

**KEMENTERIAN KEUANGAN REPUBLIK INDONESIA
BADAN PENDIDIKAN DAN PELATIHAN KEUANGAN
POLITEKNIK KEUANGAN NEGARA STAN
TANGERANG SELATAN**



SKRIPSI

**ANALISIS PENILAIAN DEBITUR KREDIT USAHA RAKYAT
(KUR) DI INDONESIA BERBASIS *DATA MINING*
KLASIFIKASI**

Diajukan oleh:
Putri Rieza Chaniago
NPM 4132210042

Untuk Memenuhi Sebagian dari Syarat-syarat
Guna Mencapai Gelar Sarjana Terapan Akuntansi
Pada Politeknik Keuangan Negara STAN
2023

**KEMENTERIAN KEUANGAN REPUBLIK INDONESIA
BADAN PENDIDIKAN DAN PELATIHAN KEUANGAN
POLITEKNIK KEUANGAN NEGARA STAN
TANGERANG SELATAN
HALAMAN PERSETUJUAN**

PERSETUJUAN PROPOSAL SKRIPSI

NAMA	:	PUTRI RIEZA CHANIAGO
NOMOR POKOK MAHASISWA	:	4132210042
PROGRAM STUDI	:	DIPLOMA IV AKUNTANSI SEKTOR PUBLIK ALIH PROGRAM
BIDANG STUDI	:	SISTEM INFORMASI
JUDUL SKRIPSI	:	ANALISIS PENILAIAN DEBITUR KREDIT USAHA RAKYAT (KUR) DI INDONESIA BERBASIS DATA MINING KLASIFIKASI

Mengetahui	Menyetujui
Ketua Program Studi,	Dosen Pembimbing,
Puji Wibowo, Ak., C.A., MIDECA NIP 19750925 199602 1 001	Agung Nugroho, S.E., Ak., MBIT NIP 19750927 199502 1 002

**KEMENTERIAN KEUANGAN REPUBLIK INDONESIA
BADAN PENDIDIKAN DAN PELATIHAN KEUANGAN
POLITEKNIK KEUANGAN NEGARA STAN
TANGERANG SELATAN**

PERNYATAAN LULUS UJIAN KOMPREHENSIF

NAMA	:	PUTRI RIEZA CHANIAGO
NOMOR POKOK MAHASISWA	:	4132210042
PROGRAM STUDI	:	DIPLOMA IV AKUNTANSI SEKTOR PUBLIK ALIH PROGRAM
BIDANG STUDI	:	SISTEM INFORMASI
JUDUL SKRIPSI	:	ANALISIS PENILAIAN DEBITUR KREDIT USAHA RAKYAT (KUR) DI INDONESIA BERBASIS DATA MINING KLASIFIKASI

Tangerang Selatan, 04 Agustus 2023

1. Ketua Penguji
Nama Lengkap dan Gelar
NIP

2. Anggota Penguji/Pembimbing
Nama Lengkap dan Gelar
NIP

3. Anggota Penguji
Agung Nugroho, S.E., Ak., MBIT
NIP 19750927 199502 1 002

**KEMENTERIAN KEUANGAN REPUBLIK INDONESIA
BADAN PENDIDIKAN DAN PELATIHAN KEUANGAN
POLITEKNIK KEUANGAN NEGARA STAN
TANGERANG SELATAN**

PERNYATAAN KEASLIAN SKRIPSI

NAMA	:	PUTRI RIEZA CHANIAGO
NOMOR POKOK MAHASISWA	:	4132210042
PROGRAM STUDI	:	DIPLOMA IV AKUNTANSI SEKTOR PUBLIK ALIH PROGRAM
BIDANG STUDI	:	SISTEM INFORMASI
JUDUL SKRIPSI	:	ANALISIS PENILAIAN DEBITUR KREDIT USAHA RAKYAT (KUR) DI INDONESIA BERBASIS <i>DATA MINING</i> KLASIFIKASI

Dengan ini menyatakan bahwa sesungguhnya Skripsi ini adalah hasil tulisan saya sendiri dan/atau tidak terdapat bagian atau keseluruhan tulisan yang saya salin atau tiru tanpa memberikan pengakuan pada penulis aslinya. Bila terbukti melakukan plagiarisme, saya siap dinyatakan tidak lulus dan dicabut gelar yang telah diberikan.

Tangerang Selatan, 04 Agustus 2023

Yang memberi pernyataan,

Putri Rieza Chaniago
NPM 4132210042

KATA PENGANTAR

Puji syukur kehadirat Allah SWT, atas berkat rahmat dan hidayah-Nya penulis dapat menyelesaian skripsi yang berjudul “Analisis Penilaian Debitur Kredit Usaha Rakyat (KUR) di Indonesia Berbasis *Data Mining* Klasifikasi”. Dalam skripsi ini dibahas mengenai model *data mining* klasifikasi dengan memanfaatkan *database* yang ada pada SIKP untuk penilaian debitur KUR agar lebih objektif menggunakan tiga algoritma yaitu *k-Nearest Neighbors* (kNN), *Random Forest*, dan *Decision Tree* berdasarkan tahapan proses CRISP-DM. Adapun penulisan skripsi ini untuk memenuhi kewajiban perkuliahan semester 10 dan syarat kelulusan gelar sarjana terapan akuntansi pada Politeknik Keuangan Negara STAN tahun akademik 2022/2023. Dalam menyelesaikan penyusunan skripsi ini penulis mendapat dukungan dan bantuan dari berbagai pihak, berkenaan dengan hal tersebut penulis mengucapkan terima kasih kepada:

1. Orang tua dan adik penulis yang senantiasa memberikan doa, dukungan, dan kekuatan kepada penulis sehingga penulis dapat menyelesaikan skripsi ini.
2. Bapak Agung Nugroho, selaku dosen pembimbing yang telah meluangkan waktu untuk membimbing dan memberikan wawasan ilmu sehingga penulis dapat menyusun skripsi ini dengan lebih baik.
3. Bapak Ahmad Iqbal Zakyuddin, selaku dosen penulis yang juga telah meluangkan waktu untuk membimbing, menyediakan data, serta memberikan wawasan ilmu sehingga penulis dapat menyusun skripsi ini dengan lebih baik.

