

Prediksi Keterlambatan Pembayaran Piutang Mahasiswa Menggunakan Algoritma Naïve Bayes (Studi Kasus Institut Teknologi dan Bisnis Muhammadiyah Wakatobi)


Predicting Late Payments of Student Receivables Using the Naive Bayes Algorithm (Case Study of the Muhammadiyah Wakatobi Institute of Technology and Business)

Anisa^{a,1,*}, Hamiruddin S^{a,2}, Kariyamin^{a,3}

^a Teknologi Informasi, Intitut Teknologi dan Bisnis Muhammadiyah Wakatobi, Wakatobi, Indonesia

¹nisa00ani@gmail.com; ²muhammadhamiruddin@gmail.com; ³karyamin28@gmail.com

*corresponding author

Informasi Artikel	ABSTRAK
<p>Diserahkan : 12 Agustus 2025 Diterima : 9 Februari 2026 Direvisi : 23 Februari 2026 Diterbitkan : 25 Februari 2026</p> <p>Kata Kunci: Prediksi Naïve Bayes Keterlambatan Pembayaran</p>	<p>Keterlambatan pembayaran piutang mahasiswa merupakan salah satu masalah penting dalam pengelolaan keuangan institusi pendidikan tinggi, termasuk di Institut Teknologi dan Bisnis Muhammadiyah Wakatobi (ITBM). Penelitian ini bertujuan memprediksi keterlambatan pembayaran menggunakan algoritma Naïve Bayes yang sederhana namun efektif. Data berasal dari mahasiswa aktif angkatan ke-2, ke-3, dan ke-4 tahun akademik 2024, dengan atribut seperti pendidikan, pekerjaan dan pendapatan orang tua, jumlah utang, serta status beasiswa. Data diolah melalui pembersihan dan transformasi numerik menggunakan TF-IDF, kemudian diuji dengan metode 5-Fold Cross Validation. Hasil evaluasi menunjukkan akurasi rata-rata 71%, presisi 82%, recall 56%, dan F1-score 65%. Presisi yang tinggi menunjukkan model cukup andal mengidentifikasi mahasiswa yang membayar tepat waktu, meskipun recall yang rendah mengindikasikan masih ada keterbatasan dalam menangkap semua kasus keterlambatan. Temuan ini menunjukkan bahwa algoritma Naïve Bayes berpotensi menjadi alat bantu prediktif dalam mendukung manajemen risiko keuangan institusi pendidikan tinggi. Disarankan pengembangan lebih lanjut dengan teknik penyeimbangan data dan perluasan atribut untuk meningkatkan kinerja model.</p>
<p>Keywords: Predict Naive Bayes Delay Payment</p>	<p>ABSTRACT</p> <p><i>Late payment of student receivables is a significant issue in the financial management of higher education institutions, including at the Muhammadiyah Wakatobi Institute of Technology and Business (ITBM). This study aims to predict late payments using the simple yet effective Naïve Bayes algorithm. The data is sourced from active students in the 2nd, 3rd, and 4th cohorts of the 2024 academic year, with attributes such as parents' education, occupation, income, total debt, and scholarship status. The data was processed through cleaning and numerical transformation using TF-IDF, and tested with the 5-Fold Cross Validation method. The evaluation results showed an average accuracy of 71%, precision of 82%, recall of 56%, and F1-score of 65%. The high precision indicates that the model is reliable in identifying students who pay on time, although the low recall suggests limitations in capturing all late payment cases. These findings suggest that the Naïve Bayes algorithm has potential as a predictive tool to support financial risk management in higher education institutions. Further development is recommended by integrating data balancing techniques and expanding data attributes to improve model performance.</i></p>
<p>This is an open access article under the CC-BY-SA license.</p> 	

I. Pendahuluan

Pendidikan tinggi berperan penting dalam mencetak sumber daya manusia berkualitas yang mampu bersaing di era global. Namun, banyak institusi pendidikan di seluruh dunia menghadapi tantangan dalam pengelolaan keuangan, salah satunya keterlambatan pembayaran biaya pendidikan oleh mahasiswa. Fenomena

ini tidak hanya mengganggu stabilitas arus kas institusi, tetapi juga dapat memengaruhi keberlanjutan layanan akademik yang optimal. Seiring meningkatnya jumlah mahasiswa dan kompleksitas manajemen keuangan, penerapan teknologi analitik berbasis data menjadi salah satu pendekatan strategis yang mulai banyak diadopsi untuk meminimalkan dampak keterlambatan pembayaran.

