Jurnal Sifo Mikroskil (JSM)
Volume 26, No 02, Oktober 2025 – Hal 275 - 288
DOI: https://doi.org/10.55601/jsm.v26i2.1803

e-ISSN: 2622-8130
ISSN: 1412-0100

Model Klasifikasi Kelayakan Pengajuan Kredit Bank Menggunakan Naïve Bayes

Joshua Valens Andrian[1], Charitas Fibriani[2]
[1,2,]Universitas Kristen Satya Wacana, Jl. Diponegoro 52-60, (0298)323212
[1,2,]Fakultas Teknologi Informasi, Sistem Informasi, Universitas Kristen Satya Wacana, Salatiga
e-mail: [1]682022008@student.uksw.edu, [2]charitas.fibriani@uksw.edu

Dikirim: ***14-07-2025*** | *Diterima:* ***07-10-2025*** | *Diterbitkan:* ***31-10-2025***

Abstrak

Pemberian pinjaman oleh Lembaga keuangan membutuhkan evaluasi yang baik untuk memastikan kelayakan nasabah sebagai calon debitur demi mengurangi resiko gagal bayar. Atribut yang dipakai dalam menentukan kelayakan calon debitur, seperti daerah tempat tinggal, jenis kelamin, status perkawinan, beban tanggungan, edukasi, status pekerjaan, pendapatan, pendapatan sampingan, jumlah pinjaman, lama pinjaman, riwayat kredit. Naïve Bayes merupakan metode klasifikasi berbasis *probabilistic* dengan asumsi independen antara atribut. Atribut– atribut tersebut dihitung untuk mendapatkan atribut kelayakan pinjaman sebagai hasil klasifikasi yang terkait dengan seberapa akurat performa model Naïve Bayes dalam memprediksi kelayakan kredit berdasarkan data historis. Kualitas model diukur menggunakan *confusion matrix*, juga dibahas terkait akurasi, presisi dan *recall*. Rumusan masalah pada penelitian ini adalah Bagaimana cara kerja model klasifikasi untuk kredit bank menggunakan Naïve Bayes. Pengujian menunjukkan algoritma Naïve Bayes mampu melakukan klasifikasi dengan tingkat akurasi 86%, *precision* sebesar 88% dan *recall* sebesar 86% dengan hasil tersebut harapan dari penelitian ini adalah model Naïve Bayes bisa bermanfaat untuk meningkatkan efisiensi dalam proses pengajuan pinjaman pada Lembaga keuangan.

Kata kunci: Klasifikasi, Naïve Bayes, Pemberian Pinjaman, *Confusion Matrix*.

Abstract

Lending by financial institutions requires a good evaluation to ensure the eligibility of customers as prospective debtors in order to reduce the risk of default. The attributes used in determining the eligibility of prospective debtors, such as area of residence, gender, marital status, burden of dependents, education, employment status, income, side income, loan amount, loan duration, credit history. Naïve Bayes is a probabilistic-based classification method with the assumption of independence between attributes. These attributes are calculated to obtain loan eligibility attributes as a classification result related to how accurately the Naive Bayes model performs in predicting creditworthiness based on historical data. Model quality is measured using a confusion matrix, also discussed regarding accuracy, recall and precision. The formulation of the problem in this study is How does the classification model work for bank credit using Naïve Bayes. Testing shows that the Naïve Bayes algorithm is able to classify with an accuracy level of 86% with these results, the hope of this study is that the naïve Bayes model can be useful for increasing efficiency in the loan application process at financial institutions.

Keywords: *Classification, Naïve bayes*, *Loan*, *Confusion matrix*

1. PENDAHULUAN

Perkembangan teknologi yang pesat sekarang ini membuat lembaga keuangan untuk implementasi teknologi di dalam proses bisnisnya. Lembaga keuangan diharuskan sangat selektif terhadap pengajuan kredit dikarenakan adanya risiko gagal bayar yang merugikan perusahaan, namun proses seleksi ini terkadang memerlukan waktu yang lama. Perusahaan membutuhkan alat yang mampu melakukan seleksi nasabah dengan cepat dan akurat untuk menentukan kelayakan penerimaan pinjaman. Model klasifikasi pengajuan kredit dengan algoritma Naïve Bayes mampu menjawab tantangan tersebut.

Algoritma Naïve Bayes memiliki prinsip bahwa setiap atribut saling independen yang dikenal dengan istilah "Naive." Penerapan algoritma Naïve Bayes memiliki asumsi bahwa tidak ada hubungan antara satu atribut dengan atribut lainnya walau ada kemungkinan atribut – atribut tersebut memiliki keterkaitan [1]. Naïve Bayes adalah algoritma berbasis hipotesis *Bayes* yang memiliki anggapan bahwa setiap variabel bersifat bebas bebas sehingga keberadaan suatu komponen tidak memiliki hubungannya dengan keberadaan elemen yang lainnya [2]. Klasifikasi Naïve Bayes merupakan algoritma bersifat probabilistik. Algoritma ini digunakan untuk klasifikasi dan efektif untuk kumpulan data besar. Naïve Bayes sangat cocok untuk memprediksi pengajuan kredit [3]. Pengembangan model klasifikasi pengajuan pinjaman dengan Naïve Bayes dilakukan dengan tahapan sistematis dimulai dari pengumpulan dan pemahaman data, kemudian dilakukan preprocessing data yaitu pembersihan data dan penyesuaian atribut, lalu dilakukan pemisahan data untuk data latih dan uji. Data tersebut akan digunakan untuk pembangunan model klasifikasi Naïve Bayes yang akan menentukan prediksi data berdasarkan atribut berbasis teorema bayes yang digunakan untuk menentukan kelayakan pinjaman berdasarkan atribut yang diberikan. Evaluasi model dibuat juga metrik evaluasi yaitu confusion matrix sehingga ada ukuran yang jelas terkait akurasi model yang membuat model ini memiliki Tingkat akurasi yang jelas dan transparansi perhitungan yang baik karena alur penelitian ini tidak hanya membahas terkait pembuatan model tetapi cara kerja model tersebut. Rumusan masalah dalam penelitian ini adalah bagaimana membangun model klasifikasi untuk kredit bank menggunakan Naïve Bayes dan apakah model yang dibangun ini bisa digunakan untuk memprediksi kreditur baru dengan baik. Tujuan dari penelitian ini adalah membangun model untuk memprediksi kelayakan pengajuan kredit untuk nasabah yang bisa diaplikasikan untuk lembaga keuangan dengan baik.

2. TINJAUAN PUSTAKA

Penerapan metode klasifikasi Naïve Bayes dalam prediksi kelayakan kredit telah banyak dilakukan oleh berbagai peneliti sebelumnya dengan hasil yang beragam, tergantung pada karakteristik data dan skema evaluasi yang digunakan. Penelitian [4] Menggunakan klasifikasi Naïve bayes sebagai algoritma website untuk memisahkan kategori berita berdasarkan judul. *Website* berhasil lolos uji coba sehingga mampu berjalan dengan baik dan algoritma klasifikasi Naïve Bayes dalam menyeleksi judul berita menorehkan akurasi sebesar 86%. Penelitian [5]membandingkan performa antara algoritma Naïve Bayes dan *Support Vector Machine* dalam melakukan analisis sentimen di twitter. Hasil pengujian menunjukkan bahwa Naïve Bayes menghasilkan akurasi sebesar 59,86%, sedikit lebih baik dibandingkan SVM, jika menggunakan fungsi *information gain* Naïve Bayes mampu memperoleh akurasi mencapai 72.02%. Akurasi tersebut lebih tinggi jika dibandingkan SVM dengan *information gain* yang memperoleh akurasi mencapai 62.38%. *Information Gain* merupakan fungsi yang meningkatkan kualitas data sebelum masuk ke algoritma. Naïve Bayes sangat membutuhkan kualitas dan kebersihan data yang maksimal agar memberikan performa yang baik.

