Jurnal SAINTEKOM (Sains dan Teknologi Komputasi) Vol. 01, No. 03, Desember, 2025
Lembaga Penelitian, Pengembangan, dan Pengabdian Kepada Masyarakat UNBIN
e-ISSN: 3109-1350

Received: Oktober 2025; Accepted: November 2025; Published: Desember 2025
DOI: 10.36350/jskom.v1i3.58

Penerapan Metode *K-Means Clustering* untuk Menentukan Debitur *Existing* Penerima Pinjaman KUR

Ariyati Wuri Kesumaningtias[1*], **Binanda Wicaksana**[2]
Program Studi Sistem Informasi, Fakultas Informatika dan Komputer, Universitas Binaniaga Indonesia
email: arya.kyoukun@yahoo.co.id
*Corresponding Author

ABSTRACT

One of the factors influencing the health level of a bank is its risk profile. The risk profile includes credit risk, which financial institutions must anticipate when providing loans to customers. This process begins with an accurate analysis of debtors. Account Officers are employees responsible for managing customer loans, which involves credit analysis, monitoring business development, and ensuring that debtors pay their principal loan installments and interest on time. To identify eligible debtors for loans, Account Officers need to first classify debtor data. This study focuses on grouping eligible debtors using the K-Means Algorithm. The K-Means Algorithm is utilized as an effective data mining technique for clustering to achieve accurate data. The data used in this study includes attributes such as loan term, current collectability, previous collectability, business inventory, business results, and loan installments. The analysis process employs the Silhouette Coefficient Method to measure cluster performance, yielding a score of 0.7299, which falls into the "Strong Structure" category in terms of data similarity. This research has undergone testing, including expert validation with a feasibility percentage of 100%, indicating that the system is viable for development, and user testing, which resulted in a score of 89.85%, categorizing the developed program as highly feasible for further enhancement.

Keywords*: data mining, k-means, credit risk, silhouette coefficient, debtor*

ABSTRAK

Salah satu faktor yang mempengaruhi tingkat kesehatan bank yaitu profil risiko. Pada profil risiko terdapat risiko kredit yang harus diantisipasi oleh lembaga keuangan dalam memberikan pinjaman kepada nasabah yang diawali dengan proses analisa debitur dengan tepat. Account Officer adalah karyawan yang bertugas mengelola kredit nasabah yang didalamnya terdapat proses analisa kredit, pemantauan perkembangan usaha dan memastikan debitur membayar tunggakan pinjaman pokok dan bunga tepat waktu. Untuk memperoleh debitur yang layak diberikan pinjaman, Account Officer perlu melakukan pengelompokan data debitur terlebih dahulu. Penelitian ini dilakukan untuk pengelompokan debitur yang layak diberikan pinjaman dengan menggunakan Algoritma K-Means. Algoritma K-Means digunakan sebagai salah satu teknik data mining yang efektif dalam pengelompokan untuk memperoleh data yang akurat. Data yang digunakan mencakup atribut seperti, jangka waktu pinjaman, kolektabilitas saat ini, kolektabilitas sebelumnya, persediaan usaha, hasil usaha dan angsuran pinjaman. Proses analisis dilakukan dengan menggunakan Metode Silhouette Coefficient untuk mengukur performa cluster dan diperoleh nilai sebesar 0.7299 yang termasuk dalam kategori Strong Structure untuk tingkat kemiripan data. Penelitian ini sudah dilakukan pengujian diantaranya uji ahli dengan nilai pesentase kelayakan sebesar 100% yang artinya sistem dapat dikembangkan, serta uji pengguna dengan nilai sebesar 89.85% yang artinya program yang dibuat termasuk kategori sangat layak untuk dikembangkan.

Kata Kunci*: data mining, k-means, risiko kredit, silhouette coeffisient, debitur*

A. PENDAHULUAN

1. LATAR BELAKANG

Menurut Undang-Undang No. 20 Tahun 2008, UMKM (Usaha Mikro Kecil dan Menengah) memiliki pengertian sebagai Usaha Mikro, yaitu usaha produktif milik orang perorangan dan atau badan usaha perorangan yang memenuhi kriteria usaha mikro sebagaimana diatur dalam undang-undang.UMKM memiliki peran strategis dalam perekonomian nasional baik dalam meningkatkan pertumbuhan ekonomi maupun dalam menyerap tenaga kerja dimana sebagian besar pengusaha atau pelaku UMKM merupakan kegiatan usaha rumah tangga yang memungkinkan menyerap banyak tenaga kerja dan membantu mengurangi tingkat pengangguran. Berdasarkan data dari Kamar Dagang dan Industri (KADIN) Indonesia, UMKM menyumbang hingga 99% dari total unit usaha di Indonesia. Pada tahun 2023, jumlah pelaku usaha UMKM mencapai sekitar 66jt, dengan kontribusi terhadap Pendapatan Domestik Bruto (PDB) melampaui 61% atau setara Rp. 9.580 Trilyun. UMKM juga mampu menyerap sekitar 117 juta tenaga kerja, yang mencakup 97% dari total tenaga kerja nasional.

