

PENGUNAAN *K-NEAREST NEIGHBORS* ALGORITHM DALAM MENGIDENTIFIKASI KELAYAKAN PINJAMAN NASABAH

Nadya Ika Septiani*, Ika Nur Laily Fitriana

Program Studi Statistika, Fakultas Sains dan Teknologi, Universitas Terbuka

*Penulis korespondensi: nadyaikaseptiani@gmail.com

ABSTRAK

Pinjaman atau kredit adalah kesepakatan antara nasabah dan kreditur atau perusahaan pemberi pinjaman. Kelayakan pinjaman seorang nasabah perlu diperhitungkan oleh perusahaan untuk mengetahui apakah seorang nasabah layak mendapatkan pinjaman yang telah diajukan guna mencegah adanya kerugian pada masa depan. Tujuan dari penelitian ini adalah untuk mengetahui proses penggunaan *K-Nearest Neighbors*, interpretasi nilai *K-Nearest Neighbors* dan evaluasi model *K-Nearest Neighbors* pada data klasifikasi kelayakan pinjaman nasabah. Data yang digunakan berupa data sekunder yang diambil dari *Kaggle* berjumlah 465 data yang memiliki 11 variabel bebas dan 1 variabel target. Metode penelitian yang digunakan adalah metode analisis klasifikasi *K-Nearest Neighbors* (KNN). Pembagian data *train* dan *testing* serta nilai k diperhitungkan untuk mengetahui mana yang menghasilkan nilai evaluasi terbaik. Hasil penelitian menunjukkan bahwa pembagian 90% data *train* dan 10% data *testing* dengan nilai $k = 10$ merupakan model yang terbaik. Model tersebut menghasilkan evaluasi performa model berupa *precision* sebesar 93.933%, *recall* sebesar 77.5%, *F1-score* sebesar 83.93% dan nilai akurasi sebesar 76.09%. Berdasarkan evaluasi tersebut *K-Nearest Neighbors* (KNN) dapat digunakan untuk mengklasifikasikan kelayakan pinjaman nasabah.

Kata kunci: klasifikasi, kelayakan pinjaman, *k-nearest neighbors*, data *train*, data *testing*

1 PENDAHULUAN

Nasabah diartikan sebagai pihak yang mempunyai rekening baik berupa rekening pinjaman maupun rekening simpanan pada Bank (Basir & Renreng, 2021). Berdasarkan UU Nomor 10 tahun 1998 tentang Perbankan pasal 1, nasabah memiliki arti sebagai seseorang yang menggunakan jasa perbankan (Nurmaulia & Sunindy, 2019). Sehingga berdasarkan kedua definisi tersebut, nasabah merupakan pihak yang memiliki rekening di bank baik berupa individu atau instansi dan menggunakan jasa atau produk yang ada disediakan oleh Bank atau pemilik jasa dan produk. Produk atau jasa yang ditawarkan oleh Bank atau pemilik jasa dan produk dapat digunakan untuk memenuhi keperluan dari nasabah itu sendiri.

Nasabah-nasabah yang menggunakan jasa atau produk memiliki sifat atau perilaku yang berbeda-beda. Perilaku atau sifat dari seorang nasabah tidaklah dapat dilihat oleh mata pihak pemilik jasa atau produk. Pola perilaku nasabah menurut Sangadji & Sopiah merupakan segala upaya untuk mendapat dan menggunakan barang-barang atau jasa guna memenuhi kebutuhan hidupnya (Zahrotun Nadiyah & Nur Aisyah, 2022). Pola perilaku nasabah memiliki sifat dinamis di mana pola perilaku akan berjalan seiring dengan waktu dan berubah untuk menempatkan diri dalam situasi terutama lingkungannya (Supriyanto et al., 2021). Terdapat faktor-faktor yang memengaruhi pola perilaku nasabah menurut Kotler dan Keller yaitu faktor budaya, faktor sosial,

faktor pribadi dan faktor psikologis (Syam et al., 2019). Pola perilaku seorang nasabah akan menjadi penilaian dari perusahaan atau bank. Penilaian tersebut juga nantinya akan memengaruhi kelancaran proses pinjaman yang diajukan oleh nasabah.

Pinjaman atau kredit yang diambil oleh nasabah merupakan pinjaman yang telah dipercayakan oleh pihak pemberi pinjaman atau kreditur kepada pihak penerima pinjaman atau debitur baik telah sepakat dalam tenggang waktu, serta besaran pokok maupun bunga yang harus dibayarkan (Agustin Wulandari, 2021). Berdasarkan Undang-Undang Republik Indonesia tentang Perbankan Pasal 1 Butir 11 No. 10 Tahun 1998, pinjaman merupakan kegiatan pengadaan uang yang didasarkan oleh keputusan bersama antara pihak peminjam atau nasabah dengan pihak yang memberikan pinjaman atau perusahaan untuk memberikan bunga dan membayar pada waktu yang telah disepakati (Retnosari, 2021). Kelayakan seorang nasabah dalam mengambil sebuah pinjaman harus menjadi konsentrasi penuh perusahaan. Karena kelayakan pinjaman dari nasabah juga menentukan perusahaan akan memberikan pinjaman kepada nasabah sesuai dengan kesepakatan yang diajukan oleh nasabah atau pemohon (Ajeng et al., 2020). Hal tersebut ditujukan agar perusahaan menghindari risiko terbesar dari pinjaman yaitu di mana kredit dari nasabah mengalami kemacetan atau tidak terbayarkan (Setiawan, 2020). Kredit macet seperti ini yang dapat merugikan perusahaan terutama dibagian keuangan. Sehingga mencari kelayakan pinjaman wajib dilakukan atau menjadi pengetahuan utama yang harus diketahui oleh perusahaan sebelum memberikan pinjaman kepada nasabah.

