Idealis: Indonesia Journal Information System
Volume 7, Nomor 2, Juli 2024
ISSN 2684-7280 (*online*)
Halaman 240-247
available online at http://jom.fti.budiluhur.ac.id/index.php/IDEALIS/index

PENERAPAN MODEL CRISP-DM PADA PREDIKSI NASABAH KREDIT MENGGUNAKAN ALGORITMA RANDOM FOREST

Dwi Bagus Saputra[1*], Vihi Atina[2], Faulinda Ely Nastiti[3]
[1,2,3] Fakultas Ilmu Komputer, Program Studi Teknik Informatika, Universitas Duta Bangsa, Surakarta, Indonesia
E-mail: [1*]dwibaguss324@gmail.com, [2]vihi_atina@udb.ac.id, [3]faulinda_ely@udb.ac.id
(* : corresponding author)

Abstrak- Kredit macet merupakan salah satu tantangan utama yang dihadapi oleh Koperasi Simpan Pinjam Baitut Tamwil Tazakka, yang berpotensi mengancam stabilitas keuangan dan kesehatan institusi. Penelitian ini bertujuan untuk mengevaluasi efektivitas algoritma *Random Forest* dalam memprediksi kredit macet pada koperasi tersebut. Metode *CRISP-DM* diterapkan dalam penelitian ini, meliputi tahap-tahap pemahaman bisnis, pemahaman data, persiapan data, pemodelan, evaluasi, dan *deployment*. Data yang digunakan terdiri dari 14 atribut dan 190 *record* yang telah dibersihkan dari nilai yang hilang. Hasil pemodelan menunjukkan bahwa algoritma *Random Forest* mampu memberikan akurasi prediksi yang sangat tinggi, dengan akurasi terbaik mencapai 94,8% pada tabel:10. Evaluasi metrik kinerja seperti *AUC, CA, F1 Score, Precision, Recall,* dan *MCC* menunjukkan nilai yang sangat baik, mengindikasikan kinerja prediktif yang kuat. Analisis *confusion matrix* juga mengonfirmasi akurasi prediksi yang tinggi dengan mayoritas prediksi benar pada kategori kredit macet, lancar, dan kurang lancar. Penelitian ini mengkonfirmasi bahwa algoritma *Random Forest* efektif dalam memprediksi kredit macet, menegaskan pentingnya penerapan *machine learning* dalam pengelolaan risiko kredit, dan memberikan kontribusi signifikan terhadap stabilitas keuangan koperasi melalui prediksi kredit macet yang lebih akurat.

Kata Kunci: Koperasi Simpan Pinjam, Kredit Macet, *Machine Learning*, Prediksi Kredit, *Random Forest*

Abstract- *Non-performing loans (NPLs) are one of the main challenges faced by Baitut Tamwil Tazakka Savings and Loan Cooperative, which can potentially threaten the financial stability and health of the institution. This study aims to evaluate the effectiveness of the Random Forest algorithm in predicting NPLs in the cooperative. The CRISP-DM method is applied in this study, encompassing the stages of business understanding, data understanding, data preparation, modeling, evaluation, and deployment. The data used consists of 14 attributes and 190 records that have been cleaned of missing values. The modeling results show that the Random Forest algorithm can provide very high prediction accuracy, with the best accuracy reaching 94.8% on a 90:10 dataset split. Performance metrics evaluation such as AUC, CA, F1 Score, Precision, Recall, and MCC indicate very good values, signifying strong predictive performance. Confusion matrix analysis also confirms high prediction accuracy with most correct predictions in the categories of non-performing, performing, and sub-performing loans. This study confirms that the Random Forest algorithm is effective in predicting NPLs, underscores the importance of applying machine learning in credit risk management, and contributes significantly to the financial stability of the cooperative through more accurate NPL predictions*.

Keywords*: Savings and Loan Cooperative, Non-Performing Loans, Machine Learning, Credit Prediction, Random Forest*

1. PENDAHULUAN

Lembaga keuangan mikro, seperti Koperasi Simpan Pinjam Baitut Tamwil Tazakka, memainkan peran vital dalam mendukung perekonomian lokal dengan memberikan akses keuangan kepada masyarakat yang kurang mampu. Namun, meskipun peran mereka sangat penting, lembaga-lembaga ini sering kali dihadapkan pada tantangan serius, salah satunya adalah meningkatnya jumlah kredit macet [1]. Peningkatan jumlah kredit macet, yang disebabkan oleh pertumbuhan portofolio kredit yang signifikan, memiliki potensi dampak negatif yang serius terhadap stabilitas keuangan dan kesehatan lembaga keuangan mikro tersebut.

