

Implementasi Algoritma Klasifikasi *Random Forest* Untuk Penilaian Kelayakan Kredit

Omar Pahlevi¹, Amrin², Yopi Handrianto³

^{1,2,3}Universitas Bina Sarana Informatika

e-mail: ¹omar.opi@bsi.ac.id, ²amrin.ain@bsi.ac.id, ³yopi.yph@bsi.ac.id

Diterima	Direvisi	Disetujui
09-05-2023	17-05-2023	31-05-2023

Abstrak - Kesalahan dalam memberikan penilaian kelayakan kredit berakibat pada resiko kredit macet. Masalah yang sering terjadi bukan pada penerapan sistem oleh pihak keuangan namun lebih banyak pada sumber daya manusia saat melakukan penilaian kelayakan kredit konsumen. Penelitian yang dilakukan merupakan implementasi salah satu algoritma klasifikasi pada *data mining*, yakni algoritma *random forest*. Himpunan data yang digunakan dalam kegiatan penelitian ini sejumlah 481 data *record* kredit kendaraan bermotor, dimana memiliki dua buah target antara lain label mengenai data kredit *bad* dan *good*. Beberapa atribut yaitu : status pernikahan, data tanggungan, usia, status tempat tinggal, kepemilikan rumah, pekerjaan, status perusahaan, pendapatan, uang muka, pendidikan terakhir, lama tinggal, dan kondisi rumah digunakan pada penelitian yang dilakukan sebagai *feature input*. Pada penelitian ini metode klasifikasi *data mining random forest* diterapkan untuk mengklasifikasikan status kelayakan kredit. Dengan menerapkan Algoritma *random forest*, didapatkan hasil tingkat akurasi sebesar 78,60% dengan nilai *Area Under the Curve* (AUC) sebesar 0,907. Mengacu pada tingkat akurasi dan nilai *Area Under the Curve* (AUC), maka model algoritma *Random Forest* termasuk kategori klasifikasi baik sekali.

Kata Kunci: klasifikasi, *random forest*, *confusion matrix*, *ROC Curve*

Abstract - Errors in assessing creditworthiness result in the risk of bad credit. The problem that often occurs is not in the application of the system by financial parties but more in human resources when assessing consumer creditworthiness. The research carried out is the implementation of one of the classification algorithms in data mining, namely the random forest algorithm. The data set used in this research activity is 481 motorized vehicle credit data records, which have two targets, including labels regarding bad and good credit data. Several attributes, namely: marital status, data on dependents, age, residence status, home ownership, occupation, employment status, company status, income, down payments, last education, length of stay, and housing conditions were used in the research conducted as input features. In this study the random forest data mining classification method is applied to classify creditworthiness status. By applying the random forest algorithm, the results obtained an accuracy rate of 78.60% with an Area Under the Curve (AUC) value of 0.907. Referring to the accuracy level and the Area Under the Curve (AUC) value, the Random Forest algorithm model is included in the very good classification category.

Keywords: classification, *random forest*, *confusion matrix*, *ROC Curve*

PENDAHULUAN

Pada tren era pertumbuhan ekonomi yang berangsur pulih pasca kejadian luar biasa pandemi Covid-19, masyarakat Indonesia pada umumnya mulai aktif melakukan kegiatan perekonomian sehari-hari. Hal ini pun berdampak pada tingkat pemanfaatan kredit dalam melakukan transaksi jual beli sehari-hari. Pengeluaran konsumen adalah salah satu pendorong utama ekonomi makro dan risiko yang bersifat sistemik. Karena itu, analisis pemberian kredit kepada konsumen menjadi relevan karena individu pada akhirnya mungkin mencari pinjaman untuk memenuhi kebutuhan konsumsinya (Aniceto et

al., 2020). Salah satu syarat utama pengajuan kredit adalah kesepakatan atau persetujuan antara pihak kreditur dan debitur tentang bagaimana mereka akan melunasi hutang mereka dalam jangka waktu tertentu. Prinsip-prinsip seperti *character*, *capacity*, *capital*, *condition*, dan *collateral* merupakan tahapan dalam penilaian kelayakan pengajuan kredit (Eriyanto & Alamsyah, 2020).