Meningkatnya industri UMKM yang sebagian besar adalah industri rumahan sangat berpengaruh pada kebutuhan modal dalam mengembangkan usaha, disinilah peran pemerintah dalam memberikan fasilitas pemberian modal usaha salah satunya dengan meluncurkan Program Kredit Usaha Rakyat (KUR). Program Kredit Usaha Rakyat (KUR) pertama kali diluncurkan pada tanggal 5 November 2007 yang bertujuan untuk memberdayakan Usaha Mikro, Kecil, Menengah dan Koperasi (UMKM) di Indonesia dengan pembiayaan yang bersumber dai dana perbankan atau lembaga keuangan yang merupakan penyalur KUR.Dalam proses penyaluran pinjaman, lembaga keuangan dan perbankan perlu melakukan analisa kredit yang bertujuan untuk meminimalisir risiko kredit macet dan memastikan bahwa kredit yang digunakan sesuai dengan perjanjian. Analisis kredit dapat dilakukan dengan menggunakan pinsip 5C (*character*, *capacity*, *capital*, *collateral*, *condition of economy*), 5P (*purpose*, *payment*, *people*, *protection*, *perspective*), dan 3R (*risk*, *return*, *relationship*)

2. PERMASALAHAN

Berdasarkan data yang ditunjukkan pada Tabel 1.2 terdapat data debitur yang diberikan penawaran berjumlah 418 dari total 425 debitur existing dengan status kolektabilitas 1 (lancar) dan kolektabilitas 2 (dalam perhatian

khusus), sedangkan sisanya, 7 debitur lainnya tidak mendapatkan penawaran karena status kolektabilitas 3 (kurang lancar), kolektabilitas 4 (diragukan), dan kolektabilitas 5 (macet). Terlihat adanya debitur atas nama Sri Retno dengan status kolektabilitas 2 yang memperoleh penawaran pinjaman, meskipun mayoritas debitur lainnya berstatus kolektabilitas 1. Selain itu, debitur atas nama Atang, yang meskipun memiliki nilai plafond dan outstanding yang masih tinggi, tetap diberikan penawaran pinjaman. Dari data pengelompokan tersebut belum tepat karena hanya dinilai dari status kolektabilitasnya saja. Dari hasil identifikasi tersebut, maka penelitian akan menambahkan atribut pendukung lainya yang masih relevan dengan kriteria pemilihan debitur, yaitu: jangka waktu, kolektabilitas sekarang, kolektabilitas sebelumnya, persediaan usaha, hasil usaha, dan angsuran pinjaman.

Tabel 1 Sampel Data Debitur Periode TW 3 Juli-September 2023

NO	NAMA DEBITUR	START_DATE	MAT_DATE	PLAFOND	OUTSTANDING	COLLECT
1	HELI SUPART	9/21/2020	9/21/2023	500.000.000	41.666.663	1
2	NURKAM	9/18/2022	9/18/2023	150.000.000	75.000.000	1
3	TOTONG	8/9/2022	8/9/2023	100.000.000	100.000.000	1
4	ABSOH	8/8/2022	8/8/2023	100.000.000	100.000.000	1
5	RONIAH	9/28/2020	9/28/2023	150.000.000	24.999.990	2
6	SITI ROAH	8/24/2021	8/24/2023	100.000.000	8.333.326	1
7	SUTARMAN	8/24/2021	8/24/2023	500.000.000	41.666.674	1
8	TURSIWAN	9/8/2022	9/8/2023	100.000.000	100.000.000	1
9	APIP SH	7/27/2021	7/27/2023	500.000.000	20.833.341	1
10	INTAN SARI	9/15/2020	9/15/2023	100.000.000	8.333.326	1
…	…	…	…	…	…	…
412	ATANG	3/13/2022	9/13/2023	486.647.500	486.647.500	1
413	TENDI SETIA	8/25/2020	8/25/2023	175.000.000	24.305.559	2
414	OJA	8/23/2020	8/23/2023	100.000.000	5.555.548	1
415	HAYATI	8/30/2020	8/30/2023	100.000.000	8.333.326	1
416	M RIKI N	8/11/2021	8/11/2023	100.000.000	8.333.326	1
417	SUTARYAT	8/8/2021	8/82023	100.000.000	8.333.326	1
418	SALIM	8/16/2022	8/16/2023	100.000.000	50.000.000	1

Berdasarkan permasalahan diatas maka dapat diidentifikasi masalah sebagai berikut:

a. Proses pengelompokan debitur yang berhak memperoleh penawaran pinjaman belum berjalan secara efektif karena tidak mempertimbangkan faktor-faktor pendukung lainnya yang lebih relevan.
b. Data debitur existing yang memenuhi kriteria untuk mendaparkan penawaran pinjaman kembali belum dikelompokkan dengan tepat, sehingga berpotensi menyebabkan penawaran diberikan kepada debitur yang kurang layak.

3. TUJUAN

Tujuan dari penelitian ini adalah:

a. Menghasilkan proses yang lebih efektif dalam pengelompokkan debitur yang layak menerima penawaran pinjaman.
b. Mengelompokkan secara akurat debitur yang layak menerima penawaran pinjaman.
c. Mengembangkan prototype aplikasi untuk pengelompokan debitur yang layak menerima penawaran pinjaman.
d. Mengukur tingkat akurasi dan efektivitas penerapan metode K-Means Clustering untuk pengelompokan debitur.

