BAB III
METODE PENELITIAN

A. Jenis Penelitian

Penelitian ini menggunakan pendekatan kuantitatif dengan memanfaatkan teknik data mining dan kerangka kerja CRISP-DM. Pendekatan kuantitatif dipilih karena penelitian ini memanfaatkan data transaksi peminjaman buku yang diolah dan dianalisis secara sistematis untuk menghasilkan klasifikasi keterlambatan pengembalian. Selanjutnya, performa model dinilai berdasarkan hasil pengukuran menggunakan sejumlah metrik evaluasi.

Metode klasifikasi yang digunakan dalam penelitian ini adalah *machine learning* dengan dua algoritma, yaitu Decision Tree dan Naïve Bayes. Kedua teknik ini diterapkan untuk menyusun model klasifikasi berdasarkan data historis transaksi peminjaman buku pada tahun 2024 di Perpustakaan UIN Imam Bonjol Padang. Langkah-langkah penelitian mencakup pengolahan informasi, pembangunan model, pengujian, serta analisis perbandingan hasil dari setiap algoritma dalam mengklasifikasikan keterlambatan pengembalian buku.

Ditinjau dari tujuannya, penelitian ini termasuk penelitian komparatif karena berfokus pada analisis perbedaan kinerja antara Decision Tree dan Naïve Bayes. Perbandingan dilakukan untuk mengidentifikasi algoritma yang menghasilkan klasifikasi paling optimal pada kumpulan data yang diterapkan. Penilaian kinerja kedua model dilakukan menggunakan *confusion matrix* dengan metrik *accuracy*, *precision*, *recall*, dan F1-score.

UIN IMAM BONJOL PADANG

B. Setting Penelitian

1. Lokasi Penelitian

Penelitian ini dilaksanakan di Perpustakaan UIN Imam Bonjol Padang sebagai tempat penelitian dan sekaligus sebagai sumber data utama. Perpustakaan ini dipilih karena memiliki sistem pencatatan transaksi sirkulasi yang terstruktur dan terdokumentasi dengan baik, sehingga memungkinkan proses pengambilan serta pengolahan data dilakukan secara sistematis. Selain itu, tingginya aktivitas

peminjaman buku menjadikan perpustakaan ini relevan untuk dianalisis dalam konteks klasifikasi pola keterlambatan pengembalian buku.

Data yang digunakan dalam penelitian ini merupakan data historis mengenai transaksi peminjaman buku pada tahun 2024 dengan jumlah keseluruhan sebanyak 13.584 data. Data tersebut diperoleh dari sistem administrasi/sirkulasi perpustakaan dan digunakan sebagai landasan dalam pengembangan serta evaluasi model klasifikasi dengan memanfaatkan algoritma Decision Tree dan Naïve Bayes.

2. Jadwal Penelitian

Penelitian ini dilaksanakan selama sekitar empat bulan dengan kegiatan yang diorganisir berdasarkan alokasi waktu setiap minggunya. Penyusunan jadwal dilakukan sebagai pedoman agar seluruh rangkaian penelitian, mulai dari tahap awal hingga penyelesaian, dapat berlangsung secara sistematis dan sesuai dengan waktu yang telah ditentukan. Rincian pelaksanaan kegiatan penelitian disajikan pada Tabel 3.1 berikut:

Tabel 3. 1 Jadwal Penelitian

Kegiatan	**Bulan 1**				**Bulan 2**				**Bulan 3**				**Bulan 4**				**Bulan 5**			
	1	2	3	4	1	2	3	4	1	2	3	4	1	2	3	4	1	2	3	4
Identifikasi Masalah																				
Studi Literatur																				
Pengumpulan Data																				
Data *Understanding*																				
Data Preparation *(Preprocessing)*																				
Penerapan Algoritma Decision Tree & Naïve Bayes																				
Evaluasi Model *(Confusion Matrix)*																				

UNIVERSITAS IMAM BONJOL PADANG

Deployment Model (Streamlit)																			
Penyusunan Laporan																			

C. Bahan/Materi Penelitian

Bahan yang digunakan dalam penelitian ini mencakup informasi dan data yang penting sebagai referensi dalam pemrosesan dan analisis untuk mendukung tujuan penelitian. Data yang dipakai berasal dari transaksi peminjaman buku yang didapatkan dari Perpustakaan UIN Imam Bonjol Padang melalui pengelola layanan sirkulasi perpustakaan. Data tersebut kemudian digunakan sebagai dataset dalam proses klasifikasi pola keterlambatan dalam mengembalikan buku. Selanjutnya, bahan dan materi yang terlibat dalam penelitian ini meliputi:

1. Data historis peminjaman buku tahun 2024 sebanyak 13.584 data, yang mencakup angkatan, hari, bulan, kategori buku, lama izin peminjaman.
2. Buku pelajaran dan artikel akademis yang relevan dengan data mining, *machine learning*, algoritma klasifikasi Decision Tree dan Naïve Bayes, serta metode evaluasi model seperti *confusion matrix*.