Data yang digunakan dan tugas yang diberikan berupa bobot nilai aplikasi dan bobot nilai perilaku merupakan dua hal dalam tahapan penilaian dalam pengajuan kredit. Pada tahap bobot nilai aplikasi akan estimasi kemungkinan pemohon pengajuan kredit gagal bayar untuk beberapa interval waktu tertentu. Data mengenai demografis dan keuangan pemohon

pengajuan kredit tercantum pada pelatihan model dengan luaran status baik atau buruk yang direkam untuk interval waktu lainnya. Pada tahapan ini terdapat kegiatan berupa penilaian aplikasi dari pemohon pengajuan kredit. Kemudian pada proses selanjutnya yaitu bobot penilaian perilaku, dimana hal ini dapat dilakukan ketika pinjaman telah diberikan kepada pemohon. Pada tahapan ini kemungkinan pemohon gagal bayar pada interval waktu tertentu akan diperiksa. Data yang digunakan untuk Pada kegiatan penilaian ini dilakukan berdasarkan pada kinerja pengembalian pinjaman oleh pemohon dengan luaran berupa status *bad* atau *good* (Ibnu & Rachmatullah, 2023).

Penilaian risiko kelayakan kredit merupakan masalah yang sangat diperlukan di lembaga keuangan. Risiko Kredit adalah didefinisikan sebagai probabilitas bahwa peminjam akan gagal membayar jumlah pinjaman. Keputusan dari pemberian atau penolakan pinjaman sangat kritis dan didasarkan pada informasi pribadi pemohon, sejarah pengajuan kredit, status hidup, loyalitas, dan lain-lain. Model klasifikasi yang digunakan pada penelitian ini telah digunakan secara luas dalam penilaian risiko kredit karena membantu menemukan hubungan antara atribut dalam menuntut pinjaman (seperti sejarah pengajuan kredit, jumlah rekening pelanggan, pendapatan tahunan, kepemilikan rumah) dan potensi tidak melunasi kredit (Arora & Kaur, 2020).

Kebutuhan pemodelan risiko kredit tumbuh dari bank yang membutuhkan perkiraan kuantitatif untuk menghitung modal berisiko yang diperlukan untuk mendukung kegiatan pinjaman mereka. Penerapan evaluasi kelayakan kredit perusahaan adalah salah satu topik utama yang dihadapi pemberi pinjaman, hal ini dilakukan sebelum memutuskan apakah perusahaan harus disetujui untuk pembiayaan. Keterbatasan ketersediaan data, fleksibilitas keuangan dan ketidakjelasan adalah sedikit dari banyak faktor yang mempersulit evaluasi kelayakan kredit untuk usaha kecil menengah, sehingga meningkatkan risiko kredit dan biaya pembiayaan (Malakauskas & Lakstutiene, 2021).

Perusahaan pembiayaan menangani berbagai jenis pinjaman, termasuk pinjaman rumah, kendaraan, pendidikan, dan pinjaman pribadi. Perusahaan ini hadir di berbagai bidang seperti di perkotaan dan pedesaan. Mekanisme kegiatannya nasabah terlebih dahulu mengajukan permohonan pinjaman dan setelah itu perusahaan pembiayaan memvalidasi kelayakan nasabah untuk persetujuan pinjaman. Data detail seperti status perkawinan, jenis kelamin, pendidikan, dan jumlah tanggungan, Pendapatan, Pinjaman Jumlah, riwayat kredit, dan lainnya diberikan dalam formulir untuk diisi oleh pemohon. Oleh karena itu, sangat dibutuhkan pemodelan *data mining* dengan mengambil detail tersebut sebagai masukan untuk memverifikasi apakah suatu pemohon layak untuk mengajukan pinjaman atau tidak. Variabel target di sini adalah Pelamar "Status Pinjaman" dan variabel lainnya adalah prediktor (Eadara & Kumar, 2023).