Dalam konteks ini, penting untuk mencari solusi yang efektif untuk mengidentifikasi dan memitigasi risiko kredit macet. Salah satu pendekatan yang menjanjikan adalah penggunaan teknik *Machine Learning*, khususnya algoritma *Random Forest* [2]. Teknik ini telah terbukti efektif dalam berbagai bidang, termasuk dalam prediksi risiko kredit. *Random Forest* mampu menggabungkan prediksi dari beberapa pohon keputusan untuk menghasilkan hasil prediksi yang lebih akurat dan stabil. Keunggulan utama dari metode ini termasuk kemampuannya untuk menangani jumlah besar data, variabel yang berkorelasi, dan menghindari *overfitting* [3].

Sejumlah penelitian telah dilakukan untuk mengatasi masalah serupa dalam konteks yang berbeda. Misalnya, penelitian oleh Salma Navisa, Luqman Hakim, dan Aulia Nabilah (2021) mengadopsi metode *CRISP-DM* untuk membandingkan berbagai algoritma klasifikasi dalam mengidentifikasi genre musik di platform *Spotify* [4]. Hasil penelitian tersebut memberikan wawasan berharga tentang performa berbagai algoritma klasifikasi dalam situasi yang mirip dengan penelitian ini. Penelitian lain yang relevan adalah "Implementasi Algoritma Klasifikasi *Random Forest* Untuk Penilaian Kelayakan Kredit" yang dilakukan oleh Omar Pahlevi, Amrin, dan Yopi Handrianto (2023). Penelitian ini fokus pada penilaian kelayakan kredit menggunakan algoritma *Random Forest*, dengan tujuan

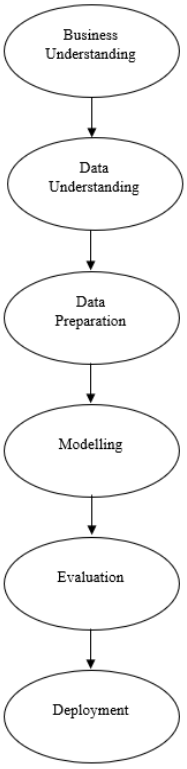
Idealis: Indonesia Journal Information System
Volume 7, Nomor 2, Juli 2024
ISSN 2684-7280 (*online*)
Halaman 240-247
available online at http://jom.fti.budiluhur.ac.id/index.php/IDEALIS/index

meningkatkan akurasi dan kecepatan dalam proses penilaian [5]. Hasil penelitian ini memberikan landasan yang kuat bagi pendekatan dalam memprediksi kredit macet.

Penelitian ini bertujuan untuk mengevaluasi efektivitas algoritma *Random Forest* dalam meningkatkan akurasi prediksi kredit macet di Koperasi Simpan Pinjam Baitut Tamwil Tazakka. Pendekatan yang mirip dengan penelitian sebelumnya diterapkan dengan menggunakan metode *CRISP-DM* untuk memandu proses analisis data. Data yang digunakan terdiri dari berbagai variabel yang relevan, seperti riwayat kredit, pendapatan, dan status pekerjaan, yang diharapkan dapat memberikan wawasan yang mendalam kepada lembaga keuangan mikro ini.

Dengan merujuk pada penelitian terdahulu yang relevan, diharapkan penelitian ini akan memberikan wawasan berharga bagi praktisi di lapangan dalam mengelola risiko kredit dan menjaga kesehatan keuangan lembaga keuangan mikro. Selain itu, metode *Random Forest* dipilih karena kemampuannya dalam memberikan prediksi yang akurat dan stabil, serta keunggulannya dalam mengatasi masalah *overfitting* dan menangani data yang kompleks dan berkorelasi tinggi.

2. METODE PENELITIAN



Gambar 1. Tahapan Metode *CRISP-DM*

Pada gambar 1 merupakan alur metode *CRISP-DM (Cross Industry Standard Process for Data Mining)*, yang dirancang untuk memberikan panduan langkah demi langkah dalam proses pengumpulan data[6]. Metode ini terdiri dari enam tahap yaitu pemahaman bisnis, pemahaman data, persiapan data, pemodelan, evaluasi, dan *deployment* [7].

