
PERBANDINGAN KLASIFIKASI DATA MINING UNTUK PREDIKSI KREDIT MACET DENGAN MENGGUNAKAN METODE NAÏVE BAYES DAN K – NEAREST NEIGHBORS (STUDI KASUS: BANK BUMN DI DAERAH JAWA BARAT)

Septiano Anggun Pratama

Fakultas Teknik, Program Studi S1 Teknik Informatika

Universitas Tadulako

Email: septiano@untad.ac.id

Arief Wibowo

Fakultas Teknologi Informasi, Program Studi Magister Ilmu Komputer

Universitas Budi Luhur

Email: arief.wibowo@budiluhur.ac.id

ABSTRAK

Bank merupakan entitas bisnis yang mengumpulkan dana dari masyarakat melalui tabungan dan menyalurkannya kembali dalam bentuk pinjaman guna meningkatkan kesejahteraan. Layanan kredit dengan agunan memang umum dijumpai, namun seringkali menghadirkan risiko tinggi akibat analisis kredit yang kurang teliti saat dilakukan secara manual. Untuk mengatasi permasalahan kredit macet atau gagal bayar tersebut, diperlukan sistem komputer yang mampu mempelajari data historis peminjam melalui penerapan teknik data mining guna mengungkap nilai tambah dari data yang sebelumnya tidak teridentifikasi. Berdasarkan latar belakang tersebut, penelitian ini bertujuan untuk membandingkan dua metode klasifikasi data mining, yakni Naïve Bayes dan K-Nearest Neighbors (KNN), dalam memprediksi nasabah yang berpotensi mengalami keterlambatan pembayaran kredit. Metode penelitian dilakukan melalui tahap praproses data, pembagian dataset (80% data training dan 20% data testing), serta evaluasi model menggunakan confusion matrix untuk mengukur akurasi, presisi, dan recall. Hasil penelitian menunjukkan bahwa algoritma Naïve Bayes mencapai tingkat akurasi sebesar 70,59%, sedangkan algoritma KNN menghasilkan akurasi sebesar 93,14%. Temuan ini mengindikasikan bahwa metode KNN lebih efektif dalam mengidentifikasi nasabah berpotensi kredit macet, sehingga dapat dijadikan dasar dalam pengambilan keputusan kredit di sektor perbankan.

Kata kunci: sistem informasi, kredit macet, bank, data mining, naïve bayes, K – NN

ABSTRACT

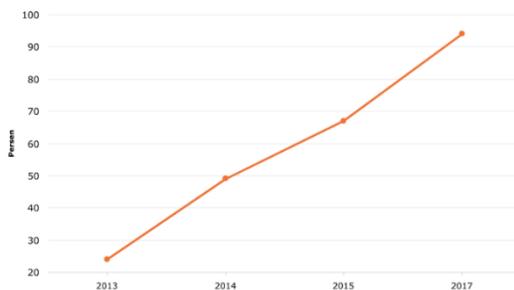
Banks are business entities that collect funds from the public through savings and channel them back as loans to improve welfare. Although credit services secured by collateral are common, they often entail high risks due to insufficient and less careful credit analysis when performed manually. To overcome the problem of bad loans or defaults, a computer system capable of learning historical borrower data through the application of data mining techniques is required to uncover hidden value in the data. Based on this background, this study aims to compare two data mining classification methods, namely Naïve Bayes and K-Nearest Neighbors (KNN), in predicting customers who are likely to experience delayed credit payments. The research methodology involves data preprocessing, partitioning the dataset into 80% training data and 20% testing data, and evaluating the models using a confusion matrix to measure accuracy, precision, and recall. The results show that the Naïve Bayes algorithm achieved an accuracy of 70.59%, while the KNN algorithm produced an accuracy of 93.14%. These findings indicate that the KNN method is more effective in identifying potential defaulters, thereby providing a sound basis for credit decision-making in the banking sector.

Keywords: information system, bad credit, bank, data mining, naïve bayes, K – NN.

1. PENDAHULUAN

Dalam dunia perbankan, layanan kredit dengan agunan atau jaminan merupakan hal yang sangat biasa dimasyarakatkan meski sering memberikan resiko tinggi. Kredit adalah suatu fasilitas keuangan yang memungkinkan seseorang atau badan usaha untuk meminjam uang untuk membeli produk dan membayarnya kembali dalam jangka waktu yang ditentukan [8].

Kredit yang mengalami masalah (kredit macet) seringkali disebabkan oleh analisis kredit yang kurang cermat, tidak hati-hati, atau bahkan karena karakter nasabah yang kurang baik. Hal ini tersebut jelas sangat mencemaskan. Tercatat melalui survey yang dilakukan oleh Pricewaterhouse Cooper (PWC) – sebuah lembaga *audit and assurance, consulting and tax services* – pada tahun 2017, dilakukan penelitian mengenai pendapat para pemimpin industri perbankan di Indonesia. Hasilnya mencatat bahwa *Non-Performing Loan* (NPL) memang sangatlah mencemaskan dan menjadi tantangan buat perkembangan dunia perkreditan bank di Indonesia saat ini.



Gambar 1. Survei NPL oleh Pricewaterhouse Cooper

Hasilnya, berdasarkan gambar 1 di atas sebanyak 94% bankir meyakini NPL akan menjadi tantangan perbankan global saat ini. Survei tersebut melibatkan 78 responden dari 58 bank di Indonesia. Responden dalam penelitian ini menduduki posisi manajemen di bank mana pun. Secara total, mereka mewakili 87% aset perbankan di Indonesia.

