

IMPLEMENTATION OF HYPERPARAMETER TUNING IN RANDOM FOREST ALGORITHM FOR LOAN APPROVAL PREDICTION

Dwi Sandhi Bhakti¹, Agung Prasetyo², Primandani Arsi³

^{1,2,3}Informatics, Computer Science Faculty, Universitas Amikom Purwokerto, Indonesia

Email: ¹sandybhakti11@gmail.com, ²pras@amikompurwokerto.ac.id, ³ukhti.prima@amikompurwokerto.ac.id

(Article received: May 01, 2024; Revision: May 17, 2024; published: July 29, 2024)

Abstract

The risk of non-performing loan is a significant issue in the financial industry, including banks and cooperatives. Loan default risks can occur due to various reasons, and one of them is the negligence of staff or subjective decision-making in loan approval. The proposed solution is to enhance an objective and accurate loan approval decision-making system through the application of machine learning technology, aiming to reduce the risk of loan default. The Random Forest algorithm has proven to be the best in predicting loan approval compared to other supervised learning models. Optimization was performed on the Random Forest algorithm through hyperparameter tuning and data balancing using SMOTE. The best accuracy obtained from several experiments was 86.2%. By implementing optimizations on the Random Forest algorithm, it is expected that the model can make loan approval predictions more objectively and accurately, serving as a reference for future loan approval system development.

Keywords: hyperparameter tuning, loan approval, machine learning, non performing loan, random forest.

IMPLEMENTASI HYPERPARAMETER TUNING PADA ALGORITMA RANDOM FOREST UNTUK PREDIKSI PERSETUJUAN

Abstrak

Risiko kredit macet menjadi masalah yang signifikan dalam industri keuangan seperti bank dan koperasi. Risiko kredit macet dapat terjadi berdasarkan beberapa alasan dan salah satunya adalah kelalaian petugas atau keputusan petugas yang subjektif dalam menetapkan persetujuan pinjaman. Solusi yang diusulkan adalah dengan meningkatkan sistem pengambilan keputusan persetujuan pinjaman yang objektif dan akurat melalui penerapan teknologi *machine learning* diharapkan mampu mengurangi risiko kredit macet. Algoritma *Random Forest* yang sudah terbukti mampu menjadi yang terbaik dalam melakukan prediksi persetujuan pinjaman dari model *supervised learning* lainnya. Dilakukan optimasi pada algoritma *Random Forest* dengan melakukan *hyperparameter tuning* dan melakukan *balancing data* menggunakan SMOTE. Didapatkan akurasi terbaik dari beberapa eksperimen dengan akurasi 86.2%. Dengan menerapkan optimasi pada algoritma *Random Forest* diharapkan model mampu melakukan prediksi persetujuan pinjaman yang lebih objektif dan akurat dan menjadi referensi untuk pengembangan sistem persetujuan pinjaman dimasa depan.

Kata kunci: kredit macet, pembelajaran mesin, penyetelan hiperparameter, persetujuan pinjaman, random forest.

1. PENDAHULUAN

Pinjaman atau kredit merupakan proses mendapatkan barang atau uang dari sebuah bank atau lembaga keuangan yang kemudian dibayarkan dikemudian hari dengan cicilan atau angsuran sesuai dengan perjanjian. pembayarannya dilakukan di kemudian hari dengan cicilan atau angsuran sesuai dengan perjanjian. Tak dapat dipungkiri bahwa bank memegang peran sentral dalam urusan kredit, bahkan kegiatan ini menjadi salah satu pilar utama dalam operasional lembaga keuangan [1].

Memberikan pinjaman kepada pelanggan merupakan aktivitas berisiko tinggi yang dilakukan secara rutin. Dalam pelaksanaannya, masalah kredit yang tidak lancar sering terjadi karena kurangnya analisis kredit yang hati-hati atau kurangnya kecermatan dalam proses penilaian kredit, serta perilaku buruk dari pelanggan. Agar dapat mencegah risiko kredit yang bermasalah, seorang analis kredit di lembaga perbankan harus mampu membuat keputusan yang tepat dalam menerima atau menolak pengajuan kredit [2]. Namun, perlu diingat bahwa

manusia memiliki sifat subjektifitas, yang dapat mempengaruhi keputusan analisis kredit.

Dalam proses pemberian kredit, bank atau lembaga keuangan akan selalu berhadapan dengan masalah kredit macet atau *Non Performing Loan* (NPL). Ratio NPL akan selalu berdampak terhadap keuntungan sebuah lembaga keuangan [3]. Serta semakin tinggi ratio NPL maka akan semakin buruk pula kualitas kredit pada bank atau lembaga keuangan tersebut [4]. Sejak 2013 hingga 2023 ratio NPL di Indonesia mengalami tren kenaikan, tercatat pada 2021 ratio NPL Indonesia sebesar 2.65% dan mendapat peringkat 4 tertinggi di ASEAN dan peringkat ke 7 dari 20 negara yang tergabung pada G20 [5]. Berdasarkan data *Global Economic Monitor* (GEM) selama sepuluh tahun terakhir, puncak kenaikan ratio NPL Indonesia terjadi pada tahun 2021 yaitu 3.35% [6].

