PENERAPAN *DATA MINING* MENGGUNAKAN ALGORITMA *K-NEAREST NEIGHBOR* PADA PREDIKSI PEMBERIAN KREDIT DI SEKTOR FINANSIAL

Nita Windy Mardiyyah[1], Nining Rahaningsih[2], Irfan Ali[3]
[1, 2] Program Studi Komputerisasi Akuntansi STMIK IKMI Cirebon
[3] Program Studi Rekayasa Perangkat Lunak STMIK IKMI Cirebon
Jln. Perjuangan No. 10B Majasem Kec. Kesambi Kota Cirebon Tlp. 0231-490480-490481
nitaawindy03@gmail.com

ABSTRAK

Dalam era dinamika finansial yang cepat, manajemen kredit menjadi esensi utama bagi lembaga-lembaga finansial untuk menjaga kestabilan dan keseimbangan. Pengambilan keputusan yang tepat dalam pemberian kredit menjadi krusial, mengingat kompleksitas risiko yang terus berkembang. Meskipun sektor finansial telah mengadopsi berbagai metode evaluasi risiko kredit, masalah persisten terkait dengan ketidakpastian dan volatilitas pasar menyulitkan perusahaan untuk membuat keputusan kredit yang tepat waktu dan akurat. Dalam permasalahan tersebut, penelitian ini bertujuan mengeksplorasi dan menerapkan algoritma *K-Nearest Neighbor (KNN)* melalui pendekatan *data mining* untuk meningkatkan prediksi pemberian kredit. Implementasi *K-Nearest Neighbor (KNN)* dilakukan dengan *tools rapidminer* yang akan membantu dalam pelatihan model dan validasi prediksi. Pada penelitian yang telah dilakukan, implementasi algoritma *K-Nearest Neighbor (KNN)* pada dataset mendapatkan hasil *accuracy* sebesar 86.15%, *precision* sebesar 90.74%, dan *recall* sebesar 92.45%. Kemampuan algoritma untuk menangani pola non-linier dan kompleks menjadikan pilihan yang sangat baik untuk menangani dataset keuangan yang sering berfluktuasi. Maka, pemanfaatan *data mining* dengan KNN dapat meningkatkan efesiensi dan akurasi keputusan pemberian pinjaman, mengurangi risiko kredit dan meningkatkan pendapatan.

Kata Kunci: *Data Mining, K-Nearest Neighbor, Klasifikasi, Kredit*

1. PENDAHULUAN

Pesatnya pertumbuhan volume data yang dihasilkan oleh sektor finansial telah menciptakan tantangan baru bagi pengelolaan dan penggunaannya yang efektif. Salah satu permasalahan penting yang dihadapi lembaga keuangan adalah pengambilan keputusan mengenai pemberian kredit. Dimana prediksi pinjaman yang akurat menjadi semakin penting dalam menghadapi dinamika pasar dan perubahan ekonomi yang cepat. *Data mining*, sebagai ilmu yang menggunakan teknik statistik dan kecerdasan buatan, menawarkan pendekatan potensial untuk mengungkap pola dan hubungan tersembunyi dalam kumpulan data kredit. Salah satunya adalah algoritma *K-Nearest Neighbor*, yang berfokus pada konsep tetangga terdekat, dalam mengidentifikasi pola kompleks yang sulit diakses dengan metode analisis konvensional.

Algortima *K-Nearest Neighbor (KNN)* merupakan sebuah metode pengklasifikasian objek baru berdasarkan (K) tetangga terdekatnya. KNN adalah algoritma *supervised learning*, dimana hasil dari *query instance* yang baru, diklasifikasikan berdasarkan mayoritas dari kategori pada KNN. K-*Nearest Neighbor* merupakan suatu pendekatan untuk menghitung kedekatan antara kasus baru dengan kasus lama, yaitu berdasarkan pada pencocokan bobot dari beberapa fitur yang ada [1].

Penelitian ini bertujuan untuk mengeksplorasi potensi algoritma *K-Nearest Neighbor* untuk meningkatkan prediksi pemberian kredit di sektor finansial. Melalui penerapan teknik *data mining* diharapkan dapat mengungkap pola-pola tersembunyi dalam data perkreditan yang dapat memberikan wawasan berharga untuk mendukung pengambilan keputusan yang lebih akurat dan cepat.

Melalui penelitian ini, kami berharap dapat memberikan kontribusi yang signifikan terhadap pengembangan metode prediksi pinjaman di sektor finansial. Hasil penerapan algoritma K-NN diharapkan dapat meningkatkan akurasi prediksi Selain itu, penelitian ini juga diharapkan dapat membuka peluang pengembangan lebih lanjut penerapan teknologi *data mining* di bidang keuangan.

2. TINJAUAN PUSTAKA

2.1. Sektor Finansial

Pada Kamus Besar Bahasa Indonesia (KBBI) pengertian sektor diartikan dengan lingkungan suatu usaha, sedangkan pengertian finansial mempunyai arti keuangan. Maka, dapat diartikan sektor finansial adalah suatu perusahaan yang bergerak pada bidang keuangan.

