

# ANALISIS SEGMENTASI PELANGGAN BANK BERDASARKAN PENGAMBILAN KREDIT DENGAN MENGUNAKAN METODE *K-MEANS CLUSTERING*

Nuril Huda Ahsina<sup>1</sup>, Fety Fatimah<sup>2</sup>, Fitria Rachmawati<sup>3</sup>

Program Studi Teknik Informatika, Fakultas Teknik dan Sains  
Universitas Ibn Khaldun Bogor

Jl.KH Sholeh Iskandar Km 2 Kota Bogor

hdnuril24@gmail.com<sup>1</sup>, fety.fatimah@uika-bogor.ac.id<sup>2</sup> fitria@uika-bogor.ac.id<sup>3</sup>

## Abstrak

Untuk mengoptimalkan pembiayaan suatu perusahaan, analisis segmentasi pelanggan ditujukan agar penawaran pembiayaan tepat sasaran. Segmentasi pelanggan sering kali menggunakan algoritma *K-means Clustering* sebagai metode untuk membantu dalam pengelompokan menjadi beberapa kluster sehingga mendapatkan visualisasi data yang signifikan hasilnya. Algoritma *K-means Clustering* merupakan salah satu dari teknik data mining yang dapat membagi data dalam suatu himpunan ke dalam beberapa kelompok yang mana kesamaan data dalam suatu kelompok lebih besar dibandingkan kesamaan data tersebut dengan data dalam kelompok lain. Penentuan jumlah *cluster* terbaik bisa dilakukan berdasarkan metode *elbow*. Berdasarkan metode *elbow* 4 cluster adalah *cluster* terbaik dari semua kemungkinan *cluster*. Hasil dari *k-means clustering* dari 1000 data yang ada didapat jumlah dari *cluster 1* dengan jumlah 286 nasabah dengan presentase 28,6% , *cluster 2* dengan jumlah 130 nasabah dengan presentase 13% , *cluster 3* dengan jumlah terbesar yaitu 542 nasabah dengan presentase 54,2% , dan *cluster 4* memiliki jumlah sebaran terendah yaitu 42 nasabah dengan presentase 4,2% dari total keseluruhan data .

Kata kunci : Segmentasi Pelanggan, *K-Means Clustering*, Metode *Elbow*, *Data Mining*

## Abstract

In a step optimizing a company's financing process, analyzing customer segmentation to make financing offers are right on target is what the company needs. Customer segmentation often uses the *K-means Clustering* algorithm method to help the grouping process into several clusters to get the significant data visualization results. *K-means*

clustering algorithm is a data mining technique that can divide the data in a set into several groups where the similarity of data in one group is greater than with data in other groups. Determination of the best number of clusters are possible to do based on the elbow method. Based on the elbow method, 4 cluster is the best cluster of all possible clusters .The results of the *k-means clustering* process from 1000 existing data, obtained from cluster 1 with a total of 286 customers with a percentage of 28.6%, cluster 2 with a total of 130 customers with a percentage of 13%, cluster 3 with the largest number of 542 customers, with a percentage of 54.2%, and cluster 4 having the lowest number of distributions, namely 42 customers with a percentage of 4.2% of the total data

Keywords : Customer Segmentation, *K-Means Clustering*, *Elbow Method*, *Data Mining*

## I.Pendahuluan

### I.1 Latar Belakang

Seiring kebutuhan hidup yang terus meningkat mendorong munculnya perusahaan yang menawarkan soal pembiayaan. Hal ini disebabkan pembiayaan saat ini sudah menjadi salah satu solusi untuk membantu meringankan kebutuhan hidup seseorang. Untuk mengoptimalkan pembiayaan suatu perusahaan, analisis segmentasi pelanggan ditujukan agar penawaran pembiayaan tepat sasaran[1].

Untuk mengetahui perilaku (*behavior*) pelanggan sehingga dapat membantu dalam penerapan strategi yang tepat adalah dengan cara mengelompokan pelanggan berdasarkan kategori variabel yang terkait untuk meningkatkan pendapatan perusahaan. Penggunaan teknik data mining

merupakan salah satu solusi untuk persoalan segmentasi konsumen [2]. Segmentasi pelanggan sering kali menggunakan algoritma *K-means Clustering* sebagai metode untuk membantu dalam pengelompokan menjadi beberapa kluster sehingga mendapatkan visualisasi data yang signifikan hasilnya.