Untuk mengetahui apakah seorang nasabah layak menerima pinjaman dari perusahaan atau pemberi pinjaman adalah dengan melakukan uji kelayakan pinjaman. Uji kelayakan pinjaman digunakan sebagai prediktor seorang nasabah akan memiliki masalah terkait dengan pinjamannya atau tidak. Nasabah diperlukan untuk mengisi data yang berisikan faktor-faktor yang diperlukan oleh perusahaan untuk melakukan uji kelayakan pinjaman tersebut. Pengambilan uji secara manual dengan jumlah data yang besar akan menyebabkan kesulitan dan memerlukan waktu yang lama, sehingga diperlukan pengujian kelayakan nasabah dengan menggunakan sistem pendukung yang dapat mempermudah perusahaan untuk menentukan kelayakan pinjaman (Putra et al., 2023). Pengujian ini akan menjadi sebuah sistem yang membantu dan mendukung perusahaan untuk memberikan keputusan. Penulis menggunakan *K-Nearest Neighbors* (KNN) karena dapat mengolah dan mengklasifikasikan data secara akurat dan cepat. Sehingga diharapkan hal tersebut menghasilkan alternatif yang mendukung keputusan perusahaan. Melihat masalah yang dihadapi oleh sebuah perusahaan dalam memutuskan sebuah keputusan pinjaman, penelitian ini bertujuan untuk mengetahui cara mengklasifikasikan data dengan model *K-Nearest Neighbors* (KNN) dalam proses penyeleksian kelayakan pinjaman nasabah agar perusahaan dapat memberikan keputusan untuk pinjaman nasabah, mengetahui interpretasi hasil klasifikasi *K-Nearest Neighbors* (KNN), dan mengetahui evaluasi model yang *K-Nearest Neighbors* (KNN) digunakan untuk mengidentifikasi kelayakan pinjaman nasabah.

2 METODE

2.1 Data

Data yang digunakan dalam penelitian ini merupakan data sekunder berupa dataset yang diambil dari *Kaggle.com* yang akan dibagi menjadi dua tabel data yaitu data *train* dan data *testing*. Data *train* atau data pelatihan merupakan data yang dibutuhkan untuk dapat membangun model kelas klasifikator (Maulana & Yahya, 2019). Sedangkan data *testing* atau data pengujian merupakan data

yang digunakan untuk menguji model kelas klasifikator. Untuk data *testing* akan diuji di mana berapa pembagian terbaik dari model tersebut dimulai dari 10% sampai dengan 90%. Terdapat 11 variabel bebas (X_i) dan 1 variabel yang menjadi target (Y).

Tabel 1. Variabel Penelitian

Variabel	Nama Variabel	Skala Pengukuran	Keterangan
X_1	Jenis Kelamin	Nominal	0 = Perempuan 1 = Laki-laki
X_2	Status Pernikahan	Nominal	0 = Tidak 1 = Ya
X_3	Tanggungan	Nominal	0 = 0 1 = 1 2 = 2 3 = 3+
X_4	Pendidikan	Nominal	0 = Lulus 1 = Tidak Lulus
X_5	Wirausaha	Nominal	0 = Tidak 1 = Ya
X_6	Penghasilan Pribadi Peminjam	Rasio	Penghasilan Pribadi Peminjam
X_7	Penghasilan Peminjam Bersama	Rasio	Penghasilan orang yang bergabung dalam pengajuan pinjaman
X_8	Pinjaman	Rasio	Jumlah pinjaman yang diajukan oleh peminjam
X_9	Tenggang Waktu	Rasio	Jangka Waktu yang Diajukan Peminjam
X_{10}	Sejarah Kredit	Nominal	0 = Tidak Pernah Melakukan Kredit 1 = Pernah Melakukan Kredit
X_{11}	Tipe Tempat Tinggal	Nominal	0 = Pedesaan 1 = Semiperkotaan 2 = Perkotaan
Y	Status Kelayakan Pinjaman	Nominal	0 = Tidak Layak 1 = Layak

