

pengambilan kredit menggunakan metode k-means clustering [6].

Dalam penelitian ini, penulis akan menggunakan algoritma k-nearest neighbor dan metode k-means clustering sebagai metode untuk menentukan kelayakan kredit. Kami akan menggunakan atribut yang telah ditentukan dan hasil kedekatan (similarity) dari data kasus sebelumnya untuk memutuskan pemberian kredit. Algoritma k-nearest neighbor sendiri merupakan algoritma yang digunakan untuk menemukan solusi dari kasus saat ini dengan mencocokkan bobot dari sejumlah fitur yang ada dengan menghitung kedekatan antara kasus baru dengan kasus sebelumnya [2].

A. Rumusan Masalah

Rumusan masalah diharapkan akan menentukan instrumen yang digunakan, fokus pembahasan, kesimpulan, dan arah pelaksanaan penelitian. Berikut ini rumusan masalah yang bisa diambil:

1. Bagaimana cara melakukan analisis risiko dalam penyaluran kredit kepada pemohon kredit?
2. Bagaimana cara menentukan apakah seorang pemohon kredit layak mendapatkan pemberian kredit berdasarkan datasets yang ada ?
3. Bagaimana cara membuat sebuah program pembelajaran mesin untuk analisis risiko pemberian kredit?
4. Bagaimana cara menerapkan suatu algoritma dan metode tertentu pada sebuah program machine learning untuk analisis risiko pemberian kredit?

B. Batasan Masalah

Batasan masalah diperlukan untuk menetapkan tujuan yang ingin dicapai dalam penelitian ini. Berikut adalah batasan masalah yang dapat diberikan:

1. Sistem untuk analisis risiko kredit pada penelitian ini hanya berfokus pada penerapan algoritma k-nearest neighbor dan metode k-means clustering.
2. Pembuatan sistem ini tidak mendukung fitur input dan output dari sisi pengguna, melainkan program machine learning yang menggunakan bahasa python.
3. Datasets yang digunakan dalam penelitian ini berasal dari situs web yang menyediakan sumber data (Kaggle), bukan dari lembaga pemberian kredit tradisional.

C. Tujuan Penelitian

Tujuan penelitian adalah untuk memahami atau memprediksi situasi atau memecahkan masalah dalam kondisi yang ada dan membuktikannya dengan tindakan yang diambil. Berikut adalah tujuan penelitian yang dapat diambil:

1. Agar seorang analis kredit mampu mengambil keputusan dalam memberikan kredit kepada calon pemohon kredit dengan cepat, tepat, dan akurat.
2. Agar seorang analis kredit mengetahui, apakah seorang pemohon kredit layak atau tidak untuk diberikan pinjaman kredit berdasarkan datasets yang ada.
3. Agar terciptanya sebuah program pembelajaran mesin untuk analisis pemberian kredit yang efektif dan efisien.
4. Agar lembaga pembiayaan kredit dapat mengurangi kesalahan manusia dengan menerapkan algoritma dan metode yang ada menggunakan pembelajaran mesin.

*) **penulis korespondensi:** Rizky Ramadhan
Email: 20081010043@student.upnjatim.ac.id

II. PENELITIAN YANG TERKAIT

A. Tinjauan Pustaka

Pada penelitian ini, penulis akan mencoba untuk menghubungkan sejumlah karya ilmiah yang telah dikembangkan oleh peneliti lain yang nantinya akan berkaitan dengan karya ilmiah yang sedang penulis kembangkan. Beberapa karya ilmiah yang dimaksud penulis sebagaitinjauan pustaka antara lain sebagai berikut:

1. Jurnal Azhari dan Wahyu, yang berjudul: Pengembangan Perangkat Lunak Agen Cerdas untuk Analisis Kelayakan Keuangan Perusahaan terhadap Pemberian Kredit Perbankan. Pada jurnal ini membahas tentang pembuatan aplikasi smart berbasis agen untuk menganalisis informasi keuangan dan menilai kelayakan manajemen. Studi kelayakan menggunakan kriteria dan bobot perusahaan, dan hasilnya menunjukkan bahwa agen mampu bekerja secara independen dan berinteraksi satu sama lain melalui pengiriman pesan, tanggapan, dan meminta agen lain untuk melakukan sesuatu yang sejalan dengan tujuan utama agen menggunakan sistem cerdas untuk menganalisis kelayakan keuangan dan mengelola perusahaan. Nilai diukur menggunakan kriteria The Five C of Credit, yang terdiri dari (Character, Capacity, Capital, Collateral, dan Condition of Economy), serta kriteria analisis keuangan perusahaan seperti (rasio likuiditas, rasio operasional, rasio struktur modal, rasio kinerja bisnis, dan rasio pertumbuhan), serta analisis manajemen perusahaan [3].
2. Jurnal Kurniawan, dkk, yang berjudul: Pemilihan Metode Predictive Analytics dengan Machine Learning untuk Analisis dan Strategi Peningkatan Kualitas Perbankan. Pada jurnal ini membandingkan berbagai metode klasifikasi machine learning untuk analisis prediktif pada kualitas kredit perbankan dengan menggunakan data yang sangat besar (big data). analitik). Hasil komputasi dari metode-metode tersebut dibandingkan untuk menghasilkan rekomendasi teknik yang paling tepat, yaitu metode Tree Ensemble dengan SMOTE. Perhitungan tersebut dapat menghasilkan rekomendasi bagi bank, dengan contoh yang diberikan dalam jurnal merekomendasikan agar bank fokus pada peningkatan kualitas kredit atau pembiayaan untuk nasabah perempuan atau nasabah wiraswasta, karena berpotensi meningkatkan NPL/NPF [4].
3. Jurnal Gultom, dkk, yang berjudul: Prediksi Tingkat Kelancaran Pembayaran Kredit Bank Dengan Menggunakan Algoritma Naive Bayes dan K-Nearest Neighbor. Pada jurnal ini dilakukan dua proses pengujian. Pengujian pertama menggunakan algoritma Naive Bayes dan K-Nearest Neighbor yang diuji secara terpisah sehingga menghasilkan akurasi yang sangat rendah. Pada pengujian kedua, algoritma Naive Bayes dan K-Nearest Neighbor digabungkan dan diuji

sebanyak 3 kali, menghasilkan akurasi tertinggi sebesar 80% [5].

4. Jurnal Ahsina, dkk, yang berjudul: Analisis Segmentasi Pelanggan Bank Berdasarkan Pengambilan Kredit Dengan Menggunakan Metode K-Means Clustering. Pada jurnal ini menjelaskan bahwa penentuan jumlah cluster terbaik dapat dilakukan dengan menggunakan metode elbow. Setelah melakukan segmentasi data pelanggan menggunakan metode k-means clustering, metode elbow menunjukkan 4 cluster terbaik dari semua kemungkinan cluster, terdiri dari cluster 1 dengan 286 pelanggan (28,6%), cluster 2 dengan 130 pelanggan (13%), cluster 3 dengan jumlah pelanggan terbesar sebanyak 542 (54,2%), dan cluster 4 dengan sebaran terendah sebanyak 42 pelanggan (4,2%). Hasil clustering dapat dipengaruhi oleh nilai centroid awal dan banyaknya data yang digunakan, perbedaan data pada centroid awal yang digunakan dapat mempengaruhi hasil centroid akhir [6].

B. Landasan Teori

Secara umum, landasan teori mengacu pada augmentasi yang sistematis dan mapan yang memiliki variabel yang kuat dan telah terbukti. Landasan teori ini juga berfungsi untuk menghubungkan pengetahuan baru dan juga memfasilitasi penelitian dalam merumuskan hipotesis dan metodologi penelitian. Beberapa landasan teori yang disebutkan oleh penulis antara lain:

1. Machine Learning

Di zaman sekarang ini, ketika algoritma cerdas dan kecerdasan buatan banyak diterapkan dalam kehidupan kita sehari-hari, mampu menjadikan pembelajaran mesin sebagai bidang studi yang signifikan dan populer. Pembelajaran mesin menggunakan berbagai algoritma matematika dan statistik yang nantinya digunakan untuk melatih data pada jumlah yang besar. Ada dua kategori besar algoritma pembelajaran mesin, yaitu algoritma pembelajaran yang diawasi dan tidak diawasi. Pada pembelajaran terawasi, telah disediakan data yang diberi label. Sedangkan, pada pembelajaran yang tidak diawasi, pelabelan data tidak diberikan. Baru-baru ini, kecerdasan buatan telah menunjukkan visibilitas dan paparan yang sangat besar dalam aplikasi yang ada di dunia nyata secara signifikan, misalnya, penanganan gambar klinis, rekomendasi produk online, model dan prediksi keuangan, analisis risiko, pemrosesan bahasa manusia, dll. Termasuk penerapan pembelajaran mesin pada penelitian ini untuk analisis risiko kredit dengan algoritma pembelajaran yang diawasi dalam bentuk data berlabel [7].