Penelitian [6] mengimplementasikan metode Naïve Bayes untuk mengevaluasi kelayakan pinjaman calon nasabah koperasi. Penelitian ini membuktikan bahwa metode tersebut efektif dan memberikan hasil yang cukup baik nilai akurasi 78.08%. Fokus utama penelitian ini adalah efisiensi metode dalam menangani dataset dengan atribut yang banyak. Penelitian [7] menerapkan algoritma Naïve Bayes pada kasus serupa, namun dengan hasil akurasi yang lebih rendah, yaitu 71%. Penelitian ini menunjukkan bahwa kemampuan prediksi sangat bergantung pada kualitas data latih dari

preprocessing yang dilakukan sebelum proses klasifikasi. Penelitian ini menjadi bukti bahwa metode Naïve Bayes tidak selalu unggul jika dataset mengandung noise atau distribusi data tidak seimbang. Penelitian [8] membandingkan Naïve Bayes dan C4.5 dalam konteks penentuan kelayakan penerima pinjaman koperasi. Hasil menunjukkan bahwa Naïve Bayes mencapai akurasi sebesar 89,90%, namun kalah dibanding C4.5 dengan 91.4% akurasi.Atribut pada penelitian ini dominan berupa data numerik. Penelitian ini menekankan pentingnya pemilihan algoritma yang sesuai dengan struktur data agar hasil prediksi lebih optimal.

Berdasarkan kelima penelitian terdahulu tersebut, dapat disimpulkan bahwa metode Naïve Bayes umumnya memberikan hasil klasifikasi yang cukup baik dalam konteks prediksi kelayakan kredit. Akan tetapi, kualitas hasil sangat bergantung pada kondisi dan struktur dataset, metode *preprocessing* yang diterapkan, serta skema evaluasi yang digunakan. Penelitian ini bertujuan untuk mengkaji lebih lanjut Algoritma Naïve Bayes dengan pendekatan berbasis confusion matrix dan *test on test data* guna menghasilkan penghitungan klasifikasi yang lebih mendalam.

Gap penelitian dalam studi ini terletak pada minimnya penelitian yang membahas pembuatan model Naïve Bayes dengan fokus mendalam pada aspek penggunaan rumus dan tahapan pemodelan secara sistematis, terutama terkait studi kasus pengajuan pinjaman. Penelitian terdahulu dominan menampilkan hasil akhir akurasi tanpa menguraikan proses klasifikasi secara rinci akibatnya kurang memberikan pemahaman menyeluruh bagi pembaca yang ingin mempelajari atau mereplikasi metode ini. Penelitian ini berupaya mengisi kekosongan tersebut dengan menyajikan analisis terperinci pada setiap langkah perhitungan Naïve Bayes, mulai dari preprocessing data, penjabaran rumus, hingga penggunaan rumus pada proses klasifikasi sehingga menghasilkan model yang tidak hanya akurat, tetapi juga transparan dan mudah dipahami.

Klasifikasi merupakan cabang dari data mining yang melakukan pembagian kelas dari objek atau atribut yang ada di dalam dataset. Klasifikasi dimulai dari pembelajaran data lalu proses klasifikasi data. Algoritma akan membuat model dengan analisis *training data* sehingga bisa muncul kelas klasifikasinya [9]. Naïve Bayes adalah salah satu algoritma yang dipilih pada penelitian ini yang memiliki rumus dasar hipotesis bayes.

$$P(X) = \frac{P(X|H).P(H)}{P(X)} \quad (1)$$

Di mana:
X = Data dengan class yang belum diketahui
H = Hipotesis data X merupakan suatu class spesifik
P(H|X) = Probabilitas hipotesis H berdasarkan kondisi x (posteriori probabilitas)
P(H) = Probabilitas hipotesis H (prior probabilitas)
P(X|H) = Probabilitas X berdasarkan kondisi tersebut
P(X) = Probabilitas dari X

Naïve Bayes adalah algoritma klasifikasi yang berbasis pada teorema Bayes (Rumus (1)) dengan asumsi naif, artinya seluruh atribut adalah independen sehingga tidak saling mempengaruhi ketika diberikan kelas target. Meskipun sederhana, Naïve Bayes sering kali efektif dalam klasifikasi teks dan data dengan atribut jenis kategorikal [10]. Naïve Bayes sendiri memiliki banyak rumus hingga mendapatkan hasil prediksi setelah membaca data latih dimulai dari rumus rata-rata:

$$\mu = \frac{\Sigma_{i=1}^{n} Xi}{n} \quad (2)$$

Rumus rata-rata digunakan pertama kali untuk mendapatkan nilai yang digunakan untuk melakukan perhitungan numerik pada rumus selanjutnya.

Rumus Standar Deviasi :

$$\sigma = \sqrt{\frac{\Sigma_{i=1}^{n} (Xi - \mu)\,2}{n-1}} \quad (3)$$

Rumus standar deviasi digunakan setelah mendapatkan nilai rata-rata pada Rumus (2). Rumus ini akan menghasilkan nilai yang akan digunakan untuk rumus selanjutnya pada data numerik.[11]

Rumus Prior

$$P(C) = \frac{Jumlah\ data\ dengan\ kelas\ C}{Total\ jumlah\ data} \quad (4)$$

Rumus prior *probability* akan menjelaskan peluang munculnya kelas tertentu berdasarkan atribut tertentu. Rumus ini akan digunakan pertama kali untuk menentukan setiap peluang data kategorikal.

Rumus Gaussian

$$P(C) = \frac{1}{\sqrt{2\pi\sigma_C^2}} e\,(\frac{(x_i - \mu C)^2}{2\sigma_c^2}) \quad (5)$$

Rumus Gaussian akan menentukan nilai yang bersifat numerik. Fungsi rumus ini adalah untuk menentukan angka probabilitas untuk perhitungan yang berbasis angka. Sebelum rumus ini dikerjakan harus menyelesaikan Rumus (2) dan Rumus (3) untuk mendapatkan nilai rata – rata dan standar deviasi.

Rumus *Conditional*

$$P(C) = \frac{jumlah\ kemunculan\ x_i dalam\ kelas\ C}{jumlah\ total\ data\ dalam\ kelas\ C} \quad (6)$$

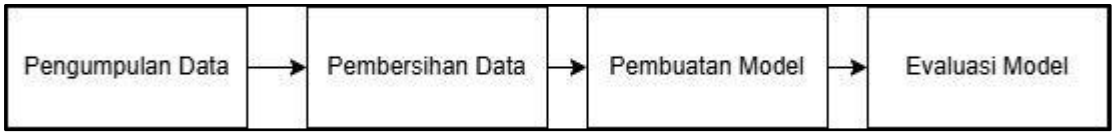
Rumus *conditional probability* adalah rumus yang digunakan setelah Rumus (4). Rumus ini akan menentukan kemunculan atribut tertentu pada kelas tertentu. Misalnya atribut gender *female* perempuan pada kelas *yes*.

Rumus Perbandingan Posterior

$$P(C_1).P(C_1) \quad vs \quad P(x\,|C_2).P(C_2) \quad (7)$$

Rumus perbandingan posterior merupakan perbandingan peluang munculnya kelas klasifikasi tertentu setelah seluruh atribut dihitung dengan Rumus (4),(5) dan (6). Rumus ini pada dasarnya sama dengan Rumus (1). Perbedaan yang paling mendasar dengan Rumus (1), Rumus (1) adalah dasar dari seluruh Naïve Bayes, sedangkan Rumus (7) adalah rumus yang disesuaikan dengan dataset ini karena hanya memiliki 2 kelas klasifikasi yaitu Y dan N. Rumus ini dikerjakan 2 kali kemudian dibandingkan mana yang memiliki nilai lebih besar itulah yang menjadi kelas klasifikasi.