Terdapat 9 masalah yang menyebabkan terjadinya kredit macet yaitu, kelemahan dalam analisis kredit, kelemahan dalam dokumen kredit, kelemahan dalam supervisi kredit, kecerobohan petugas bank, kelemahan kebijaksanaan kredit, kelemahan bidang agunan, kelemahan sumber daya manusia, kelemahan teknologi dan kecurangan petugas bank [7]. Berdasarkan kondisi tersebut, diperlukan metode penetapan persetujuan yang lebih objektif dan dapat meminimalisir terjadinya kredit macet atau NPL. *Machine learning* dapat menjadi solusi untuk melakukan klasifikasi terkait persetujuan pinjaman. *Machine learning* merupakan teknologi yang mampu melakukan proses pelatihan algoritma terhadap suatu data sehingga algoritma dapat melakukan pekerjaan secara otomatis [8]. Model *machine learning* dapat melatih model untuk melakukan klasifikasi persetujuan pinjaman berdasarkan data riwayat yang ada [9]. Dengan menggunakan *machine learning* keputusan persetujuan pinjaman akan lebih objektif dan mempermudah operasional pemberi pinjaman atau lembaga keuangan.

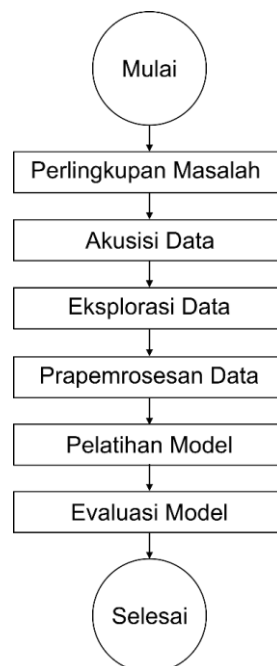
Proses klasifikasi menggunakan *machine learning* terdapat banyak algoritma yang dapat digunakan seperti *k-Nearest Neighbor* (KNN), *Naive Bayes*, *Support Vector Maching* dan *Random Forest* [10]. Pada kasus prediksi persetujuan pinjaman *Random Forest* mendapatkan hasil evaluasi yang terbaik diantara algoritma klasifikasi lainnya. Dalam komparasi berbagai model dalam melakukan klasifikasi persetujuan pinjaman, model dasar *Random Forest* mendapat akurasi tertinggi dengan skor 82% [11]. *Random Forest* juga mendapat skor evaluasi tertinggi dengan tambahan proses *balancing* data mengguna SMOTE mencapai akurasi 98% [12]. Berdasarkan penelitian tersebut algoritma dasar *Random Forest* sudah dapat melakukan klasifikasi dengan sangat baik, mungkin akan lebih baik lagi jika ditambahkan proses *hyperparameter tuning*. Seperti hasil penelitian dari (K. Sarao, 2021) terkait sistem prediksi transaksi fraud, model *random forest* dengan

hyperparameter tuning mendapat akurasi yang lebih tinggi yaitu 94.9% dibandingkan model dasar yang hanya 88.5%.

Hyperparameter tuning merupakan proses pencarian kombinasi terbaik dari nilai-nilai *hyperparameter* untuk suatu model *machine learning* dengan tujuan meningkatkan kinerja model tersebut [13]. Dengan mencari kombinasi *hyperparameter* yang optimal dapat meningkatkan kinerja model *machine learning* dan membuatnya lebih sesuai dengan data yang digunakan.

Dengan demikian, masalah kredit macet atau NPL dalam industri keuangan memerlukan solusi yang lebih objektif dalam meminimalisir kesalahan persetujuan pinjaman. Melalui penggunaan teknologi *machine learning*, khususnya algoritma *Random Forest* dengan *hyperparameter tuning*, akan menawarkan solusi yang efektif. Diharapkan dengan pendekatan ini dapat meminimalisir risiko kredit macet dan meningkatkan keputusan persetujuan yang akurat.

2. METODE PENELITIAN



Gambar 1. Metode Penelitian

Berdasarkan pada Gambar 1 dijelaskan terdapat beberapa langkah pengembangan sistem yang berpedoman dengan *AI project cycle*. *AI project cycle* atau Siklus proyek AI merupakan serangkaian langkah dan tahapan yang harus diikuti dalam pengembangan dan penerapan solusi berbasis AI [14], langkah-langkah tersebut sebagai berikut:

2.1. Perlingkupan Masalah (*Problem scoping*)

Digunakan teknik 4Ws *Canvas* untuk memudahkan tahap *problem scoping*. 4Ws *Canvas* memudahkan dalam merinci ide, memahami konteks, dan merencanakan proyek dengan lebih baik [15].