Algoritma *K-means Clustering* merupakan salah satu dari teknik data mining yang dapat membagi data dalam suatu himpunan ke dalam beberapa kelompok yang mana kesamaan data dalam suatu kelompok lebih besar dibandingkan kesamaan data tersebut dengan data dalam kelompok lain [2].

Sesuai dengan uraian latar belakang tersebut, maka dilakukan analisis segmentasi pelanggan bank berdasarkan pengambilan kredit dengan menggunakan metode *k-means clustering*. Untuk mendapatkan hasil visualisasi segmentasi dari nasabah dalam pengambilan kredit.

## I.2 Rumusan Masalah

Berdasarkan uraian latar belakang yang sudah dijelaskan, maka rumusan masalah dari penelitian ini adalah bagaimana menganalisis segmentasi data nasabah bank dalam pengambilan kredit dengan menggunakan algoritma *K-means Clustering* ?.

## I.3 Maksud dan Tujuan Penelitian

Tujuan pada penelitian ini adalah untuk mengetahui hasil visualisasi segmentasi nasabah dalam pengambilan kredit bank dengan menggunakan algoritma *K-means clustering*.

## I.4 Ruang Lingkup Masalah

Berdasarkan latar belakang dan rumusan masalah maka batasan masalah pada penelitian ini yaitu :

1. Batasan masalah dalam pembahasan penelitian ini dibatasi dengan menggunakan konsep segmentasi pelanggan berdasarkan data yang akan diteliti menggunakan algoritma *K-means Clustering* yang dimodelkan dalam bentuk visualisasi.
2. Sumber data yang digunakan pada penelitian ini adalah data pelanggan yang mengambil kredit bank pada *German Credit Data* oleh Prof. Hofmann.
3. Metode *Elbow* digunakan untuk membantu mencari masukan nilai *k* ketika proses *clustering K-Means*.

## II. Kajian Literatur

### II.1 Segmentasi Pelanggan

Segmentasi adalah cara untuk berkomunikasi lebih banyak dengan pelanggan. Tujuan segmentasi adalah untuk menyesuaikan produk, jasa, dan pesan pemasaran untuk setiap segmen [3]. Proses segmentasi menempatkan pelanggan berdasarkan dengan karakteristik kelompok pelanggan yang serupa. Segmentasi pelanggan merupakan langkah persiapan untuk mengklasifikasikan setiap pelanggan menurut kelompok pelanggan yang sudah ditentukan [4].

### II.2 Data Mining

Data mining itu sendiri adalah proses pencarian pola-pola yang tersembunyi (*hidden pattern*) berupa pengetahuan (*knowledge*) yang tidak diketahui sebelumnya dari suatu sekumpulan data yang mana data tersebut dapat berada didalam *database*, *warehouse data*, atau media penyimpanan informasi yang lain [5]. Hal penting yang terkait di dalam *data mining* adalah:

1. Data mining merupakan suatu proses otomatis terhadap data yang sudah ada.
2. Data yang akan diproses berupa data yang sangat besar.
3. Tujuan data mining adalah mendapatkan hubungan atau pola yang mungkin memberikan indikasi yang bermanfaat.

Sebagai suatu rangkaian proses, *data mining* dapat dibagi menjadi beberapa tahap. Tahap-tahap tersebut bersifat interaktif, pemakai terlibat langsung atau dengan perantaraan *knowledge base*.

### II.3 K-means clustering

Algoritma *k-means* digunakan untuk mengevaluasi kualitas partisi sehingga objek dalam satu grup memiliki kemiripan satu sama lain tetapi tidak memiliki kemiripan dengan objek pada grup lain. Algoritma *k-means* adalah fungsi obyektif yang bertujuan untuk mencapai kesamaan intra kelompok yang tinggi dan kesamaan inter kelompok yang rendah.

