
Pemilihan Metode *Predictive Analytics* dengan *Machine Learning* untuk Analisis dan Strategi Peningkatan Kualitas Kredit Perbankan

Aznovri Kurniawan*, Ahmad Rifa'i, Moch Abdillah Nafis, Nimas Sefrida, Harry Patria
Program Studi Magister Manajemen Teknologi, Sekolah Interdisiplin Manajemen dan Teknologi,
Institut Teknologi Sepuluh Nopember, Surabaya, Indonesia

*Corresponding author: 6032202012@mhs.its.ac.id

Abstract. *As a factor that determines bank's profitability, loan quality, that is categorized based on debtor's collectability classification, always gets attention, and become main analysis topic in banking industry. Through recent development of statistics and data science, especially in predictive analytics using machine learning techniques, more comprehensive analysis and prediction in loan quality can be conducted. This research is intended to give example on application of predictive analytics using machine learning technique for analysis and strategy recommendation in increasing bank's loan quality improvement. In this research, some machine learning classification methods are compared to conduct predictive analytics in loan quality with big data size (big data analytics). Computation result of different methods are compared and summarized, resulted in recommendation on most appropriate method to achieve this research objective. This research concluded that for imbalanced big data size such as bank's loan collectability, Tree Ensemble method, further development of Decision Tree method that is commonly used in machine learning, is one of appropriate methods to get satisfactory result in this research. Imbalanced data that can result in false positive may be overcome by oversampling Synthetic Minority Oversampling Technique (SMOTE). This research scope is limited to analysis and prediction of debtor's collectability for the next several months, combined with analysis and strategy recommendations based on product type, gender, and debtor's occupation. Further predictive analytics for the next several years by including external factors, such as economic growth, is not covered in this research and possible to be conducted. As machine learning application in Indonesian banking industry analysis is still in early phase, this research is expected to become one of reference in application of predictive analytics using machine learning in banking industry.*

Keywords: *loan collectability; loan quality; machine learning; predictive analytics*

1. PENDAHULUAN

Kualitas kredit atau pembiayaan dalam bidang perbankan merupakan penilaian atas kelancaran pembayaran debitur berdasarkan ketepatan waktu pembayaran yang telah diperjanjikan antara bank dan debitur. Kualitas kredit merupakan salah satu faktor yang sangat menentukan profitabilitas bank, sehingga selalu menjadi perhatian dan topik analisis utama di dunia perbankan. Menurut Peraturan Otoritas Jasa Keuangan (POJK) Nomor 40/POJK.03/2019 tentang Penilaian Kualitas Aset Bank Umum [1], bank wajib mengelola aset kredit tersebut berdasarkan prinsip kehati-hatian, dengan menilai, memantau, dan mengambil langkah yang diperlukan untuk menjaga aset bank tetap dalam kualitas baik. Berdasarkan POJK tersebut, kualitas kredit ditetapkan menjadi 5 (lima) kategori kolektibilitas, mulai dari kolektibilitas 1 yang

paling lancar sampai kolektibilitas 5 yang paling tidak lancar, yang dapat dilihat pada Tabel 1. Gabungan dari kolektibilitas 3, 4, dan 5 disebut sebagai *Non-Performing Loan* (NPL) pada perbankan umum atau *Non-Performing Financing* (NPF) pada perbankan Syariah.

Tabel 1. Kategori kolektibilitas berdasarkan POJK

Kolektibilitas	Keterangan
1	Lancar (debitur selalu membayar sesuai angsuran secara tepat waktu)
2	Dalam Perhatian Khusus (debitur menunggak pembayaran 1-90 hari)
3	Kurang Lancar (debitur menunggak pembayaran 91-120 hari)
4	Diragukan (debitur menunggak pembayaran 121-180 hari)
5	Macet (debitur menunggak pembayaran lebih dari 180 hari)

Salah satu kesulitan yang dihadapi dalam analisis di dunia perbankan, termasuk dalam analisis kualitas kredit bank adalah jumlah data yang sangat besar, yang dapat mencapai jutaan baris dan puluhan kolom, mengingat jumlah debitur bank yang dapat mencapai jutaan, tergantung dari besar kecilnya basis nasabah (*customer base*) bank tersebut. Kesulitan lainnya adalah kebutuhan akan *predictive analytics*, yaitu analisis data historis untuk melakukan prediksi kejadian di masa depan. Dengan perkembangan bidang statistika, *data science*, dan *big data analytics* saat ini, khususnya *predictive analytics* dengan menggunakan teknik komputasi *machine learning*, analisis yang lebih komprehensif terhadap kualitas kredit perbankan dapat dilakukan.

