

BAB I

PENDAHULUAN

1.1 Latar Belakang

Dalam era big data, analisis data telah menjadi salah satu komponen kunci dalam pengambilan keputusan bisnis. Data mining, sebagai salah satu teknik analisis data, memainkan peran penting dalam menemukan korelasi, pola, dan tren dari sumber data yang ada (Liu, 2007). Data mining memanfaatkan teknik algoritma dan statistik untuk menganalisis data dan menghasilkan informasi yang berguna. Tujuan utama data mining adalah untuk mengubah data menjadi pengetahuan yang dapat digunakan untuk mendukung pengambilan keputusan. Proses data mining mencakup seleksi data, preprocessing, transformasi data, pencarian informasi, dan evaluasi atau interpretasi (Kurniawansyah, 2018). Dalam konteks ini, statistik memegang peranan sentral dalam mengolah dan menganalisis data untuk menghasilkan informasi yang relevan.

Seleksi data adalah tahap awal proses data mining di mana data dipilih berdasarkan masalah yang ingin dipecahkan. Preprocessing mencakup pembersihan data, penanganan nilai yang hilang (missing value), dan skala fitur untuk menormalkan nilai variabel independen dalam dataset (Reskiyanti, 2022). Tahap transformasi bertujuan mengubah data ke dalam format yang dapat digunakan secara cross-application atau independen terhadap media pengolahan yang digunakan. Terakhir, tahap evaluasi dan interpretasi bertujuan memberikan pemahaman yang jelas.

Namun, dalam banyak penelitian, sering dijumpai kumpulan data (dataset) yang tidak seimbang. Menurut Hairani et al. (2020), data tidak seimbang adalah suatu kondisi di mana jumlah kemunculan kelas mayoritas lebih besar dari jumlah kemunculan kelas minoritas. Kondisi dataset yang tidak seimbang dapat menyebabkan model klasifikasi menjadi bias dan tidak akurat (Chawla et al., 2002). Dataset yang tidak seimbang cenderung menghasilkan prediksi yang lebih menguntungkan kelas mayoritas dan mengabaikan kelas minoritas (Anjani et al., 2023). Untuk mengatasi masalah ini, pendekatan yang efektif dapat dilakukan melalui random undersampling dan random oversampling. Teknik random undersampling melibatkan pengurangan jumlah data dari kelas mayoritas untuk menciptakan keseimbangan, sementara random oversampling meningkatkan jumlah data pada kelas minoritas dengan cara menyalin data yang sudah ada (Moustafa et al., 2019).

Classification and Regression Tree (CART) merupakan metode statistik nonparametrik yang dikembangkan untuk analisis klasifikasi, baik untuk variabel respon kontinu maupun kategorikal. CART menghasilkan pohon regresi jika variabel respon bersifat kontinu dan menghasilkan pohon klasifikasi jika variabel respon bersifat kategorikal. Tujuan CART adalah untuk mendapatkan model prediksi yang akurat dalam klasifikasi (Siswanto et al., 2022). Proses CART terdiri dari tiga tahap: pembentukan pohon klasifikasi, pemangkasan pohon klasifikasi, dan penentuan pohon klasifikasi yang optimal menggunakan estimasi validasi silang.

Penelitian tentang klasifikasi dengan model CART telah dilakukan oleh Andini dkk (2023) dari Universitas Padang dalam penelitian berjudul “Aplikasi *Synthetic Minority Oversampling Technique* (SMOTE) dalam Metode CART untuk Mengatasi Data Tidak Seimbang dalam Klasifikasi Status Pekerjaan pada Angkatan Kerja di Kota Padang.” Penelitian ini membahas masalah pengangguran di Kota Padang dengan menggunakan algoritma CART dan berbagai teknik penyeimbangan. Hasil penelitian menunjukkan bahwa penerapan SMOTE secara signifikan meningkatkan kinerja model CART dalam menangani data tidak seimbang.

Penelitian lain tentang penanganan data tidak seimbang dengan metode random oversampling dan random undersampling acak pada kinerja model prediksi telah dilakukan oleh Yang dkk. (2023) dari Erasmus University Medical Center dalam artikel berjudul “*Impact of Random Oversampling and Random Undersampling on the Performance of Prediction Models Developed Using Observational Health Data*”. Penelitian ini bertujuan untuk menilai pengaruh dua metode penyeimbangan data terhadap kinerja model prediksi klinis yang dibangun menggunakan data kesehatan observasional. Menggunakan teknik random oversampling dan random undersampling acak, penelitian ini menunjukkan bahwa pendekatan ini meningkatkan kinerja model meskipun hasilnya tidak konsisten pada beberapa algoritma.

Dalam industri perbankan, peran bank sebagai lembaga intermediasi, menghimpun dana dari masyarakat yang memiliki dana lebih dan menyalurkannya kepada masyarakat yang membutuhkan dana. Peran ini sangat penting dalam pertumbuhan perekonomian negara (Iriani et al., 2023). Permintaan kredit, yang merupakan kesepakatan antara lembaga keuangan dan nasabah, adalah salah satu aktivitas utama bank dalam menyalurkan dana. Nilai skor kredit menjadi salah satu faktor penentu persetujuan pinjaman yang diajukan oleh nasabah, di mana skor yang baik menunjukkan peluang besar untuk disetujui, sementara skor yang buruk menunjukkan bahwa nasabah dianggap kurang layak (Amelian Noli, 2015).

Dalam konteks penilaian kredit nasabah, dataset tidak seimbang seringkali terjadi, di mana kelas minoritas (nasabah yang tidak membayar kembali pinjaman) memiliki jumlah data yang lebih sedikit dibandingkan dengan kelas mayoritas (nasabah yang membayar kembali pinjaman). Oleh karena itu, penggunaan CART dengan pendekatan random undersampling dan random oversampling dapat membantu mengatasi masalah ini dan menghasilkan model penilaian kredit yang lebih akurat.

Dalam penelitian ini, kita akan menggunakan dataset informasi kredit nasabah bank yang memuat informasi pribadi nasabah, jumlah tabungan, dan hutang. Tujuan penelitian ini adalah untuk mengembangkan model penilaian kredit yang dapat memprediksi kemampuan kredit nasabah dengan akurat. Penelitian ini akan memaparkan bagaimana performa model CART dengan pendekatan random undersampling dan random oversampling dalam mengklasifikasikan status kredit nasabah bank.

