

Implementasi Algoritma *Machine Learning* untuk Pengembangan Model Prediktif atas Tingkat *Non-Performing Loan* dalam rangka Penjaminan Kredit UMKM Pemulihan Ekonomi Nasional

Reza Darmawan¹

Inspektorat Jenderal Kementerian Keuangan

reza.darmawan@kemenkeu.go.id

Dwi Purnomo²

Inspektorat Jenderal Kementerian Keuangan

d.purnomo@kemenkeu.go.id

Suryani Nilawati³

Inspektorat Jenderal Kementerian Keuangan

suryani.nilawati@kemenkeu.go.id

Fery Perdiansyah⁴

Inspektorat Jenderal Kementerian Keuangan

fery.perdiansyah@kemenkeu.go.id

Abstrak: Analisis ini bertujuan untuk membuat model prediktif NPL Kredit UMKM atas kegiatan Penjaminan Pemerintah Program PEN pada pelaku usaha UMKM. Hasil analisis ini dapat digunakan untuk melihat sektor usaha UMKM yang paling terdampak dengan adanya pandemi COVID-19, memberikan usulan tarif IJP yang akan diberikan kepada Jamkrindo dan Askrindo sebagai lembaga penjamin program PEN, serta memberikan usulan anggaran belanja subsidi IJP dan *Loss Limit* yang sesuai dan tepat. Model prediktif tersebut dikembangkan menggunakan algoritma random forest regression dengan menggunakan 2052 data panel penyaluran kredit kepada UMKM yang bersumber dari Statistik Perbankan Indonesia Otoritas Jasa Keuangan (OJK). Prediktor yang digunakan terdiri atas data sektor ekonomi serta kondisi makro ekonomi yang meliputi tingkat inflasi, tingkat pengangguran, pertumbuhan ekonomi, serta kondisi pandemi dengan target prediksi persentase NPL penyaluran kredit kepada UMKM.

Kata kunci: Non-Performing Loan (NPL), Penjaminan Kredit UMKM PEN, Machine Learning, Random Forest Regression

Abstract: This analysis aims to develop a predictive model for SMEs' credit NPL in order to support the decision making of Government SMEs Credit Guarantee program. The results of this analysis can be used to determine which SMEs business sector that is most affected by the COVID-19 pandemic, provide guarantee fees tariff suggestions that will be given to Jamkrindo and Askrindo as the guarantor institution for the National Economic Recovery program, as well as a tool for calculating the guarantee fees and loss limit budget. The predictive model was developed using a random forest regression algorithm using 2052 panel data for credit channeling to SMEs from the Indonesian Banking Statistics documents provided by,

the Indonesian Financial Services Authority (OJK). The predictors used consist of the economic sector as well as macroeconomic conditions including inflation, unemployment, economic growth, and pandemic conditions.

Keyword: *Non-Performing Loan, SMEs Credit Guarantee for Economic Recovery Program, Machine Learning, Random Forest Regression*

PENDAHULUAN

Dalam rangka mendukung kebijakan keuangan negara untuk penanganan pandemi Covid-19 dan pemulihan ekonomi nasional, Pemerintah melalui PP Nomor 23 tahun 2020 sebagaimana telah diubah dengan PP nomor 43 tahun 2020 telah mengatur 4 (empat) modalitas untuk program Pemulihan Ekonomi Nasional (PEN) yang meliputi penyertaan modal negara, penempatan dana, investasi pemerintah, dan penjaminan.

Pada kegiatan penjaminan kredit modal kerja Usaha Mikro Kecil Menengah (UMKM), pemerintah menugaskan Badan Usaha Milik Negara (BUMN) dalam hal ini Jamkrindo dan Askindo untuk bertindak sebagai penjamin bagi kredit modal kerja Usaha Mikro Kecil Menengah (UMKM). Program penjaminan ini sendiri bertujuan untuk meningkatkan minat perbankan dalam menyalurkan kredit kepada pelaku usaha agar mendapat kemudahan penjaminan saat mengajukan kredit. Selain itu, pemberian modal kerja pada UMKM penting dilakukan dalam membuat kegiatan usaha kembali menggeliat setelah terpuruk akibat dampak pandemi Covid-19.

Pemerintah telah melakukan berbagai dukungan agar program penjaminan berjalan dengan baik. Pada tahun 2020, pemerintah telah menganggarkan sejumlah Rp 6 triliun untuk memberikan dukungan pada program penjaminan pelaku Usaha Mikro Kecil

Menengah (UMKM) dengan rincian Rp 5 triliun sebagai Subsidi Belanja IJP dan Rp 1 triliun untuk dukungan penjaminan *loss limit*.

Salah satu dukungan yang dilakukan pemerintah adalah membayarkan seluruh Imbal Jasa Penjaminan (IJP) yang seharusnya ditanggung oleh pelaku usaha sebagai kreditur. Imbal Jasa Penjaminan (IJP) yang dianggarkan pemerintah akan dibayarkan ke pihak penjamin sesuai dengan perhitungan yang telah ditetapkan.

Salah satu faktor penentuan besaran Imbal Jasa Penjaminan (IJP) adalah adanya proyeksi *Non-Performing Loan* (NPL). Penentuan besaran rasio *Non-Performing Loan* (NPL) yang akurat akan berpengaruh pada ketepatan jumlah penganggaran yang dilakukan pemerintah dalam alokasi pembayaran Imbal Jasa Penjaminan (IJP). Pada penjaminan pemerintah atas kredit modal kerja pelaku usaha UMKM, penentuan tarif Imbal Jasa Penjaminan (IJP) didasari pada hasil metode perhitungan dan analisis PT Reasuransi Indonesia Utama (PT RIU) dengan mempertimbangkan proyeksi *Non-Performing Loan* (NPL).