Tujuan dari penelitian ini adalah untuk memberikan contoh aplikasi *predictive analytics* dengan *machine learning* untuk memecahkan permasalahan klasifikasi, sehingga dihasilkan analisis dan rekomendasi strategi bagi peningkatan kualitas kredit perbankan. Metode klasifikasi *predictive analytics* dengan *machine learning* yang digunakan diharapkan dapat melakukan prediksi kecenderungan penurunan kualitas kredit (kolektibilitas) berdasarkan historis pembayaran dan mendapat rekomendasi untuk periode selanjutnya. Selain itu, prediksi kelancaran pembayaran tersebut juga dapat digunakan untuk melakukan analisis kecenderungan kualitas kredit berdasarkan jenis produk, jenis kelamin, pekerjaan debitur, dan faktor lainnya.

Berdasarkan studi kasus yang diambil dari data salah satu bank nasional yang dilakukan pada penelitian ini, setelah dilakukan perbandingan berbagai metode klasifikasi *machine learning*, maka direkomendasikan metode terbaik untuk pemecahan masalah. Pada akhirnya, penelitian ini diharapkan menghasilkan contoh rekomendasi strategi bagi peningkatan kualitas kredit perbankan.

2. KAJIAN PUSTAKA

Menurut Tsai *et al.* [2] dan Elgendy *et al.* [3], dengan perkembangan teknologi digital saat ini dan berkembangnya kualitas dan kuantitas data, kebutuhan akan perubahan metode analitika data (*data analytics*) semakin berkembang, ditandai dengan pembahasan baru tentang *big data analytics*. Menurut IBM [4], *big data* ditandai dengan karakteristik 4V, yaitu *volume* (ukuran data), *variety* (keberagaman bentuk data), *velocity* (kecepatan penyebaran data), dan *veracity* (ketidakpastian data). Data di bidang perbankan, seperti data kualitas kredit sebuah bank, termasuk ke dalam kategori *big data* karena memenuhi semua karakteristik di atas. *Big data analytics* berdasarkan tujuannya mencakup empat jenis. Pertama, *descriptive analytics*, yaitu pengumpulan dan penyampaian informasi dan *insights* berdasarkan data historis, sehingga dapat

menjadi basis bagi pengambilan keputusan atau strategi bisnis lainnya. Kedua, *diagnostic analytics*, yaitu penggunaan data historis untuk menjelaskan penyebab dari suatu kejadian dari *descriptive analytics*. Ketiga, *predictive analytics*, yaitu penggunaan data historis dari *descriptive analytics* untuk identifikasi informasi dan kecenderungan tertentu, untuk kemudian dilakukan prediksi kejadian di masa depan. Keempat, *prescriptive analytics*, yaitu penggunaan hasil dari *predictive analytics* untuk memberikan rekomendasi tindakan yang harus dilakukan di masa depan.

Machine learning adalah teknik komputasi yang umum digunakan dalam melakukan *predictive analytics* untuk data dalam jumlah besar (*big data*), dengan menggunakan algoritma yang dijalankan oleh komputer. Algoritma *machine learning* membangun model berdasarkan data sampel (biasa disebut sebagai *training data*) untuk melakukan prediksi atau kejadian pada data yang dilakukan analisis (biasa disebut sebagai *test data*). Salah satu teknik *machine learning* dalam permasalahan klasifikasi dan regresi adalah *Decision Tree*, yang dengan perkembangan *data science* saat ini telah memunculkan berbagai turunan dengan kompleksitas lebih tinggi, seperti dibahas oleh Sartono *et al.* [5], yaitu *Random Forest*, *Gradient Boosting*, dan *Tree Ensemble*.

Metode *Decision Tree* atau pohon keputusan adalah metode yang umum digunakan pada *operations research* dan *operations management*, dan digunakan secara umum untuk *machine learning*. Metode ini memiliki kelebihan akan kemudahan untuk dipahami, akan tetapi hasilnya relatif kurang akurat dibandingkan metode lain, sehingga dilakukan pengembangan untuk menghasilkan prediksi yang akurat, di antaranya metode *Random Forest*, *Gradient Boosting*, dan *Tree Ensemble*.

3. METODE PENELITIAN

3.1. Metode Pengambilan dan Persiapan Data

Data yang digunakan dalam penelitian ini merupakan data sekunder yang berupa data histori pembiayaan kredit pada sebuah bank nasional di Indonesia selama bulan Maret hingga Agustus 2021. Ukuran data sebesar 1.039.161 baris dan 18 kolom. Tabel 2 menunjukkan variabel yang digunakan dalam penelitian. Sebagian data tersebut memiliki nilai data yang kosong, sehingga perlu dilakukan tahapan *data preprocessing*. Pertama, menghapus baris dari data kredit/pembiayaan yang sudah lunas. Kedua, mengganti data yang kosong dengan nilai yang seharusnya, seperti DPD (*days past due*) yang kosong dengan 0, Restruktur yang kosong dengan "Tidak Restruktur". Ketiga, mengganti data yang tidak diketahui dengan "Others", seperti Jenis Kelamin, Pekerjaan, Produk. Data tersebut tidak dihapus karena memuat informasi lain yang bermanfaat untuk analisis. Setelah dilakukan tahapan data preprocessing tersebut, dihasilkan data yang sudah siap untuk pemrosesan dengan ukuran 900.331 baris dan 18 kolom, dengan kolom yang sama dengan yang disampaikan pada Tabel 2.