Sebelum memasuki tahap klasifikasi, dataset terlebih dahulu dipersiapkan melalui proses *preprocessing*. Tahapan tersebut mencakup pembersihan (data cleaning), cek duplikat data, seleksi data, penyesuaian atau transformasi atribut, serta pemberian label berdasarkan status pengembalian buku, yaitu tepat waktu dan terlambat. Dataset yang telah disiapkan selanjutnya dimanfaatkan sebagai input untuk mengembangkan model dengan menggunakan algoritma Decision Tree serta Naïve Bayes. Keduanya digunakan untuk mengkategorikan pola keterlambatan pengembalian buku berdasarkan tujuan penelitian.

D. Alat/Instrumen Penelitian

Alat penelitian digunakan untuk mendukung pelaksanaan setiap tahapan penelitian, khususnya dalam kegiatan pengolahan, pemodelan, dan analisis data. Penggunaan instrumen tersebut disesuaikan dengan kebutuhan penelitian agar proses penerapan metode data mining dapat berjalan secara terarah. Instrumen yang

UIN IMAM BONJOL PADANG

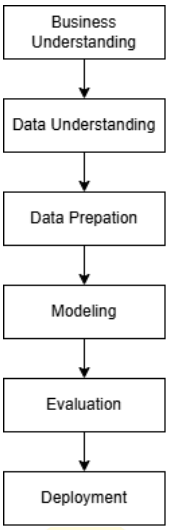
digunakan yaitu perangkat keras (*hardware*) dan perangkat lunak (*software*) sebagai sarana pendukung selama penelitian berlangsung.

Adapun alat atau instrumen yang digunakan dalam penelitian ini adalah sebagai berikut:

1. **Perangkat Keras (*Hardware)***
   a) Laptop Asus X415DAP_M415DA
   b) Processor minimal Intel Core i3 atau setara
   c) RAM minimal 8 GB
   d) Penyimpanan HDD/SSD ≥ 256 GB
   e) Mouse
2. **Perangkat Lunak *(Software)***
   a) Sistem Operasi Windows 10/11 (64-bit)
   b) Microsoft Excel, untuk proses pra-pemrosesan data *(preprocessing),* pembersihan data, serta rekapitulasi dataset peminjaman buku
   c) Google Colab dimanfaatkan untuk tahapan pra-pemrosesan data, pengembangan model klasifikasi melalui algoritma Decision Tree dan Naïve Bayes, serta penilaian kinerja model.
   d) Python, untuk proses klasifikasi keterlambatan pengembalian buku dengan memanfaatkan algoritma Decision Tree dan Naïve Bayes
   e) Library Python (Pandas, NumPy, Scikit-learn, Matplotlib), untuk pengolahan data, pembangunan model, dan evaluasi menggunakan *confusion matrix*.
   f) Microsoft Word, untuk penyusunan laporan proposal dan skripsi

UIN IMAM BONJOL PADANG

E. Alur Penelitian



Gambar 3. 1 Alur Penelitian

F. Teknik Pengumpulan Data/Prosedur Penelitian

1. Teknik Pengumpulan Data

Data penelitian diperoleh melalui studi dokumentasi. Teknik ini dilakukan dengan memanfaatkan dokumen dan arsip transaksi peminjaman yang tersimpan pada sistem sirkulasi Perpustakaan UIN Imam Bonjol Padang. Data yang diperoleh kemudian ditelaah dan disesuaikan dengan kebutuhan penelitian untuk digunakan dalam tahap pengolahan dan analisis. Dengan demikian, proses pengumpulan data tidak melibatkan survei maupun eksperimen secara langsung di lapangan, melainkan menggunakan data dokumentasi yang telah tersedia pada sistem perpustakaan.

