

BAB III

METODOLOGI PENELITIAN

Pada bab ini membahas mengenai metode yang digunakan untuk menyelesaikan masalah risiko kredit yang diawali dengan mengidentifikasi masalah, menjelaskan tahapan penelitian, dan menjelaskan penyelesaian model metode Naïve Bayes.

3.1. Metode Penelitian

Pada penelitian ini, digunakan metode penelitian dengan pendekatan kuantitatif dan menggunakan desain penelitian studi kasus. Adapun metode yang dilakukan meliputi:

1. Metode yang diterapkan dalam penelitian ini menggunakan desain studi kasus untuk menggambarkan dan menganalisis fenomena yang terjadi pada suatu perusahaan dalam konteks penggunaan model *Naïve Bayes* untuk klasifikasi risiko kredit.
2. Sumber dan teknik pengambilan sampel sumber data pada penelitian ini merupakan data sekunder yang diperoleh dari laporan keuangan dan data riwayat kredit perusahaan yang menjadi objek penelitian. Teknik pengambilan sampel yang digunakan adalah *purposive sampling*, yaitu memilih sampel dengan kriteria tertentu. Data bersumber dari perusahaan yang memiliki data kredit dan laporan keuangan lengkap tahun 2018.
3. Pengumpulan Data dan Analisis Data Pengumpulan data dilakukan dengan cara studi literatur, yaitu mengumpulkan data dari laporan keuangan dan data riwayat kredit perusahaan. Kemudian data akan dianalisis dengan menggunakan model *Naïve Bayes* untuk klasifikasi risiko kredit pada perusahaan yang menjadi objek penelitian.
4. Validitas dan reliabilitas data. Validitas data dalam penelitian ini akan dijaga dengan cara memastikan bahwa data yang digunakan valid, akurat, dan relevan dengan penelitian. Reliabilitas data dalam penelitian ini dijaga dengan cara

memastikan bahwa data yang digunakan dapat diandalkan dan konsisten. Validitas dan reliabilitas data akan diuji dengan menggunakan teknik analisis data yang tepat dan mengikuti standar penelitian yang berlaku.

3.2. Tahapan Penelitian

Penelitian ini menggunakan bantuan aplikasi pemrograman *python* yaitu jupyter notebook. Adapun tahapan penelitian yang dilakukan pada penelitian ini sebagai berikut.

1. Studi Pustaka

Pada tahap ini dilakukan pencarian dan kajian literatur yang relevan dengan topik penelitian mengenai risiko kredit dan model Naïve Bayes yang bersumber dari jurnal, buku, disertasi, dan karya tulis lainnya.

2. Pengumpulan Data

Jenis data yang dalam penelitian ini merupakan data sekunder. Data yang digunakan adalah data informasi mengenai peminjam perusahaan *Home Credit* tahun 2018. Data tersebut diperoleh dari situs web *kaggle* (<https://www.kaggle.com/>) yang diakses pada tanggal 6 Februari 2023. Dalam penelitian ini digunakan 307.511 peminjam dengan 122 variabel mengenai latar belakang peminjam dengan variabel “Target” sebagai variabel dependen dan 121 kolom lainnya sebagai variabel independen.

3. *Preprocessing* data

Tahap ini dilakukan untuk membersihkan, merapikan, dan mempersiapkan data sebelum diterapkan pada model sehingga dapat menghasilkan informasi yang berguna dan akurat. *Preprocessing* data pada penelitian ini terdiri dari beberapa tahapan, yaitu:

- a. Eksplorasi data yang bertujuan untuk memahami dan mengetahui informasi yang ada pada data.
- b. Pembersihan data dari data *outlier*, data yang tidak lengkap dan data yang tidak valid.

