

dengan menggunakan jarak euclidean. Jarak euclidean adalah jarak yang dihitung dengan menggunakan rumus matematika yang mengukur seberapa jauh jarak antara dua titik pada sebuah ruang vektor [11]. Rumus perhitungan jarak euclidean dapat dilihat dibawah ini [12]:

$$di = \sqrt{\sum_{i=1}^n (p_i - q_i)^2}$$

$$di = \sqrt{(p_1 - q_1)^2 + (p_2 - q_2)^2 + \dots + (p_n - q_n)^2}$$

Dengan keterangan:

pi: sampel data

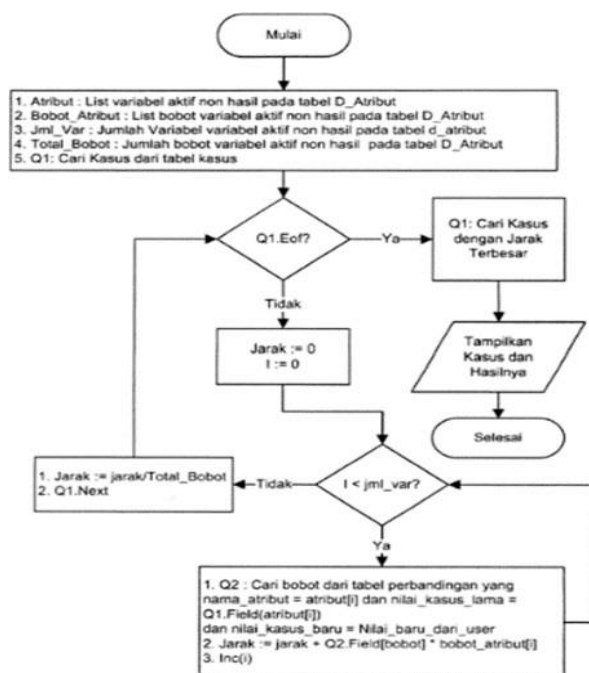
qi: data uji

i: variabel data

dist: jarak

n: dimensi data

Untuk menentukan tingkat kedekatan antara kasus baru dengan kasus yang ada dalam dataset, dilakukan perhitungan dengan algoritma yang ditunjukkan pada gambar 2.1.



Gambar 2.1. Flowchart menghitung kedekatan kasus

4. Pengertian Kredit

Kredit merupakan sistem pinjam meminjam yang sudah ada sejak zaman Yunani kuno. Yang mana berasal dari istilah "credere" yang berarti kepercayaan digunakan untuk menggambarkan kemampuan seseorang atau perusahaan untuk memenuhi kewajiban pembayaran pinjaman yang telah mereka terima dalam jangka waktu tertentu. Menurut Undang-Undang Nomor 10 Tahun 1998 tentang perubahan atas Undang-Undang Nomor 7 Tahun 1992 tentang Perbankan, kredit didefinisikan

sebagai penyediaan uang atau tagihan yang setara dengan uang, berdasarkan persetujuan atau kesepakatan pinjam meminjam antara bank dengan pihak lain yang mengharuskan peminjam untuk melunasi utangnya setelah jangka waktu tertentu dengan pemberian bunga [13].

5. Syarat Kredit

Syarat suatu kredit meliputi itikad baik nasabah (willingness to pay) dan kemampuan membayar untuk menyelesaikan pinjaman dan bunga (ability to pay). Selain itu, faktor-faktor seperti kepribadian (Character), kapasitas (Capacity), modal (Capital), jaminan (Collateral), dan keadaan perekonomian (Condition of Economy) juga dipertimbangkan dalam menilai kelayakan kredit. Kumpulan faktor-faktor tersebut disebut sebagai Analisis 5C [13].

