

Determining The Loan Feasibility of Bank Customers Using Naïve Bayes, K-Nearest Neighbors And Linear Regression Algorithms

Aniec Anafisah Pratiwi^{1*}, Wahyuning Tyas Saraswati², Rizky Firman Ardiansyah³, Erik Halma Rouf⁴, Rizki Adhi Pratama⁵

^{1,2,3,4,5}Information Systems, Faculty of Computer Science, Amikom Purwokerto University, Indonesia

Email Korespondensi: ^{1*}aniecanafisah@gmail.com.

Abstrak– Dalam industri keuangan, pemberian kredit kepada nasabah adalah salah satu kegiatan inti dalam sektor keuangan yang memiliki dampak yang signifikan terhadap ekonomi dan pertumbuhan bisnis. Kredit adalah penyediaan uang atau tagihan yang dapat dipersamakan dengan itu, berdasarkan persetujuan pinjam meminjam antar Bank dengan pihak lain yang mewajibkan pihak peminjam melunasi hutangnya setelah jangka waktu tertentu dengan pemberian bunga. Namun, dalam proses tersebut lembaga keuangan perlu melakukan penilaian kelayakan pemberian kredit kepada nasabah yang mengajukan kredit. Untuk memudahkan menentukan kelayakan pemberian kredit kepada nasabah diperlukan adanya metode analisis yang akurat dan efektif untuk dapat membantu menyelesaikan masalah dalam menentukan klasifikasi kelayakan pemberian kredit kepada nasabah dengan menerapkan algoritma *Naïve Bayes*, *K-Nearest Neighbors* (K-NN) dan *Linear Regression*. Berdasarkan hasil pengujian yang telah dilakukan dengan menggunakan ketiga algoritma tersebut menunjukkan hasil nilai akurasi pada K-NN sebesar 87,837%, perhitungan menggunakan algoritma *Naïve Bayes* didapatkan nilai akurasi sebesar 88,917%, sedangkan perhitungan menggunakan algoritma *Linear Regression* menghasilkan nilai *Mean absolute error* sebesar 6,703. Dapat disimpulkan bahwa dalam penentuan kelayakan kredit bank menggunakan metode algoritma *Naïve Bayes* lebih akurat jika dibandingkan dengan algoritma K-NN maupun *Linear Regression*.

Kata Kunci: KNN, KNIME, linear regression, naïve bayes

Abstract– In the financial industry, lending to customers is one of the core activities in the financial sector which has a significant impact on the economy and business growth. Credit is the provision of money or bills that can be equated with it, based on a loan agreement between banks and other parties that requires the agreement to repay the debt after a certain period of time by providing interest. However, the process within these financial institutions needs to assess the feasibility of granting credit to customers who apply for credit. To facilitate the determination of eligibility for granting credit to customers, an accurate and effective analytical method is needed to help solve problems in determining the eligibility classification for granting credit to customers by applying the Naive Bayes, K-Nearest Neighbors (K-NN) and Linear Regression algorithms. Based on the results of the tests that have been carried out using the three algorithms obtained, the results show an accuracy value on K-NN of 87.837%, calculations using the Naive Bayes algorithm have an accuracy value of 88.917%, while calculations using the Linear Regression algorithm produce a Mean absolute error value of 6.703. It can be concluded that in bank creditworthiness fraud using the Naive Bayes algorithm method is more accurate when compared to the K-NN and Linear Regression algorithms.

Keywords: KNN, KNIME, linear regression, naïve bayes

1. PENDAHULUAN

Dalam industri keuangan, pemberian kredit kepada nasabah adalah salah satu kegiatan inti dalam sektor keuangan yang memiliki dampak yang signifikan terhadap ekonomi dan pertumbuhan bisnis. Kredit adalah penyediaan uang atau tagihan yang dapat dipersamakan dengan itu, berdasarkan persetujuan pinjam meminjam antar Bank dengan pihak lain yang mewajibkan pihak peminjam melunasi hutangnya setelah jangka waktu tertentu dengan pemberian bunga [1]. Namun, dalam proses tersebut lembaga keuangan perlu melakukan penilaian kelayakan pemberian kredit kepada nasabah yang mengajukan kredit bertujuan untuk menentukan keputusan yang tepat apakah suatu kredit dapat diberikan kepada nasabah. Karena, penilaian kelayakan memiliki dampak yang signifikan terhadap risiko kreditur dan stabilitas sektor keuangan secara keseluruhan.

Lembaga keuangan perlu mengidentifikasi risiko yang terkait dengan pemberian kredit kepada nasabah. Risiko tersebut mencakup risiko kredit, risiko likuiditas, risiko operasional, dan risiko reputasi. Oleh karena itu, lembaga keuangan perlu menerapkan metode analisis yang akurat dan efektif dalam menilai kelayakan pemberian kredit. Beberapa riset bidang komputer untuk mengurangi risiko kredit telah banyak dilakukan dalam rangka menghindari kehancuran suatu perusahaan pembiayaan [2].