2.1 Pemahaman Bisnis *(Business Understanding)*

Pada tahap ini, akan dilakukan upaya untuk mendapatkan pemahaman yang lebih terperinci mengenai permasalahan kredit macet yang dihadapi oleh Koperasi Simpan Pinjam Baitut Tamwil Tazakka. Langkah-langkah yang akan dilaksanakan mencakup studi literatur yang menyeluruh untuk mendalami konteks permasalahan kredit macet, khususnya dalam ranah koperasi simpan pinjam mikro seperti Baitut Tamwil Tazakka. Analisis literatur ini akan melibatkan penelusuran jurnal, buku, laporan, serta studi kasus yang relevan dengan permasalahan tersebut. Selain itu, data historis kredit yang tersedia pada koperasi akan dianalisis secara cermat guna memberikan pemahaman yang lebih terperinci mengenai pola kredit macet yang terjadi di masa lalu, faktor-faktor yang terkait, serta

Idealis: Indonesia Journal Information System
Volume 7, Nomor 2, Juli 2024
ISSN 2684-7280 (*online*)
Halaman 240-247
available online at http://jom.fti.budiluhur.ac.id/index.php/IDEALIS/index

karakteristik anggota yang mungkin berperan dalam meningkatkan risiko kredit. Integrasi informasi dari berbagai sumber ini diharapkan akan membantu dalam memperoleh pemahaman yang komprehensif mengenai permasalahan kredit macet yang dihadapi oleh Koperasi Simpan Pinjam Baitut Tamwil Tazakka.

2.2 Pemahaman Data *(Data Understanding)*

Pada tahap ini, akan dilakukan proses pengumpulan, pemahaman, dan persiapan data yang akan digunakan dalam analisis. Langkah-langkah yang akan diambil meliputi pengumpulan data dari sumber-sumber yang relevan, seperti informasi tentang pinjaman yang diberikan oleh koperasi, riwayat pembayaran, dan data informasi anggota. Data yang terkumpul akan dipelajari lebih lanjut untuk memahami struktur, karakteristik, dan kualitasnya. Ini mencakup identifikasi format data, penanganan nilai yang hilang atau tidak valid, serta penentuan apakah transformasi atau normalisasi data diperlukan dengan menggunakan feature selection dari aplikasi *orange*.

2.3 Persiapan Data *(Data Preparation)*

Setelah pemahaman awal terhadap data tercapai, langkah berikutnya adalah melakukan persiapan data untuk analisis lebih lanjut, yang meliputi proses pemisahan *dataset* menjadi dua bagian, yaitu data *training* dan data *testing*. Selanjutnya, dilakukan visualisasi *scatter plot* untuk mendapatkan gambaran data variabel. Melalui proses pengumpulan, pemahaman, dan persiapan data yang cermat, dapat dipastikan bahwa data yang digunakan dalam analisis memiliki kualitas yang baik dan siap digunakan dalam pengembangan model prediksi kredit macet.

2.4 Pemodelan *(Modelling)*

Pada tahap pemodelan, dilakukan proses pembangunan model menggunakan aplikasi *Orange*, dengan fitur *outlier function* dari perangkat lunak *Orange* dapat membantu untuk mengurangi data yang tidak sesuai. Selain itu, teknik validasi silang juga digunakan untuk memvalidasi kinerja model dengan membagi data ke dalam *subset* pelatihan dan pengujian secara berulang [8].

2.5 Evaluasi *(Evaluation)*

Pada tahap evaluasi, dilakukan penilaian terhadap kinerja model yang telah dibangun untuk mengukur seberapa baik model tersebut dalam memprediksi kredit macet. Langkah-langkah evaluasi mencakup pengukuran kinerja menggunakan metrik seperti akurasi, presisi, *recall, F1-score, MCC (Matthew Correlation Coefficient)* [9], serta analisis *confusion matrix* untuk memahami jenis kesalahan yang dilakukan oleh model [10].

Dalam praktiknya, pengukuran akurasi sering kali berasal dari matriks kebingungan, yang juga dikenal sebagai matriks klasifikasi. Matriks ini menggambarkan kelompok yang benar dan salah yang dibuat oleh pengklasifikasi untuk suatu set data tertentu. Setiap baris dan kolom dari matriks kebingungan mewakili kelas yang diprediksi dan kelas sebenarnya (aktual). Makna dari setiap elemen dalam matriks kebingungan dijelaskan dengan jelas dalam analisis hasil pengklasifikasi pada Tabel 1 dibawah.

