

# Algoritma *Random Forest* pada Prediksi Status Kredit Usaha Rakyat untuk Mengurangi *Nonperforming Loan Rate*

**Wulansari<sup>1</sup> dan Diana Purwitasari<sup>2</sup>**

<sup>1</sup>Magister Manajemen Teknologi, Institut Teknologi Sepuluh Nopember, Surabaya, Indonesia

<sup>2</sup>Departemen Teknik Informatika, Fakultas Teknologi Elektro dan Informatika Cerdas, Institut Teknologi Sepuluh Nopember, Surabaya, Indonesia

**Corresponding author:** Wulansari (e-mail: wulansarirachmat27@gmail.com).

**ABSTRACT** Small and Micro credit program (KUR) is one of Indonesian government's programs in improving access to financing for small and micro segments, channeled through financial institutions with guarantees. KUR distribution funds originate 100% from KUR-distributing financial institutions. The government provides subsidies in the form of credit guarantees, whereby the government secures 70% of the total loan value and financial institutions, such as banks, assume the remaining 30%. Having a low non-performing loan (NPL) rate for KUR is crucial to averting the formation of loss-reserve expenses that will reduce bank profits. Predicting the variables that influence the borrowing impasse is essential for preventing an increase in NPLs, particularly in the early phases of lending. No research has been conducted to predict KUR default using machine learning. The purpose of this research Bank could predict through the use of machine learning and determine which variables contribute to KUR default. The SMOTE technique is also used to solve data imbalance. The results show that random forest classifiers outperform logistic regression with accuracy 88% respectively on data testing. Contributions of each variable that affects KUR default rank by feature importance are credit restructuring, availability of debt service payment account, unit operations, and business sector. The Bank may use this as a foundation to assess the risk profile of the new KUR procedure, hence mitigating the likelihood of problematic loans in the future and decreasing the NPL rate.

**KEYWORDS** KUR, Logistic Regression, NPL, *Random Forest*

**ABSTRAK** Program Kredit Usaha Rakyat (KUR) adalah salah satu program pemerintah Indonesia dalam meningkatkan akses pembiayaan kepada Usaha Mikro, Kecil, dan Menengah (UMKM) yang disalurkan melalui lembaga keuangan dengan pola penjaminan. Dana penyaluran KUR 100% berasal dari dana lembaga keuangan penyalur KUR. Pemerintah memberikan subsidi berupa pembayaran penjaminan kredit yang menjamin kredit sebesar 70% dari nilai total kredit yang ditetapkan dan 30% sisanya ditanggung lembaga keuangan salah satunya adalah bank. Mempertahankan *non performing loan* (NPL) *rate* yang rendah sangat penting untuk mencegah terjadinya pembentukan biaya pencadangan kerugian yang akan mengurangi keuntungan bank. Memprediksi variabel yang berpengaruh terhadap macetnya KUR diperlukan untuk mencegah kredit macet terutama pada tahap awal pemberian kredit. Belum terdapat penelitian untuk memprediksi kredit macet pada KUR menggunakan *machine learning*. Penelitian ini bertujuan agar Bank dapat melakukan prediksi dengan pendekatan *machine learning* dan mengetahui kontribusi variabel yang mempengaruhi KUR macet. Teknik SMOTE juga digunakan dalam penelitian untuk mengatasi ketidakseimbangan data. Hasil yang diperoleh menunjukkan bahwa klasifikasi *random forest* memberikan akurasi lebih tinggi dari regresi logistik yaitu sebesar 88% pada data uji. Urutan tingkat kepentingan dari kontribusi variabel yang mempengaruhi macet pada KUR adalah status restrukturisasi kredit, ketersediaan *debt service payment account*, unit operasional pemroses kredit, dan sektor usaha dapat menjadi dasar Bank untuk menganalisis profil resiko proses KUR baru sehingga dapat menurunkan potensi kredit bermasalah kedepannya dan menekan NPL *rate*.