3. METODE PENELITIAN



Gambar 1. Flowchart tahapan penelitian

Tahap pertama adalah pengumpulan data pada Gambar 1. Data diambil dari *website* Kaggle.com yang memiliki judul "*Loan Data Set*". Dataset ini milik *user* bernama Burak Ergun yang membahas

terkait pengajuan kredit dalam format *.csv*. Atribut dalam dataset ini berjumlah 13 dengan 614 baris data. Struktur atribut dalam data ini adalah 1 atribut bersifat unik, 11 atribut menjadi atribut penentu, dan ditutup dengan 1 atribut yang menjadi label atau target. Data ini tidak memiliki lisensi sehingga bebas untuk digunakan untuk penelitian. Atribut dan penjelasan pada Dataset yang digunakan pada penelitian ini mengacu pada Tabel 1.

Tabel 1 Daftar Atribut dan Target

No	Atribut	Sifat Atribut	Isi/Range	Keterangan
1	Loan_ID	Identifier	ID PINJAMAN	Id Pinjaman (Unik)
2	Gender	Kategorikal	Male/Female	Jenis Kelamin (Pria/Wanita)
3	Married	Kategorial	Yes/No	Berada di dalam Status Perkawinan (Ya/Tidak)
4	Dependent	Numerik	0-9999+	Jumlah tunjangan keluarga
5	Education	Kategorikal	Graduate/Not Graduate	Status kelulusan sekolah (Lulus/Tidak Lulus)
6	Self_Employeed	Kategorikal	Yes/No	Memiliki usaha sendiri (Ya/Tidak)
7	Credit_History	Numerik	0-1	Sejarah Pinjaman 0 = Belum pernah meminjam atau pernah kredit macet 1 = Kredit lancar
8	Property_Area	Kategorikal	Urban/SemiUrban/Rural	Wilayah tempat tinggal (Kota besar/Kota Kecil/ Pedesaan)
9	Applicant_Income	Numerik	0-9999+	Pendapatan utama calon kreditur
10	CoApplicant_Income	Numerik	0-9999+	Pendapatan sampingan calon kreditur
11	Loan_Amount	Numerik	0-9999+	Jumlah pinjaman kreditur
12	Loan_Amount_Term	Numerik	0-9999+	Tenor pinjaman kreditur
13	Loan_Status (Target)	Kategorikal	Y/N	Keputusan penerimaan pinjaman kreditur

Langkah kedua adalah pembersihan data. Persebaran data dapat dilihat pada Gambar 2, terdapat beberapa data yang jumlahnya tidak sampai 614 artinya terdapat null pada dataset tersebut. Dataset akan dibersihkan dengan cara menghapus baris data yang memiliki nilai null dengan cara manual sehingga seluruh data memiliki atribut yang lengkap. Data dihapus secara manual dengan tujuan mendapatkan nilai klasifikasi yang lengkap dan murni pada setiap data train karena jika disesuaikan dengan studi kasus nyata, data – data yang digunakan untuk proses pengajuan peminjaman bank harus di isi dengan lengkap. Atribut *Dependent* mengalami penyesuaian pada beberapa data yang memiliki nilai "3+" tapi tidak ada nilai "3" sehingga nilai "3+" ini di ubah menjadi "3". Setelah dilakukan pembersihan dataset yang semula sejumlah 614 baris tersisa menjadi 480 baris data. Loan Status adalah target utama dari model klasifikasi ini. *Loan Status* adalah diterima atau ditolaknya pinjaman seseorang dilambangkan dengan Y dan N artinya *YES* (Diterima) dan *NO* (Ditolak) atribut ini bersifat kategorikal. Proses pembuatan model menggunakan 11 atribut yang diperhitungkan ditambahkan 1 atribut *identifier* yang akan menghasilkan 1 atribut *final* sebagai keputusan diterima atau ditolaknya pinjaman.

2622-8130 Jurnal Sifo Mikroskil (JSM) - Volume 26, No 02, Oktober 2025
DOI: https://doi.org/10.55601/jsm.v26i2.1803

Gender	Count	Percentage
Male	489	81%
Female	112	19%

Married	Count	Percentage
Yes	398	65%
No	213	35%

Dependent	Count	Percentage
0	345	58%
1	102	17%
2	101	17%
3+	51	9%

Education	Count	Percentage
Graduate	480	78%
Not Graduate	134	22%

Self_Employed	Count	Percentage
No	500	86%
Yes	82	14%

Credit_History	Count	Percentage
1	475	84%
0	89	16%

Property_Area	Count	Percentage
Urban	202	33%
Semiurban	233	38%
Rural	179	29%

Loan_Status	Count	Percentage
Y	422	69%
N	192	31%

ApplicantIncome count	614
CoApplication count	614
Loan_amount count	592
Loan_amount term count	600

Gambar 2. Persebaran data pada dataset

Atribut data pada Gambar 2 memiliki persebaran sebagai berikut :

Gender : *Male* 489, *Female* 112. Dataset ini memiliki data pria yang jauh lebih dominan dibandingkan wanita. Rasio nya mencapai 81%:19%. Total baris pada *gender* adalah 601 baris. Artinya ada 13 data yang dihilangkan karena tidak memiliki atribut *gender*.

Married : *Yes* 398, *No* 213. Dataset ini memiliki status pernikahan yang lebih dominan adalah di dalam pernikahan. Rasio nya adalah 65%:35%. Total baris pada *married* adalah 611 sehingga ada 3 baris data yang dihilangkan karena tidak memiliki atribut ini.

Education : *Graduate* 480, *Not Graduate* 134. Dataset ini memiliki status pendidikan yang dominan lulus sekolah. Rasionya 78%:22%. Total baris pada *Education* adalah 614 sehingga tidak ada baris yang di hilangkan.

Self Employed : *No* 500, *Yes* 82. Dataset ini memiliki posisi pekerjaan dominan sebagai karyawan. Rasionya mencapai 86%:14%. Total baris pada *Self Employed* adalah 582 baris sehingga ada 32 baris yang dihilangkan karena tidak memiliki atribut ini. *Self employed* merupakan status dimana seseorang tersebut "bekerja diluar korporasi" sehingga jika statusnya *No* maka dia merupakan seorang pekerja atau karyawan.

Property Area : *Urban* 202, *Semiurban* 233, *Rural* 179. Dataset ini memiliki persebaran tempat tinggal yang cukup merata. Rasionya adalah 33%:38%:29%. Total baris pada *Property Area* adalah 614 sehingga tidak ada data yang dihilangkan karena tidak memiliki atribut ini. *Property area* merupakan atribut yang melambangkan seseorang tersebut tinggal di daerah apa.

Dependent : Jumlah baris dalam data adalah 599 yang dominan tidak memiliki beban tanggungan. *Dependent* 0 memiliki rasio 58% dari keseluruhan data. Terdapat 15 baris yang dihapus karena tidak memiliki atribut ini. Atribut ini melambangkan jumlah tanggungan di dalam keluarga, maksudnya adalah di dalam 1 keluarga berapa banyak orang yang dibiayai oleh calon peminjam.

Credit History : Jumlah baris dalam data adalah 564 yang dominan memiliki Riwayat kredit yang baik dilambangkan dengan angka 1. Rasionya mencapai 84%. Terdapat 50 baris yang dihapus karena tidak memiliki atribut ini. Atribut ini melambangkan bagaimana peminjaman orang ini sebelumnya, apakah baik, buruk, atau belum pernah meminjam sama sekali.