Di Indonesia, termasuk di Institut Teknologi dan Bisnis Muhammadiyah Wakatobi (ITBM), sistem pembayaran cicilan diterapkan untuk membantu meringankan beban mahasiswa. Meskipun demikian, keterlambatan pembayaran masih sering terjadi sehingga memerlukan strategi prediktif yang akurat untuk mendukung kebijakan penagihan dan perencanaan keuangan. Penelitian ini menggunakan data keuangan, akademik, dan demografis mahasiswa dari angkatan 2 sampai 4 pada tahun akademik 2024, yang diperoleh dari sistem administrasi kampus.

Pendekatan berbasis data mining, khususnya algoritma klasifikasi seperti Naïve Bayes, telah banyak digunakan untuk memprediksi perilaku pembayaran mahasiswa atau kasus serupa. Berbagai penelitian terdahulu menunjukkan efektivitas Naïve Bayes dalam memproses data akademik dan keuangan penelitian oleh [1] membuktikan algoritma ini mampu memprediksi piutang pendidikan dengan akurasi 92,31% pada skema data training dan testing 80:20. Penelitian lain [2] mengoptimalkan Naïve Bayes menggunakan Particle Swarm Optimization (PSO) untuk prediksi keterlambatan pembayaran uang kuliah di Politeknik TEDC Bandung dengan akurasi 73,94%. [3] menggunakan Naïve Bayes untuk memprediksi kelancaran pembayaran angsuran koperasi, menghasilkan presisi 100% dan recall 83,33%.

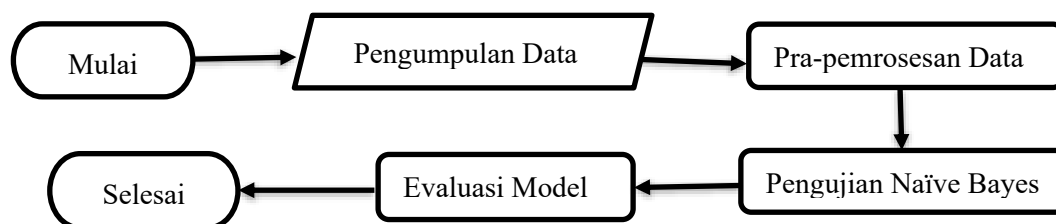
Penelitian oleh [4] juga menunjukkan penerapan Naïve Bayes pada analisis sentimen publik di media sosial X, yang memanfaatkan kombinasi preprocessing teks, representasi fitur TF-IDF, dan evaluasi model menggunakan akurasi, presisi, recall, serta F1-score. Meskipun konteksnya berbeda, yakni analisis sentimen terhadap Gerakan Muhammadiyah, penelitian ini membuktikan bahwa kombinasi metode tersebut dapat menghasilkan kinerja klasifikasi yang tinggi dengan akurasi 87,5%, precision 84,2%, recall 85,3%, dan F1-score 84,7%. Pendekatan ini relevan untuk diadaptasi dalam prediksi keterlambatan pembayaran mahasiswa karena sama-sama melibatkan pengolahan data yang kompleks dan tidak terstruktur.

Studi lain, seperti [5], [6], [7], [8], [9], dan [10], mengungkap bahwa kinerja Naïve Bayes dapat ditingkatkan melalui optimasi fitur, teknik ensemble, dan algoritma metaheuristik. Meskipun demikian, masih terdapat celah penelitian pada aspek evaluasi model yang lebih komprehensif. Banyak studi sebelumnya hanya menitikberatkan pada metrik akurasi, presisi, dan recall, tanpa mempertimbangkan keseimbangan antara keduanya yang dapat diukur melalui F1-score. Selain itu, metode validasi silang seperti K-Fold Cross Validation belum dimanfaatkan secara optimal untuk menguji konsistensi model pada berbagai subset data.

Berdasarkan kondisi tersebut, penelitian ini berfokus pada penerapan algoritma Naïve Bayes untuk memprediksi keterlambatan pembayaran mahasiswa di ITBM Wakatobi dengan pendekatan evaluasi yang lebih menyeluruh. Proses analisis melibatkan pembersihan data, transformasi numerik menggunakan Term Frequency–Inverse Document Frequency (TF-IDF), serta pengujian model dengan metode 5-Fold Cross Validation menggunakan empat metrik evaluasi utama: akurasi, presisi, recall, dan F1-Score. Kontribusi penelitian ini adalah mengembangkan model prediksi keterlambatan pembayaran dengan kombinasi TF-IDF, Naïve Bayes, dan 5-Fold Cross Validation pada dataset keuangan-akademik mahasiswa ITBM Wakatobi, yang belum pernah diterapkan sebelumnya.