1.2 Batasan Penelitian

Batasan masalah pada penelitian ini adalah data yang digunakan merupakan data sekunder hasil penelitian Dr. Hans Hofmann dari universitas Hamburg, Jerman. Data yang digunakan berasal dari 1000 nasabah yang mengisi survei dengan dua puluh variabel independent dan satu variabel dependent.

1.3 Tujuan Penelitian

Dengan memperhatikan latar belakang dan rumusan masalah di atas, maka tujuan penelitian ini ialah :

1. Untuk mendapatkan tingkat akurasi, sensitivitas, dan specifitis dengan menggunakan model CART tanpa penanganan data tidak seimbang .
2. Untuk mendapatkan tingkat akurasi, sensitivitas, dan specifitis dengan menggunakan model CART dengan penanganan *random undersampling* pada data tidak seimbang.
3. Untuk mendapatkan tingkat akurasi, sensitivitas, dan specifitis dengan menggunakan model CART dengan penanganan random over sampling pada data tidak seimbang.

1.4 Manfaat Penelitian

Manfaat yang diharapkan dari hasil penelitian ini adalah sebagai berikut :

1. Memberikan informasi mengenai penanganan data tidak seimbang dengan menggunakan metode *random oversampling* dan *random undersampling*.
2. Memberikan informasi mengenai klasifikasi CART melalui penyeimbangan kelas data.
3. Memberikan gambaran variabel yang berpengaruh dalam klasifikasi status kredit nasabah bank.

1.5 Teori

1.5.1 Data Mining

Data Mining merupakan tahapan untuk menemukan korelasi, pola, dan tren terhadap suatu hal berdasarkan sumber data yang ada (Liu, 2007). Dengan kata lain, data mining adalah proses untuk mencari sebuah informasi atau pengetahuan yang belum diketahui dari data yang ada, kemudian menghasilkan informasi yang berguna untuk keperluan pengolahan data selanjutnya. Rangkaian proses data mining secara keseluruhan meliputi seleksi data, preprocessing, transformasi data, pencarian informasi, dan evaluasi atau interpretasinya (Kurniawansyah, 2018).

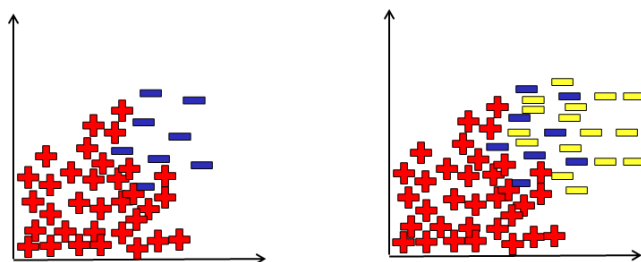
Dalam data mining sering dijumpai dataset yang tidak seimbang. Menurut Hairani et al. (2020), data tidak seimbang adalah suatu kondisi di mana jumlah kemunculan kelas mayoritas lebih besar dari jumlah kemunculan kelas minoritas.. Kelas dengan lebih banyak objek disebut kelas mayoritas, sedangkan kelas dengan lebih sedikit objek disebut kelas minoritas. Memproses dataset tidak seimbang tanpa

memperbaiki masalah ketidak seimbangan data akan cenderung memprediksi sebuah objek sebagai kelas mayoritas dan mengabaikan kelas minoritas (sani et al., 2023). Hal ini berarti model yang dibangun tidak mampu bekerja optimal dikelas minoritas.

Beberapa pendekatan untuk mengatasi ketidakseimbangan data meliputi pendekatan berbasis data, penggunaan algoritma, dan pendekatan hibrid yang menggabungkan keduanya. Pendekatan berbasis data bertujuan untuk menyeimbangkan kelas data secara eksternal dengan mengurangi jumlah sampel di kelas mayoritas atau memperbanyak data di kelas minoritas. Pendekatan ini sering digunakan untuk menangani ketidakseimbangan kelas data karena tidak bergantung pada metode analisis tertentu. Proses ini dilakukan dengan memodifikasi distribusi data pada tahap pra-pemrosesan (Yulianti et al., 2023). Terdapat tiga teknik utama dalam pendekatan berbasis data, yaitu *oversampling*, *undersampling*, dan *combine sampling*.

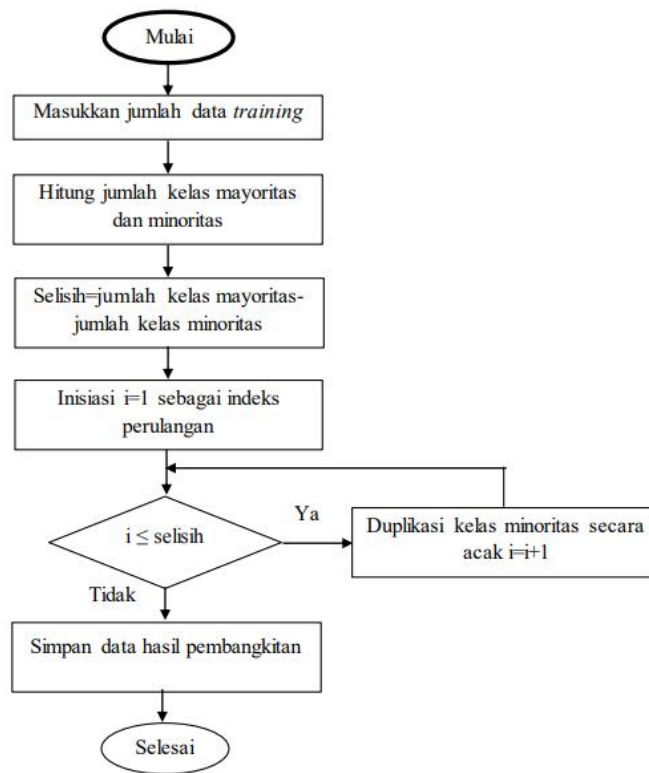
1.5.1.1 Random Oversampling

Random oversampling adalah teknik dalam data preprocessing yang digunakan untuk menangani masalah ketidakseimbangan kelas dalam dataset. Proses *random oversampling* ini dilakukan dengan cara mengulang sampel yang sudah ada secara acak pada kelas minoritas hingga jumlah sampel pada kelas minoritas sama dengan kelas mayoritas (Risk, 2023). Cara ini memungkinkan untuk menghasilkan dataset yang lebih seimbang dan meningkatkan performa model dalam mengenali kelas minoritas. Resampling dilakukan untuk mencapai hasil terbaik yang diusulkan. Ilustrasi *random oversampling* data dapat terlihat seperti pada gambar berikut :