Satu hal yang menjadi perhatian dalam penelitian ini adalah data kolektibilitas yang digunakan pada penelitian ini memiliki tingkat *data imbalance* yang tinggi, yaitu lebih dari 97,5% baris memiliki kolektibilitas 1 dan 2, sedangkan kurang dari 2,5% memiliki kolektibilitas 3-5 atau NPL (*Non-Performing Loan*)/ NPF (*Non-Performing Financing*), dimana kondisi ini adalah kondisi ideal dan umum terjadi pada bank dengan kinerja baik yang mengharuskan bank untuk menekan angka NPL/NPF serendah mungkin. Metode *Decision Tree* biasa dan metode *Naïve Bayes* yang biasa digunakan dalam klasifikasi, seperti dilakukan oleh Qurahman *et al.* [6] juga

dapat digunakan, tetapi diperkirakan tidak menghasilkan prediksi yang akurat, memperhatikan kondisi data yang *imbalanced* tersebut.

Tabel 2. Variabel penelitian

Variabel	Keterangan	Skala Data
Area	Area pemberian kredit	Nominal
Divisi	Divisi yang mengelola debitur - Konsumer - Mikro - SME (small, medium enterprises)	Nominal
Kol CIF Dec-20	Kolektibilitas Desember 2020	Ordinal
Kol CIF Last Mth	Kolektibilitas bulan sebelumnya	Ordinal
Kol CIF Current	Kolektibilitas bulan tertentu	Ordinal
Restructured	Status restruktur - Restruktur Covid - Restruktur regular - <i>Blank</i> (tidak restruktur)	Nominal
DPD	Jumlah hari tunggakan (<i>days past due</i>)	Rasio
Pencairan Pokok	Nominal pencairan awal kredit	Rasio
OS Pokok	Outstanding kredit periode tertentu	Rasio
Flag Tunggakan	Kategori tunggakan - Tidak ada tunggakan - < 50.000 - 50.000 s.d. 100.000 - 100.000 s.d. 500.000 - 500.000 s.d. 1.000.000 - 1.000.000 s.d. 2.500.000 - 2.500.000 s.d. 5.000.000 - 5.000.000 s.d. 10.000.000 - >10.000.000	Ordinal
Jenis Kelamin	Jenis kelamin debitur - Laki-laki - Perempuan	Nominal
Pekerjaan	Pekerjaan debitur	Nominal

3.2. Metode Pengolahan Data

Setelah dilakukan tahap *data preprocessing*, maka tahap berikutnya adalah pengolahan data. Metode pengolahan data dilakukan dengan tiga tahap berikut:

1. Pengolahan data dilakukan dengan aplikasi KNIME [7], aplikasi *open-source* untuk *big data analytics* yang umum digunakan oleh *data scientists* untuk melakukan komputasi *predictive analytics*. KNIME mengintegrasikan berbagai komponen dari teknik *machine learning* dan *data mining* dengan menggunakan *building blocks*, sehingga lebih memudahkan untuk mendapatkan gambaran tahapan pemrosesan data.
2. Berbagai metode pengolahan data dibandingkan untuk mendapatkan kesimpulan metode yang paling optimal untuk penggunaan pada analisis kualitas kredit ini. Metode-metode yang dibandingkan adalah turunan dari metode *Decision Tree*, yaitu *Random Forest*, *Gradient Boosting*, dan *Ensemble Tree*. *Random Forest*, yang menurut Bawono, et al. [8] adalah turunan dari metode *Decision Tree* yang dapat digunakan untuk klasifikasi kredit bank. *Gradient Boosting*, yang menurut Hatwell, et al. [9] merupakan metode *machine learning*

turunan dari *decision tree* berdasarkan *dataset* yang tabular dan terstruktur yang dapat digunakan di *financial services*. *Ensemble Tree*, yang menurut Sartono dan Syafitri [5] adalah turunan dari metode *Decision Tree* sebagai solusi pilihan mengatasi kelemahan klasifikasi tunggal.

