

Machine Learning Analisis Risiko Kredit pada Lembaga Pembiayaan Menggunakan Algoritma K-Nearest Neighbors (KNN) dan Metode K-Means Clustering

Rizky Ramadhan^{1*)}

¹Jurusan Informatika, Fakultas Ilmu Komputer, Universitas Pembangunan Nasional “Veteran” Jawa Timur, Surabaya

¹Jl. Rungkut Madya No. 1, Kec. Gn. Anyar, Kota Surabaya, 60294, Indonesia

email: ¹20081010043@student.upnjatim.ac.id

Abstract — Bad credit often becomes a problem when a credit applicant fails to make credit installment payments on time. The liquidity of credit financing institutions can be directly affected if this condition is allowed to continue. Before extending credit, it would be nice for financing institutions to carry out a creditworthiness analysis to reduce the risk of bad credit. The process of evaluating credit worthiness sometimes takes quite a lot of time and is difficult to do. Therefore, to determine which creditors are at risk of default and which are not, a system is needed to analyze creditworthiness quickly and accurately. The development of a system based on machine learning using the k-means clustering method and the k-nearest neighbor algorithm is the goal of this research. In data mining, the k-means clustering method can provide cluster descriptions in a data. Meanwhile, the k-nearest neighbor algorithm uses weight matching on a number of existing attributes to identify the current case by comparing how close the new case is to the old case. The approach used in this study is an experimental approach, which includes literature study, modeling and system development using the k-means clustering method and the k-nearest neighbor algorithm, and system implementation. The results of this study indicate that the application of the k nearest neighbor algorithm for classification, with a k-means coefficient of 30, results in an accuracy of 91.42%.

Abstrak – Kredit macet seringkali menjadi suatu masalah ketika seorang pemohon kredit gagal melakukan pembayaran cicilan kredit tepat waktu. Likuiditas lembaga pembiayaan kredit dapat terkena dampak langsung jika kondisi ini dibiarkan terus berlangsung. Sebelum menyalurkan kredit, alangkah baiknya lembaga pembiayaan perlu melakukan analisis kelayakan kredit untuk mengurangi risiko kredit macet. Proses mengevaluasi kelayakan kredit terkadang memakan waktu yang cukup banyak dan sulit dilakukan. Oleh karena itu, untuk menentukan kreditur mana yang berisiko gagal bayar dan mana yang tidak maka diperlukan sistem untuk menganalisis kelayakan kredit yang cepat dan akurat. Pengembangan sebuah sistem yang berbasis pembelajaran mesin menggunakan metode k-means clustering dan algoritma k-nearest neighbor menjadi tujuan dari penelitian ini. Dalam data mining, metode k-means clustering dapat memberikan deskripsi cluster pada sebuah data. Sedangkan, algoritma k-nearest neighbor menggunakan pencocokan bobot pada sejumlah atribut yang ada untuk mengidentifikasi kasus saat ini dengan cara membandingkan seberapa dekat kasus baru dengan kasus lama. Adapun pendekatan digunakan dalam penelitian ini yaitu pendekatan eksperimental, yang meliputi studi pustaka, pemodelan dan pengembangan sistem yang menggunakan metode k-means

clustering dan algoritma k-nearest neighbor, dan penerapan sistem. Hasil penelitian ini menunjukkan bahwa penerapan algoritma k-nearest neighbor untuk klasifikasi, dengan nilai koefisien k-means sebanyak 30, maka diperoleh hasil akurasi sebesar 91,42%.

Kata Kunci – Pembelajaran Mesin, Sistem Pakar, Analisis Risiko Kredit, K-Nearest Neighbor, Euclidean Distance, K-Means, Clustering