Gambar 1. Ilustrasi proses *random oversampling*

Gambar 1 merupakan ilustrasi setelah dilakukan *random oversampling* dimana terdapat dua kelas yaitu kelas positif yang memiliki jumlah data mayoritas dan kelas negative yang memiliki jumlah data minoritas. Gambar pada sisi kiri merupakan data original sedangkan gambar pada sisi kanan merupakan data setelah dilakukan penambahan data baru menggunakan metode *random oversampling*. Pada gambar disebelah kanan dapat dilihat bahwa kelas minoritas berwarna kuning mempresentasikan data baru. Berikut adalah flowchart algoritma *random oversampling* (Fitriani dll,2021) :

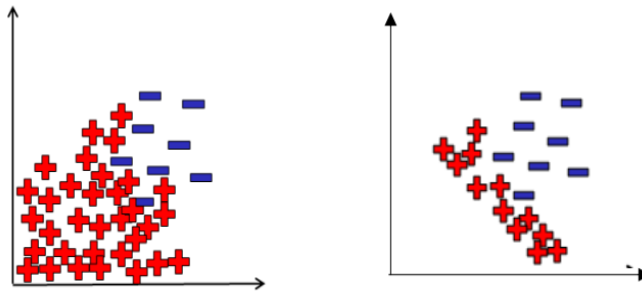


Gambar 2. Ilustrasi flowchart *random oversampling*

Kelebihan dari *random oversampling* adalah model ini dapat membantu meningkatkan jumlah data pada kelas minoritas, sehingga memungkinkan model untuk lebih memahami pola dari kelas tersebut. Teknik ini juga mempertahankan semua data asli, sehingga tidak ada informasi dari kelas mayoritas yang dihapus. Namun, kekurangan dari ROS adalah risiko overfitting pada kelas minoritas karena data yang diduplikasi terlalu sering muncul dalam proses pelatihan model.

1.5.1.2 *Random Undersampling*

Random undersampling adalah teknik dalam data preprocessing yang digunakan untuk menangani masalah ketidakseimbangan kelas dalam dataset. Proses random undersampling ini dilakukan dengan cara menghapus sampel yang sudah ada secara acak pada kelas mayoritas. Cara ini memungkinkan untuk menghasilkan dataset yang lebih seimbang dan meningkatkan performa model dalam mengenali kelas mayoritas.



Gambar 3. Ilustrasi proses random undersampling

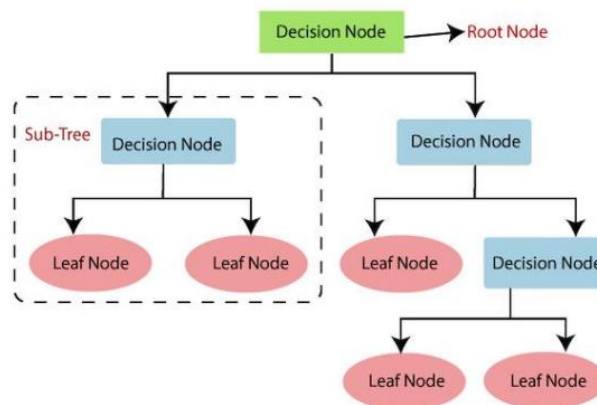
Gambar 3 merupakan ilustrasi setelah dilakukan *undersampling* dimana terdapat dua kelas yaitu kelas positif yang memiliki jumlah data mayoritas dan kelas negative yang memiliki jumlah data minoritas. Gambar pada sisi kiri merupakan data original sedangkan gambar pada sisi kanan merupakan data setelah dilakukan pembersihan menggunakan metode *undersampling*. merepresentasikan batas keputusan ideal sedangkan garis biru merupakan hasil aktual yang diperoleh. Metode *undersampling* menyebabkan beberapa informasi pada kelas negatif terhapus dan proporsi jumlah data pada kelas mayoritas serta minoritas lebih berimbang.

Kelahiran dari random undersampling adalah pengurangan waktu komputasi karena ukuran dataset yang lebih kecil, serta potensi untuk menghindari overfitting pada kelas mayoritas, yang dapat terjadi jika model terlalu banyak belajar dari data yang berlebihan. Namun, kekurangan teknik ini adalah hilangnya informasi penting, karena beberapa contoh yang mungkin berharga dihapus dari dataset. Hal ini dapat menyebabkan model kehilangan pola yang relevan, sehingga mengurangi akurasi dan kemampuan generalisasi pada data baru.

1.5.2 Decision Tree

Dalam mengaplikasikan klasifikasi, seringkali ditemukan data yang tidak memenuhi asumsi klasik metode klasifikasi, contohnya, asumsi kenormalan data dan multikolinearitas. Salah satu metode klasifikasi yang dapat menjadi solusi dari permasalahan tersebut adalah dengan menggunakan metode *decision tree*, karena *decision tree* bersifat non-parametrik. *Decision tree* adalah salah satu metode yang umum digunakan di berbagai bidang, seperti pembelajaran mesin, pemrosesan gambar, dan identifikasi pola (Fadhilah, 2023).

Klasifikasi pada metode ini menggunakan struktur pohon. Pada struktur pohon ini, terdapat node yang merepresentasikan atribut, cabangnya merepresentasikan nilai dari atribut, dan daun merepresentasikan kelas. Struktur *Decision Tree* dapat dilihat pada Gambar dibawah (Jijo & Abdulazeez, 2021):



Gambar 4. Struktur *Decision Tree*

Keterangan :

Root Node : Node teratas pada pohon keputusan yang mewakili awal dari proses pengambilan keputusan.

Decision Node : Node yang memiliki cabang ke node lain.

Leaf Node : Node yang tidak memiliki cabang, atau dahan yang mempresentasikan hasil akhir.

Visualisasi decision tree yang berbentuk pohon, terdiri dari simpul akar (root node), simpul keputusan (decision node), cabang (branches), dan simpul terminal (terminal nodes). Simpul akar (root node) merupakan simpul yang menjadi induk dari seluruh simpul pada decision tree. Setiap simpul pada decision tree mewakili variabel (attribute), simpul cabang merupakan suatu keputusan, sementara simpul terminal adalah hasil.