**KATA KUNCI** KUR, NPL, *Random Forest*, Regresi Logistik

## I. PENDAHULUAN

Bank adalah badan usaha yang menghimpun dana dari masyarakat dalam bentuk simpanan dan menyalurkannya kepada masyarakat dalam bentuk kredit dan atau bentuk-bentuk lainnya dalam rangka meningkatkan taraf hidup rakyat banyak [1]. Kredit usaha rakyat (KUR) merupakan pemberian kepada UMKM yang disalurkan melalui lembaga keuangan dengan pola penjaminan. Dana penyaluran KUR 100% berasal dari dana lembaga keuangan penyalur KUR. Pemerintah memberikan subsidi berupa pembayaran penjaminan kredit yang menjamin kredit sebesar 70% dari nilai total kredit yang ditetapkan sementara 30% sisanya ditanggung oleh penyalur KUR dalam hal ini lembaga keuangan [2]. Berdasarkan hal tersebut dari sudut pandang lembaga keuangan, mempertahankan rasio kredit macet KUR yang rendah sangat penting untuk mencegah terjadinya kerugian akibat kredit macet yang akan mengurangi keuntungan perusahaan.

*Nonperforming loan (NPL) rate* adalah rasio yang dapat menunjukkan seberapa baik kualitas kredit yang disalurkan oleh perbankan, dan jika rasio NPL semakin rendah maka penyaluran kredit suatu bank dinilai semakin berkualitas. Laba suatu bank akan tergerus apabila terjadi peningkatan NPL rate, sejalan dengan teori manajemen likuiditas *Commercial Loan Theory* yang menyatakan penyaluran kredit harus memperhatikan tingkat pengembalian kredit agar kegiatan operasional bank terus berjalan dan tetap menjaga kepercayaan nasabah [3].

Terdapat beberapa penelitian terkait prediksi kredit macet dan faktor yang menyebabkan kredit menjadi macet dengan menggunakan metode *machine learning* meskipun belum ada untuk KUR. Beberapa penelitian tersebut yaitu dengan metode Regresi Logistik dan *Random Forest*. Tujuan dari penelitian tersebut melakukan *assessment* terhadap beberapa faktor untuk membantu bank dalam melakukan seleksi terhadap permohonan kredit [4]. Penelitian lainnya untuk *loan risk prediction* model dengan menggunakan metode *random forest* adalah penelitian [5] yang menilai signifikansi masing-masing variabel (faktor), dimungkinkan untuk menentukan fitur mana yang memiliki pengaruh lebih besar terhadap hasil akhir gagal bayar, sehingga dapat dilakukan keputusan resiko kegagalan bayar yang lebih bijaksana pada bank.

Algoritma klasifikasi berguna untuk memisahkan dan mengelompokkan data untuk melakukan prediksi sebagaimana akan digunakan pada *loan prediction* [6]. Pada kasus kredit macet akan terjadi data *imbalanced* karena perbandingan yang nilainya signifikan antara klasifikasi satu kelas dengan kelas lainnya (macet/tidak macet). Salah satu metode yang digunakan untuk permasalahan *imbalanced* tersebut adalah SMOTE, teknik *oversampling* yang menghasilkan data baru untuk menyerupai data asli dari kelas yang lebih kecil sehingga dapat menciptakan keseimbangan data dan terhindar dari hasil akurasi yang bias dan menyebabkan ketidakakuratan prediksi [7].

Belum terdapat penelitian untuk memprediksi kredit macet pada KUR menggunakan *machine learning*. Penelitian ini bertujuan agar Bank dapat melakukan prediksi dengan pendekatan *machine learning* dan mengetahui kontribusi variabel yang mempengaruhi KUR macet dengan *feature importance* sehingga dapat menjadi mitigasi saat analisis permohonan kredit yang dapat mengurangi NPL rate KUR.

Dalam penelitian ini diusulkan menggunakan dua algoritma *machine learning* sebagai komparasi untuk akurasi terbaik dari model yang dibangun, metode tersebut dikombinasikan dengan SMOTE untuk mengatasi masalah data kelas yang tidak seimbang.

## II. TINJAUAN PUSTAKA

Mempertahankan NPL rate KUR yang rendah sangat penting untuk mencegah terjadinya kerugian akibat kredit macet yang akan mengurangi keuntungan perusahaan atas biaya pencadangan kerugian.

Penelitian yang telah dilakukan untuk prediksi kredit macet [8] dari data German *credit* dataset dengan jumlah data sebanyak 1.000 dengan metode *Naïve Bayes*, MLP, SMO, *K-nearest neighbour's*, *Rotation Forest*, *Bagging*, *RealADABoost*, dan *Random Forest* dengan dan tanpa SMOTE. Hasil yang diperoleh adalah *random forest* berkinerja terbaik dengan nilai akurasi sebelum SMOTE sebesar 75,3% dan nilai akurasi setelah SMOTE meningkat menjadi sebesar 83,4%. Hasil dari urutan *feature* 5 variabel yang paling berkontribusi terhadap kredit macet adalah *Over draft*, *Credit History*, *Purpose*, *Average Credit Balance* dan *Employment*.

