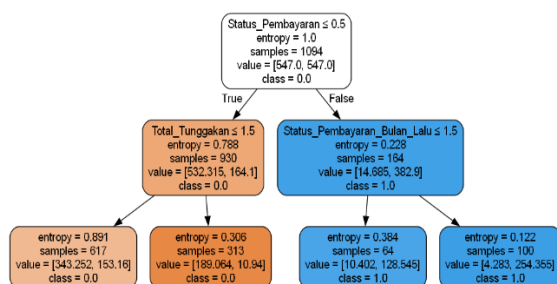
Tabel 4. Atribut Penting

Status Pembayaran	0.899692
Total Tunggakan	0.088682
Status Pembayaran Bulan Lalu	0.011626
Total Kiloliter	0.000000
Kategori Masa Tunggakan	0.000000

Analisis atribut penting mengungkapkan bahwa Status Pembayaran menjadi faktor yang paling dominan dengan skor kepentingan 0.899692. Atribut Total Tunggakan dan Total Kiloliter memiliki kontribusi yang jauh lebih kecil, masing-masing dengan skor 0.088682 dan 0.011626. Sedangkan untuk atribut seperti Kategori Masa Tunggakan dan Status Pembayaran Bulan Lalu memiliki skor kepentingan 0.000000 sehingga tidak memengaruhi pemilihan keputusan.

4.5.3. Visualisasi Decision Tree

Berikut hasil visualisasi dari *decision tree* untuk model ini.



Gambar 9. Visualisasi Decision Tree

Pohon tersebut dapat diinterpretasikan sebagai aturan atau *rules* “IF-THEN” sebagai berikut.

IF Status_Pembayaran is one of (belum lunas) AND Total_Tunggakan is one of (Tunggakan Menengah, Tunggakan Rendah) THEN

Status_Pembayaran_Bulan_Depan = belum lunas

IF Status_Pembayaran is one of (belum lunas) AND Total_Tunggakan is one of (Tunggakan Sangat Tinggi, Tunggakan Tinggi) THEN Status_Pembayaran_Bulan_Depan = belum lunas

IF Status_Pembayaran is one of (lunas) AND Status_Pembayaran_Bulan_Lalu is one of (belum lunas, lunas) THEN Status_Pembayaran_Bulan_Depan = lunas

IF Status_Pembayaran is one of (lunas) AND Status_Pembayaran_Bulan_Lalu is one of (no data) THEN Status_Pembayaran_Bulan_Depan = lunas

Berdasarkan aturan yang diekstraksi dari model pohon keputusan, logika prediksi dapat dijelaskan dalam dua skenario utama. Pertama, untuk setiap pelanggan yang status pembayarannya pada bulan berjalan adalah “Belum Lunas”, model secara konsisten akan memprediksi bahwa status pembayaran mereka pada bulan berikutnya juga akan “Belum Lunas”. Aturan ini berlaku tanpa memandang besarnya total tunggakan pelanggan, baik itu dalam kategori “Tunggakan Rendah”, “Menengah”, “Tinggi”, maupun “Sangat Tinggi”.

Di sisi lain, untuk pelanggan yang status pembayarannya pada bulan berjalan adalah “Lunas”, model akan memprediksi bahwa mereka akan tetap “Lunas” pada bulan berikutnya. Prediksi ini berlaku secara umum untuk semua pelanggan yang membayar lunas, baik mereka yang merupakan pelanggan lama dengan riwayat pembayaran sebelumnya (“Lunas” atau “Belum Lunas”) maupun pelanggan yang baru pertama kali bertransaksi dan belum memiliki riwayat pembayaran (“No Data”). Secara ringkas, model ini menggunakan status pembayaran saat ini sebagai penentu utama prediksi, dengan logika percabangan sekunder yang berbeda tergantung pada apakah pelanggan tersebut lunas atau tidak.

5. KESIMPULAN

1. Penelitian ini berhasil mencapai tujuannya untuk membangun model prediktif yang dapat mengidentifikasi faktor-faktor berpengaruh terhadap keterlambatan pembayaran di PT XYZ.