Pengembangan model prediktif atas *Non-Performing Loan* (NPL) kredit UMKM ini didasarkan pada terbatasnya informasi dalam melakukan proyeksi NPL dalam rangka penjaminan program PEN. Pada penjaminan, Pemerintah membutuhkan informasi NPL sebagai hal yang menjadi dasar perhitungan tarif IJP dan alokasi anggarannya. Sehingga, dengan mengacu

pada proyeksi NPL berdasarkan model yang dikembangkan dalam penelitian kali ini, pemerintah akan memiliki informasi alternatif dalam mengevaluasi proyeksi NPL yang diajukan oleh PT RIU.

Secara umum, ada beberapa nilai strategis dari model prediktif yang dikembangkan terhadap pengambilan keputusan oleh pemerintah:

1. Melihat sektor usaha UMKM yang paling terdampak dengan adanya pandemi COVID-19
2. Memberikan usulan tarif IJP yang akan diberikan kepada Jamkrindo dan Askrindo sebagai lembaga penjamin program PEN
3. Memberikan usulan anggaran belanja subsidi IJP dan *Loss Limit* yang sesuai dan tepat.

KAJIAN PUSTAKA

Dari sisi makro ekonomi, NPL perbankan dapat dipengaruhi oleh faktor-faktor antara lain *Gross Domestic Product* (GDP), inflasi, dan tingkat pengangguran. Terdapat sejumlah riset yang dapat dirujuk berkaitan dengan *Non-Performing Loan* (NPL) pada sektor perbankan. Penelitian terdahulu atas NPL menggunakan berbagai kondisi makro ekonomi sebagai indikator pada wilayah Uni Eropa yang berfokus pada karakteristik spesifik dari masing-masing perbankan, selain juga mempertimbangkan faktor makroekonomi serta dimensi tata kelola negara sebagai determinan atas NPL perbankan konvensional di Uni Eropa [1].

Data model dinamis digunakan untuk menentukan proyeksi NPL di Asia dengan menggunakan makroekonomi dan faktor spesifik pada bank seperti pertumbuhan kredit dan pinjaman bank yang berlebihan [2]. Tingkat NPL sendiri penting untuk menjaga stabilitas keuangan

dalam sistem keuangan global yang saling terhubung. Penelitian menggunakan data panel berdasarkan indikator makroekonomi dari tahun 1994 sampai dengan tahun 2014 pada negara Asia yang berkembang. Hasil analisis atas implikasi makro ekonomi yaitu Pertumbuhan GDP, pemberian kredit, dan tingkat pengangguran memberikan hasil yang berpengaruh signifikan pada NPL.

Di sisi lain, model prediktif atas NPL di Uni Eropa juga menggunakan data panel dengan menggunakan berbagai faktor makroekonomi dan makro finansial sebagai indikator utama [3]. Model prediktif atas NPL tersebut, selain dikembangkan untuk level Uni Eropa, juga dikembangkan dengan basis kewilayahan yakni untuk negara-negara Uni Eropa di bagian barat, serta untuk negara Uni Eropa di bagian timur dan tengah. Penggunaan *machine learning* untuk melakukan prediksi atas NPL pernah dikembangkan dengan menggunakan algoritma *Support Vector Machine* (SVM), *Bagging* dan *AdaBoost* [4]. Pengembangan model tersebut menggunakan prediktor makroekonomi dan karakteristik spesifik bank penyulur.

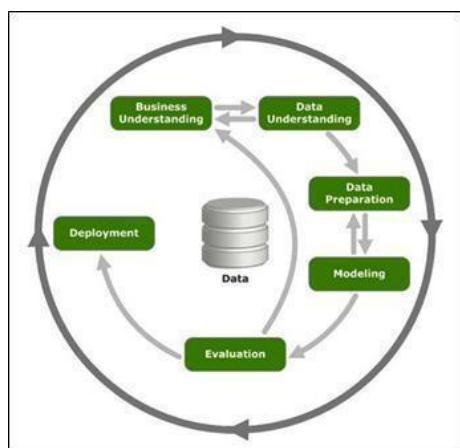
Penelitian serupa dalam konteks perbankan Indonesia juga telah dilakukan, yakni pengembangan model *classifier* untuk memprediksi NPL perbankan dengan menggunakan algoritma Naïve Bayes, *Decision Tree*, *k-NN*, *Rule Induction*, *Logistic Regression*, *Random Forest*, *Generalized Linear Model*, dan *Gradient Boost Trees*. Hasil

penelitian tersebut menghasilkan kesimpulan bahwa model dengan algoritma *Random Forest* memiliki tingkat akurasi terbaik. Selain itu, mengembangkan model prediksi juga berguna untuk penilaian risiko kredit pada BPR dengan menggunakan algoritma *decision tree*. Hanya saja, metode *machine learning* yang dikembangkan dalam

ketiga penelitian tersebut bertujuan untuk mengembangkan model *classifier* [5]. Sementara, untuk kebutuhan pengambilan keputusan dalam rangka program penjaminan PEN dibutuhkan model prediksi yang bersifat *regressor*. Namun, riset tersebut membuka peluang pemanfaatan *machine learning*, termasuk di dalamnya pohon keputusan (*decision tree*), dalam memprediksi NPL [4]. Fokus pada karakteristik spesifik dari masing-masing bank maupun basis kewilayahan dari objek penelitian dalam riset-riset tersebut menjadi perbedaan utama jika dibandingkan dengan pengembangan model prediktif atas NPL ini yang mengambil fokus pada determinan sektor usaha kredit yang dijaminkan dalam rangka program Pemulihan Ekonomi Nasional (PEN).

