ISSN 2527-3671 (*Online*)
ISSN 2087-8893 (*Print*)

Tersedia *online* di www.journal.unipdu.ac.id
Unipdu

Halaman jurnal di www.journal.unipdu.ac.id/index.php/teknologi

Perbandingan Metode Algoritma C4.5, Naïve Bayes, dan Logistic Regression untuk Penentuan Kelayakan Penerima Kredit

Nur Hafid Ibrahim [a], Laelatul Khikmah [b]
a,b Statistika, Institut Teknologi Statistika dan Bisnis Muhammadiyah Semarang, Semarang, Indonesia

email: [a,]nurhafidibrahim@gmail.com*

*Korespondensi

Dikirim 29 Mei 2024; **Direvisi** 20 Juli 2024; **Diterima** 05 Agustus 2024; **Diterbitkan** 04 September 2024

Abstrak

Kredit adalah layanan penting yang disediakan oleh sektor perbankan selain tabungan, transfer uang, dan investasi. Masalah kredit macet secara global telah mendorong penerapan pembelajaran mesin dan analitik canggih untuk manajemen risiko kredit. Tujuan dari penelitian ini adalah untuk meningkatkan efisiensi dan akurasi proses klasifikasi kelayakan pemohon kredit dengan membandingkan algoritma yang dipilih dan juga ingin mengetahui apakah algoritma yang dipilih dapat meningkatkan akurasi dan efektivitas proses penentuan kelayakan pemohon pinjaman. Algoritma klasifikasi data mining yang akan digunakan dalam perbandingan pada penelitian ini adalah algoritma C4.5, Naive Bayes, dan *Logistic Regression*. Algoritma C4.5, *Naive Bayes* dan *Logistic Regression* digunakan untuk melihat nilai akurasi, presisi dan recall yang dihasilkan antara ketiga algoritma tersebut. Nilai-nilai ini diamati dengan menggunakan model validasi silang. Hasil penelitian menunjukkan bahwa algoritma Regresi Logistik memiliki nilai akurasi tertinggi dengan nilai akurasi sebesar 81,1%. Algoritma Regresi Logistik cukup akurat dalam memprediksi data karena nilai AUC termasuk dalam predikat *Fair Classification* dengan nilai 0.76. Oleh karena itu, Algoritma Regresi Logistik dapat digunakan sebagai acuan dalam pengambilan keputusan untuk menentukan kelayakan penerima kredit.

Kata Kunci: Kredit, Algoritma, *C4.5, Naïve Bayes, Logistic Regression*

Comparison of C4.5, Naïve Bayes, and Logistic Regression Algorithm Methods for Determining Credit Recipient Eligibility

Abstract

Credit is a vital service provided by the banking sector in addition to savings, money transfers, and investments. The global problem of bad credit has encouraged the application of machine learning and advanced analytics for credit risk management. The purpose of this research is to improve the efficiency and accuracy of the credit applicant eligibility classification process by comparing the selected algorithms and also want to know whether the selected algorithm can improve the accuracy and effectiveness of the credit applicant eligibility determination process. The data mining classification algorithms that will be used in the comparison in this study are C4.5, Naive Bayes and Logistic Regression algorithms. The C4.5, Naive Bayes and Logistic Regression algorithms are used to observe the accuracy, precision and recall values generated between the three algorithms. These values observed by using a cross-validation model. The results showed that the Logistic Regression algorithm has the highest accuracy value with an accuracy value of 81.1%. The Logistic Regression algorithm is accurate enough to predict data because the AUC value is included in the Fair Classification predicate with a value of 0.76. Therefore, the Logistic Regression Algorithm can be used as a reference in decision making to determine the eligibility of credit recipients.

Keywords: *Credit; Algorithm; C4.5; Naïve Bayes; Logistic Regression*

Untuk mengutip artikel ini dengan *APA Style*:
Ibrahim, N. H., & Khikmah, L. (2024). Perbandingan Metode Algoritma C4.5, Naïve Bayes, dan Logistic Regression untuk Penentuan Kelayakan Penerima Kredit. TEKNOLOGI: Jurnal Ilmiah Sistem Informasi, 14 (2), 85 - 93: https://doi.org/10.26594/teknologi.v14i2.4650

1. Pendahuluan

Industri perbankan adalah komponen penting dalam sistem keuangan karena memungkinkan investasi, transfer uang yang aman, dan tabungan. Bank telah berkembang dari waktu ke waktu menjadi institusi yang menawarkan pinjaman kepada orang dan perusahaan selain melindungi simpanan (Hasan, 2014). Interaksi yang rumit antara analisis kredit, manajemen risiko, dan transaksi keuangan yang disebabkan oleh perkembangan ini menekankan perlunya teknik yang dapat diandalkan untuk mengevaluasi kelayakan kredit dan mengurangi kredit macet.

ISSN 2502-3357 (*Online*)
ISSN 2503-0477 (*Print*)

Kredit adalah suatu pengaturan keuangan dimana dana dipinjamkan kepada individu atau entitas yang membutuhkan, dengan ketentuan pembayaran kembali yang diatur melalui sistem cicilan (Yulianto & Andri, 2016). Meskipun proses ini bukan berarti tanpa hambatan. Evaluasi kredit yang kurang cermat atau perilaku nasabah yang kurang baik dapat menyebabkan sebagian besar kredit berubah menjadi aset bermasalah, sehingga membahayakan stabilitas lembaga keuangan dan perekonomian secara luas.