I. PENDAHULUAN

Undang-Undang Perbankan No. 10 tahun 1998 mengatur tentang bank sebagai suatu badan usaha yang mengumpulkan dana dari masyarakat melalui simpanan dan menyalurkannya kembali ke masyarakat melalui pinjaman atau bentuk lainnya untuk meningkatkan kualitas hidup masyarakat. Oleh karena itu, setiap pinjaman harus didasarkan pada perjanjian yang disetujui bersama dan keputusan diambil setelah melakukan analisis. Selain pengecekan oleh BI yang merupakan tahap awal seleksi bagi setiap nasabah yang mengajukan kredit, ketidaktepatan penilaian awal sebelum menjadi nasabah kerap kali menjadi masalah kredit macet. Sebelum memberikan kredit, lembaga pembiayaan harus melakukan analisis kelayakan karena akan mempengaruhi kelancaran dalam pengembalian pinjaman, hal inilah yang nantinya dapat mengurangi risiko kredit macet. Pengambilan keputusan yang baik dalam hal memprediksi apakah pemohon kredit dapat memperoleh kredit atau tidak, menjadi perhatian penting pada penelitian ini [1].

Evaluasi kelayakan kredit sering kali merupakan proses yang rumit dan memakan waktu yang cukup lama. Sehingga, diperlukan sistem yang efisien dan tepat dalam menganalisis kelayakan kredit dan menentukan debitur yang berisiko gagal bayar dan yang tidak. Sebelumnya, banyak peneliti telah meneliti berbagai metode analisis kredit. Azhari dan Wahyu (2006), telah mengembangkan perangkat lunak agen cerdas untuk menganalisis keuangan perusahaan dalam rangka pemberian kredit perbankan [3]. Kurniawan, dkk (2022), mengembangkan aplikasi machine learning untuk strategi peningkatan kualitas perbankan dengan metode predictive analytics [4]. Gultom, dkk (2020), memprediksi tingkat kelancaran pembayaran kredit bank menggunakan algoritma naive bayes dan k-nearest neighbor [5]. Ahsina, dkk (2022), menganalisis segmentasi pelanggan bank berdasarkan

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 sebagai tinjauan 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

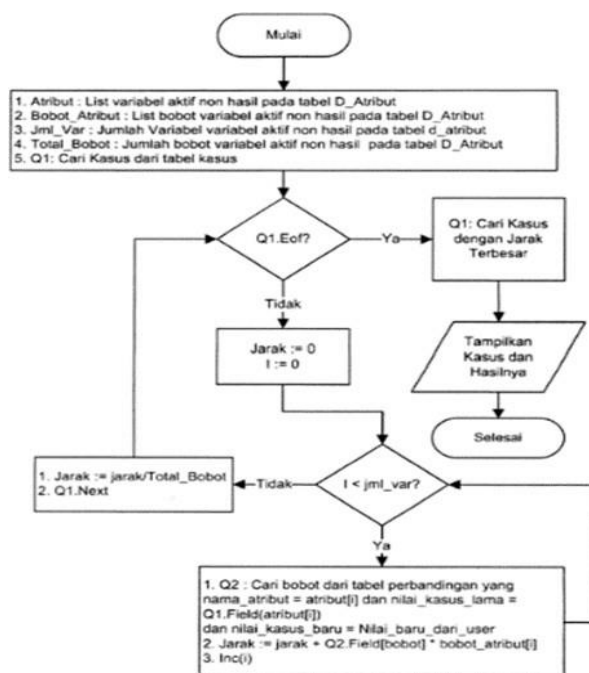
qi: data uji

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

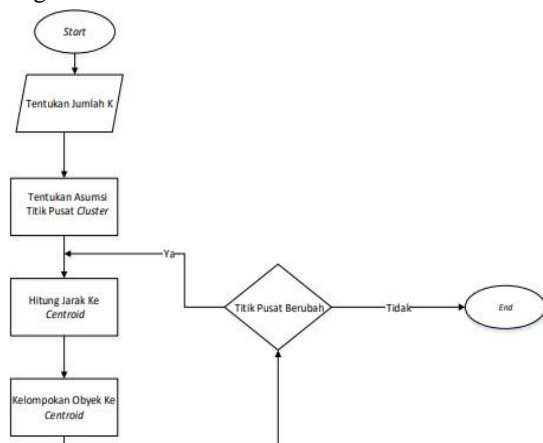
Metode k-means clustering merupakan salah satu metode klasterisasi yang sering digunakan di berbagai bidang karena sifatnya yang mudah diimplementasikan