4. Seluruh dosen Politeknik Keuangan Negara STAN yang telah memberikan ilmunya kepada penulis sehingga penulis dapat terapkan dalam skripsi ini maupun dalam dunia kerja nantinya.
5. Maldin Kurniadireja, Dina Khairunnisa, Elzami Haqie Ednoer, Tyara Laritza Gistiani, Winanda Setyaning Kridantika selaku rekan dekat penulis yang senantiasa memberikan dukungan sehingga penulis dapat menyelesaikan skripsi ini.
6. Rekan-rekan DIV Alih Program Angkatan Tahun 2021 yang telah bersama-sama penulis menjalani setiap tahapan perkuliahan sehingga penulis bisa sampai tahap ini.
7. Semua pihak yang penulis belum dapat sampaikan satu per satu yang telah memberikan dukungan dan bantuan dalam penelitian ini.

Penulis menyadari dalam penyusunan skripsi ini masih banyak kekurangan dan keterbatasan. Oleh karena itu, saran dan masukan dapat membantu penulis menjadi lebih baik kedepannya. Semoga skripsi ini memiliki manfaat selain bagi penulis juga bagi pembaca.

Tangerang Selatan, 04 Agustus 2023

Putri Rieza Chaniago

DAFTAR ISI

HALAMAN JUDUL	i
HALAMAN PERSETUJUAN	ii
PERNYATAAN LULUS UJIAN KOMPREHENSIF	iii
KATA PENGANTAR	v
DAFTAR ISI	vii
DAFTAR TABEL	ix
DAFTAR GAMBAR	x
DAFTAR LAMPIRAN	xi
ABSTRAK	xii
BAB I PENDAHULUAN	1
1.1 Latar Belakang	1
1.2 Rumusan Masalah Penelitian	2
1.3 Tujuan Penelitian.....	2
1.4 Ruang Lingkup Penelitian.....	2
1.5 Manfaat Penelitian.....	3
1.6 Sistematika Penulisan Skripsi	3
BAB II LANDASAN TEORI	5
2.1 Kredit Usaha Rakyat (KUR)	5
2.2 Usaha Mikro, Kecil, dan Menengah (UMKM)	6
2.3 Penilaian Kredit.....	6
2.4 Kolektibilitas	6
2.5 Data Mining.....	7
2.6 CRISP-DM	7
2.7 Kerangka Pemikiran.....	8
BAB III METODE PENELITIAN	9
3.1 Metodologi Penelitian	9
3.2 Jenis dan Sumber Data Penelitian	9
3.3 Teknik Pengumpulan dan Pengolahan Data	10
3.4 Sarana Pengolahan Data.....	10
BAB IV HASIL PENELITIAN DAN PEMBAHASAN	11

4.1 Bussiness Understanding.....	11
4.2 Data Understanding.....	11
4.2.1 Memilih Data.....	11
4.2.2 Pengumpulan Data	12
4.2.3 Deskripsi Data	12
4.2.4 Kualitas Data	14
4.3 Data Preparation	14
4.3.1 Penanganan Missing Value.....	14
4.3.2 Penanganan Duplikasi Data	15
4.3.3 Penanganan Tipe Data.....	15
4.3.4 Transformasi Data	15
4.3.5 Exploratory Data Analysis (EDA).....	16
BAB V SIMPULAN, SARAN, DAN KETERBATASAN 5	17
5.1 Simpulan.....	17
5.2 Keterbatasan Penelitian	18
5.3 Saran.....	19
DAFTAR PUSTAKA	20
LAMPIRAN	22

DAFTAR TABEL

Tabel II-1 Skema KUR (dalam juta rupiah)	5
--	---

DAFTAR GAMBAR

Gambar I-1 Grafik Jumlah Unit UMKM dan UB di Indonesia 2010-2019	1
Gambar II-1 Kerangka Pemikiran	8
Gambar IV-1 Sebaran Debitur Berdasarkan Tahun Akad	13

DAFTAR LAMPIRAN

Lampiran 1 Deskripsi Kolom/Atribut yang Digunakan dalam Penelitian 22

Lampiran 2 Hasil Skor Feature Importance dengan Random Forest 23

ABSTRAK

Dalam rangka mengatasi hambatan akses pembiayaan bagi UMKM pada lembaga keuangan, pemerintah Indonesia menginisiasi program Kredit Usaha Rakyat dengan penatausahaanya menggunakan sistem terintegrasi yaitu Sistem Informasi Kredit Program (SIKP). Berdasarkan RPJMN 2020-2024, diharapkan KUR dapat mendorong UMKM naik kelas/graduasi. Namun graduasi UMKM di Indonesia masih rendah dan tidak mengalami perubahan dari tahun 2010 - 2019. Salah satu penyebab adalah bank penyalur KUR dalam penilaian calon debitur KUR diterima atau ditolak masih subjektif, sehingga diperlukan *tools* penilaian calon debitur KUR. Berkenaan hal tersebut, penelitian ini membangun model *data mining* klasifikasi dengan memanfaatkan *database* yang ada pada SIKP untuk penilaian debitur KUR dengan tiga algoritma yaitu *k-Nearest Neighbors* (kNN), *Random Forest*, dan *Decision Tree* berdasarkan tahapan proses CRISP-DM agar lebih objektif. Hasil penelitian menunjukkan kinerja model yang paling baik yaitu *Decision Tree* dengan nilai akurasi sebesar 97,33%, *F1 score* sebesar 50,80%, dan ROC/AOC sebesar 50,80%.

Kata Kunci: KUR, *data mining*, klasifikasi, penilaian kredit, *k-Nearest Neighbors*, *Random Forest*, *Decision Tree*

BAB I

PENDAHULUAN

1.1 Latar Belakang

Selama bertahun-tahun usaha mikro, kecil, dan menengah (UMKM) mendominasi jumlah unit dunia usaha di Indonesia sebesar 99,99% dibandingkan usaha besar (UB) sebesar 0,01% sebagaimana pada *Gambar I.1*. Begitu juga dengan jumlah tenaga kerja UMKM yang mencapai 119,6 juta (96,9%) dibandingkan usaha besar sebanyak 3,8 juta (3,1%) pada tahun 2019 (ADB, 2022). sumber inovasi, memperluas basis pajak, serta meningkatkan daya saing.

