

PENERAPAN ALGORITMA GRADIENT BOOSTING UNTUK PREDIKSI RISIKO KETERLAMBATAN PEMBAYARAN SISWA DIGIKIDZ SOLO MENGGUNAKAN METODE CRISP-DM

Syakira Dhiya Ananda, Intan Oktaviani, Anindhiasi Ayu Kusuma

Teknik Informatika, Universitas Duta Bangsa Surakarta

Jalan Bhayangkara No.55, Tipes, Kec. Serengan, Kota Surakarta, Indonesia

220103138@mhs.udb.ac.id

ABSTRAK

Keterlambatan pembayaran biaya pendidikan merupakan masalah yang kerap dihadapi oleh lembaga pendidikan karena dapat memengaruhi kelancaran pengelolaan keuangan. Proses penagihan yang masih dilakukan secara manual sering kali menyulitkan pihak administrasi dalam mengidentifikasi siswa yang berpotensi mengalami keterlambatan pembayaran. Oleh karena itu, penelitian ini bertujuan untuk mengembangkan model prediksi risiko keterlambatan pembayaran siswa guna mendukung proses pengambilan keputusan yang lebih efektif. Penelitian dilakukan dengan menerapkan algoritma Gradient Boosting menggunakan pendekatan CRISP-DM (*Cross Industry Standard Process for Data Mining*). Data yang digunakan berupa riwayat pembayaran siswa yang mencakup informasi biaya pembayaran, metode pembayaran, jenis program, dan riwayat keterlambatan pembayaran. Tahapan penelitian meliputi preprocessing data, pembangunan model menggunakan XGBoost, serta evaluasi kinerja model menggunakan metrik accuracy, precision, recall, F1-score, dan Area Under Curve (AUC). Hasil pengujian menunjukkan bahwa model menghasilkan nilai accuracy sebesar 0,71 dan AUC sebesar 0,80, yang menunjukkan kemampuan cukup baik dalam memprediksi kemungkinan keterlambatan pembayaran siswa. Model yang dihasilkan diharapkan dapat membantu lembaga pendidikan dalam melakukan deteksi dini terhadap siswa yang berisiko mengalami keterlambatan pembayaran sehingga proses penagihan dapat dilakukan secara lebih tepat dan efisien.

Kata kunci: Gradient Boosting, XGBoost, CRISP-DM, Machine Learning, Keterlambatan Pembayaran

1. PENDAHULUAN

Pembayaran biaya pendidikan merupakan salah satu komponen penting yang mendukung keberlangsungan operasional lembaga pendidikan. Kelancaran pembayaran oleh siswa berperan dalam menjaga stabilitas keuangan serta mendukung pelaksanaan berbagai kegiatan akademik dan administrasi. Dengan meningkatnya jumlah siswa dan kompleksitas pengelolaan data keuangan, lembaga pendidikan memerlukan sistem yang mampu mendukung proses pengelolaan pembayaran secara efektif dan terstruktur.

Meskipun demikian, keterlambatan pembayaran biaya pendidikan masih menjadi permasalahan yang sering terjadi. Kondisi ini dapat berdampak pada terganggunya arus kas lembaga, meningkatnya beban administrasi dalam proses penagihan, serta menurunkan efisiensi pengelolaan keuangan. Selain itu, identifikasi terhadap siswa yang berpotensi mengalami keterlambatan pembayaran umumnya masih dilakukan secara manual sehingga membutuhkan waktu dan sumber daya yang lebih besar. Oleh karena itu, diperlukan suatu metode yang dapat membantu memprediksi risiko keterlambatan pembayaran berdasarkan data historis yang tersedia.

Perkembangan teknologi di bidang machine learning memungkinkan pemanfaatan data historis untuk menghasilkan prediksi yang lebih akurat. Sejumlah penelitian terdahulu menunjukkan bahwa informasi riwayat pembayaran siswa dapat digunakan

sebagai dasar dalam memprediksi keterlambatan pembayaran [1].

Beberapa metode klasifikasi, seperti Naive Bayes, telah diterapkan dan memberikan hasil yang cukup baik, namun masih memiliki keterbatasan dalam menangani pola hubungan data yang kompleks [1][2].

Di sisi lain, algoritma berbasis Gradient Boosting, khususnya XGBoost, memiliki kemampuan yang baik dalam memodelkan data non-linear dan menghasilkan performa prediksi yang tinggi [3][4].