Canvas ini berfokus pada empat unsur utama yang dimulai dengan huruf "W," yaitu *What* (Apa), *Who* (Siapa), *Why* (Mengapa), dan *Where* (Dimana), yang bertujuan untuk meningkatkan pemahaman yang lebih mendalam dan menyeluruh terkait dengan subjek penelitian [16]. Berdasarkan latar belakang pada penelitian tercipta perlingkupan masalah yang ditunjukkan pada Tabel 1 berikut.

Tabel 1. Perlingkupan Masalah

4Ws	Pertanyaan	Jawaban
Who	Siapa yang menjadi subjek penelitian?	Instansi / lembaga keuangan yang terlibat dalam proses persetujuan pinjaman.
What	Apa tantangan yang dihadapi oleh instansi / lembaga keuangan?	Tantangannya berkaitan dengan mengevaluasi risiko kredit macet kepada pemohon pinjaman secara efisien dan akurat untuk menentukan apakah akan menyetujui atau menolak permintaan pinjaman.
Where	Di mana proses persetujuan pinjaman berlangsung, dan di mana penerapan algoritma <i>machine learning</i> mungkin bermanfaat?	Proses persetujuan pinjaman berlangsung di lembaga keuangan seperti bank, koperasi kredit, atau platform pemberian pinjaman online. <i>Machine learning</i> dapat disematkan untuk menyederhanakan proses, mengurangi upaya manual, dan meningkatkan akurasi keputusan.
Why	Mengapa penting untuk memanfaatkan <i>machine learning</i> dalam proses persetujuan pinjaman?	Penerapan <i>machine learning</i> dalam persetujuan pinjaman meningkatkan efisiensi, mempercepat pengambilan keputusan, dan menilai risiko dengan lebih akurat. Ini dapat mengurangi risiko gagal bayar berdasarkan model yang telah dilatih sebelumnya.

2.2. Akuisi Data (Data Acquisition)

Dataset yang digunakan dalam penelitian ini merupakan dataset publik dengan judul *Loan Eligible*

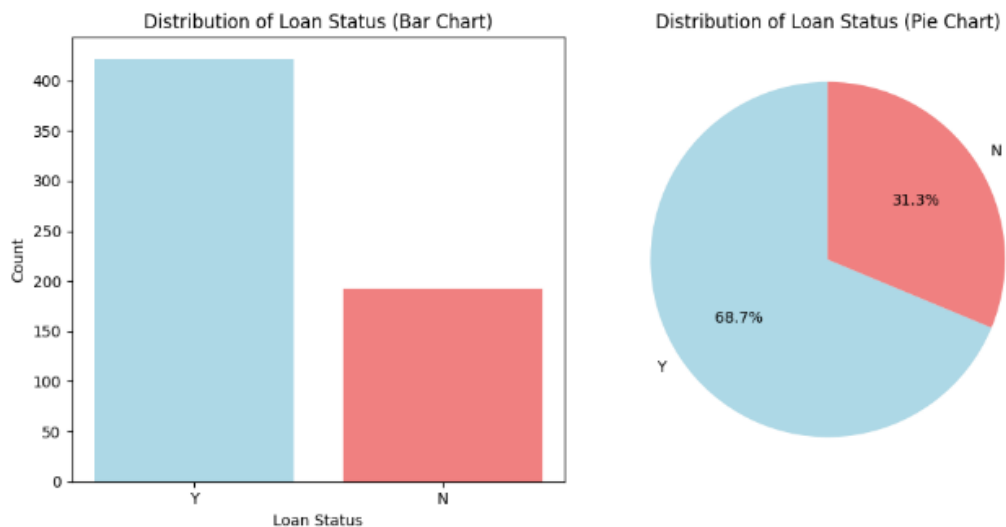
Dataset yang di rilis pada platform *Kaggle* dan dibawah lisensi *Database Contents License (DbCL)* v1.0 [17]. Dataset ini terdiri dari 614 baris data dan 13 kolom atribut, untuk spesifikasi detailnya ditunjukkan pada Tabel 2 berikut.

