

BAB I

PENDAHULUAN

1.1 Latar Belakang

Bank merupakan lembaga keuangan yang salah satu kegiatan utamanya yaitu penyaluran dana kepada masyarakat. Kegiatan ini dapat berupa pemberian kredit kepada pihak yang membutuhkan. Menurut Undang-Undang Nomor 10 Tahun 1998, kredit adalah penyediaan uang atau tagihan yang dapat dipersamakan dengan itu, berdasarkan persetujuan atau kesepakatan pinjam meminjam antara bank dengan pihak lain, yang mewajibkan pihak lain untuk melunasi hutangnya setelah jangka waktu tertentu dengan pemberian bunga.

Pemberian kredit dalam praktiknya masih seringkali terjadi adanya kredit macet. Kredit macet ini merupakan pemberian fasilitas kredit pada nasabah yang mengajukan pinjaman atau kredit mengalami masalah yaitu nasabah tersebut tidak dapat membayar kembali kredit yang telah dipinjamkan kepada pihak bank sesuai dengan perjanjian. Hal ini dapat terjadi karena beberapa hal, seperti penggunaan kredit yang salah, keadaan debitur yang dengan sengaja tidak membayar kredit, ataupun analisis kredit yang tidak memadai dan kurangnya pengawasan dari pihak bank (Winata, Wiryawan, dan Rudi, 2013).

Masalah kredit macet ini dapat mengakibatkan suatu kerugian bagi pihak bank. Kerugian ini dapat berupa kerugian finansial maupun *non* finansial (Rustam, 2013). Adanya tunggakan-tunggakan pembayaran kredit menjadi hambatan dalam kelancaran usaha bank. Masalah ini merupakan suatu hal yang perlu ditanggulangi, dengan salah satu cara, yaitu melakukan analisa kredit kepada calon debitur. Sehingga sebelum mengambil keputusan pemberian kredit, pihak bank dapat mengetahui apakah nasabah yang akan melakukan kredit layak mendapat kredit atau tidak layak mendapat kredit.

Analisis untuk menilai kelayakan kredit nasabah dapat dilakukan dengan menggunakan metode yang sesuai. Analisa yang dapat dilakukan yaitu dengan menggunakan *data mining*. *Data mining* merupakan salah satu konsep teknologi informasi yang berkaitan dengan data dan informasi. *Data mining* dapat

digunakan sebagai metode atau teknik pemodelan untuk mendapatkan pola dan hubungan antar variasi data (Ginting, 2019).

Teknik dalam *data mining* salah satunya yaitu klasifikasi. Klasifikasi merupakan pengelompokan data untuk menemukan sekumpulan model yang bertujuan untuk memperkirakan suatu kelas dari objek yang kelasnya belum diketahui (Permana dan Dewi, 2021). Metode klasifikasi yang sering digunakan yaitu *K-Nearest Neighbor* (KNN). KNN adalah suatu metode untuk melakukan klasifikasi terhadap objek berdasarkan data pembelajaran dilihat dari jarak yang paling dekat dengan objek tersebut (Dewi, Obert dan Gusmana, 2018). KNN merupakan metode klasifikasi yang sederhana, efisien, dan efektif dalam bidang pengolahan objek yang mampu melakukan *training* data dalam jumlah besar (Bhatia, 2010). Metode *K-Nearest Neighbor* (KNN) termasuk ke dalam algoritma terarah (*supervised*). *Supervised learning* sendiri adalah algoritma yang cara kerja sistemnya diberikan *training dataset* berupa informasi masukan dan keluaran yang diinginkan, sehingga sistem akan mempelajari berdasarkan data yang telah ada (Santoso, Abijono, dan Anggreini, 2021).

Metode klasifikasi *data mining* pada kenyataannya sering ditemukan permasalahan dalam prosesnya, yaitu ketika salah satu kelasnya mempunyai jumlah data yang jauh berbeda dengan kelas lainnya pada *dataset training*. Jumlah data mayoritas yang lebih banyak dibandingkan dengan jumlah data minoritas. Hal ini menyebabkan proses pengklasifikasian menjadi tidak akurat. Permasalahan ini disebut juga sebagai *imbalance class*. Model klasifikasi yang dibuat dengan menggunakan data tidak seimbang berakibat menghasilkan akurasi prediksi kelas minoritas yang rendah (Syukron dan Subekti, 2018). Data minoritas akan dianggap sebagai *noise* dan *outlier* serta dapat menurunkan kinerja dari *classifier* (Untoro, 2021).

Permasalahan *imbalance class* dapat diatasi dengan menggunakan teknik *resampling*. Teknik *resampling* adalah teknik *preprocessing* dimana distribusi data diseimbangkan kembali untuk mengurangi efek distribusi kelas tidak seimbang. Terdapat dua teknik *resampling* yang dapat digunakan untuk mengatasi ketidakseimbangan kelas dalam klasifikasi, yaitu *random oversampling* (ROS)

dan *random undersampling* (RUS) (Saifudin dan Wahono, 2015). *Random oversampling* (ROS) adalah teknik pemerataan data minoritas sehingga menjadi sebanyak data mayoritas. Sedangkan *random undersampling* (RUS) adalah teknik menyeimbangkan data dengan mengurangi data kelas mayoritas sehingga berjumlah sama dengan kelas minoritas.