Berdasarkan permasalahan tersebut, penelitian ini bertujuan untuk mengembangkan model prediksi keterlambatan pembayaran siswa menggunakan algoritma Gradient Boosting. Untuk mencapai tujuan tersebut, penelitian dilaksanakan dengan pendekatan CRISP-DM yang mencakup tahapan Business Understanding, Data Understanding, Data Preparation, Modeling, dan Evaluation. Melalui model yang dibangun, diharapkan lembaga pendidikan dapat mengidentifikasi siswa yang berisiko mengalami keterlambatan pembayaran lebih awal sehingga proses pengambilan keputusan dan penagihan dapat dilakukan secara lebih efektif.

2. TINJAUAN PUSTAKA

Berbagai penelitian terkait prediksi keterlambatan pembayaran telah dikembangkan dengan memanfaatkan beragam metode *machine learning*. Pendekatan tersebut dilakukan dengan

memanfaatkan riwayat pembayaran guna mengenali pola tertentu yang berkaitan dengan perilaku pembayaran.

Penelitian sebelumnya menunjukkan bahwa metode Naïve Bayes dapat digunakan untuk mengidentifikasi keterlambatan pembayaran dengan performa yang cukup baik [1]. Selain itu, penelitian lain menggunakan metode Logistic Regression juga menunjukkan hasil yang cukup baik dalam klasifikasi keterlambatan pembayaran [3]. Meskipun demikian, metode klasifikasi konvensional tersebut masih memiliki kelemahan dalam mengolah data yang bersifat kompleks maupun tidak seimbang.

Metode Support Vector Machine yang dikombinasikan dengan Particle Swarm Optimization (SVM-PSO) diketahui mampu meningkatkan kemampuan model dalam mengidentifikasi keterlambatan pembayaran [5]. Hal tersebut menunjukkan bahwa proses optimasi algoritma mampu meningkatkan kemampuan model dalam mengolah data minoritas. Seiring perkembangan teknologi, metode *ensemble learning* semakin banyak digunakan karena mampu meningkatkan akurasi model. Beberapa metode *ensemble*, antara lain *Random Forest*, *XGBoost*, terbukti mampu menghasilkan performa prediksi yang lebih tinggi dibandingkan algoritma klasifikasi konvensional [6][7].

Beberapa penelitian terdahulu juga telah menerapkan algoritma C4.5 untuk memprediksi keterlambatan pembayaran [8], serta melakukan perbandingan antara *XGBoost* untuk mendapatkan model terbaik [7]. Selain itu, Metode *CRISP-DM* diterapkan sebagai pendekatan sistematis dalam pengembangan model data mining karena mampu menyediakan alur proses yang terorganisasi, mulai dari tahap pemahaman bisnis sampai evaluasi model [9].

Berdasarkan berbagai studi yang telah dilakukan, metode *ensemble learning* terutama Gradient Boosting dinilai mampu menghasilkan prediksi keterlambatan pembayaran yang lebih optimal. Meskipun demikian,

masih terdapat penelitian yang belum menerapkan pendekatan *CRISP-DM* secara menyeluruh dalam proses pengembangan model. Oleh karena itu, penelitian ini menggunakan metode Gradient Boosting berbasis *CRISP-DM* guna menghasilkan model prediksi yang lebih terstruktur serta mampu meningkatkan kualitas prediksi.

3. METODE PENELITIAN

Penelitian ini menerapkan framework *CRISP-DM* (*Cross Industry Standard Process for Data Mining*) sebagai pedoman dalam proses pengembangan model prediksi. Framework tersebut terdiri atas lima tahapan utama, yaitu Business Understanding, Data Understanding, Data Preparation, Modeling, dan Evaluation.

3.1. Business Understanding

Tahap Business dilakukan untuk memahami permasalahan yang terjadi terkait keterlambatan pembayaran siswa serta dampaknya terhadap pengelolaan keuangan lembaga pendidikan. Pada tahap ini juga dirumuskan tujuan penelitian, yaitu mengembangkan model prediksi keterlambatan pembayaran siswa dengan memanfaatkan algoritma Gradient Boosting.

3.2. Data Understanding

Tahap Data Understanding bertujuan untuk memperoleh pemahaman mengenai karakteristik data yang digunakan dalam penelitian. Dataset yang digunakan merupakan data historis pembayaran siswa Digikidz Solo yang terdiri dari 1.000 data transaksi. Data tersebut memuat beberapa atribut, seperti *id_transaksi*, *id_siswa*, *kelas*, *jenis_program*, *biaya_pembayaran*, *metode_pembayaran*, *riwayat_telat_sebelumnya*, *hari_terlambat*, dan *status_telat*. Tabel 1 menunjukkan contoh data awal sebelum dilakukan proses preprocessing. Data tersebut digunakan sebagai dasar proses analisis dan pembangunan model prediksi keterlambatan pembayaran siswa.