Jenis bank yang berbeda menghadapi banyak risiko kredit yang berbeda, beberapa di antaranya mempunyai implikasi sosial karena jumlah dan keragaman pemangku kepentingan yang mempengaruhi mereka yang mempengaruhi pendapatan perusahaan kepada pemegang saham seperti manajer, pemberi pinjaman (bank), pemasok, pelanggan, lembaga keuangan, dan pemerintah, pesaing, dan badan pengatur. Jadi, untuk mengelola masalah ini dengan baik, penting untuk dimiliki sistem yang mendukung keputusan yang kuat dan canggih yang didukung oleh alat analisis, untuk mengukur, memonitor, mengelola, dan kontrol, risiko dan inefisiensi keuangan dan operasional [7]. Demi mengatasi masalah itu semua, dibutuhkanlah sebuah sistem komputer yang dapat mempelajari data historis dari peminjam kredit. Solusi yang mungkin dilakukan adalah dengan menerapkan metode data mining sebagai metode sistematis untuk mengetahui nilai tambah pengumpulan data berupa pengetahuan informal.

Fungsi dari data mining ialah untuk melakukan klasifikasi. Klasifikasi adalah proses pencarian sekumpulan model atau fungsi yang menggambarkan dan membedakan kelas data dengan tujuan agar model tersebut dapat digunakan untuk memprediksi kelas dari suatu objek yang belum diketahui kelasnya. Selain itu, klasifikasi Naïve Bayes terbukti memiliki akurasi dan kecepatan yang tinggi saat diaplikasikan ke dalam basis data dengan jumlah yang besar [5]. Naïve Bayes merupakan salah satu algoritma yang memiliki tingkat akurasi tinggi dibandingkan dengan algoritma klasifikasi lain. Bentuk klasifikasi yang dapat digunakan dalam memprediksi nasabah yang berpotensi terlambat membayar kredit. Bentuk klasifikasi tersebut disebut Naïve Bayes Classifier [3]. Prediksi tersebut memanfaatkan riwayat persetujuan kredit anggota koperasi sebagai data dasar untuk menemukan pola atau hubungan teratur dalam tingkat kelancaran pembayaran kredit oleh nasabah bank XYZ. Proses ini disebut data mining [1]. Dalam penelitian ini penulis membandingkan algoritma Naïve Bayes dengan k -Nearest Neighbor untuk melakukan prediksi kredit macet pada Bank XYZ.

Data mining juga bisa dikatakan proses ekstraksi data yang valid, dimana sebelumnya ada informasi yang sangat berharga, lebih menyeluruh namun belum diketahui, hal ini diperlukan usaha yang lebih besar

untuk menemukannya dari suatu database dimana informasi yang diperoleh dapat digunakan untuk membuat kebijakan bisnis yang penting [11].

Dalam penelitian sebelumnya yang dilakukan oleh Syifa Sintia Al Kautsar, dkk yaitu penelitian "Algoritma Naïve Bayes Untuk Memprediksi Kredit Macet Pada Koperasi Simpan Pinjam"[12]. Dalam penelitian tersebut, peneliti memproyeksikan risiko kredit yang tidak dapat dilunasi dan menilai kelayakan peminjam untuk mendapatkan pinjaman dengan memanfaatkan teknik data mining. Metode yang digunakan untuk data mining adalah klasifikasi dengan metode Naïve Bayes. Hasil pengujian akurasi model menunjukkan tingkat akurasi sebesar 59%.

Untuk memenangkan persaingan bisnis dan meningkatkan pendapatan perusahaan, maka para pemimpin dan manajemen perusahaan harus mampu mengambil keputusan yang tepat untuk menentukan strategi pemasaran. Untuk mencapai hal tersebut, dunia usaha memerlukan pusat informasi yang komprehensif untuk dapat melakukan analisis yang mendalam.

Berdasarkan penelitian lain yang dilakukan oleh Fachry Husaini yaitu penelitian "Algoritma Klasifikasi Naïve Bayes Untuk Menilai Kelayakan Kredit (Studi Kasus: Bank Mandiri Kredit Mikro)" yang dimana peneliti merancang dan membangun sistem dalam menentukan kelayakan pemberian kredit terhadap nasabah. Berdasarkan hasil penelitian membuktikan bahwa algoritma Naïve Bayes dapat diterapkan untuk menilai kelayakan kredit pada Mandiri Kredit Mikro. Pengolahan data awal merupakan tahapan yang sangat mempengaruhi hasil akurasi yang baik sehingga akurasi akhir yang dihasilkan termasuk kategori *excellent* [9].

Sementara itu penelitian sebelumnya yang dilakukan oleh Aida Indriani tentang "Analisa Perbandingan Metode Naïve Bayes Classifier dan K – Nearest Neighbor Terhadap Klasifikasi Data" ialah membandingkan kedua buah metode algoritma, yang dimana pada penelitian tersebut disimpulkan bahwa metode k – NN lebih baik tingkat akurasinya sebesar 80% disbanding metode Naïve Bayes sebesar 73% yang dihitung menggunakan *confusion matrix* [10].

Kemudian pada penelitian yang dilakukan oleh Abdul Rohman, Model Algoritma k – Nearest Neighbor (K-NN) Untuk Prediksi Kelulusan Mahasiswa. Pada penelitian ini memiliki tingkat accuracy 82,25%. Tetapi dengan menggunakan cluster data k = 5 accuracy yang dihasilkan ialah 85,15% dan nilai AUC (Area Under Curve) adalah 0,888 [15].

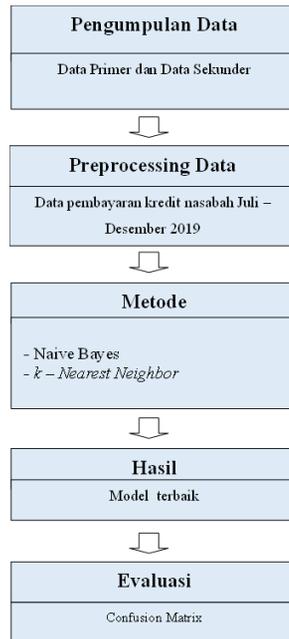
Sementara di sisi lain, penelitian terdahulu yang dilakukan oleh Yoga Religia tentang, "Analisis Algoritma K-Nearest Neighbor dan Naïve Bayes untuk klasifikasi Pembelian Mobil" yang menggunakan dataset global yang terdiri dari enam atribut dependent dan satu atribut independent. Akurasi yang dihasilkan menunjukkan untuk k- Nearest Neighbor 87.81% sedangkan untuk Naïve Bayes menunjukkan persentase 93.99% [13].

