

IMPLEMENTASI FRAMEWORK CRISP-DM UNTUK PROSES DATA MINING APLIKASI CREDIT SCORING PT. XYZ

Dwi Ahmad Dzulhijjah*, Mega Bagus Herlambang, Mohamad Haifan
Program Studi Profesi Insinyur, Institut Teknologi Indonesia, Tangerang Selatan

*Penulis korespondensi: dwiamaddzulhijjah@iti.ac.id

ABSTRAK

Proyek pengembangan *Artificial Intelligence (AI) Credit Guarantee and Collateral* untuk aplikasi *credit scoring* bertujuan meningkatkan ketepatan penilaian risiko kredit dengan memanfaatkan analisis jaminan dan agunan. Pemahaman bisnis yang mendalam mencakup tujuan bisnis, kebutuhan lembaga keuangan, dan fokus pada jaminan perorangan, jaminan kebendaan, dan analisis agunan. Metodologi CRISP-DM diterapkan mulai dari Pemahaman Bisnis, Pemahaman Data, Persiapan Data, hingga Pemodelan dan Pelatihan Data untuk menghasilkan model-machine learning yang handal. Penelitian ini berkontribusi dalam khazanah pengembangan *AI credit scoring* menggunakan metodologi CRISP-DM dengan memberikan gambaran secara luas pada tahap *modelling* dengan menyertakan perbandingan menyeluruh model klasifikasi menggunakan *library* Lazy Predict. Evaluasi kinerja model dengan menggunakan akurasi dan F1 Score mengidentifikasi tiga model terbaik: RandomForestClassifier, BaggingClassifier, dan XGBClassifier. Proses Deployment melibatkan penyimpanan model dan pembuatan API dengan Flask, diikuti pengujian API untuk memastikan keberhasilan implementasi. Keseluruhan, proyek ini memberikan solusi AI yang efisien untuk prediksi *credit score* berbasis lokasi, meningkatkan kecerdasan dan akurasi dalam pengambilan keputusan kredit.

Kata kunci: *credit score*, *crisp-dm*, data mining, *credit guarantee*, *lazy predict*.

1 PENDAHULUAN

Praktik memberikan kredit merupakan hal yang umum dan sangat penting dalam dunia keuangan, baik untuk lembaga keuangan maupun non-keuangan. Salah satu entitas keuangan yang secara rutin memberikan kredit adalah bank. Kredit memiliki peran yang signifikan dalam meningkatkan penjualan produk dan layanan yang ditawarkan oleh bank. Namun, di sisi lain, memberikan pinjaman atau kredit juga membawa risiko. Oleh karena itu, diperlukan suatu mekanisme yang dapat membantu bank dalam memilih nasabah yang memiliki kapasitas untuk membayar kembali kredit yang diberikan. Salah satu metode yang umum digunakan untuk menilai kelayakan nasabah dalam mendapatkan kredit adalah melalui *credit scoring* (Saputri & Ernita., 2019).

Credit scoring adalah suatu proses yang melibatkan pengumpulan, analisis, dan pengklasifikasian berbagai variabel terkait dengan kredit untuk menilai keputusan kredit. Penilaian ini didasarkan pada perbandingan antara informasi debitur saat ini dengan debitur sebelumnya (Anderson, 2007). *Credit scoring* dapat dianggap sebagai suatu proses statistik yang mengubah data calon debitur menjadi informasi yang dapat digunakan untuk membuat keputusan kredit. Pemberian kredit harus dilakukan dengan bijaksana, mengingat tingginya tingkat risiko yang terkait. Proses *credit scoring* tidak hanya bertujuan untuk menilai kelayakan debitur dalam membayar kembali kredit, tetapi

juga untuk menganalisis data kredit bermasalah atau Non-Performing Loans (NPL) (Yosi.,dkk, 2017). Analisis yang efektif terhadap data NPL dapat membantu bank dalam mengukur kualitas proses pengesahan kredit. Oleh karena itu, manajemen risiko kredit menjadi sangat penting dalam proses pemberian pinjaman. Beberapa model telah dikembangkan dan diterapkan untuk membantu pengambilan keputusan dalam mengevaluasi risiko terkait dengan pemberian kredit. Metode credit scoring sendiri dapat dibagi menjadi dua kategori, yaitu metode berbasis statistik dan metode berbasis kecerdasan buatan atau pembelajaran mesin (Mpofu & Mukosera, 2012).