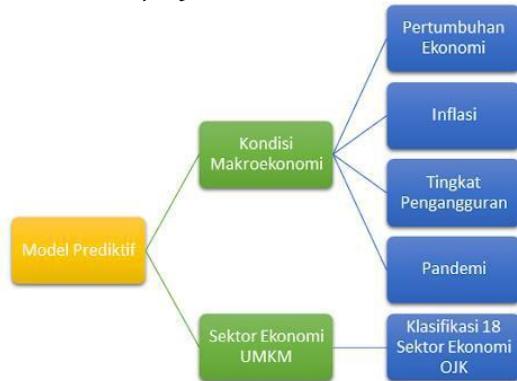
METODOLOGI

Metodologi yang jamak digunakan dalam proyek data mining adalah *Cross-Industry Standard Process for Data Mining* (CRISP-DM). CRISP-DM membagi tahapan pengembangan model machine learning ke dalam sebuah siklus yang terdiri atas enam proses.



Gambar 1. Metodologi CRISP-DM dalam Machine Learning

Tahap dalam CRISP-DM dimulai dari pemahaman proses bisnis serta data yang tersedia. Berikutnya dilakukan proses penyiapan data untuk kemudian dilakukan pengembangan model. Setelah model dikembangkan akan dilakukan evaluasi atas akurasi model tersebut. Setelah itu, barulah kemudian model tersebut dilakukan *deployment*.



Gambar 2. Skema Model Prediktif

Pengembangan model prediktif ini, dilakukan dengan target prediksi yaitu persentase NPL dari penyaluran kredit kepada UMKM per masing-masing sektor ekonomi. Prediktor yang akan digunakan, berdasarkan pada hasil telaah literatur yakni kondisi makroekonomi yang terdiri atas pertumbuhan ekonomi, inflasi, tingkat pengangguran, dan pandemi. Selain itu, digunakan juga prediktor berupa sektor ekonomi UMKM penerima penyaluran kredit. Data yang digunakan untuk pengembangan model bersumber dari dokumen Statistik Perbankan Indonesia yang diterbitkan oleh Otoritas Jasa Keuangan serta data kondisi makroekonomi yang bersumber dari Badan Pusat Statistik. Dari kedua sumber data tersebut, dihasilkan 2052 data panel penyaluran kredit kepada 18 sektor ekonomi UMKM sesuai klasifikasi OJK dan kondisi makroekonomi pada periode Januari 2011-Juni 2020.

Tabel 1. Definisi Operasional Variabel untuk pengembangan model prediktif

Variabel	Definisi
Sektor Ekonomi	Merupakan variable kategorikal yang berisi 18 sektor ekonomi berdasarkan klasifikasi OJK
PandemicTF	Merupakan variable biner yang bernilai 1 jika penyaluran kredit dan NPL termasuk ke dalam masa pandemi dan bernilai 0 jika termasuk masa sebelum pandemi
Inflasi	Tingkat inflasi bulanan. Sumber data BPS
Unemployment	Data tingkat pengangguran yang bersumber dari BPS
Economic Growth	Data pertumbuhan ekonomi bulanan year-on-year yang bersumber dari BPS
Credit Channeling	Data penyaluran kredit kepada UMKM per sector ekonomi secara bulanan yang bersumber dari SPI OJK
Percent NPL	Persentase NPL yang merupakan pembagian antara NPL kredit dan penyaluran kredit UMKM yang bersumber dari SPI OJK

Algoritma yang dipilih untuk digunakan dalam pengembangan model prediktif ini adalah *random forest*. Algoritma ini berbasis pada pohon keputusan (*decision tree*). Algoritma *random forest* akan membuat beberapa pohon keputusan dan menggabungkannya untuk menghasilkan prediksi yang lebih stabil dan akurat. Penelitian sebelumnya menunjukkan bahwa algoritma *random forest* memiliki tingkat akurasi yang paling tinggi jika dibandingkan algoritma yang *machine learning* lainnya dalam pengembangan model prediktif atas *non-performing loan* yakni mencapai 96,55%.

Dalam penelitian ini, dikembangkan dua model yang berbasis pada algoritma *random forest*, dengan mengaplikasikan metode *cross validation* yang berbeda. Yaitu *grid-search cross validation* dan *randomized-search cross validation*. Atas kedua model

tersebut, dipilih model dengan akurasi terbaik untuk dilakukan *deployment*.

HASIL DAN PEMBAHASAN

Business Understanding

Program penjaminan kredit UMKM melibatkan empat pihak, yaitu Pemerintah selaku penjamin, Badan Usaha Penjaminan yang melaksanakan penugasan penjaminan dalam hal ini dilakukan oleh PT. Jamkrindo dan PT. Askindo, perbankan selaku penerima jaminan dan pemberi fasilitas pinjaman, dan terakhir yaitu pihak terjamin dalam hal ini adalah pelaku usaha UMKM. Selain itu terdapat PT Reasuransi Indonesia Utama (RIU), BUMN yang ditugaskan oleh pemerintah untuk melaksanakan kegiatan operasional program *loss limit*, memberikan masukan dalam bentuk analisis aktuaria terhadap pelaksanaan Penjaminan Pemerintah, dan tugas lainnya yang diberikan Pemerintah.



Gambar 3. Skema alur program penjaminan kredit UMKM dalam rangka PEN.

Pada skema di atas, dapat dilihat bahwa kegiatan penjaminan program PEN dimulai saat pelaku usaha mengajukan pinjaman kepada pihak perbankan. Atas permohonan yang diajukan oleh pelaku usaha tersebut, pihak perbankan melakukan analisis syarat dan ketentuan sesuai dengan standar operasi yang berlaku di masing-masing perbankan. Jika ketentuan dan persyaratan telah terpenuhi, maka pihak perbankan dapat mengajukan

penjaminan kredit modal kerja tersebut kepada Jamkrindo/Askrindo untuk kemudian diterbitkan sertifikat penjaminan.