Secara umum data mining dapat dikelompokkan menjadi dua metode utama, yaitu: 1. *Descriptive mining*, yaitu proses pencarian ciri-ciri penting data dalam database. Teknik penambangan data yang termasuk dalam penambangan deskriptif adalah *clustering*, asosiasi, dan *sequential mining*. 2. *Predictive*, yaitu teknik penambangan data untuk menemukan pola dari data dengan menggunakan beberapa variabel di masa yang akan datang. Salah satu teknik dalam penambangan prediktif adalah klasifikasi.

Perbandingan algoritma melibatkan evaluasi dua metode untuk menentukan akurasi yang paling sesuai di antara keduanya. Dalam penelitian ini, penulis akan membandingkan dua algoritma, yaitu metode k-Nearest Neighbor dan Naïve Bayes. Parameter perbandingan ke dua metode tersebut adalah akurasi sistem dari ke dua metode tersebut.

Dari uraian diatas akan dilakukan analisis terhadap data nasabah untuk memperoleh informasi dengan membandingkan beberapa atribut dari nasabah. Penelitian ini nantinya akan menghasilkan prediksi untuk kedepannya apakah nasabah dapat menyebabkan kredit macet atau tidak berdasarkan hasil analisis kriteria masing-masing nasabah.

2. METODOLOGI PENELITIAN



Gambar 2 Tahapan Penelitian

Penelitian ini menggunakan metode eksperimen komparatif yaitu dengan mengoptimasikan dua algoritma antara lain algoritma Naïve Bayes dan algoritma k – Nearest Neighbor.

Jika dilihat dari jenis data dan informasi yang dikuasai, maka penelitian ini tergolong penelitian kuantitatif. Penelitian kuantitatif merupakan penelitian yang hipotesisnya dapat diuji dengan menggunakan metode matematika. Metode ini digunakan saat melakukan uji kualitas, dengan metode confusion matrix yang menunjukkan nilai akurasi untuk setiap metode jika dibandingkan dengannya.

Berdasarkan Gambar 2 di atas, penelitian ini menggunakan data pembayaran kredit. Data pembayaran kredit tersebut akan diolah menggunakan metode Naïve Bayes dan k – Nearest Neighbor. Dalam penelitian ini akan dilakukan beberapa langkah yang dilakukan dalam eksperimen mengikuti metodologi *Cross - Standard Industry Process for Data Mining (CRISP-DM)* [4], sebagai berikut:

a) Pemahaman Bisnis

Pada tahapan pertama diperlukan pemahaman mengenai konteks dan tujuan proses bisnis terkait penyaluran kredit, antara lain:

1. Menentukan tujuan bisnis (*determine business objectives*)
Tujuan bisnis yang ingin dicapai Bank BJB Cabang Cipanas adalah meminimalkan risiko terjadinya kredit macet.
2. Menilai situasi (*asses the situation*)
Pemberian pinjaman kepada nasabah melibatkan informasi pribadi nasabah, data aplikasi kredit, dan rekam jejak pembayaran kredit.
3. Tujuan Data Mining
Tujuan dilakukannya data mining adalah mengimplementasi metode Naïve Bayes dan k – Nearest Neighbors agar kinerjanya lebih akurat dalam memprediksi kredit macet pada Bank.

b) Pemahaman Data

Dalam pemahaman data, banyak hal yang dilakukan antara lain pengumpulan data primer, interpretasi data, analisis data, dan pengukuran kualitas data terkait proses bisnis. Permasalahan yang ada di Bank BJB Kantor Cabang Cipanas adalah tidak adanya suatu metode dalam melakukan prediksi data nasabah yang akan melakukan kredit macet [2]. Terkait dengan masalah tersebut, Peneliti mengumpulkan data yang berkaitan dengan pembuatan prediksi. Proses ini dilakukan dengan melakukan wawancara dan observasi di lapangan. Pembahasan dilakukan dengan departemen yang mempunyai kewenangan

untuk menentukan nasabah mana saja yang mempunyai kredit macet.

c) **Persiapan Data**

Dalam fase persiapan data, terdapat beberapa langkah yang dilakukan, termasuk deskripsi data set, pemilihan data, pembangunan data, integrasi data, dan pembersihan data. Data yang diambil hanya data Desember tahun 2019 yang akan dikumpulkan sebagai data training dan data testing dengan pembagian dataset sebesar 80% dan 20%. Dalam tahap persiapan *data* ini, peneliti melaksanakan wawancara dengan staf *database* untuk memahami lebih dalam data dan menentukan metode prediksi penilaian yang tengah diimplementasikan. Adapun *field* dari data training yaitu, *no_cif*, *jns_klmn*, *jlwaktu_bln*, *xangsur*, *angske*, *xtgkp*, *xtgkb*, *hr_tgkp*, *hr_tgkb*, *kreditke*

d) **Pemodelan**

Pada langkah pemodelan ini, terjadi perbandingan, pemilihan, dan implementasi teknik pemodelan klasifikasi data mining, yakni Naïve Bayes dan k-Nearest Neighbor. Data yang telah diolah pada tahap pengolahan data akan digunakan dalam proses ini.

e) **Evaluasi**

Evaluasi dilakukan untuk memverifikasi model. Evaluasi dilakukan untuk mengetahui tingkat akurasi sebenarnya dan menguji signifikansi perbedaan algoritma Naïve Bayes dengan k – Neighbors menggunakan metode *confusion matrix* [6]. Hasil uji *confusion matrix* didapatkan suatu model yang signifikan lebih akurat kinerjanya untuk memprediksi kredit macet Bank.