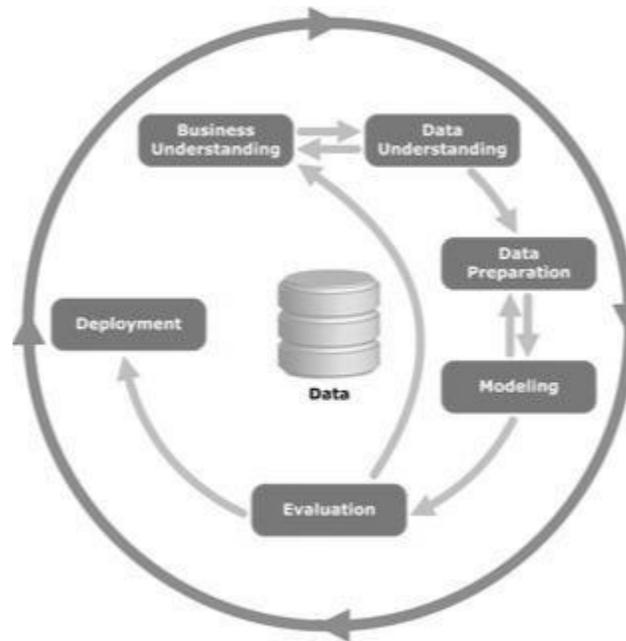
Data Mining adalah bidang yang menyatukan pola, statistik, database, dan visualisasi untuk mengambil informasi dari database yang besar (Kusrini, 20009). Metodologi CRISP-DM, yang dikembangkan oleh tiga penggagas data mining market, menjadi suatu standarisasi yang melibatkan lebih dari 300 organisasi untuk menghasilkan suatu proses pemodelan yang terdefinisi dengan baik dan dipublikasikan pada tahun 1999 (Larose, 2014). Dalam upaya meningkatkan efektivitas credit scoring, implementasi proses data mining menggunakan CRISP-DM (Cross Industry Standard Process for Data Mining) sebagai framework menjadi pilihan. CRISP-DM merupakan standarisasi pemrosesan data mining yang dikembangkan dengan tujuan menjalankan setiap fase dalam proses data mining secara terstruktur dan efisien. (Hasanah, dkk, 2021).Penelitian terdahulu terdapat implementasi metodologi CRISP-DM dalam membangun *machine learning* untuk risiko kredit dengan fokus pada masalah klasifikasi (Rawat, K., 2023) yang menjanjikan model untuk *memfilter* nasabah potensial pada dunia perbankan. Penelitian yang dilakukan oleh (Darmawan, dkk., 2020) juga menunjukkan pengklasifikasian nasabah dengan fokus model akhir untuk menguji C4.5 menggunakan metodologi CRISP-DM, pola dari pohon keputusan menjadikan referensi bank dalam memilih nasabah untuk diberikan pinjaman. Dalam ranah yang berbeda, metodologi CRISP-DM terbukti mampu menjadi pemandu dalam pembangunan sistem yang berhubungan dengan *machine learning* ataupun *data mining* misalnya untuk memprediksi *customer churn* (Yudiana, dkk., 2023), nasabah berisiko dengan fokus algoritma SVM (Wurijanto, 2022), profil risiko nasabah (Arifyan, 2022), strategi pemasaran (Singgalen, 2023), klasifikasi ulasan pengunjung (Singgalen, 2023), masalah klasterisasi dalam pengelompokkan daerah rawan bencana kebakaran (Dhewayani, 2022), prediksi curah hujan berpotensi banjir (Hasanah, 2021), sistem rekomendasi konten (Pernanda, 2023), bahkan untuk pemilihan jenis sampah dengan *neural network* (Alden, 2023).

Dari penelitian terdahulu terkait implementasi CRISP-DM, penelitian ini berfokus pada bagian spesifik dari *credit scoring* yakni variabel-variabel *credit guarantee*. Alih-alih menggunakan model yang spesifik dalam tahap pemodelan, penelitian ini menggunakan lebih dari satu model yakni dengan menggunakan *library* Lazy Predict 0.2.12 (Lazy Predict, 2022) untuk memberikan gambaran umum performa dari klasifikasi berbagai model sehingga terdapat banyak pilihan dalam menentukan model apa yang akan digunakan.

2 METODE

Penelitian ini menggunakan 32 variabel pembangun sub-sistem *credit guarantee* dalam sistem *credit scoring* dengan total data sebanyak 1000 baris. Penelitian ini menerapkan kerangka metodologi CRISP-DM (Cross Industry Standard Process for Data Mining) dengan menerapkan library Lazypredict 0.2.12 untuk memberikan gambaran klasifikasi dari proses pembelajaran mesin terhadap data yang ada.

Proses data mining, yang mengikuti framework CRISP-DM (*Cross Industry Standard Process for Data Mining*) seperti diuraikan oleh (Larose, 2006), terdiri dari enam fase utama yang nampak pada Gambar 1.



Gambar 1. Ilustrasi Alur CRISP-DM

1. Fase Pemahaman Bisnis (*Business Understanding Phase*):
 - a. Menetapkan tujuan proyek dan kebutuhan secara rinci dalam cakupan bisnis atau unit penelitian.
 - b. Menerjemahkan tujuan dan batasan menjadi formula permasalahan data mining.
 - c. Menyusun strategi awal untuk mencapai tujuan.
2. Fase Pemahaman Data (*Data Understanding Phase*):
 - a. Mengumpulkan data.
 - b. Menggunakan analisis penyelidikan data untuk mendapatkan pemahaman lebih lanjut dan mencari pengetahuan awal.
 - c. Evaluasi kualitas data.
 - d. Jika diperlukan, memilih sebagian kecil data yang mungkin mengandung pola dari permasalahan.
3. Fase Persiapan Data (*Data Preparation Phase*):
 - a. Menyiapkan data awal, mengumpulkan data yang akan digunakan pada fase-fase berikutnya. Fase ini memerlukan pekerjaan intensif.
 - b. Memilih kasus dan variabel untuk dianalisis sesuai dengan kebutuhan analisis.
 - c. Melakukan perubahan pada beberapa variabel jika diperlukan.
 - d. Menyiapkan data awal agar siap untuk proses pemodelan.