Penelitian lainnya dari [9] dengan menggunakan metode SVM, KNN, C4.5, dan *Random Forest* dengan SMOTE untuk dataset kredit dari Kaggle dengan jumlah 150.000 diperoleh hasil *random forest* dengan SMOTE berkinerja terbaik dibandingkan metode lainnya dengan nilai akurasi sebesar 93,7%.

Metode Bayesian juga digunakan dalam penelitian [10] dengan kesimpulan mode *inception games* dikembangkan dimana kreditur sebagai pemeriksa dan debitur sebagai yang diperiksa dan variabel penyebab NPL dari sisi debitur adalah utilitas kredit atau penggunaan kredit. Sepuluh literatur yang dikumpulkan dalam penelitian [11] dengan menggunakan metode SVM, regresi logistik, *decision tree*, dan *neural network* disimpulkan regresi logistik dan *decision tree* dimilai efektif dalam memprediksi gagal bayar pinjaman dan mengelola resiko kredit. Variabel yang berkontribusi terhadap kredit macet berdasarkan penelitian ini adalah usia, pendapatan, angsuran kredit, dan pendidikan.

Perbandingan metode *random forest* dan metode XGBoosting dilakukan dalam penelitian [12] dengan sumber data dari *lending club website* sebanyak 272.401. Penelitian ini menyimpulkan metode *random forest* berkinerja terbaik berdasarkan evaluasi AUC, dengan nilai AUC sebesar 0,99 lebih tinggi jika dibandingkan dengan XGBoosting classifier sebesar 0,97. Penelitian dari [13] dari German *credit* dataset

dengan jumlah data sebanyak 1.000 dan jumlah variabel yang digunakan 21 variabel, diperoleh hasil nilai akurasi model dengan menggunakan metode *random forest* sebesar 83% termasuk kedalam klasifikasi “*very good model*”.

Untuk penelitian dengan penekanan pada skenario data latih dan data uji dilakukan pada [14] dengan menggunakan tiga skenario perbandingan yaitu:

1. Data latih 60%: data uji 40%
2. Data latih 70%: data uji 30%, dan
3. Data latih 80%: data uji 20%.

Metode *machine learning* yang digunakan adalah metode *random forest* dengan sumber data dari data kredit BRI cabang Pelaihari. Kesimpulan dari penelitian ini skenario terbaik yaitu data latih berbanding data uji sebesar 70%:30% dan 80%:20% dengan nilai akurasi sama yaitu sebesar 97,1%, dan nilai *precision* sebesar 97,02% dan *recall* sebesar 100%.

Pada saat menafsirkan model, dalam penelitian [15] digunakan metode *feature importance* untuk menghitung peringkat kepentingan dan menyimpulkan kontribusi dari masing-masing variabel input. Selain itu dapat digunakan untuk mengukur kekuatan pengaruh dari suatu variabel prediktor terhadap variabel respon. Hal tersebut diperlukan untuk mengurutkan kontribusi terhadap kredit macet yang akan digunakan pada penelitian ini.

### III. METODOLOGI

Tujuan dari penelitian ini adalah membangun model klasifikasi untuk memprediksi status KUR (lancar/macet) dengan menggunakan metode *Random Forest* dengan SMOTE dan mengukur serta menganalisis kontribusi setiap variabel yang mempengaruhi kredit macet pada KUR untuk mengurangi NPL rate. Metode klasifikasi lainnya yaitu regresi logistik digunakan untuk membangun model sebagai komparasi.

#### A. DATASET

Data pada penelitian ini merupakan data sekunder yang diperoleh dari Bank XYZ regional Surabaya. Data yang digunakan adalah data pinjaman aktif KUR dari Januari 2016 sampai dengan posisi 31 Desember 2022.

