

Penerapan Data Mining *Clustering* Dalam Menentukan Tingkat Pembelian Kredit Tertinggi Algoritma *K-Means* dan *K-Medoids*

1st Dhitta Mari'ane Yusup
Universitas Buana Perjuangan
Karawang, Indonesia
If18.dhittamari'aneyusup@mhs.ubpkarawang.ac.id

2nd Amril Mutoi Siregar
Universitas Buana Perjuangan
Karawang, Indonesia
Amrilmutoi@ubpkarawang.ac.id

3rd Santi Arum Puspita Lestari
Universitas Buana Perjuangan
Karawang, Indonesia
Santi.arum@ubpkarawang.ac.id

Abstract— Pembelian kendaraan bermotor makin meningkat tiap tahunnya dikarenakan kebutuhan masyarakat akan transportasi, berniaga, sekolah dan menunjang pekerjaan. Namun bagi sebagian orang membeli kendaraan secara tunai adalah suatu hal yang sulit maka dari itu sebagian orang memilih membeli kendaraan secara kredit agar bisa memakainya terlebih dahulu dan membayarnya di kemudian hari. Pembelian kredit memiliki beberapa ketentuan dan perjanjian yang tertuang di akta jaminan fidusia dan tentunya harus disepakati oleh kedua belah pihak. Jumlah pembelian kredit di beberapa pasti berbeda. Untuk mengetahui daerah mana saja dengan pembelian kredit tertinggi diperlukan data mining untuk mengelompokan daerah pembelian kredit tinggi, daerah pembelian kredit sedang dan daerah pembelian kredit rendah. Data terdiri dari 7 daerah dan 1.770 akta yang diolah menggunakan algoritma *k-means* dan *k-medoids* sebagai pembanding. Perhitungan akan dilakukan menggunakan dua cara yaitu perhitungan manual dan perhitungan menggunakan python. Evaluasi menggunakan metode silhouette coefficient dan di dapatkan hasil tingkat akurasi *k-means* lebih besar dari algoritma *k-medoids*.

Kata kunci — *Clustering, Data Mining, Fidusia, K-Means, K-Medoids, Pembelian Kredit.*

I. PENDAHULUAN

Pertumbuhan penduduk yang bergerak sangat cepat dapat mendorong pertumbuhan ekonomi, yang artinya semakin banyak penduduk maka akan meningkatkan tingkat spesialisasi dalam perekonomian [1]. Permasalahan yang akan muncul atas pertumbuhan ekonomi adalah masyarakat akan sulit memenuhi kebutuhan hidupnya. salah satu kebutuhan masyarakat saat ini adalah kendaraan pribadi untuk berniaga, sekolah atau menunjang pekerjaan. Kendaraan pribadi juga digunakan untuk mewujudkan eksistensi seseorang dalam lingkungannya [2]. Tidak dapat dipungkiri saat ini kebutuhan masyarakat akan transportasi meningkat dari tahun ke tahun. Berdasarkan data Badan Pusat Statistik jumlah pembelian kendaraan bermotor pada tahun 2017 adalah 118.922.708 dan pada tahun 2020 menjadi 136.137.451, yang artinya ada penambahan sekitar 12%. [3] Sebagian masyarakat membeli kendaraan pribadi menggunakan sistem kredit dikarenakan pembayaran ini memudahkan masyarakat karena pembayaran dapat dilakukan secara berangsur atau bayar di kemudian hari. Pembelian kredit ini biasanya dilakukan oleh perusahaan pembiayaan konsumen dan pembeli dengan perjanjian yang telah ditentukan. Perusahaan pembiayaan konsumen merupakan kegiatan pembiayaan untuk pembelian barang atau benda yang akan digunakan konsumen dengan sistem pembayaran angsuran berjangka dan pengelolaan risiko berdasarkan perjanjian yang tertuang dalam akta jaminan fidusia atas barang atau benda tersebut [1].

Jaminan Fidusia adalah hak atas jaminan atas benda bergerak baik berwujud maupun tidak berwujud dengan benda tidak bergerak maupun bergerak khususnya bangunan yang tidak dapat dibebani hak tanggungannya [4]. Fidusia sendiri berasal dari kata "fides" yang memiliki arti kepercayaan [5]. Fidusia adalah jaminan yang berdasarkan kepercayaan antara perusahaan pembiayaan konsumen dengan penerima fidusia [6].