Tabel 1. *Confusion Matrix*

		Actual Class	
Predicted Class		C1	C2
	1	n1,1 = number of C1 records classifed correctly	n2,1 = number of C2 records classifed incorrectly as C1
	2	n1,2 = number off C1 records classifed incorrectly as C2	n2,2 = number of C2 records classifed correctly

Matriks kebingungan dari Tabel 1 tersebut mengilustrasikan bagaimana pengklasifikasi membagi data ke dalam kategori yang tepat atau tidak tepat, berdasarkan kelas yang diprediksi dan kelas aktual. Analisis ini penting dalam evaluasi hasil pengklasifikasi untuk menilai kemampuan model dalam mengidentifikasi dengan akurat berbagai kelas dalam dataset yang diberikan.[11].

3. HASIL DAN PEMBAHASAN

3.1. *Business Understanding*

Seperti yang telah dibahas sebelumnya, permasalahan utama yang dihadapi adalah meningkatnya jumlah kredit macet yang tidak terkendali. Untuk itu, diperlukan upaya pencegahan dengan mencari solusi efektif guna

Idealis: Indonesia Journal Information System
Volume 7, Nomor 2, Juli 2024
ISSN 2684-7280 (*online*)
Halaman 240-247
available online at http://jom.fti.budiluhur.ac.id/index.php/IDEALIS/index

mengidentifikasi dan mengurangi risiko kredit macet. Salah satu metode yang dapat digunakan adalah teknik *Machine Learning*, khususnya algoritma *Random Forest*. Teknik ini telah terbukti efektif dalam berbagai bidang, termasuk prediksi risiko kredit [12] . Setelah menemukan solusi yang tepat, perlu dilakukan penelitian dan pemahaman mendalam mengenai algoritma *Random Forest* yang diterapkan dalam penelitian ini untuk mencapai hasil yang diinginkan.

3.2. *Data Understanding*

Data yang digunakan dalam penelitian ini adalah data privat yang diperoleh dari Lembaga Keuangan Koperasi Simpan Pinjam Baitut Tamwil Tazakka. *Dataset* yang berhasil dikumpulkan terdiri dari 14 atribut dan 190 *record* yang telah dibersihkan dari *missing value*, seperti yang ditunjukkan pada tabel 2 yang merupakan variabel yang akan digunakan dalam penelitian. Penelitian ini akan memanfaatkan seluruh atribut dalam *dataset* dengan tujuan untuk meningkatkan akurasi prediksi kredit macet di Koperasi Simpan Pinjam Baitut Tamwil Tazakka.

Tabel 2. *Dataset*

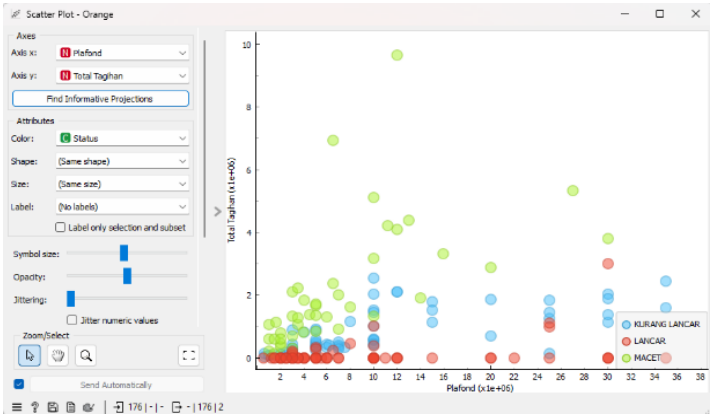
No	Nama	Type	Role	Values
1	Pekerjaan	Categorical	Feature	Buruh, Karyawan, Pengusaha
2	Jumlah tanggungan	Numeric	Feature	
3	Aset	Numeric	Feature	
4	Total pendapatan	Numeric	Feature	
5	Total pengeluaran	Numeric	Feature	
6	Plafon	Numeric	Feature	
7	Jangka (bulan)	Numeric	Feature	
8	Sisa pokok pinjaman	Numeric	Feature	
9	Bunga	Numeric	Feature	
10	Tagihan ke	Numeric	Feature	
11	Tagihan pokok	Numeric	Feature	
12	Total tagihan	Numeric	Feature	
13	Kolektibilitas	Numeric	Feature	
14	Status	Categorical	Target	Kurang Lancar, Macet, Lancar

3.3. *Data Preparation*

3.3.1 Data Latih dan Data Uji

Metode splitting data akan memisahkan data latih dan data uji. Dalam penelitian ini, akan diterapkan *pruning* dan menggunakan 10 pohon *(trees)* dengan kedalaman maksimal 5. Komposisi data latih dan data uji yang akan digunakan adalah sebagai berikut: 90% data latih dan 10% data uji, 80% data latih dan 20% data uji, 70% data latih dan 30% data uji, serta 60% data latih dan 40% data uji [13].

