

ANALISIS *CLUSTERING* UNTUK SEGMENTASI PENGGUNA KARTU KREDIT DENGAN MENGGUNAKAN ALGORITMA K-MEANS DAN PRINCIPAL COMPONENT ANALYSIS

Muhammad Nur Akbar^{1*}, Azizah Salsabila², Aldi Perdana Asri³,
Muhammad Syawir⁴

Universitas Islam Negeri Alauddin Makassar^{1,2,3,4}

muhammad.akbar@uin-alauddin.ac.id¹, azizahsalsa444@gmail.com²,
aldii.perdanaa@gmail.com³, muh.syawirflow@gmail.com⁴

Abstrak

Segmentasi pelanggan adalah proses yang digunakan oleh perusahaan untuk mengelompokkan pelanggan berdasarkan karakteristik yang sama. Tujuannya adalah memahami kebutuhan dan preferensi pelanggan dengan lebih baik, sehingga perusahaan dapat menyediakan produk dan layanan yang sesuai dengan kebutuhan pelanggan. Salah satu cara melakukan segmentasi pelanggan adalah dengan menggunakan algoritma clustering, seperti k-means. Algoritma ini mengelompokkan data ke dalam cluster yang berdekatan dengan centroid yang dipilih secara acak. Dalam kasus segmentasi pelanggan kartu kredit, algoritma k-means dapat digunakan untuk mengelompokkan pelanggan berdasarkan karakteristik seperti jumlah transaksi, jumlah pembayaran, dan riwayat kredit. Dengan demikian, perusahaan dapat memahami kebutuhan dan preferensi pelanggan kartu kredit dengan lebih baik, serta menentukan strategi pemasaran yang lebih efektif. Kelebihan dari algoritma k-means dan metode clustering adalah model yang dikembangkan dapat membantu perusahaan dalam menentukan strategi pemasaran yang lebih efektif, algoritma yang mudah digunakan dengan waktu komputasi yang cepat dan hasil yang akurat, serta digunakan pula algoritma PCA untuk pengurangan dimensi dan mempermudah visualisasi data. Berdasarkan hasil pengujian dan analisa pada data pelanggan kartu kredit, performa algoritma k-means dianggap relatif baik melakukan segmentasi dengan jumlah cluster yaitu 3 dan nilai Davies Bouldin sebesar -0.778.

Kata kunci: segmentasi pelanggan, kartu kredit, K-means, PCA

Abstract

[CLUSTERING ANALYSIS FOR CREDIT CARD USER SEGMENTATION USING K-MEANS ALGORITHM AND PRINCIPAL COMPONENT ANALYSIS] Customer segmentation is a process used by companies to group customers based on common characteristics. The goal is to understand customer needs and preferences better so that companies can provide products and services that match customer needs. One way to segment customers is to use clustering algorithms, such as k-means. This algorithm groups data into adjacent clusters with randomly selected centroids. In the case of credit card customer segmentation, the k-means algorithm can be used to group customers based on characteristics such as number of transactions, amount of payments, and credit history. Thus, companies can better understand the needs and preferences of credit card customers and determine more effective marketing strategies. The advantages of the k-means algorithm and the clustering method are that the developed models can help companies determine more effective marketing strategies, easy-to-use algorithms with fast computation time and accurate results, and the PCA algorithm is also used to reduce dimensions and makes data visualization easier. Based on the test results and analysis of credit card customer data, the performance of the k-means algorithm is considered relatively good for segmentation with the number of clusters = 3 and the Davies Bouldin value = -0.778.

Keywords: customer segmentation, credit card, K-means, PCA

1. PENDAHULUAN

Segmentasi pelanggan adalah suatu proses yang digunakan oleh perusahaan untuk mengelompokkan pelanggan mereka berdasarkan karakteristik yang serupa. Tujuan dari segmentasi pelanggan adalah untuk memahami kebutuhan dan preferensi pelanggan secara lebih baik, sehingga perusahaan dapat menyediakan produk dan layanan yang lebih sesuai dengan kebutuhan pelanggan.