Tabel 1. Contoh Data Awal Sebelum Preprocessing

Id transaksi	Id siswa	Kelas	Jenis program	Biaya pembayaran	Metode pembayaran	Riwayat telat Sebelumnya	Hari terlambat	Status telat
INV20001	DGK001	Beginner	Robotik Reguler	550000	Cash	0	0	0
INV20002	DGK001	Beginner	Robotik Reguler	550000	Cash	0	0	0
INV20003	DGK001	Beginner	Robotik Reguler	550000	Cash	0	9	1
INV20004	DGK001	Beginner	Robotik Reguler	550000	Cash	1	7	1
INV20005	DGK001	Beginner	Robotik Reguler	550000	Cash	2	11	1

Selanjutnya dilakukan proses eksplorasi data untuk mengidentifikasi tipe data, mengetahui distribusi data, serta menemukan potensi permasalahan yang dapat memengaruhi proses

pemodelan. Hasil eksplorasi ini menjadi dasar dalam tahap persiapan data sebelum dilakukan pembangunan model menggunakan algoritma *XGBoost*.

3.3. Data Preparation

Tahap Data Preparation dilakukan untuk mempersiapkan dataset sebelum digunakan pada proses pemodelan. Tahapan yang dilakukan meliputi pembersihan data (*data cleaning*) dengan menghapus data duplikat dan memastikan tidak terdapat nilai kosong (*missing value*). Selanjutnya dilakukan seleksi fitur dengan menghapus atribut yang berpotensi menyebabkan *data leakage*, kemudian atribut kategorikal dikonversi ke bentuk numerik menggunakan metode *One-Hot Encoding*. Setelah proses preprocessing selesai, dataset dibagi menjadi data training sebesar 80% dan data testing sebesar 20%.

3.4. Modeling

Pada tahap Modeling dilakukan pembangunan model prediksi menggunakan algoritma Gradient Boosting melalui implementasi XGBoost. Model dilatih menggunakan data training yang telah melalui proses preprocessing untuk mempelajari pola keterlambatan pembayaran siswa. Parameter yang digunakan dalam penelitian ini meliputi *n_estimators* = 100, *learning_rate* = 0,05, dan *max_depth* = 3.

3.5. Evaluation

Tahap Evaluation dilakukan untuk menilai kinerja model yang telah dibangun. Pengukuran performa dilakukan menggunakan metrik *accuracy*, *precision*, *recall*, *F1-score*, dan *Area Under Curve* (AUC). Selain itu, *confusion matrix* digunakan untuk mengetahui jumlah prediksi yang benar maupun salah pada setiap kelas. Hasil evaluasi digunakan untuk mengetahui kemampuan model dalam memprediksi risiko keterlambatan pembayaran siswa.

4. HASIL DAN PEMBAHASAN

4.1. Hasil Penelitian

Penelitian ini menerapkan algoritma Gradient Boosting melalui implementasi XGBoost untuk memprediksi risiko keterlambatan pembayaran siswa di Digikidz Solo. Proses pengembangan model dilakukan berdasarkan tahapan CRISP-DM yang meliputi Business Understanding, Data Understanding, Data Preparation, Modeling, dan Evaluation. Hasil yang diperoleh dari setiap tahapan dijelaskan sebagai berikut.

4.2. Business Understanding

Tahap Business Understanding dilakukan untuk memahami permasalahan bisnis yang dihadapi oleh Digikidz Solo terkait keterlambatan pembayaran biaya pendidikan siswa. Berdasarkan hasil identifikasi, keterlambatan pembayaran dapat menyebabkan terganggunya arus kas lembaga, meningkatkan beban administrasi dalam proses penagihan, serta mengurangi efisiensi pengelolaan keuangan.

Permasalahan tersebut menunjukkan perlunya suatu sistem yang mampu membantu pihak administrasi dalam mengidentifikasi siswa yang

berpotensi mengalami keterlambatan pembayaran sebelum jatuh tempo. Dengan adanya informasi tersebut, pihak lembaga dapat melakukan tindakan preventif, seperti pemberian pengingat pembayaran atau penyusunan strategi penagihan yang lebih efektif.