1.2 Rumusan Masalah

Berdasarkan dari latar belakang yang telah diuraikan maka rumusan masalah dalam penelitian ini yaitu:

1. Bagaimana mengimplementasikan metode KNN dalam analisis kelayakan pemberian kredit pada data *imbalance*?
2. Metode KNN manakah yang paling baik antara KNN tanpa penerapan ROS dan KNN dengan penerapan ROS dalam analisis kelayakan pemberian kredit pada data *imbalance*?

1.3 Tujuan Penelitian

Berdasarkan rumusan masalah yang telah dibuat maka tujuan penelitian yang ingin dicapai yaitu:

1. Untuk mengimplementasikan metode KNN dengan teknik ROS dalam analisis kelayakan pemberian kredit pada data *imbalance*.
2. Untuk membandingkan apakah metode KNN dapat lebih baik pada data *imbalance* jika dilakukan ROS dalam analisis kelayakan pemberian kredit.

1.4 Batasan Masalah

Berdasarkan uraian dari latar belakang maka batasan masalah dalam penelitian ini yaitu:

1. Pengukuran performa model pada penelitian ini digunakan *confusion matrix*.
2. Proporsi pembagian data pada penelitian ini 90% data *training* dan 10% data *testing*.

1.5 Tinjauan Pustaka

Yunita (2017) melakukan penelitian membandingkan metode *K-Nearest Neighbor* dan metode *Decision Tree* untuk analisis risiko kredit kepemilikan mobil. Kajian ini dilakukan untuk meningkatkan tingkat ketelitian kredit kepemilikan mobil (KPM) dalam memilih konsumen yang akan melakukan kredit mobil agar tidak terjadi kredit macet. Hasil penelitian ini menunjukkan metode *K-Nearest Neighbor* memiliki tingkat akurasi lebih tinggi dibandingkan dengan metode *Decision Tree*.

Syukron dan Subekti (2018) melakukan penelitian mengenai penerapan metode *random oversampling* dan *random forest* untuk klasifikasi penilaian kredit. Penelitian ini memperoleh hasil yang menunjukkan penerapan *resampling* dengan *random over-under sampling* pada algoritma *random forest* dapat meningkatkan kinerja akurasi secara efektif pada klasifikasi tidak seimbang data penilaian kredit.

Wajhilah, Ubaidallah dan Bahri (2019) melakukan penelitian analisis kredit macet dengan menggunakan metode *K-Nearest Neighbor*. Studi ini membahas akurasi penggunaan metode *K-Nearest Neighbor* dalam mengklasifikasikan data kelayakan pemberian kredit. Hasil yang diperoleh, yaitu nilai akurasi penggunaan metode *K-Nearest Neighbor* menunjukkan hasil yang cukup baik dalam proses klasifikasi data kelayakan pemberian kredit.