Kredit bermasalah merupakan masalah internasional yang menjadi perhatian dunia, melampaui batas-batas regional (Yusuf & Sestri, 2020). Dalam konteks Indonesia, data statistik yang bersumber dari Otoritas Jasa Keuangan (OJK) mengindikasikan bahwa sektor perbankan dihadapkan pada jumlah kredit bermasalah yang cukup signifikan. Fenomena ini tidak hanya terjadi di satu wilayah tertentu, namun sering kali meningkat pada saat resesi ekonomi atau ketidakstabilan keuangan, seperti yang terjadi pada pandemi global baru-baru ini (Santika, 2023).

Menanggapi masalah ini, sektor perbankan dan keuangan telah mulai menggunakan metode analitik yang canggih dan algoritma pembelajaran mesin untuk mengendalikan risiko kredit dengan lebih baik (Yudistira, 2021). C4.5, Naïve Bayes, dan Regresi Logistik adalah beberapa algoritma yang telah menunjukkan potensi dalam meningkatkan ketepatan dan keampuhan evaluasi kelayakan kredit. Algoritma-algoritma ini dapat membantu dalam pilihan persetujuan kredit dengan menilai data masa lalu, tren perilaku konsumen, dan indikator risiko yang berbeda.

Klasifikasi adalah metode analisis data yang membantu kita menentukan label atau kategori dari contoh-contoh yang ingin kita identifikasi. Ini adalah jenis pembelajaran yang menggunakan bantuan untuk menemukan pola antara informasi yang diberikan dan apa yang ingin kita ketahui. Tujuannya adalah untuk membuat hasil yang kita dapatkan dari data menjadi lebih dapat diandalkan (Hendrian, 2018).

Algoritma C4.5 termasuk ke dalam kelompok algoritma decision tree dengan dua input, yaitu sample training dan sample test. Algoritma C4.5 akan membuat sebuah pohon keputusan. Pohon keputusan adalah suatu jenis pohon yang digunakan sebagai metode penalaran deduktif untuk mencari solusi dari suatu permasalahan (Setianingrum et al., 2021).

Algoritma Naïve Bayes merupakan algoritma klasifikasi statistik yang dapat digunakan untuk memprediksi probabilitas keanggotaan kelas. Naïve Bayes didasarkan pada Teorema Bayes yang memiliki kemampuan klasifikasi yang mirip dengan beberapa algoritma klasifikasi lainnya (Kusrini & Luthfi, 2019).

Algoritma Regresi Logistik adalah metode untuk membuat model prediksi yang mirip dengan regresi linier, yang juga dikenal sebagai regresi Ordinary Least Squares (OLS). Bedanya, dalam regresi logistik, peneliti menggunakan skala dikotomi untuk memprediksi variabel dependen. Skala dikotomi ini merupakan skala data nominal dengan dua kategori, seperti Ya dan Tidak, Baik dan Buruk, atau Tinggi dan Rendah (Yualinda et al., 2020).

Tujuan dari penelitian ini adalah untuk mengevaluasi seberapa baik algoritma C4.5, Naïve Bayes, dan Regresi Logistik dalam mengklasifikasikan kelayakan pemohon pinjaman. Dengan mengevaluasi kekuatan dan kelemahan masing-masing algoritma, kami bertujuan untuk mengidentifikasi algoritma terbaik untuk klasifikasi kelayakan kredit. Selain itu, kami juga ingin mengetahui apakah algoritma yang dipilih dapat meningkatkan akurasi dan efektivitas proses penentuan kelayakan pemohon kredit.

Penelitian ini tidak hanya berfungsi sebagai eksplorasi teoritis tetapi juga memiliki implikasi praktis bagi lembaga keuangan yang ingin mengoptimalkan proses penilaian kredit mereka dan mengurangi risiko yang terkait dengan kredit bermasalah. Penelitian ini diharapkan dapat memberikan wawasan yang berharga mengenai keampuhan algoritma machine learning dalam meningkatkan praktik manajemen risiko kredit.

Beberapa penelitian yang menggunakan algoritma machine learning antara lain penelitian (Yusuf & Sestri, 2020) yang menggunakan algoritma C4.5 untuk mengklasifikasikan kredit nasabah PT Bank Perkreditan Rakyat (Studi Kasus: PT BPR Lubuk Raya Mandiri). Selain itu, penelitian dari (Novriandy, 2023) yang menganalisis kelayakan bantuan dengan menggunakan algoritma Naive Bayes dan C4.5 menunjukkan bahwa algoritma C4.5 memiliki performa yang lebih baik dibandingkan dengan algoritma Naive Bayes. Metode C4.5 memiliki akurasi sebesar 90.00%, sedangkan algoritma Naïve Bayes memiliki akurasi sebesar 70.00%. Penelitian selanjutnya dari (Siswa & Prihandoko, 2018) mengevaluasi performa terbaik dari algoritma C4.5, Naïve Bayes, K-Nearest Neighbor, Logistic Regression, dan Support Vector Machines untuk mendiagnosa penyakit kanker payudara.

ISSN 2502-3357 (*Online*)
N. H. Ibrahim dkk./TEKNOLOGI 14 (2) 2024 85-94 ISSN 2503-0477 (*Print*)

2. *State of the Art*

Berisi penjelasan penelitian sebelumnya yang membahas topik atau metode yang sama dan perbandingan dengan penelitian saat ini, serta penjelasan manfaat dari penelitian sebelumnya sehingga penulis dapat menerapkannya pada beberapa aspek yang telah dilakukan sebelumnya.