1.5.2.1 *Classification And Regression Tree*

Classification and Regression Tree (CART) merupakan salah satu metode atau algoritma dari teknik eksplorasi data yaitu pohon keputusan. CART terbilang sederhana namun merupakan metode yang kuat. CART bertujuan untuk mendapatkan suatu kelompok data yang akurat sebagai penciri dari suatu pengklasifikasian, selain itu CART digunakan untuk menggambarkan hubungan antara variabel respon dengan satu atau lebih variabel prediktor. Model pohon yang dihasilkan bergantung pada skala variabel respon, jika variabel respon data berbentuk kontinu maka model pohon yang dihasilkan adalah regression tree (pohon regresi) sedangkan bila variabel respon mempunyai skala kategorik maka pohon yang dihasilkan adalah classification tree (pohon klasifikasi).

CART mempunyai beberapa kelebihan dibandingkan metode klasifikasi lainnya, yaitu hasilnya lebih mudah diinterpretasikan, lebih akurat dan lebih cepat penghitungannya. Selain itu CART bisa diterapkan untuk himpunan data yang

mempunyai jumlah besar, variabel yang sangat banyak dan dengan skala variabel campuran melalui prosedur pemilahan biner. Data learning digunakan untuk pembentukan pohon klasifikasi optimal sedangkan data testing digunakan untuk validasi model yaitu seberapa besar kemampuan model dalam memprediksi data baru.

Metode CART yang dikembangkan oleh Leo Breiman, Jerome H. Friedman, Richard A. Olshen, dan Charles J. Stone ini merupakan teknik klasifikasi dengan menggunakan algoritma penyekatan rekursif secara biner (binary recursive partitioning) (Roger dan Lewis, 2000). Istilah “binary” berarti pemilahan dilakukan pada sekelompok data yang terkumpul dalam suatu ruang yang disebut simpul (node) menjadi dua kelompok yang disebut simpul anak (child nodes). Istilah “recursive” berarti prosedur penyekatan secara biner dilakukan secara berulang-ulang. Model pohon yang diperoleh akan terbentuk berdasarkan tipe data dari variabel dependen. Jika variabel dependen memiliki tipe data numerik maka model pohon yang diperoleh adalah regression trees. Kemudian, jika variabel dependen memiliki tipe data kategorik maka model pohon yang diperoleh adalah classification tree (Fadhilah, 2023). CART mempunyai tujuan untuk mendapatkan kumpulan data yang akurat sebagai prediktor suatu klasifikasi. Menurut Yulianti dll (2023), proses pembentukan pohon klasifikasi CART terdiri atas 3 tahapan yaitu sebagai berikut :

a). Menentukan kandidat split

Parameter yang akan digunakan dalam metode CART dapat ditentukan dengan cara melakukan pembagian kandidat split menjadi dua bagian, yaitu kandidat split kanan dan kandidat split kiri berdasarkan variabel independen. Parameter yang telah terpilih akan terbentuk menjadi suatu himpunan kelas yang disebut sebagai simpul. Simpul yang telah diperoleh akan melakukan proses pemilihan secara iterasi hingga akan didapat iterasi terakhir yang menghasilkan simpul terminal. Teknik untuk menentukan kandidat split pada algoritma CART dilakukan dengan menggunakan indeks gini $i(t)$. Indeks gini merupakan suatu pengukuran dari variasi antara probabilitas nilai-nilai variabel independen (Yulianti et al., 2023). Indeks gini digunakan untuk mendapatkan kandidat split yang paling optimal. Menurut Fadhilah (2023) persamaan indeks gini yaitu:

$$i(t) = 1 - \sum_j P^2(j|t) \quad (1)$$

dengan

$$P^2(j|t) = \frac{n_j(t)}{n_t} \quad (2)$$

Keterangan :

$i(t)$	= Indeks gini
$P^2(j t)$	= Jumlah kelas calon simpul kanan dengan j terhadap simpul t dan calon simpul kiri dengan j terhadap simpul t
$n_j(t)$	= Jumlah data kelas j pada simpul t
$n(t)$	= Jumlah data pada simpul t

Tahap berikutnya yaitu melakukan perhitungan *gini splitting index* yang akan berperan sebagai tahap evaluasi pemilihan kandidat *split s* pada *t* menggunakan persamaan dibawah ini:

$$Gini\ splitting\ index = \sum_{j=1}^n \frac{n_t}{n} i(t) \quad (3)$$

Keterangan :

- $i(t)$ = Indeks gini
- $n(t)$ = Jumlah data pada simpul *t*
- n = Jumlah data

- b). Menentukan simpul terminal

Proses untuk menentukan pola decision tree akan dihentikan ketika sampai pada jumlah iterasi dimana sebaran sampel telah berada pada satu kelas, dimana simpul terakhir disebut sebagai simpul terminal.

- c). Menentukan label kelas

Suatu proses untuk menentukan label kelas dilakukan pada tiga macam simpul, yaitu simpul akar, simpul keputusan dan juga simpul terminal. Proses menentukan label kelas ini sangat penting untuk dilakukan pada simpul terminal, karena pada simpul terminal menggunakan label untuk melakukan prediksi objek pada kelas yang terdapat pada simpul terminal tersebut (Yulianti et al., 2023). Dibawah ini merupakan persamaan yang digunakan untuk menentukan label kelas :

$$P(j|t) = \max_j \frac{n_j(t)}{n_t} \quad (4)$$

Keterangan:

- $P(j|t)$ = Jumlah kelas *j* pada simpul *t*
- $n_j(t)$ = Jumlah kelas pada simpul *t*
- n_t = Jumlah pengamatan di simpul *t*

1.5.3 Evaluasi Model

Evaluasi kinerja dilakukan dengan memeriksa sejauh mana model yang dihasilkan oleh algoritma mampu berperforma. Salah satu metode untuk melakukan evaluasi adalah dengan melihat nilai dari *confusion matrix*. *Confusion matrix* adalah informasi mengenai hasil klasifikasi aktual yang dapat diprediksi oleh sebuah sistem klasifikasi. *Confusion Matrix* dapat digunakan sebagai parameter utama untuk mengevaluasi model tersebut. Pada Tabel 1 akan ditampilkan tabel *confusion matrix* dua kelas yakni kelas negative dan positif.