4. TINJAUAN PUSTAKA

a. Data Mining Clustering

Data mining menurut David Hand, Heikki Mannila, dan Padhraic Smyth dari MIT adalah analisis terhadap data (biasanya data yang berukuran besar) untuk menemukan hubungan yang jelas serta menyimpulkannya yang belum diketahui sebelumnya dengan cara terkini dipahami dan berguna bagi pemilik data tersebut Menurut Daryl Pregibon disebutkan bahwa Data Mining adalah perpaduan dari Statistik, Artificial Intelligent dan Database (Gorunescu, 2011). Data Mining kemudian dikenal dengan nama Knowledge-Discovery in Databases (KDD) (Rahmadya dan Herlawati,2020, pp.1-2)

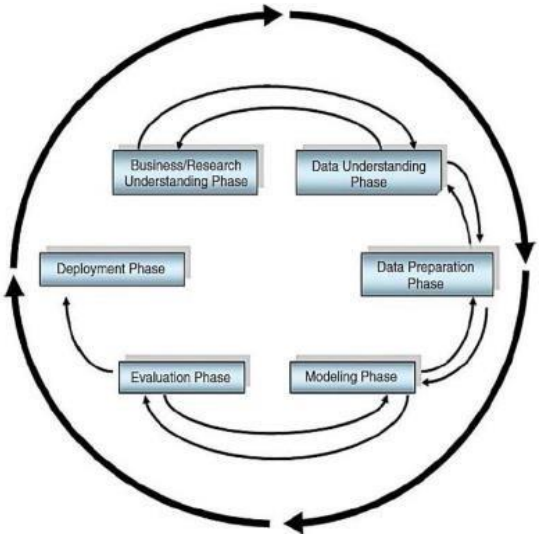
b. KUR

Menurut Agusfianto dkk (2022,p.67) secara sederhana, kredit diartikan sebagai penyaluran dana dari pihak pemilik dana kepada pihak yang memerlukan dana. Penyaluran tersebut didasarkan atas kepercayaan yang diberikan oleh pemilik dana kepada pengguna dana, dimana kredit berasal dari Bahasa latin "*Credere*" yang berarti kepercayaan. Artinya pihak yang memberikan kredit percaya kepada pihak yang menerima kredit bahwa kredit yang diberikan pasti akan dibayar.

c. CRISP-DM

Metodologi datamining CRISP-DM menjelaskan proses model proses secara hirarki, merupakan sekumpulan tugas yang dinyatakan dengan empat level abstrak (dari umum ke spesifik), tahap, tugas umum, tugas khusus dan contoh proses (proses instance). Pada level yang paling atas, proses datamining dikenal dengan empat tahap, masing-masing tahap terdiri dari beberapa tugas umum. Tugas umum digunakan

sebagai pelengkap dan setabil mungkin. Lengkap artinya mencakup baik proses data mining secara umum dan semua kemungkinan aplikasi data mining. Stabil artinya model harus valid sebelum meramalkan pengembangan secara Teknik pemodelan yang baru (Mulaab, 2021, p.8).

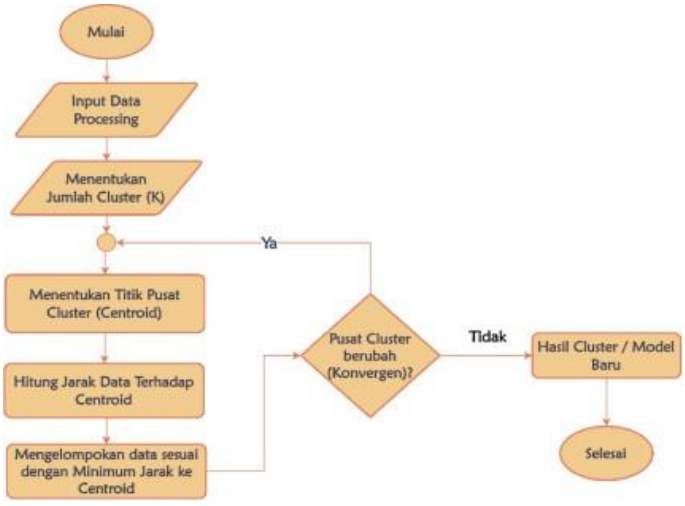


Gambar 1. Ilustrasi CRISP-DM (Sumber: Mulaab, 2021,p.8)

B. METODE

1. Metode Penelitian

Menurut Ibnu Daqiqil (2021:217) K-Means adalah salah satu "*unsupervised machine learning algorithms*" yang paling sederhana dan popular. Tujuan dari agoritma ini adalah untuk menemukan grup dalam data, dengan jumlah grup yang diwakili oleh variable K. Variabel K sendiri adalah jumlah kluster yang kita inginkan. Metode K-Means Clustering berusaha mengelompokkan data yang ada kedalam beberapa kelompok, dimana data dalam satu kelompok mempunyai karakteristik yang sama satu sama lainnya dan mempunyai karakteristik yang berbeda dengan data yang ada didalam kelompok yang lain. Karajkteristik yang sama itu ditandai dengan jarak atau *distance* yang lebih dekat, mirip seperti KNN. Dengan kata lain, metode K-Means Clustering bertujuan untuk meminimalisir *objective function* yang diset dalam proses *clustering* dengan cara meminimalkan variasi antar data didalam suatu *cluster* dan memaksimalkan variasi dengan data yang ada di *cluster* lainnya.