Tabel 2. Spesifikasi Dataset

No.	Atribut	Deskripsi	Nilai
1	<i>Loan_ID</i>	Unik ID	Acak
2	<i>Gender</i>	Jenis kelamin	<i>Male / Female</i>
3	<i>Married</i>	Status pernikahan	Y / N
4	<i>Dependents</i>	Jumlah tanggungan keluarga	
5	<i>Education</i>	Pendidikan	<i>Graduate / Under Graduate</i>
6	<i>Self_Employed</i>	Status pekerja mandiri	Y / N
7	<i>ApplicationIncome</i>	Pendapatan pemohon	150-81000
8	<i>CoapplicationIncome</i>	Pendapatan pendamping pemohon	0-41667
9	<i>LoanAmount</i>	Jumlah pinjaman diajukan	9-700
10	<i>Loan_Amount_Term</i>	Jangka waktu pinjaman	12-480
11	<i>Credit_History</i>	Riwayat kredit	0-1
12	<i>Property_Area</i>	Lokasi properti	<i>Urban / Semi-Urban / Rural</i>
13	<i>Loan_Status</i>	Status pinjaman	Y / N

2.3. Eksplorasi Data (Data Exploration)

Pada tahap ini dilakukan eksplorasi data untuk mengetahui mengidentifikasi pola, anomali, tren, dan informasi penting dalam dataset [14]. Berdasarkan atribut targetnya yaitu '*Loan_Status*' dataset ini memiliki distribusi yang tidak seimbang atau *imbalance* seperti yang ditampilkan pada Gambar 2, dengan fekuensi bernilai Y (disetujui) sebanyak 422 dan nilai N (tidak disetujui) sebanyak 192.



Gambar 2. Distribusi Atribut Target

Pada dataset ditemukan beberapa data yang terdapat *missing value*, data tersebut ditunjukkan pada Tabel 3 berikut.

Tabel 3. Temuan *Missing Value*

Atribut	Jumlah <i>Missing Value</i>
<i>Loan_ID</i>	0
Gender	13
<i>Married</i>	3
<i>Dependents</i>	15
<i>Education</i>	0
<i>Self_Employed</i>	32
<i>ApplicationIncome</i>	0
<i>CoapplicationIncome</i>	0
<i>LoanAmount</i>	22
<i>Loan_Amount_Term</i>	14
<i>Credit_History</i>	50
<i>Property_Area</i>	0
<i>Loan_Status</i>	0

2.4. Prapemrosesan Data (*Data Preprocessing*)

Pada tahap ini akan dilakukan berbagai persiapan dan pengolahan data terhadap dataset dengan tujuan memastikan kualitas data dan kebutuhan pada proses pelatihan model. Prapemrosesa data dimulai dari *data cleaning*, *balancing data* dan seleksi fitur.

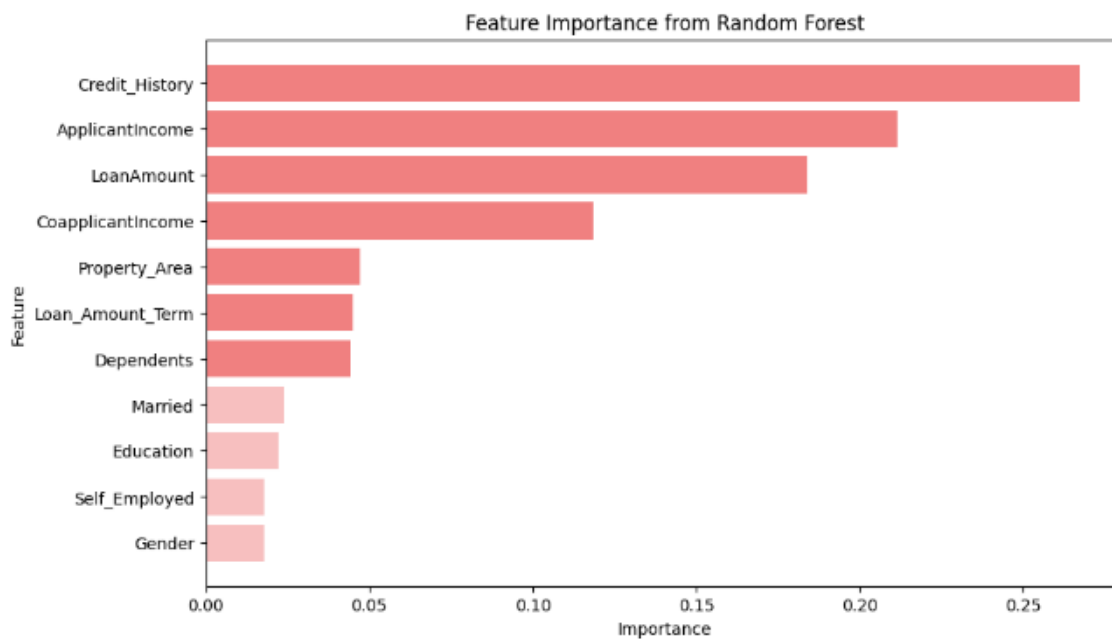
Data cleaning dilakukan dengan melakukan penghapusan atau *drop* pada atribut yang tidak digunakan dan menganani *missing value* pada data. Terdapat satu atribut yang tidak dipakai yaitu atribut

'*Loan_ID*'. Dalam penanganan *missing value* pada dataset dapat dilakukan dengan imputasi atau pengisian data berdasarkan *mean*, *median*, dan *modus* [18]. Tabel 4 berikut menampilkan penanganan imputasi pada masing-masing atribut dengan temuan *missing value*.