Penilaian kelayakan pemberian kredit adalah penting untuk meminimalkan risiko kerugian yang mungkin timbul dari pemberian kredit kepada nasabah yang tidak memiliki kemampuan atau berniat untuk tidak membayar kembali pinjaman tersebut. Adapun kriteria-kriteria yang dijadikan pihak bank dalam menentukan calon penerima kredit

diantaranya status kredit, produktivitas usaha, kondisi usaha, jaminan, koletibilitas, watak (*character*), penghasilan [3]. Kredit macet ataupun kredit bermasalah ialah sesuatu keadaan dimana pihak peminjam tidak sanggup melunasi pinjaman kepada pihak yang berikan pinjaman cocok waktu yang sudah ditetapkan [4]. Salah satu penyebab terjadinya kredit macet bisa disebabkan oleh kurang cermatnya perusahaan dalam pemberian kredit [5]. Agar menghindari terbentuknya kredit macet, seseorang analisis kredit perbankan wajib mengambil keputusan yang tepat untuk menerima maupun menolak pengajuan kredit. Buat mengenali kelayakan kredit di masa mendatang, dibutuhkan terdapatnya prediksi yang akurat, salah satunya memakai metode informasi mining [4].

Data mining adalah suatu proses yang menggunakan teknik statistik, matematika, kecerdasan tiruan, dan machine learning untuk mengekstraksi dan mengidentifikasi informasi yang bermanfaat dan pengetahuan yang terkait dari berbagai database yang besar [6]. Dalam kaitannya data mining memiliki beberapa teknik salah satunya adalah teknik klasifikasi. Klasifikasi adalah pemrosesan untuk menemukan sebuah model atau fungsi yang menjelaskan dan mencirikan konsep atau kelas data, untuk kepentingan tertentu. Klasifikasi memiliki beberapa Algoritma klasifikasi data mining salah satunya yaitu *K-Nearest Neighbors* (K-NN) [7].

Penelitian ini bertujuan untuk menerapkan algoritma *Naive Bayes* dan *K-Nearest Neighbors* (K-NN), *Naive Bayes* dan *Linear Regression* dalam menentukan kelayakan pemberian kredit kepada nasabah bank XYZ. Ketiga algoritma ini memiliki pendekatan yang berbeda namun dapat memberikan wawasan yang berharga dalam menilai risiko kredit.

Pada penelitian yang pernah dilakukan sebelumnya oleh Muryono & Irwansyah (2020) yang berjudul "*Implementasi Data Mining Untuk Menentukan Kelayakan Pemberian Kredit Dengan Menggunakan Algoritma K-Nearest Neighbors* (K-NN)". Penelitian tersebut dilakukan dengan tujuan untuk dapat membantu melakukan proses analisis kredit agar dapat menghasilkan informasi yang tepat apakah nasabah yang akan mengajukan kreditnya layak atau tidaknya sehingga dapat melihat potensi pembayaran kredit yang dilakukan nasabah. Data tersebut diambil dari hasil observasi, wawancara, dan dokumentasi, yang nantinya akan diuji dengan menggunakan metode *K-Nearest Neighbor* (K-NN). Dari hasil pengujian Algoritma *K-Nearest Neighbors* (K-NN) yang telah dilakukan menggunakan tools RapidMiner, diperoleh hasil akurasi sebesar 93.33% pada pengujian ke 5.

Penelitian oleh Gultom & Simanjuntak (2021), "*Prediksi Tingkat Kelancaran Pembayaran Kredit Bank Dengan Algoritma Naive Bayes dan K-Nearest Neighbor*" menggunakan kombinasi dua metode klasifikasi yaitu *Naive Bayes* dan *K-Nearest Neighbor*. Penelitian ini mengambil data dari dataset nasabah kredit dari Bank Danamon. Hal ini dilakukan untuk memprediksi kelancaran pembayaran kredit kedepannya. Pengujian yang dilakukan sebanyak 3 kali, menghasilkan akurasi yang tertinggi yaitu 80%.

Penelitian Muryono et al. (2021) tentang "*Perbandingan Algoritma K-Nearest Neighbors, Decision Tree, Dan Naive Bayes Untuk Menentukan Kelayakan Pemberian Kredit*" menghasilkan perbandingan tiga algoritma tersebut menunjukkan *Decision Tree* memiliki tingkat akurasi yang paling tinggi. Hal tersebut menunjukkan bahwa kinerja *Decision Tree* lebih baik dibanding dengan *K-Nearest Neighbor* dan *Naive Bayes*.

Penelitian Harlina et al. (2022) dengan judul "*Penerapan Algoritma K-Nearest Neighbor Untuk Klasifikasi kelayakan Calon Nasabah Kredit Berbasis Web*". Penelitian tersebut memiliki tujuan untuk memudahkan menentukan kelayakan calon nasabah mendapatkan kredit diperlukan adanya sistem yang dapat membantu menyelesaikan masalah dalam menentukan klasifikasi kelayakan calon nasabah untuk mendapatkan kredit dengan menerapkan algoritma *K-Nearest Neighbor* berbasis web, sehingga dapat juga membantu perusahaan dalam menentukan kelayakan calon nasabah mendapatkan kredit. Berdasarkan hasil pengujian, aplikasi sudah berjalan dengan baik sesuai perencanaan dan perancangan, begitupun dengan metode KNN menunjukkan data training dan testing yang diolah menghasilkan peringkat layak dan tidak layak dalam pemberian kredit yang diolah dari sampel yang bersumber data data obyek penelitian kami, sehingga siap untuk diimplementasikan pada Kantor PT Mandala Finance.