Gambar I-1 Grafik Jumlah Unit UMKM dan UB di Indonesia 2010-2019



Sumber: Diolah dari ADB Asia SME Monitor 2022 (2022)

Diharapkan dengan *data mining* klasifikasi tersebut dapat menjadi referensi bagi Kementerian Koordinator Bidang Perekonomian untuk penilaian calon debitur

KUR, referensi dalam pengambilan keputusan terkait KUR, serta bagi Kementerian Keuangan sebagai referensi dalam pengembangan aplikasi Sistem Informasi Kredit Program (SIKP), sehingga penyaluran KUR lebih tepat sasaran yang pada akhirnya meningkatkan jumlah UMKM naik kelas. Oleh karena itu, penulis menyusun penelitian dengan judul “Analisis Penilaian Debitur Kredit Usaha Rakyat (KUR) di Indonesia Berbasis *Data Mining* Klasifikasi”.

1.2 Rumusan Masalah Penelitian

Rumusan masalah dalam penelitian ini yaitu:

- 1 Apa faktor-faktor yang mempengaruhi tingkat pengembalian KUR?
- 2 Bagaimana kinerja model *data mining* dalam mengklasifikasikan debitur KUR?

1.3 Tujuan Penelitian

Tujuan penelitian ini yaitu:

1. Untuk mengetahui faktor-faktor yang mempengaruhi tingkat pengembalian KUR dalam rangka pembangunan model *data mining* klasifikasi debitur KUR.
2. Untuk mengetahui kinerja model *data mining* dalam mengklasifikasi debitur KUR.

1.4 Ruang Lingkup Penelitian

Penelitian ini membahas penilaian debitur KUR menggunakan *data mining* klasifikasi dengan tahapan berdasarkan *The Cross-Industry Standard Process for Data Mining* (CRISP-DM). Penelitian ini menggunakan dataset debitur KUR Indonesia hasil ekstraksi dari *database* aplikasi Sistem Informasi Kredit Program (SIKP) Kementerian Keuangan dengan akad KUR dari tahun 2015 sampai dengan tahun 2021.

1.5 Manfaat Penelitian

Penelitian ini diharapkan dapat memberikan manfaat sebagai berikut:

1. Manfaat akademis, penelitian ini diharapkan dapat memperkaya referensi dalam pemodelan *data mining* klasifikasi debitur KUR, berkenaan masih terbatas penelitian yang mengkaji terkait hal ini.
2. Manfaat praktis, penelitian ini diharapkan dapat menjadi referensi bagi Kementerian Koordinator Bidang Perekonomian dalam penilaian calon debitur KUR maupun pengambilan keputusan terkait KUR, serta bagi Kementerian Keuangan, sebagai referensi dalam pengembangan aplikasi Sistem Informasi Kredit Program (SIKP).

1.6 Sistematika Penulisan Skripsi

Skripsi ini terdiri dari 5 bab dengan sistematika serta isi sebagai berikut:

BAB I PENDAHULUAN

Pada bagian ini membahas mengenai latar belakang penelitian, rumusan masalah penelitian, tujuan penelitian, ruang lingkup penelitian, manfaat penelitian, dan sistematika penulisan skripsi.

BAB II LANDASAN TEORI

Pada bagian ini membahas mengenai teori yang relevan dalam mendukung penelitian, penelitian terdahulu, dan kerangka pemikiran.

BAB III METODE PENELITIAN

Pada bagian ini membahas mengenai metodologi penelitian, jenis dan sumber data penelitian, teknik pengumpulan dan pengolahan data, atribut/variabel penelitian, dan sarana pengolahan data.

BAB IV HASIL PENELITIAN DAN PEMBAHASAN

Pada bagian ini membahas mengenai hasil pelaksanaan tahapan penelitian sebagaimana CRISP-DM serta penjelasan yang relevan.

BAB V SIMPULAN, SARAN, DAN KETERBATASAN

Pada bagian ini akan disampaikan simpulan, saran, dan keterbatasan dalam penelitian.

BAB II

LANDASAN TEORI

2.1 Kredit Usaha Rakyat (KUR)

Kredit usaha rakyat (KUR) merupakan program pemerintah yang ditujukan untuk UMKM produktif baik individu, badan usaha, maupun kelompok usaha yang *feasible/layak* namun belum *bankable*, dengan skema subsidi bunga (Kemenko Bidang Perekonomian, 2019). Menurut Kementerian Koordinator Bidang Perekonomian (2019), usaha produktif adalah usaha yang menghasilkan barang/jasa dalam rangka memberikan *value added* (nilai tambah) serta meningkatkan kesejahteraan.

Tabel II-1 Skema KUR (dalam juta rupiah)

No.	UMKM	2015 – 2016	2017 – 2018	Plafon KUR 2019	2020	2021
1	KUR Super Mikro	Tidak Ada	Tidak Ada	Tidak Ada	0 – 10	0 – 10
2	KUR Mikro	0 – 25	0 – 25	0 – 50	>10 – 50	>10 – 100
3	KUR Kecil	>25 – 500	>25 – 500	>50 – 500	>50 – 500	>100 – 500
4	KUR Penempatan Pekerja Migran Indonesia	0 – 25	0 – 25	0 – 25	0 – 25	0 – 100
5	KUR Khusus	Tidak Ada	>25 – 500	0 – 500	0 – 500	0 – 500

Sumber: Diolah dari Kemenko Bidang Perekonomian

KUR diberikan dalam bentuk kredit modal kerja dan/atau kredit investasi berdasarkan 5 skema atau jenis penyaluran KUR. Lima skema tersebut yaitu KUR super mikro, KUR mikro, KUR kecil, KUR penempatan pekerja migran Indonesia

(PMI), dan KUR khusus. Masing-masing skema KUR tersebut memiliki plafon kredit yang berbeda. Plafon kredit tersebut telah mengalami beberapa kali perubahan.