2.1. Metode Decision Tree Dalam Klasifikasi Kredit Pada Nasabah PT Bank Perkreditan Rakyat (Studi Kasus : PT BPR Lubuk Raya Mandiri)

Penelitian yang dilakukan oleh Diana Yusuf dan Ellya Sestri pada tahun 2020 (Yusuf & Sestri, 2020) berfokus pada kasus PT BPR Lubuk Raya Mandiri dan menjelajahi penerapan algoritma data mining, khususnya C4.5, dalam analisis kredit. Penelitian ini menemukan bahwa algoritma C4.5 dapat efektif digunakan untuk mengklasifikasikan kredit pada nasabah PT Bank Perkreditan Rakyat, termasuk di PT BPR Lubuk Raya Mandiri serta pohon keputusan yang dihasilkan dari algoritma C4.5 dapat membantu dalam pengambilan keputusan terkait kelayakan pemberian kredit. Penelitian tersebit memiliki kesamaan dalam hal klasifikasi kredit pada suatu bank, menggunakan algoritma c4.5 dan menggunakan alat RapidMiner.

2.2. Perbandingan Metode Algoritma K-NN & Metode Algoritma C4.5 Pada Analisa Kredit Macet (Studi Kasus PT Tungmung Textile Bintan)

Penelitian yang dilakukan oleh Ajeng Setianingrum, Ayu Hindayanti, Dita Meilani Cahya, dan Dini Silvi Purnia pada tahun 2021 (Setianingrum et al., 2021) berfokus pada membandingkan dua metode algoritma, yaitu algoritma K-NN dan algoritma C4.5, dalam analisis kredit macet di PT Tungmung Textile Bintan. Penelitian ini menunjukkan bahwa algoritma C4.5 lebih akurat dalam menentukan kredit macet daripada algoritma K-NN. Akurasi algoritma C4.5 mencapai 61,64%, sedangkan akurasi algoritma K-NN hanya sebesar 45,21%. Dengan demikian, dapat disimpulkan bahwa algoritma C4.5 lebih efektif dalam memprediksi kredit macet berdasarkan data yang dianalisis dalam studi kasus PT Tungmung Textile. Penelitian tersebut memiliki kesamaan dalam hal klasifikasi kredit menggunakan algoritma c4.5.

2.3. Determining the Eligibility of Providing Motorized Vehicle Loans by Using the Logistic Regression, Naive Bayes and Decission Tree (C4.5)

Penelitian yang dilakukan oleh Harish Rianto, dkk pada tahun 2020 (Rianto et al., 2020) berisi tentang evaluasi kelayakan pemberian kredit kendaraan bermotor dari PT Buana Kredit Sejahtera menggunakan *algoritma Logistic Regression, Naive Bayes* dan *Decission Tree* (C4.5). Penelitian ini menunjukkan bahwa algoritma *Logistic Regression* diidentifikasi sebagai algoritma yang terbaik dengan nilai AUC 0,972 dan Akurasi 93,14%, mengungguli Naïve Bayes dan *Decision Tree* (C4.5). Dengan demikian, dapat disimpulkan bahwa model Regresi Logistik dapat digunakan oleh perusahaan untuk meningkatkan proses analisis pinjaman mereka. Penelitian tersebut memiliki korelasi terhadap penelitian ini dalam hal klasifikasi kredit menggunakan tiga algoritma yang sama dan permasalahan yang serupa yaitu tentang kredit.

2.4. Implementasi Algoritma Naive Bayes dan Algoritma C4. 5 dalam Klasifikasi Kelayakan Bantuan UMKM

Penelitian yang ditulis oleh Aldy Novriandy pada tahun 2023 (Novriandy, 2023) membahas tentang Implementasi Algoritma Naive Bayes dan Algoritma C4.5 dalam Klasifikasi Kelayakan Bantuan UMKM. Temuan penelitian menunjukkan bahwa algoritma Naive Bayes dan C4.5 dapat secara efektif mengklasifikasikan kelayakan bantuan untuk UMKM. Namun, algoritma C4.5 menunjukkan akurasi yang lebih tinggi yaitu 90% dibandingkan dengan akurasi pada algoritma Naive Bayes yaitu 70%. Dengan demikian, dapat disimpulkan bahwa algoritma C4.5 lebih efektif dalam memprediksi kelayakan bantuan UMKM berdasarkan data yang dianalisis. Korelasi terhadap penelitian ini adalah pengunaan algoritma Naïve Bayes dan C4.5.

2.5. Perbandingan Kinerja Algoritma C4.5, Naïve Bayes, K-Nearest Neighbor, Logistic Regression, dan Support Vector Machines Untuk Menimbulkan Penyakit Kanker Payudara

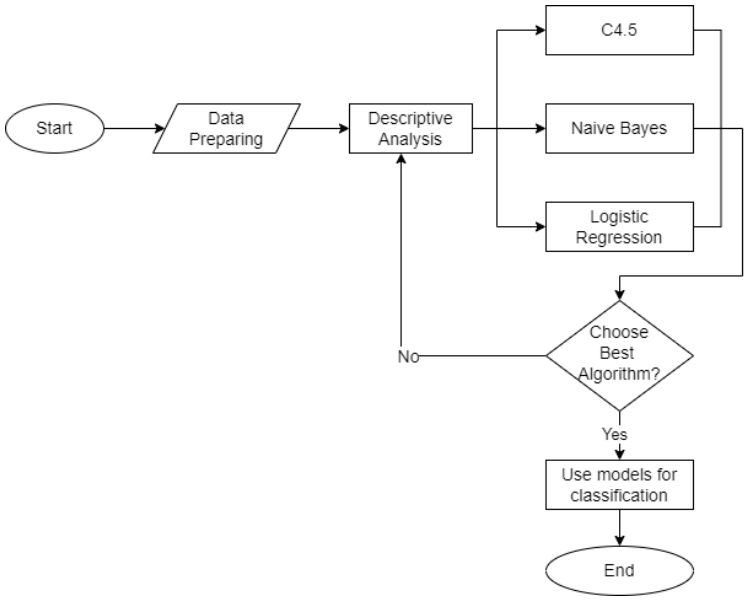
Penelitian yang ditulis oleh Taghfirul Azhima, Yoga Siswa, dan Prihandoko pada tahun 2018 (Siswa & Prihandoko, 2018) berisi tentang evaluasi kinerja berbagai algoritma klasifikasi penambangan data (C4.5, Naïve Bayes, K-Nearest Neighbor, Logistic Regression, dan Support Vector Machines) untuk mendeteksi kanker payudara. Hasil penelitian menunjukan bahwa T-Test pada lagoritma Logistic Regresson dan Support Vector Machines memiliki akurasi tertinggi 0,968, menunjukkan kinerja unggul mereka dalam deteksi kanker payudara. Dengan demilian, algoritma Logistic Regresson dan Support Vector Machines dapat digunakan dalam mendeteksi kanker payudara. Kesamaan dari penelitian ini adalah penggunaan algortma C4.5, Naïve Bayes dan Logistic Regression dalam klasifikasi data.