2.2 Metode Analisis

2.2.1 K-Nearest Neighbors (KNN)

Metode analisis yang digunakan dalam penelitian ini adalah metode klasifikasi *K-Nearest Neighbors (KNN) Algorithm*. *K-Nearest Neighbors (KNN) Algorithm* merupakan metode analisis non-parametrik dengan menggunakan teknik regresi dan klasifikasi yang diperkenalkan pertama kali oleh Evelyn Fix dan Joseph Hodges pada tahun 1951 (Sabry, 2022). Pengklasifikasian data menggunakan metode *K-Nearest Neighbors (KNN)* akan efektif pada penelitian ini, karena dengan data yang besar metode tersebut akan efektif dan cepat dalam klasifikasi. Selain itu menurut Nuraeni et al. (2023), metode analisis *K-Nearest Neighbors (KNN)* ini akan menghasilkan data yang lebih akurat karena KNN memiliki kekuatan terhadap data *train* yang *noise*. Penggunaan K-

Nearest Neighbors (KNN) Algorithm adalah dengan membandingkan jarak terdekat dari data *testing* dan data *train* (Cholil et al., 2021).

Metode analisis *K-Nearest Neighbors* seringkali digunakan untuk memprediksi sesuatu dengan melakukan uji dengan mencari jarak terdekat antara data *train* dan data *testing* (Faisal et al., 2023). Jarak antar data disebut dengan *euclidean distance*. Miftahuddin et al. (2020) menjelaskan bahwa *euclidean distance* merupakan perhitungan jarak antar dua titik menggunakan perhitungan *pythagoras* dalam satu dimensi yaitu *euclidean space*. *Euclidean space* merupakan dimensi yang mempelajari hubungan sudut dan jarak.

2.2.2 Langkah-langkah Uji Klasifikasi *K-Nearest Neighbors* (KNN)

Langkah-langkah yang diperlukan dalam uji *K-Nearest Neighbors* (KNN) adalah sebagai berikut:

1. Menentukan nilai *k* dari data. Nilai *k* merupakan nilai yang didapat dari jumlah tetangga terdekat dan keakuratan dari nilai *k* menurut Cholil et al. (2021) berdasarkan data *train* yang digunakan. Nilai *k* pada pengujian didapat dengan menggunakan persamaan berikut.

$$k = \sqrt{N}, \quad (1)$$

dengan *N* merupakan banyaknya data yang digunakan pada pengujian *K-Nearest Neighbors* (KNN). Selain dengan menggunakan rumus tersebut. Nilai *k* juga bisa didapat berdasarkan keinginan peneliti, seberapa besar nilai *k* yang akan diuji sehingga akan didapat nilai *k* yang dapat mewakili model dan memiliki nilai akurasi terbaik.

2. Mencari jarak masing-masing data (*euclidean distance*). Masing-masing data yang dimaksud adalah data *train* dan data *testing* yang akan dicari jarak terdekatnya. Persamaan untuk mencari *euclidean distance* menurut Miftahuddin et al. (2020) berikut ini.

$$d = \sqrt{(x_1 - x_2)^2 + (y_1 - y_2)^2} \quad (2)$$

Keterangan :

d = jarak atau *euclidean distance*

*x*₁ = koordinat latitudinal 1 atau koordinat garis lintang 1

*x*₂ = koordinat latitudinal 2 atau koordinat garis lintang 2

*y*₁ = koordinat longitudinal 1 atau koordinat garis bujur 1

*y*₂ = koordinat longitudinal 2 atau koordinat garis bujur 2

3. Melakukan pengurutan pada data yang telah dihitung jaraknya (*euclidean distance*). Pengurutan dilakukan dimulai dari data yang memiliki jarak atau *euclidean distance* paling kecil ke yang paling besar.
4. Melakukan pengelompokkan data menjadi kategori *Y* di mana pada penelitian ini terbagi menjadi dua kategori yaitu *Ya* dan *Tidak*. Pengelompokkan data tersebut didasarkan oleh jarak (*euclidean distance*) terkecil atau berdekatan (*nearest neighbor*).
5. Memilih nilai berdekatan (*nearest neighbor*) pada kategori *Y* yang menjadi mayoritas atau terbanyak.

2.2.3 Evaluasi Performa *K-Nearest Neighbors* (KNN)

Evaluasi performa *K-Nearest Neighbors* (KNN) perlu dilakukan dalam pengklasifikasian data dengan tujuan untuk mengetahui performa dari model analisis tersebut. Melalui evaluasi performa, manfaat yang akan didapat adalah dapat menilai model analisis yang digunakan apakah sudah baik dan dapat mengetahui model terbaik yang dapat diimplementasikan dalam keputusan yang akan

datang. Untuk mengevaluasi performa dari *K-Nearest Neighbor* (KNN) dapat dengan *confusion matrix* dan *classification report*.