- Setelah pemrosesan dengan ketiga metode tersebut, dilakukan beberapa analisis untuk memastikan akurasi data, yaitu analisis statistika, seperti *standard deviation*, *variance*, *skewness*, dan *linear correlation*, serta *confusion matrix* yang memuat informasi akurasi dan *error* dari hasil komputasi dibandingkan dengan hasil sebenarnya.

Menurut Kasanah, *et al.* [10], Bisri, *et al.* [11], Rais, *et al.* [12], dan Brown *et al.* [13], untuk mengatasi *imbalance data class*, dapat digunakan teknik *oversampling SMOTE (synthetic minority oversampling technique)*, yang juga digunakan dalam penelitian ini.

4. HASIL DAN PEMBAHASAN

4.1. Analisis Statistika

Setelah melakukan *data preprocessing*, dilanjutkan eksplorasi data untuk mengetahui gambaran data dengan cara melihat hasil perhitungan statistika deskriptif dan korelasi antar variabel.

Tabel 3. Statistika deskriptif

Variabel	Min	Max	Mean	Median	Std. Dev.
DPD	0	3.208	11,023	0	90,569
Pencairan Pokok	0	3,00x10 ¹⁰	1,31x10 ⁸	56.250.000	3,04 x10 ⁸
OS Pokok	0	3,00x10 ¹⁰	1,02x10 ⁸	40.297.140	2,53x10 ⁸
Tenor	0	360	77,283	60	50,069

Tabel 3 memperlihatkan bahwa data variabel Pencairan Pokok dan OS Pokok merupakan data yang mirip atau homogen, sedangkan variabel DPD mengandung data yang sangat bervariasi, dan variabel Tenor memiliki nilai maksimal 360 bulan dan median 60 bulan. Korelasi antar variabel dapat dilihat pada Gambar 1. Warna biru pada Gambar 1 menunjukkan korelasi positif antar variabel, dengan warna yang semakin gelap menunjukkan korelasi yang semakin mendekati nilai +1. Warna oranye menunjukkan korelasi negatif antar variabel, dengan warna yang semakin gelap menunjukkan korelasi yang semakin mendekati nilai -1.

Berdasarkan Gambar 1 dapat diketahui bahwa terdapat korelasi positif yang kuat antara kolektibilitas bulan tertentu dengan bulan sebelumnya sebesar 0,850, antara *outstanding* kredit pada bulan tertentu dengan pencairan pokok sebesar 0,944, antara kategori tunggakan dengan jumlah hari tunggakan sebesar 0,946, antara kolektibilitas bulan tertentu dengan kategori tunggakan sebesar 0,889, dan antara kolektibilitas bulan tertentu dengan jumlah hari tunggakan sebesar 0,892. Dengan demikian, dapat disimpulkan bahwa terdapat hubungan positif yang kuat antara variabel kolektibilitas bulan tertentu, kategori tunggakan, dan jumlah hari tunggakan.

Setelah analisis korelasi, dilakukan pemeriksaan *imbalanced data* pada variabel respon. Sesuai dengan tujuan penelitian ini, dimana yang dilakukan analisis adalah kolektibilitas, maka ditetapkan variabel kolektibilitas bulan tertentu menjadi variabel respon. Gambar 2 menunjukkan variabel respon Kol CIF Current yang merupakan data kolektibilitas 1-5, dimana kolektibilitas 1 mendominasi dan menunjukkan bahwa data kolektibilitas adalah *imbalanced*. Kondisi ini adalah kondisi ideal dari data kredit atau pembiayaan bank, karena kolektibilitas 1 adalah kolektibilitas dengan tingkat pembayaran paling lancar.



Gambar 1. Korelasi antar variabel



Gambar 2. *Imbalanced class* variabel respon

Kategori 1 atau kolektibilitas lancar mempunyai frekuensi yang berbeda signifikan daripada kategori lainnya, yaitu sekitar 97,5% data memiliki kolektibilitas 1, sedangkan hanya sekitar 2,5% data yang memiliki kolektibilitas 2-5. Selain *imbalanced data*, kondisi ini juga mengindikasikan bahwa akurasi model harus melebihi 97,5% untuk mengimbangi *imbalanced data* tersebut, dan perlu dilakukan minimalisasi *false positive* menjadi di bawah 0,25%, yaitu tidak melebihi 10% dari klasifikasi data yang terkecil (2,5%), sehingga menghasilkan prediksi kolektibilitas yang lebih baik. *False positive* sangat penting untuk dipertimbangkan dalam kasus ini, karena *false positive* dapat meningkatkan risiko kredit bagi bank, disebabkan karena nasabah yang diprediksi akan lancar, pada kenyataannya cenderung berstatus tidak lancar.