3. HASIL DAN PEMBAHASAN

3.1. Pengumpulan Data

Eksperimen yang dilakukan dalam penelitian ini didasarkan pada data yang dikumpulkan melalui wawancara, observasi dan makalah penelitian untuk memperoleh metode yang diperlukan untuk digunakan dalam memecahkan masalah yang diperlukan. Data yang terkumpul diolah dan dikembangkan secara rinci dengan menggunakan pilihan yang diberikan penulis untuk membantu mengembangkan proses dan merancang sistem yang diinginkan. Data uji yang diambil yaitu bulan Desember 2019 sebanyak 508 *record*. Data yang diperoleh dikumpulkan sebagai *data training* dan *data testing* yang dimana pembagian dataset sebanyak 80% dan 20% dengan 1 atribut kelas dan 9 atribut prediktor. Atribut kelas pada dataset kredit dengan agunan bank XYZ adalah status pembayaran kredit, sedangkan atribut prediktornya terdiri dari, nomor rekening nasabah, jenis kelamin, jangka waktu kredit (bulan), berapa kali angsuran, angsuran yang sudah dibayar, berapa kali tunggakan pokok, berapa kali tunggakan bunga, berapa hari tunggakan pokok, berapa hari tunggakan bunga, kredit yang telah disetujui, dan kolek.

3.2. Pra – Pemrosesan Data

Pada tahapan pra – pemrosesan data dilakukan ialah, mengolah data awal [18], mengeksplorasi data, mendeskripsikan data, dan Memeriksa kualitas data yang terkait dengan proses bisnis dan melakukan pembersihan data. Dalam konteks permasalahan yang telah dikumpulkan oleh peneliti di Bank BJB Kantor Cipanas, pendekatan dilakukan melalui serangkaian wawancara dan observasi di objek penelitian lokal. Wawancara dilaksanakan dengan divisi yang memiliki kewenangan untuk menentukan nasabah yang mengalami kredit bermasalah. Data yang terkumpul melibatkan beberapa bidang dengan total atribut sebanyak 30, yaitu seperti pada tabel 1 berikut:

Tabel 1. Atribut Data Master

<i>No.</i>	<i>Atribut</i>	<i>Keterangan</i>
1.	<i>NOREKENING</i>	<i>Nomor rekening nasabah</i>
2.	<i>NO_CIF</i>	<i>Nomor informasi file customer</i>

3.	ALAMAT	Alamat nasabah
4.	JNS_KLMN	Jenis kelamin
5.	NOTELP1	Nomor telepon nasabah 1
6.	NO TELP2	Nomor telepon nasabah 2
7.	TGL_LAHIR	Tanggal lahir nasabah
8.	KODEPOS	Kode pos nasabah
9.	TGL_BUKA	Tanggal buka kredit nasabah
10.	TGL_JT	Tanggal jatuh tempo atau kredit berakhir
11.	JWAKTU_BLN	Jangka waktu kredit dalam bulan
12.	PLAFOND	Jumlah kredit
13.	XANGSUR	Sisa angsuran nasabah
14.	ANGSKE	Angsuran yang sudah dibayar
15.	ANGSURPK	Nominal angsuran pokok
16.	ANGSURBNG	Nominal angsuran bunga
17.	KREDITKE	Kredit yang telah disetujui
18.	NOPK	Nomor perjanjian kredit
19.	XTGKP	Jumlah tunggakan pokok dalam bulan
20.	XTGKB	Jumlah tunggakan bunga dalam bulan
21.	HR_TGKP	Jumlah hari tunggakan pokok
22.	HR_TGKB	Jumlah hari tunggakan bunga
23.	REFERENSI	Referensi jaminan nasabah
24.	PEMILIK	Nama pemilik jaminan
25.	ALMT_PEMLK	Alamat pemilik jaminan
26.	JENIS_JAM	Jenis jaminan
27.	NILAIGUN	Estimasi nilai jaminan
28.	KETERANGAN	Keterangan tambahan dari jaminan
29.	TGL_APRSL	Tanggal penaksiran dari jaminan
30.	_KOLEK	Status pembayaran

a) Atribut yang digunakan

Selama analisis ini, proses verifikasi dilakukan untuk mengidentifikasi dan mengubah data sehingga dapat digunakan dalam proses entri data algoritmik dan mencapai akurasi dan kinerja yang lebih baik. Pada dataset, validasi data digunakan untuk menghapus data yang tidak lengkap atau hilang tanpa nilai (null). Setelah itu dilakukan seleksi untuk memilih atribut yang relevan dari dataset yang digunakan dalam proses analisis kredit macet pelanggan. Atribut yang diambil dari data nasabah bulan Desember yaitu: jenis kelamin/gender, jangka waktu kredit (bulan), sisa angsuran nasabah, angsuran yang sudah dibayarkan, jumlah tunggakan pokok, jumlah tunggakan bunga, jumlah hari tunggakan pokok, jumlah hari tunggakan bunga, dan jumlah kredit yang telah disetujui. Adapun untuk keterangan lebih lengkapnya dapat dilihat pada tabel 2 berikut ini:

Tabel 2. Atribut Data Penilaian

No.	Atribut	Nilai Atribut
1.	Jenis Kelamin	Laki – Laki Perempuan
2.	Jangka Waktu Kredit (Bulan)	0 s/d 72 Bulan
3.	Sisa Angsuran Nasabah (XANGSUR)	0 s/d 24 Bulan
4.	Angsuran yang Sudah Dibayar (ANGSKE)	0 s/d 25 Bulan
5.	Jumlah Tunggakan Pokok (XTGKP)	0 s/d 24 Bulan
6.	Jumlah Tunggakan Bunga (XTGKB)	0 s/d 24 Bulan
7.	Jumlah Hari Tunggakan Pokok (HR_TGKP)	0 s/d 1200 Hari
8.	Jumlah Hari Tunggakan Bunga (HR_TGKB)	0 s/d 1200 Hari
9.	Kredit yang Telah Disetujui (KREDITKE)	1 s/d 12 Kali

b) Atribut Target / Kolektabilitas

Adapun atribut target pada tabel di bawah ini atau status kredit nasabah yang digunakan pada penelitian ini terbagi menjadi 5 kategori, yaitu: Class 1 = Lancar; Class 2; Dalam Perhatian Khusus; Class 3 = Kurang Lancar; Class 4 = Diragukan; Class 5 = Macet. Dalam menentukan kredit macet atau tidak, ketika nasabah memiliki status class 1 – class 2, masih dikategorikan kredit lancar, akan tetapi jika telah berada di class 3 hingga class 5 dikategorikan kredit macet.