Definisi untuk kredit macet dalam penelitian ini adalah kredit yang memiliki tunggakan dengan hari tunggakan telah melampaui 90 hari. Dataset KUR yang diperoleh dengan total data pinjaman sebanyak 45.938 dengan variabel sebanyak 17. NPL rate KUR Bank XYZ pada posisi 31 Desember 2022 sebesar 1,0% dengan total debitur macet sebanyak 427. Terdapat 17 variabel pada data KUR yang dikumpulkan yaitu 16 variabel prediktor dan 1 variabel respon.

Dari 16 variabel prediktor awal pada tabel 1, akan dilakukan *exclude* untuk variabel TglBukaRek yang merupakan tanggal menjadi nasabah dan Tgl\_PK yang merupakan tanggal menjadi debitur, disebabkan variabel tersebut tidak dapat dilakukan “*adjustment*” di masa mendatang untuk perbaikan proses kredit sebagai masukan kepada perusahaan.

**TABEL I**  
**VARIABEL PENELITIAN**

Variabel	Nama	Keterangan	Skala
Y	Kol	Status kelancaran Kredit yaitu Lancar/Macet	Nominal
X1	Unit_Processing	Cabang Pemroses Kredit, terdapat 13 cabang pemroses	Nominal
X2	Produk	Pembayaran kewajiban secara Musiman/ Bulanan	Nominal
X3	Peruntukan	KUR Mikro: maksimum kredit sd Rp100Juta KUR Ritel: maksimum kredit sd Rp500 Juta	Nominal
X4	Maks_Krd	Maksimum kredit	Rasio
X5	Bk_Dbt	Nilai sisa pokok kredit	Rasio
X6	Suku_Bunga	Nilai suku bunga	Interval
X7	JW	Jangka waktu kredit 1- 60 bulan	Interval
X8	Jaminan	Ketersediaan Jaminan: Ada/Tidak ada jaminan	Nominal
X9	Umur_Tgk	Jumlah hari menunggak kewajiban	Interval
X10	Sektor_Ek	Sektor Ekonomi usaha debitur, Terdapat 9 sektor ekonomi usaha	Nominal
X11	Nilai_Angsuran	Kewajiban per bulan debitur	Rasio
X12	DSPA	<i>Debt service reserve account</i> yaitu ketersediaan saldo dana 1x kewajiban Tersedia/Tidak tersedia	Nominal
X13	TglBukaRek	Tanggal pembukaan rekening dana	Nominal
X14	Tgl_PK	Tanggal menjadi debitur sd 2019, 2020, 2021, 2022	Nominal
X15	R3	Restrukturisasi kredit Restruk/Non Restruk	Nominal
X16	Tipe_Debitur	Perorangan atau Non Perorangan	Nominal

#### B. ALUR PENELITIAN

Berdasarkan gambar 1 alur penelitian dimulai dengan identifikasi masalah, tinjauan literatur komprehensif dilakukan untuk mencari metodologi dan pendekatan yang ada yang dapat menyelesaikan masalah secara efektif.

Tahap selanjutnya sebelum analisis dimulai, data diproses terlebih dahulu untuk mengidentifikasi dan memperbaiki nilai yang hilang dan data outlier dihilangkan untuk memenuhi persyaratan penelitian. Tahap ini sangat penting untuk memastikan bahwa dataset dibersihkan secara menyeluruh dan dipersiapkan untuk analisis lebih lanjut dan pembuatan model prediksi. Untuk memahami karakteristik data, dilakukan *exploratory data analysis* (EDA) sehingga dapat diketahui karakteristik dari data set, fitur apa yang membuat data tersebut berbeda atau sama, bagaimana cara terbaik dalam menjelaskan data, karakteristik terpenting yang diperlukan. Kemudian dilakukan penghilangan data *outlier* yang tidak sesuai dengan karakteristik data serta pengkodean variabel (*encoding*) untuk dilakukan proses pengolahan data lebih

lanjut. Dari hasil pemeriksaan tidak ditemukan nilai *null* sehingga dapat dilanjutkan pada tahapan selanjutnya yaitu pemeriksaan *outlier* yaitu apakah terdapat maksimum KUR diatas Rp.500 Juta. Hasil yang diperoleh pada data tidak ditemukan adanya nilai *outlier*.