II. Metode

Penelitian ini menggunakan metode eksperimental dengan pendekatan kuantitatif untuk memprediksi keterlambatan pembayaran mahasiswa di Institut Teknologi dan Bisnis Muhammadiyah Wakatobi (ITBM) menggunakan algoritma Naïve Bayes dengan teknik pembobotan TF-IDF, yang umumnya digunakan untuk teks, namun diadaptasikan dalam pemrosesan atribut non-teks. Tahapan penelitian meliputi pengumpulan data, pra-pemrosesan data, penerapan algoritma, dan evaluasi kinerja model. Alur penelitian ditunjukkan pada Gambar 1 dengan penggunaan simbol flowchart standar, di mana terminator (oval) digunakan untuk proses mulai dan selesai.



Gambar 1. Alur Penelitiann

A. Pengumpulan Data

Data penelitian diperoleh dari sistem administrasi akademik dan keuangan Institut Teknologi dan Bisnis Muhammadiyah Wakatobi (ITBM). Dataset mencakup mahasiswa aktif angkatan ke-2, ke-3, dan ke-4 pada tahun akademik 2024. Data yang dikumpulkan meliputi informasi identitas, data demografis, dan data keuangan mahasiswa yang relevan untuk proses prediksi keterlambatan pembayaran dengan total data sebanyak 144 data mahasiswa pada Table 1 dibawah.

Tabel 1. Data Mahasiswa

Nim	Nama Mahasiswa	Program Studi	Angkatan	Status	Alamat Kecamatan	Gender	Pendidikan Orang Tua	Pekerjaan Orang Tua	Penghasilan Orang Tua	Jumlah Tagihan Semester	Jumlah Piutang	Jumlah Tunggakan	Beasiswa
221572010001	La Ode Ahmad Jumaidin	Teknologi Informasi	2021	Aktif	Wangi-wangi Selatan	L	SD	Petani	< Rp 500.000	Rp 1.200.000	Rp 8.400.000	Rp 600.000	Tidak
221572010002	Anisa	Teknologi Informasi	2021	Aktif	Kaledupa Selatan	P	SD	Petani	< Rp 500.000	Rp 1.200.000	Rp 8.400.000	Lunas	Tidak
221572010003	Fika	Teknologi Informasi	2021	Aktif	Wangi-wangi	P	SMP	Petani	< Rp 500.000	Rp 1.200.000	Rp 8.400.000	Lunas	Tidak
221572010007	La Ode Risman Firdaus	Teknologi Informasi	2021	Aktif	Wangi-wangi	L	SD	Petani	< Rp 500.000	Rp 1.200.000	Rp 8.400.000	Rp 7.900.000	Tidak
.....
223942020016	Safa Andika	Kewirausahaan	2023	Aktif	Wangi-wangi	L	-	-	< Rp 500.000	Rp 1.400.000	Rp 4.200.000	lunas	Tidak

Untuk memastikan kelengkapan dan relevansi, setiap entri data memuat informasi penting yang digunakan sebagai atribut penelitian, seperti pendidikan orang tua, pekerjaan orang tua, penghasilan orang tua, jumlah tagihan semester, jumlah tagihan, jumlah tunggakan, dan status beasiswa. Atribut-atribut ini digunakan untuk menentukan status pembayaran siswa, apakah Tepat Waktu atau Terlambat.

B. Pra-Pemrosesan Data

Seperti terlihat pada Gambar 2, data yang diperoleh juga telah melalui proses pembersihan untuk memastikan tidak ada kesalahan penulisan, nilai kosong, maupun inkonsistensi format, sehingga data siap digunakan pada tahap pemodelan prediksi menggunakan algoritma Naïve Bayes

Proses pembersihan data diawali dengan standarisasi nama kolom, termasuk memperbaiki kesalahan penulisan seperti mengubah "Jumlam Tunggakan" menjadi "Jumlah Tunggakan", serta memastikan tidak ada nilai yang hilang menggunakan `isnull().sum()`. Selanjutnya, `aclean_text()_combined_clean_text` diubah menjadi format numerik menggunakan **TF-IDF** agar siap diolah pada tahap pemodelan menggunakan algoritma Naïve Bayes.

```
df.columns = df.columns.str.strip()
df.rename(columns={'Jumlam Tunggakan': 'Jumlah Tunggakan'}, inplace=True)
# Menampilkan jumlah nilai yang hilang
print(df.isnull().sum())
vectorizer = TfidfVectorizer() # Fungsi pembersih teks

def clean_text(text):
    if isinstance(text, str):
        text = text.lower()
        text = re.sub(r'\d+', '', text)
        text = text.translate(str.maketrans('', '', string.punctuation))
        return text.strip()
    return ""

# Vektorisasi teks gabungan
vectorizer = TfidfVectorizer()
X = vectorizer.fit_transform(df['combined_clean_text'])
y = df['label']
```