1. *Confusion matrix* merupakan matriks dalam algoritma klasifikasi untuk melaporkan dan menginterpretasikan hasil dari *K-Nearest Neighbors* (KNN) (Permana et al., 2023). *Confusion matrix* berisi nilai yang merupakan relasi antara *actual case* dan *predicted case*. Adapun bentuk dari *confusion matrix* adalah sebagai berikut:

Tabel 2. *Confusion matrix*

Predicted case	Actual case	
	Positive	Negative
Positive	True positive	False positive (type I error)
Negative	False negatif (type II error)	True negatif

True Positive (TP) merupakan nilai yang memprediksi benar dan pada kenyataannya benar, dalam kasus ini berarti peminjam yang diprediksi status kelayakan pinjamannya Layak dan kenyataannya benar Layak. *True Negative* (TN) merupakan nilai yang memprediksi salah dan pada kenyataannya salah, dalam kasus klasifikasi peminjam tersebut dapat diartikan sebagai peminjam yang diprediksi status kelayakan pinjamannya Tidak Layak dan kenyataannya benar Tidak Layak. Dalam *confusion matrix* terdapat dua *type error* di mana *Type I Error* yaitu *False Positive* (FP) merupakan nilai yang memprediksi benar akan tetapi pada kenyataannya salah, dalam kasus ini diartikan sebagai peminjam yang diprediksi status kelayakan pinjamannya Layak dan kenyataannya Tidak Layak. Sedangkan untuk *Type II Error* adalah *False Negative* merupakan nilai yang memprediksi salah ternyata dalam kenyataannya benar, dalam kasus ini memprediksi bahwa peminjam yang diprediksi status kelayakan pinjamannya Tidak Layak dan kenyataannya Layak.

2. *Classification report* menurut Permana et al. (2023) merupakan laporan yang berisikan nilai yang didapat dari klasifikasi. *Classification report* berisikan *precision*, *recall*, *F1-score*, dan akurasi, di mana nilai tersebut dapat dihitung setelah mengetahui *confusion matrix* dan akan menjelaskan mengenai performa dari model klasifikasi yang digunakan.
 - a. *Precision*

Precision merupakan akurasi dari prediksi positif benar. Akurasi dari *precision* didapat dengan membagi prediksi positif benar (*True Positive*) dengan keseluruhan prediksi positif (*Precision Positive*). Adapun persamaan yang digunakan dalam menghitung *precision* sebagai berikut:

$$Precision = \frac{True\ Positive}{(True\ Positive + False\ Positive)} \quad (3)$$

- b. *Recall*

Recall merupakan akurasi sensitivitas prediksi yang berasal dari pembagian prediksi benar positif (*True Positive*) dengan keseluruhan yang aktual positif. Adapun persamaan yang digunakan dalam menghitung *recall* sebagai berikut:

$$Recall = \frac{True\ Positive}{(True\ Positive + False\ Negative)} \quad (4)$$

- c. *F1-Score*

F1-Score merupakan nilai yang terbentuk dari nilai *precision* dan *recall* dengan tujuan untuk dapat mendapat harmonis *threshold*. Harmonis *threshold* perlu digunakan untuk

mengetahui apakah nilai yang didapat telah melebihi standar harmonisasi. Adapun persamaan yang digunakan dalam menghitung *F1-Score* sebagai berikut:

$$F1 - Score = 2 \frac{precision \times recall}{precision + recall} \quad (5)$$

d. Akurasi

Akurasi dalam evaluasi performa sangat penting. Menurut Permana et al. (2023) akurasi diperlukan untuk mengetahui seberapa kemampuan model dalam melakukan klasifikasi terhadap data yang ada. Untuk menghitung nilai akurasi adalah sebagai berikut:

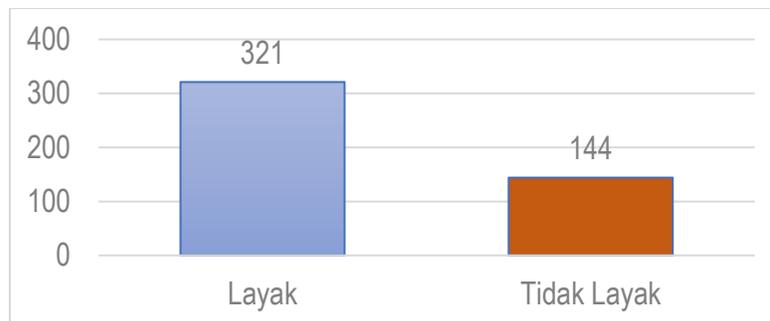
$$Akurasi = \frac{True\ Positive + True\ Negative}{True\ Positive + True\ Negative + False\ Positive + False\ Negatif} \quad (6)$$

3 HASIL DAN PEMBAHASAN

3.1 Analisis Deskriptif

Analisis deskriptif merupakan metode yang dapat mengidentifikasi dan menjelaskan terkait dengan peristiwa yang sedang diuji dan dikaji. Analisis deskriptif dapat dikatakan sebagai analisis univariat karena analisis tersebut hanya menganalisis dan mendeskripsikan sebuah variabel tanpa adanya hubungan dengan variabel lain. Analisis deskriptif dapat disajikan dalam sebuah tabel, grafik, *plot* ataupun diagram lainnya.