Berdasarkan kebutuhan tersebut, penelitian ini bertujuan membangun model prediksi risiko keterlambatan pembayaran siswa menggunakan algoritma Gradient Boosting melalui implementasi XGBoost. Model yang dihasilkan diharapkan mampu memberikan informasi mengenai kemungkinan keterlambatan pembayaran siswa berdasarkan data historis pembayaran yang tersedia sehingga dapat mendukung proses pengambilan keputusan di Digikidz Solo.

4.3. Data Understanding

Dataset yang digunakan dalam penelitian ini merupakan data historis pembayaran siswa Digikidz Solo yang terdiri atas 1.000 transaksi pembayaran. Dataset memuat 14 atribut yang mencakup informasi identitas transaksi, program pembelajaran, metode pembayaran, biaya pembayaran, riwayat keterlambatan, dan status keterlambatan pembayaran siswa.

Berdasarkan hasil eksplorasi data, diketahui bahwa dataset memiliki 1.000 baris data dan 14 atribut yang terdiri atas 5 atribut numerik dan 9 atribut kategorikal. Hasil pemeriksaan juga menunjukkan bahwa tidak terdapat nilai kosong (*missing value*) maupun data yang hilang pada seluruh atribut sehingga seluruh data dapat digunakan dalam proses analisis dan pemodelan.

Tabel 2. Karakteristik Dataset

Keterangan	Nilai
Jumlah Data	1000
Jumlah Atribut	14
Atribut Numerik	5
Atribut Kategorikal	9
Missing Value	0

Hasil eksplorasi data menunjukkan bahwa atribut yang digunakan memiliki karakteristik yang beragam, baik dalam bentuk numerik maupun kategorikal. Beberapa atribut seperti kelas, jenis_program, diskon, dan metode_pembayaran masih berbentuk kategorikal sehingga memerlukan proses transformasi ke bentuk numerik sebelum digunakan pada tahap pemodelan. Selain itu, atribut status_telat digunakan sebagai variable target yang akan diprediksi oleh model XGBoost.

4.4. Data Preparation

Tahap Data Preparation dilakukan untuk menghasilkan dataset yang siap digunakan dalam proses pemodelan. Langkah pertama yang dilakukan pembersihan data (*data cleaning*) dengan menghapus data duplikat dan memastikan tidak terdapat data yang hilang (*missing value*). Berdasarkan hasil

pemeriksaan, tidak ditemukan data kosong sehingga seluruh data dapat digunakan dalam proses analisis.

Selanjutnya dilakukan seleksi fitur dengan menghapus atribut yang berpotensi menyebabkan *data leakage*, yaitu atribut yang secara langsung berkaitan dengan status keterlambatan pembayaran. Atribut yang dihapus meliputi *id_transaksi*, *id_siswa*, *tanggal_tagihan*, *tanggal_jatuh_tempo*, *tanggal_bayar*, *hari_penyelesaian*, dan *hari_terlambat*.

Tabel 3. Atribut yang Digunakan Setelah Seleksi Fitur

Atribut
kelas
jenis_program
biaya_pembayaran
diskon
metode_pembayaran
riwayat_telat_sebelumnya
status_telat

Karena beberapa atribut masih berbentuk kategorikal, dilakukan proses *One-Hot Encoding* untuk mengubah data menjadi format numerik agar dapat diproses oleh algoritma XGBoost.

Tabel 4. Contoh Hasil Encoding Data

biaya_pembayaran	Riwayat telat sebelumnya	Kelas beginner	Kelas intermediate	Metode pembayaran transfer
550000	0	1	0	0
550000	1	1	0	0
550000	2	1	0	0

Setelah proses preprocessing selesai dilakukan, dataset dibagi menjadi data pelatihan (*training data*) dan data pengujian (*testing data*) dengan perbandingan 80:20.

Tabel 5. Pembagian Dataset

Dataset	Jumlah data
Data training	800
Data testing	200
Total	1000

4.5. Modeling

Tahap pemodelan dilakukan menggunakan algoritma XGBoost yang merupakan salah satu implementasi dari metode Gradient Boosting. Algoritma ini bekerja dengan sejumlah pohon keputusan (*decision tree*) secara bertahap, di mana setiap pohon baru digunakan untuk memperbaiki kesalahan prediksi yang dihasilkan oleh pohon sebelumnya. Melalui mekanisme tersebut, model dapat menghasilkan tingkat prediksi yang lebih baik dibandingkan penggunaan satu pohon keputusan tunggal.