Salah satu yang bergerak pada sektor finansial ialah perbankan seperti layanan pembayaran, asuransi, simpan pinjam, pengiriman uang, bahkan uang elektronik merupakan contoh produk perbankan. Saat ini, jasa keuangan dapat dibagi menjadi dua bagian. Bagian pertama non jasa, artinya, kinerja kegiatan non-keuangan. Contoh kegiatannya antara lain dana pensiun, asuransi, dan pasar modal. Kedua, jasa perbankan disebut lembaga bisnis nyata. dimana pengaturan dan pengawasan terhadap pengumpulan atau penyimpanan oleh (Ototritas Jasa Keuangan)

OJK. Uang nasabah diterima dalam bentuk produk tabungan dan dibayarkan kembali kepada masyarakat dalam bentuk pinjaman kredit [2]

2.2. Kredit

Menurut Undang – Undang No 7 1998 pengertian kredit yakni suatu penyediaan tagihan dan uang yang bisa disamakan yang berdasarkan dengan kesepakatan atau persetujuan pinjam meminjam antara pihak bank dengan pihak lainnya dan untuk mewajibkan peminjam untuk melunasi hutangnya dengan jumlah bunga, imbalan atau bagi hasilnya dalam jangka waktu yang sudah ditentukan [3].

2.3. Prediksi

Menurut Kamus Besar Bahasa Indonesia (KBBI), prediksi adalah hasil kegiatan meramalkan atau memperkirakan nilai-nilai baru dengan menggunakan data masa lalu, sehingga kesalahan (perbedaan antara sesuatu yang terjadi dan hasil yang diprediksi) dapat diminimalkan. [4] Dan definisi prediksi dapat bervariasi tergantung pada bidang kajian atau bidang ilmu.

2.4. *Data Mining*

Definisi *data mining* secara umum terbagi menjadi dua kata yaitu data dan *mining.* Data adalah kumpulan fakta atau entitas yang tidak mempunyai arti dan terabaikan. Sedangkan, *mining* adalah proses penambahan. Sehingga, *data mining* itu dapat diartikan sebagai proses penambahan data yang membuahkan hasil berupa pengetahuan [5].

Selain itu *Data mining* adalah kumpulan data hasil dari proses ekstrak data, analisa data, dan statistika data. *Data mining* adalah rangkaian data yang dapat diolah dan diproses dengan cara menggali atau menambang suatu serangkaian data yang mengandung informasi untuk menghasilkan suatu pengetahuan yang diambil melalui proses manual. [6]

2.5. Algoritma *K-Nearest Neighbor*

K-Nearest Neighbors (KNN) merupakan jenis algoritma *supervised learning* klasifikasi yang sederhana dan popular dalam *machine learning*. Prinsip kerja algoritma *K-Nearest Neighbor (KNN)* melakukan prediksi terhadap objek (data baru) berdasarkan data *training* yang jaraknya paling dekat dengan objek pada data baru tersebut [7].

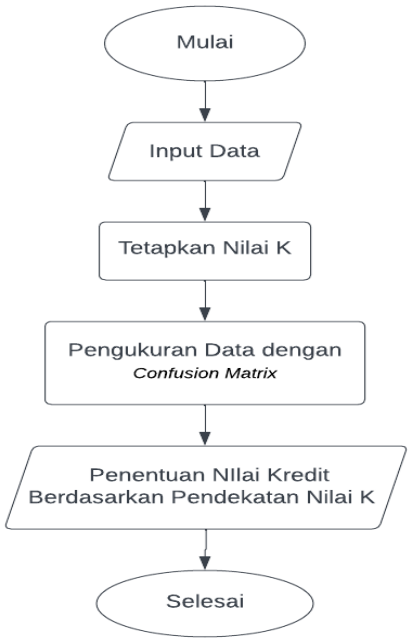
Berikut adalah kelebihan dari algoritma *K-Nearest Neighbor*:

a. Algoritma K-Nearest Neighbor sederhana serta mudah dipahami dan diimplementasikan.
b. K-Nearest Neighbor adalah algoritma non-parametrik, yang berarti tidak memerlukan tahap pelatihan yang rumit. Dimana modelnya "belajar" langsung dari data pelatihan.
c. Dapat menangani hubungan non-linier di antara fitur karena tidak membuat asumsi tentang bentuk fungsi keputusan.

Adapun beberapa kekurangan dari algoritma *K-Nearest Neighbor* sebagai berikut:

a. *K-Nearest Neighbor* memerlukan data pelatihan yang berlabel untuk mengklasifikasikan data baru. Tanpa label yang sesuai, algoritma ini tidak dapat memberikan prediksi yang baik.
b. Performa K-Nearest Neighbor sangat bergantung pada pemilihan parameter K
c. Penentuan jarak tidak jelas dan atribut apa yang harus digunakan untuk memperoleh hasil terbaik
d. Biaya komputasi yang cukup tinggi karena diperlukan perhitungan jarak dari setiap data latih derngan data ujinya

Berikut ini merupakan Flowchart Algoritma *K-Nearest Neighbors* dimana peneliti akan mengelola data.



Gambar 1. Flowchart Algoritma K-NN

Keterangan :

1. Menginput data.
2. Masukkan nilai k.
3. Menghitung Pengukuran keakuratan model data yang akan di ukur.
4. Setelah mendapatkan nilai yang akurat maka tentukan perhitungan nilai K untuk memprediksi pemberian kredit.
5. Maka hasil yang menentukan pemberian kredit dari pendekatan nilai K.