Dalam fase ini, pengujian akan menghasilkan nilai akurasi, *sensitivity*, dan *specificity*. Akurasi mencerminkan proporsi prediksi yang benar, baik yang bersifat positif maupun negatif, dari seluruh data. *Sensitivity* mencerminkan proporsi prediksi positif yang benar dari seluruh data yang sebenarnya positif. *Specificity* mencerminkan proporsi prediksi negatif yang benar dari seluruh data yang sebenarnya negative. Menurut Kusnaeni (2023), rumus untuk mencari nilai akurasi, *sensitivity*, dan *specificity* adalah dengan menggunakan Persamaan :

$$akurasi = \frac{Tp + Tn}{Tp + Fp + Fn + Tn} \times 100\% \quad (5)$$

$$sensitivity = \frac{Tp}{Tp + Fn} \times 100\% \quad (6)$$

$$specificity = \frac{Tn}{Tn + Fp} \times 100\% \quad (7)$$

Melalui evaluasi ini, dapat diperoleh pemahaman yang lebih mendalam tentang seberapa baik model mampu memprediksi dengan benar dan seberapa baik kinerja model dalam mengidentifikasi kelas tertentu, baik yang positif maupun negatif.

1.5.4 Kredit Nasabah Bank

Bank dalam kegiatan operasionalnya berperan sebagai lembaga intermediasi (mediator). Peranan ini berhubungan langsung dengan kegiatan utama bank yaitu, menghimpun dana dari masyarakat yang memiliki dana berlebih dan menyalurkannya kepada masyarakat yang membutuhkan dana. Peranan ini dilakukan oleh bank untuk memperlancar alur pembayaran dan pelayanan kepada masyarakat. Tujuan dari peranan ini adalah menghasilkan laba bagi perbankan. Selain itu, bank juga memiliki peran yang sangat penting dalam pertumbuhan perekonomian sebuah negara. Hal ini dikarenakan pengukuran pertumbuhan perekonomian suatu negara dapat dipengaruhi oleh pertumbuhan bank.

Peran perbankan sebagai lembaga intermediasi dalam menyalurkan dana kepada masyarakat dalam bentuk investasi aset produk dapat mendorong produktivitas sektor riil, akumulasi capital, dan berujung pada pertumbuhan agregat (Hung dan Cothorn, 2002). perilaku lembaga keuangan dalam menawarkan kredit yang akan diberikan kepada nasabah tergantung pada suku bunga dan risiko alokasi penyaluran kredit tersebut. Semakin tinggi suku bunga akan menurunkan minat masyarakat dalam mengajukan kredit, karena masyarakat akan memilih untuk menggunakan dananya sebagai kebutuhan lain daripada harus membayar bunga kredit yang tidak dapat dijangkau. Tingginya risiko alokasi penyaluran kredit akan menurunkan jumlah kredit yang akan ditawarkan, karena pihak lembaga keuangan akan mengurangi risiko tersebut ketika penawaran kredit dikurangi (Amelian Noli, 2015).

Permintaan kredit yang didasarkan pada perjanjian antara lembaga keuangan dan pihak lain (nasabah) yang mengharuskan peminjam untuk menyediakan mata uang yang dikenakan bunga atau klaim yang setara dengan jangka waktu tertentu (Sarifudin, 2007). Permintaan kredit diartikan sebagai pinjaman yang diberikan oleh pihak yang mempunyai dana dan pihak yang membutuhkan dana (Suseno dan Piter Abdullah, 2003).

Kredit adalah salah satu penentu persetujuan pinjaman yang dapat disetujui oleh bank. Skor kredit yang bagus artinya nasabah berpeluang besar diterima dalam pengajuan pinjaman, sementara skor kredit yang buruk artinya nasabah dianggap kurang layak. Pada penelitian ini variabel variabel yang mempengaruhi Kredit data jerman adalah :

1) Status rekening

Status rekening dalam penelitian ini memakai skala kategorik yang bermakna status rekening nasabah tersebut aktif atau tidak aktif. Ini dapat dilihat dari banyaknya jumlah tabungan (uang cash) yang ada dalam rekening nasabah tersebut. Jumlah tabungan dapat memberikan gambaran tentang stabilitas finansial individu. Berikut beberapa alasan mengapa jumlah tabungan berpengaruh kepada kredit :

- Likuiditas dan kemampuan membayar, nasabah dengan dengan jumlah tabungan yang besar cenderung memiliki cadangan dana yang bisa digunakan untuk memenuhi kewajiban finansial
- Stabilitas finansial.

2) Durasi tabungan

Durasi tabungan adalah lama nasabah menabung dalam bank tersebut. Menurut Aduda dan Obondy (2021), durasi lama tabungan dapat mempengaruhi kredit nasabah bank karena bank dapat mengetahui kemampuan nasabah dalam mengelola keuangannya dengan melihat durasi tabungan. Berikut beberapa alasan mengapa jumlah tabungan berpengaruh kepada kredit :

- Pengalaman keuangan, Nasabah yang telah menabung dalam jangka waktu yang lama mungkin lebih berpengalaman dalam membuat keputusan keuangan yang lebih bijaksana.
- Kesiapan darurat, artinya nasabah dapat menghadapi situasi mendesak tanpa berhutang.
- Persepsi pemberi pinjaman, nasabah dengan riwayat tabungan yang cukup lama dianggap memiliki resiko gagal bayar yang lebih rendah.

3) Sejarah kredit nasabah

Sejarah kredit nasabah berarti riwayat kredit yang pernah diambil nasabah. Menurut Aduda dan Obondy (2021), Sejarah kredit nasabah dapat memperlihatkan kemampuan nasabah dalam mengelola keuangannya. Penelitian ini memasukkan aspek riwayat pengambilan kredit nasabah dengan mempertimbangkan beberapa aspek yang mencerminkan tanggung jawab dan disiplin nasabah dalam mengelola kewajiban finansial mereka. Seperti riwayat pengambilan kredit dan status kredit (lunas/belum lunas).

4) Tujuan pengajuan kredit

Tujuan pengambilan kredit seseorang dapat berpengaruh signifikan terhadap nilai kreditnya karena berbagai alasan yang berkaitan dengan risiko, manajemen keuangan, dan kemampuan peminjam dalam memenuhi kewajiban. Tujuan pengambilan kredit dalam penelitian ini menggunakan skala kategorik. Alasan pengambilan kredit setiap individu tentu berbeda beda tergantung kebutuhannya.