Hal yang selanjutnya dilakukan adalah penerapan *one hot encoding* terhadap data yang bersifat kategorikal yaitu unit *processing*, produk, peruntukan, ketersediaan jaminan, sektor ekonomi, *debt service payment account*, tahun menjadi debitur, status restrukturisasi kredit, dan tipe debitur. Dilakukan pembagian data latih dan data uji dengan perbandingan yaitu 80%:20%. Data latih untuk menghasilkan model klasifikasi untuk kebutuhan prediksi, dan data uji untuk melakukan pengujian terhadap model klasifikasi.

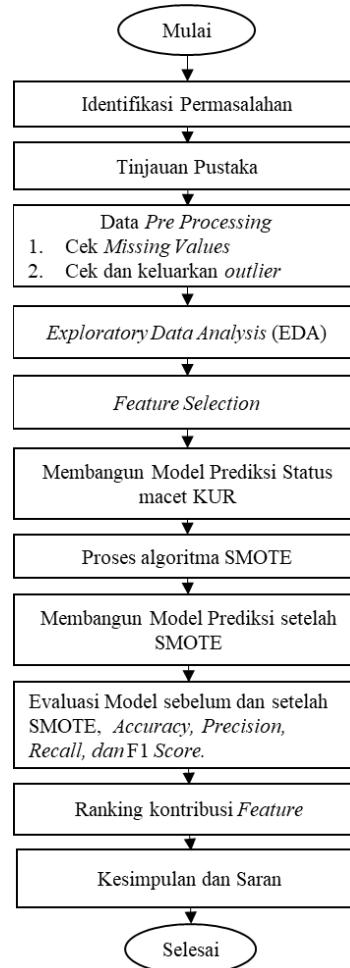
*Feature selection*, dilakukan untuk memilih fitur yang relevan dan menghilangkan atribut yang tidak perlu. *Feature selection* dilakukan dengan metode filter yaitu penerapan Lasso (*Least Absolute Shrinkage and Selection Operator*) dan menghasilkan 7 *feature* yaitu Unit\_Processing\_JEMBATAN MERAH, Unit\_Processing\_MADURA, Unit\_Processing\_GRAHA\_PANGERAN, Sektor\_Ek\_Industri Pengolahan, Sektor\_Ek\_Pertambangan, DSPA\_TERSEDIA, dan R3\_RESTRIK.

Selanjutnya dilakukan pemodelan dengan menggunakan metode regresi logistik dan *random forest*. Metode tersebut dipilih karena merupakan model klasifikasi yang dinilai memiliki akurasi yang tinggi sesuai dengan data yang diperoleh dari penelitian terdahulu yang dijadikan pedoman dalam penelitian ini.

Model selanjutnya dievaluasi berdasarkan *confusion matrix* dengan beberapa parameter yaitu *accuracy*, *precision*, *recall*, dan *F1 score*. *Accuracy* mewakili proporsi data yang diklasifikasikan dengan benar dari jumlah total data, *Precision* menggambarkan tingkat akurasi prediksi kelas positif (*default*) sebagai kelas positif (*default*). *Recall*, menunjukkan seberapa akurat model dapat memprediksi kelas positif dan disebut juga sensitivitas. *F1 score* adalah metrik yang mempertimbangkan *precision* dan *recall*. *F1 score* yang lebih tinggi menunjukkan *precision* dan *recall* yang lebih kuat, sehingga memperkuat keunggulannya dibandingkan pengukuran *accuracy*.

Untuk mencegah ketidakseimbangan dataset dan meningkatkan akurasi model klasifikasi, teknik SMOTE dilakukan untuk membuat dataset seimbang. SMOTE dilakukan pada data *training*, selanjutnya model yang dibangun sebelum dan sesudah SMOTE dievaluasi, dan model dengan nilai *recall* tertinggi dipilih. Perlu dilakukan SMOTE pada penelitian ini karena dataset tidak seimbang yaitu kelas macet sebanyak 427 dan kelas tidak macet sebanyak 45.511.

Model ini dibedah lebih lanjut untuk menentukan *feature importance* setiap fitur. Karakteristik signifikan tersebut kemudian digunakan untuk memberikan rekomendasi bagi perusahaan dalam memitigasi risiko saat menganalisis usulan KUR guna mencegah kredit macet pada KUR.



GAMBAR 1. Diagram Alir Penelitian.

#### IV. HASIL DAN PEMBAHASAN

Bagian ini menjelaskan secara rinci hasil penelitian, termasuk data dan temuan penting, dan menjawab pertanyaan dalam perumusan masalah terkait penelitian.