Pengelompokan menggunakan teknik partisi berbasis *centroid* dimana *centroid* dari setiap kelompok merepresentasikan ciri-ciri dari kelompok tersebut. Secara konseptual, *centroid* merupakan titik pusatnya. *Centroid* dapat didefinisikan dengan banyak cara, seperti dengan menghitung *mean* atau *medoid* dari objek atau titik yang dimasukkan ke

dalam kelompok. Perbedaan antar objek kemudian diukur dengan menghitung jarak *euclidean* antar dua titik. Kualitas dari suatu kelompok dapat diukur dengan variasi *within-cluster* yang merupakan jumlah kuadrat kesalahan antara semua obyek dalam suatu kelompok dengan *centroid*-nya [5].

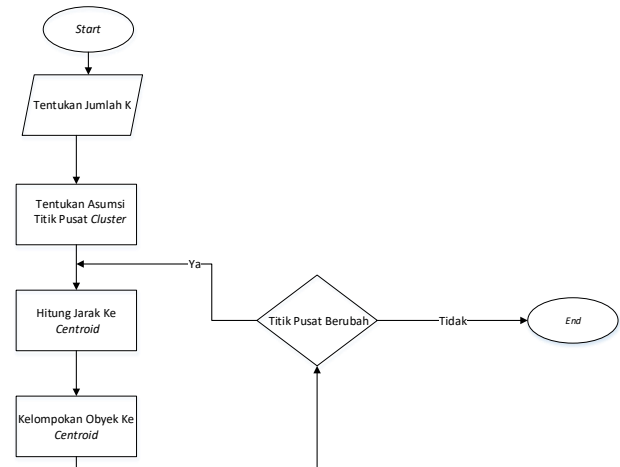
#### II.4 Algoritma K-Means

*K-Means* merupakan metode klasterisasi yang sering digunakan diberbagai bidang karena penggunaannya sederhana, mudah untuk diimplementasikan, mampu untuk mengklaster data yang besar [5]. Algoritma *K-Means* merupakan metode berbasis jarak yang membagi data kedalam sejumlah *cluster* dan dalam setiap tahapan tertentu setiap objek harus masuk dalam kelompok, pada tahap selanjutnya objek dapat berpindah ke kelompok lain. Maka dari itu algoritma *K-Means* mempunyai aturan yaitu ada jumlah *cluster* yang akan diinputkan dan hanya dapat memiliki atribut yang bertipe numerik[6].

Algoritma K-means 3 komponen didalamnya yaitu :

1. Jumlah *cluster k* Dalam metode ini jumlah *k* harus ditentukan dulu, setelah jumlah *k* didapatkan dengan melalui pendekatan metode hirarki dapat melakukan pengambilan *cluster* awal. Aturan khusus dalam menentukan jumlah *cluster k* bahkan tidak ada, namun ada juga jumlah *cluster* yang diinginkan sesuai dengan kebutuhan subjektif seseorang.
2. *Cluster* Awal Ada banyak pendapat saat melakukan pengambilan *cluster* awal untuk metode *K-Means* misalnya pemilihan terhadap interval dari jumlah setiap observasi, melalui pendekatan salah satu metode hirarki dan ada juga dengan melalui pemilihan *cluster* secara acak dari sekumpulan observasi. Maka dengan adanya beberapa cara pengambilan *cluster* awal tersebut dapat memungkinkan solusi terbaik yang dihasilkan.
3. Ukuran Jarak Pada tahapan ini ukuran jarak juga penting dalam menempatkan observasi ke dalam *cluster* berdasarkan nilai *centroid* terdekat. *Euclidian Distance* adalah jarak yang digunakan untuk mengukur jarak dalam metode *K-Means*. Berikut ini adalah rumus *Euclidian Distance*:

Tahapan algoritma *K-means*, menyatakan :



**Gambar 2. 1 Tahapan Algoritma K-Means**

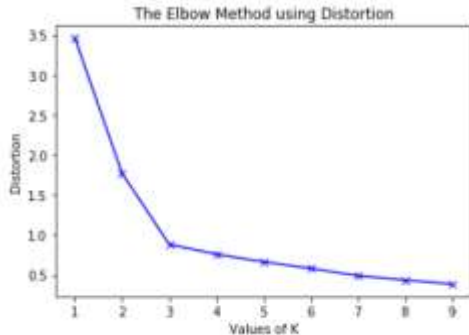
1. Memilih objek *k* secara acak, setelah mendapatkan objek *k* tersebut data akan diproses sebagai, mean pada *cluster*.
2. Setiap objek akan dimasukkan kedalam *cluster* yang mempunyai kemiripan terhadap *cluster*. Tingkat kemiripan dapat ditentukan dengan mencari jarak objek terhadap mean atau *centroid cluster* tersebut.
3. Lakukan perhitungan nilai *centroid* yang baru pada setiap *cluster*.
4. Proses perhitungan nilai *centroid* tersebut dilakukan berulang-ulang hingga didapati anggotaapada kelompok *cluster* tersebut tidak berubah.