2.6. *Confusion Matrix*

Confusion Matrix adalah satu metode yang dapat digunakan untuk mengukur seberapa baik suatu model dalam melakukan klasifikasi suatu dataset pada perangkat lunak *machine learning*. *Confusion matrix* umumnya digunakan dalam konteks masalah klasifikasi biner (dua kelas). Terdapat nilai prediksi dan nilai aktual dalam *confusion matrix*, yang masing-masing memiliki empat kombinasi istilah berbeda. [8]

Tabel 1 Tabel *confusion matrix*

Confusion Matrix		Data Aktual	
		Positif	**Negatif**
Data Prediksi	**Positif**	*True Positives (TP)*	*False Positives (FP)*
	Negatif	*False Negatives (FN)*	*True Negatives (TN)*

Berikut penjelasan mengenai empat kombinasi istilah yang terdapat pada tabel 1:

a. *True Positives (TP)* adalah data positif dan di prediksi benar.
b. *True Negatives (TN)* adalah data negatif dan di prediksi benar.
c. *False Negatives (FN)* adalah data positif tetapi di prediksi negatif.
d. *False Positives (FP)* adalah data negatif tetapi di prediksi positif.

Salah satu fungsi utama *confusion matrix* adalah untuk menentukan nilai *accuracy, precision* dan *recall* untuk mewakili prediksi dan keadaan aktual dari data yang dihasilkan oleh algoritma *machine learning*. Berikut adalah rumus dalam perhitungan performa matrik.[8]

1) *Accuracy* untuk mengukur sejauh mana model dapat mengklasifikasikan data dengan benar.

$$Accuracy = \frac{TP+TN}{TP+TN+FP+FN} \quad (1)$$

2) *Precision* untuk mengukur sejauh mana kelas yang diklasifikasikan sebagai positif oleh model adalah benar-benar positif.

$$Precision = \frac{TP+TN}{TP+FP} \quad (2)$$

3) *Recall* mengukur sejauh mana model dapat menangkap atau mendeteksi seluruh kasus positif yang seharusnya terdeteksi.

$$Recall = \frac{TP+TN}{TP+FN} \quad (3)$$

2.7. *Rapid Miner*

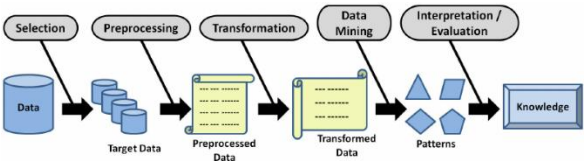
RapidMiner adalah *software* atau aplikasi *data mining* untuk mengolah kumpulan data dalam jumlah besar untuk menghasilkan sebuah informasi baru [9]. RapidMiner dapat digunakan sebagai alat, karena menyediakan berbagai macam alat berbagai metodedari evaluasi statistik sederhana seperti analisis korelasi hingga regresi, klasifikasi dan prosedur pengelompokan serta pengurangan dimensi dan optimasi parameter. [7]

3. METODE PENELITIAN

3.1. Sumber Data

Pada penelitian ini, menggunakan data sekunder yang diperoleh dari dataset kredit yang terdapat pada *website* www.kaggle.com, dengan link: https://www.kaggle.com/datasets/krishnaraj30/financ e-loan-approval-prediction-data dengan judul *"Finance Loan Approval Prediction Data"*. Diakses pada hari Rabu tanggal 1 November 2023, pukul 10.00 AM. Kaggle merupakan *platform* sumber daya ilmu data *online* yang sering digunakan para ilmuwan data dan praktisi *machine learning* untuk belajar, berkolaborasi, dan mengembangkan keterampilan mereka dalam industri yang terus berkembang ini.

3.2. Tahapan Penelitian



Gambar 2 Tahapan KDD
(Sumber: Gullo, 2015)

Pada penelitian ini menggunakan metode KDD (*Knowledge Discovery in Database*) Berikut penjelasan tahapan umum dalam proses KDD:

A. *Selection* (Seleksi Data)
Tahapan ini dimulai dengan memilih data yang relevan dan sesuai dengan tujuan analisis. Data hasil seleksi yang akan digunakan disimpan dalam suatu berkas terpisah dari basis data operasional.
B. *Preprocessing* (Pemilihan Data/Pembersihan)
Tahapan *Preprocessing* ini bertujuan untuk memastikan bahwa data yang digunakan untuk analisis bersih dan akurat. Proses ini mencakup membuang duplikasi data, membersihkan data dari noise, nilai yang hilang, atau outlier.
C. *Transformation* (Transformasi Data)
Tahapan ini dilakukan transformasi pada data untuk mendapatkan reprentasi yang sesuai. Mentransformasi bentuk data yang belum memiliki entitas yang jelas seperti pengkodean variabel kategorik dan numerik, atau transformasi lainnya yang siap diproses pada *data mining*.
D. *Data Mining*
Tahapan selanjutnya yaitu memilih metode atau algoritma yang tepat. Ini melibatkan penggunaan algoritma pembelajaran mesin, statistika, atau teknik analisis lainnya untuk mengekstrak informasi berharga.
E. *Interpretation/Evaluation* (Interpretasi/Evaluasi)
Pada tahapan terakhir adalah dilakukan menginterpretasikan hasil dari proses ekstraksi informasi dan mengevaluasi temuan. Ini melibatkan penilaian kualitas model atau temuan yang ditemukan.

4. HASIL DAN PEMBAHASAN

4.1. *Selection* (Seleksi Data)

Tahapan pertama pada penelitian ini yakni dengan memilih data yang bersumber dari *website www.kaggle.com* dengan judul "*Finance Loan Approval Prediction Data*". Pada penelitian ini menggunakan dua jenis dataset yaitu 554 data *training*

dan 324 data *testing* yang terdiri dari 13 atribut, dengan 1 *class*, 1 atribut label, 6 atribut bertipe numerik dan 6 atribut lainnya bertipe kategorik. Untuk lebih jelasnya dapat dilihat pada tabel 2 dibawah.