- c. Pemilihan fitur data yang akan digunakan sebagai variabel independen dalam model *Naïve Bayes* menggunakan korelasi Pearson dengan rumus:

$$a. \quad r = \frac{(n \sum_{i=1}^n X_i Y_i - \sum_{i=1}^n X_i \sum_{i=1}^n Y_i)}{\sqrt{(n \sum_{i=1}^n X_i^2 - (\sum_{i=1}^n X_i)^2) (n \sum_{i=1}^n Y_i^2 - (\sum_{i=1}^n Y_i)^2)}}, i = 1, 2, \dots, n$$

di mana

r : koefisien korelasi

n : jumlah data

$\sum_{i=1}^n X_i Y_i$: jumlah dari hasil perkalian antara setiap nilai x dan y

$\sum_{i=1}^n X_i$: jumlah dari setiap nilai x

$\sum_{i=1}^n Y_i$: jumlah dari setiap nilai y

Pada tahap ini, korelasi pearson hanya menganalisis pada data yang numerik, sehingga data tipe objek dikelompokkan berdasarkan nilai Target 0 dan 1.

- d. Mengkategorikan data ke dalam kelompok yang memiliki kesamaan dalam karakteristik seperti tingkat pendidikan, jenis kelamin, dan penghasilan.

- e. Penyeimbangan data menggunakan metode SMOTE. Misalkan diberikan dua data dengan p dimensi yaitu $x^T = [x_1, x_2, \dots, x_p]$ dan $y^T = [y_1, y_2, \dots, y_p]$ maka jarak euclidan $d(x, y)$ antara kedua vektor data adalah $d(x, y) =$

$$\sqrt{(x_1 - y_1)^2 + (x_2 - y_2)^2 + \dots + (x_p - y_p)^2} \quad \text{sedangkan synthetic data}$$

dilakukan dengan menggunakan persamaan berikut (Pitaloka, 2021)

$$x_{syn} = x_i + (x_{knn} - x_i)x\beta, i = 1, 2, \dots, n$$

di mana

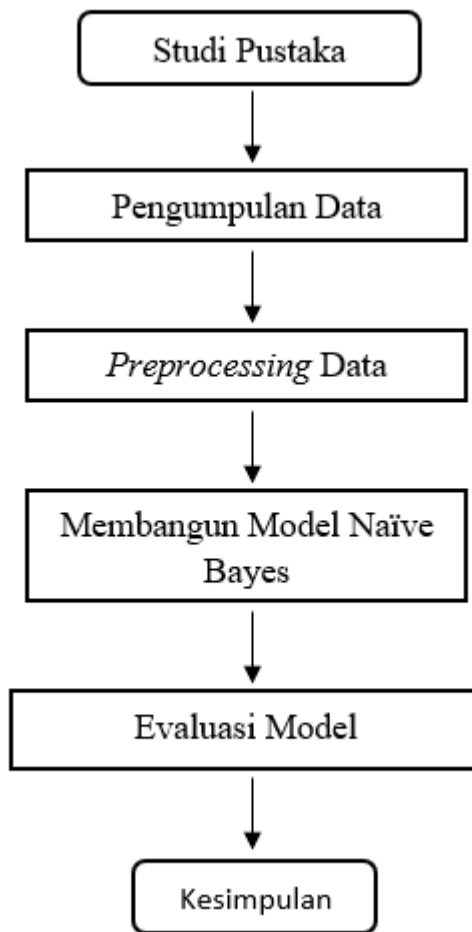
x_{syn} = data hasil replikasi

x_i = data yang akan direplikasi

x_{knn} = data yang memiliki jarak terdekat dari data yang akan di replikasi

β = bilangan *random* antara 0 sampai 1.

- f. Membagi data menjadi data *training* dan data *testing*. Data *train* digunakan untuk melatih model yang digunakan sedangkan data *test* digunakan untuk menguji kinerja model.
4. Membangun Model
Pada tahap ini, dilakukan implementasi model Naïve Bayes dengan menghubungkan *database* yang berisi data-data yang telah melalui tahap *preprocessing* data.
5. Evaluasi Model
Tahap ini dilakukan untuk mengukur performa model Naïve Bayes yang digunakan pada penelitian. *Confusion matrix* digunakan untuk membandingkan hasil prediksi model dengan data aktual yang telah diketahui.
6. Kesimpulan
Pada tahap akhir penelitian, dilakukan penarikan kesimpulan yang berisi hasil interpretasi model dan jawaban dari masalah yang telah dirumuskan.



Gambar 2. 1 Alur Penelitian