Applicant Income memiliki jumlah baris yang lengkap yaitu 614 sehingga tidak ada baris yang dihapus karena tidak memiliki atribut ini. Atribut ini melambangkan pendapatan utama per bulan.

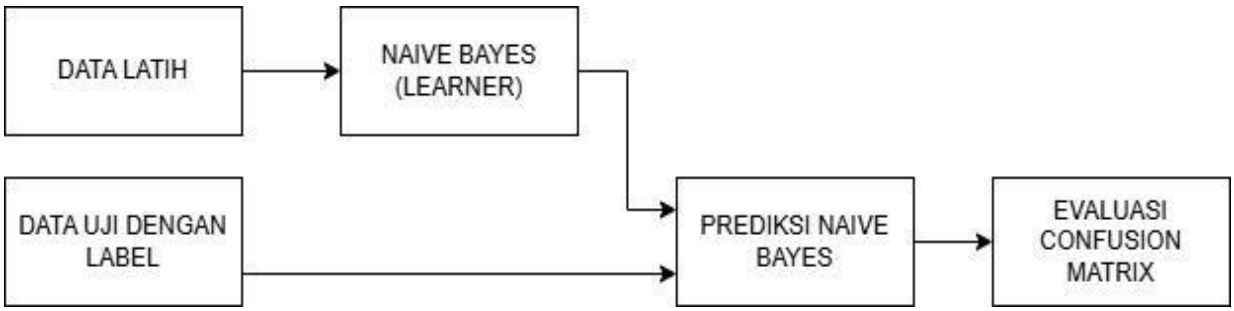
CoApplicationIncome memiliki jumlah baris yang lengkap yaitu 614 sehingga tidak ada baris yang dihapus karena tidak memiliki atribut ini. Atribut ini melambangkan pendapatan sampingan per bulan.

Loan amount memiliki 592 baris. Terdapat 85 baris yang dihapus karena tidak memiliki atribut ini. Atribut ini melambangkan jumlah pinjaman.

Loan amount term memiliki 600 baris. Terdapat 14 data yang dihapus karena tidak memiliki atribut ini. Atribut ini melambangkan tempo pinjaman dalam bulan.

Loan status adalah atribut target yang lengkap berjumlah 614 baris. Atribut ini dominan pada status *Yes*(Y) dengan rasio 68:31%. *Loan status* merupakan atribut yang menjadi kelas klasifikasi target dalam model ini. Kelas Y merepresentasikan pinjaman yang disetujui dan N ditolak. Data yang digunakan memiliki nilai Y yang lebih banyak sehingga akan lebih mudah mengeenali kelas Y sehingga untuk

mengatasi *imbalance* tersebut metrik yang digunakan tidak hanya akurasi, tetapi *precision* dan *recall* untuk memastikan model mampu mengenali kelas dengan baik.
Proses pembersihan data yang dilakukan secara manual yaitu menghilangkan baris yang mempunyai nilai null akibatnya, jumlah baris data yang semula adalah 614 baris berkurang menjadi 480 baris. Data yang sudah bersih ini digunakan untuk proses pembangunan model.



Gambar 3. Skema pembuatan model Naïve Bayes

Langkah ketiga dimulai dengan memisahkan data latih dan data uji. Data latih diambil sebanyak 429 baris data dengan data uji 51 baris data. Rasio pembagian data merupakan 85% data latih dan 15% data uji yang dipisah menjadi 2 file. Rasio tersebut dipilih karena 85:15 merupakan proporsi yang umum untuk uji coba model dalam skala kecil. 429 baris data akan dipelajari terlebih dahulu oleh Naïve Bayes learner untuk bisa menentukan prediksi data. 51 baris Data uji yang digunakan masih memiliki label tujuannya agar model prediksi Naïve Bayes bisa di evaluasi. Prediksi Naïve Bayes dapat dilakukan dengan membaca peluang kedua kelas klasifikasi terlebih dahulu menggunakan Rumus (4) kemudian algoritma akan memisahkan data yang bersifat numerik yang diolah menggunakan Rumus (5) dan data kategorikal yang diolah menggunakan Rumus (6). Klasifikasi kelas dapat dilakukan menggunakan Rumus (7). Rumus (7) akan membandingkan kedua probabilitas kelas berdasarkan perhitungan seluruh atribut dalam baris data.. Confusion matrix merupakan instrumen yang dipilih untuk melakukan evaluasi model. Alasan dipilihnya algoritma Naïve Bayes karena algoritma ini memiliki kesesuaian yang baik pada dataset. Algoritma Naïve Bayes sangat berjalan dengan baik pada dataset yang memiliki banyak atribut kategorikal yang relevan dengan studi kasus peminjaman bank.

Langkah keempat adalah melakukan prediksi menggunakan *Confusion Matri*x. Skema *Confusion Matrix* dapat melihat pada Tabel 2. Skema ini akan membandingkan data hasil prediksi dengan data label dan akan terlihat prediksi mana yang benar dan salah. Proses ini akan melakukan evaluasi prediksi model klasifikasi dengan beberapa metrik yang digunakan untuk menentukan efektivitas model. Proses evaluasi akan menghasilkan nilai berupa Akurasi, Presisi dan *Recall* [12]. Akurasi yang dihasilkan dihitung berdasarkan persebaran nilai pada *confusion matrix*. Perhitungan dilakukan dengan memanfaatkan jumlah prediksi positif yang benar (*True Positive*), prediksi positif yang salah (*False Positive*), prediksi negatif yang benar (True Negatif) dan prediksi negatif yang salah (*False Negatif*). Semakin tinggi nilai akurasi yang didapat maka semakin baik pula metode yang dihasilkan [13].

Tabel 2. Model Confusion Matrix

Klasifikasi		**Kelas Prediksi**	
		YES	**NO**
Kelas Nyata	**YES**	**TRUE POSITIVE (TP)**	**FALSE NEGATIVE (FN)**
	NO	**FALSE POSITIVE (FP)**	**TRUE NEGATIVE (TN)**

$$Precision = \frac{TP}{FP+TP} \quad (8)$$

$$Recall = \frac{TP}{TP+TN} \quad (9)$$

$$Accuracy = \frac{TP+TN}{TP+FN+FP+TN} \quad (10)$$

Precision Rumus (8) Merupakan rasio prediksi benar positif dibandingkan dengan keseluruhan hasil yang diprediksi positif. *Precision* menggunakan analogi pertanyaan "Dari semua pengajuan kredit yang diprediksi diterima, berapa banyak yang benar-benar diterima?". *Recall* Rumus (9) Merupakan rasio prediksi benar positif dibandingkan dengan keseluruhan data yang benar positif. *Recall* menggunakan analogi pertanyaan "Dari semua pengajuan kredit yang sebenarnya layak diterima, berapa banyak yang berhasil diprediksi sistem sebagai diterima?". *Accuracy* Rumus (10) merupakan rasio prediksi Benar (positif dan negatif) dengan keseluruhan data. Akurasi menggunakan analogi pertanyaan "Dari seluruh pengajuan kredit yang diperiksa, berapa banyak yang keputusannya benar?" [14].