Salah satu cara untuk melakukan segmentasi pelanggan adalah dengan menggunakan metode *clustering*. *Clustering* adalah suatu proses yang digunakan untuk mengelompokkan data ke dalam kelompok-kelompok (*cluster*) yang memiliki karakteristik yang serupa. Algoritma k-means merupakan salah satu algoritma *clustering* yang paling populer. Algoritma ini bekerja dengan cara memilih k centroid awal secara acak, kemudian mengelompokkan data ke dalam *cluster* yang berdekatan dengan centroid tersebut. Proses ini dilakukan secara iteratif sampai tidak ada perubahan lagi pada *cluster*. [1]

Dalam kasus segmentasi pelanggan kartu kredit, algoritma k-means dapat digunakan untuk mengelompokkan pelanggan kartu kredit berdasarkan karakteristik seperti jumlah transaksi, jumlah pembayaran, riwayat kredit, dan lain-lain. Dengan demikian, perusahaan dapat memahami kebutuhan dan preferensi pelanggan kartu kredit dengan lebih baik, sehingga dapat menyediakan produk dan layanan yang lebih sesuai dengan kebutuhan pelanggan tersebut. Selain itu, segmentasi pelanggan kartu kredit dengan menggunakan algoritma k-means juga dapat membantu perusahaan dalam menentukan strategi pemasaran yang lebih efektif.

Adapun kelebihan dari algoritma k-means dan metode *clustering* yaitu model segmentasi pelanggan kartu kredit yang dikembangkan membantu perusahaan menentukan strategi pemasaran yang lebih efektif dengan menggunakan algoritma K-means dan Principal Component Analysis (PCA). Algoritma *clustering* yang populer dan mudah digunakan, membantu

mengelompokkan pelanggan berdasarkan karakteristik yang serupa dengan waktu komputasi yang cepat dan hasil yang akurat yang mudah dipahami dan diinterpretasikan. PCA mempermudah pengurangan dimensi dan visualisasi data. Nilai K untuk K-means ditentukan dengan menggunakan skor Davies Boudin sebagai acuan.

2. URAIAN PENELITIAN

2.1 Data Mining

Data Mining adalah alat yang memungkinkan pengguna untuk mengakses data dalam jumlah besar dengan cepat. Pemahaman yang lebih spesifik tentang penambangan data, alat dan aplikasi analisis statistik data. Penambangan data adalah proses mengekstraksi atau mengekstraksi sejumlah besar data dan informasi dari basis data besar yang sebelumnya tidak diketahui tetapi dapat dipahami dan berguna serta digunakan untuk membuat keputusan bisnis yang sangat penting.

Penambangan data menggambarkan sekumpulan teknik yang ditujukan untuk menemukan pola yang tidak diketahui dalam data yang dikumpulkan. Penambangan data memungkinkan pengguna untuk menemukan pengetahuan dalam data basis data yang mungkin tidak diketahui keberadaannya oleh pengguna. [4]

2.2 Clustering

Clustering adalah metode analisis data, sering dimasukkan dalam metode penambangan data, yang bertujuan untuk mengelompokkan data dengan karakteristik yang mirip. *Clustering* adalah teknik penambangan data yang digunakan untuk mengkategorikan data ke dalam kelompok atau *cluster* tertentu. Setiap potongan data dikelompokkan berdasarkan kemiripannya dengan potongan data lainnya.

Data dengan karakteristik yang sama ditempatkan pada *cluster* yang sama, dan data dengan karakteristik yang berbeda ditempatkan pada *cluster* yang berbeda. Ada banyak metode yang dapat digunakan dalam metode *clustering*, salah satunya adalah K-Means.

2.3 Algoritma K-Means

K-Means adalah metode yang sering digunakan untuk pengelompokan. Algoritma K-Means merupakan metode *clustering* non-hierarchical dengan waktu komputasi yang relatif cepat. Algoritma K-Means dimulai dengan terlebih dahulu menentukan berapa banyak *cluster* yang ingin dibuat, kemudian menentukan nilai awal setiap *cluster*, kemudian menghitung jarak setiap item data menggunakan nilai awal yang telah ditentukan. Setelah jarak semua data dihitung, data tersebut dimasukkan ke dalam *cluster* berikutnya.