Tabel 4. Penganganan *Missing Value*

Atribut	Penganganan
Gender	<i>Modus</i>
<i>Married</i>	<i>Modus</i>
<i>Dependents</i>	<i>Modus</i>
<i>Self_Employed</i>	<i>Modus</i>
<i>LoanAmount</i>	<i>Median</i>
<i>Loan_Amount_Term</i>	<i>Median</i>
<i>Credit_History</i>	<i>Modus</i>

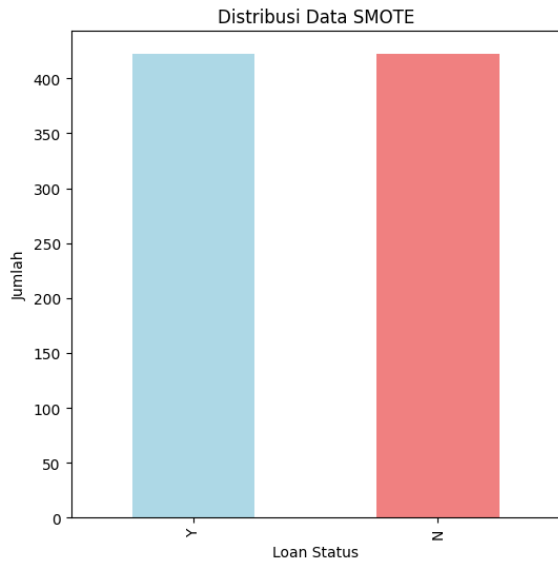
Setelah data bersih tahap terakhir prapemrosesan adalah seleksi fitur, seleksi fitur dilakukan untuk percobaan perbandingan performa antara dataset asli dengan dataset seleksi fitur. Proses seleksi fitur pada dataset mampu meningkatkan performa model [19]. Proses ini membantu menentukan variabel mana yang paling berdampak pada model prediktif, sehingga hanya variabel yang penting saja yang digunakan [20]. Seleksi fitur dilakukan dengan algoritma *Random Forest*, untuk mendapatkan *feature importance* dikarenakan RF lebih unggul dibandingkan model lain [21]. Tahap seleksi fitur menghasilkan 7 fitur terpilih seperti yang ditunjukkan pada Gambar 3 berikut.



Gambar 3. Hasil Seleksi Fitur

Tahap terakhir prapemrosesan dilakukan *balancing data*, dikarenakan dataset imbalance diperlukan proses *balancing*. Pada proses ini dilakukan menggunakan *Synthetic Minority Over-sampling Technique* (SMOTE), SMOTE bekerja dengan cara membuat sampel sintesis (buatan) untuk kelas minoritas berdasarkan sampel yang sudah ada

[22]. Penerapan SMOTE pada dataset yang *imbalance* berhasil meningkatkan akurasi pada hasil pelatihan model *Random Forest* [23]. Hasil dari SMOTE memberi penambahan data pada data minoritas sehingga jumlah baris data berubah menjadi 844 baris, hasil dari SMOTE ditunjukkan pada Gambar 4 berikut.



Gambar 4. Hasil SMOTE

2.5. Pelatihan Model (Pelatih Model)

Setelah melewati tahap prapemrosesan, didapatkan 4 dataset yang akan digunakan untuk eksperimen pada penelitian ini. Dataset tersebut adalah dataset asli (original). Dataset dengan seleksi fitur, dataset asli dengan SMOTE dan dataset dengan seleksi fitur dan SMOTE.

Sebelum memasuki tahap pelatihan model, semua dataset akan dibagi menjadi dua data dengan komposisi 80% untuk data latih dan 20% untuk data test [24]. Hasil dari split data pada masing-masing dataset ditunjukkan pada Tabel 5 berikut.

Tabel 5. Hasil Pembagian Dataset

Dataset	Data Latih	Data Uji
Dataset Asli	491	123
Dataset Seleksi Fitur	491	123
Dataset Asli Smote	675	169
Dataset Seleksi Fitur Smote	675	169

Pada proses pelatihan data *hyperparameter tuning* merupakan teknik yang penting dalam mendapatkan model *machine learning* yang tepat terhadap dataset tertentu [25]. Penelitian ini melakukan eksperimen penerapan *hyperparameter tuning* pada algoritma *Random Forest* (RF) terhadap 4 dataset yang sudah disiapkan. Detail parameter yang akan dituning ditampilkan pada Tabel 6 berikut.

Tabel 6. Hasil Pembagian Dataset

Model	Parameter	Nilai
<i>Random Forest</i>	n_estimators	[50, 100, 200]
	max_depth	[None, 10, 20, 30]
	min_samples_split	[2, 5, 10]
	min_samples_leaf	[1, 2, 4]
	criterion	['gini', 'entropy']

Ditunjukkan pada Tabel 7 didapatkan nilai parameter yang berbeda dalam setiap dataset. Hal tersebut dikarenakan model *Random Forest* menemukan metrik evaluasi terbaik pada nilai parameter tersebut disetiap datasetnya.