Maka berdasarkan masalah diatas, penelitian ini akan mengelompokan tingkat pembelian kredit berdasarkan jaminan fidusia. Penelitian ini menggunakan dua algoritma yaitu yaitu algoritma *k-means* dan *k-medoids*. Hal ini dilakukan untuk membandingkan algoritma mana yang lebih cocok atau tingkat akurasi lebih tinggi dalam *clustering*. Data yang digunakan dalam penelitian ini meliputi 7 daerah dengan 3 *cluster*, C1 yaitu daerah pembelian kredit tinggi, C2 yaitu daerah pembelian kredit sedang dan C3 yaitu daerah pembelian kredit rendah.

II. METODE PENELITIAN

A. Pembelian Kredit

Pembelian kredit adalah pembelian barang yang dilakukan seseorang dengan pembayarannya dilakukan secara bertahap kepada penjual [7]. Pembelian kredit memiliki bunga yang harus dibayar oleh debitur dan pembayarannya diatur dengan jangka waktu yang telah disepakati kedua belah pihak. Pembelian kredit meliputi pembelian motor, mobil, dan rumah [8]. Kredit memiliki beberapa unsur yaitu kreditur atau pihak yang memberikan pinjaman, kepercayaan kedua belah pihak, perjanjian antara kreditur dan debitur, resiko kerugian atas penyaluran kredit, tenggang waktu dalam membayar pinjaman, bunga atas barang yang telah dibeli sesuai dengan perjanjian yang telah disepakati [8].

B. Data Mining

Data Mining Merupakan teknik mengambil informasi berharga yang tersembunyi pada suatu database yang sangat besar sampai ditemukan suatu pola menarik yang sebelumnya tidak diketahui [9]. Tujuan data mining adalah mendapatkan pola yang mungkin memberikan indikasi yang bermanfaat. Pengolahan data mining memiliki beberapa tahapan yaitu seleksi data, pre-processing, tranformasi data, perhitungan data mining, interpretasi dan evaluasi [10].

C. Algoritma K-Means

Algoritma *k-means* merupakan metode yang mengelompokan data dalam bentuk satu atau lebih cluster. Data-data yang memiliki karakteristik berbeda dikelompokan dengan *cluster* lain sehingga data yang berada dalam satu cluster memiliki tingkat variasi [11]. Langkah – langkah perhitungan algoritma *k-means* sebagai berikut [12] :

1. Langkah pertama menentukan K sebagai jumlah cluster yang akan dibentuk. Selanjutnya memilih cluster secara acak dan mengelompokan data yang lainnya ke dalam cluster tersebut. Rumus untuk menghitung centroid cluster ke-I berikutnya yaitu ;

$$v = \frac{\sum_{i=1}^n x_i}{n}; i = 1,2,3,..n \quad (1)$$

Keterangan :

V : *centroid* pada *cluster*

x_i : objek ke-i

n : jumlah objek yang menjadi anggota *cluster*.

2. Menghitung jarak setiap objek ke masing-masing centroid dari masing-masing cluster. Untuk menghitung jarak antara objek dengan centroid menggunakan rumus euclidean, sebagai berikut ;

$$d(x, y) = \left\| x - y \right\| = \sqrt{\sum_{i=1}^n (x_i - y_i)^2}; i = 1,2,3,...,n \quad (2)$$

Keterangan :

x_i : data x ke-i

y_i : data y ke-1

n : jumlah data

3. Menempatkan masing-masing objek ke dalam *centroid* paling dekat. Kemudian lakukan iterasi pertama dan tentukan posisi *centroid* baru dengan menggunakan persamaan.
4. Jika hasil data yang didapatkan masih berpindah *cluster* dan menyebabkan perubahan nilai *centroid cluster*, maka ulangi perhitungan langkah 3 sampai tidak ditemukan data berpindah lagi.