6. Analisis Kredit (Credit Analysis)

Proses evaluasi kredit dari berbagai perspektif terkait untuk menentukan apakah suatu bisnis atau individu dapat dibiayai dengan kredit dikenal sebagai kegiatan analisis kredit. Beberapa aspek seperti ekonomi, manajemen, produksi, pemasaran, keuangan, lingkungan, hukum dan ketersediaan jaminan yang memadai, semua aspek tersebut termasuk dalam analisis kredit. Analisis kredit atau penilaian terhadap file/data, di samping berbagai rincian pendukung yang diberikan oleh pemohon kredit, dikenal sebagai analisis kredit. Kegiatan ini berfungsi sebagai dasar untuk memutuskan apakah layak untuk menerima penyaluran kredit atau tidak [13].

7. Analis Kredit (Credit Analyst)

Laporan keuangan individu, bisnis, dan data pinjaman dapat dianalisis oleh seorang analis kredit untuk menentukan tingkat risiko yang terkait dengan pemberian kredit atau peminjaman uang. Selain itu, seorang analis kredit mampu membuat laporan yang berisi data kredit untuk tujuan pengambilan keputusan [13].

III. METODE PENELITIAN

A. Objek Penelitian

Data uji berupa file excel yang berisi riwayat kredit pelanggan dari lembaga keuangan. Agendanya adalah untuk memprediksi kemungkinan kredit macet di muka dan membantu lembaga keuangan untuk mengambil langkah-langkah yang sesuai. Tujuan dari penelitian ini adalah untuk melakukan penerapan algoritma k-nearest neighbor dan metode k-means clustering untuk memutuskan apakah seorang pemohon kredit akan diberikan kredit atau tidak.

B. Metode Penelitian

1. Metode k-means clustering

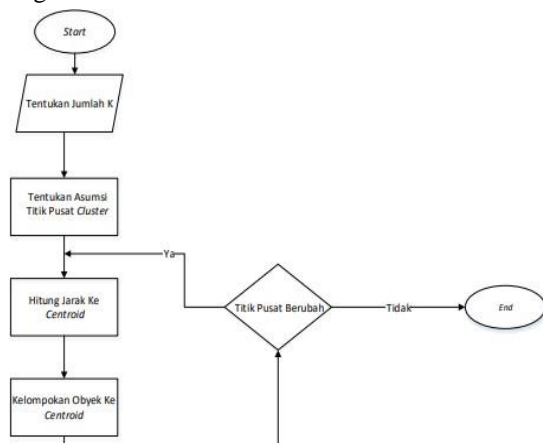
Metode k-means clustering merupakan salah satu metode klasterisasi yang sering digunakan di berbagai bidang karena sifatnya yang mudah diimplementasikan

dan efektif untuk mengklaster data yang besar [14]. Algoritma k-means berbasis pada jarak dan bertujuan untuk membagi data menjadi beberapa cluster. Dalam setiap tahap, setiap objek harus ditempatkan dalam salah satu kelompok. Pada tahap selanjutnya, objek tersebut bisa saja dipindahkan ke kelompok lain. Algoritma k-means memerlukan input jumlah cluster yang diinginkan dan hanya dapat mengolah atribut yang bertipe numerik [15].

Metode k-means clustering terdiri dari beberapa komponen, yaitu:

- Jumlah cluster "k", yang harus ditentukan terlebih dahulu. Biasanya jumlah k ditentukan melalui pendekatan metode hirarki, tetapi tidak ada aturan khusus yang harus diikuti dan jumlah k bisa disesuaikan dengan kebutuhan subjektif peneliti.
- Cluster awal, yang bisa diambil dengan berbagai cara seperti menentukan interval dari jumlah observasi, menggunakan pendekatan metode hirarki, atau dengan memilih cluster secara acak dari sekumpulan observasi.
- Ukuran jarak, yang digunakan untuk mengukur jarak antara observasi dan centroid terdekat dalam menempatkan observasi ke dalam cluster. Euclidean Distance biasanya digunakan dalam metode k-means.