2.2 Usaha Mikro, Kecil, dan Menengah (UMKM)

Usaha mikro, kecil, dan menengah (UMKM) memiliki berbagai definisi yang berbeda di berbagai negara (PRKKEK-BRIN, 2022) sesuai dengan kebutuhan serta praktis negara tersebut (Kapizionis, 2015 dikutip dalam PRKKEK-BRIN, 2022). Di Indonesia, usaha mikro, kecil, dan menengah salah satunya didefinisikan berdasarkan kriteria modal dan omset tahunan sebagaimana diatur dalam Peraturan Pemerintah (PP) Nomor 7 Tahun 2021 tentang Kemudahan, Pelindungan, dan Pemberdayaan Koperasi dan Usaha Mikro, Kecil, dan Menengah.

2.3 Penilaian Kredit

Perbankan memiliki fungsi utama sebagai perantara keuangan (*financial intermediary*) antara pihak yang kekurangan/membutuhkan dana (*lack of funds*) dengan pihak yang memiliki kelebihan dana (*surplus of funds*). Berkenaan dengan dana masyarakat dipercayakan kepada bank, maka dalam menjalankan bisnisnya bank perlu hati-hati dan melakukan manajemen risiko untuk melindungi dana masyarakat tersebut. Hal tersebut sudah menjadi kewajiban bank sebagaimana diatur dalam pasal 20A Undang-Undang Republik Indonesia Nomor 4 Tahun 2023 tentang Pengembangan dan Penguatan Sektor Keuangan.

2.4 Kolektibilitas

Prinsip kehati-hatian dan manajeman risiko dilakukan untuk mencegah terjadinya kredit bermasalah atau dikenal dengan istilah *Non-Performing Loan*

(NPL) (Disemadi, 2019). *Non-Performing Loan* (NPL) menjadi salah satu indikator dalam menilai kesehatan aset bank umum. Berdasarkan Peraturan Bank Indonesia Nomor 15/ PBI/2013 tentang Penetapan Status dan Tindak Lanjut Pengawasan Bank Umum Konvensional, kredit bermasalah adalah kredit dengan status kolektibilitas kurang lancar, diragukan, dan macet.

2.5 Data Mining

Data mining merupakan proses analisis untuk menemukan informasi/pola yang berguna (Pradnyana *et al.*, 2020; Liu, 2002) pada data yang berjumlah besar (Larose & Larose, 2014) secara otomatis atau semi otomatis (Liu, 2002) dengan bantuan teknologi asosiasi pola, statistik, dan matematika (Larose, 2006 dikutip dalam Rahayu, 2020) yang mulai dikenalkan pada tahun 1990 (Pradnyana *et al.*, 2020). *Data mining* memadukan metode analisis data tradisional dengan algoritma yang canggih dalam pengolahan data yang besar (Tan *et al.*, 2019; Pradnyana *et al.*, 2020).

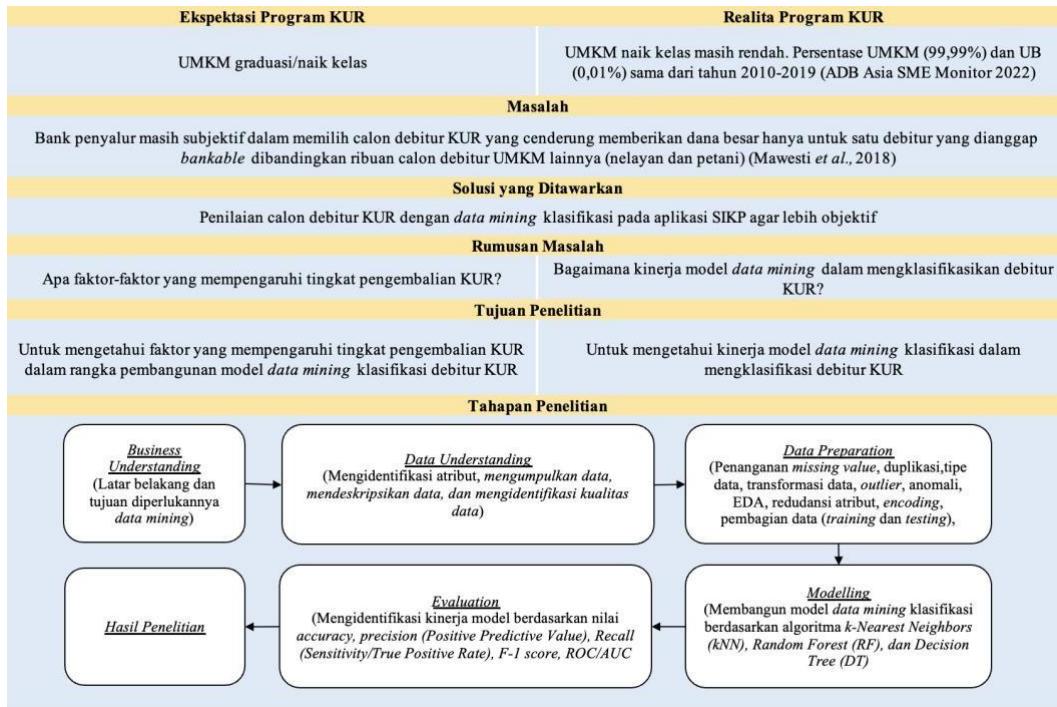
2.6 CRISP-DM

Terdapat beberapa model proses dalam data mining seperti Knowledge Discovery Databases (KDD), Sample, Explore, Modify, Model and Assess (SEMMA), dan The Cross-Industry Standard Process for Data Mining (CRISPDM). CRISP-DM menjadi model yang paling banyak digunakan dalam data mining (Piatetsky, 2014; Schröer *et al.*, 2021). Oleh karena model tersebut menyediakan tahapan business understanding dan deployment sehingga bisa menangkap keadaan sebenarnya (Pyvovar *et al.*, 2019 dikutip dalam Clancy & Bruton, 2023), iteratif, reversibel (Palacios *et al.*, 2017 dikutip dalam Clancy & Bruton, 2023), mudah digunakan, dan terstruktur (Schröer *et al.*, 2021).