Gambar 2. Pembersihan Data

C. Pengujian Naïve Bayes

Pengujian model dilakukan menggunakan algoritma Naïve Bayes dengan metode 5-Fold Cross Validation untuk memastikan evaluasi yang lebih akurat dan stabil. Pada proses ini, dataset dibagi menjadi lima bagian, di mana setiap bagian secara bergantian digunakan sebagai data uji, sementara empat bagian lainnya digunakan sebagai data latih. Model dibor menggunakan data hasil vektorisasi TF-IDF, kemudian menghasilkan prediksi yang dibandingkan dengan label sebenarnya untuk menghitung metrik evaluasi. Seperti terlihat pada Gambar 3 dibawah

```
# Inisialisasi KFold dan model
kf = KFold(n_splits=5, shuffle=True, random_state=42)
classifier = MultinomialNB()
# Cross-validation
for fold, (train_idx, test_idx) in enumerate(kf.split(X), 1):
    X_train, X_test = X[train_idx], X[test_idx]
    y_train, y_test = y.iloc[train_idx], y.iloc[test_idx]
    classifier.fit(X_train, y_train)
    y_pred = classifier.predict(X_test)
```

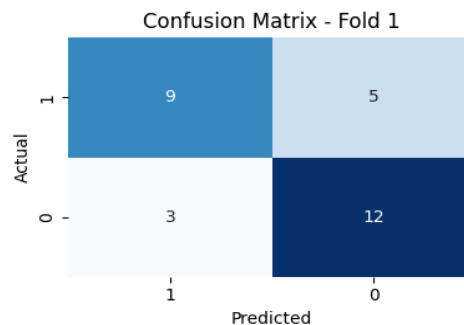
Gambar 3. Pengujian Model

D. Evaluasi Model

Evaluasi model adalah tahap untuk menilai kinerja model yang telah dibangun dalam memprediksi atau mengklasifikasikan data. Tahap ini dilakukan setelah proses pengolahan data dan pembentukan model, dengan tujuan untuk mengetahui tingkat akurasi, ketepatan, dan kemampuan generalisasi model terhadap data baru. Evaluasi biasanya menggunakan metrik tertentu sesuai dengan jenis permasalahan, seperti accuracy, precision, recall, dan F1-score untuk klasifikasi, atau MAE, MSE, dan R² untuk regresi. Selain itu, teknik seperti train-test split dan cross validation juga digunakan untuk memastikan model tidak mengalami overfitting dan benar-benar mampu memberikan hasil yang optimal.

III. Hasil dan Pembahasan

Evaluasi model algoritma Naïve Bayes dilakukan menggunakan metode 5-Fold Cross Validation dengan menggunakan metrik akurasi, presisi, recall, dan F1-score yang dihitung berdasarkan matriks kebingungan, sedangkan hasil perhitungan setiap fold disajikan di bagian Hasil dan Pembahasan.



Gambar 4. Confusion Matrix- Fold 1

Tabel 2. Hasil Fold 1

TP	FP	TN	FN
9	3	12	5
Akurasi	Presisi	Recall	F1-score
72%	75%	64%	69%

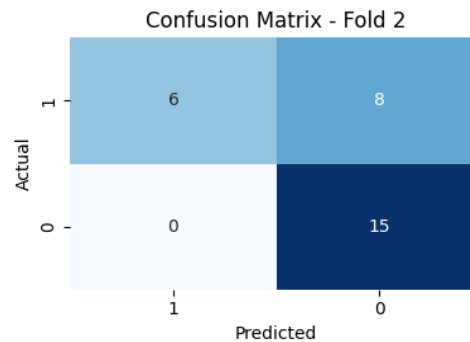
Pada Fold 1 bisa dilihat pada Gambar 4 dan Table 2, model menghasilkan True Positive (TP) sebanyak 9, False Positive (FP) sebanyak 3, True Negative (TN) sebanyak 12, dan False Negative (FN) sebanyak 5. Nilai ini menghasilkan akurasi sebesar 72%, presisi 75%, recall 64%, dan F1-score 69%. Hasil ini menunjukkan model cukup baik dalam mengklasifikasikan mahasiswa tepat waktu, tetapi masih terdapat 5 kasus keterlambatan yang salah diklasifikasikan sebagai tepat waktu.