#### II.5 Metode Elbow

Metode *Elbow* merupakan ide atau gagasan yang digunakan untuk menghasilkan informasi dalam memilih nilai *cluster* terbaik dengan cara melihat presentasi hasil perbandingan antara jumlah *cluster* yang membentuk siku pada suatu titik[7]. Selain itu presentasi perhitungan yang dihasilkan menjadi perbandingan antara jumlah *cluster* yang ditambahkan. Hasil presentasi yang berbeda dari setiap nilai *cluster* pertama dengan nilai *cluster* kedua memberikan sudut dalam grafik atau nilai *cluster* tersebut yang terbaik. Untuk mendapatkan perbandingan nya adalah dengan menghitung SSE (sum of Square Error) dari masing-masing nilai *cluster*. Karena semakin besar jumlah *cluster k* maka nilai SSE semakin kecil[8].

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

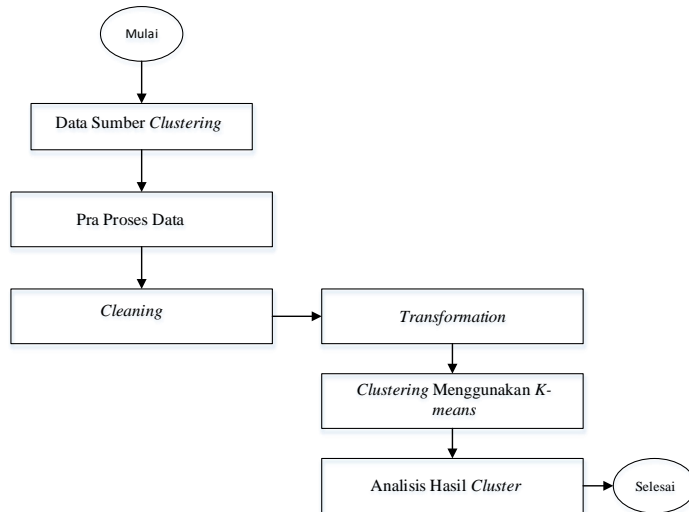
Setelah dilihat akan ada beberapa nilai K yang mengalami penurunan paling besar dan selanjutnya hasil dari nilai K akan turun secara perlahan-perlahan sampai hasil dari nilai K tersebut stabil. Berikut contoh gambar dari metode *elbow* [9].



Gambar 2. 2 Metode Elbow

### III.METODE PENELITIAN

Metode penelitian adalah prosedur atau langkah-langkah dalam mendapatkan pengetahuan ilmiah, secara sistematis untuk menyusun ilmu pengetahuan [10].



Gambar 3. 1 Bagan Proses Analisis

Adapun tahapan yang akan dilakukan untuk menganalisis data seperti alur pada gambar 3.1 yaitu:

1. Data Sumber *Clustering*

Data Sumber *clustering* terdapat pada Lampiran 1. Tabel Data Sumber *Clustering*.

2. Pra Proses Data

Pada bagian ini data telah dipersiapkan agar memudahkan dalam proses *mining*. Proses pra proses data akan mencakup tiga hal mendasar, yaitu:

a. *Cleaning*

Pada proses ini data akan dibersihkan mencakup kelengkapan data, peghapusan data redundand dan menghilangkan noise.

b. *Transformation*

Proses *transformation* ini yaitu data akan di seleksi dan di transformasikan agar nantinya data siap untuk di *clustering*.