2.7 Kerangka Pemikiran

Kerangka pemikiran pada penelitian ini sebagaimana *Gambar II.1.*

Gambar II-1 Kerangka Pemikiran



Sumber: Diolah Penulis

BAB III

METODE PENELITIAN

3.1 Metodologi Penelitian

Sebagaimana Saunders *et al.* (2006), metodologi penelitian menjelasakan bagaimana penelitian akan dilakukan. Dalam penelitian ini, untuk dapat menjawab pertanyaan serta mencapai tujuan penelitian dibutuhkan penyelidikan empiris melalui pengamatan terhadap data histori pinjaman debitur KUR atau kasus-kasus pinjaman debitur KUR sebelumnya yang berasal dari *database* SIKP serta studi literatur seperti jurnal untuk mengetahui faktor-faktor yang mempengaruhi tingkat pengembalian KUR yang telah dikaji oleh para peneliti. Menurut Robson (2002) dikutip dalam Saunders *et al.* (2007), strategi penelitian dengan melakukan penyelidikan empiris terhadap fenomena kontemporer dalam konteks kehidupan nyata menggunakan berbagai sumber bukti yaitu studi kasus. Berkenaan dengan hal tersebut maka strategi penelitian utama yang digunakan penulis adalah studi kasus, dengan juga dilakukan studi literatur untuk mengetahui faktor-faktor yang mempengaruhi tingkat pengembalian KUR.

3.2 Jenis dan Sumber Data Penelitian

Jenis data dalam penelitian ini adalah data sekunder yang didapatkan dari *database* aplikasi Sistem Informasi Kredit Program (SIKP) Kementerian Keuangan

berupa Dataset Debitur KUR tahun 2015 - 2021. Disebut data sekunder karena data yang penulis gunakan telah tersedia (Oktaviani, 2019) pada *database* aplikasi SIKP dan tidak dikumpulkan oleh penulis (Azwar, 2004 dikutip dalam Lesmana *et al.*, 2017) melainkan telah dikumpulkan oleh Kementerian Keuangan.

3.3 Teknik Pengumpulan dan Pengolahan Data

Teknik pengumpulan data yang penulis gunakan adalah ekstraksi data pada *database* aplikasi SIKP Kementerian Keuangan. Esktraksi data adalah proses mengambil data mentah dari suatu sumber (Yang, 2023) dan mereplikasi ke tempat lainnya (Stitch, 2022). Teknik pengumpulan data tersebut digunakan oleh karena data yang penulis butuhkan terdapat pada *database* aplikasi SIKP Kementerian Keuangan, sehingga perlu dilakukan ekstraksi data. Kemudian, teknik pengolahan data berupa *data mining* klasifikasi dengan tahapan proses pemodelan menggunakan tahapan pada CRISP-DM.

3.4 Sarana Pengolahan Data

Pengolahan data dilakukan dengan aplikasi Google Collab. Kemudian, perangkat yang digunakan penulis dalam penelitian ini adalah Laptop Macbook Air M1 2020 dengan spesifikasi Chip Apple M1, 8 GB RAM, CPU dan GPU 8-core, dan sistem operasi MacOS Ventura versi 13.4.1.

BAB IV

HASIL PENELITIAN DAN PEMBAHASAN

4.1 Bussiness Understanding

Dalam rangka mengatasi hambatan akses pembiayaan bagi UMKM pada lembaga keuangan, pemerintah Indonesia menginisiasi program Kredit Usaha Rakyat pada tahun 2007. KUR menggunakan skema subsidi bunga/marjin. Dengan skema subsidi bunga/marjin pemerintah membayar sebagian bunga yang seharusnya ditanggung debitur yaitu sebesar selisih bunga yang diterima penyalur KUR dengan yang dibebankan kepada debitur KUR.

4.2 Data Understanding

Data understanding merupakan tahapan untuk memahami data yang akan ditambang (*mining*), dengan tujuan agar familier dengan dengan data, mengetahui masalah kualitas data, maupun mendeteksi adanya informasi menarik dari data yang dimiliki untuk diteliti lebih lanjut (Chapman *et al.*, 2000). Secara rinci berikut proses *data understanding* pada penelitian ini.

4.2.1 Memilih Data

Berdasarkan hasil studi literatur, didapatkan faktor-faktor yang mempengaruhi tingkat pengembalian KUR sebagaimana terdapat pada *Tabel IV.1*.

Beberapa peneliti menggunakan karakteristik individu/personal (seperti usia, pendidikan, dan lain sebagainya), karakteristik usaha (seperti jenis/sektor usaha, lama usaha, dan lain sebagainya), dan karakteristik pinjaman/kredit (seperti jumlah kredit, agunan, dan lain sebagainya) sebagai *variabel* dalam menilai faktor-faktor yang mempengaruhi tingkat pengembalian KUR sebagaimana yang dilakukan dalam penelitian Putra *et al.* (2023), Abadi (2014), Rochmawati (2014), dan Lubis & Rachmina (2011) dan kredit lainnya yang dilakukan dalam penelitian Kusumaningtyas (2017) dan Arinta (2014).

4.2.2 Pengumpulan Data

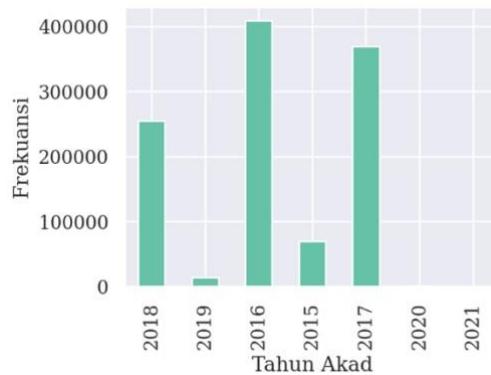
Dataset diperoleh dari *database* SIKP Kementerian Keuangan dalam format .csv. Atribut yang digunakan penulis berasal dari tiga tabel yang berbeda pada *database* SIKP. Atribut tanggal lahir, jenis kelamin, status pernikahan, pendidikan, pekerjaan, kabupaten/kota, mulai usaha, modal usaha, jumlah pekerja, pola KUR *linkage*, agunan, dan subsidi sebelumnya berasal dari tabel debitur. Selanjutnya, atribut bank penyalur KUR, status akad, penjamin, tanggal akad, tanggal jatuh tempo, nilai akad, nilai dijamin, sektor, negara tujuan, tingkat bunga berasal dari tabel akad. Kemudian, atribut koletibilitas berasal dari tabel transaksi.