D. Algoritma K-Medoids

Algoritma *k-medoids* merupakan algoritma yang mempunyai tujuan membagi data menjadi beberapa cluster (riyanto). Algoritma *k-medoids* menggunakan objek data sebagai perwakilan [13]. Langkah-langkah perhitungan algoritma *k-medoids* clustering sebagai berikut [14] ;

1. Inisialisasi pusat cluster sebanyak data sampel lalu menghitung data menggunakan persamaan euclidean dengan persamaan sebagai berikut ;

$$d_{ij} = \sqrt{\sum_{a=1}^p (x_{ia} - x_{ja})^2} = \sqrt{(x_i - x_j)'(x_i - x_j)} \quad (3)$$

Keterangan :

i : objek 1, 2, 3, n

j : medoids

p : banyaknya variable

v : matriks varian kovarian

2. Memilih secara acak objek yang ada pada setiap cluster dengan kandidat medoids baru. Kemudian lakukan iterasi kedua.
3. Menghitung total simpangan (S) dengan rumus total *distance* baru – total *distance* lama.
4. Ulangi langkah kedua sampai ketiga, sehingga tidak terjadi perubahan *medoids*, maka didapatkan cluster beserta anggota cluster masing-masing.

E. Evaluasi

Evaluasi *silhouette coefficient* digunakan untuk menentukan tingkat akurasi clustering algoritma *k-means* dan *k-medoids*. Evaluasi *silhouette coefficient* merupakan gabungan dari dua metode yaitu metode *cohesion* untuk mengukur seberapa dekat relasi antara objek data sebuah cluster dan *separation* digunakan untuk mengukur jarak *cluster* terpisah dengan *cluster* lain [15]. Rumus *silhouette coefficient* adalah sebagai berikut [16] ;

$$\text{sil}(c) = \text{sil}(k) \frac{1}{|k|} \sum_{i=1}^k (\text{sil}(c_i)) \quad (4)$$

Keterangan :

Sil(k) : nilai *silhouette* seluruh *cluster*

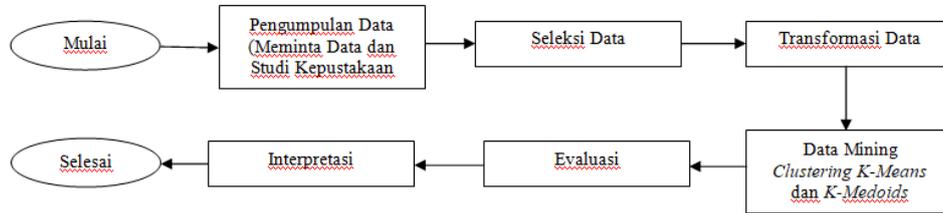
|k| : jumlah *cluster* k

Sil(c_i) : rata-rata nilai *silhouette*

Jika hasil dari *silhouette coefficient* mendekati nilai 1 maka artinya pengelompokan data dalam satu *cluster* adalah baik. Namun jika hasil *silhouette coefficient* bernilai -1 maka artinya pengelompokan data dalam satu *cluster* adalah buruk [16].

F. Prosedur Penelitian

Langkah – langkah dalam penelitian ini adalah mengumpulkan data, melakukan seleksi data, kemudian tahap data mining clustering menggunakan algoritma *k-means* dan *k-medoids*, lalu hasil tersebut akan di evaluasi untuk mengetahui tingkat akurasi. Alur prosedur penelitian sebagai berikut :



Gambar 1 Prosedur Penelitian

III. HASIL DAN PEMBAHASAN

A. Pengumpulan Data

Data yang digunakan untuk penelitian ini adalah data akta fidusia dari Januari – Desember pada tahun 2021. Data yang diproses adalah data pembuatan akta fidusia yang dibuat untuk konsumen dari PT Indomobil Finance Indonesia berupa tabel - tabel dalam Microsoft Excel yang didapat dari *database* kantor Notaris Endang Noviyanti., S.H., M.Kn. Data yang diperoleh berjumlah 2.055 data.