Penelitian Neighbor et al. (2020) tentang "*Klasifikasi Penentuan Pengajuan Kartu Kredit Menggunakan K-Nearest Neighbor*". Dengan melakukan penelitian ini diharapkan dapat mempermudah pihak bank atau pihak analis untuk menentukan kategori kartu kredit untuk nasabah bank secara tepat. Metode penelitian yang digunakan yaitu dengan menerapkan metode *K-Nearest Neighbor* untuk mengklasifikasikan calon nasabah dalam pembuatan kartu kredit sesuai dengan kategori nasabah dengan menggunakan data nasabah pada Bank BNI Syariah Surabaya. Hasil dari penelitian ini menunjukkan bahwa rata-rata nilai *precision* sebesar 92%, nilai *recall* sebesar 83%, dan nilai *accuracy* sebesar 93%.

Dan penelitian Dewi (2019) dengan judul penelitian "*Komparasi Metode Algoritma Data Mining pada Prediksi Uji Kelayakan Credit Approval pada Calon Nasabah Kredit Perbankan*". Dan dalam penelitian ini dilakukan pembuatan model menggunakan algoritma Klasifikasi yaitu *Neural Network*, *Naive Bayes* dan K-NN, menggunakan data pemasararan pada Bank. Algoritma *Neural Network* memiliki Akurasi yang lebih tinggi dengan nilai 89.81% dibandingkan dengan 2 algoritma lain nya sedangkan di urutan ke dua yaitu algoritma K-NN dengan nilai 87.79% dan yang terakhir algoritma *Naive Bayes* dengan Nilai 84.60%. Dengan demikian algoritma *Neural Network* dapat memberikan pemecahan untuk permasalahan penentuan kelayakan konsumen yang mendapatkan pembiayaan kredit.

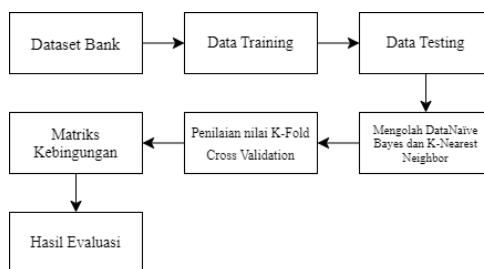
2. METODOLOGI PENELITIAN

Algoritma *Naive Bayes* adalah metode algoritma pengklasifikasian statistik yang biasanya digunakan untuk memprediksi sebuah gudang suatu kelas [11]. *Naive Bayes* merupakan machine learning yang menggunakan perhitungan probabilitas yang menggunakan konsep pendekatan Bayesian. Penggunaan teorema Bayes pada algoritma *Naive Bayes* yaitu dengan mengkombinasikan prior probability dan probabilitas bersyarat dalam sebuah rumus yang bisa digunakan untuk menghitung probabilitas tiap klasifikasi yang mungkin [12]. Algoritma *K-Nearest Neighbor* (KNN) ialah algoritma yang dapat melakukan prediksi. Prinsip kerja algoritma ini sangat sederhana yakni dengan menghitung jarak paling dekat. Artinya jika ada masukkan objek yang baru yang tidak dikenali, maka algoritma K-NN akan mencari objek yang terdekat dengan objek yang baru dimasukkan pada basis data, dan kemudian melakukan tindakan yang sama kepada objek yang baru dimasukkan kepada objek yang terdekat [8]. Sedangkan Regresi linier adalah metode statistika yang digunakan untuk membentuk model hubungan antara variabel terikat (dependen; respon; Y) dengan satu atau lebih variabel bebas (independen, prediktor, X). Apabila banyaknya variabel bebas hanya ada satu, disebut sebagai regresi linier sederhana, sedangkan apabila terdapat lebih dari 1 variabel bebas, disebut sebagai regresi linier berganda [13].

Setiap algoritma memiliki aspek yang unik dan mempengaruhi kinerja dan interpretabilitas model. Dengan pemahaman yang baik tentang kekuatan dan kelemahan masing-masing algoritma, pemangku kepentingan industri keuangan dapat membuat keputusan yang lebih baik dalam menentukan algoritma yang paling cocok untuk penilaian kelayakan pemberian kredit. Dalam dunia yang semakin tergantung pada analisis data dan kecerdasan buatan, pemahaman tentang penggunaan algoritma *Naive Bayes*, *K-Nearest Neighbors* (K-NN) dan *Linear Regression* ini dapat memberikan wawasan yang berharga bagi institusi keuangan dalam mengambil keputusan yang lebih baik dan mengelola risiko kredit dengan lebih efektif.

Tahap Penelitian

Penelitian ini dilakukan untuk mengetahui performa dari beberapa algoritma klasifikasi yaitu *Naive Bayes*, *K-Nearest Neighbors* (K-NN) dan *Linear Regression*.



Gambar 1. Diagram Alur [7]

Pada diagram alur di atas dapat dijelaskan sebagai berikut:

3.1. Dataset Bank

Tabel 1. Dataset Bank

No	Age	Job	Marital	Education	Loan	Contact	Duration	Response
1	56	H M	M	Basic 4y	No	Telp	261	N o
2	57	SV	M	High school	No	Telp	149	N o
3	37	SV	M	High school	No	Telp	226	N o
...

...
41.1	56	RT	M	Univer	No	Cell	189	N	
86				sity				o	
				gedree					
41.1	44	TC	M	Profes	No	Cell	442	N	
87		H		sional				o	
				couse					
41.1	74	RT	M	Profes	No	Cell	239	N	
88				sional				o	
				couse					

⇒ Keterangan Job: Housemaid (HM), Services (SV), Retired (RT), Technician (TCH).