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

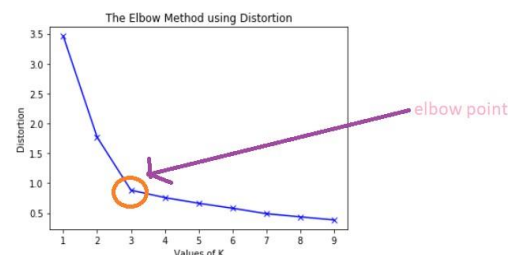
- Proses pemilihan objek 'k' yang acak diikuti dengan pemrosesan data sebagai mean dari cluster.
- Kemudian, setiap objek akan ditempatkan dalam cluster yang memiliki kemiripan dengan cluster lainnya, dengan menghitung jarak objek terhadap mean atau centroid dari cluster tersebut.
- Selanjutnya, akan 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)

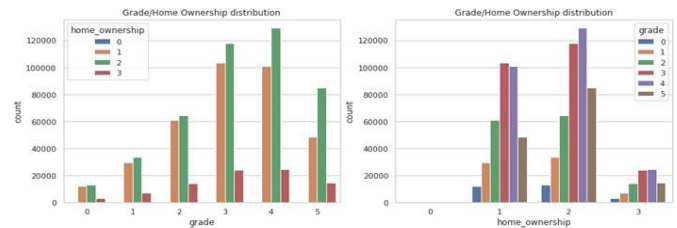
| 1. label | |
|---|----------|
| Keterangan | Kategori |
| "Lunas", "Terkini", "Tidak memenuhi kebijakan kredit. Status: Fully paid" (kondisi yang tidak gagal bayar). | 0 |
| "Terlambat (31-120 hari)", "Terlambat (16-30 hari)", "Dalam Masa Grace Period", "Ditolak", "Tidak memenuhi kebijakan kredit.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

| | |
|--------------------------|-----------------|
| “E” | 1 |
| “D” | 2 |
| “C” | 3 |
| “B” | 4 |
| “A” | 5 |
| 3. emp_length | |
| Keterangan | Kategori |
| ‘< 1 tahun’ | 1 |
| ‘1 tahun’ | 2 |
| ‘2 tahun’ | 3 |
| ‘3 tahun’ | 4 |
| ‘4 tahun’ | 5 |
| ‘5 tahun’ | 6 |
| ‘6 tahun’ | 7 |
| ‘7 tahun’ | 8 |
| ‘8 tahun’ | 9 |
| ‘9 tahun’ | 10 |
| ‘10 tahun’ | 11 |
| ‘10+ tahun’ | 12 |
| 4. home_ownership | |
| Keterangan | Kategori |
| “RENT” | 1 |
| “MORTGAGE” | 2 |
| “OWN” | 3 |

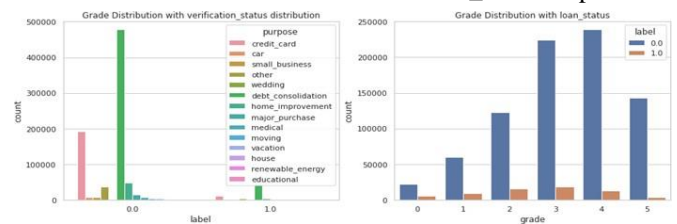
C. Eksplorasi dan Analisis Data

Exploratory data analysis (EDA) atau biasa disebut sebagai Analisis Data Eksploratif merupakan langkah penting dalam setiap analisis penelitian. Tujuan utamanya adalah untuk memeriksa data untuk mencari distribusi, outlier, dan anomali yang dapat mengarahkan pengujian hipotesis tertentu. Selain itu, EDA juga memberikan alat untuk membuat hipotesis dengan memvisualisasikan dan memahami data secara lebih mendalam, biasanya melalui grafik. EDA juga bertujuan untuk membantu analis menemukan pola-pola alami dari data

yang dianalisis. Berikut ini beberapa contoh diagram yang akan dianalisis dalam EDA:

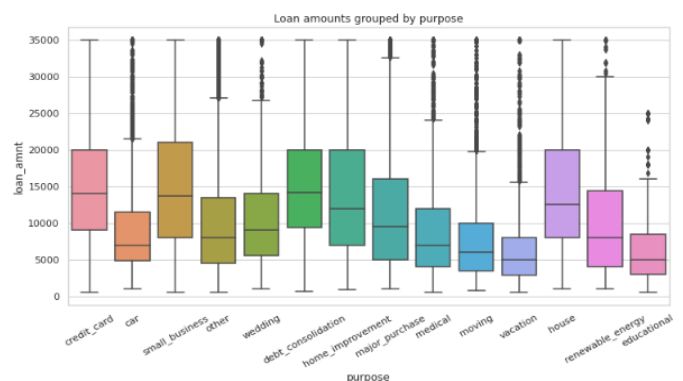


Gambar 4.1 EDA untuk atribut home_ownership

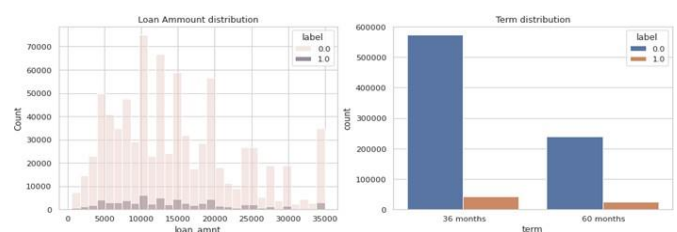


Gambar 4.2 EDA untuk atribut grade

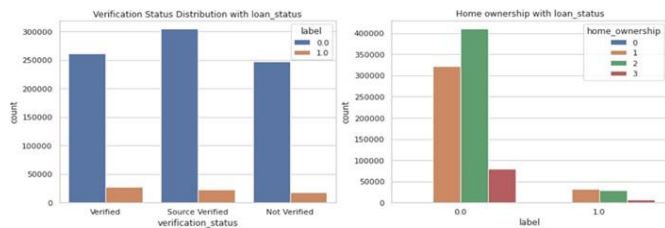
Berdasarkan data yang ditunjukkan pada gambar 4.1 dan gambar 4.2 tersebut, dapat dilihat bahwa orang-orang yang memiliki nilai grade yang lebih tinggi (yang lebih baik) cenderung berjumlah yang lebih sedikit dibandingkan dengan orang-orang yang memiliki grade yang lebih rendah (yang lebih buruk). Selain itu, kebanyakan orang yang meminjam uang dengan label 0 dan 1 menggunakannya untuk konsolidasi utang yang telah mereka miliki. Sementara itu, orang-orang dengan grade tertinggi yang mampu melunasi pinjaman mereka ada pada grade 4, sedangkan yang paling gagal melunasi pinjaman mereka terdapat pada grade 3.



Gambar 4.3 EDA untuk atribut purpose dan loan_amnt. Berdasarkan diagram diatas dapat dianalisis bahwa terdapat 5 kategori dalam jumlah pemohon kredit tertinggi dengan tujuan sebagai berikut: Kartu kredit, bisnis UKM, konsolidasi utang, perbaikan rumah, dan membeli rumah.

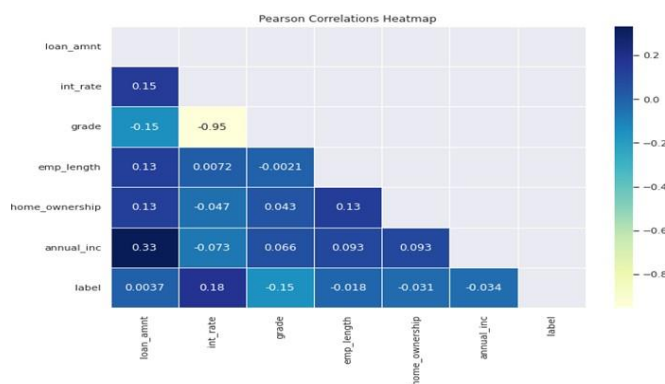


Gambar 4.4 EDA untuk atribut loan_amnt dan term



Gambar 4.5 EDA untuk atribut verification status dan label

Berdasarkan diagram diatas dapat dianalisis bahwa nilai nominal utang terbesar adalah 10000 USD. Jatuh tempo maksimum adalah 36 bulan, sedangkan untuk jatuh tempo 60 bulan jumlahnya hampir sepertiga. Sebagian besar kredit yang dapat dibayar penuh diperoleh dari pemohon kredit dengan status verifikasi “Terverifikasi”.