Dalam pelaksanaan penugasan Penjaminan Program PEN, PT Jamkrindo dan PT Askrindo berhak mendapatkan Imbal Jasa Penjaminan (IJP) yang dibayarkan seluruhnya oleh Pemerintah. Besaran IJP yang diberikan dihitung dengan formula besaran IJP = tarif IJP x plafon Pinjaman, di mana tarif IJP tersebut ditetapkan oleh Pemerintah dan ditetapkan dengan memperhatikan:

- keputusan mengenai kebijakan penjaminan;
- laporan keuangan PT Jamkrindo dan/atau PT Askrindo;
- kemampuan Pemerintah melalui Menteri Keuangan menyediakan alokasi belanja pembayaran IJP; dan/atau
- data dan informasi pendukung lainnya seperti proyeksi *non-performing loan* (NPL), besaran porsi penjaminan, batasan *loss limit*, biaya *overhead*, jangka waktu Pinjaman, dan *margin*.

Model prediktif atas NPL kredit UMKM yang dikembangkan melalui penelitian ini dapat dipergunakan sebagai alat bantu bagi pemerintah dalam menentukan tarif IJP dalam rangka penjaminan kredit UMKM.

Data Understanding

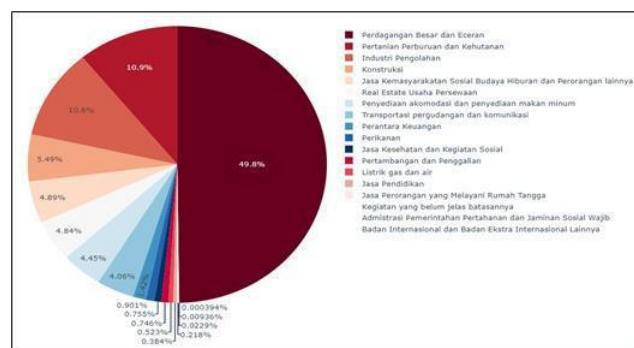
Berdasarkan analisis atas data penyaluran kredit kepada UMKM, per Juni tahun 2020, diketahui bahwa terdapat total Rp 1.019,73 triliun kredit yang disalurkan oleh perbankan kepada 18 sektor usaha UMKM. Dari total penyaluran kredit

tersebut, terdapat NPL sebesar Rp 41,99 triliun.



Grafik 1. Grafik Total Penyaluran Kredit dan Total NPL Kredit UMKM Tahun 2020 (s.d. Juni)

Jika dianalisis berdasarkan sektor ekonomi UMKM, diketahui bahwa Rp 505,66 triliun atau sebesar 49,8% dari total kredit tersebut disalurkan oleh perbankan kepada UMKM dengan sektor ekonomi perdagangan besar dan eceran. Dua sektor ekonomi UMKM terbesar setelahnya, adalah sektor ekonomi pertanian, perburuan, dan kehutanan dengan penyaluran kredit sebesar Rp 110,98 triliun (10,9%) serta sektor ekonomi industri pengolahan dengan penyaluran kredit sebesar Rp 107,34 triliun. Ketiga sektor ekonomi UMKM tersebut sejak tahun 2013 menjadi penerima mayoritas dari kredit yang disalurkan oleh perbankan.

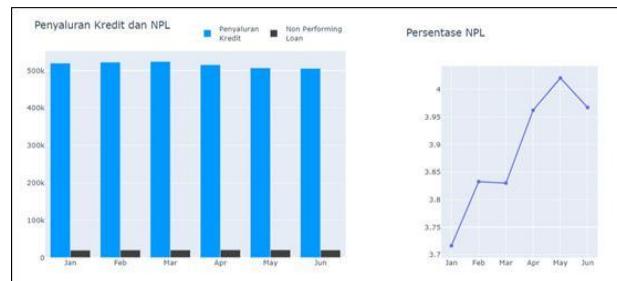


Grafik 2. Komposisi penyaluran kredit perbankan kepada UMKM per sektor ekonomi bulan Juni 2020.

Dari sisi kinerja kredit, sebagai penerima penyaluran mayoritas kredit UMKM, adanya pandemi Covid-19 memberikan pengaruh yang berbeda terhadap sektor ekonomi perdagangan besar dan eceran, sektor industri

pengolahan, sektor pertanian dan kehutanan, serta sektor konstruksi.

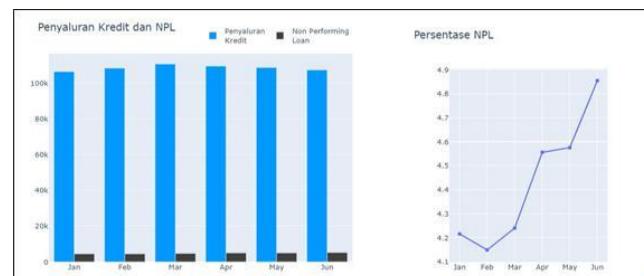
Sektor ekonomi perdagangan besar dan eceran terlihat mengalami kenaikan persentase NPL jika dibandingkan dengan periode saat kondisi sebelum pandemi. Sektor ini pernah mencatatkan NPL sebesar 4,02% atau Rp 20,38 triliun dari penyaluran kreditnya pada bulan Mei 2020. Meskipun secara persentase mengalami kenaikan pada bulan Mei, namun secara jumlah terjadi penurunan nilai NPL jika dibandingkan dengan bulan April yaitu senilai Rp 20,42 triliun. Penurunan ini terus berlanjut, hingga bulan Juni 2020, penyaluran kredit UMKM pada sektor ekonomi perdagangan besar dan eceran memiliki NPL sebesar 3,97% atau senilai Rp 20,06 triliun.