Pada Tabel 1, data yang didapatkan berupa dataset bank dengan jumlah data yang dipakai sebanyak 41.189 data, dan variable yang digunakan sebagai bahan penelitian yaitu variable age, job, marital, education, loan, contact, duration, dan y. Kemudian, hasil seleksi yang digunakan untuk proses *data mining*, di simpan dalam suatu berkas, terpisah dari basis data operasional.

3.2. Analisis Data

Analisis data merupakan proses pengumpulan, mengolah serta mengubah data, dengan tujuan untuk menemukan informasi yang bermanfaat agar memberikan petunjuk dalam pengambilan suatu keputusan [14]. Pada tahap analisis data ini dilakukan untuk menentukan variable pada dataset bank yang akan diimplementasikan ke dalam prediksi kelayakan kredit bank. Pada tahap analisis data terdiri dari data pelatihan (training data) dan data uji (testing data). Di mana training data akan digunakan untuk melatih algoritma dalam mencari model yang sesuai, sedangkan testing data dipakai untuk menguji dan mengetahui performa model yang didapatkan pada tahap testing.

3.3. Mengolah Data

Pada tahapan ini dilakukan penyiapan data awal, pemilihan atribut atau variabel yang akan dianalisis, validasi variabel yang ada serta persiapan transformasi data [4]. Tahap pengolahan data dilakukan dengan menyediakan data nasabah bank yang mengajukan kredit dalam format excel dan memilih variable yang akan digunakan untuk pengujian algoritma *Naïve Bayes*, *K-Nearest Neighbor* dan *Linear Regression* dengan menggunakan aplikasi K-NIME. Dilakukan pengolahan data khusus untuk masing-masing algoritma. Misalnya, pada algoritma K-NN, fitur-fitur numerik dapat dinormalisasi agar memiliki skala yang sama. Pada *Naïve Bayes*, fitur-fitur kategori perlu diubah menjadi representasi numerik. Sedangkan pengolahan data menggunakan algoritma regresi linear melibatkan langkah-langkah penting untuk menganalisis hubungan linier antara variabel independen dan dependen. Pengolahan data menggunakan algoritma regresi linear memberikan wawasan tentang hubungan linier antara variabel dan dapat membantu dalam pemberian kelayakan kredit dengan memprediksi kemungkinan pembayaran kembali oleh peminjam.

3.4. Penentuan nilai K-Fold Cross Validation.

Menurut penelitian sebelumnya, alur yang digunakan dalam penelitian adalah dengan memasukkan variabel data dalam perhitungan sehingga didapatkan parameter k (jumlah tetangga terdekat), menghitung kuadrat jarak *eucliden* pada data training, mengurutkan hasil dari kuadrat jarak *eucliden* secara *ascending*, menentukan variabel Y berdasarkan nilai pada k, menentukan probabilitas tertinggi sehingga dapat dihitung *precision*, *recall*, dan *accuracy* pada aplikasi tersebut.

3.5. Confusion Matriks

Dari confusion matrix, kita dapat menghitung metrik evaluasi seperti akurasi, presisi, *recall*, dan *F1-score*. Akurasi mengukur sejauh mana model dapat mengklasifikasikan dengan benar seluruh data. Presisi mengukur sejauh mana model dapat mengidentifikasi dengan benar data positif dari keseluruhan data yang diprediksi positif. Recall mengukur sejauh mana model dapat mengklasifikasikan dengan benar data positif dari keseluruhan data aktual yang positif. *F1-score* merupakan rata-rata harmonik antara presisi dan *recall*, memberikan gambaran yang seimbang tentang performa model. Didapat bahwa dalam aplikasi tersebut dengan algoritma KNN maka didapatkan nilai klasifikasi berupa jenis kartu kredit yang sesuai dengan nasabah yang mengajukan kredit kepada bank.

3.6. Hasil Evaluasi

Berdasarkan hasil evaluasi berdasarkan *confusion matriks* yang didapatkan, kita dapat membuat kesimpulan tentang performa model dalam penyediaan kelayakan kredit bank. Nilai akurasi yang tinggi menunjukkan bahwa model

mampu mengklasifikasikan dengan benar sebagian besar data. Nilai presisi yang tinggi menunjukkan bahwa model cenderung memberikan prediksi yang akurat untuk data positif. Nilai *recall* yang tinggi menunjukkan bahwa model mampu mengklasifikasikan dengan benar sebagian besar data aktual yang positif. Nilai *F1-score* yang tinggi menunjukkan keseimbangan antara presisi dan *recall*. Penelitian ini dapat disimpulkan bahwa hasil dari proses klasifikasi menggunakan aplikasi K-NIME ini yaitu untuk mengetahui akurasi data klasifikasi nasabah pengajuan kredit pada suatu bank.

3. HASIL DAN PEMBAHASAN

4.1. Pengumpulan Data

Pengumpulan data dilakukan untuk memperoleh informasi yang dibutuhkan dalam rangka mencapai tujuan penelitian. Metode pengumpulan data yang umum digunakan dalam suatu penelitian adalah: kuesioner, observasi, wawancara [15]. Pengumpulan data kelayakan kredit bank melibatkan identifikasi variabel relevan, seperti pendapatan, riwayat kredit, dan jumlah tanggungan. Data yang digunakan adalah data bank yang berisi data nasabah bank yang mengajukan kredit kepada bank. Selanjutnya, data diverifikasi, divalidasi, dan disimpan dalam format yang sesuai. Penting untuk menjaga kerahasiaan data dan mematuhi kebijakan privasi. Software atau aplikasi yang digunakan untuk penelitian dataset adalah aplikasi K-NIME.