3.3.2 Visualisasi Scatter Plot



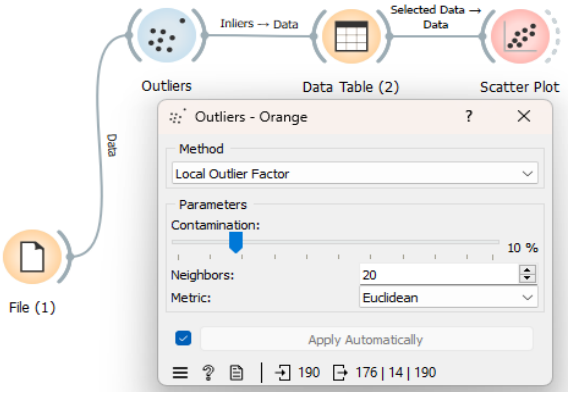
Gambar 2. Visualisasi Scatter Plot

Idealis: Indonesia Journal Information System
Volume 7, Nomor 2, Juli 2024
ISSN 2684-7280 (*online*)
Halaman 240-247
available online at http://jom.fti.budiluhur.ac.id/index.php/IDEALIS/index

Pada gambar 2 menunjukan bahwa terdapat hubungan antara status kelancaran kredit dengan plafon pinjaman dan total tagihan menggunakan perangkat lunak *Orange* Data Mining. Kumpulan titik merah atau peminjam dengan status lancar mendominasi area dengan plafon pinjaman kecil dan total tagihan kecil. Sementara itu, kumpulan titik biru atau peminjam dengan status kurang lancar mendominasi area dengan plafon pinjaman besar dan total tagihan besar. Dari analisis deskriptif ini dapat disimpulkan bahwa status kelancaran kredit dapat dipengaruhi oleh plafon pinjaman dan total tagihan. Kemudian hasil *scatter plot* juga menunjukkan bahwa terdapat beberapa data yang terlalu jauh dari kelompok data lainnya. Beberapa data ini perlu dikeluarkan agar algoritma pembelajaran mesin dapat bekerja secara optimal.

3.3.3 Data Reduction

Dengan bantuan fungsi *outlier* dari *Orange Software* dapat membantu mengurangi data yang berada di luar jangkauan. Hasil dari *outlier* diperoleh 14 jumlah data yang tidak valid untuk digunakan pada gambar 3 dibawah.

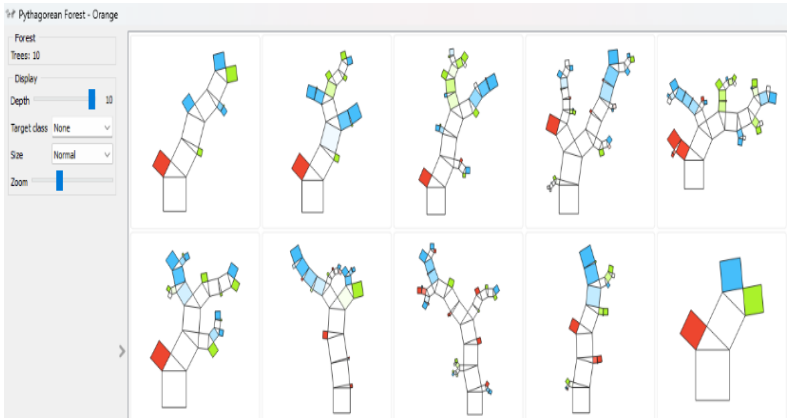


Gambar 3. *Outlier - Orange*

Setelah dilakukan reduksi data, data tersebut siap digunakan dalam algoritma *machine learning*. Dalam penelitian ini, peneliti menggunakan algoritma *random forest* untuk melatih dan pengujian data.

3.4. *Modelling*

3.4.1 *Pythagorean Random Forest*



Gambar 4. *Pythagorean Random Forest*

Gambar 4 diatas merupakan hasil visualisasi dari *Pythagorean* forest dengan total 10 pohon menggambarkan mekanisme kerja algoritma *random forest* dalam menentukan prediksi.