Tabel 2 Atribut dan tipe data

Atribut	Tipe	Keterangan
Load_ID	Kategorik	Id Peminjam
Gender	Kategorik	Jenis Kelamin
Married	Kategorik	Status Pernikahan
Dependents	Numerik	Tanggungan
Education	Kategorik	Pendidikan Terakhir
Self_Employe	Kategorik	Pekerjaan
ApplicantIncome	Numerik	Pendapatan
CoapplicantIncome	Numerik	Pendapatan Bersama
LoanAmount	Numerik	Jumlah Pinjaman
Loan_Amount_Term	Numerik	Jangka Waktu Pinjaman
Credit_History	Numerik	Riwayat Kredit
Property_Area	Kategorik	Kawasan Properti
Loan_Status	Label	Status Pinjaman

Tabel 3 Sampel Data *Training* Sebelum Diseleksi

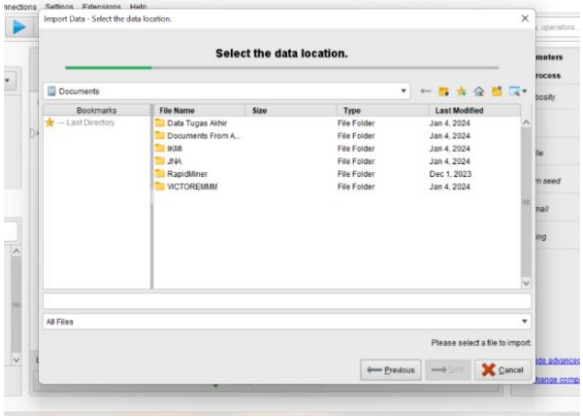
Loan_Id	Gender	Married	Dependents	Self_employe	…	…	Loan_Status
LP001002	Male	No	0	No			Y
LP001003	Male	Yes	1	No			N
LP001005	Male	Yes	0	Yes			Y
LP001006	Male	Yes	0	No			Y
LP001008	Male	No	0	No			Y
LP001011	Male	Yes	2	Yes			Y
LP001013	Male	Yes	3+	No			Y
LP001014	Male	Yes	2	No			N
LP001018	Male	Yes	1	No			Y
LP001020	Male	Yes	2	No			N

Tabel 4 Sampel Data *Testing* Sebelum Diseleksi

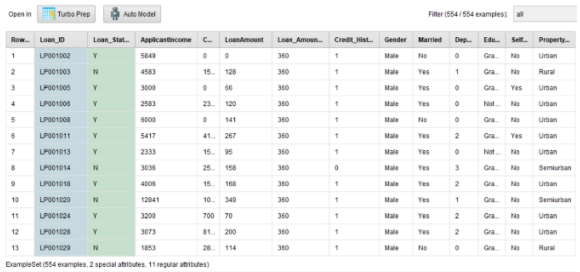
Loan_Id	Gender	Married	Dependents	Self_employe	…	…	Loan_Status
LP001015	Male	Yes	0	No			?
LP001022	Male	Yes	1	No			?
LP001031	Male	Yes	2	No			?
LP001035	Male	Yes	2	No			?
LP001051	Male	No	0	No			?
LP001054	Male	Yes	0	Yes			?
LP001055	Female	No	1	No			?
LP001056	Male	Yes	2	No			?
LP001059	Male	Yes	2				?
LP001067	Male	No	0	No			?

Tabel 3 dan tabel 4 merupakan tampilan sampel data yang belum diseleksi dan masih berisi data nilai yang tidak dibutuhkan. Sebelum dilakukan seleksi data pada *RapidMiner*, data yang akan digunakan di import seperti pada gambar 3, dilanjutkan dengan gambar 4 untuk mencari lokasi atau letak penyimpanan data.

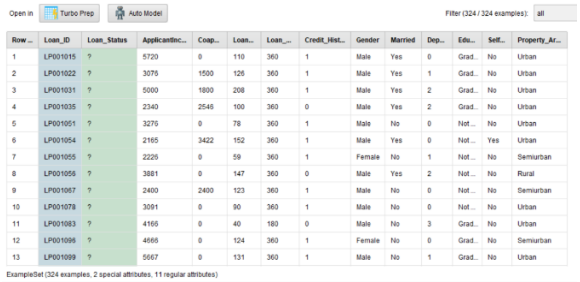
Gambar 3. Import data



Gambar 4. Mencari lokasi data



Gambar 5. Hasil seleksi data *training*



Gambar 6. Hasil seleksi data *testing*

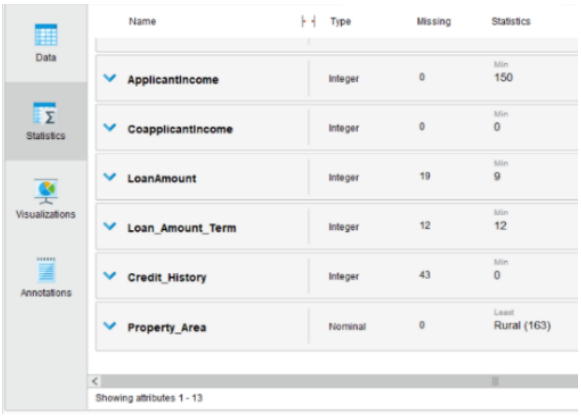
Gambar 5 dan 6 menunjukan bahwasannya hasil dari seleksi data pada *RapidMiner* dan sudah sesuai dengan apa yang dibutuhkan untuk penelitian.