2. Model *Decision Tree Classifier* yang dibangun menggunakan *criterion entropy* menunjukkan performa yang sangat baik dengan tingkat akurasi keseluruhan mencapai 92.70% dalam memprediksi status pembayaran pelanggan pada bulan berikutnya.
3. Atribut yang paling berpengaruh dalam memengaruhi prediksi keterlambatan pembayaran adalah Status Pembayaran bulan berjalan, dengan skor kepentingan sebesar 0.899692. Faktor lain seperti Total Tunggalan dan Status Pembayaran Bulan Lalu memiliki pengaruh yang jauh lebih kecil.
4. Kelebihan model ini terletak pada kemampuannya yang sangat tinggi dalam mengidentifikasi pelanggan yang akan membayar tepat waktu, yang dibuktikan dengan nilai Precision 0.94, Recall 0.97, dan F1-Score 0.96 untuk kelas "Lunas". Selain itu, model ini menghasilkan aturan "IF-THEN" yang logis dan mudah diinterpretasikan untuk pengambilan keputusan.
5. Kekurangan model ini adalah kemampuannya yang lebih rendah dalam menangkap semua pelanggan yang berpotensi "Belum Lunas". Nilai Recall sebesar 0.74 untuk kelas "Belum Lunas" menunjukkan bahwa model masih belum dapat mengidentifikasi sekitar 26% dari total pelanggan yang sebenarnya tidak melakukan pembayaran.
6. Pengembangan selanjutnya dapat difokuskan untuk meningkatkan kinerja model dalam memprediksi kelas minoritas ("Belum Lunas"). Hal ini dapat dilakukan dengan mengeksplorasi teknik *feature engineering* yang lebih mendalam, menggunakan algoritma klasifikasi lain, atau memperkaya dataset dengan lebih banyak variabel yang relevan untuk mengatasi kekurangan yang ada.

UCAPAN TERIMA KASIH

Ucapan terima kasih sebesar-besarnya penulis sampaikan kepada para dosen dari Universitas Muhammadiyah Banjarmasin yang telah memberikan bimbingan, arahan, dan dukungan yang tak ternilai selama proses penelitian ini berlangsung. Penghargaan setinggi-tingginya juga penulis haturkan kepada

seluruh tim editorial Jurnal Informatika dan Teknik Elektro Terapan (JITET) atas masukan konstruktif dan kesempatan yang diberikan sehingga penelitian ini dapat dipublikasikan dan menjangkau khalayak yang lebih luas.

DAFTAR PUSTAKA

- [1] A. A. Putri dan A. Noviandry, "Analisis Pengelolaan Arus Kas Dalam Menjaga Likuiditas di PT. Marisa Jaya Abadi," *Jurnal Manajemen dan Ekonomi Kreatif*, vol. 3, no. 1, hlm. 24–40, Jan 2025, doi: 10.59024/jumek.v3i1.508.
- [2] R. Rachman, "Analisa Pengendalian Piutang Terhadap Resiko Piutang Tak Tertagih Pada PT Enseval Putra Megatrading Tbk Cabang Bogor," *Jurnal Ilmiah Akuntansi Kesatuan*, vol. 7, hlm. 243–250, 2019.
- [3] E. B. F. Damayanty dan U. Lasiyono, "Analisis Kebijakan Penjualan Kredit DAN Pengelolaan Piutang Untuk Meminimalkan Piutang Tidak Tertagih Pada PT. Subur Agro Pratama," *Journal of Sustainability Business Research*, vol. 3, no. 4, hlm. 54–61, Des 2022.
- [4] R. A. Aztari, "Pengaruh Risiko Kredit, Struktur Modal, Dan Likuiditas Terhadap Profitabilitas Pada Perusahaan Perbankan Yang Terdaftar di Bursa Efek Indonesia," *Jurnal Ilmu dan Riset Akuntansi*, vol. 12, no. 5, 2023.
- [5] Modestus La'a dan Vip Pramarta, "Pengembangan Sistem Informasi Laboratorium Untuk Meningkatkan Efisiensi dan Akurasi," *Jurnal ilmiah Sistem Informasi dan Ilmu Komputer*, vol. 3, no. 2, hlm. 244–255, Jul 2023, doi: 10.55606/juisik.v3i2.517.
- [6] T. Connolly dan C. Begg, *Database Systems: a practical approach to design, implementation, and management*, 5 ed. Pearson Education.
- [7] R. N. Ramadhon, A. Ogi, A. P. Agung, R. Putra, S. S. Febrihartina, dan U. Firdaus, "Implementasi Algoritma Decision Tree untuk Klasifikasi Pelanggan Aktif atau Tidak Aktif pada Data Bank," *Jurnal Karimah Tauhid*, vol. 3, no. 2, 2024, doi: <https://doi.org/10.30997/karimahtauhid.v3i2.11952>.
- [8] M. H. Sukri dan Y. Handrianto, "Penerapan Algoritma C4.5 Dalam Menentukan Prediksi Prestasi Siswa Pada SMPN 51 Jakarta," *Informatics and Computer Engineering Journal*, vol. 4, no. 1, 2024, [Daring]. Tersedia pada: <https://jurnal.bsi.ac.id/index.php/ijec/>
- [9] A. Arif dan A. Yulianto, "Klasifikasi Penjualan Makanan Hewan Peliharaan di Muezza Petshop01 Bogor Menggunakan Metode Algoritma C4.5," *Jurnal Riset dan E-Jurnal Manajemen Informatika Komputer*, vol. 6, no.