4. HASIL DAN PEMBAHASAN

Visualisasi distribusi data pada Gambar 2 menyatakan sebagian besar pemohon pinjaman berjenis kelamin laki-laki (81%), sudah menikah (65%) dan berpendidikan sarjana (78%). Mayoritas tidak bekerja secara mandiri (86%) dan memiliki catatan kredit yang baik (*Credit_History* = 1) sebesar 84%, yang menjadi indikasi penting karena atribut ini sangat berpengaruh dalam penentuan kelayakan pinjaman. Distribusi jumlah tanggungan (*dependent*) didominasi oleh angka 0 (58%). Atribut seperti *Loan_Amount, Loan_Amount_Term* dan *applicant income* terlihat memiliki nilai kosong (null), sehingga dilakukan pembersihan data dengan menghapus baris yang tidak lengkap secara manual. Setelah proses ini, data tersisa sebanyak 480 baris yang sepenuhnya dan siap digunakan untuk pembuatan model Naïve Bayes. Langkah ini penting untuk memastikan bahwa algoritma dapat bekerja secara optimal tanpa gangguan dari atribut yang kosong.

Pembuatan model dilakukan dengan memisahkan data terlebih dahulu. Data latih berjumlah 429 baris dan data uji berjumlah 51 baris. Sampel masing – masing data dapat dilihat pada Gambar 4 dan Gambar 5.

Loan_ID	Gender	Married	Dependents	Education	Self_Employe	Applicanti	Coapplica	LoanAmo	Loan_Amoun	Credit_History	Property_	Loan_Statu
LP001003	Male	Yes	1	Graduate	No	4583	1508	128	360	1	Rural	N
LP001005	Male	Yes	0	Graduate	Yes	3000	0	66	360	1	Urban	Y
LP001006	Male	Yes	0	Not Gradu	No	2583	2358	120	360	1	Urban	Y
LP001008	Male	No	0	Graduate	No	6000	0	141	360	1	Urban	Y
LP001011	Male	Yes	2	Graduate	Yes	5417	4196	267	360	1	Urban	Y
LP001013	Male	Yes	0	Not Gradu	No	2333	1516	95	360	1	Urban	Y
LP001014	Male	Yes	3	Graduate	No	3036	2504	158	360	0	Semiurba	N
LP001018	Male	Yes	2	Graduate	No	4006	1526	168	360	1	Urban	Y
LP001020	Male	Yes	1	Graduate	No	12841	10968	349	360	1	Semiurba	N
LP001024	Male	Yes	2	Graduate	No	3200	700	70	360	1	Urban	Y
LP001028	Male	Yes	2	Graduate	No	3073	8106	200	360	1	Urban	Y
LP001029	Male	No	0	Graduate	No	1853	2840	114	360	1	Rural	N
LP001030	Male	Yes	2	Graduate	No	1299	1086	17	120	1	Urban	Y
LP001032	Male	No	0	Graduate	No	4950	0	125	360	1	Urban	Y
LP001036	Female	No	0	Graduate	No	3510	0	76	360	0	Urban	N
LP001038	Male	Yes	0	Not Gradu	No	4887	0	133	360	1	Rural	N

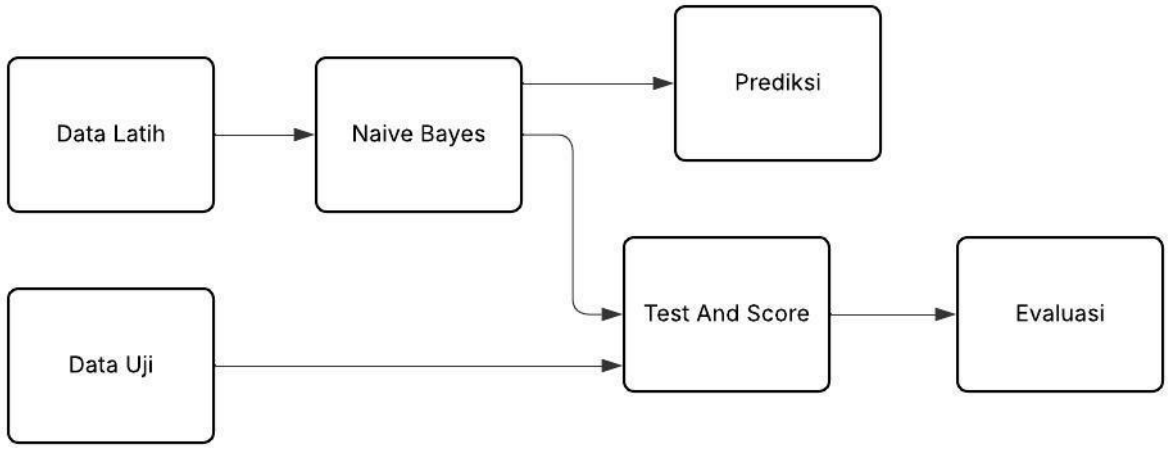
Gambar 4. Sampel Data latih

Data latih pada Gambar 4 berjumlah 381 baris. Data ini berfungsi untuk melatih Naïve Bayes agar bisa memprediksi data. Data latih harus memiliki label karena Naïve Bayes akan mempelajari seluruh atribut dimulai dari *Gender, Married, Dependents, Education, Self_Employee, Applicant_Income, CoApplicantIncome, LoanAmount, Loan_Amount_Term, Credit_history, dan Property_Area*. Rumus pertama yang digunakan adalah Rumus (4) untuk membaca peluang dari *loan_status* dan akan dilanjutkan dengan rumus lainnya untuk mempelajari data.

Loan_ID	Gender	Married	Dependen	Education	Self_Empl	Applicant	Coapplica	LoanAmo	Loan_Am	Credit_His	Property_	Loan_Status
LP002777	Male	Yes	0	Graduate	No	2785	2016	110	360	1	Rural	Y
LP002785	Male	Yes	1	Graduate	No	3333	3250	158	360	1	Urban	Y
LP002788	Male	Yes	0	Not Gradu	No	2454	2333	181	360	0	Urban	N
LP002789	Male	Yes	0	Graduate	No	3593	4266	132	180	0	Rural	N
LP002792	Male	Yes	1	Graduate	No	5468	1032	26	360	1	Semiurba	Y
LP002795	Male	Yes	3	Graduate	Yes	10139	0	260	360	1	Semiurba	Y
LP002798	Male	Yes	0	Graduate	No	3887	2669	162	360	1	Semiurba	Y
LP002804	Female	Yes	0	Graduate	No	4180	2306	182	360	1	Semiurba	Y
LP002807	Male	Yes	2	Not Gradu	No	3675	242	108	360	1	Semiurba	Y
LP002813	Female	Yes	1	Graduate	Yes	19484	0	600	360	1	Semiurba	Y
LP002820	Male	Yes	0	Graduate	No	5923	2054	211	360	1	Rural	Y
LP002821	Male	No	0	Not Gradu	Yes	5800	0	132	360	1	Semiurba	Y
LP002832	Male	Yes	2	Graduate	No	8799	0	258	360	0	Urban	N
LP002836	Male	No	0	Graduate	No	3333	0	70	360	1	Urban	Y
LP002837	Male	Yes	3	Graduate	No	3400	2500	123	360	0	Rural	N

Gambar 5. Sampel Data uji

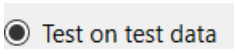
Data uji digunakan untuk menguji model setelah pelatihan selesai. Data ini adalah data yang tidak terlihat. Model maupun manusia tidak dapat melihat pola ini selama pelatihan. Dataset pelatihan, validasi, dan pengujian harus merupakan sampel yang representatif dari masalah yang akan dipecahkan [6]. Data uji pada penelitian ini berjumlah 51 baris yang memiliki label dengan tujuan bisa dievaluasi menggunakan *confusion matrix*. Sampel data uji harus memiliki jumlah atribut sama dengan data latih yang dapat dilihat contohnya pada Gambar 5.