$$\text{Akurasi: } \frac{TP+TN}{TP+TN+FP+FN} = \frac{9+12}{9+12+3+5} = \frac{21}{29} = 0,7241 \text{ atau } 72\%$$

$$\text{Presisi: } \frac{TP}{TP+FP} = \frac{9}{9+3} = \frac{9}{12} = 0,75 \text{ atau } 75\%$$

$$\text{Recall: } \frac{TP}{TP+FN} = \frac{9}{9+5} = \frac{9}{14} = 0,6428 \text{ atau } 64\%$$

$$\text{F1 Score: } \frac{2 \cdot (0,75 \cdot 0,6428)}{0,75 + 0,6428} = \frac{0,9642}{1,3928} = 0,6924 \text{ atau } 69\%$$



Gambar 5. Confusion Matrix- Fold 2

Tabel 3. Hasil Fold 2

TP	FP	TN	FN
6	0	15	8
Akurasi	Presisi	Recall	F1-score
72%	100%	43%	60%

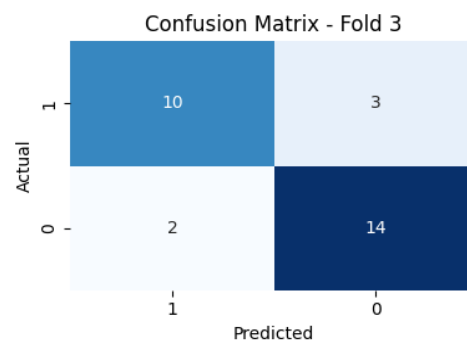
Fold 2 menunjukkan peningkatan kinerja dengan TP = 6, FP = 0, TN = 15, dan FN = 8, menghasilkan akurasi 72%, presisi 100%, recall 43%, dan F1-score 60%, sesuai dengan nilai yang disajikan pada Tabel 3.

$$\text{Akurasi: } \frac{TP+TN}{TP+TN+FP+FN} = \frac{6+15}{6+15+0+8} = \frac{21}{29} = 0,7241 \text{ atau } 72\%$$

$$\text{Presisi: } \frac{TP}{TP+FP} = \frac{6}{6+0} = \frac{6}{6} = 1,00 \text{ atau } 100\%$$

$$\text{Recall: } \frac{TP}{TP+FN} = \frac{6}{6+8} = \frac{6}{14} = 0,4285 \text{ atau } 43\%$$

$$\text{F1 score: } \frac{2 \cdot (1,0 \cdot 0,4285)}{1,0 + 0,4285} = \frac{0,857}{1,4285} = 0,60 \text{ atau } 60\%$$



Gambar 6. Confusion Matrix- Fold 3

Tabel 4. Hasil Fold 3

TP	FP	TN	FN
10	2	14	3
Akurasi	Presisi	Recall	F1-score
83%	83%	77%	80%

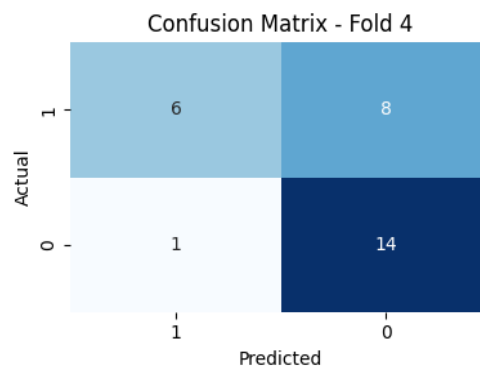
Fold 3 memperoleh TP = 10, FP = 2, TN = 14, dan FN = 3, dengan akurasi 87%, presisi 83%, recall 77%, dan F1-score 80%. Performa pada fold ini merupakan salah satu yang terbaik karena keseimbangan antara presisi dan recall yang relatif baik.

$$\text{Akurasi: } \frac{TP+TN}{TP+TN+FP+FN} = \frac{10+14}{10+14+2+3} = \frac{24}{29} = 0,8276 \text{ atau } 83\%$$

$$\text{Presisi: } \frac{TP}{TP+FP} = \frac{10}{10+2} = \frac{10}{12} = 0,8333 \text{ atau } 83\%$$

$$\text{Recall: } \frac{TP}{TP+FN} = \frac{10}{10+3} = \frac{10}{13} = 0,7692 \text{ atau } 77\%$$

$$\text{F1 score: } \frac{2 \cdot (0,8333 \cdot 0,7692)}{0,8333 + 0,7692} = \frac{1,2819}{1,6025} = 0,7999 \text{ atau } 80\%$$