Tabel 7. Hasil *Hyperparameter Tuning*

Dataset	Parameter	Nilai
Dataset Asli	n_estimators	50
	max_depth	10
	min_samples_split	2
	min_samples_leaf	2
	criterion	gini
Dataset Seleksi Fitur	n_estimators	50
	max_depth	10
	min_samples_split	5
	min_samples_leaf	1
	criterion	gini
Dataset Asli Smote	n_estimators	50
	max_depth	10
	min_samples_split	2
	min_samples_leaf	2
	criterion	gini
Dataset Seleksi Fitur Smote	n_estimators	50
	max_depth	None
	min_samples_split	5
	min_samples_leaf	2
	criterion	gini

2.6. Evaluasi (Evaluation)

Dalam melakukan evaluasi terhadap performa model, digunakanlah laporan klasifikasi dengan metrik evaluasi yang mencakup akurasi, presisi, *recall* dan *f1-score* [26]. Sasaran dari metrik evaluasi adalah untuk menilai kualitas kesesuaian antara prediksi model dengan nilai sebenarnya yang terjadi dalam suatu sistem [27]. Metrik evaluasi ditunjukkan pada persamaan (1) - (4).

$$Akurasi = \frac{Jumlah\ prediksi\ benar}{Total\ jumlah\ prediksi} \tag{1}$$

$$Precision = \frac{True\ Positive\ (TP)}{True\ Positive\ (TP)+False\ Negative\ (FN)} \tag{2}$$

$$Recall = \frac{True\ Positive\ (TP)}{True\ Positive\ (TP)+False\ Positive\ (FP)} \tag{3}$$

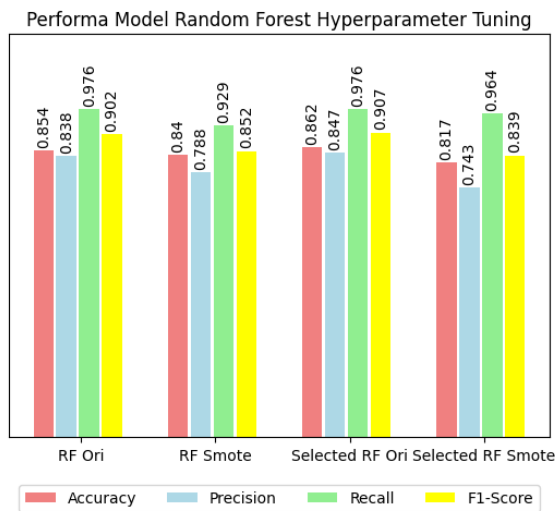
$$F1 - score = 2 \times \frac{Precision \times Recall}{Precision + Recall} \tag{4}$$

3. HASIL DAN PEMBAHASAN

Eksperimen dilakukan dengan menguji model hasil pelatihan terhadap masing-masing data uji sesuai model tersebut.

Dari hasil evaluasi yang ditampilkan pada Gambar 5, terlihat bahwa performa terbaik dari empat model yang dievaluasi bervariasi tergantung pada dataset yang digunakan. Model pada dataset Selected RF Ori menonjol dengan akurasi tertinggi yaitu 86.2%, menunjukkan kemampuannya dalam memberikan prediksi yang benar secara keseluruhan pada dataset yang diberikan. Model pada dataset Selected RF Ori juga menampilkan presisi terbaik, menekankan kemampuannya dalam mengidentifikasi instance-instance yang relevan dengan lebih baik dan mengurangi jumlah *false positive*. Model pada dataset RF Ori dan Selected RF Ori masing-masing menunjukkan *recall* yang lebih tinggi, menandakan

kemampuannya dalam mengidentifikasi sebagian besar instance dari kelas positif dengan sangat baik. Sedangkan untuk *F1-score*, Selected RF Ori mendominasi dengan keseimbangan yang baik antara presisi dan *recall*, menunjukkan bahwa model ini cocok untuk kasus yang membutuhkan penekanan pada kedua aspek tersebut.



Gambar 5. Performa Model

Secara keseluruhan model dari dataset Selected RF Ori mendapatkan performa tertinggi pada penelitian ini. Analisis ini menegaskan pentingnya mempertimbangkan karakteristik dataset dan tujuan penggunaan model dalam memilih model yang paling sesuai untuk mengoptimalkan kinerja dalam tugas klasifikasi.

4. DISKUSI

Random Forest sudah banyak digunakan dalam berbagai bidang penelitian dan sudah banyak terbukti dapat melakukan prediksi dengan baik. Seperti dalam penelitian sebelumnya *Random Forest* mendapat akurasi terbaik dari algoritma lain dalam memprediksi persetujuan pinjaman dengan akurasi 80% [28]. *Random Forest* berhasil mendapatkan akurasi 82% dan lebih unggul dari algoritma *supervised learning* lainnya [29].