Dalam penelitian ini, analisis deskriptif akan menganalisis variabel target atau variabel Y yaitu status kelayakan pinjaman. Tujuannya untuk mengetahui perbandingan antara banyak data yang menghasilkan status kelayakan pinjaman Layak dan status kelayakan pinjaman Tidak Layak. Hasil analisis deskriptif pada variabel Y akan ditampilkan dalam diagram batang pada Gambar 1 berikut.



Gambar 1. Diagram Batang Status Kelayakan Pinjaman (Variabel Y)

Berdasarkan Gambar 1, didapat dua kelas yaitu status kelayakan pinjaman Layak dan status kelayakan pinjaman Tidak Layak. Dari 465 data total, didapat status kelayakan pinjaman paling banyak adalah 321 data dengan status kelayakan pinjaman Layak, sedangkan pada status kelayakan pinjaman Tidak Layak hanya berjumlah 144 data. Hal tersebut menunjukkan bahwa data yang digunakan didominasi oleh data dengan status kelayakan pinjaman Layak.

3.2 Praproses

Pada tahap praproses data, data akan disiapkan terlebih dahulu sebelum siap dilakukan proses klasifikasi. Hal tersebut dilakukan dengan tujuan agar program komputer yang dijalankan tidak mendeteksi adanya *error* atau kesalahan karena data.

3.3.1 Pengecekan data

Praproses yang pertama adalah mengecek apakah data telah bersih dari hal-hal yang mengganggu proses klasifikasi seperti adanya *blank*, data tidak lengkap atau duplikasi. Hal-hal yang mengganggu tersebut akan menyebabkan program akan *error* dan proses klasifikasi akan tidak membuahkan hasil. Langkah yang dapat dilakukan adalah dengan melakukan penghapusan pada data yang tidak lengkap atau duplikasi. Setelah melakukan pemeriksaan didapat bahwa terdapat data yang hilang atau *blank* sebanyak 149 data dari 614 data. Sehingga data yang bisa digunakan dalam proses klasifikasi hanya berjumlah 465 data.

3.3.1 Melakukan Transformasi Data

Pada praproses data, transformasi data yang dimaksud adalah mengubah data yang sebelumnya merupakan data yang tidak berskala numerik menjadi data yang berskala ukur numerik. Sebagai contoh adalah pada variabel X_1 yang berisikan jenis kelamin Laki-laki dan Perempuan, variabel tersebut tidak akan bisa dibaca oleh program komputer dalam proses klasifikasi. Sehingga hal yang harus dilakukan adalah mengubah data tersebut ke skala ukur nominal dengan 0 = Perempuan dan 1 = Laki-laki. Begitupun seterusnya dilakukan pada data-data yang tidak berskala numerik.

3.3 Klasifikasi

Pada proses klasifikasi data akan dibagi menjadi beberapa proporsi dan akan ditentukan nilai k yang terdiri dari beberapa nilai. Hal tersebut ditujukan untuk mendapatkan proporsi pembagian dan nilai k terbaik dalam proses pengidentifikasian kelayakan pinjaman nasabah dengan metode *K-Nearest Neighbors (KNN)*.

3.3.1 Pembagian Data

Pembagian data pada penelitian ini ditujukan membagi dataset menjadi dua kelas yaitu data *train* dan data *testing*. Pembagian data akan disimulasikan menjadi beberapa proporsi guna mengetahui komposisi data mana yang terbaik. Sehingga nantinya juga akan memberikan evaluasi model *K-Nearest Neighbors (KNN)* terbaik.

Tabel 3. Pembagian Data *Train* dan Data *Testing*

Proporsi	Jumlah Data	
	Data <i>Train</i>	Data <i>Testing</i>
Data <i>Train</i> 90%, Data <i>Testing</i> 10%	419	46
Data <i>Train</i> 80%, Data <i>Testing</i> 20%	372	93
Data <i>Train</i> 70%, Data <i>Testing</i> 30%	326	139
Data <i>Train</i> 60%, Data <i>Testing</i> 40%	279	186
Data <i>Train</i> 50%, Data <i>Testing</i> 50%	233	232
Data <i>Train</i> 40%, Data <i>Testing</i> 60%	186	279
Data <i>Train</i> 30%, Data <i>Testing</i> 70%	139	326
Data <i>Train</i> 20%, Data <i>Testing</i> 80%	93	372
Data <i>Train</i> 10%, Data <i>Testing</i> 90%	46	419

3.3.2 Penentuan Nilai k

Nilai k pada pengujian ini akan sekaligus menjadi pembuktian apakah dengan menggunakan persamaan (1) yaitu $k = \sqrt{N}$ merupakan rumus dalam penentuan nilai k . Jumlah data *train* terbesar berada di komposisi Data *Train* 90% dan Data *Testing* 10%. Di mana pada pembagian data tersebut

data *train* berjumlah 419 data. Maka menurut persamaan (1), maksimal nilai k yang akan diuji adalah $\sqrt{419} = 20.47 \sim 21$. Sehingga dalam penelitian ini, nilai k yang akan digunakan adalah 1,2,3, ..., 21.