3. Menentukan *cluster* terbaik menggunakan metode *elbow*.

4. *Clustering* menggunakan algoritma *K-means*

Setelah proses sebelumnya dilakukan yaitu pra poses data, tahapan berikutnya data akan di *cluster* menggunakan algoritma *K-Means*.

a. Dalam pemilihan pusat *cluster* ada berbagai cara namun kali ini dilakukan dengan cara pegambilan *cluster* awal secara *random*.

b. Setelah pusat *cluster* awal ditentukan, tempatkan setiap data dengan cluster terdekat. Kedekatan antara data ke suatu *cluster* diperlukan perhitungan jarak setiap data terhadap pusat *cluster*. Jarak terdekat itulah yang nantinya akan menjadi penentu suatu data masuk kedalam *cluster* mana. Dalam perhitungan jarak penelitian ini menggunakan rumus *Euclidean*. *Euclidean* sendiri merupakan suatu metode pencarian kedekatan nilai jarak dari 2 buah variabel.

$$D_{ik} = \sqrt{\sum_j^M = 1(x_{ij} - c_{kj})^2}$$

Keterangan :

$D_{ik}$  = titik, dokumen/jarak data ke-i

$M$  = jumlah variabel

$X_{ij}$  = data yang akan di *cluster* kan

$C_{kj}$  = pusat dari *cluster*

c. Langkah selanjutnya setelah melakukan perhitungan jarak, kemudian hitung *centroid* baru dengan menggunakan nilai mean dari semua data dalam sel *cluster* tertentu.

Adapun rumus yang dapat digunakan untuk menentukan *centroid* baru yaitu:

$$V_{ij} = \frac{1}{N_i} \sum_{k=0}^{N_i} X_{kj}$$

Berikut merupakan rumus penentu *centroid* baru :

Keterangan :

$V_{ij}$  = *centroid*/rata-rata *cluster* ke-i untuk variabel ke-j

$N_i$  = jumlah data yang menjadi anggota *cluster* ke-i

$I, k$  = indeks dari *cluster*

$J$  = indeks dari variabel

$X_{kj}$  = nilai data ke-k didalam *cluster* tersebut untuk variabel ke-j

- d. Bila perhitungan kelompok data hasil mengalami perubahan maka lakukan pengulangan setiap data dengan *centroid* yang baru dengan cara seperti langkah b hingga c jika didapati perhitungan kelompok data hasil tidak mengalami perubahan, maka proses dalam pengclusteran dianggap selesai.
  - e. Pada tahapan selanjutnya dilakukan perhitungan nilai *SSE* (*Sum Square Error*). Jumlah *cluster* dan data yang sudah dikelompokkan dalam *cluster* dapat mempengaruhi nilai *SSE*. Dalam penilaian *SSE* dimana semakin kecil nilai *SSE* hasil clustering semakin bagus .
5. Analisis hasil *cluster* Dalam melakukan analisis, hasil dapat dilakukan setelah diperoleh pada proses *clustering*.

## IV.HASIL DAN PEMBAHASAN

### IV.1 Pengumpulan Data

Bahan atau data yang digunakan pada

Kategori	keterangan
<i>Job</i>	Jenis pekerjaan nasabah/pelanggan yang ingin mengambil kredit
<i>Age</i>	Umur nasabah/pelanggan
<i>Sex</i>	Jenis kelamin nasabah/pelanggan
<i>Housing</i>	Jenis hunian nasabah/pelanggan
<i>Purpose</i>	Tujuan nasabah/pelanggan untuk pengambilan kredit
<i>Credit Amount</i>	Jumlah pengambilan kredit dalam mata uang Jerman <i>Deutsche Mark (DM)</i> sebelum sekarang euro
<i>Duration</i>	Lama nya durasi pengambilan kredit dalam hitungan bulan

penelitian ini adalah data nasabah (pelanggan bank),

dan data transaksi nasabah pada saat melakukan pengambilan kredit pada data *German Credit Data* yang di *publish* oleh Prof.Hofmann. terdapat 9 kategori yang bisa dilihat pada tabel 4.1 :

**Tabel 4. 1 kategori data**

### IV.2 Transformasi Data

Dari semua kategori data yang sudah ada diberi inialisasi untuk mempermudah pengelompokan ke dalam *cluster*, kemudian inialisasi dari bentuk *categorical* ke bentuk *numerical*. Berikut hasil inialisasi dari beberapa data *categorical*.

