

BAB III METODOLOGI PENELITIAN

3.1 Subjek dan Objek Penelitian

Sesuai dengan latar belakang yang telah dijelaskan sebelumnya, penelitian ini bertujuan untuk mengukur Skor Kredit seseorang apakah layak untuk mengajukan kredit atau tidak dengan menggunakan algoritma *eXtreme Gradient Boosting* (XGBoost). Subjek penelitian dalam konteks ini adalah individu atau entitas yang menjadi calon peminjam (debitur) di sebuah lembaga keuangan seperti bank atau lembaga pembiayaan. Sementara objek penelitiannya adalah variabel yang digunakan untuk menilai kelayakan kredit dari subjek. Objek penelitian mencakup informasi dan data yang digunakan dalam proses pengambilan keputusan tentang pemberian kredit kepada subjek.

3.2 Alat dan Bahan Penelitian

Pada penelitian ini peneliti menggunakan alat yang digunakan untuk membantu proses penelitian, di antaranya adalah *Hardware* (perangkat keras) dan *Software* (perangkat lunak), serta bahan dari penelitian ini adalah v

3.2.1 Alat

3.2.1.1 Hardware (Perangkat Keras)

Hardware (perangkat keras) yang digunakan pada penelitian ini terdapat pada Tabel 3.1 yaitu sebagai berikut:

Tabel 3. 1 Perangkat Keras yang Digunakan

No	Nama <i>Hardware</i>	Fungsi
1	Laptop Acer Arpire 3 A315-21-948E (Prosesor AMD A-Series A9-9420, RAM: 12 GB, OS:	Membuat instrumen penelitian dan mengumpulkan serta mengolah data hasil penelitian

	Windows 10)	
--	-------------	--

3.2.1.2 *Software* (Perangkat Lunak)

Software (perangkat lunak) yang digunakan pada penelitian ini terdapat pada Tabel 3.2 yaitu sebagai berikut:

Tabel 3. 2 Perangkat Lunak yang Digunakan

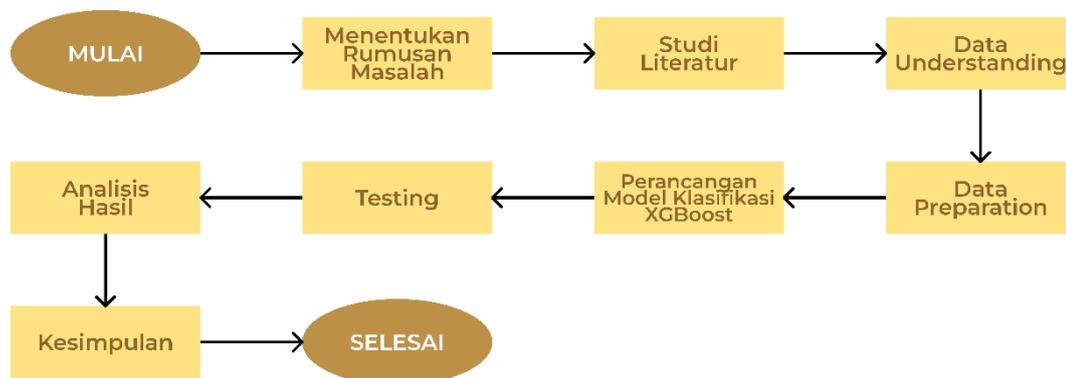
No	Nama <i>Software</i>	Fungsi
1	Google Document	Membuat, mencatat, dan menyimpan instrumen penelitian secara online
2	Microsoft Word	Membuat, mencatat, dan menyimpan instrumen penelitian secara <i>offline</i>
3	Jupyter Notebook	Mengolah data, perancangan model klasifikasi XGBoost, dan melakukan Exploratory Data Analysis (EDA)

3.2.2 Bahan

Bahan yang digunakan merupakan data histori kredit seseorang yang diperoleh dari platform Kaggle. Data yang digunakan sebanyak 150.000 data.