Gambar 2. Alur kerja Algoritma K-Means
(Sumber: Putra dkk, 2023)

Berikut ini merupakan penjelasan lebih lanjut terkait proses K-Means dari Gambar 2.2:

a. Pilih nikai K untuk menentukan jumlah *cluster* yang akan dibentuk
b. Menetapkan titik K secara acak yang akan bertindak sebagai pusat *cluster (centroid)*. Menentukan nilai *centroid* dengan cara mengambil dari nilai rata-rata (mean) semua nilai data pada setiap fiturnya. Jika *M* menyatakan jumlah data pada suatu kelompok, maka *i* menyatakan fitur ke-*i* dalah sebuah kelompok.

$$C_i = \frac{1}{M}\sum_{j=i}^{M} x_j$$

Perhitungan jarak dari *centroid cluster*. Untuk mengukur jarak antar data dengan *centroid*menggunakan formula Euclidean distance.

$$D(x_2, x_1) = \sqrt{\sum_{j=1}^{p} |x_{2j} - x_{1j}|^2}$$

a. Tetapkan setiap titik data, berdasarkan jaraknya dari titik yang dipilih secara acak (*centroid*), ke *centroid* terdekat, yang akan membentuk cluster yang telah ditentukan sebelumnya.
b. Tempatkan *centroid* baru dari setiap *cluster*.
c. Ulangi Langkah 3 sampai Langkah 5 hingga nilai titik *centroid* tidak berubah (konvergen)
d. Jika telah konvergen, maka model baru siap digunakan.

$$a_{ji} = \begin{cases} 1, & d = min(x_j, C_i)) \\ 0, & lainnya \end{cases}$$

2. Teknik Analisis Data

Dalam penelitian ini menggunakan metode K-Means, uji hasil akan digunakan adalah *Silhouette coefisien*. Metode ini digunakan untuk mengetahui tingkat kemiripan data melalui perhitungan jarak antar data, semakin kecil jarak antar data maka semakin tinggi kemiripan data tersebut. Perhitungan nilai Silhouette Coeffisien terdapat dua komponen yaitu *a(i)* dan *b(i)*. *a(i)* adalah rata-rata jarak data ke-*i* dengan semua data lainnya dalam satu cluster, sedangkan *b(i)* didapaktkan dengan menghitung rata-rata jarak data ke-*i* terhadap semua data lainnya dalam satu cluster yang lain yang tidak dalam satu cluster dengan data ke-*i*, kemudian diambil yang terkecil.

$$a_i^j = \frac{1}{m_j - 1}\sum_{\substack{r=1 \\ r \neq 1}}^{m_j} d\left(x_i^j, x_r^j\right)$$

Keterangan:
j = cluster
i = index data (i = 1,2 … mj)a^j = rata-rata jarak data ke-i terhadap semua data dalam satu *cluster*
M_j = jumlah data dalam *cluster* ke-j
d (x^j, x^j) = jarak data ke-i dengan data ke-i dalam satu *cluster* j
i r
Berikut ini adalah rumus perhitungan untuk mendapatkan nilai *b(i)*:

$$b_i^j = min_{\substack{n=1,\dots k \\ n \neq j}} \left\{ \frac{1}{m_n}\sum_{\substack{r=1 \\ r \neq 1}}^{m_n} d\left(x_i^j, x_r^n\right) \right\}$$

Keterangan:
j = cluster
i = index data (i = 1,2 … mj)
b^j = rata-rata jarak data ke-i terhadap semua data yang tidak dalam satu cluster dengan data ke-i
M_j = jumlah data dalam *cluster* ke-n
d (x^j, x^n) = jarak data ke-i dengan data ke-j dalam satu *cluster* n

Kriteria subjektif menggunakan pengelompokan berdasarkan Silhouette Coeffisient menurut Kaufman dan Roesseeuw (1990) dapat dilihat pada tabel 2.

Tabel 2. Silhouette Coefisien

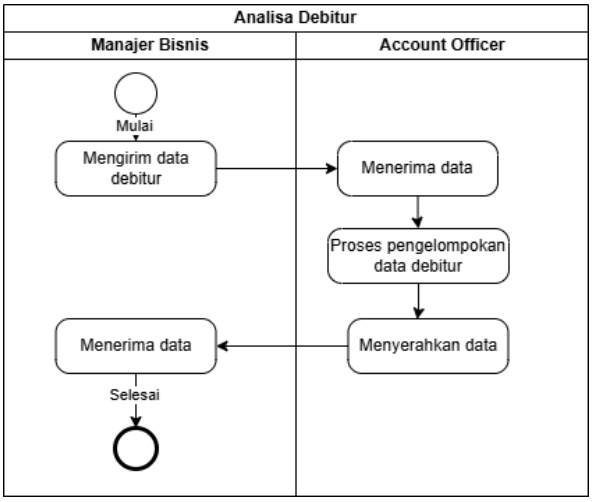
Skala	Keterangan
0.71 – 1.0	Strong Structure
0.52 – 0.70	Medium Structure
0.26 – 0.50	Weak Structure
<= 0.25	No Structure