4.2. Seleksi Data

Dalam pemberian kelayakan kredit bank menggunakan algoritma K-NN, *Naive Bayes* dan *Linear Regression*, seleksi data dilakukan untuk memilih variabel yang relevan dan mengumpulkan data yang diperlukan. Seleksi data adalah proses menganalisis data-data yang relevan dari database karena sering ditemukan bahwa tidak semua data dibutuhkan dalam proses *data mining*. Data tersebut dipilih dan diseleksi dari database untuk di analisis. Sumber data yang digunakan dalam penelitian ini berasal dari dataset yang diambil dari kaggle. Data kemudian di *preprocessing*, termasuk pembersihan dan persiapan data. Setelah itu, data dibagi menjadi subset training dan testing. Selanjutnya, dilakukan seleksi fitur untuk memilih subset variabel yang paling relevan. Jika perlu, normalisasi data dilakukan untuk mengubah skala variabel menjadi seragam.

4.3. Analisa Sistem

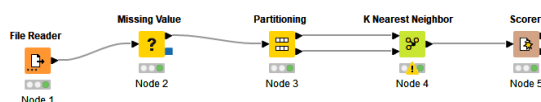
Analisis sistem pada pemberian kelayakan kredit bank menggunakan algoritma K-NN, *Naive Bayes* dan *Linear Regression* melibatkan pemahaman tentang arsitektur sistem, pengelolaan data, *preprocessing* data, seleksi fitur, evaluasi model, kinerja sistem, keamanan dan privasi data, serta pemeliharaan dan pengembangan sistem. Analisis dan perancangan sistem dari suatu sistem yang utuh kedalam bagian komponen dengan maksud mengidentifikasi dan mengevaluasi permasalahan yang terjadi dan kebutuhan untuk perbaikan. Hasil akhir dari sistem ini adalah solusi sesuatu dalam bentuk spesifikasi sistem berikutnya, dengan terlebih dahulu mempersiapkan data dan membagi menjadi data training dan data testing.

4.4. Perhitungan K-NN, *Naive Bayes* dan *Linear Regression*

Perhitungan algoritma pada pemberian kelayakan kredit bank menggunakan algoritma K-NN, *Naive Bayes* dan *Linear Regression* melibatkan langkah-langkah yang berbeda. Pada algoritma K-NN, perhitungan melibatkan penghitungan jarak antara data uji dengan data latih, pemilihan tetangga terdekat berdasarkan jarak, dan penentuan kelas mayoritas dari tetangga terdekat untuk mengklasifikasikan data. Sementara itu, algoritma *Naive Bayes* melibatkan perhitungan probabilitas kelas dan probabilitas fitur untuk mengklasifikasikan data. Dalam *Naive Bayes*, probabilitas kelas dihitung berdasarkan data latih, dan probabilitas fitur dihitung untuk setiap fitur dalam data uji. Kemudian, menggunakan aturan Bayes, probabilitas posterior dihitung untuk setiap kelas, dan kelas dengan probabilitas posterior tertinggi dipilih sebagai hasil klasifikasi. *Linear Regression* digunakan untuk memodelkan hubungan antara variabel dependen (variabel kelayakan kredit) dengan satu atau lebih variabel independen (fitur-fitur lainnya).

Hasil Perhitungan Algoritma

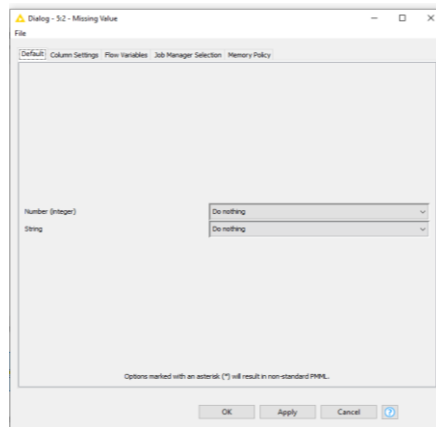
1) Hasil perhitungan KNIME menggunakan algoritma K-NN



Gambar 2. K-Nearest Neighbor (KNN)

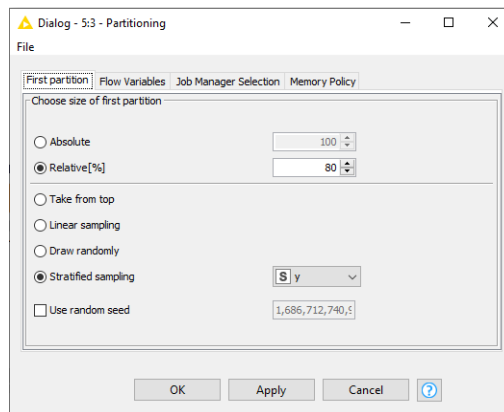
Pada Gambar 3.1, node yang digunakan yaitu *File Reader*, *Missing Value*, *Partitioning*, *K Nearest Neighbor*, dan *Scorer*. Setelah dataset dibaca oleh *File Reader* pada saat proses *preprocessing* data, kemudian gunakan *Missing Value*

sebelum masuk ke proses *Partitioning*. *K-Nearest Neighbor* (KNN) yaitu fokus pada *Partitioning* dengan memilih *configure* dan atur *Partitioning*.