3.4 Evaluasi Model *K-Nearest Neighbors* (KNN)

Evaluasi menurut KBBI merupakan kegiatan yang dilakukan dengan tujuan untuk menyediakan informasi sejauh mana sebuah proses tersebut telah berjalan dan telah mencapai nilai berdasarkan kriteria yang telah ditentukan. Sehingga evaluasi model *K-Nearest Neighbors* adalah kegiatan untuk mengecek model dan hasil dari penerapan model *K-Nearest Neighbors* (KNN). Dengan evaluasi model *K-Nearest Neighbors* (KNN) peneliti akan mengetahui sejauh mana model *K-Nearest Neighbors* (KNN) dapat mengidentifikasi kelayakan pinjaman nasabah.

Setelah dilakukan pemrosesan klasifikasi akan menghasilkan *confusion matrix* dengan nilai akurasi dari setiap k dan pembagian data yang digunakan. Sebelum menentukan *confusion matrix*, akan dicari terlebih dahulu proporsi pembagian data dan nilai k yang menghasilkan nilai akurasi tertinggi. Dari proporsi dan pembagian data tersebut akan diketahui *confusion matrix* yang akan menghasilkan *classification report* model *K-Nearest Neighbors* (KNN) terbaik juga. Adapun nilai akurasi model *K-Nearest Neighbors* (KNN) berdasarkan nilai k dan pembagian data ditampilkan pada Tabel 4 berikut ini.

Tabel 4. Akurasi Model *K-Nearest Neighbors* (KNN) berdasarkan nilai K dan pembagian data

K	Train								
	90%, Testing 10%	80%, Testing 20%	70%, Testing 30%	60%, Testing 40%	50%, Testing 50%	40%, Testing 60%	30%, Testing 70%	20%, Testing 80%	10%, Testing 90%
1	0.5217	0.6022	0.5971	0.5914	0.5776	0.5735	0.5859	0.5430	0.5298
2	0.6304	0.5806	0.5612	0.5591	0.5733	0.5914	0.6074	0.5484	0.5060
3	0.5870	0.5914	0.5971	0.5914	0.5991	0.6237	0.6104	0.5672	0.5084
4	0.6522	0.6667	0.5971	0.5914	0.5991	0.6129	0.6104	0.5780	0.5394
5	0.6304	0.6344	0.6187	0.6183	0.5862	0.6308	0.6258	0.5968	0.5298
6	0.6522	0.6452	0.6115	0.6452	0.5905	0.6559	0.6442	0.6290	0.5346
7	0.6739	0.6559	0.6187	0.6290	0.6293	0.6631	0.6472	0.6505	0.5227
8	0.6957	0.6452	0.6547	0.6290	0.6466	0.6738	0.6380	0.6344	0.5346
9	0.6957	0.6559	0.6906	0.6344	0.6466	0.6846	0.6810	0.6505	0.5370
10	0.7609*	0.6667	0.6475	0.6452	0.6379	0.6631	0.6748	0.6559	0.4845
11	0.7391	0.7097	0.6763	0.6505	0.6595	0.6631	0.6840	0.6559	0.4726
12	0.7391	0.7204	0.6763	0.6667	0.6724	0.6452	0.6810	0.6667	0.4678
13	0.7391	0.7204	0.6691	0.6613	0.6595	0.6703	0.7086*	0.6667	0.4773
14	0.7391	0.7312*	0.6619	0.6720*	0.6681	0.6810	0.7086*	0.6720	0.5513
15	0.7174	0.7204	0.6691	0.6613	0.6509	0.7025	0.7055	0.6801	0.6277
16	0.7174	0.6882	0.6763	0.6667	0.6595	0.7133*	0.7055	0.6909	0.6229
17	0.7174	0.7204	0.6763	0.6613	0.6509	0.7061	0.7086*	0.6989*	0.6539
18	0.7174	0.6882	0.6691	0.6613	0.6552	0.7097	0.7086*	0.6962	0.6659*
19	0.6957	0.7097	0.7050*	0.6559	0.6638	0.7025	0.7086*	0.6989	0.6444
20	0.7174	0.6989	0.6763	0.6613	0.6767	0.6953	0.7055	0.6935	0.6635
21	0.7174	0.7097	0.6547	0.6613	0.6853*	0.7025	0.7086*	0.6989*	0.6539

Berdasarkan Tabel 4, nilai k dan pembagian data yang memiliki nilai akurasi tertinggi berada di pembagian data 90% data *train* 10% data *testing* dengan nilai k = 10 dan nilai akurasi sebesar 76.09%. Melihat kembali Tabel 4, nilai k yang memiliki nilai akurasi tertinggi bukanlah berada pada nilai $k = \sqrt{\text{banyaknya data train}}$. Sehingga dapat disimpulkan bahwa rumus nilai k terbaik $k = \sqrt{\text{banyaknya data train}}$ tidak dapat menjadi pedoman sehingga diperlukan percobaan untuk melihat nilai k terbaik yang menghasilkan akurasi tertinggi.