(Sumber: Kauffman dan Roesseeuw, 1990

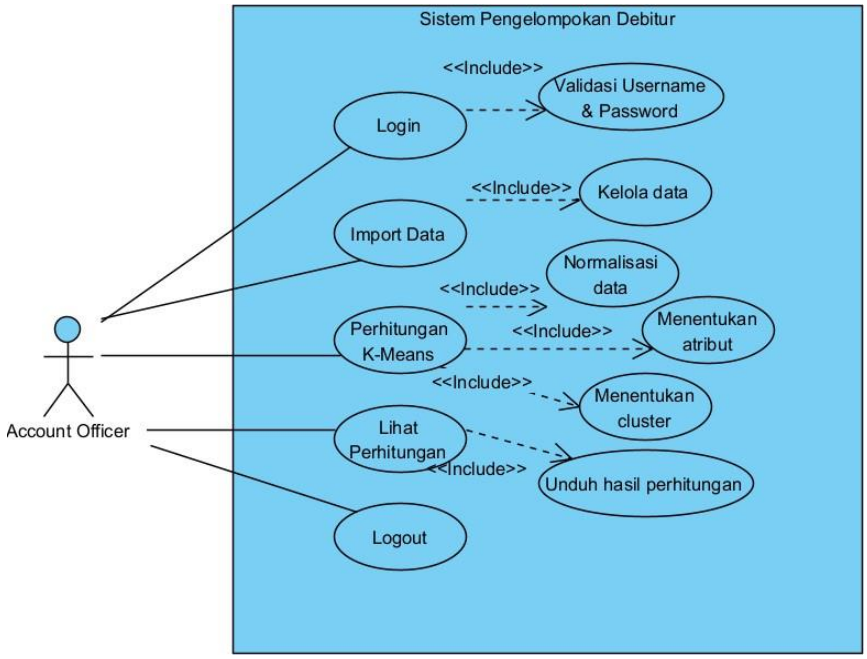
C. HASIL DAN PEMBAHASAN

1. Hasil

a. Hasil Analisis Kebutuhan



Gambar 3. Proses Bisnis yang sedang berjalan



Gambar 4. Usecase Diagram

Berdasarkan Gambar, terdapat kelemahan pada sistem lama, pada sistem baru, yang digambarkan dalam use case diagram, dijelaskan bahwa account officer selaku actor dalam sistem melakukan import data dengan 6 atribut yang digunakan untuk analisa data diantaranya jangka waktu, kolektabilitas sebelumnya, kolektabilitas sekarang, persediaan usaha, hasil usaha, dan angsuran pinjaman.

b. Perhitungan dengan Metode K-Means

1) Menyiapkan dataset

Dataset yang digunakan dapat dilihat pada tabel 4.2 yang berisi 418 data dengan 6 variabel.

Tabel 3. Penentuan Centroid Awal

DB	JW	CN	CP	PU	HU	AK
DB1	3	1	31	0.857	0.057	0.526
DB2	3	1	15	0.714	0.206	0.429
DB3	4	1	0	0.857	0.202	0.116
DB4	1	1	25	0.000	0.108	0.145

DB5	1	1	0	0.000	0.007	0.006
…	…	…	…	…	…	…
DB415	1	1	0	0.000	0.000	0.000

2) Menentukan jumlah cluster
Berdasarkan hasil analisa kebutuhan pengguna, dati tabel dataset akan dibuat 2 (dua) kelompok yaitu debitur yang menerima penawaran pinjaman dan debitur yang tidak menerima penawaran pinjaman.

3) Menentukan Centroid Awal
Untuk penentuan centroid awal dilakukan secara acak dengan memilih data yang ada dalam dataset untuk dijadikan centroid awal. Dalam pengelompokan data debitur ini, data yang dipilih menjadi centroid awal dalam perhitungan ini adalah data ke-115 dan data ke 356 yang dapat dilihat pada tabel berikut.

Tabel 4. Titik Awal Centroid

Centroid 1	1	1	31	0.314	0.347	0.016
Centroid 2	4	1	44	0.743	0.818	0.084

4) Menghitung Jarak dengan Pusat Data (centroid)
Menghitung jarak setiap data input terhadap masing-masing pusat cluster menggunakan rumus euclidean distance untuk menghitung jarak objek dengan masing-masing centroid dengan rumus persamaan sebagai berikut:

$$d\,(x_i c_k) = \sqrt{\sum_{i=1}^{r}(x_{ij} - c_{kj})^2}$$

Keterangan:
i = jumlah data (row/baris)
j = jumlah fitur
k = jumlah Centroid

Iterasi I
Data ke-1 pada iterasi I

$d\,(x_1, c_1) = \sqrt{(3-1)^2 + (1-1)^2 + (31-31)^2 + (0.857-0.314)^2 + (0.057-0.347)^2 + (0.526-0.016)^2} = 2.154$

$d\,(x_1, c_2) = \sqrt{(3-4)^2 + (1-1)^2 + (31-44)^2 + (0.857-0.743)^2 + (0.057-0.818)^2 + (0.526-0.084)^2} = 13.069$

5) Mengklasifikasikan setiap data berdasarkan kedekatannya dengan centroid (jarak terkecil)

Tabel 5. Hasil Pengelompokan pada Iterasi-I

DB	C1	C2	JARAK	CLUSTER
DB1	2.154	13.069	2.154	C1
DB2	16.135	29.026	16.135	C1
DB3	31.150	44.004	31.150	C1
DB4	6.014	19.263	6.014	C1
DB5	31.003	44.116	31.003	C1
DB6	31.002	44.113	31.002	C1
DB7	31.003	44.115	31.003	C1
DB8	11.006	24.207	11.006	C1
DB9	31.003	44.116	31.003	C1
…	…	…	…	…
DB391	28.020	15.155	15.155	C2
DB392	28.073	15.060	15.060	C2