Melalui teknik ini, peneliti memperoleh data historis peminjaman buku tahun 2024 sebanyak 13.584 data. Dengan demikian, proses pengumpulan data yang dapat dilakukan secara efektif dan efisien sesuai dengan kebutuhan penelitian. Dataset yang diperoleh dari proses dokumentasi tersebut memuat beberapa atribut penting yang berkaitan dengan keterlambatan pengembalian buku, antara lain:

a) Angkatan
b) Hari

UIN IMAM BONJOL PADANG

c) Bulan

d) Kategori Buku

e) Lama Izin

Data yang diperoleh dari proses pengumpulan kemudian diolah untuk menghasilkan informasi yang sesuai kebutuhan penelitian. Tahap analisis dilakukan dengan menerapkan algoritma Decision Tree dan Naïve Bayes untuk mengidentifikasi serta mengelompokkan status keterlambatan pengembalian buku. Penerapan kedua algoritma tersebut dilakukan berdasarkan karakteristik data transaksi perpustakaan sehingga hasil klasifikasi dapat diperoleh secara terukur dan dapat digunakan untuk membandingkan kinerja masing-masing metode.

2. Prosedur Penelitian

Prosedur penelitian menggambarkan rangkaian kegiatan yang dilakukan peneliti secara terarah untuk mencapai tujuan penelitian. Penelitian ini menggunakan kerangka kerja CRISP-DM (*Cross Industry Standard Process for Data Mining*) sebagai pedoman dalam menjalankan semua langkahnya. Pemilihan CRISP-DM berlandaskan pada struktur kerjanya yang teratur, sehingga mampu mengarahkan penelitian yang berkaitan dengan penggalian data dari tahap penentuan masalah, memahami serta mempersiapkan data, pengembangan model, sampai dengan penilaian dan penerapan hasil yang diperoleh. Selanjutnya, urutan proses yang dilalui dalam penelitian ini dijelaskan sebagai berikut:

a. Business Understanding

Tahap *Business Understanding* dilakukan untuk mendapatkan kejelasan tentang permasalahan yang menjadi inti dari penelitian dan tujuan yang ingin diraih. Dalam penelitian ini, isu yang diteliti berkaitan dengan keterlambatan pengembalian buku di Perpustakaan UIN Imam Bonjol Padang. Data historis transaksi peminjaman yang tersedia dimanfaatkan untuk mengidentifikasi pola yang berkaitan dengan status pengembalian buku.

Berdasarkan permasalahan tersebut, penelitian diarahkan pada pembangunan model klasifikasi yang mampu membedakan transaksi berdasarkan status pengembaliannya, yaitu tepat waktu dan terlambat. Untuk

UIN IMAM BONJOL PADANG

mencapai tujuan tersebut, digunakan dua algoritma klasifikasi, yaitu Decision Tree dan Naïve Bayes. Selanjutnya, kinerja dari kedua algoritma tersebut dibandingkan untuk menentukan metode yang paling sesuai dengan karakteristik data transaksi perpustakaan serta menghasilkan informasi yang dapat digunakan sebagai pendukung pengambilan keputusan berbasis data..

b. Data Understanding

Pada tahap *data understanding,* peneliti mengumpulkan dan menelaah data historis peminjaman buku tahun 2024 sebanyak 13.584 data yang diperoleh dari sistem sirkulasi perpustakaan. Data tersebut dipelajari untuk mengetahui struktur, jenis atribut, serta karakteristik data yang digunakan. Variabel yang digunakan dalam penelitian ini meliputi angkatan, hari, bulan, kategori buku, lama izin. Selain itu, dilakukan pemeriksaan awal untuk memastikan data lengkap, konsisten, dan layak digunakan pada tahap selanjutnya.

c. Data Preparation

Tahap *data preparation* merupakan proses yang dilakukan untuk mempersiapkan dan mengolah data sebelum digunakan dalam tahap analisis dan pemodelan. Data awal yang diperoleh dari transaksi peminjaman buku di Perpustakaan UIN Imam Bonjol Padang tahun 2024 berjumlah 13.584 data. Pada tahap ini, dilakukan beberapa proses *preprocessing* serta analisis terhadap kualitas data yang meliputi beberapa tahapan berikut:

1) *Data Cleaning* (Pembersihan Data)

   Pada tahap ini dilakukan pemeriksaan terhadap kelengkapan data pada setiap atribut yang digunakan dalam penelitian.