**Tabel 4. 2 Inialisasi Data**

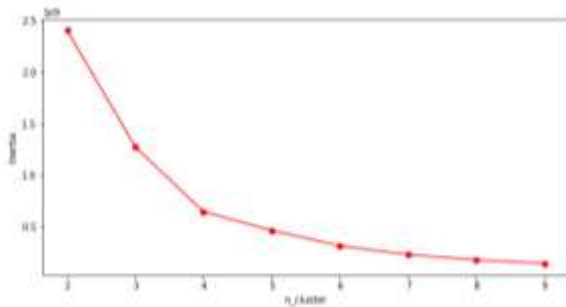
1. Job	
Keterangan	Kategori
<i>Highly Qualified Employee</i>	0
<i>Skilled Employee</i>	1
<i>Unemployed- Non Resident</i>	2
<i>Unskilled - Resident</i>	3
3. Sex	
Keterangan	Kategori
Laki Laki	0
Perempuan	1
4. Housing	
Keterangan	Kategori
<i>Free</i>	0
<i>Own</i>	1
<i>Rent</i>	2
4. Purpose	
Keterangan	Kategori
<i>Business</i>	0
<i>Car</i>	1
<i>Domestic Appliances</i>	2
<i>Education</i>	3
<i>Furniture/Equipment</i>	4
<i>Radio/TV</i>	5
<i>Repairs</i>	6
<i>Vacation</i>	7

### IV.3 Pemodelan K-means Clustering

Metode yang digunakan untuk pemodelan clustering dalam penelitian ini menggunakan k-means clustering dengan jumlah cluster yang ditentukan oleh metode elbow. Aplikasi yang digunakan untuk mengolah data nasabah adalah aplikasi jupyter notebook. Berikut dibawah ini penjelasan tahapan pemodelan tersebut.

1. Metode *elbow*

Dari hasil menggunakan metode *elbow* yang bisa dilihat pada gambar 4.1 bisa terlihat bahwa cluster terbaik ada di cluster 4 berdasarkan hasil penurunan paling besar dan penurunan selanjutnya yang stabil dengan nilai penurunan yang sama . selanjutnya pembuatan *cluster* dengan menggunakan metode *k-means clustering* dengan 4 cluster seperti gambar 4. 1



Gambar 4. 1 Cluster Terbaik

2. Hasil K-Means Clustering

Ada 4 kriteria pelanggan yang bisa dilihat dari gambar 4. pada titik pusat terakhir, jika dibandingkan dengan data rata-rata (mean) pada gambar 4.3.

	Job	Sex	Housing	Purpose	Age	Credit amount	Duration
0	1.947735	0.707317	1.087108	2.770035	34.648084	3576.184669	23.540070
1	2.192308	0.776923	0.953848	2.184615	36.800000	7127.523077	33.346154
2	1.776340	0.654344	1.110906	3.133087	35.634011	1465.476895	15.011091
3	2.357143	0.761905	0.809524	2.476190	36.666667	12511.714286	40.261905

Gambar 4. 2 Titik Pusat Terakhir

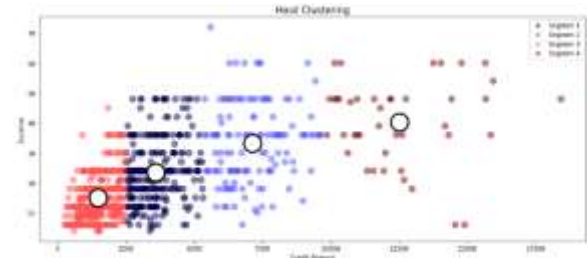
	Job	Sex	Housing	Purpose	Age	Credit amount	Duration
credit	1000.000000	1000.000000	1000.000000	1000.000000	1000.000000	1000.000000	1000.000000
mean	1.904000	0.690000	1.071000	2.378000	35.548000	3271.258000	20.893000
std	0.663614	0.462725	0.631264	1.978138	11.375488	2822.738876	12.688814
min	0.000000	0.000000	0.000000	0.000000	16.000000	250.000000	4.000000
25%	2.000000	0.000000	1.000000	1.000000	27.000000	1365.000000	12.000000
50%	2.000000	1.000000	1.000000	3.000000	33.000000	2319.000000	18.000000
75%	2.000000	1.000000	1.000000	5.000000	42.000000	3972.250000	24.000000
max	3.000000	1.000000	2.000000	7.000000	75.000000	18424.000000	72.000000

Gambar 4. 3 Rata-Rata Data

Karena tidak ada perbedaan yang signifikan dalam kategori *job*, *age*, *sex*, *housing*, dan *purpose* maka perbandingan dilakukan pada kategori *credit amount* dan *duration*.