Gambar 6. Skema Naïve Bayes untuk pengujian model

Skema Naïve Bayes pada Gambar 6 merupakan pengembangan dari rancangan model pada Gambar 3. Skema ini digunakan untuk melakukan pengujian model yang nantinya akan dievaluasi menggunakan *confusion matrix*, Skema ini harus menggunakan data uji yang sudah memiliki label karena adanya fungsi *test and score* dan *confusion matrix*.
Fungsi data terdapat 2 jenis data yaitu data latih dan data uji yang dapat diperhatikan pada Gambar 6. Data latih dan data uji sepenuhnya dipisah sehingga proses pengujian tidak terdapat kebocoran data. Data uji akan langsung masuk ke fungsi *test and score* untuk dievaluasi secara langsung tanpa "dipelajari" oleh Naïve Bayes terlebih dahulu. Fungsi Naïve Bayes pada Gambar 6 digunakan untuk mempelajari data latih agar bisa mengenali mana kategori pinjaman yang bisa dikategorikan Yes and

No. Rumus standar dari Naïve Bayes adalah Rumus (1) yang digunakan sebagai dasar dari pembuatan model ini. Fungsi ini juga menentukan peluang kedua kelas menggunakan Rumus (4).



Gambar 7. Skema *test on test data* pada fungsi *test and score*

Fungsi *test and score* pada Gambar 6 akan melakukan pengujian pada data latih dan membandingkan dengan labelnya. Pada fungsi ini akan muncul nilai akurasi, *precision*, dan nilai lainnya yang digunakan untuk melakukan evaluasi. Fungsi ini memanfaatkan banyak rumus. Prediksi data menggunakan Rumus (5) untuk data numerik, Rumus (6) untuk data kategorikal, dan Rumus (7) sebagai penentu kelas. Evaluasi data pada fungsi ini memanfaatkan Rumus (8),(9) dan (10). Fungsi ini menggunakan skema *test on test data* seperti pada Gambar 7. Skema tersebut merupakan skema yang memungkinkan untuk melakukan evaluasi berdasarkan 2 dataset yang terpisah. Dataset yang dipisah menjadi 2 memungkinkan evaluasi model menjadi lebih akurat karena tidak terjadi kebocoran data yang membuat proses pengujian menjadi lebih jujur karena karena model tidak mengetahui terlebih dahulu jawaban dari atributnya.

	Naive Bayes	error	Loan_Status	Loan_ID	Naive Bayes	Naive Bayes (N)	Naive Bayes (Y)	Fold	Gender
27	0.19 : 0.81 → Y	0.188	Y	LP002894	Y	0.18824	0.81176	1	Female
28	0.84 : 0.16 → N	0.164	N	LP002911	N	0.836342	0.163658	1	Male
29	0.42 : 0.58 → Y	0.580	N	LP002912	Y	0.419572	0.580428	1	Male
30	0.11 : 0.89 → Y	0.115	Y	LP002916	Y	0.114748	0.885252	1	Male
31	0.35 : 0.65 → Y	0.351	Y	LP002917	Y	0.350523	0.649477	1	Female
32	0.68 : 0.32 → N	0.319	N	LP002926	N	0.680628	0.319372	1	Male
33	0.16 : 0.84 → Y	0.159	Y	LP002928	Y	0.158565	0.841435	1	Male
34	0.16 : 0.84 → Y	0.841	N	LP002931	Y	0.158744	0.841256	1	Male
35	0.23 : 0.77 → Y	0.233	Y	LP002936	Y	0.232744	0.767256	1	Male
36	0.33 : 0.67 → Y	0.326	Y	LP002938	Y	0.326259	0.673741	1	Male

Gambar 8. Fungsi prediksi

Fungsi prediksi pada Gambar 6 yang digambarkan pada Gambar 8 akan menunjukkan hasil prediksi Naïve Bayes. Fungsi ini akan menampilkan seluruh hasil prediksi Naïve Bayes, dan menampilkan seluruh atribut sehingga dapat dilihat potensial baris data tertentu masuk ke kelas apa. Fungsi ini akan menampilkan perbandingan kelas prediksi dimana kelas yang lebih tinggi nilainya akan dijadikan label oleh Naïve Bayes. Fungsi ini akan menampilkan hasil perhitungan Rumus (7) dimana selisih dari perbandingan kedua kelas akan dinyatakan dalam nilai error. Gambar 8 akan memberikan juga kelas prediksi Naïve Bayes yang langsung dibandingkan dengan label pada data uji. Perbandingan ini akan diakumulasikan dan di rangkum pada fungsi *Confusion Matrix* pada Gambar 6 akan melakukan rekap dan visualisasi kategori prediksi pada fungsi *test and score*. Fungsi ini akan menampilkan apakah data tersebut terkategori sebagai prediksi yang benar atau salah. Detail pada fungsi ini dapat dilihat pada Tabel 2 dan 4.

	Naive Bayes	error	Loan_Status	Loan_ID	Naive Bayes	Naive Bayes (N)	Naive Bayes (Y)
1	0.17 : 0.83 → Y	0.172	Y	LP002777	Y	0.17214	0.82786
2	0.18 : 0.82 → Y	0.185	Y	LP002785	Y	0.184677	0.815323
3	0.90 : 0.10 → N	0.097	N	LP002788	N	0.902635	0.0973653
4	0.89 : 0.11 → N	0.114	N	LP002789	N	0.886297	0.113703
5	0.09 : 0.91 → Y	0.089	Y	LP002792	Y	0.088555	0.911445

Gambar 9. Hasil Prediksi Naïve Bayes pada fitur *Predictions*

Gambar 9 adalah sampel evaluasi hasil prediksi Naïve Bayes dari 51 data test. Perhitungan prediksi Naïve Bayes yang pertama kali adalah menghitung probabilitas atribut target. Atribut Target dihitung berdasarkan Rumus (4). Rumus (4) dilakukan 2 kali karena ada 2 jenis kelas target yaitu Y dan N. Penghitungan dilanjutkan dengan menghitung data yang bersifat numerik yang memanfaatkan Rumus (5) perhitungan ini dilakukan pada atribut numerik pada Tabel 1. Proses dilanjutkan dengan menghitung data kategorikal menggunakan Rumus (6) yang dilakukan pada atribut kategorikal Tabel 1 kemudian ditutup dengan Rumus (7) untuk mendapatkan kelas klasifikasinya. Metrik evaluasi dilakukan jika seluruh data sudah dihitung dan mendapatkan nilai klasifikasi yang pada Gambar 6 ditunjukkan pada kolom Naïve Bayes N dan Y. *Confusion matrix* adalah tabel yang digunakan untuk mengevaluasi kinerja suatu model klasifikasi. Tabel ini akan membandingkan hasil prediksi dengan label pada target kemudian akan dikategorikan ke dalam metrik evaluasi sehingga akan terlihat mana kelas yang terprediksi benar dan salah [15]. Prediksi Naïve Bayes yang dilakukan pada Gambar 10 dan divisualisasikan menggunakan Gambar 11 dengan detail pada Gambar 13 akan direkap pada Tabel 3.

Tabel 3. Tabel prediksi Naïve Bayes

Naïve Bayes (Predicted Class)	Label (Actual Class)	Confusion Matrix
Yes	Yes	True Positive
Yes	No	False Positive
No	Yes	False Negative
No	No	True Negative

Berdasarkan Gambar 13, bagian Naïve Bayes merupakan hasil prediksi dari model dan *Loan_Status* merupakan label. Evaluasi Naïve Bayes memanfaatkan *confusion matrix* yang aturannya ada pada Tabel 3 dan menyatukan hasilnya dalam Tabel 4.