Grafik 3. Kinerja kredit UMKM sektor ekonomi perdagangan besar dan eceran tahun 2020

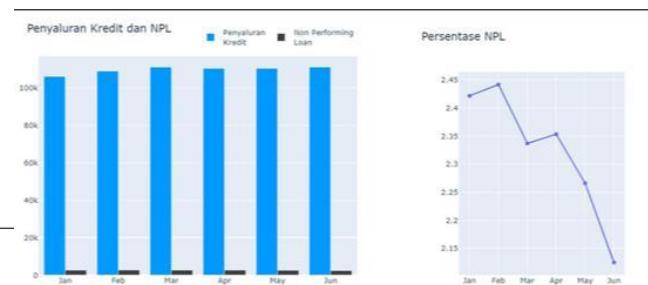
Penyaluran kredit UMKM kepada sektor ekonomi industri pengolahan juga tampak mengalami kenaikan persentase NPL sejak bulan Maret 2020. Terakhir, pada bulan Juni 2020, NPL penyaluran kredit sektor ini mencapai 4,85% atau sebesar Rp

5,21 triliun. Kenaikan ini cukup signifikan jika dibandingkan dengan sebelum pandemi atau bulan Februari 2020 yaitu sebesar Rp 4,49 triliun atau 4,15% dari penyaluran kredit.



Grafik 4. Kinerja kredit UMKM sektor industri pengolahan tahun 2020

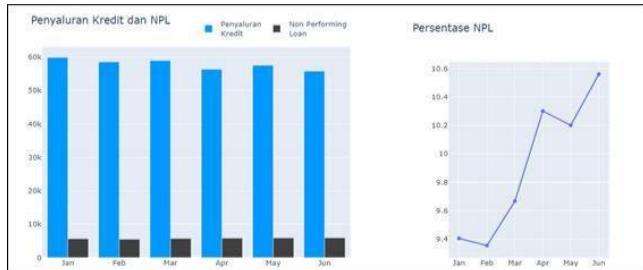
Menariknya, penyaluran kredit UMKM untuk sektor pertanian, perburuan, dan kehutanan, justru terlihat mengalami penurunan persentase NPL. NPL penyaluran kredit untuk UMKM sektor tersebut pada bulan Juni 2020 adalah 2,12% atau Rp 2,36 triliun. Jumlah NPL ini menurun jika dibandingkan sejak sebelum pandemi, yakni bulan Februari 2020 sebesar Rp 2,66 triliun atau 2,44% dari penyaluran kredit.



Grafik 5. Kinerja kredit UMKM sektor pertanian, perburuan, dan kehutanan 2020

Sementara itu, pada bulan Juni 2020, kredit kepada UMKM sektor konstruksi memiliki NPL 10,56% atau Rp 5,89 triliun dari total penyaluran. Pada tahun 2020, kredit UMKM sektor konstruksi mengalami

peningkatan NPL sejak sebelum pandemi Covid-19 atau bulan Februari 2020, yang berada pada level 9,35%. atau sebesar Rp 5,47 triliun dari total penyaluran kredit.



Grafik 6. Kinerja kredit UMKM sektor konstruksi tahun 2020

Bila dilakukan perbandingan data NPL atas penyaluran kredit kepada UMKM antar sektor ekonomi pada bulan Juni 2020, maka dapat dilihat bahwa penyaluran kredit kepada UMKM dengan sektor ekonomi konstruksi memiliki tingkat NPL yang paling tinggi. Sektor ekonomi dengan persentase NPL tertinggi kedua adalah penyaluran kredit kepada UMKM untuk kegiatan yang belum jelas batasannya. Dan yang ketiga adalah penyaluran kredit untuk sektor ekonomi industri pengolahan.



Grafik 7. Perbandingan Persentase NPL atas penyaluran kredit UMKM per sektor ekonomi Juni 2020

Data Preparation

Prediktor-prediktor (*features*) dalam model prediktif dapat dibagi ke dalam dua jenis berdasarkan bentuknya, yakni *continuous features* dan *categorical features*. *Continuous features* adalah prediktor yang bersifat numerik. Sementara *categorical features* adalah prediktor yang bersifat diskrit. Persiapan data dilakukan dengan melakukan transformasi atas prediktor yang bersifat kategorikal, yakni sektor ekonomi UMKM penerima penyaluran kredit.

Ada beragam cara untuk melakukan *encoding* atas *categorical features*. Salah satunya, yang paling banyak digunakan adalah *One Hot Encoder*. *One Hot Encoder* mengubah setiap nilai dalam fitur *categorical* menjadi fitur-fitur individual. Tujuannya agar algoritma *machine learning* tidak salah menerjemahkan variabel *categorical* tersebut sebagai variabel numerik.

Data Modelling

Langkah selanjutnya adalah *data modelling* dengan tahapan seperti disajikan pada Gambar 4. Kertas kerja pengembangan model terdokumentasi pada tautan berikut:



Gambar 4. Tahapan *data modelling*

a. Dataset Splitting

Untuk melakukan pengembangan model, maka dataset harus dibagi menjadi dua, agar model prediktif yang dihasilkan dapat dievaluasi. Metode ini dikenal sebagai *holdout method*, yakni dataset dibagi menjadi dataset *training* dan dataset *testing* [6].

Dataset *training*, adalah dataset yang akan diolah oleh algoritma untuk membentuk model prediktif. Sementara dataset *testing*, akan digunakan untuk mengevaluasi akurasi model prediktif yang dibentuk menggunakan dataset *testing* [7]. *Training* dan *testing* harus bersifat *mutually exclusive* agar model prediktif yang dihasilkan tidak *overfitting*.