Hasil dari *k-means clustering* dari 1000 data yang ada didapat jumlah dari *cluster 1* dengan jumlah 286 nasabah dengan presentase 28,6% , *cluster 2* dengan jumlah 130 nasabah dengan presentase 13%, *cluster 3* dengan jumlah terbesar yaitu 542 nasabah dengan presentase 54,2% , dan *cluster 4* memiliki jumlah sebaran terendah yaitu 42 nasabah dengan presentase 4,2% dari total keseluruhan data .

Plot hasil sebaran data *cluster* dari *credit amount* dan *duration* dengan *centroid* akhir dari *cluster*. Sesuai dengan perbandingan data *centroid* akhir dan rata-rata dari keseluruhan data *credit amount* dan *duration* menjadi 2 katagori pembanding untuk mendapatkan hasil dari pelanggan potensial untuk pengambilan strategi perusahaan dalam peningkatan kinerja atau peningkatan nasabah baru.



Gambar 4. 4 Sebaran Cluster Dengan Centroid Cluster

Proses segmentasi menempatkan pelanggan berdasarkan dengan karakteristik kelompok pelanggan yang serupa. Segmentasi pelanggan merupakan langkah persiapan untuk mengklasifikasikan setiap pelanggan menurut kelompok pelanggan yang sudah ditentukan. Maka dari itu segmentasi pelanggan ini dibedakan berdasarkan 4 katagori yang telah disebutkan diatas yaitu :

1. Most Valuble Customer.  
*Cluster* yang termasuk kriteria ini yaitu *cluster 3* karena nasabah pada *cluster* ini yang paling banyak yang pengambilan kredit meskipun dengan jumlah pengambilan kredit sangat kecil dengan durasi paling singkat dari semua kriteria nasabah.
2. Most Growable Customer.  
*Cluster* yang termasuk kriteria ini yaitu *cluster 4* karena nasabah pada *cluster* ini melakukan jumlah pengambilan kredit tinggi walaupun dengan durasi terpanjang.
3. Below Zeros.

*Cluster* yang termasuk kriteria ini yaitu *cluster 2* karena nasabah pada *cluster* ini termasuk sedang dalam melakukan pengambilan kredit dengan durasi yang sedang dari semua kriteria nasabah.

4. Migrators.

*Cluster* yang termasuk kriteria ini yaitu *cluster 1* karena nasabah pada *cluster* ini termasuk kecil dalam melakukan pengambilan kredit dengan durasi singkat dari semua kriteria nasabah, namun memiliki jumlah sebaran nasabah cukup banyak.

Hasilnya perusahaan sebaiknya lebih kredit kepada mempromosikan kepada pelanggan yang pada dengan umur 36 dengan pekerjaan *unemployee*, segmen *male* mempunyai rumah dengan status *own* atau kepemilikan pribadi dan dengan tujuan pengambilan kredit *radio/tv*. Sesuai dengan jumlah sebaran segmen 3 yang memiliki presentase sebaran 54,2 % .

## V. Penutup

### V.1 Kesimpulan

Berdasarkan hasil penelitian yang dilakukan, dapat diambil beberapa kesimpulan diantaranya :

1. Penentuan jumlah *cluster* terbaik bisa dilakukan berdasarkan metode *elbow*. Setelah dilakukan segmentasi data nasabah dengan menggunakan metode *k-means clustering*, dengan berdasarkan metode *elbow* 4 *cluster* adalah *cluster* terbaik dari semua kemungkinan *cluster* yang terdiri dari *cluster 1* dengan jumlah 286 nasabah dengan presentase 28,6% , *cluster 2* dengan jumlah 130 nasabah dengan presentase 13%, *cluster 3* dengan jumlah terbesar yaitu 542 nasabah dengan presentase 54,2% , dan *cluster 4* memiliki jumlah sebaran terendah yaitu 42 nasabah dengan presentase 4,2%.
2. Hasil dari *clustering* dapat dipengaruhi dari nilai centroid awal dan jumlah data yang dipakai, perbedaan data pusat *centroid* awal yang digunakan dapat mempengaruhi hasil dari *centroid* akhir.