Gambar 4.6 EDA korelasi dari masing-masing atribut

Dari diagram di atas, dapat dilihat bahwa terdapat korelasi antara setiap variabel yang menunjukkan bahwa variabel jumlah kredit sangat tergantung pada variabel pendapatan tahunan dari seorang peminjam. Hal ini dibuktikan oleh hasil korelasi yang terdapat diantara kedua variabel tersebut yang memiliki korelasi sebesar 0.33.

D. Pasca-Pemrosesan Data

Pasca-pemrosesan data kurang lebih sama seperti pra-pemrosesan sebelumnya. Berikut hasil transformasi pada pasca-pemrosesan:

Tabel 4.3

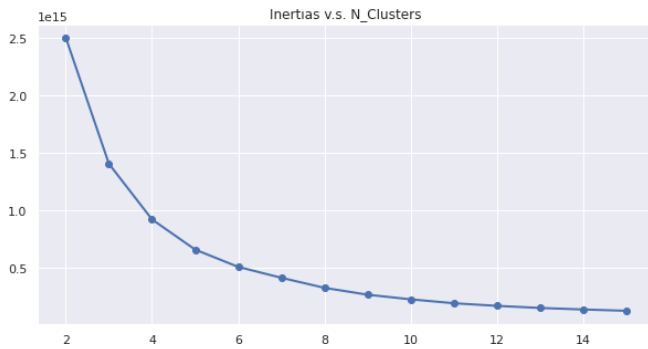
TABEL TRANSFORMASI (PASCA-PEMROSESAN)

| 1. verification_status | |
|------------------------|----------|
| Keterangan | Kategori |
| “verified” | 0 |
| “source verified” | 1 |
| “not verified” | 2 |
| 2. purpose | |

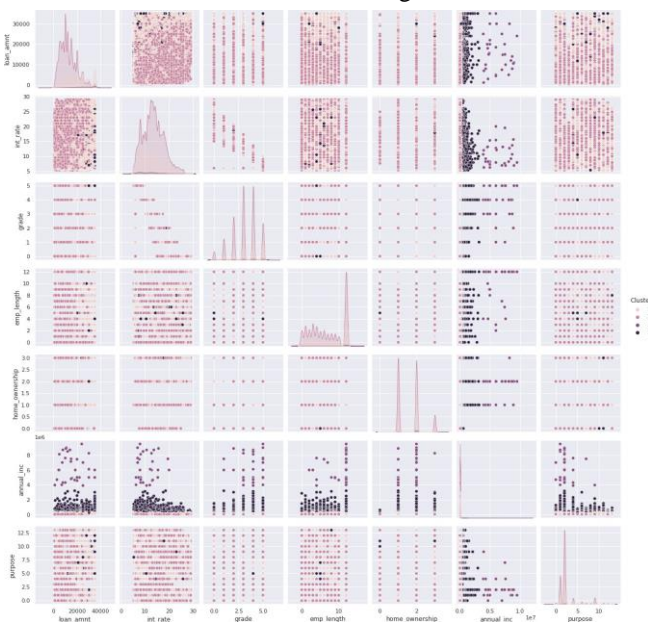
| Keterangan | Kategori |
|----------------------|----------|
| “credit_card” | 0 |
| “car” | 1 |
| “small_business” | 2 |
| “other” | 3 |
| “wedding” | 4 |
| “debt_consolidation” | 5 |
| “home_improvement” | 6 |
| “major_purchase” | 7 |
| “medical” | 8 |
| “moving” | 9 |
| “vacation” | 10 |
| “house” | 11 |
| “renewable_energy” | 12 |
| “educational” | 13 |
| 3. term | |
| Keterangan | Kategori |
| “36 months” | 0 |
| “60 months” | 1 |

E. Metode Elbow

Berdasarkan hasil uji menggunakan metode elbow yang ditunjukkan pada Gambar 4.7, terlihat bahwa cluster terbaik terdapat pada cluster ke-4 hal ini didasarkan pada penurunan terbesar dan angka penurunan selanjutnya yang stabil jika dibandingkan dengan nilai penurunan yang sama. Visualisasi plot dari sebaran data cluster akhir dari semua atribut dengan centroid akhir dari cluster dapat dilihat pada Gambar 4.8.