5) Jumlah kredit sekarang

Jumlah kredit yang dimiliki oleh nasabah saat ini adalah faktor penting yang mempengaruhi keputusan kredit. Rasio kredit sekarang terhadap permintaan kredit baru disebut *Debt-to-Income Ratio*. Ini adalah rasio penting yang digunakan oleh bank untuk menilai risiko pemberian kredit baru. Rasio ini menghitung jumlah kredit yang saat ini dimiliki seseorang terhadap total pendapatan mereka, serta terhadap kredit baru yang diminta. Jika nasabah sudah memiliki kredit yang besar, kemungkinan risiko gagal bayar akan meningkat. Bank cenderung lebih berhati-hati memberikan kredit tambahan untuk menghindari peningkatan risiko kredit macet.

$$Debt - to - Income Ratio = \frac{Jumlah\ Kredit\ Sekarang}{Pendapatan\ Kotor\ Tahunan} \times 100 \quad (8)$$

6) Jumlah tabungan/obligasi

Jumlah tabungan dan obligasi mencerminkan likuiditas dan stabilitas keuangan nasabah. Rasio yang digunakan dalam mengukur jumlah obligasi terhadap permintaan kredit baru disebut *Bond-To-Loan Ratio*. Jika seseorang atau entitas memiliki investasi dalam obligasi, rasio ini dapat menunjukkan hubungan antara nilai obligasi yang dimiliki dan kredit baru yang diminta. Semakin besar jumlah aset yang dimiliki, semakin rendah risiko gagal bayar, karena nasabah memiliki aset yang bisa dijadikan jaminan (Berger dan Udell, 2006).

$$Bond - To - Loan Ratio = \frac{Jumlah\ Obligasi}{Permintaan\ Kredit\ Baru} \quad (9)$$

7) Durasi lama kerja (kualitatif)

Durasi lama kerja menunjukkan stabilitas pekerjaan nasabah. Menurut Muellbauer (2012), Nasabah yang telah bekerja dalam jangka waktu lama menunjukkan stabilitas ekonomi yang lebih baik dan Stabilitas pekerjaan dianggap sebagai tanda bahwa nasabah memiliki penghasilan tetap.

8) Tingkat angsuran dalam persentase pendapatan yang diterima

Tingkat angsuran dalam persentase pendapatan yang diterima bermakna perbandingan angsuran kredit yang diambil dengan jumlah pendapatan sekarang kemudian dijadikan dalam bentuk persentase. Rasio pendapatan angsuran kredit dengan pendapatan yang lebih tinggi dapat menunjukkan

ketidakmampuan nasabah untuk memenuhi kewajiban keuangan lainnya (Siregar et al., 2024).

$$\text{Tingkat Angsuran} = \frac{\text{Jumlah Angsuran Bulanan}}{\text{Pendapatan Bulanan}} \times 100 \quad (10)$$

Keterangan :

Jumlah Angsuran Bulanan = Total pembayaran kredit bulanan yang harus dilakukan oleh debitur, termasuk pokok dan bunga.

Pendapatan Bulanan = Total pendapatan bulanan yang diterima oleh individu atau rumah tangga, sebelum pajak (pendapatan bruto).

Tingkat angsuran yang sehat biasanya berada di bawah 30% dari pendapatan bulanan. Jika tingkat angsuran berada di atas 30%, maka hal ini menunjukkan bahwa debitur mungkin menghadapi risiko kesulitan keuangan dalam memenuhi kewajiban kredit (Mae, 2024).

- 9) Status pernikahan dan jenis kelamin
Status pernikahan dan jenis kelamin juga dapat mempengaruhi penilaian kredit. Penelitian menunjukkan bahwa pasangan yang menikah lebih kecil kemungkinannya untuk mengalami gagal bayar karena adanya dukungan finansial ganda (Siregar et al., 2024).
- 10) Debitur lain/Penjamin
Debitur lain/Penjamin dalam analisis kredit nasabah bank mengacu pada keberadaan pihak ketiga yang turut terlibat dalam perjanjian kredit, seperti pemohon bersama atau penjamin. Nasabah yang memiliki permohonan kredit bersama atau penjamin dianggap lebih aman, karena bank memiliki lebih dari satu pihak yang bertanggung jawab atas pembayaran kembali kredit.
- 11) Sejak kapan tinggal di tempat tinggal sekarang
Durasi tempat tinggal menunjukkan stabilitas sosial dan finansial nasabah. Nasabah yang sering berpindah-pindah memiliki profil risiko lebih tinggi karena kurang menunjukkan keterikatan dan stabilitas.
- 12) Jenis properti yang dimiliki
Kepemilikan properti merupakan tanda bahwa nasabah memiliki aset yang dapat digunakan sebagai jaminan. Properti biasanya menjadi pertimbangan penting karena menunjukkan kemampuan ekonomi yang baik
- 13) Usia nasabah saat ini
Usia nasabah menjadi pertimbangan bank karena berkaitan dengan stabilitas pendapatan dan risiko hidup. Nasabah yang lebih muda mungkin dianggap kurang stabil secara finansial, sementara nasabah yang lebih tua mungkin dianggap berisiko tinggi karena mendekati masa pensiun. Berikut beberapa alasan mengapa jumlah tabungan berpengaruh kepada kredit :
 - Pengalaman keuangan, Nasabah yang lebih tua umumnya memiliki pengalaman lebih banyak dalam mengelola keuangan dan kredit.
 - Stabilitas keuangan, Usia yang lebih tua sering kali berkaitan dengan stabilitas finansial yang lebih baik. Nasabah yang telah berada di pasar kerja lebih lama biasanya memiliki penghasilan yang lebih stabil, yang mempengaruhi kemampuan mereka untuk membayar kembali utang.