4.2.3 Deskripsi Data

Dataset merupakan debitur KUR yang tersimpan dalam *database* SIKP Kementerian Keuangan sejak diluncurkan tahun 2017 dengan sebaran debitur berdasarkan tahun akad sebagaimana diagram batang pada *Gambar IV.1*. Debitur pada tahun 2020 dan 2021 jumlahnya sangat sedikit sehingga tidak terlihat pada diagram batang. Dataset terdiri dari 23 atribut dengan 1.115.797 *records*. Penjelasan

masing-masing atribut/kolom sebagaimana pada *Lampiran 1*. Tampilan sebagian isi dataset dapat dilihat pada *Gambar IV.2*. Dataset merupakan hasil gabungan tabel debitur, tabel akad, dan tabel transaksi (tabel utama penggabungan yaitu tabel transaksi) dan hasil pengambilan sampel sebanyak 10% dari populasi berdasarkan atribut kolektibilitas, sektor, pendidikan, dan pekerjaan (*stratified sampling*). Pengambilan *sample* 10% berkenaan keterbatasan sumber daya penulis dalam mengolah seluruh populasi.

Gambar IV-1 Sebaran Debitur Berdasarkan Tahun Akad



Sumber: Diolah Penulis

Walaupun dataset merupakan debitur KUR yang tersimpan dalam *database* SIKP Kementerian Keuangan sejak diluncurkan tahun 2017, namun tidak sepenuhnya lengkap data seluruh debitur KUR. Ada beberapa kemungkinan penyebab data debitur KUR tidak lengkap, diantaranya penyalur KUR belum lengkap dalam mengunggah data KUR pada aplikasi SIKP baik atas satu tabel maupun ketiga tabel (tabel debitur, tabel akad, dan tabel transaksi), adanya ketidaksamaan tanggal tarikan data setiap tabel, karena memang penambahan debitur tidak terlalu banyak, maupun penyebab lainnya.

4.2.4 Kualitas Data

Berdasarkan Gambar IV.3, dari total 23 atribut, terdapat 18 atribut yang memiliki tipe data numerik (*float64* dan *int64*) dan 5 atribut yang memiliki tipe data *object*. Beberapa atribut memiliki tipe yang belum sesuai, seperti tanggal lahir yang seharusnya merupakan tanggal, dan atribut lainnya. Kemudian, terdapat *missing value* berdasarkan adanya ketidakseragaman jumlah kolom yang terisi, seperti salah satunya tanggal lahir berjumlah 1.115.797, sedangkan kabupaten/kota berjumlah

4.3 Data Preparation

Pada tahap *data preparation* merupakan tahapan untuk membuat dataset siap diproses pada tahap *modeling*. *Data preparation* perlu dilakukan, karena data mentah yang tidak lengkap, mengandung *outlier*, terduplikasi, dan hal lainnya (Larose & Larose, 2014) yang bisa menghambat kinerja model.

4.3.1 Penanganan Missing Value

Missing value merupakan sebagian informasi atribut yang tidak tersedia pada suatu *record*. Kehadiran *missing value* dapat menyebabkan data yang tidak akurat (Han *et al.*, 2012), karena kehilangan informasi yang bisa berpotensi signifikan dalam membentuk pola pada *modeling*, sehingga perlu ditangani. Terdapat 7 atribut yang memiliki *missing value* berdasarkan Gambar IV.4 yaitu kabupaten/kota, mulai usaha, modal usaha, jumlah pekerja, agunan, negara tujuan, dan tingkat bunga. *Missing value* pada atribut kabupaten/kota, modal usaha, jumlah pekerja, agunan, dan negara tujuan diisi nilainya dengan *k-Nearest Neighbors (kNN) Imputation*.

4.3.2 Penanganan Duplikasi Data

Duplikasi data dalam hal ini pada level *record* (Han *et al.*, 2012). Duplikasi data perlu ditangani, karena dapat menyebabkan *overweighting* nilai data pada *record* yang terduplikasi tersebut (Larose & Larose, 2014). Menurut Larose & Larose (2014), jika terdapat duplikasi, maka satu saja yang digunakan. Penulis melakukan pengecekan duplikasi data berdasarkan kesamaan seluruh atribut pada setiap *record* dengan perintah sebagaimana *Gambar IV.6* dan hasilnya terdapat duplikasi atas 21.787 *record*. Duplikasi data pada setiap *record* hanya digunakan satu *record* saja dengan perintah sebagaimana *Gambar IV.6*.

4.3.3 Penanganan Tipe Data

Tipe data/*data type* dalam *data mining* merupakan data yang memiliki karakteristik masing-masing, seperti tipe data numerik yang merupakan angka dan mengukur kuantitas, sedangkan data nominal atau bisa disebut kategori yang merupakan simbol atau nama atas sesuatu yang nilainya setara/tidak memiliki urutan tertentu (Han *et al.*, 2012), seperti hewan yaitu kucing, kerbau, dan zebra. Berkenaan dengan karakteristik yang berbeda akan mempengaruhi proses *preparation* yang digunakan, sehingga perlu disesuaikan.