4. Fase Pemodelan (*Modelling Phase*):
 - a. Memilih dan menerapkan teknik pemodelan yang sesuai.
 - b. Mengkalibrasi aturan model untuk mengoptimalkan hasil.
 - c. Memperhatikan bahwa beberapa teknik dapat digunakan untuk permasalahan data mining yang sama.
 - d. Jika diperlukan, proses dapat kembali ke fase persiapan data untuk memenuhi spesifikasi teknik data mining tertentu.
5. Fase Evaluasi (*Evaluation Phase*):
 - a. Mengevaluasi satu atau lebih model yang digunakan dalam fase pemodelan untuk menentukan kualitas dan efektivitas sebelum disebarkan untuk digunakan.
 - b. Menetapkan apakah model memenuhi tujuan awal.
 - c. Menetapkan apakah terdapat permasalahan penting dari bisnis atau penelitian yang belum ditangani dengan baik.
 - d. Mengambil keputusan terkait penggunaan hasil dari data mining.
6. Fase Penyebaran (*Deployment Phase*):
 - a. Menggunakan model yang dihasilkan. Pembentukan model tidak menandakan selesainya proyek.
 - b. Contoh sederhana penyebaran: Pembuatan laporan.

3 HASIL DAN PEMBAHASAN

3.1 Fase Pemahaman Bisnis (*Business Understanding Phase*)

Pemahaman Bisnis (*Business Understanding*) dalam konteks *AI Credit Guarantee and Collateral* pada aplikasi *credit scoring* melibatkan pemahaman mendalam mengenai tujuan bisnis dan kebutuhan lembaga keuangan terkait dengan analisis kondisi dalam proses penilaian risiko kredit. Beberapa komponen yang tercakup dalam pemahaman bisnis tersebut mencakup:

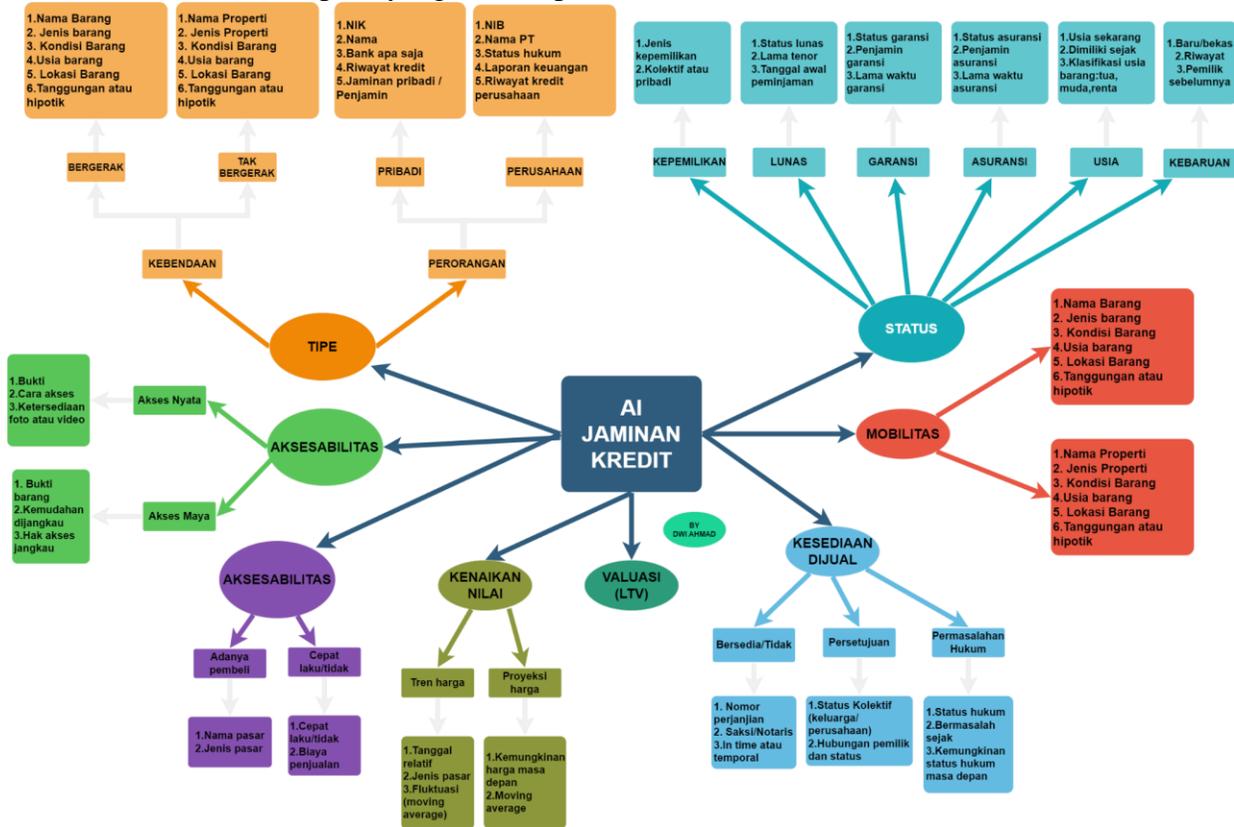
1. Tujuan Bisnis
 - Meningkatkan ketepatan penilaian risiko kredit dengan mempertimbangkan faktor kondisional yang mempengaruhi kemampuan penerima kredit untuk membayar.
 - Mengurangi risiko kredit yang tidak diinginkan yang dapat membahayakan stabilitas keuangan lembaga keuangan.
2. Kebutuhan Lembaga Keuangan
 - Menyesuaikan penilaian risiko kredit dengan faktor kondisional yang memengaruhi pelanggan atau perusahaan dalam membayar kewajiban finansial.
 - Memperkuat keputusan kredit dengan memasukkan analisis kondisi terkait konteks ekonomi, industri, dan personal yang relevan.
 - Menghadapi risiko dan ketidakpastian terkait dengan kondisi ekonomi, industri, dan personal pelanggan.
3. Fokus *AI Credit Guarantee and Collateral*
 - Jaminan Perorangan: Analisis terhadap jaminan perorangan sebagai penjamin pembayaran kewajiban pinjaman.
 - Jaminan Kebendaan: Penilaian terhadap nilai dan kelayakan aset berwujud yang diajukan sebagai agunan.