Tabel 3. Atribut Target

<i>Kategori</i>	<i>Keterangan</i>
<i>1</i>	<i>Lancar</i>
<i>2</i>	<i>Dalam Perhatian Khusus</i>
<i>3</i>	<i>Kurang Lancar</i>
<i>4</i>	<i>Diragukan</i>
<i>5</i>	<i>Macet</i>

3.3. Pemodelan dengan Algoritma K-NN (K-Nearest Neighbors)

Setelah ditentukan jumlah data, atribut prediktor, dan atribut target, langkah selanjutnya ialah melakukan pemodelan pada data yang sudah digunakan menggunakan algoritma Naïve Bayes dan algoritma k - Nearest Neighbor. Pemodelan algoritma K-NN melakukan perhitungan jarak menggunakan *Euclidean Distance* pada Persamaan 1 di bawah ini.

$$d(x, y) = \sqrt{\sum_{i=1}^n (x_i - y_i)^2} \tag{1}$$

Pada Persamaan 1, Perhitungan dilakukan menggunakan 15 *dataset* terpilih untuk diteruskan ke proses *sorting*. Selanjutnya menggunakan persamaan *Euclidean Distance* untuk mencari jarak dari yang terkecil hingga terbesar pada *dataset* menggunakan rumus di bawah ini.

$$d(\text{sample}, \text{uji}) = \sqrt{(\text{sample}_{\text{attribute1}} - \text{uji}_{\text{attribute1}})^2 + \dots + \sqrt{(\text{sample}_n - \text{uji}_n)^2}} \tag{2}$$

Berdasarkan Persamaan 2, digunakan untuk mencari hasil perhitungan jarak seluruh dataset, didapat nilai *Euclidean Distance* setiap dataset yang telah di urutkan dari yang terkecil hingga terbesar dengan contoh perhitungan *data testing* pada Tabel 4 di bawah ini.

Tabel 4. Contoh Data Testing

<i>NO_CIF</i>	<i>JNS KLMN</i>	<i>JWAKTU BLN</i>	<i>XANGSUR</i>	<i>ANGS KE</i>	<i>XTG KP</i>	<i>XTG KB</i>	<i>HRTG KP</i>	<i>HRTG KB</i>	<i>KREDIT KE</i>
<i>005675</i>	<i>1</i>	<i>24</i>	<i>3</i>	<i>4</i>	<i>0</i>	<i>0</i>	<i>0</i>	<i>0</i>	<i>1</i>

Berdasarkan contoh data testing pada Tabel 4 dilakukan proses perhitungan Euclidean Distance dari 406 dataset dan nilai tersebut di sorting dari terkecil hingga terbesar, maka selanjutnya dilakukan proses split dataset dengan mengambil 15 dataset dengan nilai Euclidean Distance terkecil. Hasil split dataset dapat dilihat pada Tabel 5 berikut.

Tabel 5. Hasil Split Dataset

<i>NO CIF</i>	<i>EUCLIDEAN DISTANCE</i>	<i>KOLEKTABILITAS</i>
<i>005570</i>	<i>0</i>	<i>1</i>
<i>000836</i>	<i>0</i>	<i>1</i>

005594	0	2
005547	1	1
005646	1	1
005558	1	2
005484	2	5
005001	3	1
005494	3	1
004684	3	3
005752	3	1
005024	3	1
003409	4	2
003711	4	1
000821	6	1

Berdasarkan Tabel 5, dari hasil split dataset diatas maka dihitung Majority setiap class kolek. Jumlah class yang muncul paling banyak adalah hasil dari klasifikasi. Hasil perhitungan Majority dapat dilihat pada Tabel 6 di bawah ini.

Tabel 6. Hasil Perhitungan Majority

<i>Class Kolek</i>	<i>Jumlah</i>
1	10
2	3
3	1
5	1

Berdasarkan Tabel 6, hasil perhitungan tabel majority, didapat bahwa class kolek 1 memiliki jumlah terbanyak dengan muncul 10 kali. Maka dari hasil klasifikasi kredit macet didapat bahwa hasilnya adalah class kolek 1.

Untuk *data test* yang ke-2 sampai data test ke-102 dilakukan proses yang sama dari hitung jarak dengan Euclidean distance hingga menghitung majority. Hasil prediksi seluruh data testing dengan metode KNN ialah terdapat 73 data dengan kategori Kredit LANCAR (kolek 1), 14 data dengan kategori kredit DALAM PERHATIAN KHUSUS (kolek 2), 7 data dengan kategori kredit KURANG LANCAR (kolek 3), 5 data dengan kategori kredit DIRAGUKAN (kolek 4), dan 3 data dengan kategori KREDIT MACET (kolek 5).

3.4. Pemodelan dengan *Algoritma Naïve Bayes*

Selanjutnya dilakukan pemodelan menggunakan algoritma Naïve Bayes yaitu menghitung probabilitas setiap kelas berdasarkan 406 dataset dan 102 data uji yang telah dimiliki dengan menggunakan Persamaan 3 di bawah ini.

$$P(C_i) = \frac{\text{Jumlah kelas}_i}{\text{Total Sampel}} \quad (3) [14]$$

Pada Persamaan 3 di atas, akan diimplementasikan dengan data testing pada Tabel 4 di atas. Dengan menggunakan rumus probabilitas kelas (Ci) maka penerapannya untuk 406 dataset adalah sebagai berikut yang dapat dilihat hasilnya pada Tabel 7 berikut.