3.3 Diagram Alir Penelitian

Diagram alir penelitian merupakan diagram yang berfungsi untuk menggambarkan urutan suatu proses yang bertujuan untuk membantu menjelaskan suatu penelitian. Diagram alir disusun secara urut agar memudahkan peneliti melakukan suatu penelitian. Selain itu, diagram alir yang dibuat harus terstruktur, jelas, dan mudah dipahami.



Gambar 3. 1 Diagram Alir

Pada Gambar 3.1 menggambarkan alur proses penelitian yang dimulai dari tahap pertama (menentukan rumusan masalah) hingga tahap terakhir (kesimpulan) dengan keterangan dan penjelasan sebagai berikut:

3.3.1 Menentukan Rumusan Masalah

Tahap pertama pada penelitian ini adalah menentukan rumusan masalah, tujuan penelitian, serta manfaat penelitian. Perumusan masalah harus dilakukan dengan membaca referensi dan jurnal yang berkaitan dengan credit scoring, khususnya yang menggunakan metode algoritma XGBoost. Setelah memahami tentang credit scoring dan XGBoost, penentuan rumusan masalah, tujuan penelitian, dan manfaat penelitiannya dapat dibuat.

3.3.2 Studi Literatur

Tahap kedua adalah melakukan studi literatur yang bertujuan untuk mencari teori dan informasi yang berasal dari jurnal/penelitian sebelumnya dengan subjek dan objek yang relevan dengan penelitian ini.

3.3.3 Data Understanding

Data understanding adalah proses penting untuk menggali informasi tambahan dari data. Langkah pertama melibatkan pengumpulan data yang kemudian diikuti oleh tahap eksplorasi. Eksplorasi yang dilakukan yaitu mengidentifikasi tipe data yang ada pada dataset, kemudian melihat apakah

ada missing values pada data tersebut, serta mengidentifikasi apakah data tersebut seimbang atau tidak.

3.3.4 Data Preparation

Data preparation adalah proses yang krusial dalam pengolahan data. Tujuannya adalah memastikan data yang tersedia siap untuk diolah secara efektif. Persiapan data penting sebelum membangun model guna memastikan kualitas model yang dihasilkan. Langkah-langkah preprocessing melibatkan penanganan masalah seperti ketidakseimbangan data, yang dapat mengganggu performa model. Ketidakseimbangan data merupakan keadaan di mana jumlah kelas data yang satu lebih banyak atau lebih sedikit dari kelas data yang lain. Data dapat dikatakan seimbang apabila jumlah kelas data yang satu dengan yang lainnya memiliki jumlah yang relative sama. Proses penanganan ketidakseimbangan data dilakukan dengan proses pruning atau pemangkasan.

Selain menangani ketidakseimbangan data, pada proses ini juga dilakukan penanganan missing values. Missing values merupakan keadaan di mana data tersebut tidak memiliki nilai. Untuk menangani missing values dilakukan penghapusan pada semua data yang tidak bernilai. Dengan persiapan data yang tepat, hasil model dapat lebih akurat dan dapat diandalkan. Oleh karena itu, tahapan persiapan data perlu diberikan perhatian khusus untuk memastikan bahwa data yang diolah mewakili masalah secara akurat dan menghasilkan hasil analisis yang bernilai.

3.3.5 Perancangan Model XGBoost

Setelah semua data dipersiapkan, langkah selanjutnya adalah pengolahan. Proses ini mencakup pembentukan model klasifikasi menggunakan algoritma XGBoost. Pembentukan model klasifikasi XGBoost dilakukan dengan menentukan nilai-nilai yang akan digunakan di dalam parameter. Kemudian nilai-nilai tersebut akan dilakukan optimasi dengan menggunakan GridSearchCV untuk mendapatkan nilai parameter yang paling optimal untuk menentukan tes kelayakan kredit. Selanjutnya

juga dilakukan evaluasi model untuk menentukan kualitas dan performanya. Dalam tahap awal, data disiapkan dengan teliti. Setelah itu, data diolah melalui algoritma XGBoost yang kuat. Hasilnya dievaluasi untuk memilih model optimal. Evaluasi melibatkan berbagai metrik untuk mengukur performa. Tujuannya adalah menemukan model terbaik yang sesuai dengan memiliki akurasi yang baik.