Pemisahan (*split*) dataset *training* dan *testing* menggunakan *library train_test_split* dari *scikit-learn*. Dalam pemisahan dataset dilakukan *setting* parameter *test-size* dan *stratify*. Parameter *test-size* digunakan untuk mendefinisikan besaran porsi dari dataset yang akan digunakan sebagai dataset *testing*. Dalam pengembangan model prediktif kali ini, parameter *test-size* diset sebesar 0,2, artinya 20% dataset menjadi dataset *testing*, dan 80% menjadi dataset *training*.

b. Pemilihan dan *fitting* algoritma

Dalam pengembangan model, diaplikasikan metodologi *ensemble learning*. *Ensemble learning* menggunakan beberapa model pembelajaran mesin (*machine learning model*) untuk membuat prediksi yang lebih baik atas dataset.

Algoritma *ensemble learning* yang digunakan dalam model prediksi ini adalah *Random Forest Regression*. *Random forest* merupakan *ensemble learning* yang bertipe *bootstrap aggregation (bagging)*. Penggunaan algoritma berbasis pohon keputusan seperti *random forest*, alih-alih menggunakan algoritma regresi linear, didasarkan pada pertimbangan bahwa model prediksi ini memiliki fitur yang bersifat kategorikal yakni sektor ekonomi UMKM.

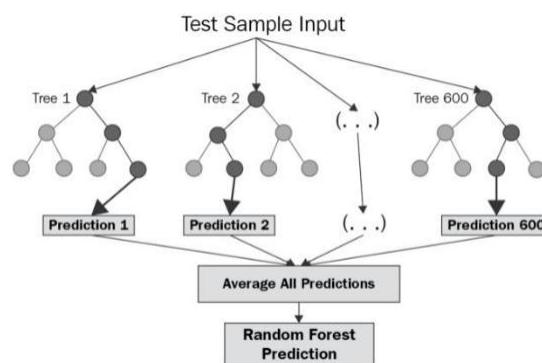
Bootstrap berarti *ensemble learning* akan dilakukan dengan mengambil *random sampling* dengan penggantian. Melalui metode ini, *training set* yang sudah

dihadirkan melalui proses *splitting*, akan dipilih kembali secara *sampling* untuk menghasilkan sejumlah dataset baru, yang pada akhirnya akan memungkinkan adanya beberapa observasi.

Bagging merupakan prosedur umum untuk mengurangi varians pada algoritma yang memiliki varians tinggi seperti *decision tree* [6]. *Bagging* membuat setiap model yang dihasilkan akan berjalan secara independen untuk kemudian mengagregasikan hasilnya tanpa preferensi atas masing-masing model.

Algoritma *Random Forest Regression* akan menghasilkan beberapa model *decision tree* [8]. Tidak ada interaksi antara setiap model *decision tree* tersebut. Perlunya algoritma *random forest* adalah karena algoritma *decision tree* sangat sensitif terhadap data *training set* yang digunakan dalam membentuk model.

Konsekuensinya, jika *training set* diubah, maka *decision tree* yang dihasilkan juga akan berbeda sehingga akan menghasilkan prediksi yang berbeda pula. Ini disebut sebagai risiko *overfitting*. *Random forest regression* dimanfaatkan sebagai upaya untuk menanggulangi risiko *overfitting* tersebut [8].



Gambar 5. Konsep *Random Forest Regression*

Model *random forest* merupakan sebuah *meta-estimator*, artinya ia mengkombinasikan hasil dari beberapa

prediksi, yang dalam hal ini dihasilkan oleh beberapa *decision tree*. Dalam konsep regresi menggunakan *random forest*, maka prediksi final yang dihasilkan dari model tersebut adalah rata-rata dari hasil prediksi individual setiap *decision tree*.

c. Hyperparameter tuning

Dalam mengembangkan model prediktif, setiap algoritma memiliki satu set *hyperparameter* yang dapat dilakukan *tuning*. *Tuning* parameter adalah serangkaian proses untuk menemukan parameter yang paling optimal untuk meningkatkan akurasi model prediktif yang dihasilkan.

Untuk melakukan *tuning*, dapat digunakan modul *Cross Validation* dari *scikit-learn*. Cara kerja utama dari *cross validation* adalah dengan membagi dataset training secara random ke dalam beberapa kelompok sebagaimana ditentukan. Dari kelompok-kelompok tersebut, salah satu grup akan diperlukan sebagai *test-set*, dan sisanya akan digunakan sebagai *training-set*. Model kemudian akan dilatih menggunakan

training-set tersebut dan dinilai menggunakan *test-set*. Proses tersebut kemudian diulang hingga setiap kelompok telah digunakan sebagai *test-set* [6]. Oleh karena itulah, setelah dilakukan fit dataset *training* pada algoritma *Random Forest Regressor*, dilakukan satu langkah tambahan untuk mengambil model dengan parameter estimator terbaik.

Dalam pengembangan model, digunakan dua jenis *cross validation* untuk kemudian dalam tahap evaluasi model dipilih model dengan tingkat akurasi terbaik, yaitu *grid search* dan *randomized search*. Meskipun secara konsep sama-sama bekerja dengan melakukan pemilihan kombinasi *hyperparameter* terbaik untuk model prediksi, *grid search* dan *randomized*

search memiliki cara kerja yang berbeda. *Grid search* bekerja dengan melakukan *training* secara berulang kali atas model yang dikembangkan sebanyak parameter yang ditentukan [7]. Dengan cara ini, model yang sedang dikembangkan dapat diuji menggunakan tiap-tiap nilai *hyperparameter* untuk kemudian menemukan nilai optimal guna mendapatkan model dengan estimator terbaik.