Gambar 7. Confusion Matrix- Fold 4

Tabel 5. Hasil Fold 4

TP	FP	TN	FN
6	1	14	8
Akurasi	Presisi	Recall	F1-score
69%	86%	43%	57%

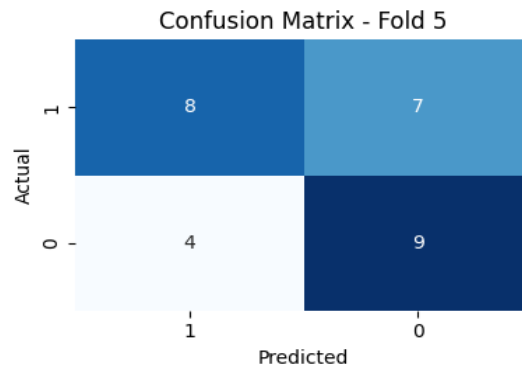
Sebaliknya, Fold 4 menunjukkan penurunan kinerja dengan TP = 6, FP = 1, TN = 14, dan FN = 7, menghasilkan akurasi 69%, presisi 86%, recall 46%, dan F1-score 57%. Nilai recall yang rendah menunjukkan bahwa model kurang optimal dalam mendeteksi mahasiswa terlambat, kondisi ini mengindikasikan adanya ketidakseimbangan distribusi kelas pada sunset data.

$$\text{Akurasi: } \frac{TP+TN}{TP+TN+FP+FN} = \frac{6+14}{6+14+1+8} = \frac{20}{29} = 0,6896 \text{ atau } 69\%$$

$$\text{Presisi: } \frac{TP}{TP+FP} = \frac{6}{6+1} = \frac{6}{7} = 0,8571 \text{ atau } 86\%$$

$$\text{Recall: } \frac{TP}{TP+FN} = \frac{6}{6+8} = \frac{6}{14} = 0,4285 \text{ atau } 43\%$$

$$\text{F1 score: } \frac{2 \cdot (0,8571 \cdot 0,4285)}{0,8571 + 0,4285} = \frac{0,7345}{1,2856} = 0,5713 \text{ atau } 57\%$$



Gambar 8. Confusion Matrix- Fold 5

Tabel 6. Hasil Fold 5

TP	FP	TN	FN
8	4	9	7
Akurasi	Presisi	Recall	F1-score
60%	66%	53%	59%

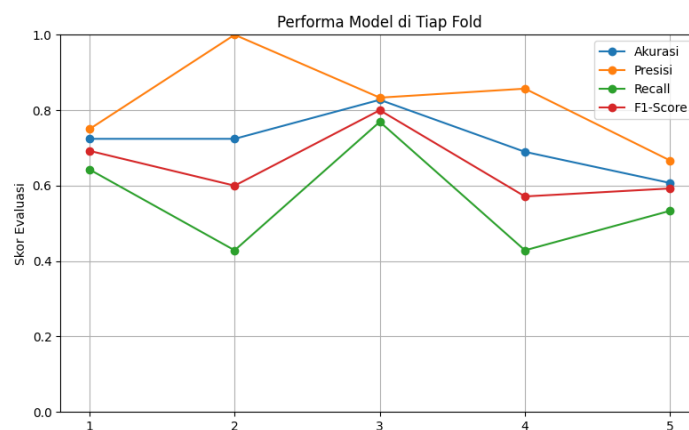
Sedangkan, Fold 5 memberikan hasil TP = 8, FP = 4, TN = 9, dan FN = 7, dengan akurasi 64%, presisi 67%, recall 53%, dan F1-score 59%. Rendahnya presisi mengindikasikan model cukup sering salah mengklasifikasikan mahasiswa yang terlambat menjadi tepat waktu.

$$\text{Akurasi: } \frac{TP+TN}{TP+TN+FP+FN} = \frac{8+9}{8+9+4+7} = \frac{17}{28} = 0,6071 \text{ atau } 60\%$$

$$\text{Presisi: } \frac{TP}{TP+FP} = \frac{8}{8+4} = \frac{8}{12} = 0,6666 \text{ atau } 66\%$$

$$\text{Recall: } \frac{TP}{TP+FN} = \frac{8}{8+7} = \frac{8}{15} = 0,5333 \text{ atau } 53\%$$

$$\text{F1 score: } \frac{2 \cdot (0,6666 \cdot 0,5333)}{0,6666 + 0,5333} = \frac{0,7109}{1,1999} = 0,5924 \text{ atau } 59\%$$