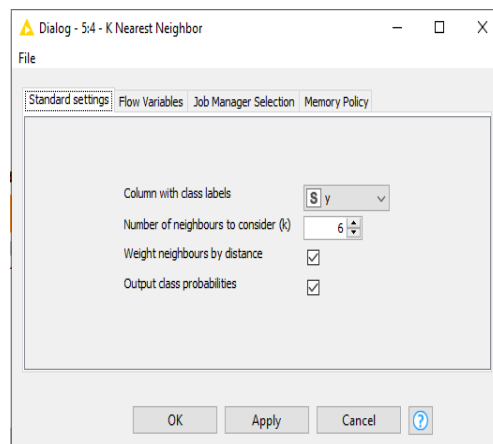
Gambar 3. Missing Value

Pada Gambar 3.2, fungsi dari *Missing Value* ini yaitu untuk mengisi data kosong dengan perintah mengisi data dengan data terbanyak atau yang sering muncul agar terhindar dari adanya kesalahan data atau error.



Gambar 4. Partitioning

Pada Gambar 3.3 yaitu *Partitioning* atur relative 80% kemudian pilih *stratified sampling* dengan variabel *y* sebagai prediksi yang digunakan.



Gambar 5. Node K Nearest Neighbor (KNN)

Pada Node *K-Nearest Neighbor* (KNN) column yang digunakan ini adalah *y* kemudian pada *wight neighbours by distance* dan *output class probabilitas* di aktifkan. Setelah itu klik Apply dan OK.

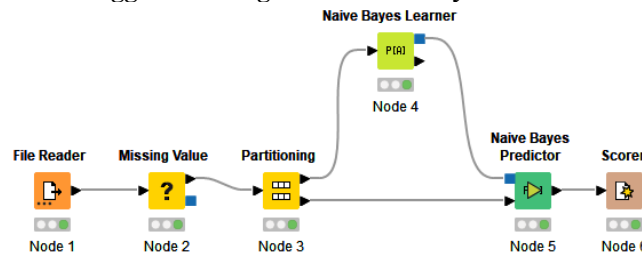
y \ Class [kNN]	no	yes
no	6977	333
yes	669	259

Correct classified: 7,236 Wrong classified: 1,002
 Accuracy: 87.837% Error: 12.163%
 Cohen's kappa (κ): 0.277%

Gambar 6. Confusion Matrix

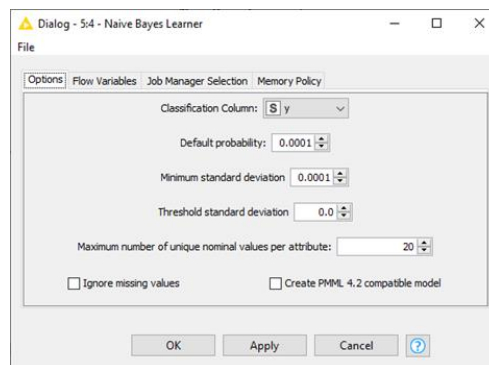
Dari gambar 3.5 hasil *confusion matrix* menampilkan nilai akurasi sebanyak 87,837% dan nilai error sebanyak 12,163% dari 8.238 data yang di training terdapat 7.236 data dinyatakan benar dan 1.002 data dinyatakan salah.

2) Hasil perhitungan KNIME menggunakan Algoritma *Naïve Bayes*



Gambar 7. Naive Bayes

Pada Gambar 3.6 node yang digunakan yaitu *File Reader*, *Missing Value*, *Partitioning*, *Naive Bayes Learning*, *Naive Bayes Predictor* dan *Scorer* yang kemudian disambungkan seperti pada gambar. Pada node *File Reader* dilakukan import data dengan format data berupa csv yang sudah tersedia, kemudian sebelum ke proses selanjutnya gunakan *missing value* untuk mengisi data kosong dengan perintah mengisi data dengan data terbanyak atau yang sering muncul. Selanjutnya pada *Partitioning* memilih *configure* dan atur *relative* sebanyak 80% kemudian pilih *stratified* sampling yaitu kolom *y* sebagai prediksi yang digunakan.



Gambar 8. Naive Bayes Learner

Selanjutnya pada Gambar 3.7 node *Naive Bayes Learner* untuk menghasilkan *Gaussian distribution* setiap *cluster*. Pada *Classification Column* pilih *y* dan untuk pengaturan lainnya diatur secara *default* dari KNIME itu sendiri.

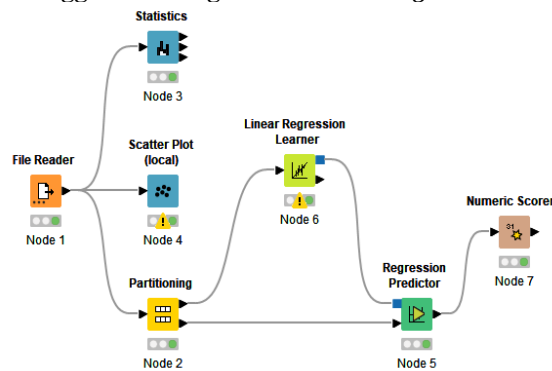
y \ Prediction (y)	no	yes
no	7261	73
yes	840	64

Correct classified: 7,325 Wrong classified: 913
 Accuracy: 88.917% Error: 11.083%
 Cohen's kappa (κ): 0.097%

Gambar 9. Confusion Matrix

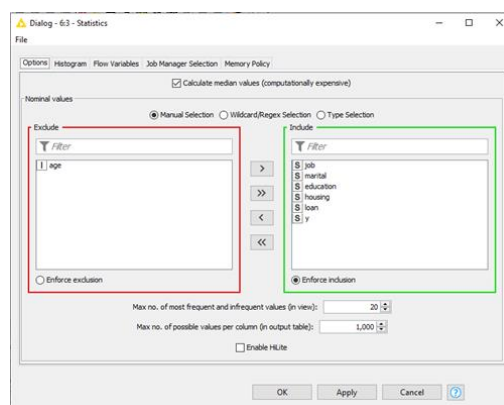
Dari hasil *confusion matrix* diatas menampilkan nilai akurasi sebanyak 88.917 % dan nilai *error* sebanyak 11.038 % dari 8.238 data yang di training terdapat 7.325 data dinyatakan benar dan 913 data dinyatakan salah.