Langkah-langkah yang dilakukan dalam proses clustering menggunakan metode K-Means adalah sebagai berikut:



Gambar 3.1. Tahapan metode k-means clustering

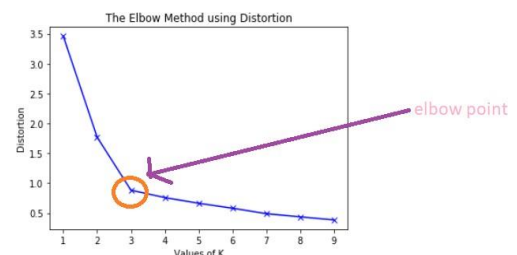
- Proses pemilihan objek 'k' yang acak diikuti dengan pemrosesan data sebagai mean dari cluster.
- Kemudian, setiap objek akan ditempatkan dalam cluster yang memiliki kemiripan dengan cluster lainnya, dengan menghitung jarak objek terhadap mean atau centroid dari cluster tersebut.
- Selanjutnya, akan dilakukan perhitungan centroid baru untuk setiap cluster,
- Sehingga, proses perhitungan tersebut akan diulang sampai anggota dari kelompok cluster tidak berubah lagi.

2. Metode elbow

Metode elbow merupakan teknik yang digunakan untuk memilih nilai cluster terbaik dengan cara mengamati presentasi hasil perbandingan antara jumlah cluster yang membentuk siku pada suatu titik [16]. Presentasi perhitungan yang dihasilkan merupakan perbandingan antara jumlah cluster pertama dan kedua. Untuk mengetahui perbandingannya, dilakukan perhitungan SSE (sum of square error) dari setiap nilai cluster. Semakin banyak jumlah cluster "k", maka nilai SSE akan semakin kecil [17]. SSE dihitung dengan menggunakan rumus sebagai berikut:

$$SSE = \sum_{k=1}^K \sum_{x \in S_k} \|X_i - C_k\|_2^2$$

Setelah dilihat, beberapa nilai K akan mengalami penurunan yang signifikan dan selanjutnya hasil dari nilai "k" akan terus menurun secara perlahan sampai mencapai stabilitas [18]. Berikut adalah contoh gambar dari metode elbow:



Gambar 3.2. Contoh titik elbow

C. Metode Pengumpulan Data

1. Observasi

Observasi adalah metode pengumpulan data dengan cara melakukan pengamatan secara langsung terhadap objek yang diteliti, yaitu dataset riwayat kredit pelanggan yang meliputi data kualitatif dan data kuantitatif. Data kualitatif meliputi aspek-aspek seperti 5C yang terdiri dari (character, capacity, capital, collateral, dan condition). Sementara itu, data kuantitatif terdiri dari informasi mengenai pendapatan, jangka waktu, dan kondisi pemohon kredit (Credit Term and Condition) [19].

2. Studi Pustaka

Studi pustaka atau studi literatur merupakan proses pencarian informasi yang terkait dengan masalah yang sedang diteliti melalui sumber-sumber tertulis seperti buku-buku ilmiah, laporan penelitian, tesis, disertasi, peraturan, undang-undang, dan sumber elektronik lainnya. Hasil dari studi pustaka ini adalah informasi terbaru mengenai algoritma k-nearest neighbor, metode k-means clustering, metode elbow, dan teori-teori pendukung lainnya [20].

3. Datasets statistik

Datasets statistik merupakan jenis data yang biasanya digunakan dalam penelitian kuantitatif, dimana data

tersebut sudah tersedia dan biasanya telah dikumpulkan oleh pihak ketiga yang memiliki otoritas. Penggunaan dataset ini biasanya lebih cepat karena peneliti hanya perlu mengakses dataset yang sudah tersedia, tidak perlu melakukan penyebaran kuesioner ke lapangan.