Pada penelitian ini, model XGBoost dikembangkan menggunakan parameter *n_estimators* = 100, *learning_rate* = 0,05, dan *max_depth* = 3. Parameter tersebut digunakan untuk mengatur jumlah

pohon yang dibangun, laju pembelajaran model, serta kedalaman maksimum setiap pohon keputusan.

Persamaan dasar Gradient Boosting dapat dituliskan sebagai berikut:

$$F_m(x) = F_{m-1}(x) + \eta h_m(x)$$

Keterangan:

$F_m(x)$: model pada iterasi ke- m

$F_{m-1}(x)$: model pada iterasi sebelumnya

η : learning rate

$h_m(x)$: pohon keputusan yang ditambahkan pada iterasi ke- m .

Setelah proses preprocessing selesai dilakukan, model dilatih menggunakan 800 data training dan diuji menggunakan 200 data testing. Proses pelatihan bertujuan untuk mempelajari pola hubungan antara atribut yang digunakan, seperti kelas, jenis program, biaya pembayaran, metode pembayaran, dan riwayat keterlambatan sebelumnya terhadap status keterlambatan pembayaran siswa.

4.6. Evaluation

Tahap evaluasi dilakukan untuk mengukur kinerja model XGBoost dalam memprediksi risiko keterlambatan pembayaran siswa. Pengukuran performa dilakukan menggunakan *confusion matrix* serta beberapa metrik evaluasi, yaitu *accuracy*, *precision*, *recall*, *F1-score*, dan *Area Under Curve* (AUC).

Tabel 6. Confusion Matrix Hasil Prediksi XGBoost

	Prediksi Tidak Telat	Prediksi Telat
Aktual Tidak Telat	96	25
Aktual Telat	32	47

Berdasarkan hasil pengujian menggunakan 200 data testing, diperoleh nilai pada *confusion matrix* sebagai berikut:

True negative (TN) = 96

False Positive (FP) = 25

False Negative (FN) = 32

True Positive (TP) = 47

Nilai accuracy dihitung menggunakan persamaan berikut:

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN}$$

Dengan memasukkan nilai confusion matrix diperoleh:

$$Accuracy = \frac{47 + 96}{47 + 96 + 25 + 32}$$

$$Accuracy = \frac{143}{200}$$

$$Accuracy = 0,715 \text{ atau } 71,5\%.$$

Nilai accuracy sebesar 71,5% menunjukkan bahwa model mampu mengklasifikasikan 71,5% data pengujian dengan benar.

Tabel 7. Hasil Evaluasi Model XGBoost

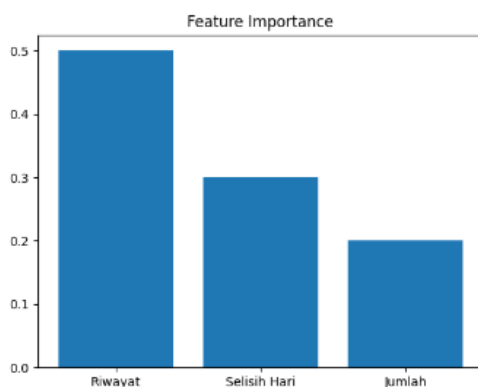
Metrik	Nilai
Accuracy	0,715
Precision	0,653
Recall	0,595
F1-Score	0,623
AUC	0,801

Berdasarkan hasil evaluasi, model XGBoost memperoleh nilai *accuracy* sebesar 0,715 yang menunjukkan bahwa model mampu mengklasifikasikan data dengan tingkat ketepatan sebesar 71,5%. Nilai *precision* sebesar 0,653 menunjukkan bahwa 65,3% data yang diprediksi mengalami keterlambatan benar-benar termasuk dalam kategori terlambat. Nilai *recall* sebesar 0,595 menunjukkan bahwa model mampu mengidentifikasi 59,5% dari seluruh kasus keterlambatan pembayaran yang terjadi.

Selain itu, nilai *F1-score* sebesar 0,623 menunjukkan keseimbangan yang cukup baik antara *precision* dan *recall*. Sementara itu, nilai *Area Under Curve* (AUC) sebesar 0,801 menunjukkan bahwa model memiliki kemampuan yang baik dalam membedakan siswa yang berisiko mengalami keterlambatan pembayaran dengan siswa yang melakukan pembayaran tepat waktu.

4.7. Analisis Feature Importance

Analisis *feature importance* dilakukan untuk mengetahui tingkat kontribusi masing-masing atribut dalam proses prediksi keterlambatan pembayaran siswa. Fitur ini merupakan salah satu keunggulan algoritma XGBoost yang dapat menunjukkan atribut mana yang paling berpengaruh dalam pembentukan model.