4.2. Penggunaan *Random Forest*, *Gradient Boosting*, *Tree Ensemble*, dan *SMOTE*

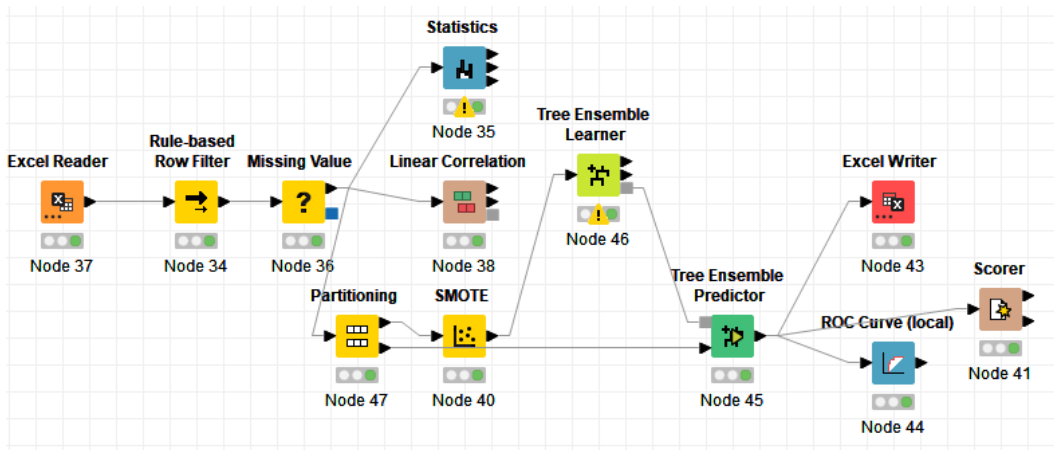
Berdasarkan deskripsi data, untuk meningkatkan ketepatan klasifikasi, diterapkan metode *Random Forest*, *Gradient Boosted Trees*, dan *Tree Ensemble*. Sementara itu, untuk mengatasi *imbalanced data* digunakan metode *SMOTE* (*synthetic minority oversampling technique*). Hasil klasifikasi dari metode tersebut dibandingkan dengan metode *Decision Tree* dan *Naïve Bayes*, untuk menunjukkan bahwa akurasi di atas 90% belum tentu tepat untuk model ini, karena ada permasalahan *false positive* yang dijelaskan pada bagian 4.1. Sebelum dilakukan analisis, data dibagi menjadi dua, yaitu data *training* dan data *testing*. Data *training* merupakan sebagian data

kredit atau pembiayaan selama masa pandemi COVID-19 dari bulan Maret hingga Juli 2021, yaitu sebanyak 750.057 baris. Data *testing* merupakan data pembiayaan kredit pada bulan Agustus 2021, yaitu sebanyak 150.274 baris. Pembentukan model dilakukan dengan menggunakan data *training*, kemudian diuji ketepatan klasifikasinya dengan menggunakan data *testing*. Dari model yang terbentuk, diperoleh nilai akurasi masing-masing model seperti pada Tabel 4.

Tabel 4. Ketepatan model klasifikasi

Model	Akurasi (ROC)	% False Positive
<i>Decision Tree</i>	94,94%	2,68%
<i>Naïve Bayes</i>	98,84%	0,77%
<i>Random Forest</i>	99,53%	0,39%
<i>Gradient Boosted Trees</i>	99,52%	0,37%
<i>Ensemble Tree</i>	99,51%	0,31%

Dari Tabel 4 dapat disimpulkan bahwa metode *Decision Tree* dan *Naïve Bayes* menghasilkan akurasi terendah, yaitu di bawah 99%. Klasifikasi pada model *Random Forest*, *Gradient Boosted Trees*, dan *Tree Ensemble* memberikan prediksi yang lebih akurat, yaitu akurasi sekitar 99,5% dan persentase *false positive* antara 0,3% dan 0,4%. Antara ketiga model *Random Forest*, *Gradient Boosted Trees*, dan *Tree Ensemble*, nilai akurasi berimbang, akan tetapi nilai *false positive* terendah terdapat pada model *Tree Ensemble*, sehingga model ini dapat digunakan untuk analisis selanjutnya. Untuk mengatasi *imbalanced data* yang cukup signifikan, kita perlu melakukan permodelan tambahan dengan menggunakan teknik *oversampling SMOTE*. Penggunaan teknik *Tree Ensemble* dengan *SMOTE* tersebut selengkapny dapat dilihat pada *KNIME workflow* pada Gambar 3.