- 4, hlm. 840–854, Okt 2022, doi: 10.33395/remik.v6i4.11859.
- [10] S. Dewi Nugraha, M. Hikmatyar, J. Teknik Informatika, F. Teknik, dan U. Perjuangan Tasikmalaya, “Penerapan Model Klasifikasi Pada Kelulusan Mahasiswa di Universitas Perjuangan Tasikmalaya Menggunakan Algoritma C4.5,” *Jurnal Informatika dan Teknik Elektro Terapan*, vol. 12, no. 3, hlm. 3402–3411, Agu 2024, doi: 10.23960/jitet.v12i3.5265.
- [11] M. Mubarakah, D. Aditya Nugraha, dan A. Yunus, “Penerapan Algoritma C4.5 dalam Memprediksi Keterlambatan Pembayaran Sumbangan Pembinaan Pendidikan,” *Jurnal Riset Mahasiswa Bidang Teknologi Informasi*, vol. 7, no. 1, 2024, doi: <https://doi.org/10.21067/bimasakti.v7i1.10368>.
- [12] A. Rohman dan A. Rufiyanto, “Implementasi Data Mining Dengan Algoritma Decision Tree C4.5 Untuk Prediksi Kelulusan Mahasiswa di Universitas Pandanaran,” dalam *Proceeding SINTAK 2019*, 2019, hlm. 134–139.
- [13] N. Y. L. Gaol, “Prediksi Mahasiswa Berpotensi Non Aktif Menggunakan Data Mining dalam Decision Tree dan Algoritma C4.5,” *Jurnal Informasi & Teknologi*, hlm. 23–29, Mar 2020, doi: 10.37034/jidt.v2i1.22.
- [14] D. N. Batubara, A. P. Windarto, dan E. Irawan, “Analisis Prediksi Keterlambatan Pembayaran Listrik Menggunakan Komparasi Metode Klasifikasi Decision Tree dan Support Vector Machine,” *JURIKOM (Jurnal Riset Komputer)*, vol. 9, no. 1, hlm. 102, Feb 2022, doi: 10.30865/jurikom.v9i1.3833.
- [15] I. A. Subekti, C. B. Andriano, D. Nurdiansyah, dan R. Hidayat, “Penerapan Algoritma C4.5 untuk Memprediksi Harga Rumah,” *Jurnal Sistem Informasi, Teknologi Informasi, dan Komputer*, vol. 15, no. 2, hlm. 350, Jan 2025.
- [16] M. Bramer, *Principles of Data Mining*, Fourth Edition. London: Springer, 2020.