Dengan melakukan beberapa optimasi pada proses prapemrosesan data dan pelatihan model, penelitian ini dapat mengungguli penelitian sebelumnya. Algoritma *Random Forest* dapat mencapai akurasi 86.2% dengan melakukan seleksi fitur dan *balancing data* dengan SMOTE kemudian melakukan *hyperparameter tuning* pada pelatihan model. Penelitian ini dapat menjadi acuan dimasa depan, dengan menerapkan prapemrosesan lain seperti standarisasi, reduksi dimensi dan sebagainya atau dengan melakukan *hyperparameter tuning* dengan nilai parameter yang berbeda, dengan begitu diharapkan algoritma *Random Forest* mampu mendapatkan akurasi yang maksimal dalam prediksi pengajuan pinjaman.

5. KESIMPULAN

Dengan melakukan beberapa eksperimen terhadap *Loan Eligible Dataset* menghasilkan 4 dataset yang menjadi bahan penelitian. Masing-masing dataset tersebut berhasil dilatih dengan model *Random Forest* yang dioptimasi dengan *hyperparameter tuning* mendapatkan hasil yang baik, semua evaluasi hasil pelatihan mendapatkan akurasi diatas 80%, dan akurasi tertinggi didapatkan oleh dataset dengan penanganan *balancing data* dengan akurasi 86.2%. Dengan model *machine learning* yang sudah dilatih tersebut berharap mampu menjadi alat bantu pengambilan keputusan dalam persetujuan pengajuan pinjaman yang objektif dan tepat akurat sehingga lebih meminimalisir terjadinya resiko kredit macet.

DAFTAR PUSTAKA

- [1] Kasmir, *Manajemen Perbankan*. Jakarta: Rajawali Pers, 2017.
- [2] Kurniawan, D. Ahmad, and D. Kriestanto, "Penerapan Naïve Bayes untuk Prediksi Kelayakan Kredit," *J. Inform. dan Komput.*, vol. 1, no. 1, pp. 19–23, 2016.
- [3] A. Malik, "Pengaruh Loan To Deposit Ratio (LDR) Terhadap Profitabilitas Dengan Non Performing Loan (NPL) Sebagai Variabel Intervening Pada Subsektor Perbankan," *Sains Manaj. J. Manaj. Unsera*, vol. 6, no. 1, pp. 13–22, 2020.
- [4] S. P. Dewi, "Pengaruh Capital Adequacy Ratio, Non Performing Loan, Loan to Deposit Ratio dan Efisiensi Operasional Terhadap Profitabilitas Perbankan Yang Terdaftar Di Bursa Efek Indonesia," *J. Akunt.*, vol. 18, no. 3, pp. 422–437, 2014.
- [5] F. Sandria, "NIM Perbankan RI Super Tinggi, Kredit Macet Apa Kabar?," 2023. <https://www.cnbcindonesia.com/market/20230306131512-17-419214/nim-perbankan-ri-super-tinggi-kredit-macet-apa-kabar> (accessed Mar. 16, 2024).
- [6] CEIC DATA, "Indonesia Ratio Kredit Bermasalah," 2023. <https://www.ceicdata.com/id/indicator/indonesia/non-performing-loans-ratio> (accessed Mar. 16, 2024).
- [7] Mahmoeddin, *Melacak kredit bermasalah*. Jakarta: Pustaka Sinar Harapan, 2010.
- [8] E. Alpaydin, *Machine Learning*. Cambridge: MIT press, 2021.
- [9] M. A. Sheikh, A. K. Goel, and T. Kumar, "An approach for prediction of loan approval using machine learning algorithm," in *2020 International Conference on Electronics and Sustainable Communication Systems (ICESC)*, 2020.