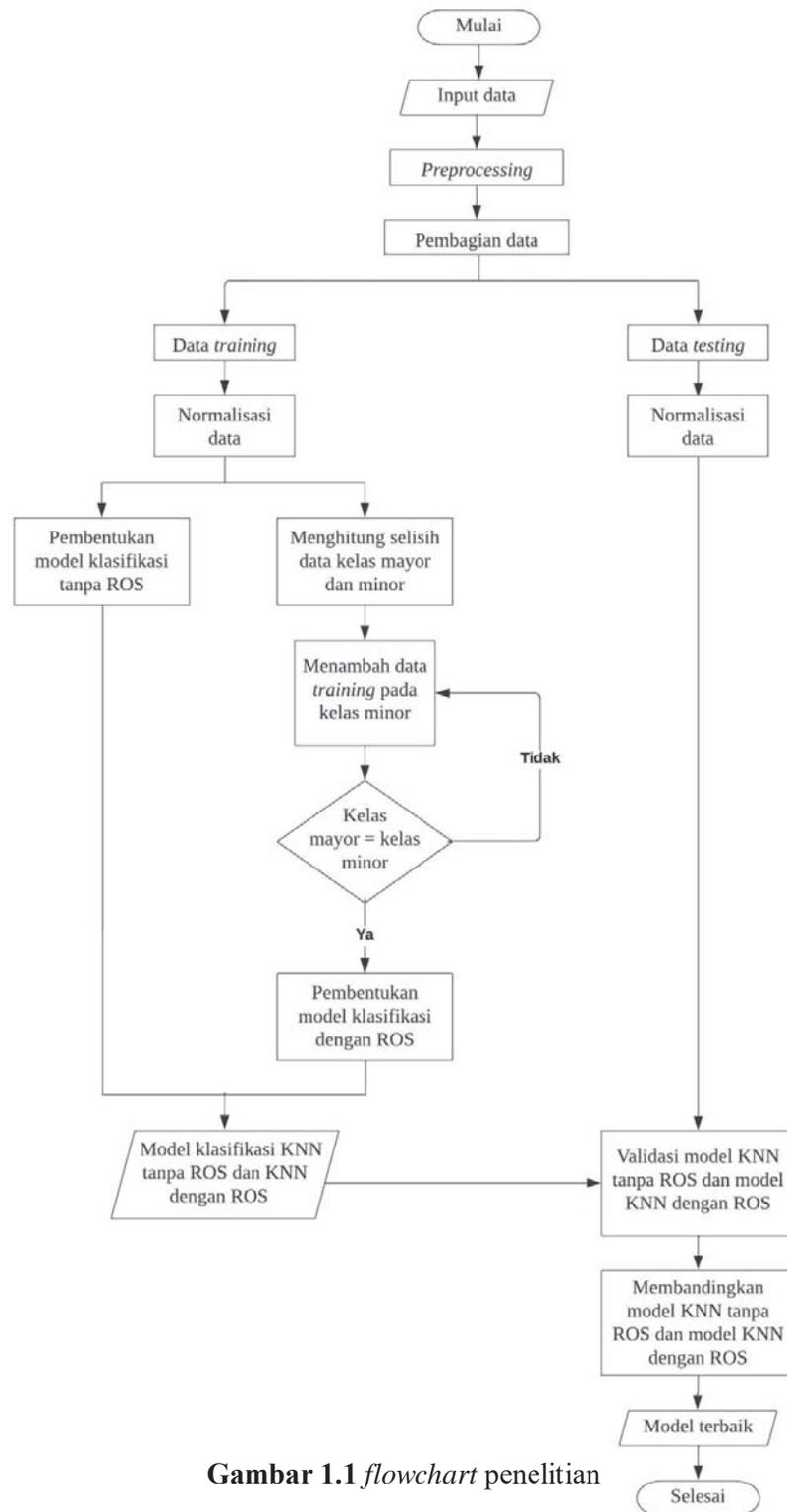
Hamami dan Dahlan (2022) melakukan penelitian mengenai klasifikasi cuaca menggunakan algoritma *random forest* dengan teknik *oversampling*. Penelitian ini membahas tentang penggunaan algoritma *random forest* yang dikombinasikan dengan teknik *oversampling* untuk menangani ketidakmerataan jumlah data dari setiap kelas cuaca. Hasil yang diperoleh menunjukkan bahwa model *random forest* dapat mencapai akurasi sebesar 70% dengan data minoritas ditingkatkan dengan rata-rata sebesar 50%.

1.6 Metodologi Penelitian

Pada penelitian ini data yang digunakan berupa data sekunder yang diperoleh dari bank X. Berupa data debitur bank X dalam kelompok debitur lancar dan tidak

lancar. Metode yang digunakan diawali dengan studi literatur yaitu untuk mencari referensi yang relevan terhadap studi kasus yang diteliti. Setelah itu proses analisis dilakukan menggunakan metode *K-Nearest Neighbor* (KNN) dengan teknik *resampling*, *random oversampling* (ROS). Pada penelitian ini menggunakan 11 atribut yang terdiri dari jumlah pinjaman, suku bunga, jangka waktu penyelesaian pinjaman, kolektibilitas kredit, total angsuran, tujuan kredit, usia, gaji, premi dan admin, instansi, dan jenis kredit. Proses analisis dilakukan dengan bantuan *software* R studio. Adapun proses analisis yang dilakukan adalah sebagai berikut:

- a) Melakukan *preprocessing* data yang akan digunakan. *Preprocessing* data terdiri dari *cleaning* data, *selection* data, dan *transformasi* data.
- b) Membagi data menjadi dua yaitu data *training* dan data *testing*. Data *training* digunakan untuk pembentukan model dan data *testing* digunakan untuk menguji akurasi ketepatan model.
- c) Melakukan normalisasi data *training* dan data *testing* sebelum digunakan untuk pembentukan model.
- d) Pembentukan model KNN tanpa ROS menggunakan data *training*.
- e) Melakukan proses ROS dengan menambahkan data kelas minoritas secara acak ke dalam *dataset training*.
- f) Pembentukan model KNN dengan ROS menggunakan data *training* yang telah di *resampling* dengan ROS menjadi *balance*.
- g) Menentukan klasifikasi calon debitur model KNN tanpa ROS dengan menggunakan data *testing*.
- h) Menentukan klasifikasi calon debitur model KNN dengan ROS dengan menggunakan data *testing*.
- i) Membandingkan tingkat akurasi ketepatan model yang dihasilkan dari metode KNN tanpa ROS dan KNN dengan ROS.
- j) Menentukan metode terbaik untuk mengklasifikasikan calon debitur.
- k) Menarik kesimpulan.



Gambar 1.1 flowchart penelitian