Idealis: Indonesia Journal Information System
Volume 7, Nomor 2, Juli 2024
ISSN 2684-7280 (*online*)
Halaman 240-247
available online at http://jom.fti.budiluhur.ac.id/index.php/IDEALIS/index

3.4.2 *Cross Validation*

Pada pemodelan dengan *cross-validation* menggunakan 10 *fold*, data latih dan data uji akan dibagi menjadi 10 partisi yang berbeda [14]. Proses *cross-validation* akan menggunakan algoritma yang sama yang digunakan pada pembagian *dataset* untuk *random forest*.

Tabel 3. Hasil Pembagian Data dan *Cross Validation*

Dataset	Akurasi *Random Forest*
90 - 10	94.8%
80 - 20	93.1%
70 - 30	93.8%
60 - 40	94.3%
Cross Validation	94.4%

Pada Tabel 3 diatas menunjukan hasil dari algoritma *Random Forest* dengan pembagian *dataset* 90:10 memberikan hasil akurasi terbaik tanpa menggunakan *cross validation*. Oleh karena itu, dalam penelitian ini, *cross validation* tidak diperlukan.

3.5. *Evaluation*

Kinerja prediksi dari algoritma *random forest* dapat diukur menggunakan berbagai metrik, termasuk *AUC (area under the curve)*, *CA* (akurasi), *F1 Score*, *Precision*, *Recall*, dan *MCC*. Berikut hasil pengukuran pada tabel dibawah dari berbagai metrik ini menunjukkan kinerja prediktif yang cukup baik, mendekati nilai 1 [15].

Tabel 4. Hasil Model *Random Forest*

Model	AUC	CA	F1	Prec	Recall	MCC
Random Forest	0.999	0.974	0.974	0.976	0.974	0.961

Tabel 4 menggambarkan hasil evaluasi performa model *Random Forest* dalam penelitian ini. Model ini mencapai kinerja yang sangat baik dengan nilai *AUC* mencapai 0.999, mencerminkan kemampuan model dalam membedakan kelas-kelas dengan presisi tinggi. Akurasi (*CA*) sebesar 0.974 menunjukkan tingkat keakuratan model secara umum. *F1 Score* dan *Precision* masing-masing mencapai 0.974 dan 0.976, menunjukkan keseimbangan yang baik antara presisi dan *recall* dalam mengklasifikasikan kelas positif. *Recall* sebesar 0.974 menunjukkan kemampuan model dalam mengenali sebagian besar kasus positif. *Matthews Correlation Coefficient (MCC)* mencapai 0.961, mengindikasikan tingkat korelasi yang kuat antara prediksi model dengan label aktual, menegaskan kehandalan model *Random Forest* dalam analisis ini.

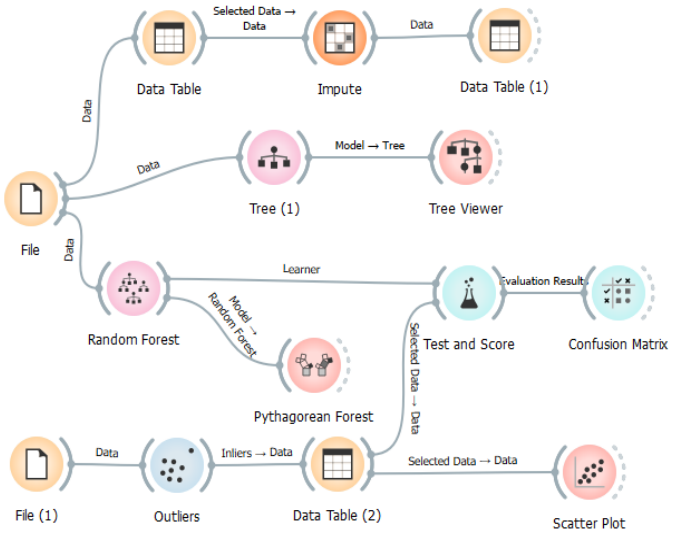
Tabel 5. Hasil *Confusion Matrix*

		Predicted			
		Kurang Lancar	Lancar	Macet	∑
Actual	Kurang Lancar	63	1	0	64
	Lancar	2	67	1	70
	Macet	1	0	41	42
	∑	66	68	42	176

Berdasarkan analisis Tabel 5 *confusion matrix*, dapat disimpulkan bahwa prediksi algoritma *Random Forest* dari total 176 data menunjukkan hasil yang akurat. Pada kategori kredit macet, terdapat prediksi 41 benar dari 42 data. Pada kategori lancar, terdapat 67 prediksi benar dari 70 data. Sementara itu, pada kategori kredit kurang lancar, terdapat 63 prediksi benar dari 64 data.