Tabel 7. Hasil Perhitungan Probabilitas Class (Ci)

<i>Class Kolek</i>	<i>Jumlah</i>	<i>Probabilitas Class (Ci)</i>
1	382	0.94088669950739
2	10	0.024630541871921
3	2	0.0049261083743842
4	1	0.0024630541871921
5	11	0.027093596059113

Berdasarkan Tabel 7, didapatkan hasil sebanyak 382 dataset yang berada pada Class KOLEK 1, 10 dataset berada pada Class KOLEK 2, 2 dataset pada Class KOLEK 3, dan 1 dataset berada pada Class KOLEK 4. Tahap kedua yaitu menghitung probabilitas attribute setiap kelas menggunakan Persamaan 4 berikut ini.

$$P(X_i) = \frac{\text{Jumlah Attribute Kelas}_i \text{ yang sama dengan Attribute Uji}_i}{\text{Jumlah Kelas}_i} \quad (4) [17]$$

Dengan menggunakan Persamaan 4 diatas, maka di dapat hasil probabilitas atribut setiap kelas yang dapat dilihat pada Tabel 8 di bawah ini.

Tabel 8. Tabel persamaan Atribut dataset dengan data uji

Class	JNS	JWAKTU	XANGSUR	ANGSKE	XTGKP	XTGKB	HR	HR	KREDIT
	KLMN	BLN					TGKP	TGKB	
1	260	97	38	39	329	328	329	328	207
2	9	1	2	1	0	0	0	0	4
3	2	0	0	0	0	0	0	0	0
4	1	0	0	0	0	0	0	0	0
5	10	3	1	0	0	0	0	0	8

Setelah mencari persamaan setiap atribut pada Tabel 8, maka di hitung nilai probabilitas setiap atribut menggunakan Persamaan 5 sebagai berikut:

$$(\text{Attribute}_i) = \frac{\text{Jumlah Attribute}_i}{\text{Jumlah Kelas}_i} \quad (5)$$

Hasil perhitungan dapat dilihat pada Tabel 9 di bawah ini.

Tabel 9. Probabilitas Setiap Atribut

Class	JNS	JWAKTU	XANGSUR	ANGSKE	XTGKP	XTGKB	HR	HR	KREDIT
	KLMN	BLN					TGKP	TGKB	
1	0.68062	0.25392	0.09947	0.10209	0.86125	0.85863	0.86125	0.85863	0.54188
2	0.9	0.1	0.2	0.1	0	0	0	0	0.4
3	1	0	0	0	0	0	0	0	0
4	1	0	0	0	0	0	0	0	0
5	0.90909	0.27272	0.090909	0	0	0	0	0	0.72727

Setelah menghitung probabilitas pada Tabel 9 setiap attribute, maka mendapatkan probabilitas nya (XCi) dengan Persamaan 6 berikut.

$$P(XCi) = (\text{Attribute}_0 \times \text{Attribute}_1 \times \text{Attribute}_n) + 1/1 \quad (6)$$

Penambahan 1 lalu di bagi 1 digunakan oleh *hukum laplace* untuk menghindari nilai 0 pada probabilitas. Dengan menggunakan rumus tersebut maka didapat hasil probabilitas pada Tabel 10.

Tabel 10. Probabilitas Attribute (XCi)

Class Kolek	Probabilitas Attribute
1	3.0827017979042
2	1.4
3	1
4	1
5	1.7272727272727

Selanjutnya untuk tahap terakhir yaitu mengkalikan probabilitas class (C_i) dengan probabilitas attribute (XC_i) menggunakan Persamaan 7 sebagai berikut

$$P_i = P(C_i) \times P(XC_i) \tag{7}$$

Berdasarkan Persamaan 7 diatas, maka didapat hasil probabilitas yang dapat dilihat pada Tabel 10 berikut ini.

Tabel 11. Probabilitas Final

<i>Class Kolek</i>	<i>Probabilitas Final</i>
1	2.90047312019558
2	0.03448275862068
3	0.00492610837438
4	0.00246305418719
5	0.04679802955664

Dari hasil perhitungan probabilitas pada Tabel 10, maka karena class kolek 1 memiliki nilai probabilitas terbesar, disimpulkan bahwa data uji tersebut diklasifikasikan pada kolek 1.

Untuk data uji selanjutnya hingga akhir dilakukan perhitungan yang sama dari perhitungan probabilitas C_i , probabilitas XC_i , hingga probabilitas akhirnya.

3.5. Hasil Komparasi

Multi class confusion matrix di gunakan pada penelitian ini untuk menguji kinerja 2 metode yang di terapkan yaitu KNN dan Naïve Bayes [16]. Yang akan di cari adalah akurasi, presisi, dan recall. Untuk melihat table confusion matrix untuk metode KNN dapat dilihat pada tabel 12 dan untuk metode Naïve Bayes dapat dilihat pada tabel 13.