6) Memperbaharui Nilai centroid baru di peroleh dari rata-rata cluster yang bersangkutan dengan menggunakan rumus:

$$c_j = \frac{1}{Nk}\sum_{i=1}^{Nk} x_{ji}$$

Keterangan:
Nk = jumlah data yang tergabung dalam cluster

Tabel 6 Titik Centroid Baru Iterasi ke-II

Centroid	JW	CN	CP	PU	HU	AK
C1	1.543	1.017	10.424	0.128	0.271	0.052
C2	1.614	1.018	49.158	0.100	0.267	0.030

7) Melakukan perulangan dari Langkah 4 hingga 6, sampai anggota tiap *cluster* tidak ada yang berubah

Iterasi ke-II
Data ke-1 pada iterasi II

$d\,(x_1, c_1) = \sqrt{(3 - 1.543)^2 + (1 - 1.017)^2 + (31 - 10.424)^2 + (0.857 - 0.128)^2 + (0.057 - 0.271)^2 + (0.526 - 0.052)^2} = 20.647$

$d\,(x_1, c_2) = \sqrt{(3 - 1.614)^2 + (1 - 1.018)^2 + (31 - 49.158)^2 + (0.857 - 0.100)^2 + (0.057 - 0.267)^2 + (0.526 - 0.030)^2} = 18.234$

….

Iterasi ke-IV
Data ke-1 pada iterasi IV

$d\,(x_1, c_1) = \sqrt{(3 - 1.471)^2 + (1 - 1.010)^2 + (31 - 5.311)^2 + (0.857 - 0.115)^2 + (0.057 - 0.263)^2 + (0.526 - 0.046)^2} = 25.750$

$d\,(x_1, c_2) = \sqrt{(3 - 1.736)^2 + (1 - 1.031)^2 + (31 - 38.992)^2 + (0.857 - 0.146)^2 + (0.057 - 0.288)^2 + (0.526 - 0.055)^2} = 8.140$

Tabel 7. Hasil Iterasi ke-IV

DB	C1	C2	JARAK	CLUSTER
DB2	3	1	15	0.714
DB3	4	1	0	0.857
DB5	1	1	0	0.000
DB6	1	1	0	0.000
DB7	1	1	0	0.000
DB8	1	1	20	0.000
DB9	1	1	0	0.000
DB10	1	1	0	0.000
DB11	1	1	0	0.000
…	…	…	…	…
DB416	5	1	26	0.314
DB418	1	1	33	0.286

Tabel 8. Centroid Baru Iterasi 4

Centroid	JW	CN	CP	PU	HU	AK
C1	1.471	1.010	5.311	0.115	0.263	0.046
C2	1.736	1.031	38.992	0.146	0.288	0.055

8) Hasil Cluster
   a. Cluster 1 (Kelompok Debitur Penerima Penawaran Pinjaman)
   Dari hasil perhitungan yang telah dilakukan dengan metode K-Means Clustering, diperoleh hasil sebanyak **289** debitur yang masuk dalam kelompok debitur yang mendapatkan penawaran pinjaman, Debitur yang mendapatkan penawaran pinjaman adalah:

Tabel 9. Kelompok Cluster 1 (Debitur Penerima Penawaran Pinjaman)

NO	DB	JW	CN	CP	PU	HU	AK	C1	C2	JARAK	CLUSTER
1	DB2	3	1	15	0.714	0.206	0.429	9.834	24.035	9.834	C1
2	DB3	4	1	0	0.857	0.202	0.116	5.930	39.065	5.930	C1
3	DB5	1	1	0	0.000	0.007	0.006	5.340	39.001	5.340	C1
4	DB6	1	1	0	0.000	0.171	0.006	5.334	39.000	5.334	C1
5	DB7	1	1	0	0.000	0.044	0.006	5.338	39.000	5.338	C1
…	…	…	…	…	…	…	…	…	…	…	…
285	DB412	3	1	18	0.000	0.111	0.079	12.782	21.032	12.782	C1
286	DB413	4	1	19	0.714	0.398	0.047	13.934	20.128	13.934	C1
287	DB414	4	1	0	0.000	0.000	0.036	5.890	39.059	5.890	C1

NO	DB	JW	CN	CP	PU	HU	AK	C1	C2	JARAK	CLUSTER
288	DB415	1	1	0	0.000	0.000	0.000	5.340	39.001	5.340	C1
289	DB417	4	1	0	0.314	0.000	0.000	5.892	39.059	5.892	C1

b. Cluster 2 (Kelompok Debitur Tidak Menerima Penawaran Pinjaman)
Dari hasil perhitungan yang telah dilakukan dengan metode K-Means Clustering, diperoleh hasil sebanyak **129** debitur yang masuk dalam kelompok debitur yang tidak mendapatkan penawaran pinjaman, Debitur yang tidak mendapatkan penawaran pinjaman adalah:

Tabel 10 Kelompok Cluster 1 (Debitur Penerima Penawaran Pinjaman)