2) Pemeriksaan Data Duplikat

   Tahap berikutnya adalah memeriksa dataset untuk mengidentifikasi adanya pencatatan transaksi yang sama secara berulang. Pemeriksaan ini dilakukan guna memastikan setiap data transaksi peminjaman tercatat secara unik dan tidak terdapat duplikasi yang dapat memengaruhi proses analisis.

UIN IMAM BONJOL PADANG

3) Transformasi dan Penyesuaian Format Data

Pada tahap ini dilakukan penyesuaian dan transformasi data agar sesuai dengan kebutuhan proses analisis dan pemodelan. Data yang berupa atribut tanggal diolah menjadi variabel turunan, yaitu bulan peminjaman, hari peminjaman, dan lama izin. Selain itu, atribut yang bersifat kategorikal, seperti angkatan, bulan, hari, dan kategori buku, ditransformasikan ke dalam format yang dapat diproses oleh algoritma klasifikasi. Pada algoritma Decision Tree, transformasi dilakukan menggunakan One-Hot Encoding, sedangkan pada algoritma Naïve Bayes digunakan Ordinal Encoding.

4) Pembentukan Variabel Target (*Labeling*)

Pembentukan variabel target dilakukan untuk menentukan status keterlambatan pengembalian buku. Penentuan label keterlambatan dilakukan dengan membandingkan tanggal pengembalian dengan tanggal jatuh tempo. Jika tanggal pengembalian lebih besar dari tanggal jatuh tempo, maka data dikategorikan sebagai "Terlambat", sedangkan jika tanggal pengembalian kurang dari atau sama dengan tanggal jatuh tempo, maka dikategorikan sebagai "Tidak Terlambat". Variabel target ini kemudian digunakan sebagai kelas dalam proses klasifikasi dengan memanfaatkan atribut angkatan, bulan peminjaman, hari peminjaman, kategori buku, dan lama izin.

5) Pembagian Data Training dan Testing

Setelah dilakukan transformasi data, tahap selanjutnya adalah pembagian data menjadi data training dan data testing. Pembagian data dilakukan menggunakan tiga skema, yaitu 70:30, 80:20, dan 90:10 dengan metode stratified split. Data training digunakan sebagai data yang akan digunakan dalam proses pembentukan model, sedangkan data testing disiapkan untuk digunakan pada tahap evaluasi model. Penggunaan stratified split bertujuan untuk mempertahankan proporsi kelas pada data training dan data testing.

UIN IMAM BONJOL PADANG

6) Analisis Distribusi Kelas (Ketidakseimbangan Data)

Setelah seluruh data memperoleh label, dilakukan pemeriksaan terhadap komposisi masing-masing kelas untuk mengetahui distribusi data pada kelas Terlambat dan Tidak Terlambat. Tahap ini penting karena ketidakseimbangan jumlah data antar kelas dapat memengaruhi proses pembelajaran dan kemampuan model dalam melakukan klasifikasi. Apabila terdapat ketidakseimbangan kelas, dilakukan penyeimbangan pada data training menggunakan metode Random Undersampling (RUS). Proses tersebut dilakukan dengan mengurangi sebagian data dari kelas mayoritas sehingga distribusi kelas pada data training menjadi lebih seimbang, sedangkan data testing tetap menggunakan data asli tanpa proses penyeimbangan.

Hasil dari seluruh proses *preprocessing* dan analisis menghasilkan dataset akhir yang telah memenuhi kebutuhan penelitian. Dataset tersebut selanjutnya digunakan sebagai data masukan pada tahap pemodelan dengan menerapkan algoritma Decision Tree dan Naïve Bayes untuk proses klasifikasi.

d. Modeling

Tahap modeling merupakan proses penerapan metode klasifikasi untuk membangun model yang dapat mengklasifikasikan status keterlambatan pengembalian buku. Dalam penelitian ini, algoritma yang digunakan terdiri dari dua metode, yaitu Decision Tree dan Naïve Bayes. Data yang telah melalui tahap Data Preparation dibagi menjadi data training dan data testing menggunakan tiga skema pembagian, yaitu 70:30, 80:20, dan 90:10 dengan metode stratified split. Data training digunakan untuk membangun dan melatih model, sedangkan data testing digunakan untuk mengukur kemampuan model dalam melakukan klasifikasi terhadap data yang tidak digunakan selama proses pelatihan.