Gambar 1. Feature Importance Model XGBoost

Berdasarkan hasil analisis *feature importance*, diketahui bahwa setiap atribut memiliki tingkat kontribusi yang berbeda dalam proses prediksi. Atribut dengan nilai *importance* yang lebih tinggi menunjukkan pengaruh yang lebih besar terhadap keputusan model dalam menentukan status keterlambatan pembayaran siswa.

Hasil analisis menunjukkan bahwa atribut *riwayat_telat_sebelumnya* memiliki nilai *importance* tertinggi dibandingkan atribut lainnya. Temuan ini mengindikasikan bahwa riwayat keterlambatan pembayaran pada periode sebelumnya merupakan faktor yang paling berpengaruh dalam memprediksi kemungkinan keterlambatan pembayaran pada periode berikutnya. Dengan kata lain, siswa yang pernah mengalami keterlambatan pembayaran memiliki kecenderungan lebih tinggi untuk kembali mengalami keterlambatan pada periode selanjutnya.

Selain atribut tersebut, atribut *biaya_pembayaran*, *metode_pembayaran*, dan *jenis_program* juga memberikan kontribusi terhadap proses prediksi, meskipun tingkat pengaruhnya relatif lebih rendah. Hal ini menunjukkan bahwa risiko keterlambatan pembayaran tidak hanya dipengaruhi oleh riwayat pembayaran sebelumnya, tetapi juga dipengaruhi oleh karakteristik transaksi dan program pembelajaran yang diikuti siswa.

Hasil analisis *feature importance* memberikan informasi yang bermanfaat bagi pihak Digikidz Solo dalam memahami faktor-faktor yang memengaruhi keterlambatan pembayaran siswa. Informasi tersebut dapat digunakan sebagai dasar dalam penyusunan strategi penagihan, pemberian pengingat pembayaran, maupun pengambilan kebijakan yang lebih tepat sasaran untuk meminimalkan risiko keterlambatan pembayaran.

Tabel 8. Peringkat Feature Importance

Peringkat	Atribut
1	riwayat_telat_sebelumnya
2	biaya_pembayaran
3	metode_pembayaran
4	jenis_program
5	kelas

4.8. Pembahasan

Hasil penelitian menunjukkan bahwa algoritma XGBoost yang merupakan salah satu metode Gradient Boosting mampu digunakan untuk memprediksi risiko keterlambatan pembayaran siswa di Digikidz Solo. Berdasarkan hasil evaluasi, model memperoleh nilai *accuracy* sebesar 71,5%, *precision* sebesar 65,3%, *recall* sebesar 59,5%, *F1-score* sebesar 62,3%, dan *AUC* sebesar 0,801.

Nilai *accuracy* sebesar 71,5% menunjukkan bahwa model mampu mengklasifikasikan sebagian besar data pengujian dengan benar. Sementara itu, nilai *precision* sebesar 65,3% mengindikasikan bahwa sebagian besar siswa yang diprediksi berisiko mengalami keterlambatan memang termasuk ke dalam kategori terlambat. Nilai *recall* sebesar 59,5% menunjukkan bahwa model mampu mengidentifikasi lebih dari setengah jumlah siswa yang benar-benar mengalami keterlambatan pembayaran.

Selain itu, nilai *F1-score* sebesar 62,3% menunjukkan keseimbangan yang cukup baik antara *precision* dan *recall*. Nilai *AUC* sebesar 0,801

mengindikasikan bahwa model memiliki kemampuan yang baik dalam membedakan siswa yang berisiko mengalami keterlambatan pembayaran dengan siswa yang melakukan pembayaran tepat waktu.

Hasil tersebut menunjukkan bahwa penerapan algoritma Gradient Boosting melalui XGBoost dapat dimanfaatkan untuk membantu proses identifikasi risiko keterlambatan pembayaran siswa. Dengan adanya informasi prediksi yang dihasilkan model, pihak Digikidz Solo dapat melakukan tindakan antisipatif lebih awal sehingga proses pengelolaan pembayaran dan penagihan dapat dilakukan secara lebih efektif.

4.9. Analisis

Meskipun model berhasil mencapai nilai accuracy sebesar 71,5% dan AUC sebesar 0,801, masih terdapat sejumlah kesalahan prediksi yang terlihat pada confusion matrix. Dari 79 siswa yang benar-benar mengalami keterlambatan pembayaran, model berhasil mengidentifikasi 47 siswa, sedangkan 32 siswa lainnya masih diklasifikasikan sebagai tidak terlambat. Kondisi ini menyebabkan nilai recall yang diperoleh belum optimal.