3) Hasil perhitungan KNIME menggunakan Algoritma *Linear Regression*



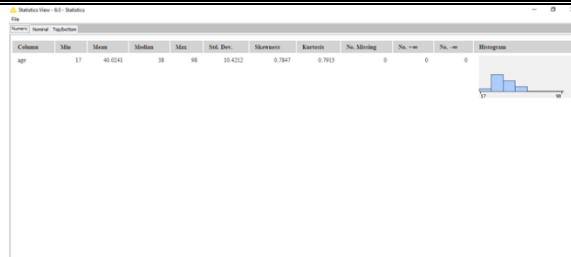
Gambar 10. Linear Regression

Pada gambar 3.9, node yang digunakan yaitu *File Reader*, *Statistic*, *Scatter Plot (local)*, *Partitioning*, *Linear Regression Learner*, *Regression Predictor* dan *Scorer* yang kemudian disambungkan seperti pada gambar. setelah dataset diimport di *file reader*, kemudian proses dari selanjutnya yaitu perhitungan *statistics* untuk mengetahui nilai-nilai *mean*, *median*, nilai minimal, nilai maksimal dan lainnya.



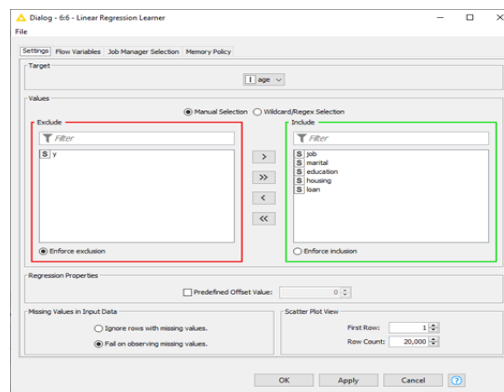
Gambar 11. Statistic

Pada Gambar 3.10 merupakan proses pengolahan dataset menjadi data yang dapat dibaca secara statistika, sebanyak 6 variabel digunakan.



Gambar 12. Hasil Perhitungan *Statistic*

Gambar 3.11 merupakan hasil dari perhitungan statistik dengan kolom *age* yang memiliki nilai *mean* 40,0241, *median* 38 dengan nilai minimal 17 dan nilai maksimal 98.



Gambar 13. *Linear Regression Learner*

Gambar 3.12, menyatakan proses pengolahan dataset menjadi data yang dapat dibaca secara statistika, variabel yang digunakan sebanyak 5 variabel.

Variable	Coeff.	Std. Err.	T-value	P-value
job-bless collar	-1.5787	NA&S	NA&S	NA&S
job-management	1.1749	NA&S	NA&S	NA&S
job-motivated	-2.7388	NA&S	NA&S	NA&S
job-management	2.5754	NA&S	NA&S	NA&S
job-oriented	19.2644	NA&S	NA&S	NA&S
job-well motivated	0.7603	NA&S	NA&S	NA&S
job-ambitious	-1.0144	NA&S	NA&S	NA&S

Gambar 14. Hasil *Linear Regression Learner*

Pada data gambar 3.13 menampilkan bahwa hasil dari kolom *coeff* menunjukkan setiap peningkatan satu unit dalam variabel. Kemudian *Std err* merupakan perkiraan standar deviasi dari koefisien. Pada kolom *T-value* menggambarkan sejauh mana koefisien tersebut berbeda secara signifikan. Semakin tinggi nilai *t-value*, maka semakin tinggi signifikan pengaruh variabel independen terhadap variabel dependen. Kolom $P > |t|$ adalah nilai *p* yang menunjukkan bahwa tingkat signifikansi statistik.

R ² :	0.369
Mean absolute error:	6.703
Mean squared error:	68.472
Root mean squared error:	8.275
Mean signed difference:	-0.038
Mean absolute percentage error:	0.174
Adjusted R ² :	0.369

Gambar 15. hasil permodelan

Pada Gambar 3.14 diatas menunjukkan hasil pemodelan, R^2 yang memiliki nilai 0,369 menunjukkan bahwa model yang digunakan tidak sesuai dengan data. Nilai *Mean absolute error* sebesar 6,703 menunjukkan adanya selisih antara nilai yang diamati dan nilai yang diprediksi. Semakin rendah nilai *Mean absolute error*, semakin baik kualitas prediksi model.

4. KESIMPULAN

Berdasarkan pembahasan, implementasi dan pengujian dari system untuk menentukan kelayakan kredit bank menggunakan algoritma K-NN, *Naive Bayes* dan *Linear Regression* diperoleh beberapa kesimpulan yaitu, Sistem data mining yang diteliti dapat membantu pihak bank untuk proses penentuan calon kelayakan nasabah kredit dalam pemberian kredit dengan cepat dan akurat, sebagai bahan pertimbangan dan alat bantu dalam proses penentuan klasifikasi kelayakan calon nasabah kredit.