3. Metode Penelitian

Algoritma klasifikasi data mining yang akan digunakan dalam perbandingan pada penelitian ini adalah algoritma C4.5, Naive Bayes dan Logistic Regression. Algoritma C4.5, Naive Bayes dan Logistic Regression digunakan untuk melihat nilai akurasi, presisi dan recall yang dihasilkan antara ketiga algoritma tersebut. Nilai-nilai ini diamati dengan menggunakan model Cross Validation. Berdasarkan hasil yang diperoleh, nilai

ISSN 2502-3357 (*Online*)
N. H. Ibrahim dkk./TEKNOLOGI 14 (2) 2024 85-94 ISSN 2503-0477 (*Print*)

akurasi, presisi, dan recall tertinggi akan digunakan sebagai dasar dan acuan pengambilan keputusan. Selanjutnya algoritma terbaik dengan nilai akurasi tertinggi akan digunakan untuk melakukan klasifikasi, yang akan menunjukkan atribut-atribut yang akan berpengaruh terhadap keputusan dalam pemberian kredit apakah diterima atau ditolak. Gambar 1 menampilkan alur penelitian untuk perbandingan tiga algoritma klasifikasi.



Gambar 1. Alur Penelitian

3.1. Menyiapkan data

Data dalam penelitian ini diperoleh dengan mengambil dari website kaagle, yaitu data "Loan Prediction Problem" (Unknown, 2019). Data "Loan Prediction Problem" dibersihkan dari data yang hilang atau kesalahan yang mungkin ada. Proses ini melibatkan identifikasi dan penanganan nilai yang hilang, duplikat, atau tidak valid. Data yang telah dibersihkan kemudian diproses untuk menyiapkannya ke dalam format yang sesuai untuk analisis.

3.2. Analisis Deskriptif

Analisis deskriptif memberikan gambaran umum mengenai karakteristik data, termasuk distribusi dan statistik deskriptif. Hal ini membantu dalam memahami data secara keseluruhan sebelum analisis lebih lanjut menggunakan algoritme yang telah disebutkan sebelumnya.

3.3. Algoritma C4.5

Algoritma C4.5 termasuk dalam kelompok algoritma pohon keputusan dengan dua input, yaitu data training dan data testing. Algoritma C4.5 akan membuat sebuah pohon keputusan (Setianingrum et al., 2021). Pada algoritma C4.5, langkah pertama yang dilakukan adalah menghitung nilai entropy dan nilai gain dari setiap atribut. Nilai entropy ini akan digunakan untuk mencari nilai gain, dan atribut dengan gain tertinggi akan menjadi akar pertama dalam pohon keputusan. Rumus untuk menghitung nilai entropy dan gain adalah sebagai berikut

$$Entropy(S) = \sum_{i=1}^{n} -p_i . \log_2 p_i$$

$$Gain(S, A) = Entropy(S) \sum_{i=1}^{n} \frac{|S_i|}{|S|} \times Entropy(S_i)$$

3.4. Algoritma Naïve Bayes

Algoritma Naive Bayes merupakan algoritma klasifikasi berdasarkan konsep probabilitas yang menerapkan teorema Bayes pada aplikasinya dengan independensi yang tinggi (Rianto et al., 2020). Teorema Bayes dapat dituliskan pada persamaan berikut

$$P(Y|X) = \frac{P(X|Y).P(Y)}{P(X)}$$

3.5. Algoritma Logistic Regression

ISSN 2502-3357 (*Online*)
N. H. Ibrahim dkk./TEKNOLOGI 14 (2) 2024 85-94 ISSN 2503-0477 (*Print*)

Logistic Regression adalah metode statistik yang digunakan untuk membangun hubungan antara variabel hasil biner dan satu atau lebih variabel prediktor. Variabel independen dapat diklasifikasikan sebagai kontinu, diskrit, biner, atau kombinasi dari tipe-tipe tersebut (van Smeden et al., 2019). Logistic Regression menggunakan variabel yang telah ditentukan atau variabel yang dikategorikan menjadi 2 variabel. Seperti pada prediksi sukses atau gagal, hidup atau mati, sakit atau tidak sakit, dan lain sebagainya (Rianto et al., 2020). Model logistic regression dapat dituliskan dengan persamaan di bawah ini

$$\pi(x) = \frac{e^{(\beta_0+\beta_1 x_1+\beta_2 x_2+ \dots +\beta_n x_n)}}{1 + e^{(\beta_0+\beta_1 x_1+\beta_2 x_2+ \dots +\beta_n x_n)}}$$

3.6. Cross Validation

Cross Validation adalah metode tambahan dari metode data mining untuk mengembangkan model split validation yang bertujuan untuk mendapatkan hasil akurasi yang maksimal dimana validasi tersebut mengukur training error dengan melakukan pengujian dengan data testing. Dari hasil percobaan dengan pengujian sebanyak n kali dengan k-fold cross validation, nilai evaluasi performa model akan dicatat dengan menggunakan confusion matrix (Tanjung & Fujiati, 2021).