Pada algoritma Decision Tree, pembentukan model menggunakan kriteria entropy untuk menentukan pemisahan atribut dalam membentuk struktur pohon keputusan. Sementara itu, algoritma Naïve Bayes menggunakan pendekatan probabilistik dalam menentukan kelas berdasarkan karakteristik data. Setelah

UIN IMAM BONJOL PADANG

proses tuning dan pembentukan model dilakukan, masing-masing model digunakan untuk memprediksi data testing. Hasil prediksi dari kedua algoritma pada setiap skema pembagian data selanjutnya digunakan pada tahap Evaluation untuk menghitung Accuracy, Precision, Recall, F1-Score, serta Confusion Matrix sebagai dasar perbandingan kinerja kedua algoritma.

e. Evaluation

Tahap *evaluation* bertujuan untuk menilai kinerja model klasifikasi yang telah dibangun pada tahap *modeling*. Pada tahap ini, hasil klasifikasi algoritma Decision Tree dan Naïve Bayes terhadap data uji dibandingkan dengan nilai aktual pada dataset. Proses evaluasi dilakukan menggunakan *confusion matrix* untuk mengetahui jumlah klasifikasi yang benar maupun salah pada masing-masing kelas.

Selanjutnya, hasil *confusion matrix* digunakan untuk menghitung beberapa metrik evaluasi, meliputi *accuracy*, *precision*, *recall*, dan F1-score. Metrik tersebut digunakan untuk menilai tingkat ketepatan dan kemampuan model dalam mengklasifikasikan data keterlambatan pengembalian buku. Nilai evaluasi yang dihasilkan oleh algoritma Decision Tree dan Naïve Bayes kemudian dibandingkan untuk menentukan algoritma yang memiliki kinerja terbaik dalam mengklasifikasikan pola keterlambatan pengembalian buku berdasarkan data perpustakaan yang digunakan.

UIN IMAM BONJOL
PADANG

f. Deployment

Tahap *Deployment* merupakan tahap akhir dalam kerangka kerja CRISP-DM yang berfokus pada penerapan model klasifikasi agar dapat digunakan dalam sistem. Tahap ini dilakukan setelah proses modeling dan evaluation terhadap algoritma Decision Tree dan Naïve Bayes. Model dan skema pembagian data yang akan digunakan pada tahap deployment ditentukan berdasarkan hasil evaluasi menggunakan metrik Accuracy, Precision, Recall, dan F1-Score, dengan F1-Score sebagai dasar utama dalam menentukan model yang akan diterapkan. Model yang terpilih selanjutnya akan diimplementasikan ke dalam aplikasi berbasis web menggunakan Streamlit.

Aplikasi yang dibangun akan dirancang agar pengguna dapat memasukkan data peminjaman buku berupa angkatan, hari, bulan, kategori buku, dan lama izin. Data masukan akan melalui proses transformasi sesuai dengan tahapan pengolahan data yang telah diterapkan pada model sebelum digunakan untuk melakukan klasifikasi. Selanjutnya, aplikasi akan menampilkan hasil klasifikasi status pengembalian buku ke dalam dua kategori, yaitu "Terlambat" dan "Tidak Terlambat". Tahap deployment ini bertujuan untuk menerapkan model hasil penelitian ke dalam bentuk aplikasi sehingga model dapat digunakan untuk melakukan klasifikasi berdasarkan data yang dimasukkan oleh pengguna..

G. Teknik Analisis Data

Teknik analisis data dalam penelitian ini menggunakan pendekatan data mining dengan metode klasifikasi. Analisis dilakukan untuk membangun model yang mampu mengklasifikasikan keterlambatan pengembalian buku di perpustakaan menggunakan algoritma Decision Tree dan Naïve Bayes. Seluruh tahapan pengolahan data, pembentukan model, hingga evaluasi kinerja algoritma dilakukan dengan menggunakan bahasa pemrograman *python*. Adapun tahapan analisis untuk masing-masing algoritma diuraikan sebagai berikut:

1. Analisis Menggunakan Algoritma Decision Tree dan Naïve Bayes

a. Analisis menggunakan Decision Tree

Analisis data dilakukan dengan menggunakan algoritma Decision Tree untuk membangun model klasifikasi yang dapat mengidentifikasi keterlambatan pengembalian buku. Adapun tahapan analisis yang diterapkan adalah sebagai berikut:

1) Menyiapkan Data Training

Data yang telah melalui tahap *preprocessing* pada proses sebelumnya digunakan sebagai *data training* untuk membangun model klasifikasi. Dataset tersebut berisi beberapa atribut yang berkaitan dengan aktivitas peminjaman buku di perpustakaan, serta satu atribut kelas yang menunjukkan kategori terlambat atau tidak terlambat. *Data training* ini digunakan oleh algoritma untuk mempelajari pola keterlambatan pengembalian buku.