Setelah mengetahui nilai akurasi tertinggi dari setiap nilai k dan pembagian data, kita akan melihat *confusion matrix* dari nilai k dan pembagian terbaik untuk dapat mengetahui evaluasi performa model *K-Nearest Neighbors* (KNN) dalam bentuk *classification report* yang akan ditampilkan pada Tabel 5.

Tabel 5. *Confusion matrix* 90% data *train* dan 10% data *testing* dengan nilai k = 10

<i>Predicted Case</i>	<i>Actual Case</i>	
	1	0
1	31	2
0	9	4

Berdasarkan Tabel 5, yang berisikan *confusion matrix* dari nilai k dan pembagian data yang memiliki nilai akurasi tertinggi akan diubah menjadi *classification report* guna melihat performa model *K-Nearest Neighbors* (KNN) dan akan dievaluasi terkait sejauh mana model yang digunakan dapat mengidentifikasi kelayakan pinjaman nasabah. Adapun *classification report* yang didapat akan dijelaskan dalam Tabel 6 berikut.

Tabel 6. *Classification report*

Pembagian Data	k	<i>Precision</i>	<i>Recall</i>	<i>F1-Score</i>	Akurasi
90% Data <i>Train</i> , 10% Data <i>Testing</i>	10	0.93939	0.775	0.8493	0.7609

Berdasarkan *Classification Report* pada Tabel 6 didapat nilai *precision* sebesar 93.939%, nilai *recall* sebesar 77.5%, nilai *F1-Score* sebesar 84.93%, dan nilai akurasi 76.09%. Nilai-nilai yang ditampilkan dalam *classification report* pada Tabel 6 memiliki nilai yang tinggi. Perhitungan pembagian data dan nilai k perlu diperhatikan agar dapat menghasilkan nilai akurasi yang tinggi, sehingga *K-Nearest Neighbors* (KNN) dapat digunakan untuk mengklasifikasikan kelayakan pinjaman nasabah.

4 KESIMPULAN

Berdasarkan hasil dan pembahasan dapat disimpulkan bahwa proses klasifikasi data status kelayakan pinjaman dalam mengidentifikasi kelayakan pinjaman nasabah diawali dengan menganalisis data status kelayakan pinjaman secara deskriptif, dilanjut dengan praproses data berupa pengecekan dan perubahan data, lalu proses klasifikasi yaitu pembagian data dan penentuan nilai, dan ditutup dengan evaluasi performa model *K-Nearest Neighbors* (KNN). Interpretasi nilai yang didapat dari proses klasifikasi *K-Nearest Neighbors* (KNN) adalah pembagian data dan nilai k terbaik yang dilakukan simulasi adalah pembagian data 90% data *train* dan 10% data *testing* serta nilai k terbaiknya adalah 10. Berdasarkan pembagian data dan nilai k terbaik tersebut menghasilkan evaluasi performa model *K-Nearest Neighbors* (KNN) yang tinggi yaitu *precision* 93.939%, *recall* 77.5%, *F1-score* 84.93% dan nilai akurasi sebesar 76.09%. Dengan demikian K-

Nearest Neighbors (KNN) dapat digunakan untuk mengklasifikasikan kelayakan pinjaman nasabah.

UCAPAN TERIMA KASIH

Di dalam artikel ini, penulis telah memperoleh banyak petunjuk, bantuan, dan bimbingan dari berbagai pihak dari mulai memilih data, mengolah data hingga artikel ini selesai. Selain itu penulis juga mendapatkan banyak dukungan dan nasihat yang membangun dari berbagai pihak ketika penulis mengalami kesulitan. Penulis menyampaikan terima kasih tak terkira kepada pihak-pihak yang telah membantu penulis sehingga artikel ini dapat terselesaikan dengan baik.