3.7. Confusion Matrix

Confusion Matrix merupakan alat yang efektif untuk menilai seberapa baik pengklasifikasi yang kita gunakan dapat mengidentifikasi pola dari berbagai data. Confusion Matrix adalah sekumpulan baris dan kolom yang menampilkan hasil klasifikasi dalam bentuk tabel. Baris-baris menampilkan data aktual sedangkan kolom-kolom menampilkan data yang diprediksi oleh model (Peco Chacón et al., 2023). Confusion matrix akan mengahsilkan nilai accuracy, precision, dan recall. Tabel Confusion Matrix yang memberikan gambaran lengkap mengenai hasil klasifikasi model dapat dilihat pada Tabel 1.

Tabel 1. Confusion Matrix

Class	Predict Yes	Predict No	Total
Actual Yes	True Positive (TP)	False Negative (FN)	Positive (P)
Actual No	False Positive (FP)	True Negative (TN)	Negative (N)
Total	P'	N'	P+N

3.8. ROC (Receiver Operating Characteristic) Curve

Kurva ROC adalah plot grafis yang menggambarkan kemampuan diagnostik sistem klasifikasi biner dengan memplotkan true positif rate (TPR) terhadap false positif rate (FPR) pada berbagai ambang batas. ROC juga dapat dianggap sebagai fungsi yang menunjukan kekuatan dari kesalahan tipe I dari aturan keputusan(Purwanto & Nugroho, 2023). AUC (Area Under Curve) pada kurva ROC digunakan untuk mengukur akurasi klasifikasi, yang diklasifikasikan sebagai berikut(Moolayil, 2018).

0.90 - 1.00 = Excellent classification

0.80 - 0.90 = Good classification

0.70 - 0.80 = Fair classification

0.60 - 0.70 = Poor classification

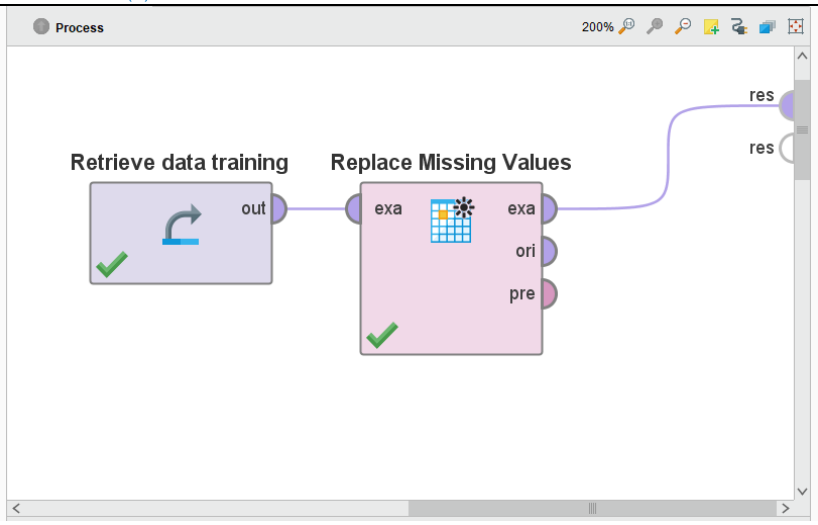
0.50 - 0.60 = Failure

4. Hasil dan Pembahasan

4.1. Menyiapkan data

Dalam penelitian ini, tahapan persiapan data terdiri dari seleksi data, pembersihan data, dan transformasi data menggunakan RApidMiner 9.0. Atribut yang dihasilkan pasca pembersihan adalah Gender, Married, Dependents, Education, Self Employed, Applicant Income, Co applicant Income, Loan Amount, Loan Amount Term, Credit History, Property Area, dan Loan Status. Data ini kemudian ditransformasi menjadi variabel input dan output, dengan 'Loan Status' sebagai variabel output yang memiliki nilai 'Passed' dan 'Not Passed'. Tahapan proses data dapat dilihat pada gambar 2.

ISSN 2502-3357 (*Online*)
N. H. Ibrahim dkk./TEKNOLOGI 14 (2) 2024 85-94 ISSN 2503-0477 (*Print*)