Kemudian menghitung rata-rata setiap *cluster*, rata-rata ini nantinya akan digunakan sebagai nilai awal baru saat menghitung jarak semua data. Dalam k-means ini, pusat atau centroid *cluster* dipilih secara acak dari sekumpulan kumpulan data (populasi) pada tahap awal. K-Means kemudian menguji setiap komponen dalam populasi data dan memberi label pada komponen di salah satu centroid yang telah ditentukan berdasarkan jarak minimum antara komponen (data) dan setiap centroid. Posisi centroid dihitung kembali hingga semua komponen data dikelompokkan ke dalam masing-masing centroid dan akhirnya terbentuk posisi centroid baru. Langkah-langkah pengelompokan menggunakan metode K-Means *clustering* adalah sebagai berikut[9]:

1. Tentukan nilai K jumlah *cluster* yang ingin dibentuk
2. Pilih nilai pusat awal (titik tengah) *cluster* hingga K
3. Hitung jarak dari semua data input ke setiap centroid menggunakan rumus jarak

$$D(i, j) = \sqrt{(x_{1j} - x_{1i})^2 + (x_{2j} - x_{2i})^2 + \dots + (x_{kj} - x_{ki})^2} \quad (1)$$

Keterangan: $D(i, j)$ = jarak data ke i ke pusat cluster j
 x_{ki} = data ke i pada atribut data ke k
 x_{kj} = data ke j pada atribut data ke k

2.4 Principal Component Analysis (PCA)

Principal Component Analysis (PCA) merupakan teknik yang berguna dalam mengekstraksi struktur dari set data dengan dimensi yang cukup tinggi. Teknik ini sering digunakan dalam penelitian dengan tujuan

untuk meringkas data dengan jumlah variabel yang lebih kecil. Dengan menggunakan PCA, kita dapat dengan mudah mengurangi dimensi data tanpa kehilangan banyak informasi yang berguna dari data tersebut, sehingga dapat membantu kita dalam melakukan visualisasi data dan menemukan pola-pola yang terdapat didalamnya. [8]

Dengan mengaplikasikan Principal Component Analysis (PCA) pada data kita, kita dapat mengubah data yang memiliki 17 dimensi menjadi data dengan 2 dimensi yang lebih mudah untuk divisualisasikan. PCA merupakan teknik yang membantu mengurangi jumlah variabel data dengan cara mengubah satu set besar variabel menjadi sejumlah variabel yang lebih kecil, namun masih memuat sebagian besar informasi yang terdapat dalam set besar tersebut.

2.5 Rapid Miner

RapidMiner merupakan perangkat lunak yang bersifat terbuka (open source) dan menyediakan solusi untuk melakukan analisis terhadap data mining, text mining, dan analisis prediksi. RapidMiner menggunakan beragam teknik deskriptif dan prediksi untuk memberikan wawasan kepada penggunanya, sehingga pengguna dapat membuat keputusan yang tepat. Selain itu, RapidMiner juga memiliki keunggulan dalam pengolahan data yang cepat dan mudah digunakan, serta dapat bekerja dengan berbagai jenis data seperti data kuantitatif maupun kualitatif. Dengan demikian, RapidMiner merupakan pilihan yang tepat bagi para pengguna yang ingin melakukan analisis terhadap data mining, text mining, dan analisis prediksi. [2]

3. PERANCANGAN SISTEM

3.1 Data

Data yang digunakan bersumber dari kaggle yang berisi 8950 pemegang kartu aktif selama 6 bulan terakhir dengan memiliki 17 atribut-atribut yang menjadi tolak ukur dalam melakukan *clustering*.

Sumber data:

<https://www.kaggle.com/datasets/arjunbhasin/2013/ccdata>

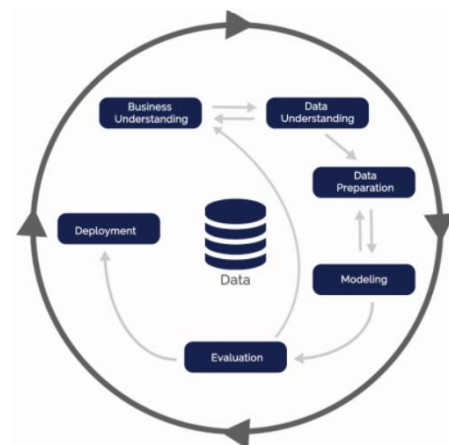
Tabel 1. Deskripsi Dataset

No	Atribut	Tipedata	Deskripsi
1	cust_id	Nominal	ID pemegang kartu kredit
2	balance	Real	Jumlah saldo tersisa dalam akun
3	balance_frequency	Real	Seberapa sering balance CC diperbarui, (1 = sering diperbarui, 0 = tidak sering diperbarui)
4	purchases	Real	Jumlah pembelian yang dilakukan dari akun
5	oneoff_purchases	Real	Jumlah pembelian maksimum dilakukan dalam sekali transaksi
6	installments_purchases	Real	Jumlah pembelian yang dilakukan dengan cicilan
7	cash_advance	Real	Uang muka yang diberikan oleh pengguna
8	purchases_frequency	Real	Seberapa sering pembelian dilakukan, (1 = sering dibeli, 0 = tidak sering dibeli)
9	oneoff_purchases_frequency	Real	Seberapa sering pembelian terjadi dalam satu transaksi (1 = sering dibeli, 0 = tidak sering dibeli)
10	purchases_installments_frequency	Real	Seberapa sering pembelian dengan cara mencicil dilakukan (1 = sering dilakukan, 0 = tidak sering dilakukan)
11	cash_advance_frequency	Real	Seberapa sering uang muka dibayarkan
12	cash_advance_trx	integer	Jumlah transaksi yang dilakukan dengan uang muka
13	purchases_trx	integer	Jumlah transaksi pembelian yang dilakukan
14	credit_limit	integer	Batas Kartu Kredit untuk pengguna
15	payments	Real	Jumlah Pembayaran yang dilakukan oleh pengguna

16	minimum_payments	Real	Jumlah minimum pembayaran yang dilakukan oleh pengguna
17	prc_full_payment	Real	Persentase pembayaran penuh yang dibayarkan oleh pengguna
18	tenure	integer	Masa berlaku layanan kartu kredit bagi pengguna

3.2 Metodologi Penelitian

Penelitian ini menggunakan metode Cross Industry Standard Process for Data Mining atau yang biasa disingkat menjadi CRISP-DM. CRISP-DM dianggap teknologi yang netral, industri independen dan merupakan standar de-facto untuk data mining [3].



Gambar 1. Alur CRISP-DM

Proses data mining berdasarkan CRISP-DM terdiri dari 6 fase. Yaitu:

1. Business Understanding

Merupakan pemahaman tentang substansi dari kegiatan data mining yang akan dilakukan, kebutuhan dari perspektif bisnis. Kegiatannya antara lain: menentukan sasaran atau tujuan bisnis, memahami situasi bisnis, menentukan tujuan data mining dan membuat perencanaan strategi serta jadwal penelitian.

2. Data Understanding

Fase mengumpulkan data awal, mempelajari data untuk bisa mengenal data yang akan dipakai, mengidentifikasi masalah yang berkaitan dengan kualitas data,

mendeteksi subset yang menarik dari data untuk membuat hipotesa awal.

3. Data Preparation

Fase yang padat karya. Aktivitas yang dilakukan antara lain memilih table dan field yang akan ditransformasikan ke dalam database baru untuk bahan data mining (set data mentah).

Dilakukan *preprocessing* data menggunakan RapidMiner sebagai berikut:

- Proses *cleansing* dengan menganalisa kualitas dari suatu data dengan mengubah, mengoreksi atau menghapus data-data yang tidak sesuai dengan kebutuhan penelitian.
- Mencari nilai yang hilang atau data yang kosong kemudian diisi berdasarkan nilai rata-rata pada data tersebut sesuai dengan lokasinya.

4. Modeling

Fase menentukan tehnik data mining yang digunakan, menentukan tools data mining, teknik data mining, algoritma data mining, menentukan parameter dengan nilai yang optimal.