Tabel 4. Hasil Prediksi Model dalam *confusion matrix*

CLASSIFICATION		PREDICTED CLASS	
		YES	NO
ACTUAL CLASS	YES	36	0
	NO	7	8

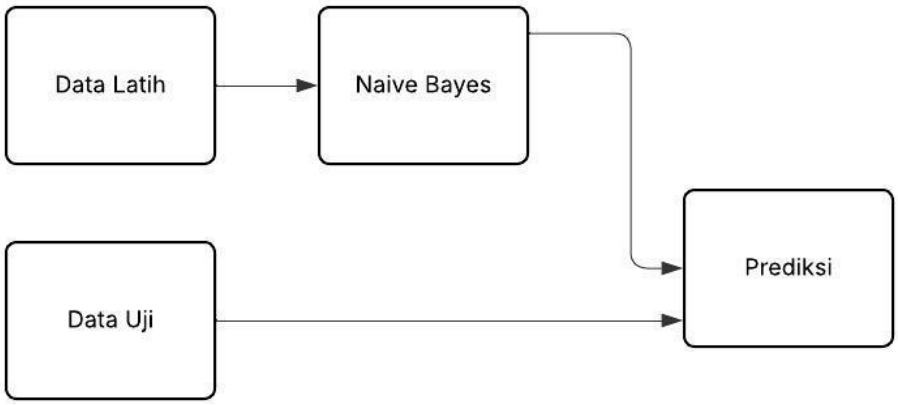
Tabel 4 menjelaskan adanya *true positive* yang menjelaskan tentang kelas benar yang diprediksi sebagai nilai benar (kelas diterima) sebanyak 36 baris. Ada lagi *true negative* dimana menjelaskan tentang kelas salah yang diprediksi sebagai salah (kelas ditolak), ada lagi kelas *false negative* yang menggambarkan kelas yang seharusnya diterima diprediksi sebagai ditolak. Kelas false positive merupakan kelas yang seharusnya ditolak tapi dikategorikan sebagai kelas yang diterima.

Tabel 5. Skor *precision, recall dan accuracy*

Precision	Recall	Accuracy
88%	86%	86%

Confusion Matrix pada Tabel 5 menunjukkan hasil prediksi dengan distribusi *True Positive* (TP) sebanyak 36, *False Negative* (FN) sebanyak 0, *False Positive* (FP) sebanyak 7, dan *True Negative* (TN) sebanyak 8 dari total 51 data uji. Hasil ini menunjukkan bahwa model mampu mengidentifikasi seluruh nasabah yang layak kredit dengan tepat (FN = 0), sehingga tidak ada nasabah potensial yang ditolak secara keliru. Kondisi ini sangat penting bagi bank, karena mengurangi risiko kehilangan calon debitur yang sebenarnya sehat secara finansial.. False positive sendiri terdapat 7 kasus di mana sistem mengklasifikasikan nasabah sebagai layak kredit padahal sebenarnya tidak. Dari sisi perbankan, hal ini dapat membuat resiko terjadinya kredit macet alias gagal bayar. Kelas FP tergolong relatif kecil yang

masih dalam batas wajar untuk model yang termasuk sederhana. Metrik evaluasi yang dihitung menggunakan Rumus (8), (9), dan (10) menghasilkan nilai akurasi 86%, precision 88%, dan recall 86%. Nilai precision yang tinggi menunjukkan bahwa sebagian besar nasabah yang dinyatakan layak oleh sistem memang benar-benar memenuhi kriteria, sementara nilai recall yang tinggi memastikan hampir semua nasabah layak dapat terdeteksi. Kombinasi ini menjadikan model Naïve Bayes cukup efektif untuk digunakan sebagai *screening tool* awal, yang dapat membantu debitur untuk membuat keputusan yang lebih cepat dan efisien.



Gambar 10. Skema Naïve Bayes untuk memprediksi data tanpa label

Perbandingan skema pada Gambar 6 dan Gambar 9 terlihat pada tidak adanya fungsi *test and score* dan *confusion matrix*. Kedua fungsi tersebut tidak bisa digunakan apabila data tidak memiliki label. Skema ini dibuat untuk membuktikan bahwa model ini bisa melakukan prediksi walaupun tidak memiliki label data. Fungsi keseluruhan pada Gambar 10 Sama persis dengan Gambar 6.

	Naive Bayes	oan_Statu	Loan_ID	Gender	Married	pende	Education	f_Employ	pplicantIncor	pplicantInc	anAmou	oan_Amount_Terr	redit_Histo	Property_Area
1	Y	?	LP002893	Male	No	0	Graduate	No	1836	33837.0	90.0	360.0	1.0	Urban
2	Y	?	LP002912	Male	Yes	1	Graduate	No	4283	3000.0	172.0	84.0	1.0	Rural
3	N	?	LP002911	Male	Yes	1	Graduate	No	2787	1917.0	146.0	360.0	0.0	Rural
4	Y	?	LP002916	Male	Yes	0	Graduate	No	2297	1522.0	104.0	360.0	1.0	Urban
5	Y	?	LP002892	Male	Yes	2	Graduate	No	6540	0.0	205.0	360.0	1.0	Semiurban
6	Y	?	LP002894	Female	Yes	0	Graduate	No	3166	0.0	36.0	360.0	1.0	Semiurban

Gambar 11. Prediksi model untuk data tanpa label

Hasil Prediksi pada Gambar 11 dapat dilihat bahwa atribut *loan_status* tidak memiliki label (?). Namun, kolom Naïve Bayes memberikan prediksinya. Perhitungan di skema ini sama seperti skema yang tertera pada Gambar 6. Skema ini apabila mengacu pada hasil evaluasi yang ada pada Tabel nomor 4, artinya data prediksi pada Gambar 11 tingkat akurasi data adalah 86%.

Pembahasan ini memperkuat penelitian [6] bahwa klasifikasi Naïve Bayes mampu digunakan untuk melakukan analisis pinjaman bank. Pada penelitian sebelumnya, akurasi Naïve Bayes mampu menorehkan akurasi sebesar 81%. Penelitian sebelumnya juga menyatakan bahwa "metode Naïve Bayes mengasumsikan bahwa semua fitur dalam data adalah independen, yang dalam situasi nyata mungkin tidak selalu terpenuhi dan dapat mempengaruhi akurasi prediksi. Kualitas data yang digunakan dalam

analisa ini sangat penting, karena data yang buruk atau tidak representatif dapat menyebabkan prediksi yang tidak akurat…sebaiknya dilakukan evaluasi lebih lanjut dengan data yang lebih besar dan representatif" [6]. Data uji dengan atribut yang lebih banyak mampu meningkatkan akurasi model untuk studi kasus pinjaman bank sebesar 8% dan meningkatkan nilai metrik lain seperti *Precision* sebanyak 20%. Terdapat nilai metrik yang turun yaitu nilai *Recall* sebanyak 12% hal ini dikarenakan perbedaan banyaknya jumlah data uji yang digunakan. Penelitian sebelumnya menggunakan 5 data uji berbanding penelitian ini dengan 51 data uji. Angka *recall* menurun merupakan hal wajar karena data uji yang jauh lebih banyak. Keterbatasan dalam penelitian ini adalah dataset yang tidak seimbang dari kelas klasifikasinya sehingga model cenderung lebih baik dalam mengenali satu jenis kelas saja.Jumlah baris data yang sedikit juga menjadi keterbatasan penelitian karena data latih yang sedikit menyebabkan model tidak terlalu maksimal dalam mengenali kelas. Penelitian ini juga tidak membahas terkait perbandingan dengan algoritma lain yang memungkinkan untuk memberikan hasil yang lebih baik.