Gambar 2. Langkah Pembersihan data

4.2. Analisis Deskriptif

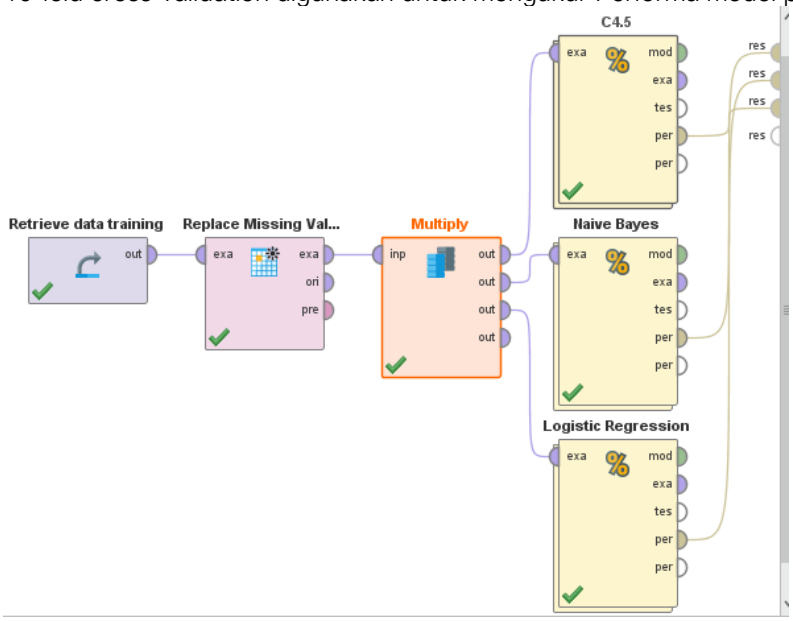
Analisis deskriptif dalam penelitian ini dilakukan untuk memahami secara singkat mengenai kondisi penerima kredit. Data menunjukan bahwa 68,7% pelamar kredit diterima dan 31,3% pelamar kredit ditolak. Terdapat kebutuhan untuk memahami lebih dalam tentang faktor-faktor yang mempengaruhi hasil ini. Penelitian lebih lanjut dan analisis menggunakan tiga algoritma ini dapat membantu dalam meningkatkan proses seleksi dan memberikan peluang yang lebih baik bagi pelamar di masa depan. Untuk hasil analisis yang lebih lengkap dapat dilihat pada tabel 2.

Tabel 2. Analisis Deskriptif pada data label

Variabel	Keterangan	Frekuensi	%
Loan Status	Y(Accepted)	422	68,7%
	N(Rejected)	192	31,3%

4.3. Cross Validation

Analisis selanjutnya yaitu klasifikasi data dengan membandingkan ketiga algoritma klasifikasi yang digunakan, yaitu C4.5, Naïve Bayes, dan Rogistic regression. Gambar 3 menampilkan model validasi silang pada rapidminer untuk ketiga algoritma tersebut, dimana model validasi silang ditunjukkan pada gambar dengan keterangan sesuai dengan nama algoritma. Penelitian ini melakukan 10 kali Cross Validation. Teknik 10-fold cross-validation digunakan untuk mengukur Performa model pada penelitian ini.

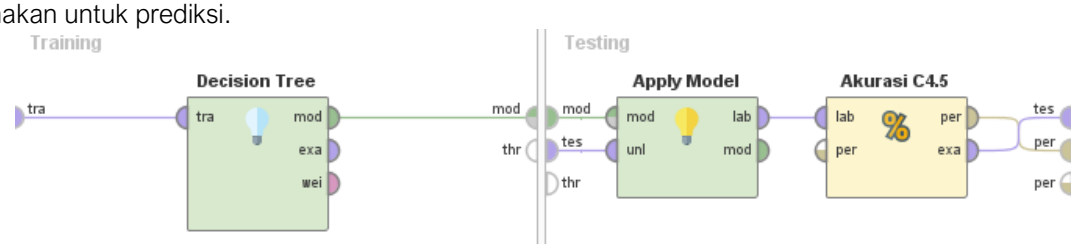


Gambar 3. Tampilan klasifikasi data dengan Cross Validation

ISSN 2502-3357 (*Online*)
N. H. Ibrahim dkk./TEKNOLOGI 14 (2) 2024 85-94 ISSN 2503-0477 (*Print*)

4.4. Algoritma C4. 5

Susunan model Algoritma C4.5 pada Rapidminer dengan Cross Validation ditampilkan pada Gambar 4, di mana proses ini melibatkan aplikasi dari 10-fold Cross Validation untuk mengukur kinerja model sebelum digunakan untuk prediksi.



Gambar 4. Struktur model algoritma C4.5

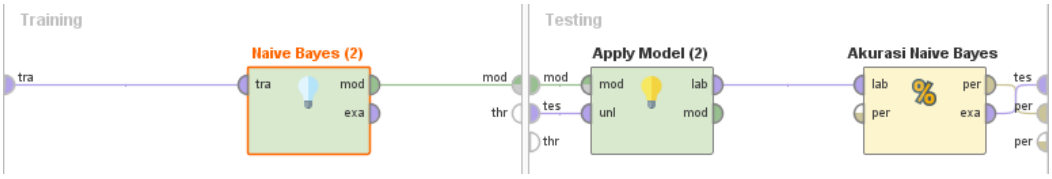
Berdasarkan penelitian yang telah dilakukan, pengujian Algoritma C4.5 menunjukkan performa yang signifikan dalam analisis data. Algoritma ini memperlihatkan evaluasi performa dalam klasifikasi dan prediksi data, seperti yang dapat dilihat dalam Tabel 3. Hasil tersebut tersebut mencakup berbagai parameter seperti tingkat keberhasilan, dan tingkat kesalahan, yang semuanya berkontribusi dalam mengukur efektivitas algoritma.

Tabel 3. Evaluasi Algoritma C4.5

	True Y	True N	Class Precision
Pred. Y	405	113	78.19%
Pred. N	17	79	82.29%
Class Recall	95.97%	41.15%	

4.5. *Algoritma Naïve Bayes*

Susunan model Algoritma Naïve Bayes pada Rapidminer dengan Cross Validation ditampilkan *pada Gambar 5*.