B. Seleksi Data

Data asli yang telah di peroleh kemudian melalui tahap seleksi data, karena data yang dibutuhkan hanya data akta jamina fidusia maka didapatkan hasil seleksi data terdapat 1.770 data. Data asli memiliki 7 atribut yang terdiri dari nomor urut, nomor bulanan, tanggal akta, sifat akta, nama penghadap dan kuasa (nama kepala cabang, kota cabang, nama penghadap/pembeli), nama kepala cabang, dan nama daerah. Lalu dilakukanlah seleksi data dan hasilnya adalah ada 5 atribut yang dibutuhkan dalam proses perhitungan ini yaitu nomor urut, tanggal akta, nama penghadap/pembeli, cabang, jenis kendaraan. Hasil seleksi data sebagai berikut :

Tabel 1 Seleksi Data

Nomor	Tanggal	Nama Penerima Fidusia	Cabang	Jenis
1	04-01-2021	Tn. Islahuddin Ny. Mahmudah Kurniati, S.Psi	Bekasi	R2
2	04-01-2021	Tn. Suharto Ny. Liana Samawati	Bekasi	R2
3	04-01-2021	Tn. Sultan Ny. Reny Budiarty. Y	Bekasi	R2
4	04-01-2021	Tn. Sahrul	Bekasi	R2
5	04-01-2021	Tn. Davi Sriyono Ny. Ruminah	Bekasi	R2
6	05-01-2021	Ny. Siti Nurjanah	Matraman	R4
.....
.....
227	31-12-2021	Tn. Ridwan Arifin Ny. Muthia Septianita	Bekasi	R2
228	31-12-2021	Tn. Arvan Maulana Ny. Fenny Fitriyanti	Bekasi	R2
229	31-12-2021	Tn. Amiko	Bekasi	R2

C. Tranformasi Data

Tahap transformasi data ini dilakukan secara manual menggunakan Microsoft excel. Hasil seleksi data akta Fidusia dari Januari sampai Desember dimasukan ke dalam tabel excel yang sama. Kemudian untuk nama cabang/nama daerah diganti menjadi numerik. Hasil transformasi data sebagai berikut :

Transformasi Data												
Kota	Januari	Februari	Maret	April	Mei	Juni	Juli	Agustus	September	Oktober	November	Desember
Karawang	2	4	6	4	8	6	7	17	15	18	20	24
Cikampek	0	1	2	2	1	0	2	6	4	2	3	1
Matraman	41	51	57	58	72	90	68	62	112	97	70	110
Kelapa Gading	9	22	26	21	17	22	15	11	14	20	20	24
Cikarang	0	0	0	0	0	3	0	2	13	3	7	14
Bekasi	19	20	20	26	31	36	17	45	41	54	39	52
Purwakarta	4	4	4	5	3	9	2	3	5	13	7	5

Gambar 2 Hasil Transformasi Data

D. Hasil Perhitungan Data

1. Perhitungan Manual Menggunakan Microsoft Excel Algoritma *K-Means*

Perhitungan Manual Algoritma *k-means* menghasilkan perhitungan sebanyak dua iterasi dengan langkah-langkah perhitungan yang telah dijelaskan pada teori sebelumnya. Dalam perhitungan ini didapatkan hasil *cluster 1* adalah daerah pembelian kredit tinggi, *Cluster 2* adalah daerah pembelian kredit sedang dan *Cluster 3* adalah daerah kredit rendah. Lampiran hasil iterasi adalah sebagai berikut :

Centroid 1	Centroid 2	Centroid 3	K1	K2	K3
227.36	19.03	81.58		Klaster 2	
260.83	23.23	116.34		Klaster 2	
0.00	238.73	150.13	Klaster 1		
206.60	39.23	68.90		Klaster 2	
252.38	18.98	108.99		Klaster 2	
150.13	94.57	0.00			Klaster 3
248.70	13.66	104.36		Klaster 2	
Jumlah			1	5	1

Gambar 3 Hasil Perhitungan Manual Algoritma *K-Means*

2. Perhitungan Menggunakan Python Algoritma *K-Means*

Perhitungan Algoritma *k-means* menggunakan python digunakan untuk memperkuat hasil perhitungan yang telah dilakukan sebelumnya yaitu perhitungan manual. Perhitungan algoritma *k-means* menggunakan python memiliki hasil yang sama dengan perhitungan manual. Lampiran hasil adalah sebagai berikut :

```

Hasil clustering
cluster 1 = 1
cluster 2 = 5
cluster 3 = 1
    
```