- Risiko kredit, Pemberi pinjaman cenderung melihat nasabah yang lebih tua sebagai risiko yang lebih rendah. Hal ini disebabkan oleh kemungkinan yang lebih kecil untuk mengalami perubahan drastis dalam situasi finansial, seperti kehilangan pekerjaan.
- 14) Paket cicilan lainnya
Menurut Iriani et al (2023), Jumlah paket cicilan lainnya yang dimiliki nasabah akan mempengaruhi rasio utang terhadap pendapatan. Jika terlalu banyak cicilan, risiko gagal bayar semakin tinggi, dan bank mungkin ragu untuk memberikan pinjaman tambahan.
 - 15) Status rumah saat ini
Nasabah yang memiliki rumah cenderung dianggap lebih stabil, sementara mereka yang menyewa rumah mungkin dianggap memiliki tanggungan finansial lebih besar (Muellbauer, 2012).
 - 16) Jumlah kredit yang ada di bank saat ini
Jumlah kredit yang ada di bank saat ini menjadi indikator seberapa besar komitmen keuangan nasabah terhadap bank tersebut. Menurut Amelia (2015), Semakin besar jumlah kredit yang ada, semakin hati-hati bank dalam memberikan pinjaman tambahan.
 - 17) Pekerjaan yang di laksanakan saat ini
Jenis pekerjaan memberikan indikasi stabilitas pendapatan. Pekerjaan dengan pendapatan tetap dan stabilitas karir, seperti pekerjaan di sektor formal, akan memberikan nilai lebih baik dalam analisis risiko kredit (Baensens et al., 2003).
 - 18) Jumlah anggota keluarga yang jadi tanggungan
Jumlah tanggungan mempengaruhi kemampuan nasabah dalam membayar pinjaman. Semakin banyak tanggungan, semakin besar beban keuangan nasabah, yang dapat meningkatkan risiko gagal bayar (Berger dan Udell, 2006).
 - 19) Apakah memakai telepon
Penggunaan telepon sebagai variabel mungkin berfungsi sebagai indikator stabilitas kontak dan keandalan nasabah. Nasabah yang memiliki sarana komunikasi yang stabil mungkin dianggap lebih mudah dihubungi jika terjadi masalah dengan pembayaran (Hand dan Henley, 1997).
 - 20) Apakah nasabah merupakan pekerja asing
Kewarganegaraan dapat memiliki pengaruh signifikan terhadap kredit dan pinjaman. Dalam beberapa kasus, kewarganegaraan asing dapat membuat posisi perlindungan hukum lebih lemah, sehingga mempengaruhi kemampuan untuk mendapatkan kredit atau pinjaman. Berikut contoh beberapa pengaruh kewarganegaraan terhadap kredit dan pinjaman:
 - Kemampuan untuk mendapatkan kredit dengan suku bunga yang lebih rendah bagi warga negara lokal.
 - Warga negara asing mungkin tidak dapat memberikan jaminan hukum yang sama dengan warga negara lokal.
 - Kewarganegaraan dapat mempengaruhi pengawasan yang diterapkan oleh lembaga keuangan. Warga negara asing mungkin lebih banyak diawasi daripada warga negara lokal.

21) Status kredit nasabah

Status kredit mengacu pada kondisi keuangan dan perilaku pembayaran nasabah dalam mengelola pinjaman atau kredit yang telah diberikan. Ini mencakup berbagai informasi mengenai riwayat pembayaran, jumlah utang, dan lainnya. Lembaga keuangan menggunakan status kredit untuk menilai risiko pemberian kredit kepada nasabah. Klasifikasi status kredit dalam penelitian ini dibagi menjadi dua yakni status kredit buruk dan status kredit baik.

BAB II

METODE PENELITIAN

2.1 Jenis dan Sumber Data

Penelitian ini menggunakan data sekunder yang diperoleh dari website pendidikan yang banyak memuat hasil penelitian (<https://archive.ics.uci.edu/>). Data yang digunakan merupakan hasil penelitian Professor Hans Hoffman terhadap 1000 nasabah bank Jerman pada tahun 1996, dimana data yang dihasilkan adalah data Status kredit nasabah yang terdiri dari 20 variabel independent dan 1 variabel dependen. Status kredit nasabah bank dikategorikan menjadi 2 yaitu status kredit buruk dan status kredit baik.

2.2 Variabel Penelitian

Variabel yang digunakan dalam penelitian ini ada pada tabel berikut :

Tabel 2 Variabel Penelitian

Nama Variabel	Variabel	Skala
X_1	Status Rekening	Kategorik 1 = Tidak ada rekening giro 2 = $0 < \dots \leq 100$ Deutsche Mark 3 = $100 < \dots \leq 200$ Deutsche Mark 4 = $\dots > 200$ Deutsche Mark
X_2	Durasi tabungan	Numerik
X_3	Sejarah kredit nasabah	Kategorik 1 = tidak ada kredit yang diambil 2 = Semua kredit telah dibayar lunas 3 = Masih terdapat kredit yang belum lunas 4 = Memiliki riwayat keterlambatan pelunasan kredit dimasa lalu 5 = Akun kritis, artinya nasabah sempat mengalami gagal bayar
X_4	Tujuan pengajuan kredit	Kategorik 1 = pembelian mobil baru 2 = pembelian mobil bekas 3 = pembelian perabotan rumah tangga seperti lemari, kursi, meja dan lain lain 4 = Pembelian radio/ televisi 5 = Pembelian peralatan rumah tangga seperti kompor, mesin cuci dan lain lain

		6 = Perbaikan/renovasi rumah 7 = Pendidikan 8 = Liburan 9 = Kursus/pelatihan kompetensi 10 = Bisnis 20 = Lainnya
X_5	Jumlah kredit sekarang	Numerik
X_6	Jumlah tabungan/obligasi	Kategorik 1 = ... $\leq$ 100 Deutsche Mark 2 = $100 < \dots \leq 500$ Deutsche Mark 3 = $500 < \dots \leq 1000$ Deutsche Mark 4 = ... > 1000 Deutsche Mark 5 = Tidak memiliki tabungan
X_7	Durasi lama kerja (kualitatif)	Kategorik 1 = Belum bekerja 2 = ... < 1 tahun 3 = $1 \leq \dots < 4$ tahun 4 = $4 \leq \dots < 7$ tahun 5 = ... ≥ 7 tahun
X_8	Tingkat angsuran dalam persentase pendapatan yang diterima	Numerik
X_9	Status pribadi dan jenis kelamin	Kategorik 1 = Laki – laki (cerai/pisah) 2 = Perempuan (cerai/pisah/menikah) 3 = Laki – laki (lajang) 4 = Laki – laki (menikah/duda ditinggal mati) 5 = Perempuan (lajang)
X_{10}	Penjamin nasabah	Kategorik 1 = Tidak ada penjamin 2 = Permohonan bersama (<i>co-applicant</i>) 3 = Ada penjamin
X_{11}	Sejak kapan tinggal di tempat tinggal sekarang	Numerik
X_{12}	Jenis properti yang dimiliki	Kategorik 1 = real estate 2 = tabungan / asuransi jiwa 3 = mobil atau lainnya 4 = tidak punya property