4.3.4 Transformasi Data

Transformasi data dibutuhkan diantaranya untuk membuat data mudah dipahami model maupun efisiensi proses penambangan data (Han *et al.*, 2012). Atribut tanggal lahir, mulai usaha, tanggal akad, tanggal jatuh tempo, ditransformasi menjadi usia dalam satuan tahun (tanggal akad dan tanggal lahir), lama usaha dalam satuan tahun (mulai usaha dan tanggal lahir) dan jangka waktu

kredit dalam satuan bulan (tanggal jatuh tempo dan tanggal akad) untuk lebih mudah dipahami model dan menyesuaikan sebagaimana literatur faktor-faktor yang mempengaruhi pengembalian KUR. Perintah yang digunakan untuk membuat atribut baru tersebut sebagaimana *Gambar IV.8* dan berlaku sama untuk lama usaha dan jangka waktu kredit.

4.3.5 Exploratory Data Analysis (EDA)

Berdasarkan Larose & Larose (2014), mempertimbangkan untuk melakukan proses pembersihan data terlebih dahulu sebelum melakukan analisis data eksplorasi. Sebagaimana pada dataset penulis, terdapat *record* yang bukan merupakan KUR, terdapat atribut yang perlu diubah agar bisa diperoleh informasi yang lebih berguna, seperti usia, lama usaha, dan jangka waktu kredit dan lain sebagainya, sehingga *exploratory data analysis* dilakukan setelah data dibersihkan. Lebih lanjut, EDA bermanfaat untuk mengetahui hubungan antar atribut, mengidentifikasi hal menarik, maupun menemukan gagasan awal hubungan antara variabel prediktor dan variabel target.

BAB V

SIMPULAN, SARAN, DAN KETERBATASAN

5.1 Simpulan

Penelitian ini bertujuan untuk membuat model *data mining* klasifikasi debitur KUR agar dalam penilaian calon debitur KUR lebih objektif atas seluruh bank penyalur, sehingga penyaluran KUR lebih tepat sasaran yang pada akhirnya meningkatkan jumlah UMKM naik kelas. Berdasarkan penelitian dapat disimpulkan, sebagai berikut:

- 1) Faktor-faktor yang mempengaruhi tingkat pengembalian KUR berdasarkan studi literatur yaitu karakteristik individu/personal yang terdiri dari usia, pendidikan, pekerjaan, karakteristik usaha yang terdiri dari sektor usaha, lama usaha, dan karakteristik kredit yang terdiri dari jumlah kredit, jangka waktu kredit, agunan, dan tingkat bunga.
- 2) Berdasarkan *exploratory data analysis* ditemukan bahwa:
 - a. Atribut kategorikal yaitu jenis kelamin, status pernikahan, agunan, dan subsidi sebelumnya memiliki proporsi kredit bermasalah relatif seimbang pada setiap kategori sebagaimana analisis menggunakan *stacked bar*.

- b. Atribut pendidikan, proporsi terjadinya kredit bermasalah terbesar ada pada debitur dengan pendidikan dibawah diploma dan sarjana yaitu SD,
 - c. SMP, dan SMA. Kredit bermasalah terjadi lebih dari 50% pada masingmasing tingkat pendidikan tersebut, sebagaimana analisis menggunakan *stacked bar* dan *pie chart*.
- 3) Hasil *feature importance* dengan *random forest* untuk melihat peringkat atribut/variabel dari yang terpenting sampai yang kurang penting dalam membangun model, didapatkan 30 variabel teratas (sesuai urutan) berdasarkan nama kolom
- 4) Kinerja model yang paling baik diantara kNN, *random forest*, dan *decision tree* berdasarkan nilai *accuracy*, *F1 score*, dan ROC/AOC yaitu *Decision Tree* dengan nilai akurasi sebesar 97,33%, *F1 score* sebesar 50,80%, dan ROC/AOC sebesar 50,80%. Namun skor *F1 score* sebesar 50,80% atau 0,508 masih rendah untuk model bisa memprediksi dengan baik dikedua kelas.

5.2 Keterbatasan Penelitian

Dalam penyusunan penelitian ini masih terdapat kekurangan dan keterbatasan, sebagai berikut:

- 1) Data penelitian yang digunakan untuk membuat model berupa sampel berupa 10% bukan populasi, karena keterbatasan sumber daya untuk memproses seluruh populasi.
- 2) Tidak melakukan *hypetuning parameters* untuk meningkatkan kinerja model, karena keterbatasan keterbatasan sumber daya untuk memproses model.
- 3) Algoritma hanya tiga, karena keterbatasan waktu untuk dapat menyelesaikan pada skripsi ini.

5.3 Saran

Saran terkait hasil penelitian ini, diantaranya:

1. Bagi praktisi yaitu Kementerian Koordinator Bidang Perekonomian dan Kementerian Keuangan
 - a. Perlu memperbanyak informasi kuantitatif terkait debitur yang dikumpulkan pada SIKP seperti penghasilan, pengeluaran perbulan, dan lain sebagainya serta kajian lebih mendalam atas penanganan *imbalance class* untuk bisa mengoptimalkan penggunaan SIKP sebagai alat penilaian debitur KUR yang lebih objektif yaitu dengan *data mining* klasifikasi.
 - b. Perlu mempertimbangkan kajian lebih lanjut atas temuan:
 - 1) Kredit bermasalah dengan nilai akad diatas sekitar Rp30 Juta lebih sering terjadi pada debitur berusia antara sekitar 30 - 50 tahun.
 - 2) Kredit bermasalah dengan tingkat bunga 9,5% – 10,5% lebih sering terjadi pada KUR dengan jangka waktu kredit dibawah sekitar 12 bulan.
2. Bagi akademisi/peneliti selanjutnya, dapat menambahkan atribut kuantitaif, menggunakan berbagai teknik penanganan *imbalance class*, hypertuning parameters, dan beragam algoritma untuk bisa meningkatkan kinerja model *data mining* klasifikasi KUR