Gambar 4.7 Cluster terbaik dengan metode elbow



Gambar 4.8 Sebaran cluster dengan centroid cluster

F. Metode K-Mean Clustering

Terdapat 4 kriteria pemohon kredit yang dapat dilihat berdasarkan cluster yang diperoleh sebelumnya menggunakan metode elbow jika dibandingkan dengan titik pusat terakhir pada Gambar 4.9.

| | loan_amnt | term | int_rate | installment | grade | emp_length | home_ownership | annual_inc | verification_status | purpose | label |
|---|-----------|------|----------|-------------|-------|------------|----------------|------------|---------------------|---------|-------|
| 0 | 5000.0 | 0 | 10.65 | 162.87 | 4 | 12 | 1 | 24000.0 | 2 | 1 | 0.0 |
| 1 | 2500.0 | 1 | 15.27 | 59.83 | 3 | 1 | 1 | 30000.0 | 1 | 0 | 1.0 |
| 2 | 2400.0 | 0 | 15.96 | 84.33 | 3 | 12 | 1 | 12252.0 | 0 | 11 | 0.0 |
| 3 | 10000.0 | 0 | 13.49 | 339.31 | 3 | 12 | 1 | 49200.0 | 1 | 9 | 0.0 |

Gambar 4.9 Titik Pusat Terakhir

G. Algoritma K-Nearest Neighbors

Hasil penelitian ini menunjukkan bahwa dengan menerapkan algoritma k-nearest neighbor untuk klasifikasi, jika nilai koefisien k-means sebesar 30, diperoleh hasil akurasi sebesar 91,42% yang dapat dilihat pada gambar di bawah ini:

If we use K-Nearest Neighbors Classification, then the value of K is 30 to get the best prediction, then the average accuracy is 0.9142870821862111

Gambar 4.10 Hasil Akurasi Dengan Algoritma KNN

V. KESIMPULAN

Pada penelitian ini memfokuskan penerapan metode k-means clustering dan algoritma k-nearest neighbor dalam proses penentuan kelayakan pemberian kredit, yang mana menunjukkan bahwa proses penentuan kelayakan kredit menjadi lebih cepat dan akurat. Hal ini dibuktikan dengan penerapan algoritma k-nearest neighbor pada klasifikasi, untuk nilai koefisien k-means sebanyak 30, maka diperoleh hasil akurasi yang sangat baik yaitu sebesar 91,42%. Akan tetapi, proses klasifikasi yang menggunakan model k-nearest neighbors (KNN) membutuhkan banyak waktu dan memori, sehingga untuk penelitian selanjutnya mungkin dapat dilakukan pengujian untuk membandingkan hasil akurasi dengan model machine learning lainnya seperti Support Vector Machine, Decision Tree, Random Forest, dan Gaussian Naive Bayes, dapat digunakan untuk mendapatkan hasil yang mirip dengan model k-nearest neighbor.