Gambar 4 Hasil Perhitungan Python Algoritma *K-Means*

3. Perhitungan Manual Menggunakan Microsoft Excel Algoritma *K-Medoids*

Perhitungan manual algoritma *k-medoids* menggunakan Microsoft excel menghasilkan perhitungan sebanyak dua iterasi. Perhitungan dilakukan dengan langkah – langkah yang telah dijelaskan pada teori sebelumnya. Dalam perhitungan ini didapatkan 3 cluster, *cluster 1* = 1 daerah pembelian kredit tertinggi, *cluster 2* = 1 daerah pembelian kredit sedang, *cluster 3* = 5 daerah pembelian kredit rendah. Berikut adalah hasil iterasi algoritma *k-medoids* :

Centroid 1	Centroid 2	Centroid 3	K1	K2	K3	Kedekatan
81.58	227.36	30.32			Klaster 3	30.32
116.34	260.83	16.91			Klaster 3	16.91
150.13	0.00	248.70		Klaster 2		0.00
68.90	206.60	48.45			Klaster 3	48.45
108.99	252.38	19.18			Klaster 3	19.18
0.00	150.13	104.36	Klaster 1			0.00
104.36	248.70	0.00			Klaster 3	0.00
Jumlah			1	1	5	114.86

Gambar 5 Hasil Perhitungan Python Algoritma *K-Medoids*

4. Perhitungan Menggunakan Python Algoritma *K-Medoids*

Perhitungan algoritma *k-medoids* menggunakan python mendapatkan hasil yang berbeda dengan perhitungan manual. Dalam perhitungan python ini hasil yang di dapatkan adalah *cluster* 1 adalah 2 daerah pembelian kredit tinggi, *cluster* 2 adalah 2 daerah pembelian kredit sedang, dan *cluster* 3 daerah pembelian kredit rendah. Berikut adalah lampiran hasil perhitungan python :

```
Hasil clustering
cluster 1 = 2
cluster 2 = 2
cluster 3 = 3
```

Gambar 6 Hasil Perhitungan Python Algoritma K-Medoids

E. Evaluasi Menggunakan Metode *Silhouette Coefficient*

Evaluasi pada penelitian ini menggunakan metode *silhouette coefficient*, fungsi dari evaluasi ini adalah untuk melihat tingkat akurasi dari algoritma *k-means* dan *k-medoids*. Dari hasil perhitungan python didapatkan akurasi algoritma *k-means* adalah 0,4297 dan hasil dari algoritma *k-medoids* adalah 0,2307, maka artinya algoritma *k-means* lebih baik dalam *clustering* untuk penelitian ini dibandingkan algoritma *k-medoids*

IV. KESIMPULAN

Berdasarkan hasil dan pembahasan diatas dapat disimpulkan algoritma *k-means* lebih baik dalam *clustering* dibandingkan algoritma *k-medoids* dibuktikan dengan hasil evaluasi menggunakan metode *silhouette coefficient* nilai akurasi algoritma *k-means* lebih tinggi dibandingkan *k-medoids*. Hasil evaluasi algoritma *k-means* adalah 0,4297 sedangkan hasil evaluasi algoritma *k-medoids* adalah 0,2307. Kemudian untuk hasil *clustering* algoritma *k-means* yang dilakukan menggunakan 2 cara yaitu manual dan python memiliki hasil yang sama yaitu ada 1 daerah dengan pembelian kredit tinggi yaitu Matraman, 5 daerah pembelian kredit sedang yaitu Karawang, Cikampek, Kelapa Gading, Cikarang, dan 1 daerah pembelian kredit rendah yaitu Bekasi. Hasil perhitungan ini berdasarkan data akta *repertorium* atau data fidusia yang didapat di Kantor Notaris Endang Noviyanti, S.H., M.Kn dapat menentukan tingkat pembelian kredit tertinggi.

PENGAKUAN

Naskah Ilmiah ini adalah sebagian dari penelitian Tugas Akhir milik Dhitta Mari'ane Yusup dengan judul Model *Clustering* Pembelian Kredit Tertinggi Menggunakan Algoritma *K-Means* Dan *K-Medoids* yang dibimbing oleh Amril Mutoi Siregar dan Santi Arum Puspita Lestari.