##### A. MODEL PREDIKSI SEBELUM DILAKUKAN SMOTE

Untuk mengetahui efektivitas dan kinerja model prediktif perlu dilakukan evaluasi atas model tersebut dengan menggunakan berbagai metrik atau estimator. *Precision* menunjukkan akurasi prediksi kasus positif dibandingkan dengan total prediksi kelas positif, *recall* menunjukkan akurasi prediksi kasus positif dibandingkan dengan jumlah total kelas positif terjadi, *accuracy* menunjukkan tingkat akurasi kedua kelas diprediksi dibandingkan dengan semua catatan, akan menjadi ukuran yang baik ketika kumpulan data seimbang, *F1-score* yaitu mean harmonik dari *precision* dan *recall*. Performa model prediksi, seperti terlihat pada tabel 2 model klasifikasi *random forest* menunjukkan tingkat kinerja yang lebih tinggi dibandingkan model regresi logistik.

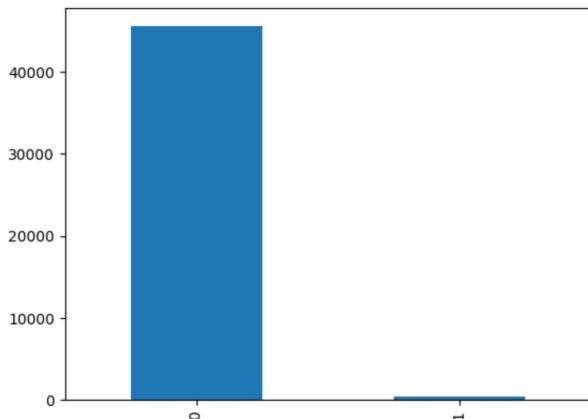
**TABEL II**  
EVALUASI MODEL PREDIKSI SEBELUM SMOTE

Metrik	Regresi Logistik	Random Forest
Accuracy	0.99	0.99
Precision	0.00	0.60
Recall	0.00	0.07
F1-score	0.00	0.13

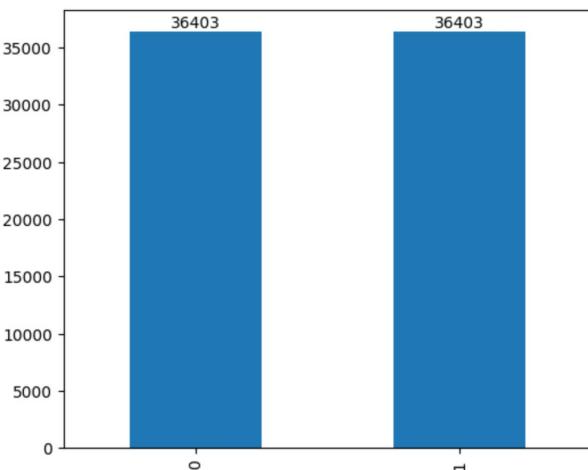
### B. MODEL PREDIKSI SETELAH DILAKUKAN SMOTE

Untuk mencegah ketidakseimbangan dataset dan meningkatkan akurasi model klasifikasi, teknik SMOTE digunakan dalam model prediksi untuk membuat dataset seimbang dengan menambahkan data ke kelas minor.

Gambar 2 dan gambar 3 menggambarkan variabel respon sebelum dan sesudah dilakukan SMOTE. Pada saat dilakukan kembali model prediksi hasil yang diharapkan tidak bias karena perbedaan antar kelas klasifikasi yang terlalu jauh.



**GAMBAR 2.** KUR Macet Sebelum SMOTE.



**GAMBAR 3.** KUR Macet Setelah SMOTE.

Setelah melakukan algoritma SMOTE, model klasifikasi dibuat ulang menggunakan regresi logistik dan *random forest*, sehingga menghasilkan *confusion matrix* berikutnya

dengan *accuracy*, *precision*, *recall*, dan *F1 score* baru yang ditunjukkan pada tabel 3.

**TABEL III**  
EVALUASI MODEL PREDIKSI SETELAH SMOTE

Metrik	Regresi Logistik	Random Forest
Accuracy	0.80	0.88
Precision	0.03	0.06
Recall	0.75	0.86
F1-score	0.06	0.11

Meskipun terjadi penurunan *accuracy*, *precision*, *recall*, dan *F1 score* setelah dilakukan SMOTE, perlu dilihat bahwa algoritma *random forest* terus menunjukkan kinerja yang lebih unggul dibandingkan dengan regresi logistik. Sehingga untuk model prediksi yang dipilih adalah model dari *random forest*. Selanjutnya dilakukan evaluasi untuk model dengan metode *random forest* yaitu analisis perbandingan model *random forest* sebelum dan sesudah penerapan teknik SMOTE pada kelas 1 (macet).