4.2. *Preprocessing* (Pemilihan Data/)

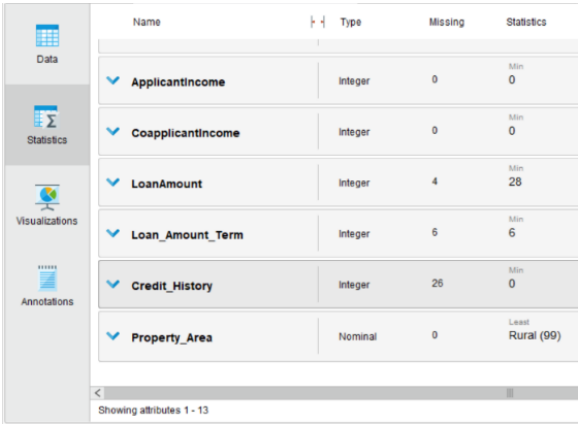
Tahap *preprocessing data* adalah dilakukannya pembersihan data *missing value* atau atribut yang memiliki data kosong. Sebelum melakukan pembersihan data, objek data harus terlebih dahulu dimuat kedalam proses *RapidMiner* dengan menggunakan dua operator *retrieve* untuk data *train* dan data *test* seperti pada gambar 7 dibawah.

Gambar 7. Menambahkan Operator *Retrieve*

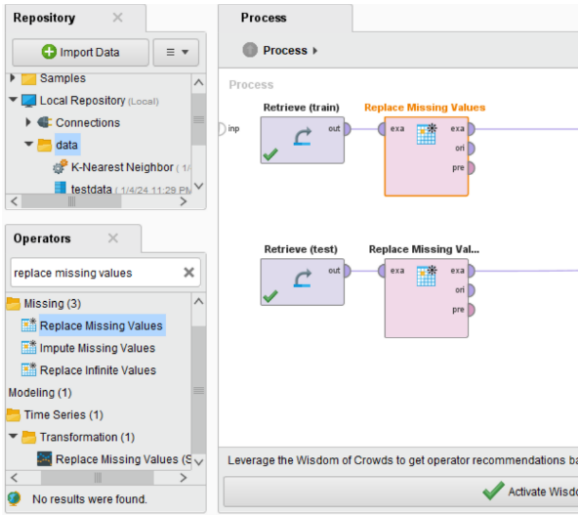
Gambar 8 dan 9 menunjukan data tersebut terdapat beberapa atribut yang memiliki data kosong atau *missing value*, maka dari itu akan dilakukan proses pembersihan data dengan menggunakan operator *replace missing values* seperti pada gambar 10.



Gambar 8. Cek Missing Value Data *Training*

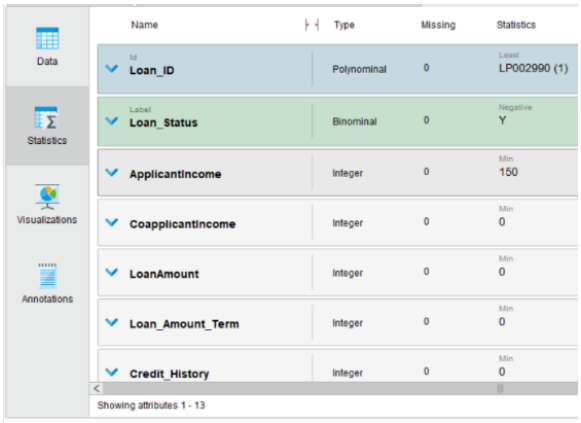


Gambar 9. Cek *Missing Value* Data *Testing*

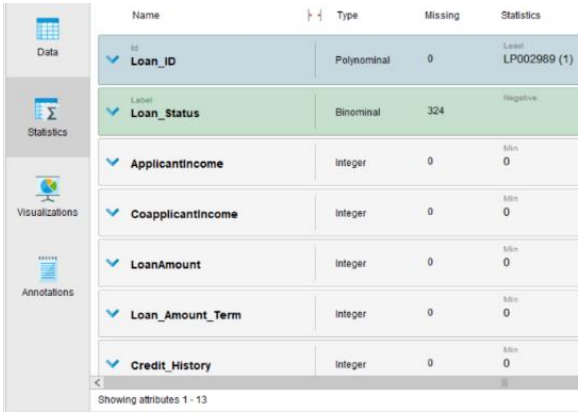


Gambar 10. Memasukan Operator *Replace Missing Value*

Setelah melakukan pembersihan data dengan memasukan operator *replace missing values* sehingga akan menampilakan statistik hasil pembersihan seperti pada gambar 11 untuk data *training*, begitupun pada gambar 12 untuk data *testing* tetapi pada atribut Loan_Status (data *testing)* masih terdapat missing dikarenakan pada tahap ini data belum mendapatkan nilai hasil prediksi.