Gambar 5. Struktur model algoritma Naive Bayes

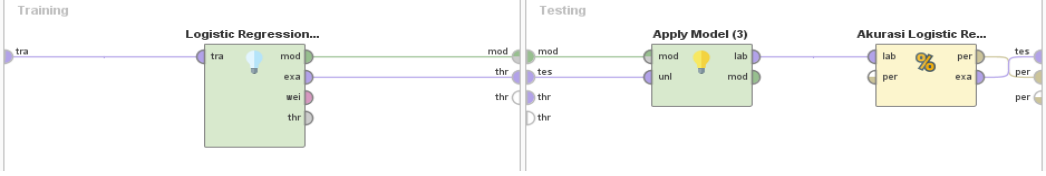
Berdasarkan penelitian yang telah dilakukan, pengujian Algoritma Naïve Bayes menunjukkan performa yang signifikan dalam analisis data. Algoritma ini memperlihatkan evaluasi performa dalam klasifikasi dan prediksi data, seperti yang dapat dilihat dalam Tabel 4. Hasil tersebut tersebut mencakup berbagai parameter seperti tingkat keberhasilan, dan tingkat kesalahan, yang semuanya berkontribusi dalam mengukur efektivitas algoritma.

Tabel 4. Evaluasi Algoritma Naïve Bayes

	True Y	True N	Class Precision
Pred. Y	396	104	79.20%
Pred. N	26	88	77.19%
Class Recall	93.84%	45.83%	

4.6. *Algoritma Logistic Regression*

Susunan model Algoritma Naïve Bayes pada Rapidminer dengan Cross Validation ditampilkan pada Gambar 6.



Gambar 6. Struktur model algoritma Logistic Regression

ISSN 2502-3357 (*Online*)
N. H. Ibrahim dkk./TEKNOLOGI 14 (2) 2024 85-94 ISSN 2503-0477 (*Print*)

Berdasarkan penelitian yang telah dilakukan, pengujian Algoritma Naïve Bayes menunjukkan performa yang signifikan dalam analisis data. Algoritma ini memperlihatkan evaluasi performa dalam klasifikasi dan prediksi data, seperti yang dapat dilihat dalam Tabel 5. Hasil tersebut tersebut mencakup berbagai parameter seperti tingkat keberhasilan, dan tingkat kesalahan, yang semuanya berkontribusi dalam mengukur efektivitas algoritma.

Tabel 5. Evaluasi Algoritma Logistic Regression

	True Y	True N	Class Precision
Pred. Y	414	108	79.31%
Pred. N	8	84	91.30%
Class Recall	98.10%	43.75%	

4.7. *Confision Matrix dan AUC*

Confision Matrix juga digunakan untuk mengevaluasi kinerja algoritma dalam klasifikasi, yang memungkinkan untuk melihat secara lebih terperinci bagaimana model mengklasifikasikan data dari setiap atribut. Setiap algoritma dievaluasi berdasarkan nilai akurasi, presisi, recall, dan AUC untuk menentukan algoritma mana yang memiliki nilai tertinggi dan dapat digunakan untuk klasifikasi. Nilai akurasi, presisi, recall, dan AUC dapat dilihat dalam Tabel 6.

Tabel 6. Perbandingan Performa

Algoritma	Accuracy	Precision	Recall	AUC
C4.5	78,51%	83,78%	39,05%	0,701
Naïve Bayes	79,33%	80,04%	45,87%	0,732
Logistic Regression	81,1%	90,82%	43,68%	0,760

Berdasarkan tabel 6, diperoleh hasil bahwa algoritma C4.5 memiliki akurasi sebesar 78,51%, Algoritma Naive Bayes memiliki akurasi sebesar 79,33% dan Logistic Regression memiliki akurasi yang paling tinggi yaitu 81,1%. Kemudian Untuk nilai AUC, algoritma C4.5 memiliki Nilai AUC sebesar 0,701, Algoritma Naive Bayes memiliki nilai AUC sebesar 0,732 dan Logistic Regression memiliki nilai AUC yang paling tinggi yaitu 0,76. Dengan demikian, dapat diketahui bahwa algoritma Regresi Logistik merupakan model algoritma yang memiliki nilai akurasi dan AUC tertinggi di antara algoritma lainnya, sehingga cukup akurat untuk memprediksi data. Hal ini didasarkan pada kriteria bahwa nilai AUC-nya termasuk dalam kategori Klasifikasi yang Fair dengan nilai antara 0,70 hingga 0,80.

5. Kesimpulan

Berdasarkan analisis dan pembahasan hasil perbandingan tiga algoritma klasifikasi C4.5, Naive Bayes dan Logistic Regression pada data “Loan Prediction Problem”, maka dapat disimpulkan bahwa berdasarkan nilai akurasi, algoritma regresi logistik merupakan algoritma yang paling baik untuk melakukan klasifikasi dan algoritma regresi logistik cukup akurat dalam memprediksi sebuah model karena memiliki nilai AUC sebesar 0.76 yang termasuk dalam predikat klasifikasi “Fair Classification”. Oleh karena itu, Algoritma Regresi Logistik dapat digunakan sebagai acuan dalam pengambilan keputusan untuk menentukan kelayakan penerima kredit. Namun, nilai AUC pada algoritma masih dalam kategori Fair Classification. Hasil evaluasi pada model tersebut perlu ditingkatkan dari segi akurasi, presisi, recall, dan AUC. Oleh karena itu, saran terkait dengan penelitian ini adalah menerapkan algoritma klasifikasi data mining lainnya untuk membuat keputusan yang lebih baik terkait atribut yang digunakan. Selain itu, disarankan untuk menggunakan dataset lain yang relevan dengan kredit dalam pembuatan model klasifikasi. Pengembangan lainnya yang berkaitan dengan penelitian ini termasuk menentukan atribut pendukung keputusan yang memiliki pengaruh lebih besar terhadap akurasi.