Tabel 12. Confusion Matrix untuk Algoritma K-NN

<i>Class</i>	<i>Prediksi</i>				
	1	2	3	4	5
1	72	0	0	0	0
2	1	14	2	0	0
3	0	0	5	0	0
4	0	0	0	1	0
5	0	0	0	4	3

Tabel 13. Confusion Matrix untuk Algoritma Naïve Bayes

<i>Class</i>	<i>Prediksi</i>				
	1	2	3	4	5
1	72	0	0	0	0
2	17	0	0	0	0
3	5	0	0	0	0
4	1	0	0	0	0
5	7	0	0	0	0

Berdasarkan evaluasi dengan menggunakan confusion matrix yang merujuk pada Tabel 11 dan Tabel 12, dapat disimpulkan bahwa prediksi kredit macet dengan menggunakan metode KNN lebih unggul dibandingkan dengan metode Naïve Bayes. Hal ini terlihat dari fakta bahwa KNN memberikan akurasi, presisi, dan recall yang lebih tinggi jika dibandingkan dengan Naïve Bayes. Informasi lebih rinci dari confusion matrix dapat ditemukan di Tabel 14

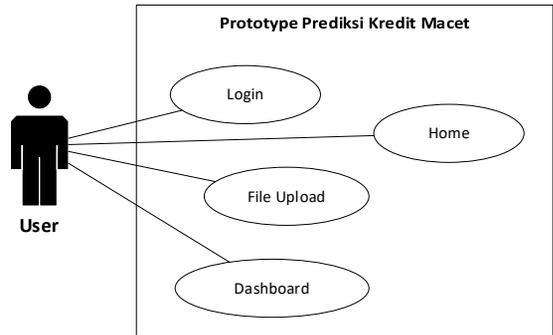
Tabel 14. Hasil Confusion Matrix

	<i>KNN</i>	<i>Naïve Bayes</i>
<i>Accuracy</i>	93,14%	70,59%
<i>Precision</i>	78%	14%

<i>Recall</i>	85%	20%
---------------	-----	-----

3.6. Penerapan Aplikasi *Data Mining*

Berdasarkan tujuan dari penelitian ini ialah menghasilkan suatu prototipe yang dapat memprediksi nasabah yang berpotensi melakukan kredit macet di kemudian hari. Sehingga *rule* yang dihasilkan dari tujuan penelitian ini ialah pembuatan prototipe aplikasi dengan menggunakan metode Naive Bayes dan k – Nearest Neighbors sehingga dapat memudahkan dalam memprediksi kredit macet.



Gambar 3. Use Case Diagram

Berdasarkan Gambar 3, *user* dapat melakukan 3 aktivitas di dalam prototipe yang dibuat, seperti melakukan login, mengunduh file, dan kembali ke *home* atau *dashboard* awal. Prototipe yang digunakan dalam penelitian ini dibuat berbasis web dan menggunakan PHP serta database dibuat menggunakan PostgreSQL. Tampilan dari layar *login* utama *Graphical User Interface* (GUI) prototipe hingga hasil prediksi kredit macet dapat dilihat pada gambar dibawah ini.



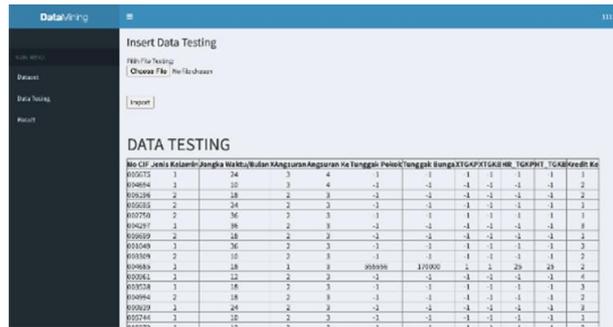
Gambar 4. Tampilan Layar Login

Berdasarkan Gambar 4, pada halaman login, *user* harus melakukan login terlebih dahulu dengan akun yang sudah terdaftar yaitu *user id* dan juga *password*.

No CIP	Jenis Kulkas	Jumlah Waktu/Bulan	Ketersediaan	Anggaran	Tanggal	Tanggal	STORP	STOKS	HB_TURKP	HT_TKAS	Kredit	Kalk
000643	1	30	7	30	400000	3000000	1	10	850	1190	1	5
000666	1	18	8	18	3842895	2750000	5	8	875	1095	2	5
008432	1	24	1	24	22000000	16780000	18	23	1123	1178	1	5
008433	1	24	17	24	3170000	8600000	6	13	749	902	1	5
008268	1	12	3	12	248311	675000	1	8	800	1142	2	5
000972	2	15	4	15	948337	2790000	1	11	885	989	1	5
000906	1	18	2	18	5511394	2370000	15	16	840	1052	1	5
000675	1	18	11	18	1087461	1190000	2	7	485	548	1	5

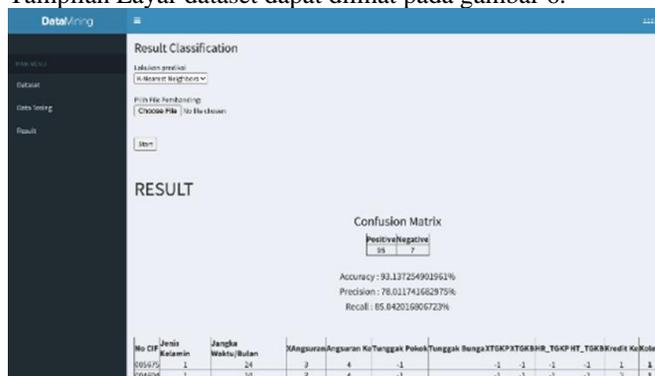
Gambar 5. Tampilan Layar Dataset

Berdasarkan Gambar 5, Pada halaman *dataset*, *user* dapat melakukan *upload* berkas Microsoft excel sebagai *dataset* nya untuk langkah awal melakukan prediksi kredit macet. Microsoft excel yang *diupload* tentunya harus memiliki *format table* yang sesuai.



Gambar 6. Tampilan Layar Data Testing

Berdasarkan Gambar 6, pada halaman data testing, *user* dapat melakukan *upload* berkas Microsoft excel sebagai *data testing* atau *data* yang ingin di prediksi untuk langkah ke dua melakukan prediksi kredit macet. Microsoft excel yang di *upload* tentunya harus memiliki *format table* yang sesuai. Tampilan Layar dataset dapat dilihat pada gambar 6.



Gambar 7. Tampilan Layar Hasil Prediksi

Berdasarkan Gambar 7, pada halaman ini *user* akan memulai dan melihat hasil dari prediksi. *User* dapat memilih metode prediksi apa yang ingin di gunakan. Bisa menggunakan metode K-Nearest Neighbors atau Naïve Bayes. Setelah *user* memilih metode, tekan tombol *Start* untuk memulai proses prediksi kredit macet, dan hasil akan muncul dalam bentuk tabel.