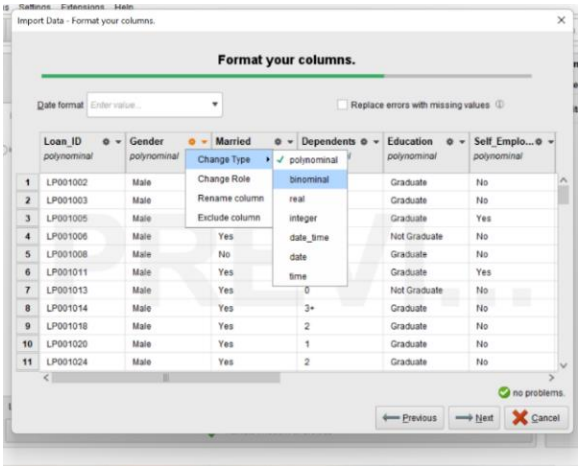


Gambar 11. Hasil *Cleansing* Data *Training*



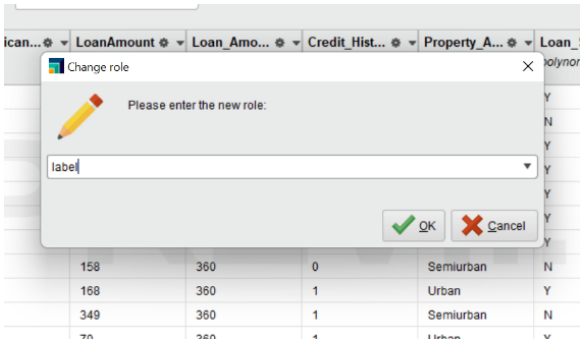
Gambar 12. Hasil *Cleansing* Data *Testing*

4.3. *Transformation* (Transformasi Data)



Gambar 13. Merubah Tipe Data

Gambar 13 menampilkan proses pengaturan untuk memastikan tipe atribut sudah sesuai dengan yang dibutuhkan agar dapat di proses, dan mengatur atau mengubah tipe atribut yang belum disesuaikan. Tipe binominal untuk atribut yang memiliki dua nilai, polynominal untuk atribut yang memiliki lebih dari dua nilai, adapun untuk nilai data yang digunakan untuk bilangan bulat (angka tanpa koma) dengan integer, dan tipe data real untuk bilangan riil (angka dengan koma).

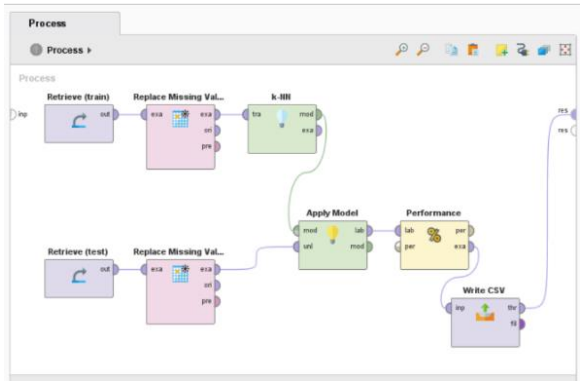


Gambar 14. Merubah Role Atribut

Pada gambar 14 merupakan proses transformasi data yang akan diberikan label yang dibutuhkan oleh algoritma *K-Nearest Neighbor,* label tersebut nantinya digunakan sebagai hasil dari prediksi kelayakan pemberian kredit.

4.4. *Data Mining*

Untuk mendapatkan hasil data prediksi pemberian kredit pada data *testing* yang belum memiliki nilai label maka dilakukan proses *data mining* algoritma *K-Nearest Neighbor (KNN)* menggunakan *tools RapidMiner* dengan model berikut:



Gambar 15. Model *K-Nearest Neighbor* 1

Berdasarkan gambar 15 diatas adalah model dari implementasi algoritma *K-Nearest Neighbor (KNN)* dengan *tools RapidMiner* yaitu dengan menjalankan beberapa operator diantaranya:

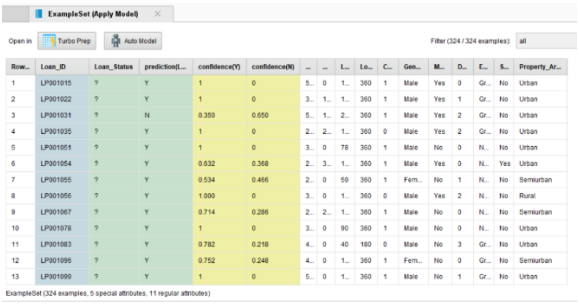
1) *Retrieve (train)*
   Operator ini digunakan untuk memuat data latih yang akan digunakan. Data latih adalah data yang digunakan untuk melatih algoritma KNN sehingga dapat mengenali pola dan hubungan dalam data.
2) *Retrieve (test)*
   Operator ini digunakan untuk memuat data uji yang akan digunakan. Data uji adalah data yang digunakan untuk menguji kinerja algoritma KNN setelah dilatih dengan data latih.
3) *Replace Missing Values*
   Operator ini digunakan untuk memberikan solusi pada nilai yang hilang dalam dataset. Jika terdapat nilai yang hilang pada dataset, operator

ini akan menggantikan nilai tersebut dengan nilai yang sesuai, seperti nilai rata-rata atau nilai median.

4) *K-Nearest Neighbor (KNN)*
   Algoritma data mining yang digunakan pada penelitian ini. Algoritma KNN digunakan untuk mengklasifikasikan data baru berdasarkan mayoritas dari kategori pada K tetangga terdekatnya.
5) *Apply Model*
   Operator ini digunakan untuk menerapkan model pada data baru yang akan diuji. Setelah algoritma KNN dilatih dengan data latih, model ini akan digunakan untuk memprediksi kategori data baru.
6) *Performance*
   Operator ini digunakan untuk mengevaluasi kinerja model atau algoritma. Operator ini akan menghasilkan metrik evaluasi seperti akurasi, presisi, dan recall untuk menentukan seberapa baik algoritma KNN dalam melakukan klasifikasi data.
7) *Write* CSV
   Operator yang digunakan untuk menyimpan data atau hasil analisis dalam format CSV. Hasil prediksi dari algoritma KNN akan disimpan dalam format CSV untuk digunakan dalam analisis selanjutnya.