- Agunan: Analisis menyeluruh terhadap berbagai aspek agunan, termasuk tipe, status, kemungkinan digadaikan, dan riwayat agunan sebelumnya.

Dengan fokus pada jaminan perorangan, jaminan kebendaan, dan analisis agunan secara keseluruhan, AI Credit Guarantee and Collateral memungkinkan lembaga keuangan untuk membuat keputusan kredit yang lebih cerdas dan berdasarkan informasi yang akurat. Sistem ini memberikan kepercayaan tambahan kepada lembaga keuangan dalam menilai risiko kredit dan mengelola portofolio kredit mereka secara efisien.

3.2 Fase Pemahaman Data (Data Understanding Phase)

Pada tahap Pemahaman Data, informasi yang relevan terkait dengan variabel data yang akan digunakan dikumpulkan. Setelah pengumpulan informasi, langkah selanjutnya melibatkan pembuatan *mind map* untuk fitur *AI Credit Guarantee and Collateral*. *Mind map* ini menjadi alat visual yang memetakan secara rinci komponen-komponen utama dan hubungan antar variabel dalam fitur tersebut, seperti yang terlihat pada Gambar 2.



Gambar 2. Mind Map AI Collateral and Credit Guarantee (AI Jaminan Kredit)

Secara garis besar, AI Credit Guarantee and Collateral terbagi menjadi tiga kondisi utama, yaitu kondisi ekonomi, kondisi industri, dan kondisi sosial. Penjelasan rinci mengenai variabel dapat dilihat pada Tabel 3.1.

Tabel 3.1 Hasil Artikel yang Relevan dengan Inklusi

No	Nama Variabel	Deskripsi	Tipe Data	Elemen/Value	Contoh
1	first_Name	Merupakan variabel berisikan nama depan.	Varchar	Unique	Dwi
2	last_name	Merupakan variabel berisikan nama belakang.	Varchar	Unique	Ahmad
3	job	Pekerjaan dari pengaju jaminan.	Varchar	Daftar pekerjaan ; Pengacara, Manajer Operasi, Programmer, Gamer, Nelayan, Petani	Programmer
4	salary	Penghasilan dari pengaju jaminan.	Integer	Penghasilan ; 15000000,1000000	15000000
5	tujuan_pinjaman	Merupakan tujuan pengaju pinjaman untuk penggunaan bisnis atau untuk individu.	Varchar	Pribadi atau bisnis	Pribadi
6	loan_amount	Jumlah nilai pinjaman yang diajukan berupa nominal seberapa banyak uang yang diajukan	Integer	Jumlah pinjaman; 15000000	20000000
7	agunan	Merupakan item yang diajukan sebagai agunan	Varchar	Nominal-Unique	Sepeda motor
8	jumlah_agunan	Total item agunan yang diajukan	Integer	Jumlah agunan (diskrit)	1
9	tipe_agunan	Tipe agunan merupakan klasifikasi jenis dari agunan yang diajukan	Varchar	Property, Saham, NFT, Kendaraan, barang seni, gadget, alat rumah tangga, perhiasan,	Kendaraan