### V.2 Saran

Berdasarkan penelitian yang telah dilakukan maka peneliti memberikan saran untuk penelitian selanjutnya yaitu :

1. Untuk analisis selanjutnya diharapkan untuk menambah katagori lain seperti rekap data dalam

pembayaran kredit dan juga penentuan *cluster* yang diharapkan.

2. Hasil dari proses analisis *k-means clustering* ini dapat digunakan dalam mengambil keputusan lebih lanjut dalam penetapan pelanggan potensial kedepannya.
3. Penelitian ini juga dapat dikembangkan dengan menggunakan *tools* yang lain nya selain dengan menggunakan *jupyter notebook* untuk membantu mencari hasil yang yang lebih optimal lagi.

## Daftar Pustaka

- [1] D. Meriska and Noviyanti, “ANALISIS SEGMENTASI NASABAH MENGGUNAKAN ALGORITMA K-MEANS CLUSTERING ( Studi Kasus di PT . Buana Sejahtera Multidana Cabang Cikampek ),” no. November 2017, 2017.
- [2] B. E. Adiana, I. Soesanti, and A. E. Permanasari, “Analisis Segmentasi Pelanggan Menggunakan Kombinasi Rfm Model Dan Teknik Clustering,” *J. Terap. Teknol. Inf.*, vol. 2, no. 1, pp. 23–32, 2018, doi: 10.21460/jutei.2018.21.76.
- [3] M. I. Istiana, “Segmentasi Pelanggan menggunakan Algoritma K-Means Sebagai Dasar Strategi Pemasaran pada LAROIBA Seluler,” vol. 1, pp. 3–4, 2013.
- [4] A. V. Angelic, “Segmentasi Pelanggan Menggunakan Clustering K-Means Dan Model Rfm ( Studi Kasus : Pt . Bina Adidaya Surabaya ) Customer Segmentation Using K-Means Clustering and Rfm Model ( Case Study: Pt . Bina Adidaya Surabaya Segmentasi Pelanggan Menggunakan Clusteri,” p. 163, 2017.
- [5] M. A. Akbar, F. Fatimah, and J. Jaenudin, “Penerapan Data Mining Untuk Pengelompokan Posisi Pemain Sepak Bola Menggunakan Algoritma K-Means Clustering,” *Semnati*, vol. x, No.x, pp. 278–282, 2019.
- [6] D. P. Hidayatullah<sup>1</sup>, R. I. Rokhmawati, and ..., “Analisis Pemetaan Pelanggan Potensial Menggunakan Algoritma K-Means dan LRFM Model Untuk Mendukung Strategi Pengelolaan Pelanggan (Studi Pada ...,” ... *Inf. dan Ilmu ...*, vol. 2, no. 8, 2018, [Online].

Available: <http://j-ptiik.ub.ac.id/index.php/j-ptiik/article/view/1769>.

- [7] D. K. Pane, “Implementasi Data Mining Pada Penjualan Produk Elektronik Dengan Algoritma Apriori ( Studi Kasus : Kreditplus ),” *Pelita Inform. Budi Darma*, vol. volume : I, pp. 25–29, 2013.
- [8] A. D. Savitri, F. A. Bachtiar, and N. Y. Setiawan, “Segmentasi Pelanggan Menggunakan Metode K-Means Clustering Berdasarkan Model RFM Pada Klinik Kecantikan (Studi Kasus: Belle Crown Malang),” *J. Pengemb. Teknol. Inf. dan Ilmu Komput. Univ. Brawijaya*, vol. 2, no. 9, pp. 2957–2966, 2018.
- [9] N. Putu, E. Merliana, and A. J. Santoso, “Analisa Penentuan Jumlah Cluster Terbaik pada Metode K-Means,” pp. 978–979.
- [10] Ms. Prof. Dr. Suryana, “Metodologi Penelitian: Metodologi Penelitian Model Praktis Penelitian Kuantitatif dan Kualitatif,” *Univ. Pendidik. Indones.*, pp. 1–243, 2012, doi: 10.1007/s13398-014-0173-7.2.