2. Clustering

Clustering adalah proses mempartisi sekumpulan objek data menjadi sekumpulan bagian yang disebut cluster. Objek dalam suatu cluster memiliki karakteristik yang mirip satu sama lain dan berbeda dengan cluster lainnya. Pemisahan tidak dilakukan secara manual melainkan dengan algoritma clustering. Oleh karena itu, clustering

sangat berguna dan dapat menemukan grup atau cluster yang tidak diketahui dalam data. Clustering banyak digunakan dalam berbagai aplikasi seperti business intelligence, image pattern recognition, pencarian web, biologi, keamanan, dan analisis risiko kredit pada lembaga pembiayaan [8].

3. Algoritma K-Nearest Neighbor (KNN)

Algoritma k-nearest neighbor (KNN) menggunakan penyesuaian bobot pada beberapa atribut untuk menemukan kedekatan antara objek yang akan diklasifikasikan dengan objek pada data latih. Penyesuaian bobot ini bertujuan untuk memberikan bobot yang lebih besar pada atribut yang lebih penting atau relevan dalam proses klasifikasi. Algoritma KNN menghitung jarak antara objek yang akan diklasifikasikan dengan objek pada data latih dengan menggunakan rumus yang disesuaikan dengan jenis atribut yang digunakan. Misalnya, menggunakan solusi yang diinginkan dari pemohon kredit sebelumnya untuk diterapkan kepada pemohon kredit baru. Kedekatan kasus pemohon kredit saat ini dengan sejumlah kasus pemohon kredit sebelumnya yang nantinya dihitung untuk memilih kasus pemohon kredit mana yang akan diterapkan. Keputusan akan diambil untuk menerapkan solusi kepada pemohon kredit baru dalam kasus pemohon kredit sebelumnya dengan nilai kemiripan terbesar [9]. Beberapa rumus yang dapat digunakan untuk menghitung kemiripan antara objek yang akan diklasifikasikan dengan objek pada data latih antara lain sebagai berikut [10]:

$$\text{Similarity}(T, S) = \frac{\sum_{i=1}^n f(T_i, S_i) * W_i}{W_i}$$

$$\text{Similarity}(T, S) = \frac{s_1 * w_1 + s_2 * w_2 + \dots + s_n * w_n}{w_1 + w_2 + \dots + w_n}$$

Dengan keterangan:

T: klasifikasi objek baru

S: data latih yang tersimpan dalam memori

n: jumlah atribut pada data latih

i: indeks yang bernilai antara 1 sampai n

f: fungsi similarity untuk menghitung tingkat kemiripan antara atribut di kasus T dan kasus S

s: nilai kemiripan (similarity)

w: bobot yang dapat diberikan sesuai dengan tingkat relevansi atribut

Algoritma KNN (K-Nearest Neighbor) bertujuan untuk mengklasifikasikan objek baru dengan menggunakan atribut dari data latih (training samples) yang telah ada. Objek yang akan diklasifikasikan akan dikelompokkan ke dalam kategori yang sama dengan mayoritas dari objek terdekat yang terdapat pada data latih. Algoritma KNN menggunakan klasifikasi ketetanggaan sebagai dasar untuk menentukan kelompok objek yang akan diklasifikasikan. Salah satu metode yang digunakan untuk menghitung ketetanggaan antar objek adalah

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

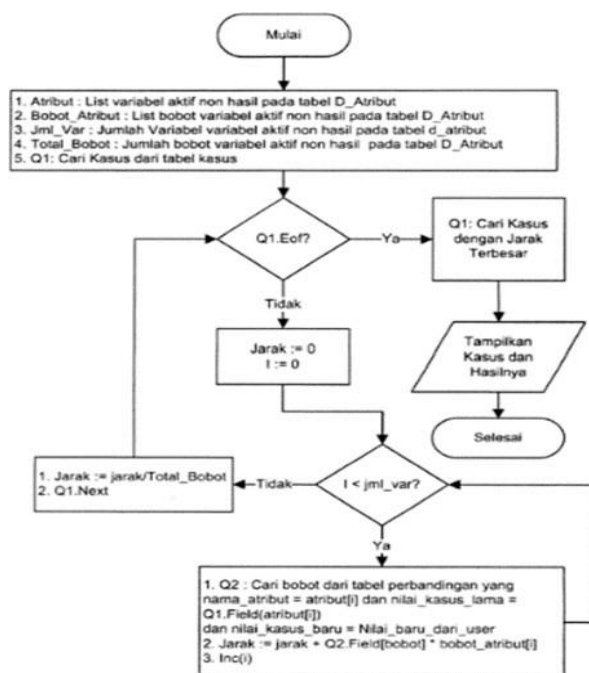
qi: data uji

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