Pada era industri modern saat ini *data mining*

memiliki peran strategis. *Data mining* dapat digunakan untuk menilai risiko kredit karena memiliki kemampuan untuk memprediksi hasil pengamatan masa depan. Hal ini memang sangat penting mengingat risiko kredit merupakan 60% dari total risiko dari suatu perusahaan. Selain itu, Untuk menggunakan teknologi dan otomasi dalam meminimalkan biaya operasional pemberi pinjaman kontemporer, diperlukan model penilaian kredit yang akurat. Masalah yang lebih kompleks dalam pemberian kredit perlu dibentuknya sistem *credit scoring* otomatis yang dapat menghindari hilangnya kesempatan atau kerugian kredit untuk mengurangi potensi kerugian bagi masing-masing lembaga pemberi pinjaman (Aji & Dhini, 2019).

Data mining adalah kumpulan prosedur yang digunakan untuk mendapatkan nilai tambahan dari informasi dalam suatu basis data yang sebelumnya tidak diketahui secara manual. Proses ini menggunakan proses ekstraksi dan menggali pola penting dari data yang sudah ada untuk mendapatkan model atau fungsi yang menggambarkan dan membedakan konsep atau label data. (Ismanto & Novalia, 2021). Merupakan salah satu *Induction-Based Learning* dimana melakukan kegiatan pengamatan pada beberapa contoh spesifik dari berbagai konsep yang akan dipelajari. *Knowledge Discovery in Databases* (KDD) adalah salah satu penerapan metode *scientific* pada *data mining*. Satu langkah KDD dijelaskan oleh *data mining*. Kegiatan pencarian pola atau tren yang diharapkan dalam basis data berukuran besar adalah manfaat dari *data mining* untuk membantu *decision-making* di masa depan. Pada tahapan kegiatan ini menghasilkan *pattern* yang dapat diidentifikasi oleh sebuah *device* tertentu, dimana menghasilkan solusi dalam pengambilan keputusan yang baik (Amir & Abijono, 2019).

Buku Larose "*Discovering Knowledge in Data: An Introduction to Data Mining*", membahas mengenai *pembagian* kelompok yang terdapat pada *data mining*, antara lain (Pahlevi, 2018):

1. *Description*

Pada tahap ini membahas mengenai teknik dalam menjelaskan *pattern* dan isi spesifik pada sebuah data. Kemungkinan penjabaran untuk suatu *pattern* atau kecenderungan pembahasan mengenai data merupakan kegiatan utama pada tahapan deskripsi.

2. *Estimation*

Kecuali penilaian variabel target bersifat numerik dan bukan kategoris, penilaian mirip dengan klasifikasi. Model diciptakan menggunakan daftar lengkap atau deretan data sehingga memberikan nilai variabel target untuk nilai prediksi. Nilai variabel prediksi kemudian digunakan untuk membuat perkiraan nilai variabel target, yang pada gilirannya didasarkan pada nilai variabel prediksi.

3. *Prediction*

Prediksi dan klasifikasi terdapat kemiripan, kecuali prediksi yakni memprediksi nilai hasil selanjutnya. Mengingat keadaan yang tepat, beberapa aturan dan cara yang dipakai untuk klasifikasi dan

evaluasi juga dapat dipakai untuk peramalan.

4. *Classification*

Pada klasifikasi, target variabel kategori digunakan. Contohnya, pendapatan dapat diklasifikasikan menjadi kategori rendah, sedang, dan tinggi.

5. *Clustering*

Clustering atau pengelompokan adalah proses pengelompokan atau pengamatan kumpulan data, melakukan pengamatan dan membentuk kategori.

Ada 3 (tiga) langkah utama dari proses data mining :

1. *Data preparation*

Selanjutnya dilakukan pemilihan data, kemudian lakukan pembersihan data, lalu masuk kedalam *preprocessed* sesuai instruksi dan *knowledge*.