UIN IMAM BONJOL PADANG

2) Menghitung Nilai Entropy

Tahap berikutnya dilakukan dengan menentukan nilai *entropy* pada dataset. Perhitungan ini digunakan untuk mengetahui tingkat ketidakpastian serta keragaman kelas yang terdapat dalam suatu himpunan data. nilai *entropy* yang tinggi menunjukkan bahwa komposisi data dalam himpunan tersebut semakin beragam atau kurang homogen. Sebaliknya, nilai *entropy* yang rendah menunjukkan bahwa data memiliki kecenderungan yang lebih seragam dalam suatu kelas. Perhitungan *entropy* yang digunakan dalam penelitian ini mengikuti persamaan 2.1.

3) Pembentukan Pohon Keputusan

etelah atribut dengan nilai *gain* tertinggi ditetapkan sebagai akar pohon, pembentukan decision tree dilanjutkan dengan membagi data berdasarkan nilai atribut tersebut. Pembagian data dilakukan secara rekursif dengan menghitung kembali nilai *entropy* dan *gain* pada setiap cabang hingga mencapai kondisi ketika seluruh data pada suatu node memiliki kelas yang sama atau tidak terdapat atribut lain yang dapat digunakan untuk melakukan pemisahan. Proses tersebut menghasilkan struktur pohon keputusan yang memuat aturan-aturan untuk melakukan klasifikasi.

4) Proses Klasifikasi

Setelah pohon keputusan terbentuk, model kemudian digunakan untuk melakukan klasifikasi terhadap *data testing*. Data uji dimasukkan ke dalam model dan akan mengikuti aturan yang terdapat pada pohon keputusan hingga mencapai node daun (leaf) yang menunjukkan hasil klasifikasi. Melalui proses ini, sistem dapat mengklasifikasikan apakah suatu data peminjaman buku termasuk dalam kategori terlambat atau tidak terlambat dalam pengembalian buku.

b. Analisis menggunakan Naïve Bayes

Selain algoritma Decision Tree, penelitian ini menerapkan algoritma Naïve Bayes untuk melakukan klasifikasi status pengembalian buku ke dalam dua kategori, yaitu terlambat dan tidak terlambat. Tahapan analisis yang digunakan

UIN IMAM BONJOL PADANG

mengacu pada landasan teori yang telah diuraikan pada persamaan 2.3, dengan tahapan sebagai berikut:

1) Menentukan Peluang Prior

   Probabilitas prior merupakan probabilitas awal dari masing-masing kelas sebelum mempertimbangkan atribut data. Pada penelitian ini, probabilitas prior dihitung berdasarkan proporsi jumlah data pada setiap kelas (terlambat dan tidak terlambat) dalam data latih.

2) Menentukan Peluang Kondisional

   Peluang kondisional dihitung untuk mengetahui kemungkinan munculnya kategori suatu atribut terhadap masing-masing kelas, yaitu terlambat atau tidak terlambat. Atribut yang digunakan meliputi angkatan, hari, bulan, kategori buku, dan lama izin. Perhitungan dilakukan berdasarkan frekuensi kemunculan setiap kategori atribut pada masing-masing kelas dalam data latih.

3) Perhitungan Menggunakan Persamaan Naïve Bayes

   Setelah diperoleh nilai probabilitas prior dan probabilitas kondisional, langkah selanjutnya adalah menghitung probabilitas posterior menggunakan persamaan Naïve Bayes. Hasil perhitungan probabilitas untuk setiap kelas selanjutnya dibandingkan, kemudian kelas yang memiliki nilai probabilitas paling tinggi ditetapkan sebagai hasil klasifikasi terhadap data uji.

UIN IMAM BONJOL PADANG

2. Evaluasi Hasil Klasifikasi

Setelah model Decision Tree dan Naïve Bayes menghasilkan hasil klasifikasi, tahap selanjutnya adalah melakukan evaluasi terhadap kinerja masing-masing model. Evaluasi dilakukan untuk mengetahui kemampuan model dalam memberikan hasil klasifikasi yang tepat terhadap pola keterlambatan pengembalian buku. Dalam penelitian ini, evaluasi kinerja model dilakukan menggunakan *confusion matrix* yang disertai dengan beberapa metrik evaluasi. Hasil evaluasi tersebut digunakan untuk mengetahui kemampuan masing-masing algoritma dalam melakukan klasifikasi terhadap data uji.