DAFTAR PUSTAKA

- [1] Susilo, L., Apriani, R., & Zubaedah, R. 2021. Kedudukan Jaminan Fidusia Serta Perlindungan Hukum Bagi Lembaga Pembiayaan Konsumen. *SUPREMASI: Jurnal Pemikiran, Penelitian Ilmu-Ilmu Sosial, Hukum Dan Pengajarannya*, 16(1), 108.
- [2] Hakim, R. A. 2019. Perlindungan Hukum oleh Perusahaan Aplikasi terhadap Pemilik Kendaraan Angkutan Sewa Khusus. *Adliya*, 13(1), 58–72.
- [3] Badan Pusat Statistik. Perkembangan Jumlah Kendaraan Bermotor Menurut Jenis Unit, 2018-2020. bps.go.id. [Badan Pusat Statistik \(bps.go.id\)](https://www.bps.go.id)
- [4] Pemerintah Indonesia. 1999. Undang - Undang Republik Indonesia Nomor 42 Tahun 1999, Pasal 1 (2) Tentang Jaminan Fidusia. Sekretariat Negara. Jakarta
- [5] Sabir, M., & Tunnsia, R. 2020. Jaminan Fidusia Dalam Transaksi Perbankan; Studi Komparatif Hukum Positif dan Hukum Islam. *Mazahibuna : Jurnal Perbandingan Mazhab*.
- [6] Widjaja, G., Yani, A. 2000. *Seri Hukum Bisnis : Jaminan Fidusia*. Cetakan Pertama. RajaGrafindo Persada. Jakarta.
- [7] Ahdan, S., Latih, H. S., & Ramadona, S. 2018. Aplikasi Mobile Simulasi Perhitungan Kredit Pembelian Sepeda Motor pada PT. Tunas Motor Pratama. *Jurnal Tekno Kompak*, 12(1), 29-33.
- [8] Drs. Ismail, MBA., AK. 2010. *Manajemen Perbankan : Dari Teori Menuju Aplikasi*. Edisi Pertama Cetakan Kelima. Predanamedia. Jakarta
- [9] Siregar, A.M., Kom, S., Puspabhuana, M. K. D. A., Kom, S., & Kom, M. 2017. *Data Mining : Pengolahan Data Menjadi Informasi dengan Rapid Miner*. Cetakan Pertama. CV. Kekata Grup. Surakarta
- [10] Kusriani & Emha Taufiq Luthfi. 2009. *Algoritma Data Mining*. Penerbit Andi. Yogyakarta
- [11] Sagala, R. M. 2021. Prediksi Kelulusan Mahasiswa Menggunakan Data mining Prediction of college subject using K-means Algorithm in Data mining. 131–142.
- [12] Winarta, A., & Kurniawan, W. J. 2021. Optimasi Cluster K-Means Menggunakan Metode Elbow Pada Data Pengguna Narkoba Dengan Pemrograman Python. *Jurnal Teknik Informatika Kaputama (JTIK)*, 5(1), 113–119.
- [13] Riyanto, B. 2019. Penerapan Algoritma K-Medoids Clustering Untuk Pengelompokan Penyebaran Diare Di Kota Medan (Studi Kasus : Kantor Dinas Kesehatan Kota Medan). *KOMIK (Konferensi Nasional Teknologi Informasi Dan Komputer)*, 3(1).
- [14] Kamila, I., Khairunnisa, U., & Mustakim, M. 2019. Perbandingan Algoritma K-Means dan K-Medoids untuk Pengelompokan Data Transaksi Bongkar Muat di Provinsi Riau. *Jurnal Ilmiah Rekayasa Dan Manajemen Sistem Informasi*, 5(1), 119.
- [15] Falahi, F. M. 2019. Penerapan Metode Clustering Untuk Pengelompokan Mahasiswa Potensial Drop Out Menggunakan Algoritma *K-Means++* (Doctoral Dissertation, Universitas Islam Negeri Sunan Ampel Surabaya).
- [16] Paembonan, S., & Abduh, H. 2021. Penerapan Metode Silhouette Coefficient untuk Evaluasi Clustering Obat. *PENA TEKNIK: Jurnal Ilmiah Ilmu-Ilmu Teknik*, 6(2), 48.