2. *Data Mining algorithm*

Pada fase ini, algoritma data mining digunakan untuk memeriksa data terintegrasi dan membantu mengidentifikasi informasi berharga.

3. *Data Analysis Phase*

Hasil *data mining* dilakukan *review* untuk memastikan *knowledge domain* dapat dilihat dalam bentuk *rule* yang telah di ekstrak dari jaringan.

Pada *data mining*, pengukuran kinerja klasifikasi memiliki tujuan untuk memperlihatkan tingkat akurasi sistem dalam mengklasifikasikan data. Tujuan dari tahapan evaluasi adalah melakukan evaluasi model klasifikasi data mining sehingga kinerja sistem dapat diketahui. Salah satu metode yang *familiar* dalam melakukan evaluasi kinerja klasifikasi data mining adalah *k-fold Cross Validation*, yakni teknik untuk melakukan validasi keakuratan model berdasarkan kumpulan data yang diberikan. Dalam metode *k-fold Cross Validation*, kumpulan data terbagi kedalam 2 bagian, yakni *data training* dan *data testing* (Anam & Santoso, 2018).

Kemudian teknik *confusion matrix* yang dipakai sebagai perhitungan untuk memastikan konsep data mining akurat. Informasi yang diperoleh melalui *confusion matrix* dibutuhkan dalam menetapkan kinerja model klasifikasi, Dimana informasi ini selanjutnya diganti menjadi nilai untuk membandingkan kinerja berbagai model, hal ini dapat dikerjakan dengan menggunakan *performace metric* (Sugiyarti & Maseleno, 2018). Terdapat beberapa istilah yang digunakan pada *confusion matrix* diantaranya adalah TP (*True Positive*) merupakan hasil nilai yang diperoleh dari hasil prediksi yang sesuai dengan kelas sesungguhnya dan TN (*True Negative*) merupakan nilai sungguh negatif yang diperoleh dari hasil prediksi oleh model klasifikasi sesuai dengan kelas sesungguhnya. Selanjutnya istilah FP (*False Positive*) yang merupakan kelas aktual yang memiliki label negatif tetapi setelah dilakukan model klasifikasi diprediksi positif dan FN (*False Negative*) merupakan kelas aktual yang memiliki label positif tetapi setelah dilakukan model klasifikasi diprediksi negatif (Simarmata et al., 2022).

Selain menggunakan 2 (dua) metode yang disebutkan diatas, penelitian ini menggunakan Kurva *Receiver Operating Characteristics* (ROC).

Merupakan salah satu mekanisme dalam menggambarkan atau visualisasi, mengelompokkan atau mengorganisasikan serta menetapkan pengklasifikasi sesuai kinerja setiap algoritma. Kurva ROC diaplikasikan dalam mengukur nilai *Area Under Curve* (AUC) (Bahri, 2020).

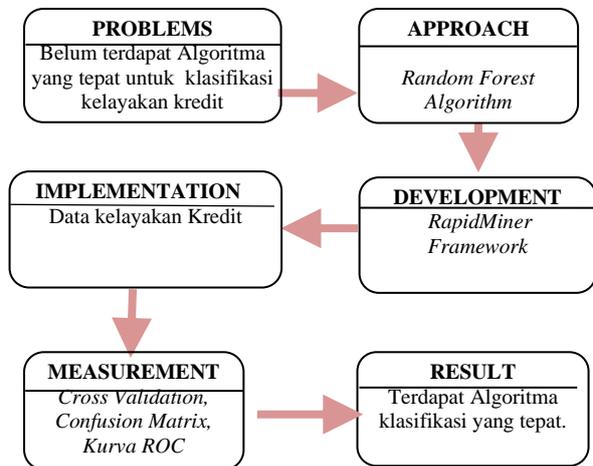
Pada penelitian ini, peneliti akan menggunakan algoritma *random forest* dengan tujuan untuk melakukan klasifikasi kelayakan kredit. Alasan peneliti menggunakan algoritma *random forest* dikarenakan algoritma ini diadopsi untuk membangun model untuk memprediksi tingkat kegagalan melunasi pinjaman di lembaga yang bergerak di bidang pembiayaan. Pada hasil pengujian terlihat bahwa algoritma *random forest* bekerja lebih optimal daripada algoritma *logistic regression*, *decision tree* dan *support vector machine*. dalam memperkirakan gagal bayar pinjaman dan memiliki kemampuan generalisasi yang kuat (Zhu et al., 2019). Terdapat lima langkah utama yang diperlukan untuk menerapkan model *random forest*. Langkah pertama adalah menentukan berapa banyak *decision tree* yang akan digunakan untuk membentuk sejumlah-k, langkah berikutnya adalah membuat sampel *random bootstrap* dan kemudian membuat *decision tree* untuk setiap sampel. Selanjutnya adalah melakukan validasi jumlah *decision tree* dengan jumlah K yang sudah ditetapkan pada langkah sebelumnya. Selanjutnya, hasil prediksi dari setiap *tree* yang dibentuk digabungkan. (Dominicus et al., 2020).