Pada gambar 5 menjelaskan proses secara keseluruhan prediksi menggunakan algoritma *Random Forest.* Proses penelitian dimulai dengan langkah awal pembersihan data untuk menangani nilai yang hilang dan mengurangi outlier yang berpotensi mempengaruhi akurasi prediksi. Selanjutnya, data dipersiapkan untuk analisis lebih lanjut.

Idealis: Indonesia Journal Information System
Volume 7, Nomor 2, Juli 2024
ISSN 2684-7280 (*online*)
Halaman 240-247
available online at http://jom.fti.budiluhur.ac.id/index.php/IDEALIS/index



Gambar 5. Alur Prediksi Algoritma *Random Forest* Menggunakan *Orange*

Algoritma *Random Forest* kemudian diterapkan dengan konfigurasi yang telah ditentukan sebelumnya, termasuk spesifikasi jumlah pohon (*trees*) dan kedalaman maksimal. Model dilatih menggunakan data latih guna mengidentifikasi pola yang signifikan dalam dataset. Evaluasi kinerja model dilakukan dengan menggunakan data uji, yang menghasilkan prediksi yang dapat diandalkan untuk memahami dan mengantisipasi fenomena yang relevan dalam data kredit. Dengan harapan hasil dari penelitian ini dapat memberikan kontribusi yang signifikan dalam pengelolaan risiko kredit yang lebih proaktif, berbasis pada prediksi kredit macet yang lebih akurat dan efektif dalam konteks keuangan mikro.

4. KESIMPULAN

Kesimpulan dari studi ini menunjukkan bahwa algoritma *Random Forest* efektif dalam meramalkan kemungkinan kredit macet di Koperasi Simpan Pinjam Baitut Tamwil Tazakka, dengan mencapai tingkat akurasi prediksi yang tinggi hingga 94,8% pada pembagian *dataset* 90:10. Implikasi dari penelitian ini sangat relevan dalam konteks manajemen risiko kredit di lembaga keuangan mikro, menawarkan pendekatan yang lebih akurat dan efisien untuk menjaga stabilitas keuangan institusi. Meskipun demikian, studi ini terbatas oleh keterbatasan dalam mengakses data yang lebih luas dan fokus hanya pada satu jenis algoritma *Machine Learning*. Untuk penelitian berikutnya, disarankan untuk memperluas cakupan data dan menggabungkan beberapa algoritma untuk analisis yang lebih menyeluruh. Dengan demikian, kontribusi penelitian ini diharapkan dapat meningkatkan strategi manajemen risiko yang lebih proaktif dan terapan, memberikan manfaat yang lebih besar bagi masyarakat umum.

DAFTAR PUSTAKA

[1] Y. Rahmawati *et al.*, "Analisis Penyelesaian Kredit Bermasalah Melalui Rescheduling Untuk Meningkatkan Profitabilitas Pada Masa Pandemi Covid-19 Pada Koperasi Simpan Pinjam Dan Pembiayaan Syariah (KSPPS) Cisaat Analysis Of Problem Loan Solution Through Rescheduling To Increase Profitability During The Covid-19 Pandemic In Sharia Save And Finance Cooperatives (KSPPS) CISAAT," :*Journal of Economic, Business and Accounting*, vol. 5, 2022, Accessed: Jun. 28, 2024. [Online]. Available: https://journal.ipm2kpe.or.id/index.php/COSTING/article/view/2512

[2] E. Wahyu and H. Budianto, "Bank Bukopin Syariah dan Konvensional: Studi Pustaka (Library Research) dan Bibliometrik VOSviewer," 2023, doi: 10.5281/zenodo.10049572.

[3] J. Sanjaya, E. Renata, V. E. Budiman, F. Anderson, and M. Ayub, "Prediksi Kelalaian Pinjaman Bank Menggunakan Random Forest dan Adaptive Boosting," *Jurnal Teknik Informatika dan Sistem Informasi*, vol. 6, no. 1, Apr. 2020, doi: 10.28932/jutisi.v6i1.2313.