DAFTAR PUSTAKA

- [1] Wahyuningsih, S., and Utari, D. R., "Perbandingan Metode K-Nearest Neighbor, Naive Bayes dan Decision Tree untuk Prediksi Kelayakan Pemberian Kredit," Konferensi Nasional Sistem Informasi (KNSI) 2018.
- [2] Saputro, M. I., and Mardiana, T., "Agen cerdas untuk penentuan kelayakan pemberian kredit koperasi simpan pinjam," Jurnal Teknik Komputer AMIK BSI, vol. 1, no. 2, pp. 245-251, 2015.
- [3] Azhari, A., and Watimena, W. R., "Pengembangan Perangkat Lunak Agen Cerdas untuk Analisis Kelayakan Keuangan Perusahaan terhadap Pemberian Kredit Perbankan," Seminar Nasional Aplikasi Teknologi Informasi (SNATI).
- [4] Kurniawan, A., Rifa'i, A., Nafis, M. A., Sefrida, N., & Patria, H. (2022). Pemilihan Metode Predictive Analytics dengan Machine Learning untuk Analisis dan Strategi Peningkatan Kualitas Kredit Perbankan. Indonesian Journal of Applied Statistics, 5(1), 1-11.
- [5] Kurniawan, A., Rifa'i, A., Nafis, M. A., Sefrida, N., and Patria, H., "Pemilihan Metode Predictive Analytics dengan Machine Learning untuk Analisis dan Strategi Peningkatan Kualitas Kredit Perbankan," Indonesian Journal of Applied Statistics, vol. 5, no. 1, pp. 1-11, 2022.
- [6] Ahsina, N., Fatimah, F., and Rachmawati, F., "Analisis Segmentasi Pelanggan Bank Berdasarkan Pengambilan Kredit Dengan Menggunakan Metode K-Means Clustering," Jurnal Ilmiah Teknologi Infomasi Terapan, vol. 8, no. 3.
- [7] Quantum Machine Learning. Germany: De Gruyter, 2020.
- [8] Irwansyah, E., and Faisal, M., "Advanced Clustering: Teori dan Aplikasi," Deepublish.
- [9] Kusri, E. T. L., and Taufiq, E., "Algoritma data mining," Yogyakarta: Andi Offset, 2009.
- [10] Utomo, B., "Rekomendasi Film Berbasis Web Pada Bioskop Mini Menggunakan Algoritma Nearest-neighbor," 2015
- [11] Hidayah, U., Sifaunajah, A., and Kom, M., "Cara Mudah Memahami Algoritma K-Nearest Neighbor Studi Kasus Visual Basic 6.0," Lembaga Penelitian dan Pengabdian kepada Masyarakat Universitas KH. A. Wahab Hasbullah, 2019.
- [12] Sutoyo, M. N., "Rancangan Bangun Aplikasi Untuk Memprediksi Status Gizi Balita," Klik-Kumpulkan Jurnal Ilmu Komputer, vol. 5, no. 2, pp. 136-146, 2018.
- [13] Keuangan, A. L., "Analisis Laporan Keuangan," Dasar-Dasar Analisis Laporan Keuangan, 2021, pp. 37.
- [14] Akbar, M. A., Fatimah, F., and Jaenudin, J., "Penerapan Data Mining Untuk Pengelompokan Posisi Pemain Sepak Bola Menggunakan Algoritma K-Means Clustering," Seminar Nasional Teknologi Informasi, vol. 2, pp. 278-282, 2019, October.
- [15] Hidayatullah, D. P., "Analisis Pemetaan Pelanggan Potensial Menggunakan Algoritma K-Means dan LRFM Model Untuk Mendukung Strategi Pengelolaan Pelanggan (Studi Pada Maninjau Center Kota Malang)," Doctoral dissertation, Universitas Brawijaya, 2017.

- [16] Pane, D. K., "Implementasi Data Mining Pada Penjualan Produk Elektronik Dengan Algoritma Apriori (Studi Kasus: Kreditplus)," *Pelita Informatika Budi Dharma*, vol. 4, no. 3, pp. 25-29, 2013.
- [17] Savitri, A. D., "Segmentasi pelanggan menggunakan metode k-means clustering berdasarkan model RFM pada klinik kecantikan (studi kasus: Belle Crown Malang)," Doctoral dissertation, Universitas Brawijaya, 2018.
- [18] Putu, N., Merliana, E., and Santoso, A. J., "Analisa penentuan jumlah cluster terbaik pada metode K-Means," *Pros. Semin. Nas. Multi Disiplin Ilmu Call Pap. UNISBANK*, pp. 978-979, 2015.
- [19] Rachmawati, T., "Metode Pengumpulan Data dalam Penelitian Kualitatif," UNPAR Press, Bandung, 2017.
- [20] Azizah, A., "Studi kepustakaan mengenai landasan teori dan praktik konseling naratif," Doctoral dissertation, State University of Surabaya, 2017