Penelitian ini berkaitan dengan penelitian sebelumnya (Zailani & Hanun, 2020). Pada hasil pengujian akhir diperoleh nilai akurasi sebesar 87,88% dan terbukti tepat dalam hal melakukan analisis kredit baik untuk debitur yang bermasalah maupun yang tidak bermasalah. Penelitian lain yang terkait dilakukan oleh (Prasojo & Haryatmi, 2021). Pada kegiatan penelitian ini menggunakan klasifikasi algoritma *Random Forest*, penelitian dilakukan untuk melihat implementasi metode klasifikasi dengan algoritma *Random Forest* pada *german data credit*. Hasil dari tahapan pengujian akhir diperoleh nilai performa dengan menggunakan AUC mendapatkan nilai sebesar 0,83 (83%) dan termasuk ke dalam *Very Good Model*. Penelitian lain yang berkaitan dilakukan oleh (Widjiyati, 2021). Penelitian ini menghasilkan bahwa *data mining* dapat digunakan sebagai *problem solving* untuk klasifikasi pemberian kredit bagi nasabah. Kemudian hasil pengujiannya menghasilkan bahwa algoritma *Random Forest* dapat memperlihatkan kinerja dari sisi akurasi *true positive rate* dan *true negative rate* yang lebih akurat apabila dibandingkan *single classifier* lainnya.

METODE PENELITIAN

Pada Gambar 1. menunjukkan *Research Framework* penelitian ini memiliki beberapa langkah tahapan. Permasalahan penelitian ini bahwa belum diketahuinya algoritma yang tepat untuk klasifikasi kelayakan kredit, sehingga dibuatlah suatu *approach*

(model) yaitu algoritma *random forest* untuk mengatasi permasalahan yang dihadapi. Selanjutnya adalah proses uji terhadap kinerja dari metode tersebut dengan metode *Cross Validation*, *Confusion Matrix* dan kurva ROC. Setelah itu, untuk proses *application development* sesuai dengan model yang dibuat, gunakan aplikasi *RapidMiner*.



Sumber: Penelitian (2023)
Gambar 1. *Research Framework*

HASIL DAN PEMBAHASAN

A. Data Analysis

Total *dataset* dalam penelitian ini meliputi 481 data record berupa informasi kredit kendaraan bermotor dari PT. Buana kredit Sejahtera bagi yang memiliki masalah atau tidak. Sedangkan variabel input penelitian adalah: 1) Status Pernikahan, 2) Data Tanggungan, 3) Usia, 4) Status Tempat Tinggal, 5) Kepemilikan Rumah, 6) Pekerjaan, 7) Status Pekerjaan, 8) Status Perusahaan, 9) Pendapatan, 10) Uang Muka, 11) Pendidikan Terakhir, 12) Lama Tinggal, dan 13) Kondisi Rumah. Gambar 2 dibawah ini menunjukkan sampel *dataset* untuk digunakan dalam pengujian algoritma,