Dalam model ini, data latih dan data uji dimuat menggunakan operator *Retrieve*. Kemudian, operator *Replace Missing Values* digunakan untuk mengatasi nilai yang hilang dalam dataset. Setelah itu, algoritma KNN dilatih menggunakan data latih dan diuji menggunakan data uji dengan menggunakan operator *K-Nearest Neighbor* dan *Apply Model*. Hasil prediksi dari algoritma KNN akan dievaluasi menggunakan operator *Performance* untuk menentukan seberapa baik kinerja algoritma. Terakhir, hasil prediksi akan disimpan dalam format CSV menggunakan operator *Write* CSV untuk digunakan dalam analisis selanjutnya.

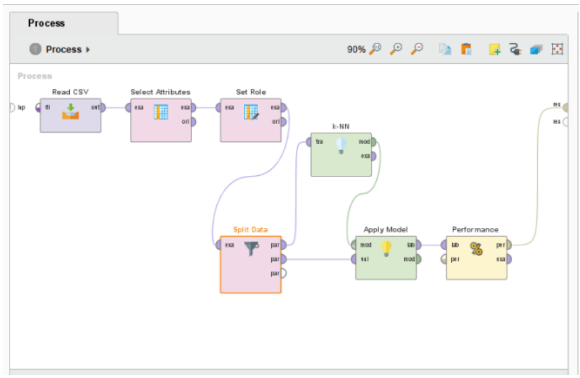
Pada data *training* diterapkannya algoritma klasifikasi yakni *K-Nearest Neighbor (KNN)*, sedangkan pada data *testing* guna menerapkan model KNN serta menghasilkan dataset prediksi untuk digunakan dalam analisis selanjutnya.



Gambar 16. *Example set*

Pada gambar 16 menunjukan tampilan yang didapatkan setelah dilakukan proses *data mining* yaitu *Example Set*, yang dimana berisi informasi prediksi yang dapat menentukan keputusan dalam pemberian kredit. Jika hasil prediksi tertulis "Y", maka pinjaman kredit dapat disetujui oleh pihak debitur. Sebaliknya jika hasil prediksi tertulis "N", maka pinjaman kredit tidak dapat disetujui oleh pihak debitur. Selain prediksi yang terdapat pada *Example Set* yakni parameter *confidence* sebagai besar persentase keyakinan dalam persetujuan pemberian kredit tersebut.

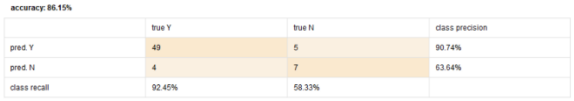
4.5. *Evaluation* (Evaluasi)



Gambar 17. Model *K-Nearest Neighbor* 2

Gambar 17 menujukan model algoritma *K-Nearest Neighbor* dalam pengujian untuk menganalisis lanjutan yang dapat mengukur kinerja algoritma *K-Nearest Neighbor (KNN)* menggunakan *tools RapidMiner* dengan menjalankan beberapa operator antara lain:

1) *Read CSV*
   Operator ini digunakan untuk mengimpor data dari file CSV ke dalam *RapidMiner*.
2) *Select Attributes*
   Operator ini digunakan untuk memilih atau mengabaikan atribut tertentu dari dataset.
3) *Set Role*
   Operator ini digunakan untuk menentukan peran atau jenis data dari suatu atribut.
4) *Split Data*
   Operator ini digunakan untuk membagi dataset menjadi dua bagian, pada penelitian ini memakai rasio 8:2 untuk data latih dan data uji.
5) *K-Nearest Neighbor (KNN)*
   Algoritma *data mining* yang digunakan pada penelitian ini dengan menggunakan nilai K=3.
6) *Apply Model*
   Operator ini digunakan sebagai penghubung algoritma KNN dengan *performance*.
7) *Performance*
   Operator ini digunakan untuk mengevaluasi kinerja model atau algoritma.



Gambar 18. *Performance vector*

Setelah proses model kedua dilakukan, maka akan didapatkan *performance vector* yang dapat mengukur sejauh mana model mampu mengidentifikasi kelas-kelas. Gambar 18 menampilkan hasil dari pengujian menggunakan algoritma *K-Nearest Neighbor* dengan nilai K = 3, hasil eksperimen menunjukkan bahwa nilai K = 3 menghasilkan tingkat akurasi, presisi, dan *recall* yang baik dalam memprediksi pemberian kredit di sektor finansial. Oleh karena itu, nilai K = 3 dipilih sebagai nilai optimal untuk algoritma KNN pada penelitian ini. Pada dataset penelitian mencapai tingkat *accuracy* sebesar 86.15%, *precision* sebesar 90.74%, dan *recall* sebesar 92.45%. dengan rincian prediksi Y dan *true* Y sebanyak 49 data, prediksi N dan *true* Y sebanyak 4 data, prediksi Y dan *true* N sebanyak 5 data, prediksi N dan *true* N sebanyak 7 data.