Evaluasi model dilakukan menggunakan *confusion matrix* dan beberapa metrik evaluasi, yaitu sebagai berikut:

a. Confusion Matrix

Confusion matrix digunakan untuk melihat kesesuaian antara hasil klasifikasi model dan kelas aktual pada data uji. Melalui matriks ini, dapat diketahui jumlah data yang diklasifikasikan dengan benar maupun yang mengalami kesalahan klasifikasi. Informasi tersebut kemudian digunakan sebagai dasar untuk menghitung berbagai metrik yang diperlukan dalam mengukur kinerja model:

1) *True Positive* (TP): Data yang di klasifikasi terlambat dan memang benar terlambat.
2) *True Negative* (TN): Data yang di klasifikasi tepat waktu dan memang benar tepat waktu.
3) *False Positive* (FP): Data yang di klasifikasi terlambat, namun sebenarnya tepat waktu.
4) *False Negative* (FN): Data yang di klasifikasi tepat waktu, namun sebenarnya terlambat.

b. *Accuracy*

Accuracy merupakan metrik yang digunakan untuk mengukur tingkat ketepatan model dalam mengklasifikasikan keseluruhan data uji. Nilai *accuracy* dihitung dengan membandingkan jumlah klasifikasi yang benar dengan total data yang diuji. Semakin tinggi nilai *accuracy* yang dihasilkan, semakin baik tingkat ketepatan model dalam melakukan klasifikasi.

c. *Precision*

Precision digunakan untuk mengukur ketepatan model dalam mengklasifikasikan data ke dalam kelas terlambat. Nilai *precision* dihitung berdasarkan perbandingan antara jumlah *True Positive* (TP) dengan seluruh data yang di klasifikasi sebagai kelas terlambat, yaitu *True Positive* (TP) dan *False Positive* (FP). Semakin tinggi nilai *precision* yang diperoleh, semakin rendah kemungkinan model menghasilkan klasifikasi yang salah pada data yang dikategorikan sebagai terlambat.

UIN IMAM BONJOL PADANG

d. *Recall*

Recall digunakan untuk mengukur kemampuan model dalam mengenali data yang sebenarnya termasuk dalam kategori terlambat. Nilai *recall* dihitung dengan membandingkan jumlah *True Positive* (TP) dengan seluruh data yang secara aktual berada pada kelas terlambat, yaitu *True Positive* (TP) dan *False Negative* (FN). Semakin tinggi nilai *recall* yang diperoleh, semakin baik kemampuan model dalam mengidentifikasi kasus keterlambatan pengembalian buku.

e. F1-Score

F1-Score digunakan untuk mengukur keseimbangan antara nilai *precision* dan *recall* pada model klasifikasi. Nilai F1-Score diperoleh melalui perhitungan rata-rata harmonis dari kedua metrik tersebut, sehingga dapat memberikan gambaran yang lebih menyeluruh mengenai kinerja model dalam melakukan klasifikasi. Metrik ini juga sesuai digunakan ketika distribusi jumlah data antar kelas mengalami ketidakseimbangan.

3. Deployment Model

Tahap *deployment* merupakan proses penerapan model klasifikasi yang menghasilkan kinerja paling baik. Sebelum diterapkan, model Decision Tree dan Naïve Bayes dibandingkan berdasarkan nilai *accuracy*, *precision*, *recall*, dan F1-score yang diperoleh pada tahap evaluasi. Selanjutnya, model dengan kinerja terbaik dipilih untuk diimplementasikan ke dalam aplikasi berbasis web menggunakan Streamlit.

Aplikasi yang dikembangkan berfungsi sebagai media visualisasi dan pengujian model terpilih. Data yang dimasukkan kemudian diproses oleh model terbaik untuk menghasilkan klasifikasi apakah pengembalian buku termasuk dalam kategori terlambat atau tidak terlambat. Aplikasi yang dikembangkan dalam penelitian ini bersifat prototipe dan digunakan sebagai media visualisasi serta pengujian model, sehingga belum terintegrasi secara langsung dengan sistem operasional perpustakaan.

UIN IMAM BONJOL PADANG