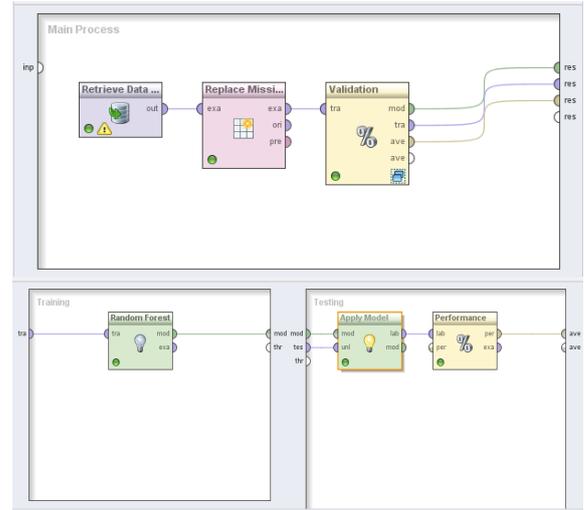
status pernikahan	jumlah tanggungan	pendidikan terakhir	usia	kepemilikan rumah	lama tinggal	kondisi rumah
menikah	2-3	SLTP	21-55	ortu	>5	permanen
belum menikah	tidak ada	SLTA	21-55	ortu	1-3	permanen
menikah	2-3	SLTA	21-55	KPR	3-5	permanen
belum menikah	tidak ada	SLTA	<21-50	ortu	1-3	permanen
menikah	>3	SD	21-55	tidak ada	>5	non permanen
menikah	2-3	SLTP	21-55	tidak sendiri	>5	permanen
menikah	2-3	SLTA	21-55	tidak sendiri	>5	permanen

Sumber: Penelitian (2023)

Gambar 2. Contoh *Dataset*

B. Model Testing

Dilakukan melalui uji eksperimen terhadap model yang diajukan. Selanjutnya adalah proses evaluasi dan validasi model sehingga dihasilkan nilai *accuracy* dan AUC. Proses uji dilakukan dengan *RapidMiner* dengan operator *10-fold cross-validation* agar menghasilkan nilai *accuracy* dan AUC pada algoritma yang dilakukan pengujian. Gambar 3 dibawah ini merupakan model yang dibangun.



Sumber: Penelitian (2023)
Gambar 3. *Built Models*

Selanjutnya adalah proses evaluasi menggunakan *Confusion Matrix* dan ROC Curve atau *Area Under Curve (AUC)*.

1. Confusion Matrix

Tabel 1 merupakan *Confusion Matrix* bagi algoritma *Random Forest*. Terlihat 92 data diklasifikasikan “bad” diperkirakan sama dengan data riil, kemudian 87 data diperkirakan “good” akan tetapi dinyatakan “bad”. Selanjutnya 286 data diklasifikasikan “good” diperkirakan sesuai, dan 16 data diperkirakan “bad” ternyata “good”.

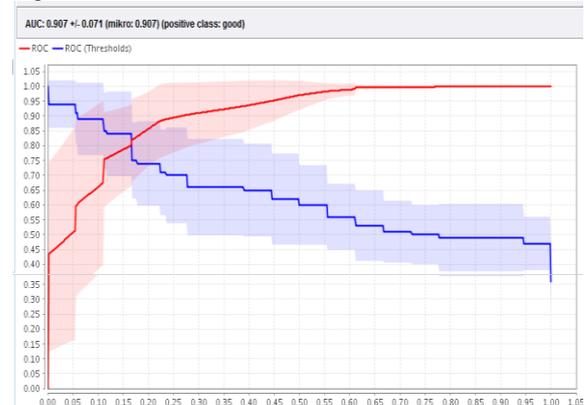
Tabel 1. *Confusion Matrix Random Forest Model*

Ketepatan	78.60%	+/- 8.46%	(mikro:78.59%)
	Benar Buruk	Benar Baik	Presisi Kelas
Pred.Buruk	92	16	85.19%
Pred.Baik	87	286	76.88%
class recall	51.40%	94.70%	

Sumber: Penelitian (2023)

2. Kurva ROC

Gambar 4 dibawah ini menunjukkan kurva ROC algoritma Random Forest.