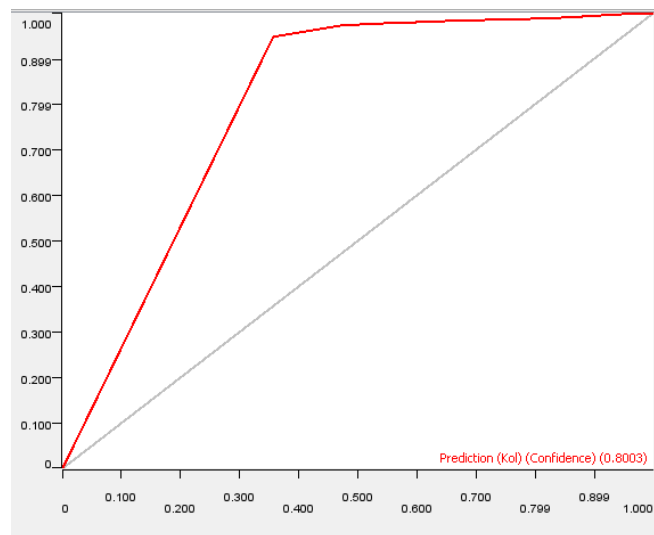
Gambar 3. Workflow klasifikasi *Tree Ensemble* dengan *SMOTE*

Tabel 5. Ketepatan model klasifikasi dengan/tanpa *SMOTE*

Model	Akurasi/ROC	% False positive
<i>Ensemble Tree</i> tanpa <i>SMOTE</i>	99,51%	0,31%
<i>Ensemble Tree</i> dengan <i>SMOTE</i>	99,44%	0,22%

Hasil komputasi yang didapatkan dengan teknik *Tree Ensemble* tersebut disampaikan pada Tabel 5. Meskipun akurasi sedikit berkurang dari 99,51% (tanpa menggunakan *SMOTE*)

menjadi 99,44% (dengan menggunakan SMOTE), tetapi persentase *false positive* jauh berkurang menjadi di bawah 0,25%, seperti dipersyaratkan oleh desain model, sehingga dapat disimpulkan bahwa penggunaan metode *Tree Ensemble* dengan SMOTE adalah model terbaik untuk memprediksi kelancaran pembiayaan kredit. Kurva ROC (*receiver operating characteristics*) dari metode *Tree Ensemble* dengan SMOTE ini menghasilkan *confidence level* 80% dan dapat dilihat pada Gambar 4.



Gambar 4. Kurva ROC klasifikasi *Tree Ensemble* dengan SMOTE

4.3. Predictive Analytics Kolektibilitas Debitur

Pada tahapan selanjutnya, setelah dilakukan pemilihan metode, dilakukan *predictive analytics* kelancaran pembayaran selama bulan September-Desember 2021, dengan menjalankan metode *Tree Ensemble* dengan SMOTE secara berulang. Tabel 6 menunjukkan hasil prediksi kolektibilitas untuk bulan September-Desember 2021 menggunakan model *Tree Ensemble* dengan SMOTE.

Tabel 6. Prediksi kolektibilitas September-Desember

Bulan (Tahun 2021)	Kolektibilitas					% NPL/ NPF*
	1	2	3	4	5	
Maret	95,01%	2,43%	0,67%	0,33%	1,57%	2,56%
April	94,70%	2,88%	0,65%	0,23%	1,54%	2,43%
Mei	94,81%	2,71%	0,62%	0,29%	1,57%	2,48%
Juni	95,18%	2,37%	0,66%	0,41%	1,38%	2,44%
Juli	94,41%	2,97%	0,73%	0,46%	1,44%	2,62%
Agustus	94,61%	2,70%	0,74%	0,44%	1,51%	2,70%
September (Prediksi)	94,16%	3,16%	0,72%	0,47%	1,49%	2,68%
Oktober (Prediksi)	93,97%	3,41%	0,69%	0,41%	1,53%	2,62%
November (Prediksi)	93,69%	3,62%	0,70%	0,43%	1,56%	2,68%
Desember (Prediksi)	93,56%	3,91%	0,64%	0,42%	1,47%	2,54%

*) NPL/NPF: *Non-Performing Loan/Financing*, yaitu kredit dengan kolektibilitas 3,4,5.

Tabel 6 memperlihatkan bahwa, meskipun terdapat kecenderungan kenaikan NPL/NPF pada Juli dan Agustus 2021, pada Desember 2021 diprediksi akan terjadi penurunan NPL/NPF. Hal ini menjadi indikasi yang menggembirakan, karena di tengah pandemi COVID-19, rasio NPL/NPF diprediksi menurun, yang menunjukkan potensi peningkatan kemampuan bayar dari nasabah. Prediksi NPL/NPF berdasarkan jenis produk, jenis kelamin, dan pekerjaan debitur dapat dilihat pada Tabel 7, 8, dan 9.