3.3.6 Testing

Tahap keenam dalam proses pengembangan model adalah Testing. Setelah melalui proses pelatihan, model akan diuji dengan menggunakan data test untuk mengukur performanya. Pengujian ini bertujuan untuk menghasilkan nilai akurasi dan f1-score yang akan divisualisasikan melalui confusion matrix.

Pada tahap Testing, model akan menghadapi data yang belum pernah dilihat sebelumnya. Dengan cara ini, kita dapat mengevaluasi sejauh mana model dapat melakukan prediksi yang tepat terhadap data yang baru. Akurasi adalah metrik yang mengukur sejauh mana model berhasil memprediksi data dengan benar dibandingkan dengan total data yang ada.

Selain akurasi, f1-score juga digunakan sebagai ukuran performa model. F1-score adalah harmonik rata-rata dari precision dan recall. Precision mengukur seberapa baik model dalam mengidentifikasi positif yang sebenarnya dari semua prediksi positif yang dibuat. Sementara itu, recall mengukur seberapa baik model dalam mengklasifikasikan semua kasus positif yang sebenarnya.

Visualisasi menggunakan confusion matrix membantu kita memahami hasil pengujian dengan lebih jelas. Confusion matrix adalah tabel yang menggambarkan jumlah prediksi yang benar dan salah untuk setiap kelas. Dari sini, kita dapat melihat seberapa baik model dalam mengklasifikasikan data ke dalam kelas yang benar dan mengidentifikasi kesalahan yang mungkin perlu diperbaiki.

Pentingnya tahap Testing ini adalah untuk memastikan bahwa model yang telah dilatih memiliki kinerja yang baik dan dapat diandalkan dalam menghadapi data dunia nyata. Dengan evaluasi yang tepat, kita dapat mengidentifikasi kelemahan model dan meningkatkan kualitasnya sehingga sesuai dengan kebutuhan dan tujuan yang diinginkan.

3.3.7 Analisis Hasil

Tahap ketujuh merupakan tahap penting dalam pengembangan model. Pada tahap ini, dilakukan proses analisis menyeluruh terhadap model yang telah dibuat sebelumnya. Proses analisis bertujuan untuk mengevaluasi performa model dalam mencapai tujuan yang telah ditetapkan.

Salah satu aspek utama dari analisis adalah mengukur akurasi model. Akurasi merupakan ukuran sejauh mana model mampu memberikan prediksi yang tepat dan akurat terhadap data baru. Hasil akurasi akan dibandingkan dengan harapan yang telah ditetapkan sebelumnya untuk menilai kesesuaian model.

Selain akurasi, analisis juga melibatkan penelusuran berbagai metrik evaluasi lainnya, seperti presisi, recall, dan F1-score. Semua hasil ini akan membantu dalam mengidentifikasi kelemahan dan kelebihan model yang perlu diperbaiki atau diperkuat.

Proses analisis ini akan memberikan wawasan yang berharga untuk mengembangkan model yang lebih baik dan andal dalam menghadapi tantangan data di masa depan.

3.3.8 Kesimpulan

Pada tahap akhir penelitian, penulis menarik kesimpulan untuk menjawab rumusan masalah yang tercantum di Bab 1. Dalam tahapan ini, peneliti akan membahas kelebihan dan kekurangan dari metode yang digunakan, yaitu Algoritma XGBoost. Kesimpulan ini penting untuk menyimpulkan hasil penelitian secara keseluruhan dan mengambil pandangan yang komprehensif tentang topik yang diteliti. Dalam proses penarikan kesimpulan, peneliti juga akan menganalisis temuan penelitian

dan mengaitkannya dengan tujuan awal penelitian. Selain itu, di tahap ini, saran dan rekomendasi berdasarkan hasil penelitian dapat diajukan untuk pengembangan penelitian selanjutnya atau penerapan praktis dalam bidang terkait.