5. Evaluation

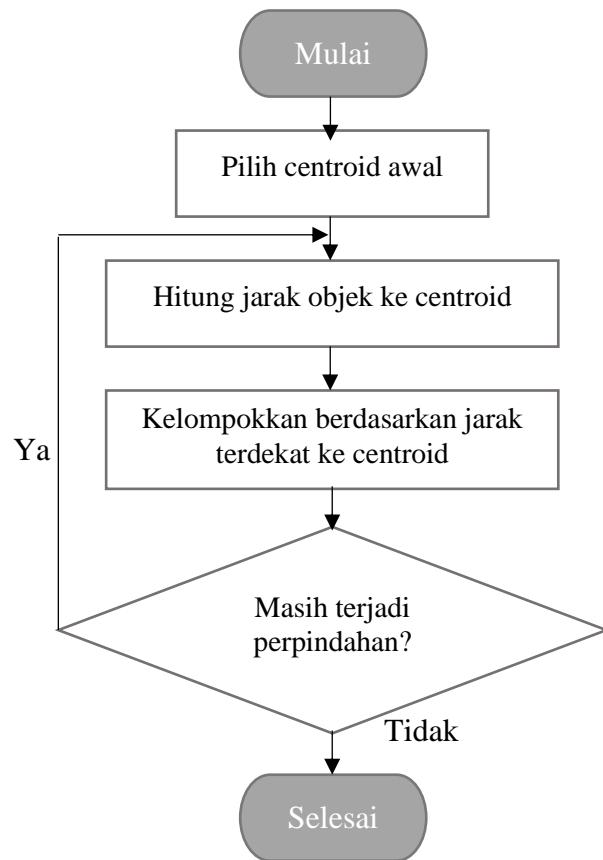
Fase interpretasi terhadap hasil data mining yang ditunjukkan dalam proses pemodelan pada fase sebelumnya. Evaluasi dilakukan secara mendalam dengan tujuan menyesuaikan model yang didapat agar sesuai dengan sasaran yang ingin dicapai dalam fase pertama.

Menurut Permatadevi, et al. [6] jika proses pengklasteran untuk masing-masing k selesai, maka untuk menentukan jumlah klaster yang terbaik dapat dilakukan penilaian menggunakan Davies-Bouldin Index (DBI). Pendekatan pengukuran ini bertujuan untuk memaksimalkan jarak antara klaster yang satu dengan yang lainnya dan pada waktu yang sama mencoba untuk meminimalkan jarak antara objek dalam sebuah klaster [5]. Pengklasteran dengan jumlah klaster yang terbaik adalah pengklasteran yang memiliki nilai DBI minimum.

6. Deployment

Fase penyusunan laporan atau presentasi dari pengetahuan yang didapat dari evaluasi pada proses data mining[7].

3.3 Cara Kerja Algoritma K-Means



Gambar 2. Alur kerja algoritma K-Means

- Tentukan jumlah *cluster* yang diinginkan (k).
- Tentukan *centroid* awal dari masing-masing *cluster*.
- Centroid* ini bisa dipilih secara acak atau dengan menggunakan metode seperti Forgy atau Random Partition.
- Tentukan keanggotaan setiap data terhadap masing-masing *cluster* berdasarkan jarak data tersebut ke *centroid cluster* tersebut.
- Data yang paling dekat dengan *centroid* tertentu akan dianggap sebagai anggota dari *cluster* tersebut. Hitung *centroid* baru dari masing-masing *cluster* dengan menghitung rata-rata dari setiap data yang merupakan anggota dari *cluster* tersebut.

4. HASIL DAN ANALISIS

4.1 Hasil Perancangan

1. Pengolahan Data

Data awal yang digunakan menggunakan format .csv, kemudian pada RapidMiner kita menggunakan *Read CSV* dan memasukkan data awalnya. Pada gambar berikut kita memilih atribut yang akan digunakan.

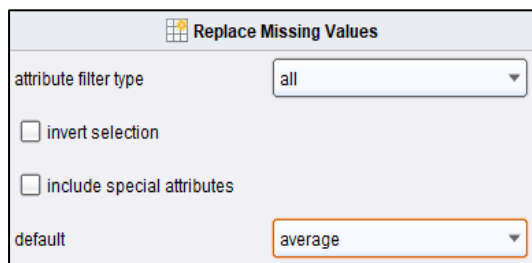
Attribute	Type
CUST_ID	integer
BALANCE	real
BALANCE_F	real
PURCHASE	real
ONEOFF_P	real
INSTALLMENT	real
CASH_ADV	real
PURCHASE	real
ONEOFF_P	real
PURCHASE	real
CASH_ADV	real
CASH_ADV	integer

Gambar 3. Proses import data pada *RapidMiner*.

Data yang di *import* memiliki atribut masing-masing, disini kita mengatur tipe data berdasarkan atributnya. Untuk *id costumer* kita merubah atributnya menjadi *id* dan 17 atribut lainnya kita sesuaikan berdasarkan isi datanya, data yang memiliki koma (,) memiliki tipe data *real* dan yang berupa angka biasa memiliki tipe data *integer*.