Salah satu faktor yang memengaruhi hasil tersebut adalah keterbatasan variabel yang digunakan dalam penelitian. Model hanya memanfaatkan informasi historis pembayaran dan karakteristik transaksi yang tersedia dalam dataset. Faktor lain yang kemungkinan berpengaruh terhadap keterlambatan pembayaran, seperti kondisi ekonomi orang tua, tingkat kehadiran siswa, maupun faktor eksternal lainnya belum tersedia sehingga belum dapat dianalisis oleh model.

Selain itu, proses pemodelan belum menerapkan optimasi parameter secara mendalam. Penggunaan teknik hyperparameter tuning seperti Grid Search atau Random Search berpotensi meningkatkan performa model sehingga menghasilkan prediksi yang lebih akurat.

Hasil analisis feature importance juga menunjukkan bahwa atribut riwayat_telat_sebelumnya merupakan faktor yang paling berpengaruh dalam proses prediksi. Temuan ini menunjukkan bahwa siswa yang pernah mengalami keterlambatan pembayaran pada periode sebelumnya cenderung memiliki risiko lebih tinggi untuk mengalami keterlambatan kembali pada periode berikutnya.

Secara keseluruhan, hasil penelitian menunjukkan bahwa algoritma XGBoost mampu mempelajari pola keterlambatan pembayaran siswa dengan cukup baik dan memiliki potensi untuk diterapkan sebagai model prediksi pada lingkungan lembaga pendidikan.

4.10. Implikasi Penelitian

Hasil penelitian ini memberikan manfaat praktis bagi Digikidz Solo dalam mendukung pengelolaan

administrasi pembayaran siswa. Model prediksi yang dihasilkan dapat digunakan sebagai sistem pendukung keputusan untuk membantu pihak administrasi dalam mengidentifikasi siswa yang memiliki risiko keterlambatan pembayaran berdasarkan data historis yang tersedia.

Selain itu, model dapat dimanfaatkan sebagai alat deteksi dini terhadap potensi keterlambatan pembayaran sehingga pihak lembaga dapat melakukan tindakan pencegahan sebelum keterlambatan terjadi. Langkah tersebut dapat berupa pemberian pengingat pembayaran, komunikasi kepada orang tua siswa, maupun penyusunan strategi penagihan yang lebih efektif.

Lebih lanjut, hasil prediksi yang dihasilkan model dapat digunakan sebagai dasar dalam penyusunan kebijakan pengelolaan keuangan lembaga pendidikan. Dengan mengetahui tingkat risiko keterlambatan pembayaran siswa, pihak Digikidz Solo dapat merencanakan langkah antisipatif yang lebih tepat guna menjaga stabilitas arus kas dan meningkatkan efisiensi proses administrasi pembayaran.

5. KESIMPULAN DAN SARAN

Berdasarkan hasil penelitian yang telah dilakukan, algoritma Gradient Boosting melalui implementasi XGBoost berhasil diterapkan untuk memprediksi risiko keterlambatan pembayaran siswa di Digikidz Solo dengan menggunakan pendekatan CRISP-DM. Proses pengembangan model dilakukan melalui tahapan Business Understanding, Data Understanding, Data Preparation, Modeling, dan Evaluation menggunakan data historis pembayaran siswa sebanyak 1.000 transaksi.

Hasil pengujian menunjukkan bahwa model memperoleh nilai accuracy sebesar 71,5%, precision sebesar 65,3%, recall sebesar 59,5%, F1-score sebesar 62,3%, dan Area Under Curve (AUC) sebesar 0,801. Nilai tersebut menunjukkan bahwa model memiliki kemampuan yang cukup baik dalam mengklasifikasikan siswa yang berpotensi mengalami keterlambatan pembayaran maupun siswa yang melakukan pembayaran tepat waktu. Selain itu, hasil analisis feature importance menunjukkan bahwa atribut riwayat_telat_sebelumnya merupakan faktor yang paling berpengaruh dalam proses prediksi keterlambatan pembayaran siswa.

Berdasarkan hasil yang diperoleh, model yang dikembangkan dapat dimanfaatkan sebagai alat bantu untuk mengidentifikasi risiko keterlambatan pembayaran siswa sehingga dapat mendukung proses pengambilan keputusan dan pengelolaan administrasi pembayaran di Digikidz Solo.