Gambar 9. Hasil Performa Tiap Fold

Pada Gambar 9 grafik “Performa Model di Tiap Fold” menampilkan hasil evaluasi algoritma *Naïve Bayes* pada lima kali pengujian *cross-validation* menggunakan empat metrik: akurasi, presisi, recall, dan F1-score. Pada Fold 1, performa model relatif seimbang dengan akurasi sekitar 72%, presisi 75%, recall 64%, dan F1-score 69%. Fold 2 menunjukkan lonjakan presisi mencapai 100%, meskipun recall turun menjadi 43%, yang menandakan model sangat tepat dalam mengidentifikasi mahasiswa tepat waktu namun kurang mampu

mengetahui keterlambatan. Fold 3 merupakan salah satu hasil terbaik, dengan presisi dan recall yang relatif seimbang (83% dan 77%), serta akurasi 87% dan F1-score 80%. Sebaliknya, Fold 4 memperlihatkan penurunan kinerja, terutama pada recall yang kembali rendah (43%), meski presisi tetap cukup tinggi di 86%. Pada Fold 5, semua metrik menurun, dengan akurasi 64%, presisi 67%, recall 53%, dan F1-score 59%. Secara umum, grafik ini memperlihatkan bahwa model cenderung memiliki presisi yang lebih tinggi dibanding recall pada sebagian besar fold, yang berarti model lebih andal dalam memprediksi mahasiswa tepat waktu dibanding mendeteksi keterlambatan pembayaran. Perbedaan performa antar fold dipengaruhi oleh variasi distribusi data latih dan data uji pada setiap pembagian, khususnya karena jumlah data mahasiswa terlambat relatif lebih sedikit dibandingkan mahasiswa tepat waktu.

Tabel 7. Hasil Nilai Rata-rata 5-Fold

Akurasi	Presisi	Recall	F1-score
71%	82%	56%	65%

Hasil rata-rata evaluasi model berdasarkan 5-Fold Cross Validation ditunjukkan pada Tabel 7. menghasilkan akurasi sebesar 71%, presisi 82%, recall 56%, dan F1-score 65%. Nilai presisi yang tinggi mengindikasikan bahwa model lebih mampu memprediksi mahasiswa yang tepat waktu dengan tingkat kesalahan yang rendah. Namun, nilai recall yang relatif rendah menunjukkan adanya keterbatasan model dalam mendeteksi seluruh kasus keterlambatan pembayaran, sehingga masih terdapat mahasiswa yang terlambat namun diprediksi tepat waktu.

Pola ini mengindikasikan adanya kemungkinan distribusi data yang tidak seimbang (*class imbalance*), di mana jumlah mahasiswa tepat waktu lebih banyak dibanding mahasiswa terlambat. Hal ini sejalan dengan temuan [3] dan [5], yang juga mencatat perbedaan signifikan antara presisi dan recall pada kasus prediksi keterlambatan berbasis data keuangan. Perbedaan metrik ini dapat memengaruhi efektivitas penerapan model di dunia nyata, khususnya jika tujuan utama adalah meminimalkan risiko piutang.

Jika dibandingkan dengan penelitian [6] pada prediksi kelancaran pembayaran kredit yang mencapai akurasi 85%, hasil pada penelitian ini masih di bawahnya. Perbedaan ini kemungkinan disebabkan oleh variasi atribut dataset dan distribusi kelas target. Untuk meningkatkan performa, disarankan penerapan teknik penyeimbangan data seperti *SMOTE* atau kombinasi model (*ensemble learning*) yang telah terbukti efektif pada penelitian [2] dan [9].