**TABEL IV**  
EVALUASI RANDOM FOREST: MODEL PREDIKSI SEBELUM DAN SETELAH SMOTE UNTUK KELAS 1 (MACET)

Metrik	Random Forest sebelum SMOTE	Random Forest setelah SMOTE
Accuracy	0.99	0.88
Precision	0.60	0.06
Recall	0.07	0.86
F1-score	0.13	0.11

Berdasarkan tabel 4, terlihat bahwa penggunaan SMOTE menyebabkan penurunan nilai *accuracy*, *precision*, dan *F1 score*, sedangkan nilai *recall* mengalami peningkatan. *Recall* atau sensitivitas mengukur kapasitas model untuk secara efektif mengurangi kejadian *false negative*, yang mengacu pada prediksi macet pada KUR yang benar-benar macet namun secara keliru diklasifikasikan sebagai tidak macet oleh model atau diprediksi sebagai kredit yang baik. Hal ini penting bagi Bank XYZ karena adanya peningkatan biaya pencadangan kerugian yang akan mengurangi keuntungan bank.

Penurunan akurasi pada model setelah dilakukan SMOTE, karena pada data yang tidak seimbang memungkinkan model akan memilih prediksi pada kelas mayoritas dimana dalam kasus kredit macet, kelas mayoritas adalah kredit lancar. Nilai *recall* memiliki peningkatan yang signifikan setelah dilakukan SMOTE sehingga model prediksi yang dipilih adalah *random forest* setelah SMOTE.

### C. FEATURE IMPORTANCE

Melakukan interpretasi klasifikasi dalam penelitian sebelumnya dengan menggunakan metode *feature importance* untuk menghitung peringkat kontribusi dari masing-masing variabel prediktor terhadap variabel respon.

**TABEL V**  
**FEATURE IMPORTANCE DARI VARIABEL PREDIKSI**

<i>Feature</i>	<i>Importance</i>
0 R3 RESTRIK	0,36
1 DSPA TERSEDIA	0,24
2 Unit Processing MADURA	0,21
3 Sektor_Ek_Industri Pengolahan	0,15

Fitur yang berulang kali menunjukkan kecenderungan untuk mengurangi ketidakmurnian dan meningkatkan performa model akan diberi skor relevansi yang lebih tinggi, sehingga menandakan signifikansinya dalam menghasilkan prediksi yang akurat. Tabel 5 menampilkan kontribusi dari 4 karakteristik teratas yang memiliki pengaruh paling besar, dan mewakili 96% dari keseluruhan fitur. Fitur yang paling berpengaruh terhadap gagal bayar/macet KUR adalah restrukturisasi kredit, ketersediaan *debt service payment account*, unit pemroses kredit dari Madura, dan sektor usaha dari sektor industri pengolahan.

## V. KESIMPULAN

Setelah menggunakan teknik SMOTE, algoritma *random forest* konsisten menunjukkan kinerja yang lebih unggul dibandingkan dengan regresi logistik. Pengujian data performa untuk memprediksi kelas 1 (macet) nilai *recall* meningkat setelah SMOTE yang menunjukkan kapasitas model secara efektif mengurangi kejadian *false negative*. Hasil dari model terpilih dapat menjadi input untuk Bank XYZ sebagai pertimbangan dalam memprediksi kredit macet pada KUR.

Fitur dengan kontribusi tertinggi berdasarkan hasil *feature important* yang paling besar pengaruhnya terhadap macet pada KUR adalah status restrukturisasi kredit, ketersediaan *debt service payment account*, unit pemroses kredit dari Madura, dan sektor usaha dari sektor industri pengolahan dapat dijadikan pertimbangan oleh Bank XYZ untuk mengkaji bobot risiko saat ini.

Untuk mengurangi NPL *rate* dalam ekspansi KUR baru yang berkualitas tindakan mitigasi yang dapat dilakukan oleh Bank XYZ adalah pemberian bobot resiko yang lebih tinggi dalam hal aplikasi permohonan KUR yang memiliki historikal restrukturisasi kredit pada bank lain, sektor ekonomi industri pengolahan, dan unit proses Madura.