No	Nama Variabel	Deskripsi	Tipe Data	Elemen/Value	Contoh
				dokumen penting, dll	
10	status_gadai	Status gadai merupakan kesediaan suatu agunan untuk digadaikan oleh pengaju	Varchar	Bersedia, Tergadai, Tidak digadaikan, tidak bersedia	Tidak bersedia
11	tipe_jalan	Tipe jalan merupakan jenis jalan berdasarkan akses dan skala jalan, digunakan untuk agunan sejenis properti	Varchar	Jalan raya, gang, jalan sedang, tidak di jalan, tidak diketahui, bukan property	bukan property
12	alamat	Alamat merupakan data lengkap tentang lokasi dari suatu properti	Varchar	Alamat lengkap dari barang yang digunakan	bukan property
13	kodepos	Lokasi kodepos dari suatu properti	Varchar	Kodepos	bukan property
14	status_agunan_terakhir	Merupakan status agunan sebelumnya yang diajukan apakah memiliki reputasi yang baik, buruk, tidak disebutkan?	Varchar	Disita, masih digadai, ditebus, tidak lunas, diambil paksa	Ditebus
15	status_agunan_lainnya_1	Merupakan status agunan sebelumnya yang diajukan apakah memiliki reputasi yang baik, buruk, tidak disebutkan?	Varchar	Disita, masih digadai, ditebus, tidak lunas, diambil paksa	Ditebus
16	status_agunan_lainnya_2	Merupakan status agunan sebelumnya yang diajukan apakah memiliki reputasi yang baik, buruk, tidak disebutkan?	Varchar	Disita, masih digadai, ditebus, tidak lunas, diambil paksa	Ditebus

No	Nama Variabel	Deskripsi	Tipe Data	Elemen/Value	Contoh
17	status_agunan_lainnya_3	Merupakan status agunan sebelumnya yang diajukan apakah memiliki reputasi yang baik, buruk, tidak disebutkan?	Varchar	Disita, masih digadai, ditebus, tidak lunas, diambil paksa	Ditebus
18	status_agunan_terburuk	Merupakan status agunan terburuk yang pernah dilakukan oleh pemilik agunan	Varchar	Disita, masih digadai, ditebus, tidak lunas, diambil paksa, tidak ada record buruk	tidak ada record buruk
19	nilai_agunan	Merupakan nilai nominal agunan dalam rupiah	Integer	Kuantitas	17000000
20	rasio_nilai_agunan_to_pinjaman	Merupakan hasil perbandingan antara nilai agunan terhadap nilai pinjaman yang diajukan	Float	Nilai tergantung dari nilai agunan dan nilai pinjaman	0.85
21	wujud_agunan	Merupakan wujud dari agunan apakah berupa nyata atau maya, dan bergerak atau tidak bergerak	Varchar	Maya, Nyata, Tak berwujud, bergerak, tidak bergerak	bergerak
22	liquiditas_agunan	Liquiditas agunan merupakan seberapa mudah agunan untuk dijual kembali di pasar	Varchar	Sedang, rendah, tinggi, tak laku, tidak stabil	tidak stabil
23	aksesibel	Merupakan kemudahan akses suatu agunan untuk diketahui atau dinyatakan perwujudannya	Varchar	Mudah diakses, sulit diakses, tidak dapat diakses, akses maya, akses tertentu, akses terbatas, akses terlarang,	sulit diakses

No	Nama Variabel	Deskripsi	Tipe Data	Elemen/Value	Contoh
24	Kondisi_agunan	Merupakan status barang agunan baru atau bekas	Varchar	Bekas, Baru, tidak dapat ditentukan, tidak diketahui, rusak	Bekas
25	Usia_agunan	Merupakan usia agunan saat pertama kali dibeli berdasarkan tahunnya	Varchar	<2000 dan >2000	2020
26	Asuransi agunan	Merupakan variabel apakah sebuah agunan memiliki asuransi yang ditanggung atau tidak	Varchar	Memiliki asuransi atau tidak memiliki asuransi	Memiliki
27	Status lunas agunan	Apakah agunan merupakan barang yang wajib dicicil atau tidak	Varchar	Sudah lunas, pernah dicicil, cicilan aktif	Sudah lunas
28	credit history	This includes factors such as payment history, outstanding debt, length of credit history, and credit utilization. It provides an indication of how likely the applicant is to repay their debts.	Integer	0 : lancar 1 : dalam perhatian khusus 2 : kurang lancar 3 : diragukan 4 : macet	0
29	employment history	Stable employment with a good income can provide assurance that the applicant is capable of repaying the loan.	Varchar		5 Tahun
30	Education level	A higher education level can provide an indication of the applicant's earning potential and financial literacy.	Varchar	Tidak bersekolah, SD, SMP, SMA, D3, D4, S1, S2, S3	S4