Sementara itu, dalam *randomized search*, mengatur seperangkat nilai *hyperparameter* untuk kemudian secara random memilih kombinasi *hyperparameter* tersebut dan menghasilkan model prediksi dengan estimator terbaik. Hal ini untuk keperluan *cross validation*, terdapat beberapa *hyperparameter* yang ditetapkan sebagai *grid* untuk diuji [7]. *Hyperparameter* tersebut adalah:

- Jumlah *decision tree* yang akan digunakan oleh algoritma *random forest* (*n_estimators*)
- Jumlah maksimum fitur yang digunakan untuk setiap *split* (*max_features*)
- Jumlah maksimum level dalam setiap *decision tree* yang dihasilkan (*max_depth*)
- Sampel minimum yang digunakan untuk membuat pemisahan *node* (*min_samples_split*)
- Jumlah sampel minimum untuk setiap *leaf node* (*min_samples_leaf*)
- Metode seleksi sampel untuk melatih setiap *decision tree* (*bootstrap*)

Hyperparameter untuk *randomized search* adalah sebagaimana dirinci Tabel 1, sementara untuk *grid search* adalah sebagaimana disajikan pada Tabel 2. Terlihat bahwa *hyperparameter* untuk *randomized search* diatur untuk memiliki lebih banyak nilai untuk mengakomodasi pemilihan *hyperparameter* secara acak. Sementara *grid*

search memiliki opsi *hyperparameter* yang lebih spesifik untuk efisiensi sumber daya pengujian.

Tabel 2. Rincian Pengaturan *Hyperparameter Random Forest* untuk *Randomized Search*

Hyperparameter	Value
N_estimators	200, 400, 600, 800, 100, 1200, 1400, 1600, 1800, 2000
Max_features	Auto, sqrt
Max_depth	10, 20, 30, 40, 50, 60, 70, 80, 90, 100, 110, None
Min_samples_split	2, 5, 10
Min_samples_leaf	1, 2, 4
Bootstrap	True, False

Tabel 3. Rincian Pengaturan *Hyperparameter Random Forest* untuk *Grid Search*

Hyperparameter	Value
N_estimators	50, 100, 200
Max_features	Auto
Max_depth	None
Min_samples_split	2
Min_samples_leaf	1
Bootstrap	True

Berdasarkan hasil *tuning* menggunakan *Grid Search Cross Validation*, diketahui parameter yang menghasilkan akurasi terbaik adalah model *random forest* dengan 200 *decision tree*. Sementara hasil *tuning* menggunakan *Randomized Search Cross Validation*, diketahui parameter yang menghasilkan akurasi terbaik adalah model *random forest* dengan 400 *decision tree*.

Model Evaluation

Setelah model dikembangkan, perlu dilakukan evaluasi model untuk menghitung akurasi model. Akurasi merupakan metriks yang merepresentasikan fraksi dari prediksi model yang benar. Akurasi diukur dengan membandingkan jumlah prediksi yang benar dengan jumlah total prediksi. Atas model *ensemble* yang

dibangun menggunakan *Random Forest Regression*, dilakukan pengecekan nilai akurasi dengan menggunakan *Mean Absolute Error* (MAE) masing-masing pada modul *Grid Search Cross Validation* dan *Randomized Search Cross Validation*. MAE adalah salah satu metode untuk mengukur tingkat keakuratan suatu model. Nilai MAE merepresentasikan rata-rata kesalahan (*error*) absolut antara hasil prediksi dengan nilai sebenarnya [7].

Selain MAE, dalam evaluasi model prediktif digunakan metriks *Root Mean Squared Error* (RMSE). RMSE mengukur perbedaan antara nilai hasil prediksi dengan nilai actual. Sehingga, semakin kecil nilai RMSE, maka semakin bagus prediksi yang dihasilkan oleh model tersebut. Perbedaannya dengan MAE, nilai *error* yang digunakan untuk menghitung RMSE tidak diukur berdasarkan nilai absolutnya.

Metriks yang terakhir adalah *R-Squared*. Metriks ini merepresentasikan proporsi dari varian dalam variabel target yang mampu dijelaskan oleh fitur-fitur yang digunakan dalam model prediksi. Sehingga, semakin tinggi nilai *R-Squared* maka semakin bagus model tersebut karena fitur yang digunakan mampu menjelaskan sebagian besar adanya varian dalam variabel target. Diketahui bahwa model *random forest* yang dihasilkan menggunakan *Grid Search Cross Validation*, jika diukur menggunakan dataset testing menghasilkan:

- *Mean Absolute Error* sebesar 0,0068, artinya rata-rata prediksi yang dihasilkan atas persentase NPL memiliki deviasi sebesar 0,69 persen dari persentase NPL aktual.
- *Root Mean-Squared Error* sebesar 0.0248, artinya rata-rata prediksi yang dihasilkan atas persentase NPL memiliki

- deviasi 2,48 persen dari persentase NPL aktual.
- *R-Squared* sebesar 0,48 artinya model prediktif mampu menjelaskan 48% faktor penyebab tingkat persentase NPL.

Diketahui bahwa model random forest yang dihasilkan menggunakan *Randomized Search Cross Validation*, jika diukur menggunakan dataset testing menghasilkan:

- *Mean Absolute Error* sebesar 0.0067, artinya rata-rata prediksi yang dihasilkan atas persentase NPL memiliki deviasi sebesar 0,67 persen dari persentase NPL aktual.
- *Root Mean-Squared Error* sebesar 0.0225, artinya rata-rata prediksi yang dihasilkan atas persentase NPL memiliki deviasi 2,25 persen dari persentase NPL aktual.
- *R-Squared* sebesar 0,57 artinya model prediktif mampu menjelaskan 57% faktor penyebab tingkat persentase NPL

Sehingga, dengan mempertimbangkan metriks-metriks pengukuran akurasi model, maka untuk keperluan deployment, digunakan model prediktif hasil dari *randomized search cross validation*.