Untuk penelitian kedepannya, data secara nasional dapat ditambahkan sehingga dapat meningkatkan akurasi prediksi kredit macet pada Bank XYZ.

## PERAN PENULIS

**Wulansari:** Analisis Formal, Pengumpulan dataset, Metodologi, Administrasi Proyek, Perangkat Lunak, Validasi, Visualisasi, Penulisan Penyusunan Draf Asli.

**Diana Purwitasari:** Penulisan Review & Penyuntingan;

## COPYRIGHT



This work is licensed under a Creative Commons Attribution-NonCommercial-ShareAlike 4.0 International License.

## DAFTAR PUSTAKA

- [1] Bank Indonesia, "Undang-Undang Republik Indonesia Nomor 10 Tahun 1998 Tentang Perbankan." Jakarta, 1998.
- [2] Menteri Koordinator Bidang Perekonomian, "Peraturan Menteri Koordinator Bidang Perekonomian Nomor 1 Tahun 2022 tentang Pedoman Pelaksanaan Kredit Usaha Rakyat." Jakarta, 2022.
- [3] S. B. Haryanto and E. T. Widarty, "Analisis Pengaruh NIM, NPL, BOPO, BI Rate dan CAR terhadap Penyaluran Kredit Bank Umum Go Public Periode Tahun 2012-2016," vol. 6, no. Nomor 4, pp. 1–11, 2017.
- [4] M. Madaan, 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, Jan. 2021, doi: 10.1088/1757-899X/1022/1/012042.
- [5] Q. Zhang, "Loan Risk Prediction Model based on Random Forest," *Advances in Economics, Management and Political Sciences*, vol. 5, no. 1, pp. 216–222, Apr. 2023, doi: 10.54254/2754-1169/5/20220082.
- [6] 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.
- [7] 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, doi: 10.1016/j.procs.2019.12.017.
- [8] M. K. Pandey, M. Mittal, and K. Subbiah, "Optimal balancing & efficient feature ranking approach to minimize credit risk," *International Journal of Information Management Data Insights*, vol. 1, no. 2, Nov. 2021, doi: 10.1016/j.jjimei.2021.100037.
- [9] L. Zhou and H. Wang, "Loan Default Prediction on Large Imbalanced Data Using Random Forests," *TELKOMNIKA Indonesian Journal of Electrical Engineering*, vol. 10, no. 6, Sep. 2012, doi: 10.11591/Telkomnika.v10i6.1323.
- [10] E. Widodo, O. A. Rochmadhan, Lukmandono, and Januardi, "Modeling Bayesian inspection game for non-performing loan problems," *Operations Research Perspectives*, vol. 9, p. 100218, 2022, doi: 10.1016/j.orp.2021.100218.
- [11] U. Aslam, H. I. T. Aziz, A. Sohail, and N. K. Batcha, "An empirical study on loan default prediction models," *J Comput Theor Nanosci*, vol. 16, no. 8, pp. 3483–3488, 2019, doi: 10.1166/jctn.2019.8312.
- [12] A. K. Sharma, L.-H. Li, and R. Ahmad, "Default Risk Prediction Using Random Forest and XGBoosting Classifier," 2023, pp. 91–101, doi: 10.1007/978-3-031-05491-4\_10.
- [13] B. Prasojo and E. Haryatmi, "Analisa Prediksi Kelayakan Pemberian Kredit Pinjaman dengan Metode Random Forest," *Jurnal Nasional Teknologi dan Sistem Informasi*, vol. 7, no. 2, pp. 79–89, Sep. 2021, doi: 10.25077/TEKNOSI.v7i2.2021.79-89.
- [14] M. I. Putra, A. Yusuf, and N. Yalina, "Klasifikasi Kelancaran Kredit Dengan Metode Random Forest," *Systemic: Information System and Informatics Journal*, vol. 5, no. 2, pp. 7–12, Mar. 2020, doi: 10.29080/Systemic.v5i2.713.
- [15] G. K. Rajbahadur, S. Wang, Y. Kamei, and A. E. Hassan, "The impact of feature importance methods on the interpretation of defect classifiers," Feb. 2022, doi: 10.1109/TSE.2021.3056941.