NO	DB	JW	CN	CP	PU	HU	AK	C1	C2	JARAK	CLUSTER
1	DB1	3	1	31	0.857	0.057	0.526	25.750	8.140	8.140	C2
2	DB4	1	1	25	0.000	0.108	0.145	19.695	14.014	14.014	C2
3	DB25	2	1	35	0.714	1.000	0.526	29.712	4.130	4.130	C2
4	DB28	1	1	35	0.000	0.107	0.006	29.693	4.067	4.067	C2
5	DB29	1	1	35	0.000	0.195	0.006	29.693	4.064	4.064	C2
…											
125	DB404	3	1	33	1.000	0.983	0.093	27.754	6.222	6.222	C2
126	DB409	3	1	33	1.000	0.920	0.157	27.753	6.216	6.216	C2
127	DB411	1	1	33	0.000	0.067	0.000	27.694	6.043	6.043	C2
128	DB416	5	1	26	0.314	0.096	0.139	20.989	13.399	13.399	C2
129	DB418	1	1	33	0.286	0.000	0.000	27.694	6.046	6.046	C2

2. Pembahasan

Berdasarkan hasil pengembangan terkait metode yang diterapkan telah dilakukan uji hasil dengan menerapkan metode Silhouette Coeffisient. Silhouette Coeffisient adalah metrik untuk mengukur seberapa baik objek telah dikluster. Nilainya berkisar antara -1 hingga 1, dimana nilai tinggi menunjukkan bahwa objek cocok dengan baik di dalam klusternya dan tidak cocok dengan kluster tetangga (Prihandoko dkk,2024,p.31)

Tabel 11. Hasil Clustering

NO	DATA	JW	CN	CP	PU	HU	AK	CLUSTER
1	DB2	3	1	15	0.714	0.206	0.429	1
2	DB3	4	1	0	0.857	0.202	0.116	1
3	DB5	1	1	0	0.000	0.007	0.006	1
4	DB6	1	1	0	0.000	0.171	0.006	1
5	DB7	1	1	0	0.000	0.044	0.006	1
6	DB8	1	1	20	0.000	0.170	0.006	1
…	…	…	…	…	…	…	…	…
413	DB400	1	1	33	0.000	0.075	0.081	2
415	DB409	3	1	33	1.000	0.920	0.157	2
416	DB411	1	1	33	0.000	0.067	0.000	2
417	DB416	5	1	26	0.314	0.096	0.139	2
418	DB418	1	1	33	0.286	0.000	0.000	2

Pada tabel 11 menunjukkan nilai cluster dari masing-masing data yang berjumlah 418 dengan anggota Cluster 1 sebanyak 289 debitur dan anggota Cluster 2 sebanyak 129 debitur. Kemudian dari hasil perhitungan data diatas akan dilakukan perhitungan untuk mencari nilai silhouette coefficient dengan langkah-langkah sebagai berikut:

1. Menghitung jarak Euclidean antar data:

Tabel 12 Hasil perhitungan Jarak Euclidean

NO	DATA	CLUSTER	Jarak DB2	Jarak DB3	Jarak DB5	…	Jarak DB400	Jarak DB404	Jarak DB409
1	DB2	1	0	15.04	15.16	…	18.13	18.02	18.02
2	DB3	1	15.04	0	3.13	…	33.15	33.02	33.02
3	DB5	1	15.16	3.13	0	…	33.00	33.09	33.09
…	…	…	…	…	…	…	…	…	…
413	DB400	2	18.13	33.15	33.00	…	0	2.41	2.39
414	DB404	2	18.02	33.02	33.09	…	2.41	0	0.09
415	DB409	2	18.02	33.02	33.09	…	2.39	0.09	0

2. Menghitung nilai a(i) yaitu jarak rata-rata data dalam kelompok Cluster yang sama:

a. Nilai a(i) untuk data DB2. Menghitung jarak data DB2, DB3 s.d DB417 (data pada Cluster 1 dengan jumlah sebanyak 289 data):

$$a(1) = \frac{d(DB2,DB2) + d(DB2,DB3) + \cdots + d(DB2,DB417)}{289} = \frac{3024.55}{289} = 10.466$$

b. Nilai a(i) untuk data DB400. Menghitung jarak data mulai dari data DB1, DB4 s.d DB418 (data pada Cluster 2 dengan jumlah sebanyak 129 data):

$$a(2) = \frac{d(DB400,DB1) + d(DB400,DB4) + \cdots + d(DB400,DB418)}{129} = \frac{1128.91}{129} = 8.751$$

3. Menghitung niai b(i) yaitu jarak rata-rata data ke Cluster lain:
   a. Nilai b(i) untuk data DB2. Menghitung jarak ke Cluster 2

$$b(1) = \frac{d(DB2,DB1) + d(DB2,DB4) + \cdots + d(DB2,DB418)}{129} = \frac{3105.08}{129} = 24.070$$

   b. Nilai b(i) untuk data DB400. Menghitung jarak ke Cluster 1

$$b(2) = \frac{d(DB400,DB2) + d(DB400,DB3) + \cdots + d(DB400,DB417)}{289} = \frac{7994.47}{289} = 27.663$$

4. Menghitung nilai Silhouette Coeffisient:

Tabel 13. Hasil perhitungan nilai a(i), b(i) dan s(i)