Deployment

Model yang telah dipilih berdasarkan hasil evaluasi kemudian disimpan menggunakan *library pickle* dan dilakukan *deployment* pada *dashboard* untuk keperluan pengambilan keputusan. Untuk pengembangan *dashboard*, dipergunakan *library python plotly*. *Source code* untuk *dashboard* tersebut memuat 4 (empat) fitur utama yakni:

0,0067 dan RMSE sebesar 0,0225, serta *R-Squared* sebesar 0,57.

- Informasi Umum, berisi grafik-grafik interaktif yang membantu melakukan analisis deskriptif atas penyaluran kredit dari perbankan kepada UMKM.
- Evaluasi Sektor Terdampak, yang berguna untuk melakukan prediksi atas sektor-sektor yang terdampak oleh pandemi sesuai kondisi makro ekonomi terkini.
- Penganggaran IJP dan Loss Limit, berisi perhitungan anggaran untuk subsidi IJP dan loss limit yang berdasar pada proyeksi atas NPL kredit UMKM sesuai kondisi makro ekonomi.
- Evaluasi Tarif IJP, berisi perhitungan tarif IJP sesuai dengan proyeksi NPL kredit UMKM sesuai kondisi makro ekonomi.

Terkait dengan formulasi perhitungan tarif IJP yang dipergunakan adalah sebagai berikut:

$$\frac{0,8) - 0,01}{0,9}$$

Adapun formulasi perhitungan IJP *loss limit* yang dibayarkan oleh Jamkrindo/Askrindo kepada Pemerintah adalah sebesar 1% dari pinjaman yang diberikan, dengan asumsi jumlah *Loss Limit* sama dengan premi IJP *loss limit* yang diterima. Kedua formula tersebut menjadi acuan bagi perhitungan tarif IJP dan perhitungan penganggaran untuk IJP dan *loss limit* dari NPL yang telah diprediksi.

PENUTUP

Riset ini menghasilkan model prediktif atas NPL penyaluran kredit kepada UMKM dengan algoritma *Random Forest Regresion* dan penerapan *randomized search cross validation* dengan tingkat MAE sebesar

Beberapa keterbatasan yang terdapat dalam analisis NPL UMKM ini yang pertama adalah data panel yang

dipergunakan adalah data penyaluran kredit kepada UMKM secara sektoral yang merupakan data agregat dari penyaluran kredit dari seluruh bank kepada masing-masing sektor ekonomi. Hal ini menyebabkan analisis yang dilakukan tidak mampu melihat bagaimana karakteristik dari setiap bank penyalur kredit juga menjadi faktor yang berperan penting dalam kinerja kredit UMKM.

REFERENSI

- [1] Y. Y. Lee, M. H. Dato Haji Yahya, M. S. Habibullah, and Z. Mohd Ashhari, "Non-performing loans in European Union: country governance dimensions," *J. Financ. Econ. Policy*, vol. 12, no. 2, pp. 209–226, 2020, doi: 10.1108/JFEP-01-2019-0027.
- [2] S. Urach *et al.*, "Statistical analysis of Goal Attainment Scaling endpoints in randomised trials," *Stat. Methods Med. Res.*, vol. 28, no. 6, pp. 1893–1910, 2019, doi: 10.1177/0962280218777896.
- [3] K. Staehr and L. Uusküla, "Macroeconomic and macro-financial factors as leading indicators of non-performing loans: Evidence from the EU countries," *J. Econ. Stud.*, vol. 48, no. 3, pp. 720–740, 2021, doi: 10.1108/JES-03-2019-0107.
- [4] R. D. Morris, H. Kang, and J. Jie, "The determinants and value relevance of banks' discretionary loan loss provisions during the financial crisis," *J. Contemp. Account. Econ.*, vol. 12, no. 2, pp. 176–190, 2016, doi: 10.1016/j.jcae.2016.07.001.
- [5] I. G. N. N. Mandala, C. B. Nawangpalipi, and F. R. Praktikto, "Assessing Credit Risk: An Application of Data Mining in a Rural Bank," *Procedia Econ. Financ.*, vol. 4, no. Icsmed, pp. 406–412, 2012, doi: 10.1016/s2212-5671(12)00355-3.
- [6] Y. Ren, "Python Machine Learning : Machine Learning and Deep Learning With Python , " *Int. J. Knowledge-Based Organ.*, vol. 11, no. 1, pp. 67–70, 2021.
- [7] Chris Albon, *Machine Learning with Python Cookbook Practical Solutions from Preprocessing to Deep Learning*. 2018.
- [8] S. Shalev-Shwartz and S. Ben-David, *Understanding machine learning: From theory to algorithms*. 2013.
- [9] D. Xu and M. Qu, "Energy, environmental, and economic evaluation of a CCHP system for a data center based on operational data," *Energy Build.*, vol. 67, pp. 176–186, 2013, doi:

Dari sisi teknis algoritma yang digunakan, *random forest* yang berbasis pohon keputusan memiliki keterbatasan dalam dua hal. Keterbatasan pertama adalah *random forest* tidak dapat secara efisien mempertimbangkan hubungan antara target dengan prediktor. Selain itu, keterbatasan yang dimiliki *random forest* adalah problem ekstrapolasi [9]. Ketika prediksi diperlukan untuk poin di luar dataset *training*. Salah satu alternatif yang dapat digunakan adalah menggunakan *regression-enhanced random forest* sebagaimana dikembangkan oleh [10].

10.1016/j.enbuild.2013.08.021.

- [10] H. Zhang, D. Nettleton, and Z. Zhu, "Regression-Enhanced Random Forests," *Sect. Stat. Learn. Data Sci.*, vol. 1, no. 1, pp. 1–13, 2019, [Online]. Available: <http://arxiv.org/abs/1904.10416>.