Sumber: Penelitian (2023)

Gambar 4. Kurva ROC Algoritma *Random Forest*

Kurva ROC pada Gambar 4 menyajikan *confusion matrix*. Garis horizontal merupakan *false positives* dan garis vertikal merupakan *true positives*.

3. Analisis Hasil

Nilai *accuracy*, *precision*, *recall* dan AUC untuk algoritma *Random Forest* ditunjukkan oleh tabel 2 di bawah ini.

Tabel 2. Evaluasi dan Validasi Model

Algoritma	<i>Accuracy</i>	<i>Precision</i>	<i>Recall</i>	AUC
<i>Random Forest</i>	78.60%	78.20%	94.72%	0.907

Sumber: Penelitian (2023)

Menurut (Gorunescu, 2011) dalam (Amrin et al, 2021) mendeskripsikan bahwa berdasarkan klasifikasi data mining, nilai AUC dapat terbagi menjadi berbagai kelompok yaitu,

- a. 0.90-1.00 = baik sekali
- b. 0.80-0.90 = baik
- c. 0.70-0.80 = cukup
- d. 0.60-0.70 = buruk
- e. 0.50-0.60 = salah

Dari pengelompokkan diatas, maka dapat ditarik kesimpulan bahwa model algoritma *Random Forest* termasuk kategori klasifikasi baik sekali.

KESIMPULAN

Dari hasil penelitian dan pengujian, performa model algoritma *Random Forest* untuk klasifikasi kelayakan kredit memberikan tingkat akurasi kebenaran sebesar 78,60% dan memiliki nilai *Area Under the Curve* (AUC) sebesar 0,907. Bersumber pada tingkat akurasi dan nilai *Area Under the Curve* (AUC), maka model algoritma *Random Forest* termasuk kategori klasifikasi baik sekali.

REFERENSI

Aji, N. A., & Dhini, A. (2019). Credit scoring through data mining approach: A case study of mortgage loan in Indonesia. *2019 16th International Conference on Service Systems and Service Management, ICSSSM 2019*, 1–5. <https://doi.org/10.1109/ICSSSM.2019.8887731>

Amir, S., & Abijono, H. (2019). Penerapan Data Mining untuk Mendukung Pemasaran Produk Pinjaman Bank. *CAHAYATECH*, 7(2), 161. <https://doi.org/10.47047/ct.v7i2.102>

Amrin, Pahlevi, O., & Satriadi, I. (2021). Optimasi Algoritma C4.5 dan Naïve Bayes Berbasis Particle Swarm Optimization Untuk Diagnosa Penyakit Peradangan Hati. *Insantek*, 2(1), 10–14. <http://jurnal.bsi.ac.id/index.php/insantek/article/view/399>

Anam, C., & Santoso, H. B. (2018). Perbandingan Kinerja Algoritma C4.5 dan Naive Bayes untuk Klasifikasi Penerima Beasiswa. *Energy - Jurnal Ilmiah Ilmu-Ilmu Teknik*, 8(1), 13–19. <https://ejournal.upm.ac.id/index.php/energy/article/view/111>

Aniceto, M. C., Barboza, F., & Kimura, H. (2020). Machine learning predictivity applied to consumer creditworthiness. *Future Business Journal*, 6(1), 1–14. <https://doi.org/10.1186/s43093-020-00041-w>

Arora, N., & Kaur, P. D. (2020). A Bolasso based consistent feature selection enabled random forest classification algorithm: An application to credit risk assessment. *Applied Soft Computing Journal*, 86, 105936. <https://doi.org/10.1016/j.asoc.2019.105936>

Bahri, S. (2020). Implementasi Data Mining Untuk Memprediksi Keterlambatan Jam Masuk Kerja Menggunakan Algoritma Klasifikasi. *Sistem Informasi (JUSIN)*, 1(1), 11–20.