[4] S. N. Luqman *et al.*, "Komparasi Algoritma Klasifikasi Genre Musik pada Spotify Menggunakan CRISP-DM," 2021. Accessed: Jun. 28, 2024. [Online]. Available: https://apic.id/jurnal/index.php/jsc/article/download/162/86

[5] O. Pahlevi and Y. Handrianto, "Implementasi Algoritma Klasifikasi Random Forest Untuk Penilaian Kelayakan Kredit," *Jurnal Infortech*, vol. 5, no. 1, 2023, [Online]. Available: http://ejournal.bsi.ac.id/ejurnal/index.php/infortech

[6] H. Indrawan, B. Irawan, and T. Suprapti, "Klasifikasi Ulasan Pengguna Aplikasi Access by KAI Berbasis Aspek Dengan Algoritma Naïve Bayes dan SVM," 2023. Accessed: Jun. 28, 2024. [Online]. Available: https://ejournal.itn.ac.id/index.php/jati/article/view/8234
[7] C. Cahyaningtyas, D. Manongga, and I. Sembiring, "Algorithm Comparison And Feature Selection For Classification Of Broiler Chicken Harvest," *Jurnal Teknik Informatika (Jutif)*, vol. 3, no. 6, pp. 1717–1727, Dec. 2022, doi: 10.20884/1.jutif.2022.3.6.493.
[8] A. B. Prasetyo and T. G. Laksana, "Optimasi Algoritma K-Nearest Neighbors dengan Teknik Cross Validation Dengan Streamlit (Studi Data: Penyakit Diabetes)," 2022. [Online]. Available: http://jurnal.polibatam.ac.id/index.php/JAIC
[9] N. Hasdyna and R. Kesuma Dinata, "Analisis Matthew Correlation Coefficient pada K-Nearest Neighbor dalam Klasifikasi Ikan Hias," 2020. Accessed: Jun. 28, 2024. [Online]. Available: https://jurnal.unej.ac.id/index.php/INFORMAL/article/download/18907/8555
[10] N. Khoirunnisaa, K. Nabila, N. Kesuma, S. Setiawan, A. Yunizar, and P. Yusuf, "Klasifikasi Teks Ulasan Aplikasi Netflix Pada Google Play Store Menggunakan Algoritma Naive Bayes dan SVM," *SKANIKA: Sistem Komputer dan Teknik Informatika*, vol. 7, no. 1, pp. 64–73, 2024, Accessed: Jun. 28, 2024. [Online]. Available: https://jom.fti.budiluhur.ac.id/index.php/SKANIKA/article/view/3138
[11] T. H. Hasibuan and D. Mahdiana, "Prediksi Kelulusan Mahasiswa Tepat Waktu Menggunakan Algoritma C4.5 Pada Uin Syarif Hidayatullah Jakarta," *SKANIKA: Sistem Komputer dan Teknik Informatika*, vol. 6, pp. 61–74, 2023, Accessed: Jun. 28, 2024. [Online]. Available: https://jom.fti.budiluhur.ac.id/index.php/SKANIKA/article/view/2976
[12] Intan Permata and Esther Sorta Mauli Nababan, "Application Of Game Theory In Determining Optimum Marketing Strategy In Marketplace," *Jurnal Riset Rumpun Matematika Dan Ilmu Pengetahuan Alam*, vol. 2, no. 2, pp. 65–71, Jul. 2023, doi: 10.55606/jurrimipa.v2i2.1336.
[13] A. Citra Mawani, L. Li Hin, and D. Anubhakti, "Deteksi Dini Gejala Awal Penyakit Diabetes Menggunakan Algoritma Random Forest," 2023. [Online]. Available: http://jom.fti.budiluhur.ac.id/index.php/IDEALIS/indexAjengCitraMawani|http://jom.fti.budiluhur.ac.id/index.php/IDEALIS/index|
[14] R. Rizqi Robbi Arisandi, B. Warsito, and A. Rachman Hakim, "Aplikasi Naïve Bayes Classifier (Nbc) Pada Klasifikasi Status Gizi Balita Stunting Dengan Pengujian K-Fold Cross Validation," *Jurnal Gaussian*, vol. 11, no. 1, pp. 130–139, 2022, [Online]. Available: https://ejournal3.undip.ac.id/index.php/gaussian/
[15] E. Budiawan and M. Tryana Sembiring, "Utilizing Science Data to Increasing The Number Msme Debtors at PT.Bank Central Asia.Tbk (Case Study of PT. Bank Central Asia.Tbk Kcu Tebing Tinggi)," *Journal of Accounting Research, Utility Finance and Digital Assets*, vol. 2, 2023, [Online]. Available: https://jaruda.org