DAFTAR PUSTAKA

- Agustin Wulandari, R. (2021). Fungsi Perjanjian Kredit Bagi Bank Selaku Kreditur pada PT. BPR Dharma Nagari. *JAH Jurnal Analisis Hukum*, 2(2), 34.
- Ajeng, N., Sari, B. W., & Prabowo, D. (2020). Prediksi Pemberian Kelayakan Pinjaman Dengan Metode Fuzzy Tsukamoto. *Information System Journal (INFOS)*, 3(1), 19.
- Basir, M., & Renreng, M. (2021). Analisis Pembiayaan Produktif Terhadap Jumlah Nasabah Pada Koperasi Karyawan Samudera Tonasa Lines Kabupaten Pangkep. *PAY Jurnal Keuangan Dan Perbankan*, 3(1), 16–20. <https://doi.org/10.46918/pay.v3i1.960>
- Cholil, S. R., Handayani, T., Prathivi, R. & Ardianita. T., (2021). Implementasi Algoritma Klasifikasi K-Nearest Neighbor (KNN) untuk Klasifikasi Seleksi Penerima Beasiswa. *IJCIT (Indonesian Journal on Computer and Information Technology)*, 6(2), 118-127. <http://dx.doi.org/10.31294/ijcit.v6i2.10438>
- Faisal, M., Utami, W. S., Parmica, S., Studi, P., Informatika, T., Sains, F., & Raharja, U. (2023). Implementasi Algoritma K-Nearest Neighbor (KNN) Dalam Memprediksi Indeks Kemiskinan. *Journal Sensi: Strategic of Education in Information System*, 9(1), 11-23. <https://doi.org/10.33050/sensi.v9i1.2616>
- Maulana, D., & Yahya, R. (2019). IMPLEMENTASI ALGORITMA NAÏVE BAYES UNTUK KLASIFIKASI PENDERITA PENYAKIT JANTUNG DI INDONESIA MENGGUNAKAN RAPID MINER. *SIGMA Information Technology Journal*, 10 (2), 191–197.
- Miftahuddin, Y., Umaroh, S., & Karim, F. (2020). PERBANDINGAN METODE PERHITUNGAN JARAK EUCLIDEAN, HAVERSINE, DAN MANHATTAN DALAM PENENTUAN POSISI KARYAWAN. *Jurnal Tekno Insentif*, 14(2), 69-77. <https://doi.org/https://doi.org/10.36787/jti.v14i2.270>
- Nuraeni, S., Putri, S., Syam, A., Wajdi, M. F., & Malkan, M. (2023). *G-Tech : Jurnal Teknologi Terapan Implementasi Metode K-NN Untuk Menentukan Jurusan Siswa di SMAN 02 Manokwari*. 7(1), 89–95.
- Nurmaulia, S. A., & Sunindyo, A. (2019). Analisis Pengaruh Kualitas Layanan Terhadap Tingkat Kepuasan Nasabah Tabungan Simpedes Pada Pt Bank Rakyat Indonesia (Persero), Tbk Kantor Cabang Pattimura Semarang. *Keunis*, 7(1), 5. <https://doi.org/10.32497/keunis.v7i1.1527>
- Putra, I. M. A. W., Gunawan, I. M. A. O., & Sudiatmika, I. P. G. A. (2023). Implementasi Metode ELECTRE dalam Sistem Pendukung Keputusan Penentuan Kelayakan Pinjaman. *Journal of Information System Research (JOSH)*, 4(3), 785–793. <https://doi.org/10.47065/josh.v4i3.3237>
- Permana, I. G. T & Dwidasmara, I. B. G. (2023). Evaluasi Performance dengan Grid Search Terhadap K Nearest Neighbor (KNN) untuk Klasifikasi Penderita Diabetes Melitus. *Jurnal*

Elektronik Ilmu Komputer Udayana, 12(1), 71-82.
<https://doi.org/10.24843/JLK.2023.v12.i01.p09>.

- Retnosari, R. (2021). ANALISIS KELAYAKAN KREDIT USAHA MIKRO BERJALAN PADA PERBANKAN DENGAN METODE NAIVE BAYES. *Jurnal PROSISKO*, 8 (1), 53. <https://doi.org/10.30656/prosisko.v8i1.2848>
- Sabry, F. (2022). *K Nearest Neighbor Algorithm : Fundamentals and Applications*. One Billion Knowledgeable.
- Setiawan, R. (2020). Analisis Kelayakan Pemberian Kredit Nasabah Koperasi Menggunakan Algoritma C4.5. *Techno Xplore : Jurnal Ilmu Komputer Dan Teknologi Informasi*, 5(2), 74–78. <https://doi.org/10.36805/technoxplore.v5i2.1175>
- Supriyanto, D., Aravik, H., & Choiriyah, C. (2021). Analisis Perilaku Milenial Dalam Keputusan Menjadi Nasabah Bank Syariah Di Era Revolusi 4.0 (Studi Kasus Mahasiswa Prodi Perbankan Syariah Stebis Igm Palembang). *Jurnal Ilmiah Mahasiswa Perbankan Syariah (JIMPA)*, 1(2), 185–192. <https://doi.org/10.36908/jimpa.v1i2.29>
- Syam, M. K., Dilla, A. M., & Musa, M. I. (2019). FAKTOR-FAKTOR YANG MEMPENGARUHI PERILAKU KONSUMEN TERHADAP KEPUTUSAN PEMBELIAN SEPEDA MOTOR YAMAHA PADA PT. SURACIJAYA ABADI MOTOR CABAN BATUA RAYA DI KOTA MAKASSAR. Departemen Fakultas Ekonomi Universitas Negeri Makasar. 22 hlm.
- Zahrotun Nidiyah, S., & Nur Aisyah, E. (2022). Analisis Perilaku Nasabah Dalam Melakukan Pembiayaan Pada Koperasi Syariah Murni Amanah Sejahtera Kota Malang. *Jurnal Tabarru' : Islamic Banking and Finance*, 5(2), 398–411. [https://doi.org/10.25299/jtb.2022.vol5\(2\).10872](https://doi.org/10.25299/jtb.2022.vol5(2).10872)