Dominicus, D. A., Setiawan, N. Y., & Wicaksono, S. A. (2020). Prediksi Kecenderungan Pelanggan Telat Bayar pada Layanan Pembiayaan Adira Finance Saluran E-Commerce. *Jurnal Pengembangan Teknologi Informasi Dan Ilmu Komputer*, 4(4), 1300–1307.

Eadara, S., & Kumar, I. P. (2023). *Prediction Of Customer Loan Eligibility Using Random Forest Algorithm*. 14(01), 482–492.

Eriyanto, R. D., & Alamsyah, A. (2020). Analisis Kelayakan Kredit Pada Peer To Peer Lending Menggunakan Data Sosial Media (Studi Kasus: LinkedIn). *E-Proceeding of Management*, 7(2), 2045–2050.

Ibnu, M., & Rachmatullah, C. (2023). Penerapan SMOTE untuk Meningkatkan Kinerja Klasifikasi Penilaian Kredit. *JURIKOM (Jurnal Riset Komputer)*, 10(1), 2407–389. <https://doi.org/10.30865/jurikom.v10i1.5612>

Ismanto, E., & Novalia, M. (2021). Komparasi Kinerja Algoritma C4.5, Random Forest, dan Gradient Boosting untuk Klasifikasi Komoditas. *Techno.Com*, 20(3), 400–410. <https://doi.org/10.33633/tc.v20i3.4576>

Malakauskas, A., & Lakstutiene, A. (2021). Financial distress prediction for small and medium enterprises using machine learning techniques. *Engineering Economics*, 32(1), 4–14. <https://doi.org/10.5755/j01.ee.32.1.27382>

Pahlevi, O. (2018). Data Mining Penentuan Aturan Asosiasi Penjualan Makanan di Amaria Hotel Jakarta Menggunakan Algoritma Apriori. *Jurnal Sistem Informasi STMIK Antar Bangsa*, 2, 137–142.

Prasojo, B., & Haryatmi, E. (2021). Analisa Prediksi Kelayakan Pemberian Kredit Pinjaman dengan Metode Random. *Jurnal Nasional Teknologi Dan Sistem Informasi*, 07(02), 79–89.

Simarmata, K. B., Hartomo, K. D., & Hartomo, K. D.

- (2022). Analisa Rekomendasi Fitur Persetujuan Pinjaman Perusahaan Financial Technology Menggunakan Metode Random Forest. *JATISI (Jurnal Teknik Informatika Dan Sistem Informasi)*, 9(3), 2055–2070. <https://doi.org/10.35957/jatisi.v9i3.2258>
- Sugiyarti, E., & Maseleno, A. (2018). Sistem Pendukung Keputusan (DSS) Penyeleksian Pemilihan Penerima Beasiswa SMA N 1 Ulubelu Tanggamus Menggunakan Data Mining. *Konferensi Mahasiswa Sistem Informasi*, 6(1), 62–69.
- Widjiyati, N. (2021). Implementasi Algoritme Random Forest Pada Klasifikasi Dataset Credit Approval. *Jurnal Janitra Informatika Dan Sistem Informasi*, 1(1), 1–7. <https://doi.org/10.25008/janitra.v1i1.118>
- Zailani, A. U., & Hanun, N. L. (2020). Penerapan Algoritma Klasifikasi Random Forest Untuk Penentuan Kelayakan Pemberian Kredit Di Koperasi Mitra Sejahtera. *Infotech: Journal of Technology Information*, 6(1), 7–14. <https://doi.org/10.37365/jti.v6i1.61>
- Zhu, L., Qiu, D., Ergu, D., Ying, C., & Liu, K. (2019). A study on predicting loan default based on the random forest algorithm. *Procedia Computer Science*, 162(Itqm 2019), 503–513. <https://doi.org/10.1016/j.procs.2019.12.017>