No	Nama Variabel	Deskripsi	Tipe Data	Elemen/Value	Contoh
31	Age	The applicant's age can be a factor in determining creditworthiness, as older individuals may be seen as more financially stable.	Integer	>18 thn	23
32	NIK	Merupakan nilai unik dari sebuah entitas data nasabah yang akan digunakan sebagai primary key and foreign	Varchar	Unique value based on ID CARD	6201011604 994

3.3 Fase Persiapan Data (*Data Preparation*)

Pada tahap Persiapan Data (*Data Preparation*), serangkaian langkah kritis dilakukan untuk memastikan kualitas dan kesesuaian data sebelum memasuki proses analisis. Proses ini melibatkan beberapa langkah penting, pertama-tama adalah pembersihan data. Langkah ini mencakup identifikasi dan penanganan missing values, outlier, atau anomali dalam dataset, dengan tujuan untuk menjaga kebersihan dataset dan menghilangkan potensi gangguan dari data yang tidak standar. Selanjutnya, dilakukan transformasi variabel, yang melibatkan normalisasi data. Normalisasi membantu membawa variabel ke skala yang seragam, memastikan bahwa perbedaan skala antar variabel tidak mempengaruhi analisis secara tidak proporsional. Proses ini juga melibatkan encoding kategori atau pengelompokan variabel, bergantung pada kebutuhan analisis.

Pentingnya tahap Persiapan Data adalah untuk memastikan bahwa data yang digunakan dalam pembuatan fitur AI Credit Guarantee and Collateral adalah akurat, lengkap, dan sesuai dengan kebutuhan analisis. Dengan data yang telah dipersiapkan dengan baik, proses analisis selanjutnya dapat menghasilkan model yang lebih dapat diandalkan dan memberikan output yang lebih akurat. Langkah berikutnya adalah melakukan EDA (*Exploratory Data Analysis*), di mana data dikumpulkan dan dieksplorasi untuk memahami struktur, distribusi, dan karakteristiknya. EDA merupakan teknik yang digunakan dalam langkah ini untuk melakukan analisis lebih lanjut terhadap data. Langkah pertama dalam melakukan EDA adalah melihat descriptive statistics, seperti bentuk data, informasi data, dan deskripsi statistik data untuk memahami secara lebih mendalam struktur data yang akan diolah.

3.4 Fase Pemodelan (*Modelling Phase*)

Pada tahap Pemodelan dan Pelatihan Data, langkah awal yang diambil adalah proses train-test split data. Proses ini dilakukan untuk membagi dataset menjadi dua bagian utama, yaitu data training dan data testing. Data training digunakan untuk melatih model Machine Learning, sementara data testing digunakan untuk menguji performa dan generalisasi model yang telah dilatih. Proses train-test split data dilakukan dengan pembagian dataset menjadi 70% data training dan 30% data testing. Pembagian ini dapat memberikan evaluasi yang baik terhadap performa model pada data yang tidak pernah dilihat selama proses pelatihan.

Selanjutnya, dilakukan pemodelan menggunakan Lazy Predict, sebuah library Python yang menyediakan fungsi-fungsi berguna untuk eksplorasi awal dan evaluasi model Machine Learning. Lazypredict menyediakan ringkasan cepat dari performa berbagai model machine learning tanpa perlu mengkonfigurasi dan melatih setiap model secara terpisah.

3.5 Fase Evaluasi (*Evaluation Phase*)

Pada tahap evaluasi, dilakukan penilaian kinerja berbagai model menggunakan metrik akurasi (*Accuracy*) dan skor F1 (*F1 Score*) yang terdokumentasi dalam Tabel 3.2. Tabel tersebut mencakup batas atas dan batas bawah untuk kedua metrik pada berbagai model seperti *RandomForestClassifier*, *BaggingClassifier*, *XGBClassifier*, dan lainnya.

Tabel 3.2 Batas atas dan batas bawah faktor

Model	Accuracy	F1 Score
<i>RandomForestClassifier</i>	0.97	0.99
<i>BaggingClassifier</i>	0.94	0.94
<i>XGBClassifier</i>	0.94	0.94
<i>SGDClassifier</i>	0.94	0.94
<i>LogisticRegression</i>	0.94	0.94
<i>LGBMClassifier</i>	0.91	0.91
<i>DecissionTreeClassifier</i>	0.91	0.91
<i>RidgeClassifierCV</i>	0.91	0.91
<i>NearestCentroid</i>	0.91	0.91
<i>Perceptron</i>	0.88	0.88
<i>PassiveAggresiveClassifier</i>	0.88	0.88
<i>LinearSVC</i>	0.88	0.88
<i>ExtraTreeClassifier</i>	0.76	0.83
<i>RidgeClassifier</i>	0.79	0.80
<i>LinearDiscriminantAnalysis</i>	0.76	0.77
<i>ExtraTreesClassifier</i>	0.85	0.85
<i>LabelSpreading</i>	0.36	0.49
<i>LabelPropagation</i>	0.36	0.49
<i>NuSVC</i>	0.91	0.89
<i>SVC</i>	0.91	0.89
<i>CalibratedClassifierCV</i>	0.88	0.86
<i>QuadraticDiscriminantAnalysis</i>	0.48	0.49
<i>BernoulliNB</i>	0.82	0.82
<i>KNeighborsClassifier</i>	0.79	0.79
<i>AdaBoostClassifier</i>	0.73	0.79
<i>DummyClassifier</i>	0.70	0.57
<i>GaussianNB</i>	0.39	0.46