Algoritma K-NN, *Naive Bayes* dan *Linear Regression* dapat digunakan sebagai pendekatan yang efektif dalam pemberian kelayakan kredit bank, dan pemilihan algoritma yang tepat tergantung pada kebutuhan spesifik dan karakteristik data yang ada. Dari berbagai hasil perhitungan yang sudah dilakukan menggunakan algoritma K-NN didapatkan nilai akurasi sebesar 87,837%, perhitungan menggunakan algoritma *Naive Bayes* didapatkan nilai akurasi sebesar 88,917%, sedangkan perhitungan menggunakan algoritma *Linear Regression* menghasilkan nilai *Mean absolute error* sebesar 6,703. Dapat disimpulkan bahwa dalam penentuan kelayakan kredit bank menggunakan metode algoritma *Naive Bayes* lebih akurat jika dibandingkan dengan algoritma K-NN maupun *Linear Regression*

REFERENCES

- [1] N. Nuraeni, "No Title," *Penentuan Kelayakan Kredit Dengan Algoritma Naive Bayes Classif. Stud. Kasus Bank Mayapada Mitra Usaha Cab. PGC.*, vol. III(1), pp. 9–15, 2017.
- [2] S. Harlina, S. Suryani, and M. Oton Kadang, "Penerapan Algoritma K-Nearest Neighbor Untuk Klasifikasi kelayakan Calon Nasabah Kredit Berbasis Web," *Semin. Nas. Tek. Elektro, Inform. dan Sist. Inf.*, vol. 1, no. 1, 2022, doi: 10.35842/sintaks.v1i1.18.
- [3] S. Shalev-Shwartz and S. Ben-David, "Nearest Neighbor," *Underst. Mach. Learn.*, vol. 7, no. 1, pp. 219–227, 2014, doi: 10.1017/cbo9781107298019.020.
- [4] T. T. Muryono, A. Taufik, and I. Irwansyah, "Perbandingan Algoritma K-Nearest Neighbor, Decision Tree, Dan Naive Bayes Untuk Menentukan Kelayakan Pemberian Kredit," *Infotech J. Technol. Inf.*, vol. 7, no. 1, pp. 35–40, 2021, doi: 10.37365/jti.v7i1.104.
- [5] M. H. Rifqo and A. Wijaya, "Implementasi Algoritma Naive Bayes Dalam Penentuan Pemberian Kredit," *Pseudocode*, vol. 4, no. 2, pp. 120–128, 2017, doi: 10.33369/pseudocode.4.2.120-128.
- [6] L. P. D. Tampubolon, "Efektifitas Faktor Penentu Putusan Penilaian Kelayakan Kredit Berbasis Data Mining," pp. 1057–1064.
- [7] T. T. Muryono and I. Irwansyah, "Implementasi Data Mining Untuk Menentukan Kelayakan Pemberian Kredit Dengan Menggunakan Algoritma K-Nearest Neighbors (K-Nn)," *Infotech J. Technol. Inf.*, vol. 6, no. 1, pp. 43–48, 2020, doi: 10.37365/jti.v6i1.78.
- [8] F. Gultom and T. Simanjuntak, "Prediksi Tingkat Kelancaran Pembayaran Kredit Bank Dengan Menggunakan Algoritma Naive Bayes Dan K-Nearest Neighbor," *METHOMIKA J. Manaj. Inform. dan Komputerisasi Akunt.*, vol. 4, no. 2, pp. 98–102, 2021, doi: 10.46880/jmika.vol4no2.pp98-102.
- [9] K. Neighbor *et al.*, "Klasifikasi Penentuan Pengajuan Kartu Kredit Menggunakan," vol. 22, no. 1, pp. 73–82, 2020.
- [10] S. Dewi, "Komparasi Metode Algoritma Data Mining pada Prediksi Uji Kelayakan Credit Approval pada Calon Nasabah Kredit Perbankan," *J. Khatulistiwa Inform.*, vol. 7, no. 1, pp. 59–65, 2019, doi: 10.31294/jki.v7i1.5744.
- [11] A. Prawira, D. Arisandi, and T. Sutrisno, "Penerapan Algoritma Naive Bayes dan Multiple Linear Regression Untuk Prediksi Status dan Plafon Kredit (Studi Kasus: Bank ABC)," *J. Educ.*, vol. 5, no. 1, pp. 1075–1087, 2022, doi: 10.31004/joe.v5i1.720.

- [12] S. Wahyuningsih, D. R. Utari, U. B. Luhur, D. Tree, and K. Validation, “Perbandingan Metode K-Nearest Neighbor , Naïve Bayes dan Decision Tree untuk Prediksi Kelayakan Pemberian Kredit,” pp. 8–9, 2018.
- [13] D. Kurniawan, “No Title,” *Regresi linier*, 2008.
- [14] S. Khatib, J., & Dalam, “No Title,” *Indones. J. Comput. Sci.*, vol. 11(1), pp. 635–644, 2022.
- [15] R. Retnosari, “Analisa Kelayakan Kredit Usaha Mikro Berjalan pada Perbankan dengan Metode Naive Bayes,” *PROSISKO J. Pengemb. Ris. dan Obs. Sist. Komput.*, vol. 8, no. 1, pp. 53–59, 2021.