Untuk penelitian selanjutnya, disarankan menggunakan dataset yang lebih besar dan lebih beragam agar model dapat mempelajari pola keterlambatan pembayaran secara lebih optimal. Selain itu, penambahan variabel yang relevan, penerapan teknik penanganan data tidak seimbang,

serta optimasi parameter melalui metode hyperparameter tuning dapat dilakukan untuk meningkatkan performa prediksi. Model yang dihasilkan juga dapat dikembangkan lebih lanjut dalam bentuk aplikasi berbasis web atau sistem pendukung keputusan sehingga dapat dimanfaatkan secara langsung oleh pihak Digikidz Solo.

DAFTAR PUSTAKA

- [1] M. Khoil, H. Oktavianto, L. A. Muharom, and N. Qodariyah, "Prediksi Keterlambatan Pembayaran Sumbangan Pembinaan Pendidikan Sekolah Menggunakan Metode Naive Bayes," vol. 1, pp. 109–118, 2025, doi: 10.63232/jstem.v1i2.81.
- [2] T. Azhima and Y. Siswa, "Prediction of Late Tuition Fees at Muhammadiyah University Kalimantan Timur Using the Logistic Regression Method," vol. 1, no. 1, pp. 30–34, 2023, doi: 10.30650/jse.v1i1.3435.
- [3] S. H. Nugraha, M. Sholeh, E. Roesminingsih, K. Amalia, and A. Khamidi, "Pemanfaatan Teknologi Informasi dalam Manajemen Keuangan Sekolah : Sebuah Kajian Literatur," *J. Ilm. ilmu Pendidik.*, vol. 8, no. 6, pp. 6732–6741, 2025, doi: 10.54371/jiip.v8i6.8325.
- [4] O. Calvin, A. Novrandika, M. A. Herlambang, Y. A. Prasetyo, and P. Studi, "Penerapan Teknologi Internet of Things Untuk Meningkatkan Efisiensi Pencarian Tempat Parkir Mobil Di Lingkungan Pusat Perbelanjaan," *Tek. dan Multimed.*, vol. 2, no. 1, pp. 38–34, 2024, [Online]. Available: <https://journal.mediapublikasi.id/index.php/Biner>
- [5] H. Nurdin, I. Nawawi, A. Wuryanto, D. Yuliandari, and H. Sugiarto, "Prediksi Keterlambatan Pembayaran Mahasiswa untuk Mitigasi Risiko Cuti Menggunakan SVM Optimasi PSO," vol. 7, no. 1, 2025, doi: 10.26905/jasiek.v7i1.15483.
- [6] V. Wijaya, V. Joeanca, H. Yap, and S. Lim, "Methodical Approach : Building a Web-Based Warehouse Management System Using the Waterfall Method," *Int. J. Comput. Sci. infomation Technol.*, vol. 1, no. 1, pp. 8–19, 2024, doi: 10.55123/ijisit.
- [7] P. S. Rizky *et al.*, "Perbandingan Metode LightGBM dan XGBoost dalam Menangani Data dengan Kelas Tidak Seimbang Universitas Hamzanwadi Selong," *J. Stat.*, vol. 15, no. 2, pp. 228–236, 2022, doi: 10.36456/jstat.vol15.no2.a5548.
- [8] S. Kasus, U. Muhammadiyah, R. R. Rahmah, I. R. Setiawan, and F. F. Az-zahra, "Prediksi Keterlambatan Mahasiswa dalam Membayar Biaya Kuliah Menggunakan Algoritma C4.5," *J. Mhs. Tek. Inform.*, vol. 8, no. 5, pp. 10792–10800, 2024, doi: 10.36040/jati.v8i5.10979.
- [9] R. Sahara, S. Abdullah, C. R. Hassolthine, and M. I. Saputra, "Employee Credit Classification Analysis Using Decision Tree Based CRISP-DM Model (Case Study of Samsung Indonesia Company)," *J. Inform. Rekayasa Elektron.*, vol. 7, no. 2, pp. 293–300, 2024, [Online]. Available: <http://e-journal.stmiklombok.ac.id/index.php/jire%0AVolume>
- [10] H. Umar *et al.*, "Klasifikasi Keterlambatan Pembayaran Sumbangan Pembinaan Pendidikan Menggunakan Algoritma Naïve Bayes dan Support Vector Machine TIN: Terapan Informatika Nusantara," *Terap. Inform. Nusantara*, vol. 4, no. 11, pp. 709–718, 2024, doi: 10.47065/tin.v4i11.4969.