IV. Kesimpulan dan saran

Kesimpulan dari penelitian ini menunjukkan Berkontribusi pada penerapan kombinasi *TF-IDF*, algoritma Naïve Bayes, dan evaluasi *5-Fold Cross Validation* untuk memprediksi keterlambatan pembayaran piutang mahasiswa di ITBM Wakatobi, dengan akurasi 71%, presisi 82%, recall 56%, dan F1-score 65%, di mana Kontribusi utama penelitian ini adalah penerapan kombinasi TF-IDF, algoritma Naïve Bayes, dan evaluasi 5-Fold Cross Validation pada data keuangan-akademik mahasiswa ITBM Wakatobi untuk prediksi keterlambatan pembayaran piutang. Presisi yang tinggi menandakan kemampuan model yang baik dalam mengidentifikasi mahasiswa tepat waktu, meskipun recall yang rendah mengindikasikan keterbatasan dalam mendeteksi seluruh kasus keterlambatan oleh karena itu, untuk penelitian selanjutnya disarankan penerapan teknik penyeimbangan data seperti *SMOTE* untuk mengatasi class imbalance, menambahkan atribut dataset yang lebih variatif, serta menguji pendekatan hybrid atau ensemble learning dan membandingkan performa dengan algoritma lain seperti *Random Forest* atau *Gradient Boosting* guna memperoleh model prediksi yang lebih optimal dan aplikatif dalam mendukung manajemen risiko keuangan institusi.

Daftar Pustaka

- [1] S. F. Hamka, K. Kusriani, and K. Kusnawi, "Prediksi Piutang Biaya Pendidikan Mahasiswa Tak Tertagih menggunakan Algoritma Naïve Bayes di Institut Teknologi dan Bisnis Muhammadiyah Wakatobi," *J. Inform. Univ. Pamulang*, vol. 8, no. 2, pp. 199–208, 2023, doi: 10.32493/informatika.v8i2.30711.
- [2] D. Rohmayani, "Analysis of Student Tuition Fee Pay Delay Prediction Using Naive Bayes Algorithm With Particle Swarm Optimization Optimazation (Case Study : Politeknik Tedc Bandung)," *J. Teknol. Inf. dan Pendidik.*, vol. 13, no. 2, pp. 1–8, 2020, doi: 10.24036/tip.v13i2.317.
- [3] Suwati, R. Yesputra, and A. Sapta, "Prediksi Kelancaran Pembayaran Angsuran Pada Koperasi Dengan Metode Naive Bayes Classifier," *Indones. J. Comput. Sci.*, vol. 11, no. 2, pp. 635–644, 2022, doi: 10.33022/ijcs.v11i2.3080.
- [4] S. A. Putri, N. W. Kencana, and A. Khoirudin, "Analisis Sentimen Media Sosial X terhadap Gerakan

- Muhammadiyah Menggunakan Algoritma Naïve Bayes,” *Bul. Sist. Inf. dan Teknol. Islam*, vol. 6, no. 1, pp. 9–17, 2025, doi: 10.33096/busiti.v6i1.2569.
- [5] T. A. Y. Siswa and R. P. Wibowo, “Komparasi Metode Seleksi Fitur Dalam Prediksi Keterlambatan Pembayaran Biaya Kuliah,” *Teknika*, vol. 12, no. 1, pp. 73–82, 2023, doi: 10.34148/teknika.v12i1.601.
- [6] F. Gultom and T. Simanjuntak, “Prediksi Tingkat Kelancaran Pembayaran Kredit Bank Dengan Menggunakan Algoritma Naïve Bayes Dan K-Nearest Neighbor,” *METHOMIKA J. Manaj. Inform. dan Komputerisasi Akunt.*, vol. 4, no. 2, pp. 98–102, 2021, doi: 10.46880/jmika.vol4no2.pp98-102.
- [7] A. Rahmayanti, L. Rusdiana, and S. Suratno, “Perbandingan Metode Algoritma C4.5 Dan Naïve Bayes Untuk Memprediksi Kelulusan Mahasiswa,” *Walisongo J. Inf. Technol.*, vol. 4, no. 1, pp. 11–22, 2022, doi: 10.21580/wjit.2022.4.1.9654.
- [8] F. Fitriyani, “Prediksi Diabetes Menggunakan Algoritma Naive Bayes dan Greedy Forward Selection,” *J. Nas. Teknol. dan Sist. Inf.*, vol. 7, no. 2, pp. 61–69, 2021, doi: 10.25077/teknosi.v7i2.2021.61-69.
- [9] A. Nugroho and Y. Religia, “Analisis Optimasi Algoritma Klasifikasi Naive Bayes menggunakan Genetic Algorithm dan Bagging,” *J. RESTI (Rekayasa Sist. dan Teknol. Informasi)*, vol. 5, no. 3, pp. 504–510, 2021, doi: 10.29207/resti.v5i3.3067.
- [10] S. Indriyani, M. Fatchan, and A. Firmansyah, “Prediksi Harga Logam Mulia Dengan Pendekatan Algorithma Naïve Bayes Dan Pso,” *J. Inform. Teknol. dan Sains*, vol. 5, no. 1, pp. 179–182, 2023, doi: 10.51401/jinteks.v5i1.2230.