Tabel 7. Prediksi NPL/NPF berdasarkan Produk/Segmen*

Bulan (Tahun 2021)	Produk/Segmen			Total
	Konsumer	Mikro	SME	
Mei	1,88%	2,81%	5,23%	2,48%
Juni	1,92%	2,59%	5,02%	2,44%
Juli	2,05%	2,69%	5,59%	2,62%
Agustus	2,05%	2,77%	6,10%	2,70%
September (Prediksi)	2,14%	2,78%	5,59%	2,68%
Oktober (Prediksi)	2,01%	2,71%	5,89%	2,62%
November (Prediksi)	2,02%	2,68%	6,19%	2,68%
Desember (Prediksi)	1,94%	2,62%	5,61%	2,54%

*) SME: *small and medium enterprises* atau usaha kecil dan menengah (UKM)

Tabel 8. Prediksi NPL berdasarkan Jenis Kelamin*

Bulan (Tahun 2021)	% Non NPL/NPF (Kolektibilitas 1-2)		% NPL/NPF (Kolektibilitas 3-5)		Total Komposisi (Kolektibilitas 1-5)	
	L	P	L	P	L	P
	Mei	65,6%	34,4%	71,0%	29,0%	65,7%
Juni	63,3%	36,7%	71,9%	28,1%	63,5%	36,5%
Juli	65,2%	34,8%	72,2%	27,8%	65,4%	34,6%
Agustus	65,3%	34,7%	69,7%	30,3%	65,4%	34,6%
September (Prediksi)	65,2%	34,8%	72,1%	27,9%	65,4%	34,6%
Oktober (Prediksi)	65,2%	34,8%	70,1%	29,9%	65,3%	34,7%
November (Prediksi)	65,2%	34,8%	71,4%	28,6%	65,4%	34,6%
Desember (Prediksi)	65,2%	34,8%	70,0%	30,0%	65,3%	34,7%

*) L = laki-laki, P = perempuan

Tabel 7 menunjukkan bahwa NPL/NPF pada Desember 2021 diperkirakan akan turun untuk semua segmen (Konsumer, Mikro, dan SME). Tabel 8 menunjukkan bahwa NPL/NPF pada Desember 2021 diperkirakan akan turun untuk nasabah berjenis kelamin laki-laki, tetapi cenderung naik untuk nasabah perempuan. Dengan demikian, bank perlu melakukan rencana tindakan untuk mencegah penurunan kualitas kredit dari nasabah perempuan.

Tabel 9 menunjukkan bahwa NPL/NPF pada Desember 2021 diperkirakan akan turun untuk semua jenis pekerjaan, tetapi masih ada potensi peningkatan untuk jenis pekerjaan pegawai swasta. Hal ini dapat dipahami, karena adanya pandemi COVID-19 yang menyebabkan banyaknya perusahaan swasta yang melakukan program efisiensi. Akan tetapi, nilai NPL/NPF tertinggi tetap dihasilkan oleh pekerjaan wiraswasta, yang menyentuh angka 5%. Dengan demikian, bank perlu melakukan rencana tindakan untuk mencegah penurunan kualitas kredit dari nasabah dengan pekerjaan wiraswasta.

Tabel 9. Prediksi NPL berdasarkan Pekerjaan

Bulan (Tahun 2021)	Pekerjaan			
	Pegawai Negeri	Pegawai Swasta	Wiraswasta	Lain-lain
Mei	0,63%	2,03%	5,59%	2,15%
Juni	0,35%	1,06%	4,59%	2,72%
Juli	0,68%	2,06%	6,05%	1,79%
Agustus	0,81%	1,22%	5,20%	2,62%
September (Prediksi)	0,76%	1,07%	5,38%	2,56%
Oktober (Prediksi)	0,62%	1,10%	5,28%	2,53%
November (Prediksi)	0,67%	1,04%	5,31%	2,70%
Desember (Prediksi)	0,64%	1,11%	4,90%	2,57%

Berdasarkan data hasil komputasi yang disampaikan pada Tabel 7, 8 dan 9, dapat disimpulkan bahwa untuk data yang diambil pada penelitian ini, baik produk Konsumer, Mikro, maupun SME, meskipun memiliki basis kualitas kredit yang berbeda-beda, tetapi memiliki kecenderungan untuk dapat membaik. Pertama, meskipun memiliki basis kualitas yang lebih baik daripada debitur laki-laki, debitur perempuan cenderung akan mengalami pemburukan kualitas kredit. Kedua, seiring dengan membaiknya prediksi kualitas kredit, demikian juga halnya dengan prediksi penurunan NPL debitur dengan profesi wiraswasta yang secara historis cenderung tinggi, sedangkan debitur pegawai negeri dan swasta tetap memiliki NPL yang rendah. Dengan dilakukannya *predictive analytics* ini, bank dapat membuat strategi produk dan target market yang tepat untuk meningkatkan kualitas kredit.