6. Kontribusi Penulis

Nur Hafid Ibrahim : *Pencarian Data, Konseptualisasi, Analisis Data, Methodologi, Software, Visualisasi, dan Menulis draft*. MLaelatul Khikmah : *Pengawasan, Validasi, review & editing*.

7. Declaration of Competing Interest

Penulis menyatakan tidak ada konflik kepentingan.

ISSN 2502-3357 (*Online*)
N. H. Ibrahim dkk./TEKNOLOGI 14 (2) 2024 85-94 ISSN 2503-0477 (*Print*)

8. Referensi

Hasan, N. I. (2014). Pengantar Perbankan. In *Referensi (Gaung Persada Press Group)*. Referensi (Gaung Persada Press Group).

Hendrian, S. (2018). Algoritma Klasifikasi Data Mining Untuk Memprediksi Siswa Dalam Memperoleh Bantuan Dana Pendidikan. *Faktor Exacta*, *11*(3), 266–274. https://doi.org/10.30998/faktorexacta.v11i3.2777

Kusrini, & Luthfi, E. taufiq. (2019). Algoritma Data Mining. In *ANDI*. ANDI.

Moolayil, J. (2018). Learn Keras for Deep Neural Networks: A Fast-Track Approach to Modern Deep Learning with Python. In *Apress*. https://doi.org/10.1007/978-1-4842-4240-7

Novriandy, A. (2023). Implementasi Algoritma Naive Bayes dan Algoritma C4. 5 dalam Klasifikasi Kelayakan Bantuan UMKM. *KLIK: Kajian Ilmiah Informatika Dan Komputer*, *4*(1), 208–217. https://doi.org/10.30865/klik.v4i1.1099

Peco Chacón, A. M., Segovia Ramírez, I., & García Márquez, F. P. (2023). K-nearest neighbour and K-fold cross-validation used in wind turbines for false alarm detection. *Sustainable Futures*, *6*(September), 0–5. https://doi.org/10.1016/j.sftr.2023.100132

Purwanto, A., & Nugroho, H. W. (2023). Analisa Perbandingan Kinerja Algoritma C4.5 Dan Algoritma K-Nearest Neighbors Untuk Klasifikasi Penerima Beasiswa. *Jurnal Teknoinfo*, *17*(1), 236. https://doi.org/10.33365/jti.v17i1.2370

Rianto, H., Amrin, Rudianto, Pahlevi, O., Kusumawardhani, P., & Hadi, S. S. (2020). Determining the Eligibility of Providing Motorized Vehicle Loans by Using the Logistic Regression, Naive Bayes and Decission Tree (C4.5). *Journal of Physics: Conference Series*, *1641*(1). https://doi.org/10.1088/1742-6596/1641/1/012061

Santika, E. F. (2023). Kucuran Kredit Perbankan Naik 7,76% per Juni 2023, Pembayaran Macetnya Turun. *Katadata*. https://databoks.katadata.co.id/datapublish/2023/09/25/kucuran-kredit-perbankan-naik-776-per-juni-2023-pembayaran-macetnya-turun

Setianingrum, A., Hindayanti, A., Cahya, D. M., & Purnia, D. S. (2021). Perbandingan Metode Algoritma K-NN & Algoritma C4 . 5 Pada Analisa Kredit Macet (Studi Kasus PT Tungmung Textile Bintan). *Evolusi: Jurnal Sains Dan Manajemen*, *9*(2), 78–92.

Siswa, T. A. Y., & Prihandoko. (2018). Perbandingan Kinerja Algoritma C4. 5, Naïve Bayes, K-Nearest Neighbor, Logistic Regression, Dan Support Vector Machines Untuk Mendeteksi Penyakit Kanker Payudara. *Jurnal Teknologi Informasi Dan Komunikasi*, *7*(2), 1–10.

Tanjung, D. Y. H., & Fujiati. (2021). Analisis Perbandingan Algoritma ID3 dan C4.5 Terhadap Data Pengisian Uang ATM. *CSRID Journal*, *13*(3A), 231–242.

Unknown. (2019). *Loan Prediction Problem Dataset*. Kaagle.

van Smeden, M., Moons, K. G. M., de Groot, J. A. H., Collins, G. S., Altman, D. G., Eijkemans, M. J. C., & Reitsma, J. B. (2019). Sample size for binary logistic prediction models: Beyond events per variable criteria. *Statistical Methods in Medical Research*, *28*(8), 2455–2474. https://doi.org/10.1177/0962280218784726

Yualinda, S., Hernawati, E., & Wijaya, D. R. (2020). Application To Predict Poverty Based on E-Commerce Data Using Logistic. *E-Proceeding of Applied Science*, *6*(2), 3109–3122.

Yudistira, N. (2021). Peran Big Data dan Deep Learning untuk Menyelesaikan Permasalahan Secara Komprehensif. *EXPERT: Jurnal Manajemen Sistem Informasi Dan Teknologi*, *11*(2), 78. https://doi.org/10.36448/expert.v11i2.2063

Yulianto, A., & Andri, S. (2016). Analisis Penerapan 5 C Dalam Pemberian Kredit Konsumtif Pada Pt. Adira Dinamika Multifinance Cabang Nangka Pekanbaru. *Jurnal Online Mahasiswa (JOM) Bidang Ilmu Sosial Dan Ilmu Politik*, *3*(1), 1–12.

Yusuf, D., & Sestri, E. (2020). Metode Decision Tree Dalam Klasifikasi Kredit Pada Nasabah PT Bank Perkreditan Rakyat (Studi Kasus : PT BPR Lubuk Raya Mandiri). *Jurnal Sistem Informasi (JUSIN)*, *1*(1), 21–28. https://doi.org/10.32546/jusin.v1i1.855