- [10] A. Nata and S. Suparmadi, "Analisis Sistem Pendukung Keputusan Dengan Model Klasifikasi Berbasis Machine Learning Dalam Penentuan Penerima Program Indonesia Pintar," *J. Sci. Soc. Res.*, vol. 5, no. 3, 2021.
- [11] P. Tumuluru, L. R. Burra, M. Loukya, S. Bhavana, H. M. H. CSaiBaba, and N. Sunanda, "Comparative Analysis of Customer Loan Approval Prediction using Machine Learning Algorithms," in *2022 Second International Conference on Artificial Intelligence and Smart Energy (ICAIS)*, 2022.
- [12] L. Zhu, D. Qiu, D. Ergu, C. Ying, and K. Liu, "A study on predicting loan default based on the random forest algorithm," *Procedia Comput. Sci.*, vol. 162, pp. 503–513, 2019.
- [13] J. Wu, X.-Y. Chen, H. Zhang, L.-D. Xiong, H. Lei, and S.-H. Deng, "Hyperparameter optimization for machine learning models based on Bayesian optimization," *ournal Electron. Sci. Technol.*, vol. 17, no. 1, pp. 26–40, 2019.
- [14] D. De Silva and D. Alahakoon, "An artificial intelligence life cycle: From conception to production," *Patterns*, vol. 3, no. 6, 2022.
- [15] M. Solekhah, "PEMANFAATAN TEKNOLOGI ARTIFICIAL INTELLIGENCE CUBATBOT (CULTURE BALINESE CHATBOT) SEBAGAI INFORMASI KEBUDAYAAN BALI," *Kreat. Pada Pengabd. Masy.*, vol. 1, no. 2, 2023.
- [16] R. Chakma, J. Paul, and S. Dhir, "Organizational Ambidexterity: A Review and Research Agenda," *IEEE Trans. Eng. Manag. (Early Access)*, pp. 1–17, 2021, doi: 10.1109/TEM.2021.3114609.
- [17] V. Ukani, "Loan Eligible Dataset," 2020. <https://www.kaggle.com/datasets/vikasukani/loan-eligible-dataset> (accessed Mar. 21, 2024).
- [18] Suyebaanjum, "Mastering Null Value Handling: A Comprehensive Guide to Replacing Missing Data in Your Dataset," 2023. <https://medium.com/@suyebaanjum98/mastering-null-value-handling-a-comprehensive-guide-to-replacing-missing-data-in-your-dataset-1a0bf711e531> (accessed Apr. 30, 2024).
- [19] Iswanto, Tulus, and Poltak, "COMPARISON OF FEATURE SELECTION TO PERFORMANCE IMPROVEMENT OF K-NEAREST NEIGHBOR ALGORITHM IN DATA CLASSIFICATION," *J. Tek. Inform.*, vol. 3, no. 6, pp. 1709–1716, 2022.
- [20] M. H. Ariansyah, E. N. Fitri, and S. Winarno, "IMPROVING PERFORMANCE OF STUDENTS' GRADE CLASSIFICATION MODEL USES NAÏVE BAYES GAUSSIAN TUNING MODEL AND FEATURE SELECTION," *J. Tek. Inform.*, vol. 4, no. 3, pp. 493–501, 2023.
- [21] A. Altmann, L. Toloşi, O. Sander, and T. Lengauer, "Permutation importance: a corrected feature importance measure," *Bioinformatics*, vol. 26, no. 10, pp. 1340–1347, 2010.
- [22] N. V. Chawla, K. W. Bowyer, L. O. Hall, and W. P. Kegelmeyer, "SMOTE: Synthetic Minority Over-sampling Technique," *J. Artif. Intell. Res.*, vol. 16, pp. 321–357, 2002.
- [23] N. N. Sholihah and A. Hermawan, "IMPLEMENTATION OF RANDOM FOREST AND SMOTE METHODS FOR ECONOMIC STATUS CLASSIFICATION IN CIREBON CITY," *J. Tek. Inform.*, vol. 4, no. 6, pp. 1387–1397, 2023.
- [24] R. Faurina, M. J. Gazali, and I. D. A. Herani, "Optimization Of Disease Prediction Accuracy Through Artificial Neural Network (ANN) Algorithms In Diagnese Application," *J. Tek. Inform.*, vol. 5, no. 2, pp. 339–347, 2024.
- [25] S. Ambesange, R. Nadagoudar, R. Uppin, V. Patil, S. Patil, and S. Patil, "Liver Diseases Prediction using KNN with Hyper Parameter Tuning Techniques," in *2020 IEEE Bangalore Humanitarian Technology Conference (B-HTC)*, pp. 1–6, 2020.
- [26] A. S. Munir and R. Waluyo, "Optimasi Prediksi Kematian pada Gagal Jantung: Analisis Perbandingan Algoritma Pembelajaran Ensemble dan Teknik Penyeimbangan Data pada Dataset," *J. Sist. dan Teknol. Inf.*, vol. 12, no. 2, pp. 365–372, 2024.
- [27] F. P. Utama, R. Faurina, T. Mardiansyah, and A. Vatesia, "Scientific Articles Recommendation System Based On User's Relatedness Using Item-Based Collaborative Filtering Method," *J. Tek. Inform.*, vol. 4, no. 3, pp. 467–475, 2023.
- [28] Madaan, Mehul, A. Kumar, C. Keshri, R. Jain, and P. Nagrath, "Loan default prediction using decision trees and random forest: A comparative study," *IOP Conf. Ser. Mater. Sci. Eng.*, vol. 1022, no. 1, p. 012042, 2021.
- [29] L. U. Bhanu and S. Narayana, "Customer Loan Prediction Using Supervised Learning Technique," *Int. J. Sci. Res. Publ.*, vol. 11, no. 6, pp. 403–407, 2021.