NO	DATA	a(i)	b(i)	s(i)
1	DB2	10.466	24.070	0.565
2	DB3	6.966	39.084	0.822
3	DB5	5.697	39.019	0.854
4	DB6	5.671	39.018	0.855
5	DB7	5.684	39.019	0.854
6	DB8	14.770	19.064	0.225
…	…	…	…	…
415	DB409	9.007	27.723	0.675
416	DB411	8.752	27.663	0.684
417	DB416	13.832	21.020	0.342
418	DB418	8.757	27.642	0.683

Tabel 13 menunjukkan nilai a(i) jarak data antar cluster, b(i) jarak data dengan cluster lain, dan s(i) nilai silhouette coefficient masing-masing data, sehingga diperoleh nilai s (i) rata-rata dari keseluruhan data berada pada skala 0.7299 dan berdasarkan kategori nilai silhouette coefficient menurut Prasetyo (2014) nilai 0.71 – 1.0 masuk kategori Strong Structure artinya terdapat ikatan yang sangat baik (strong structure) antara objek dan kelompok yang terbentuk.

D. KESIMPULAN

Berdasarkan hasil penelitian yang dilakukan, kesimpulan yang bisa diuraikan antrara lain:

1. Penelitian menggunakan algoritma K-Means menghasilkan pengelompokan debitur menjadi dua cluster: 289 debitur layak menerima penawaran pinjaman (Cluster 1) dan 129 debitur tidak layak menerima penawaran pinjaman (Cluster 2). Validasi hasil penelitian menunjukan bahwa uji hasil menggunakan *Silhouette Coefficient* menghasilkan nilai 0.7299, yang masuk dalam kategori *strong structure*, menegaskan bahwa jumlah cluster dalam penelitian ini optimal. Oleh karena itu, penelitian ini layak digunakan.
2. Hasil pengelompokan membantu petugas *Account Officer* mempercepat proses analisa debitur, memberikan hasil yang lebih objektif serta menghasilkan informasi yang mendukung Manager Bisnis dalam menyusun strategi pemasaran dan mitigasi risiko secara lebih efektif.
3. Perbedaan atribut antara proses manual dengan perhitungan K-Means bertujuan untuk mendapatkan segmentasi debitur yang lebih mendalam. Atribut seperti hasil usaha, persediaan usaha, dan angsuran pinjaman digunakan untuk menilai kesanggupan bayar debitur, sehingga meningkatkan relevansi hasil pengelompokan.

D. DAFTAR PUSTAKA

[1] Daqiqil Ibnu. (2021). *Machine Learning: Teori, Studi Kasus Dan Implementasi Menggunakan Python*. Riau: UR Press. https://books.google.co.id/books?id=JvBPEAAAQBAJ [Diakses tanggal 20 Oktober 2024]
[2] Ikatan Bankir Indonesia. 2013. Modul Sertifikasi Tingkat I. General Banking. LSPP, Jakarta. https://books.google.co.id/books?id=LKBLDwAAQBAJ [Diakses tanggal 10 Oktober 2024]

[3] Kaufman, L & Rousseeuw, P.J. (2005). Finding Group in Data: An Introduction to Cluster Analysis. New York: John Wiley & Sons, Inc
[4] Mulaab. (2021). *Data Mining: Konsep Dan Aplikasi*. Malang: Media Nusa Creative. https://books.google.co.id/books?id=X1FKEAAAQBAJ [Diakses tanggal 10 Oktober 2024]
[5] Pressman, S.R. (2012). Rekayasa Perangkat Lunak. Yogyakarta: Andi
[6] Prasetyo E, (2014). *Data Mining Mengolah Data Menjadi Informasi Menggunakan Matlab*. Yogyakarta: Andi.
[7] Rahayu PW dkk. (2018). *Buku Ajar Data Mining*. Malang: PT Sonpedia Publishing Indonesia.
[8] Riri Fitri S & Ardiati Utami S. (2021). *Rekayasa Perangkat Lunak Berorientasi Objek Menggunakan Php*. Yogyakarta: Andi. https://books.google.co.id/books?id=x8xEEAAAQBAJ [Diakses tanggal 10 Oktober 2024]
[9] Sudikan SY, Indarti T, and Faizin. (2023). *Metode Penelitian Dan Pengembangan (Research & Development) Dalam Pendidikan Dan Pembelajaran*. Malang: Penerbit Universitas Muhammadiyah Malang. https://books.google.co.id/books?id=ZY3kEAAAQBAJ [Diakses tanggal 12 Oktober 2024]
[10] Sugiyono. (2018). Metode Penelitian Kuantitatif, Kualitatif, dan R & D. Bandung: Alfabeta
[11] Sugiyono. (2019). Metode Penelitian Kuantitatif, Kualitatif, dan R & D. Bandung: Alfabeta
[12] Sugiyono. (2019). Metode Penelitian dan Pengembangan (Research and Development / R&D). Bandung: Alfabeta
[13] Swastika R dkk. (2023). *Implementasi Data Mining (Clastering, Association, Prediction, Estimation, Classification)*. Indramayu: Penerbit Adab CV. Adanu Abimata. https://books.google.co.id/books?id=LsOqEAAAQBAJ [Diakses tanggal 10 Oktober 2024]
[14] Tarigan WJ dkk. (2024). *Kewirausahaan*. Batam: Yayasan Cendikia Mulia Mandiri.