Berikut adalah perhitungan manual dari tingkatan akurasi yang dihasilkan *rapidminer* yang:

1) *Accuracy*

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

$$\frac{49 + 7}{49 + 7 + 4 + 5}$$

$$\frac{56}{65} = \underline{86.15\%}$$

2) *Precision*

$$\frac{TP}{TP + FP}$$

$$\frac{49}{49 + 5}$$

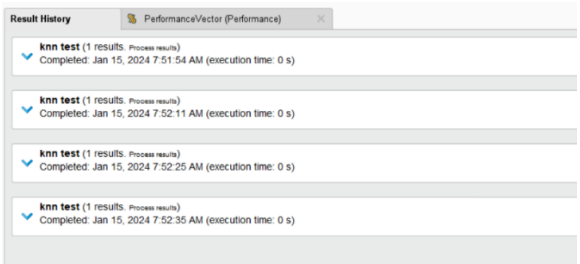
$$\frac{49}{53} = \underline{90.74\%}$$

3) *Recall*

$$\frac{TP}{TP + FN}$$

$$\frac{49}{49 + 4}$$

$$\frac{49}{54} = \underline{92.45\%}$$



Gambar 19. Result history

Gambar 19 merupakan daftar riwayat pengujian menggunakan algoritma *K-Nearest Neighbor* dengan *tools RapidMiner*, daftar tersebut dimana berisi informasi waktu eksekusi yang hanya membutuhkan waktu 0 *second* atau 0 detik, sehingga dapat diartikan bahwa eksekusi dengan menggunakan algoritma *K-Nearest Neighbor* ini berjalan dengan cepat.

5. KESIMPULAN DAN SARAN

Data yang bersumber dari *www.kaggle.com* dengan judul "*Finance Loan Approval Prediction Data*" sebanyak 324 data *testing*, yang terdiri dari 13 atribut, 12 sebagai penentu dalam perhitungan pada *data mining* dan 1 atribut sebagai label hasil. Setelah diimplementasikan pada algoritma *K-Nearest Neighbor* dengan nilai K 3 maka menghasilkan tingkatan *accuracy* sebesar 86.15%, *precision* sebesar 90.74%, dan *recall* sebesar 92.45%. sehingga dapat dikatakan bahwa algoritma KNN membantu untuk mengambil keputusan dalam pemberian kredit pada sektor finansial dengan mudah dan cepat. Dari hasil penelitian yang sudah dibahas ini menyarankan agar diadakan penelitian lebih lanjut guna meningkatkan nilai akurasi untuk menentukan kelayakan pemberian kredit. Selain itu jumlah data yang lebih besar atau metode yang berbeda dapat digunakan pada penelitian selanjutnya agar dapat menghitung nilai performa yang baru.

DAFTAR PUSTAKA

[1] M Syukri Mustafa and Wayan Simpen, "Implementasi Algoritma K-Nearest Neighbor (KNN) Untuk Memprediksi Pasien Terkena Penyakit Diabetes Pada Puskesmas Manyampa Kabupaten Bulukumba," *Seminar Ilmiah Sistem Informasi dan Teknologi Informasi*, vol. 8, pp. 1–10, Feb. 2019.

[2] A. Susilo, "Perbandingan Kinerja K-Nearest Neighbors dan Naive Bayes Untuk Klasifikasi Perilaku Nasabah Pada Pembayaran Kredit Bank," *Jurnal Sains dan Teknologi (JSIT)*, vol. 3, no. 3, pp. 364–379, Nov. 2023, doi: 10.47233/jsit.v3i3.1264.

[3] S. Harlina, "Data Mining Pada Penentuan Kelayakan Kredit Menggunakan Algoritma K-NN Berbasis Forward Selection," *Creative Communication and Innovative Technology Journal*, vol. 11, pp. 236–244, Jul. 2018.

[4] Rahmadini, E. E. L. Lubis, A. Priansyah, Yolanda R.W.N, and T. Meutia, "Penerapan Data Mining Untuk Memprediksi Harga Bahan Pangan Di Indonesia Menggunakan Algoritma K-Nearest Neighbor," *Jurnal Mahasiswa Akuntansi Samudra (JMAS)*, vol. 4, pp. 223–225, Aug. 2023.

[5] S. Kom. , M. Kom. Dicky Nofriansyah and M. Sc. DR. Ir. Gunadi Widi Nurcahyo, *Algoritma Data Mining dan Pengujian*. Yogyakarta: Deepublish, 2019.

[6] M Faisal, Wiranti Sri Utami, and Sarah Parmica, "Implementasi Algoritma K-Nearest Neighbor (KNN) Dalam Memprediksi Indeks Kemiskinan," *Journal Sensi Online*, vol. 9, no. 1, pp. 11–23, Feb. 2023.

[7] Edi Junaedi, Amril Mutoi Siregar, and Euis Nurlaelasari, "Implementasi C4.5 Dan Algoritma K Nearest Neighbor Untuk Prediksi Kelayakan Pemberian Kredit Menggunakan RapidMiner

Studio," *Scientific Student Journal for Information, Technology and Science*, vol. 3, no. 1, pp. 83–90, Jan. 2022.

[8] Hardiana Said, Nurhafifah Matondang, and Helena Nurramdhani Irmanda, "Penerapan Algoritma K-Nearest Neighbor Untuk Memprediksi Kualitas Air Yang Dapat Dikonsumsi," *Jurnal Teknologi Informasi Techno.com*, vol. 23, pp. 256–267, May 2022.

[9] Taufik Hidayat, Yuni Handayani, and Ahmad Syaifudin, "Implementasi Algoritma K-Nearest Neighbor Untuk Meningkatkan Penjualan Produk Meuble dan Furniture," *Jurnal Ilmiah Rekayasa dan Manajemen Sistem Informasi*, vol. 9, pp. 118–124, Feb. 2023.

[10] F. Gullo, "From Patterns in Data to Knowledge Discovery: What Data Mining Can Do," *Phys Procedia*, vol. 62, pp. 18–22, Feb. 2015, doi: 10.1016/j.phpro.2015.02.005.