IV. HASIL DAN PEMBAHASAN

A. Pengumpulan Data

Pada tahap ini, peneliti telah memperoleh 887,379 data yang berisi riwayat kredit seorang pemohon kredit dari lembaga pembiayaan. Agendanya adalah untuk memprediksi kemungkinan kredit macet di muka dan membantu lembaga pembiayaan untuk mengambil langkah-langkah yang sesuai. Beberapa atribut yang akan dianalisis dalam penelitian ini dapat dilihat pada tabel 4.1.

Tabel 4.1

TABEL ATRIBUT STATISTIK PINJAMAN

| No. | Atribut | Deskripsi |
|-----|----------------|--|
| 1. | loan_amnt | Jumlah pinjaman yang terdaftar dan yang diajukan oleh peminjam. Jika pada suatu saat, departemen kredit mengurangi jumlah pinjaman, maka itu akan tercermin dalam nilai ini. |
| 2. | term | Jumlah pembayaran pinjaman. Nilai dalam bulan dapat berupa 36 atau 60. |
| 3. | int_rate | Suku bunga pinjaman |
| 4. | installment | Pembayaran bulanan terutang oleh peminjam jika pinjaman berasal. |
| 5. | grade | Surat kredit yang diberikan peringkat pinjaman. |
| 6. | emp_length | Lama bekerja dalam satu tahun. Nilai yang mungkin berkisar antara 0 dan 10, di mana 0 berarti kurang dari satu tahun dan 10 berarti sepuluh tahun atau lebih. |
| 7. | home_ownership | Status kepemilikan rumah yang dilaporkan oleh peminjam saat mendaftar. Nilai-nilai untuk variabel ini adalah: Sewa (Rent), Milik Sendiri (Own), Hipotek (Mortgage), Lainnya. |
| 8. | annual_inc | Pendapatan tahunan yang dilaporkan oleh peminjam sendiri saat mendaftar. |

| | | |
|-----|---------------------|--|
| 9. | verification_status | Menunjukkan apakah pendapatan bersama peminjam telah diverifikasi melalui surat kredit atau tidak, atau apakah sumber pendapatan telah diverifikasi. |
| 10. | loan_status | Status pinjaman saat ini. |
| 11. | purpose | Kategori yang diberikan oleh peminjam untuk tujuan permohonan pinjaman. |

B. Pra-Pemrosesan Data

Pada tahap ini, data telah dipersiapkan untuk mempermudah proses pembelajaran mesin. Pra-pemrosesan data yang dilakukan meliputi:

1. Cleaning

Pada proses ini dilakukan pembersihan data dengan mengecek kelengkapan data, menghapus data yang tidak perlu, dan menghilangkan noise. Ada banyak nilai yang kosong di kolom "emp_length" dan sedikit di "annual_inc", untuk menghilangkan nilai null, maka kolom "emp_length" dan "annual_inc" diisi dengan nilai 0.

2. Transformation

Pada proses ini data akan dipilih dan ditransformasikan sehingga siap untuk diklasifikasikan ke dalam cluster. Dari semua kategori data yang ada, diberikan inisialisasi untuk mempermudah pengelompokannya ke dalam cluster. Kemudian, inisialisasi dari data bertipe categorical diubah menjadi data numerik. Berikut ini adalah hasil transformasi dari beberapa data categorical, yang dapat dilihat pada tabel 4.2:

Tabel 4.2

TABEL TRANSFORMASI (PRA-PEMROSESAN)

| 1. label | |
|---|----------|
| Keterangan | Kategori |
| "Lunas", "Terkini", "Tidak memenuhi kebijakan kredit. Status: Fully paid" (kondisi yang tidak gagal bayar). | 0 |
| "Terlambat (31-120 hari)", "Terlambat (16-30 hari)", "Dalam Masa Grace Period", "Ditolak", "Tidak memenuhi kebijakan kredit. Status: Ditolak" (kondisi yang gagal bayar). | 1 |
| 2. grade | |
| Keterangan | Kategori |