Akurasi mengukur sejauh mana model dapat memberikan prediksi yang benar secara keseluruhan, sedangkan F1 Score mengukur keseimbangan antara presisi dan recall (Zhang, dkk., 2019).. Melihat Tabel 3, beberapa model seperti RandomForestClassifier menunjukkan akurasi dan F1 Score tertinggi, dengan batas atas mencapai 0.97 dan 0.99, menandakan tingkat kinerja yang sangat baik. Terdapat model-model dengan batas bawah yang lebih rendah, seperti LabelSpreading dan LabelPropagation yang memiliki akurasi dan F1 Score sekitar 0.36 hingga 0.49. Hal ini menunjukkan variasi kinerja antar model, dan pemilihan model yang sesuai dengan kebutuhan aplikasi menjadi penting.

Dari hasil pemodelan menggunakan Lazypredict, dapat disimpulkan bahwa tiga model terbaik yang dapat digunakan pada dataset adalah RandomForestClassifier, BaggingClassifier, dan XGBClassifier. Pemilihan ketiga model ini didasarkan pada hasil komparasi sebelumnya yang menunjukkan bahwa ketiganya memberikan performa terbaik dalam konteks aplikasi Credit Guarantee and Collateral. Dalam konteks evaluasi ini, pemilihan model untuk implementasi fitur AI Credit Guarantee and Collateral harus mempertimbangkan trade-off antara akurasi dan kinerja secara keseluruhan (Zhang, dkk., 2019). Evaluasi ini menjadi dasar untuk mengidentifikasi model-model yang paling sesuai dengan tujuan pengembangan fitur pada aplikasi credit scoring.

3.6 Fase Deployment

Proses implementasi setelah membangun model machine learning melibatkan penyimpanan pipeline dan model menggunakan library seperti *pickle*. Dengan menyimpan objek pipeline dan model dalam file, dapat digunakan kembali tanpa perlu membangun ulang dari awal. Pembuatan API dengan menggunakan framework web seperti Flask menjadi langkah berikutnya, memungkinkan pembuatan endpoint untuk menerima input data dan mendapatkan prediksi dari model machine learning. Pengujian dilakukan dengan mengirimkan permintaan ke endpoint yang telah dibuat dan memeriksa respons yang diterima untuk memastikan kesesuaian dengan harapan. Langkah terakhir dalam implementasi adalah melakukan pengujian API. Pengujian ini bertujuan untuk memastikan bahwa API berfungsi dengan baik dan memberikan prediksi yang akurat berdasarkan data input yang diberikan. Proses ini menjadi krusial untuk memverifikasi keberhasilan implementasi dan memastikan bahwa model machine learning dapat beroperasi secara efektif dalam lingkungan produksi. Melalui tahap pengimplementasian ini, model machine learning dapat diintegrasikan ke dalam aplikasi atau sistem yang lebih kompleks. Ini memungkinkan penggunaan model secara praktis untuk melakukan prediksi credit score berdasarkan data lokasi. Pengujian API menjadi langkah penting untuk memastikan keberhasilan implementasi dan menjaga kinerja model dalam lingkungan produksi.

4 KESIMPULAN

Proyek AI Credit Guarantee and Collateral untuk credit scoring telah sukses menghadirkan solusi yang inovatif dan efisien dalam meningkatkan ketepatan penilaian risiko kredit. Melalui pemahaman bisnis yang mendalam, proyek ini bertujuan utama untuk memperbaiki ketepatan penilaian risiko dengan mempertimbangkan faktor kondisional yang memengaruhi kemampuan penerima kredit untuk membayar, serta fokus pada analisis jaminan perorangan, jaminan kebendaan, dan agunan secara menyeluruh. Tahapan metodologi pengembangan CRISP-DM, mulai dari Pemahaman Bisnis, Pemahaman Data, Persiapan Data, hingga Pemodelan dan Pelatihan Data, telah dijalankan secara konsisten untuk menghasilkan model-machine learning yang dapat diandalkan.

Proses evaluasi model menggunakan metrik akurasi dan F1 Score memberikan pemahaman yang mendalam tentang kinerja masing-masing model, dan tiga model terbaik, yaitu RandomForestClassifier, BaggingClassifier, dan XGBClassifier, terpilih untuk diimplementasikan. Pada tahap Deployment, model-machine learning disimpan dan diuji melalui pembuatan API menggunakan Flask. Keberhasilan implementasi ini membuka jalan bagi penggunaan praktis model dalam prediksi credit score berdasarkan data lokasi, dengan pengujian API menjadi langkah krusial untuk memastikan kinerja model dalam lingkungan produksi. Dengan demikian, proyek ini berhasil menghadirkan solusi AI yang dapat memberikan kontribusi positif terhadap pengambilan keputusan kredit yang lebih cerdas dan akurat.