4. KESIMPULAN

Berdasarkan penelitian di atas dapat disimpulkan bahwa metode K – Nearest Neighbors dapat memprediksi kredit macet nasabah dengan tingkat akurasi yang baik, sesuai dengan pedoman yang telah ditetapkan oleh lembaga perbankan.

Kemudian ditemukan juga hasil dari pengujian dua buah metode yang ada dalam klasifikasi data mining, yaitu K – Nearest Neighbors dan Naïve Bayes di dapat bahwa metode data mining dengan menggunakan K – Nearest Neighbors memiliki tingkat akurasi yang baik yaitu sebesar 93,14% untuk prediksi kredit macet nasabah sedangkan metode Naïve Bayes sebesar 70,59%.

Penelitian selanjutnya sebaiknya mengintegrasikan berbagai algoritma klasifikasi seperti *ensemble methods* (misalnya *Random Forest* atau *Gradient Boosting*) serta mengoptimasi fitur melalui teknik *feature engineering* dan penanganan ketidakseimbangan kelas untuk meningkatkan akurasi model, selanjutnya menguji model dengan *dataset* yang lebih luas dan beragam guna memastikan generalisasi hasil, serta mengembangkan prototipe aplikasi real-time yang terintegrasi ke dalam sistem informasi perbankan guna mendukung pengambilan keputusan kredit yang lebih tepat, disertai analisis mendalam mengenai dampak ekonomi dan strategi mitigasi risiko sebagai upaya komprehensif untuk mengurangi kredit macet.

DAFTAR PUSTAKA

-
- [1] Antaristi, M., & Kurniawan, Y. I. (2017). Aplikasi Klasifikasi Penentuan Pengajuan Kartu Kredit Menggunakan Metode Naive Bayes di Bank BNI Syariah Surabaya. *Jurnal Teknik Elektro*, 9(2), 45–52
- [2] Bank BJB. (n.d.). Tentang bank bjb. Retrieved March 20, 2020, from <https://www.bankbjb.co.id/tentang/ina>
- [3] Eko Prasetyo. (2013). Data Mining : Konsep Dan Aplikasi Menggunakan Matlab. In *Journal of Chemical Information and Modeling*. <https://doi.org/10.1017/CBO9781107415324.004>
- [4] Fadilah, A. N., Cholissodin, I., & Mahmudy, W. F. (2015). Implementasi analytical hierarchy process (AHP) dan algoritma genetika untuk rekomendasi dan optimasi pemupukan berimbang tanaman hortikultura. *Repository Jurnal Mahasiswa PTIIK Universitas Brawijaya*, 5.
- [5] Han, J., & Kamber, M. (2006). *Data Mining Concept and Tehniques*. In San Fransisco: Morgan Kauffman.
- [6] Han, Jiawei, Kamber, M., & Pei, J. (2012). *Data Mining: Concepts and Techniques*. In *Data Mining: Concepts and Techniques*. <https://doi.org/10.1016/C2009-0-61819-5>
- [7] Hasan, M. (2017). Prediksi Tingkat Kelancaran Pembayaran Kredit Bank Menggunakan Algoritma NAIVE BAYES Berbasis Forward Selection. *ILKOM Jurnal Ilmiah*. <https://doi.org/10.33096/ilkom.v9i3.163.317-324>
- [8] Hasibuan, M. S. P. (2009). *Manajemen Dasar Pengertian dan Masalah*. In Jakarta: Bumi Aksara.
- [9] Husaini, F. (2016). Algoritma Klasifikasi Naive Bayes Untuk Menilai Kelayakan Kredit (Studi Kasus : Bank Mandiri Kredit Mikro). *Program Studi Teknik Informatika*, 1(3).
- [10] Indriani, A. (2018). Analisa Perbandingan Metode Naive Bayes Classifier dan K- Nearest Neighbor Terhadap Klasifikasi Data. *SEBATIK*, 1–7.
- [11] Ponniah, P. (2001). Ponniah, Paulraj. 2001. *Data Warehouse Fundamentals*. John Wiley & Sons. New York. *Jurnal Ilmiah Matematika Terapan*.
- [12] Puspitasari, D., Al Khautsar, S. S., & Mustika, W. P. (2019). Algoritma Naive Bayes Untuk Memprediksi Kredit Macet Pada Koperasi Simpan Pinjam. *Jurnal Informatika Upgris*, 4(2). <https://doi.org/10.26877/jiu.v4i2.2919>
- [13] Religia, Y. (2015). *Analisis Algoritma k-NN dan Naive Bayes Untuk Klasifikasi Pembelian Mobil*. UDINUS, Semarang.
- [14] Rifqo, & Wijaya. (2017). IMPLEMENTASI ALGORITMA NAIVE BAYES DALAM PENENTUAN PEMBERIAN KREDIT. *Pseudocode*. <https://doi.org/10.33369/pseudocode.4.2.120-128>
- [15] Rohman, A. (2015). Model Algoritma K-Nearest Neighbor (K-NN) Untuk Prediksi Kelulusan Mahasiswa. *Neo Teknika*. <https://doi.org/10.37760/neoteknika.v1i1.350>
- [16] Wicaksana, P. D. (2015). Perbandingan Algoritma K-Nearest Neighbors Dan Naive Bayes Untuk Studi Data " Wisconsin Diagnosis Breast Cancer ". 7.
- [17] Wijayanti, R., & Sulastri. (2018). Analisa Klasifikasi Kartu Kredit Menggunakan Algoritma Naive Bayes. *Prosiding SINTAK*.
- [18] Xiang-wei, L., & Yian-fang, Q. (2012). A Data Preprocessing Algorithm for Classification Model Based On Rough Sets. *Physics Procedia*, 25, 2025–2029. <https://doi.org/10.1016/j.phpro.2012.03.345>