5. KESIMPULAN

Pada penelitian ini dilakukan perbandingan berbagai metode klasifikasi *machine learning* untuk melakukan *predictive analytics* terhadap kualitas kredit perbankan dengan data dalam jumlah yang sangat besar (*big data analytics*). Hasil komputasi berbagai metode tersebut dibandingkan untuk kemudian menghasilkan rekomendasi teknik yang paling tepat, yaitu metode *Tree Ensemble* dengan SMOTE. Komputasi dapat menghasilkan rekomendasi strategi bagi bank, dengan contoh yang disampaikan pada penelitian ini menghasilkan rekomendasi bagi bank untuk memberikan fokus terhadap peningkatan kualitas kredit atau pembiayaan untuk nasabah berjenis kelamin perempuan atau nasabah dengan pekerjaan wiraswasta, karena adanya potensi peningkatan NPL/NPF.

Predictive analytics dapat dilakukan secara rutin oleh bank untuk membuat strategi produk dan target market yang tepat untuk meningkatkan kualitas kredit, yang pada akhirnya akan meningkatkan profitabilitas bank. Perlu diperhatikan bahwa basis data setiap bank berbeda-beda, sehingga kesimpulan kualitas kredit berdasarkan produk, jenis kelamin, dan pekerjaan di atas hanyalah contoh penerapan, dan tidak dapat disimpulkan untuk basis data yang berbeda. Pengembangan lebih lanjut dari metode ini dimungkinkan, seperti *predictive analytics* yang lebih ke masa depan, analisis profil debitur, dan analisis lainnya.

DAFTAR PUSTAKA

- [1] Otoritas Jasa Keuangan (OJK): Penilaian Kualitas Aset Bank Umum: <https://www.ojk.go.id/id/ regulasi/Pages/Penilaian-Kualitas-Aset-Bank-Umum.aspx>. 2019.

- [2] C. Tsai, C. Lai, H. Chao, and A.V. Vasilakos, “Big Data Analytics: A Survey”, *Journal of Big Data*, vol. 2, no. 21, 2015.
- [3] N. Elgendy and A. Elragal, “Big Data Analytics: A Literature Review Paper”, P. Perner Ed., *LNAI 8557*, pp. 214–227, 2014.
- [4] IBM, “Journey to AI”, <https://www.ibm.com/blogs/journey-to-ai/>, 2021.
- [5] B. Sartono and U.D. Syafitri, “Metode Pohon Gabungan: Solusi Pilihan untuk Mengatasi Kelemahan Pohon Regresi dan Klasifikasi Tunggal”, *Forum Statistika dan Komputasi*, vol. 15, no. 1, 2010.
- [6] T. Qurahman, M. Mustakim, and A. Jaini, “Penerapan Algoritma *Naïve Bayes Classifier* dan *Probabilistic Neural Network* untuk Klasifikasi Nasabah Bank dalam Membayar Kredit”, *Seminar Nasional Teknologi Informasi, Komunikasi, dan Industri*, pp. 205-213, 2019.
- [7] KNIME, “KNIME: End to End Data Science”, <https://www.knime.com>, 2021.
- [8] B. Bawono and R. Wasono, “Perbandingan Metode *Random Forest* dan *Naïve Bayes*, untuk Klasifikasi Debitur Berdasarkan Kualitas Kredit”, *Seminar Nasional Edusaintek*, vol. 3, pp. 343-348, 2019.
- [9] J. Hatwell, M. Gaber, and M. Azad, “gbt-HIPS: Explaining the Classifications of Gradient Boosted Tree Ensembles”, *Appl. Sci.*, vol. 11, no. 2511, pp. 1-17, <https://doi.org/10.3390/app11062511>, 2021.
- [10] A. Kasanah, Muladi, and U. Pujiyanto, “Penerapan Teknik SMOTE untuk mengatasi *Imbalance Class* dalam Klasifikasi Objektivitas Berita Online Menggunakan Algoritma KNN”, *Jurnal Rekayasa Sistem dan Teknologi Informasi*, vol. 3, no. 2, pp. 196-201, 2019.
- [11] A. Bisri and R. Rachmatika, “Integrasi *Gradient Boosted Trees* dengan SMOTE dan *Bagging* untuk Deteksi Kelulusan Mahasiswa”, *JNTETI*, vol. 8, no. 4, pp. 309-314, 2019.
- [12] A. Rais, and A. Subekti, “Integrasi SMOTE dan Ensemble AdaBoost untuk Mengatasi *Imbalance Class* pada Data Bank *Direct Marketing*”, *Jurnal Informatika*, vol. 6, no. 2, pp. 278-285, 2019.
- [13] I. Brown and C. Mues, “An experimental comparison of classification algorithms for imbalanced credit scoring data sets”, *Expert Systems with Applications*, vol. 39, pp. 3446-3453, 2012.