X_{13}	Usia nasabah saat ini	Numerik
X_{14}	Paket cicilan lainnya	Kategorik 1 = Bank 2 = Pembiayaan 3 = Tidak ada
X_{15}	Status rumah saat ini	Kategorik 1 = Sewa 2 = Milik sendiri 3 = Gratis
X_{16}	Jumlah kredit yang ada di bank saat ini	Numerik
X_{17}	Pekerjaan yang di laksanakan saat ini	Kategorik 1 = Menganggur / tidak terampil (bukan penduduk) 2 = Tidak terampil (penduduk) 3 = Pegawai / pejabat terampil 4 = Manajemen / wiraswasta / petugas yang berkualifikasi tinggi
X_{18}	Jumlah anggota keluarga yang jadi tanggungan	Numerik
X_{19}	Apakah memakai telepon	Kategorik 1 = Tidak ada 2 = Ya, terdaftar atas nama pelanggan
X_{20}	Apakah nasabah merupakan pekerja asing	Kategorik 1 = Ya 2 = Tidak
Y	Status kredit nasabah	Kategorik 0 = Buruk 1 = Baik

2.3 Metode Penelitian

Langkah – langkah yang dilakukan pada penelitian untuk analisis data dengan metode CART adalah sebagai berikut :

1. Mengumpulkan data sekunder status nasabah bank yang diperoleh dari situs <https://archive.ics.uci.edu>.
2. Melakukan *pra-processing* data meliputi pengumpulan data, transformasi manual untuk mengkodekan variabel kategorik dan numerik dan pemisahan variabel

independen dan dependen. Proses ini melibatkan pemberian bobot atau nilai numerik yang sesuai untuk setiap kategori dalam kuesioner.

3. Membagi data menjadi data latih dan data uji. Data latih digunakan untuk melatih model, sedangkan data uji digunakan untuk menguji model yang telah dilatih untuk mengetahui seberapa baik mereka berfungsi. Pada penelitian ini data latih dan data uji akan dibagi dengan proporsi 70% data latih dan 30%.
4. Menggunakan data latih untuk memodelkan data dalam metode CART. Pada tahap ini, data latih diproses melalui algoritma CART untuk menghasilkan pohon keputusan yang menggambarkan hubungan antara variabel input (prediktor) dan variabel target (respon). Model yang dihasilkan dari data latih ini selanjutnya akan diuji untuk melihat performanya menggunakan data uji.
5. Menampilkan confusion matriks. Confusion matriks adalah tabel yang digunakan untuk mengevaluasi kinerja model klasifikasi.
6. Menghitung selang kepercayaan model. Selang kepercayaan digunakan untuk memberikan perkiraan yang lebih akurat tentang seberapa baik model kita akan bekerja pada data yang tidak terlihat. Dalam konteks klasifikasi, ini membantu mengukur seberapa stabil kinerja model ketika diterapkan pada data baru.
7. Selesai.

Langkah – langkah yang dilakukan pada penelitian untuk analisis data dengan metode CART dengan *random oversampling* adalah sebagai berikut :

1. Mengumpulkan data sekunder status nasabah bank yang diperoleh dari situs <https://archive.ics.uci.edu>.
2. Melakukan *pra-processing* data meliputi pengumpulan data, transformasi manual untuk mengkodekan variabel kategorik dan numerik dan pemisahan variabel independen dan dependen. Proses ini melibatkan pemberian bobot atau nilai numerik yang sesuai untuk setiap kategori dalam kuesioner.
3. Membagi data menjadi data latih dan data uji. Data latih digunakan untuk melatih model, sedangkan data uji digunakan untuk menguji model yang telah dilatih untuk mengetahui seberapa baik mereka berfungsi. Pada penelitian ini data latih dan data uji akan dibagi dengan proporsi 70% data latih dan 30%.
4. Data latih diolah dengan *random oversampling* sehingga menghasilkan data sintesis baru.
5. Menggunakan data latih untuk memodelkan data dalam metode CART. Pada tahap ini, data latih diproses melalui algoritma CART untuk menghasilkan pohon keputusan yang menggambarkan hubungan antara variabel input (prediktor) dan variabel target (respon). Model yang dihasilkan dari data latih ini selanjutnya akan diuji untuk melihat performanya menggunakan data uji.
6. Menampilkan confusion matriks. Confusion matriks adalah tabel yang digunakan untuk mengevaluasi kinerja model klasifikasi.
7. Menghitung selang kepercayaan model. Selang kepercayaan digunakan untuk memberikan perkiraan yang lebih akurat tentang seberapa baik model kita akan

bekerja pada data yang tidak terlihat. Dalam konteks klasifikasi, ini membantu mengukur seberapa stabil kinerja model ketika diterapkan pada data baru.

8. Selesai.

Langkah – langkah yang dilakukan pada penelitian untuk analisis data dengan metode CART dengan *random undersampling* adalah sebagai berikut :

1. Mengumpulkan data sekunder status nasabah bank yang diperoleh dari situs <https://archive.ics.uci.edu>.
2. Melakukan *pra-processing* data meliputi pengumpulan data, transformasi manual untuk mengkodekan variabel kategorik dan numerik dan pemisahan variabel independen dan dependen. Proses ini melibatkan pemberian bobot atau nilai numerik yang sesuai untuk setiap kategori dalam kuesioner.
3. Membagi data menjadi data latih dan data uji. Data latih digunakan untuk melatih model, sedangkan data uji digunakan untuk menguji model yang telah dilatih untuk mengetahui seberapa baik mereka berfungsi. Pada penelitian ini data latih dan data uji akan dibagi dengan proporsi 70% data latih dan 30%.
4. Data latih diolah dengan *random undersampling* sehingga menghasilkan data sintesis baru.
5. Menggunakan data latih untuk memodelkan data dalam metode CART. Pada tahap ini, data latih diproses melalui algoritma CART untuk menghasilkan pohon keputusan yang menggambarkan hubungan antara variabel input (prediktor) dan variabel target (respon). Model yang dihasilkan dari data latih ini selanjutnya akan diuji untuk melihat performanya menggunakan data uji.
6. Menampilkan confusion matriks. Confusion matriks adalah tabel yang digunakan untuk mengevaluasi kinerja model klasifikasi.
7. Menghitung selang kepercayaan model. Selang kepercayaan digunakan untuk memberikan perkiraan yang lebih akurat tentang seberapa baik model kita akan bekerja pada data yang tidak terlihat. Dalam konteks klasifikasi, ini membantu mengukur seberapa stabil kinerja model ketika diterapkan pada data baru.
8. Selesai.