5. KESIMPULAN

Pembuatan model klasifikasi kelayakan pinjaman menggunakan Naïve Bayes memanfaatkan 1 Atribut identifier yaitu Loan_ID yang disusul dengan 11 Atribut lain yaitu *Gender, Married, Dependent, Education, Self_Employeed, Applicant Income, Co Applicant Income, Loan_Amount, Loan_Amount_Term, Credit_History, Property_Area* dengan atribut *Loan_Status* sebagai target klasifikasi yang dilengkapi dengan data latih sebanyak 429. Model Naïve Bayes ini dibuat dengan skema *test on test data* yang menghasilkan skor akurasi sebesar 86%, dan metrik lain seperti Tingkat presisi sebesar 88% dan *recall score* sebesar 86%. Skor tersebut dihasilkan dari *confusion matrix* sebagai instrumen evaluasi. Proses evaluasi dari 51 data uji menunjukkan 44 data berhasil diprediksi dengan benar dan 7 data diprediksi secara salah. Model ini juga memungkinkan untuk memprediksi data tanpa label sehingga bisa digunakan untuk membantu menentukan kelayakan pinjaman kreditur dengan hanya mengisi atribut yang dibutuhkan. Penelitian ini berhasil menjawab rumusan masalah, yakni pembangunan model klasifikasi berbasis Naïve Bayes yang dapat digunakan untuk memprediksi status kelayakan pinjaman tanpa atribut target (label) sehingga nasabah baru bisa langsung diprediksi oleh model apakah layak atau tidak. Metrik evaluasi juga memperlihatkan angka yang sangat baik sehingga model ini bisa dikembangkan lebih lanjut untuk digunakan di dunia nyata.

6. SARAN

Pengembangan model lebih lanjut sangat mungkin untuk dilakukan dengan menambahkan atribut lain yang lebih relevan. Perlu adanya pembanding dengan metode klasifikasi lain seperti *decision tree* atau *random forest* yang tujuannya mengetahui model yang lebih baik dalam menentukan klasifikasi. Pengembangan model juga sangat mungkin untuk dilakukan dalam bentuk aplikasi, maupun *website*. Model ini juga bisa ditingkatkan akurasinya dengan menyesuaikan data menggunakan data yang berkualitas dari lembaga keuangan atau menambah jumlah data latih. Model ini hanya sebagai alat bantu dalam memprediksi kelayakan pinjaman kreditur artinya, tidak menjadi satu – satunya dasar keputusan dan seluruh keputusan dikembalikan kepada pihak bank.

.

DAFTAR PUSTAKA

[1] M. E. Purba, A. Z. Situmorang, and M. W. P. Lubis, "Perbandingan Pengklasifikasian Penyakit Parkinson Menggunakan Algoritma Naïve Bayes, Random Forest, dan Regresi Logistik," *Jurnal Sifo Mikroskil*, vol. 26, no. 1, pp. 21–36, Apr. 2025, doi: 10.55601/jsm.v26i1.1466.

[2] E. Martantoh and N. Yanih, "Implementasi Metode Naïve Bayes Untuk Klasifikasi Karakteristik Kepribadian Siswa Di Sekolah MTS Darussa'adah Menggunakan PHP MySQL," *Jurnal Teknologi dan Sistem Informasi (JTSI)*, vol. 3, no. 2, pp. 166–175, Sep. 2022, doi : 10.35957/jtsi.v3i2.

[3] J. K. Sulaiman, A. Syakur, R. P. Putra, and C. Juliane, "Optimalisasi Metode Naïve Bayes Classifier Untuk Prediksi Persetujuan Kredit," *Indonesian Journal of Computer Science Attribution*, vol. 13, no. 1, pp. 1091–1099, 2024, doi: 10.33022/ijcs.v13i1.

[4] H. Hartono, A. Hajjah, and Y. N. Marlim, "Penerapan Metode Naïve Bayes Classifier Untuk Klasifikasi Judul Berita," *Jurnal SimanteC*, vol. 12, no. 1, pp. 37-46, Des. 2023, doi : 10.21107/simantec.v12i1.19398.

[5] R. Vincent, I. Maulana, and O. Komarudin, "Perbandingan Klasifikasi Naïve Bayes dan Support Vector Machine dalam Analisis Sentimen dengan Multiclass di Twitter," *JATI (Jurnal Mahasiswa Teknik Informatika)*, vol. 7, no. 4, pp. 2496-2505, Agu. 2023, doi: 10.36040/jati.v7i4.7152.

[6] S. Lestari and D. Mayasari, "Analisa Penerapan Metode Naïve Bayes Dalam Memprediksi Kelayakan Calon Nasabah Dalam Melakukan Pinjaman," *Jurnal Elektronika dan Teknik Informatika Terapan (JENTIK)*, vol.1,no. 2, pp. 30–37, Jun. 2023, doi: 10.59061/jentik.v1i2.351.

[7] A. Astofa and E. Sutono, "Implementasi Algoritma Naïve Bayes Untuk Memprediksi Kelayakan Kredit Nasabah," *LOGIC : Jurnal Ilmu Komputer dan Pendidikan*, vol. 2, no. 5, pp. 766–775, Agu. 2024.

[8] T. Ardiyanto and H. W. Nugroho, "Penentuan kelayakan penerima pembiayaan kredit menggunakan metode Naïve Bayes dan algoritma C4.5 Studi kasus KSPPS. BMT adil berkah sejahtera," *Jurnal SIMADA (Sistem Informasi dan Manajemen Basis Data)*, vol. 6, no. 2, pp. 1–11, Mar. 2023, doi : 10.30873/simada.v6i2.

[9] H. Susana and N. Suarna, "Penerapan model klasifikasi metode Naïve Bayes terhadap penggunaan akses internet," *Jursistekni (Jurnal Sistem Informasi dan Teknologi Informasi)*, vol. 4, no. 1, pp. 1–8, 2022, doi : 10.52005/jursistekni.v4i1.96.

[10] M. R. B. Keliat *et al.*, "Komparasi Algoritma Support Vector Machine dan Naïve Bayes pada Klasifikasi Jenis Buah Kurma berdasarkan Citra Hue Saturation Value," *Sistemasi: Jurnal Sistem Informasi*, vol. 14, no. 1, pp. 470–481, Feb. 2025, doi : 10.32520/stmsi.v14i1.4978.

[11] W. E. Nugroho, A. Sofyan, and O. Somantri, "Metode Naïve Bayes Dalam Menentukan Program Studi Bagi Calon Mahasiswa Baru," *Infotekmesin*, vol. 12, no. 1, pp. 59–64, Mar. 2021, doi: 10.35970/infotekmesin.v12i1.491.

[12] A. Jalil, A. Homaidi, and Z. Fatah, "Implementasi Algoritma Support Vector Machine Untuk Klasifikasi Status Stunting Pada Balita," *G-Tech: Jurnal Teknologi Terapan*, vol. 8, no. 3, pp. 2070–2079, Jul. 2024, doi: 10.33379/gtech.v8i3.4811.

[13] K. Nurnasikha, S. Farisa, and C. Haviana, "Klasifikasi Bidang Ilmu Publikasi Ilmiah Terindeks SINTA Menggunakan Metode Naïve Bayes," *Jurnal Transistor Elektro dan Informatika (TRANSISTOR EI)*, vol. 5, no. 3, pp. 147–154, 2023.

[14] W. I. Rahayu, C. Prianto, and E. A. Novia, "Perbandingan algoritma K-means dan naïve bayes untuk memprediksi prioritas pembayaran tagihan rumah sakit berdasarkan tingkat kepentingan pada PT. Pertamina (Persero)," *Jurnal Teknik Informatika*, vol. 13, no. 2, pp. 1–8, Apr. 2021.

[15] L. Hakim, A. Sobri, L. Sunardi, and D. Nurdiansyah, "Prediksi penyakit jantung berbasis mesin learning dengan menggunakan metode k-nn," *Jurnal Digital dan Teknologi Informasi*, vol. 7, no. 2, pp. 14–20, 2024, doi: 10.32502/digital.v7i2.9429.