DAFTAR PUSTAKA

- Abadi, A. F., & Purwanti, E. Y. (2014). *Analisis Pengaruh Karakteristik Peminjam, Besar Pinjaman, Jenis Usaha, dan Lama Usaha Terhadap Tingkat Kelancaran Pengembalian Kredit Usaha Rakyat (KUR) Mikro (Studi Kasus pada Debitur KUR Mikro BRI Unit Kendal Kota)* (Doctoral dissertation, Fakultas Ekonomika dan Bisnis). <http://eprints.undip.ac.id/43426/>
- Abhigyan. (14 Juni 2020). Importance of Dimensionality Reduction!!.. Medium. Diakses tanggal 18 Mei 2023, dari <https://medium.com/analyticsvidhya/importance-of-dimensionality-reduction-d6a4c7289b92>
- Allwring 2022. (12 Juli 2022) F1 Score vs AUC, Which Is The Best Classification Metric? Diakses tanggal 30 Mei 2023, dari [https://stephenallwright.com/f1 score- vsauc/#:~:text=F1%20should%20be%20used%20for,F1%20in%20compar is on%20with%20AUC](https://stephenallwright.com/f1-score-vs-auc/#:~:text=F1%20should%20be%20used%20for,F1%20in%20comparis on%20with%20AUC).
- Asian Development Bank. (2022). Asian Development Bank (ADB) Asia SME Monitor 2022 – Indonesia. <https://data.adb.org/media/10461/download>
- AssemblyAI. (22 Juli 2022). Neural Networks Explained in 60 Seconds!. Diakses tanggal 16 Mei 2023, dari <https://www.youtube.com/shorts/kQl45ophSVQ>
- Ayungtyas, D. A. (2017). *Klasifikasi Menggunakan Metode Regresi Logistik Dan Support Vector Machine* (Doctoral dissertation, Universitas Brawijaya). <http://repository.ub.ac.id/id/eprint/3988>
- Budiarsih, R., & Sony, H. (2022). Persepsi Pelaku UMKM Terhadap Alternatif Model Pemungutan Pajak UMKM dengan Pemanfaatan Teknologi
- Chapman, P., Clinton, J., Kerber, R., Khabaza, T., Reinartz, T., Shearer, C., & Wirth, R. (2000). CRISP-DM 1.0: Step-by-Step Data Mining Guide. *SPSS inc*, 9(13), 1-73. <https://www.kde.cs.uni-kassel.de/wpcontent/uploads/lehre/ws2012-13/kdd/files/CRISPWP-0800.pdf>
- Cismondi, F., Fialho, A. S., Vieira, S. M., Reti, S. R., Sousa, J. M., & Finkelstein, S. N. (2013). Missing data in medical databases: Impute, delete or classify?. *Artificial intelligence in medicine*, 58(1), 63-72. <https://doi.org/10.1016/j.artmed.2013.01.00>
- Jiang, Y. (2009, March). Credit scoring model based on the decision tree and the simulated annealing algorithm. In *2009 WRI world congress on computer*

science and information engineering (Vol. 4, pp. 18-22). IEEE.
<https://doi.org/10.1109/csie.2009.481>

Kasmir, S.E., M.M. (2017). *Manajemen Perbankan: Edisi Revisi*. Rajawali Pers.

Kusumaningtyas, I. (2017). Pengaruh Karakteristik Personal, Karakteristik Usaha, Karakteristik Kredit, dan Jaminan terhadap Tingkat Pengembalian Kredit di BPR Nusamba Adiwerna Kabupaten Tegal Tahun 2016. *Multiplier: Jurnal Magister Manajemen*, 1(2). <https://doi.org/10.24905/mlt.v1i2.772>

Larose, D. T. (2006). *Data mining methods & models*. John Wiley & Sons.

Larose, D. T., & Larose, C. D. (2014). *Discovering Knowledge In Data: An Introduction to Data Mining* (Vol. 4). John Wiley & Sons.

Wang, Y., & Yao, Y. (2002). Market reforms, Technological Capabilities and The Performance of Small Enterprises in China. *Small Business Economics*, 18, 195-209. <https://doi.org/10.1023/A:1015191608288>

Wulandari, N. W. F., Yuesti, A., & Sudiartana, I. M. (2021). Analisis Faktor-Faktor Yang Mempengaruhi Kolektibilitas Kredit Usaha Rakyat (KUR) Mikro Bri Unit Renon. *KARMA (Karya Riset Mahasiswa Akuntansi)*, 1(4), 1145-1152. <https://e-journal.unmas.ac.id/index.php/karma/article/view/3521>

Yang, Z. (16 Februari 2023). Streamlining Your Business Operations with Data Extraction. Coupler.io. Diakses tanggal 17 Mei 2023, dari https://blog.coupler.io/data-extraction-guide/#What_is_data_extraction

LAMPIRAN

Lampiran 1 Deskripsi Kolom/Atribut yang Digunakan dalam Penelitian

No	Nama Kolom pada SIKP	Nama Atribut	Tipe Data	Deskripsi
1	TGL_LAHIR	Tanggal Lahir	Tanggal	Tanggal lahir debitur dengan format ddmmmyy
2	JNS_KELAMIN	Jenis Kelamin	Kategori	Jenis kelamin bagi debitur individu, sedangkan bagi badan usaha berupa keterangan badan usaha dalam bentuk kode. Referensi kode sebagai berikut: 1 = Laki-Laki, 2 = Perempuan, dan 9 = Badan Usaha
3	MARITAS_STS	Status Pernikahan	Kategori	Status pernikahan/marital status bagi debitur individu, sedangkan bagi badan usaha berupa keterangan badan usaha dalam bentuk kode. Referensi kode sebagai berikut: 0 = Tidak Kawin, 1 = Kawin, dan 9 = Badan Usaha
4	PENDIDIKAN	Pendidikan	Kategori	Jenjang pendidikan terakhir debitur dalam bentuk kode. Referensi kode sebagai berikut: 1 = SD, 2 = SMP, 3 = SMU, 4 = Diploma, 5 = Sarjana, 6 = Lainnya, dan 9 = Badan Usaha.
5	PEKERJAAN	Pekerjaan	Kategori	Pekerjaan debitur dalam bentuk kode. Referensi kode sebagai berikut: 1 = PNS, 2 = TNI/Polri, 3 = Pensiunan/Purnawirawan, 4 = Profesional, 5 = Karyawan Swasta, 6 = Wiraswasta, 7 = Petani, 8 = Pedagang, 9 = Nelayan, dan 99 = Lain-Lain/Badan Usaha

Lampiran 2 Hasil Skor Feature Importance dengan Random Forest