UCAPAN TERIMAKASIH

Penulis mengucapkan terimakasih kepada Program Studi Program Profesi Insinyur, Institut Teknologi Indonesia untuk segala fasilitas dalam perkuliahan dan penunjang riset. Penulis juga berterimakasih kepada PT. XYZ yang sudah memberikan peluang untuk studi kasus keinsinyuran.

DAFTAR PUSTAKA

- Saputri, S. D., & Ermatita, E. (2019). Credit Scoring Kelayakan Debitur Menggunakan Metode Hybrid ANN Backpropagation dan TOPSIS. *Jurnal RESTI (Rekayasa Sistem Dan Teknologi Informasi)*, 3(1), 73-78.
- Anderson, R. (2007). The credit scoring toolkit: theory and practice for retail credit risk management and decision automation. *Oxford university press*.
- Mpofu, T. P., & Mukosera, M. (2014). Credit scoring techniques: a survey. *International Journal of Science and Research (IJSR)*, 2319-7064.
- Kusrini, E. T. L., & Taufiq, E. (2009). Algoritma data mining. *Yogyakarta: Andi Offset*.
- Larose, D. T., & Larose, C. D. (2014). Discovering knowledge in data: an introduction to data mining (Vol. 4). John Wiley & Sons.
- Hasanah, M. A., Soim, S., & Handayani, A. S. (2021). Implementasi CRISP-DM Model Menggunakan Metode Decision Tree dengan Algoritma CART untuk Prediksi Curah Hujan Berpotensi Banjir. *Journal of Applied Informatics and Computing*, 5(2), 103-108.
- Rawat, K. (2023, July). Applying CRISP-DM Methodology in Developing Machine Learning Model for Credit Risk Prediction. In *Science and Information Conference* (pp. 522-538). Cham: Springer Nature Switzerland.
- Darmawan, T., Birawa, A. S., Eryanto, E., & Mauritsius, T. (2020). Credit classification using CRISP-DM method on Bank ABC customers. *International Journal of Emerging Trends in Engineering Research*, 8(6).
- Yudiana, Y., Agustina, A. Y., & Khofifah, N. (2023). Prediksi Customer Churn Menggunakan Metode CRISP-DM Pada Industri Telekomunikasi Sebagai Implementasi Mempertahankan Pelanggan. *Indonesian Journal of Islamic Economics and Business*, 8(1), 1-20.
- Wurijanto, T., Setiawan, H. B., & Tjandrarini, A. B. (2022). Penerapan Model CRISP-DM pada Prediksi Nasabah Kredit yang Berisiko Menggunakan Algoritma Support Vector Machine. *Jurnal Ilmiah Scroll (Jendela Teknologi Informasi)*, 10(1), 1-6.
- Arifyan, A. (2022). Analisis Perbandingan Optimasi berbasis Evolutionary pada Algoritma Klasifikasi Penentuan Profile Resiko Nasabah. *Techno. Com*, 21(3), 565-578.
- Singgalen, Y. A. (2023). Penerapan Metode CRISP-DM untuk Optimalisasi Strategi Pemasaran STP (Segmenting, Targeting, Positioning) Layanan Akomodasi Hotel, Homestay, dan Resort. *JURNAL MEDIA INFORMATIKA BUDIDARMA*, 7(4), 1980-1993.

- Singgalen, Y. A. (2023). Analisis Sentimen dan Sistem Pendukung Keputusan Menginap di Hotel Menggunakan Metode CRISP-DM dan SAW. *Journal of Information System Research (JOSH)*, 4(4), 1343-1353.
- Dhewayani, F. N., Amelia, D., Alifah, D. N., Sari, B. N., & Jajuli, M. (2022). Implementasi K-Means Clustering untuk Pengelompokan Daerah Rawan Bencana Kebakaran Menggunakan Model CRISP-DM. *Jurnal Teknologi dan Informasi*, 12(1), 64-77.
- Pernanda, S. T., Lavindi, E. E., Yobioktabera, A., & Karima, A. (2023). IMPLEMENTASI CRISP-DM PADA SISTEM REKOMENDASI BERBASIS KONTEN UNTUK PENENTUAN JENIS IKAN HIAS DALAM AKUARIUM. *Orbith: Majalah Ilmiah Pengembangan Rekayasa dan Sosial*, 19(3), 255-262.
- Alden, S., & Sari, B. N. (2023). Implementasi Algoritma CNN Untuk Pemilahan Jenis Sampah Berbasis Android Dengan Metode CRISP-DM. *Jurnal Informatika*, 10(1), 62-71.
- S. R. Pandala. lazypredict. (2022, September 28). PyPI. <https://pypi.org/project/lazypredict/>
- Zhang, H., Yu, Y., Jiao, J., Xing, E., El Ghaoui, L., & Jordan, M. (2019, May). Theoretically principled trade-off between robustness and accuracy. *In International conference on machine learning* (pp. 7472-7482). PMLR.