2. Preprocessing Data

Data yang telah dimasukkan kemudian kita bersihkan, terdapat banyak data yang hilang atau biasa disebut *missing value* pada beberapa atribut. Maka kita memasukkan operator *Replace Missing Values* pada bagian *Data Cleansing*. Kemudian kita ubah data yang hilang menjadi rata-rata berdasarkan atributnya masing-masing.



Gambar 4. Penanganan *missing value* pada data.

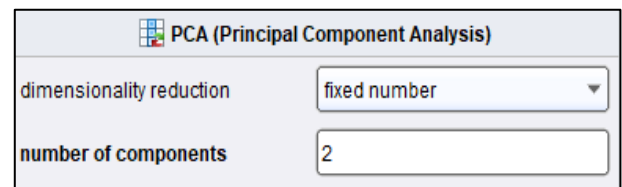
Kemudian terdapat banyak data yang *outlier*, tetapi kita tidak menghapusnya karena dataset ini berisi informasi pribadi pada setiap

customer sehingga kita dapat menganggap *outlier* ini sebagai nilai ekstrim.

4.2 Hasil Pengujian Sistem

1. Reduksi dimensi data

Pada tahapan ini, dilakukan reduksi data terlebih dahulu sebelum melakukan *clustering* karena akan lebih sulit jika data divisualisasikan dengan 17 atribut. Maka digunakan *Principal Component Analysis* (PCA) untuk mengubah data menjadi 2 dimensi saja menggunakan operator *Principal Component Analysis* pada bagian *Transformation* di *RapidMiner*.

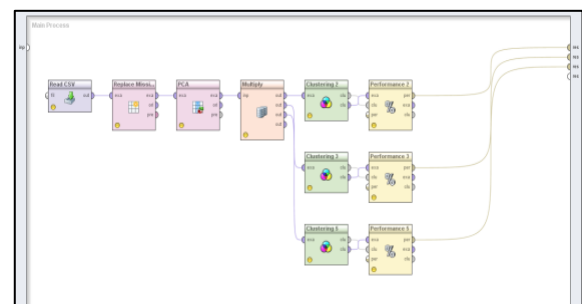


Gambar 5. Reduksi dimensi data menggunakan PCA

Dilakukan reduksi data yang tadinya memiliki tujuh belas dimensi berdasarkan atributnya menjadi 2 dimensi agar visualisasinya lebih baik.

2. Clustering

Pada *clustering* dataset ini, dilakukan konfigurasi data dan operator seperti berikut:



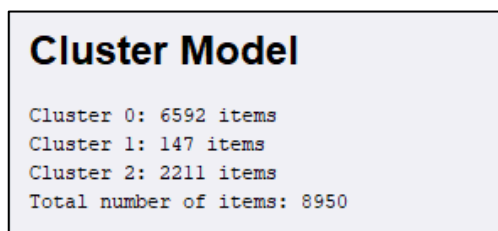
Gambar 6. Alur proses clustering K-Means menggunakan *RapidMiner*.

Pada tahapan ini ditentukan banyaknya *cluster* berdasarkan performa dari beberapa *cluster* yang diinginkan dengan mengukur evaluasi *cluster* dengan metode *Davies-bouldin*. Berikut hasil pengukuran evaluasinya.

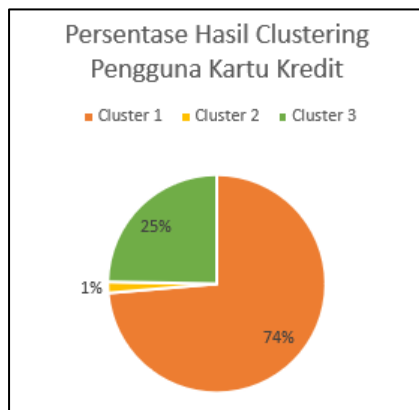
Tabel 2. Hasil evaluasi nilai K pada K-Means

K	Davies Bouldin
2	-0.838
3	-0.778
5	-0.782

Dari hasil performa pada tiap *cluster*, diambil $k = 3$ karena memiliki nilai Davies Bouldin terkecil. Artinya, jumlah *cluster* terbaik dalam hal ini adalah 3. Kemudian $k = 3$ untuk algoritma K-Means digunakan untuk analisa *cluster* lebih lanjut. Berikut hasil *clustering*-nya:



Gambar 7. Hasil *clustering* dengan $K=3$



Gambar 8. Persentase anggota *cluster*

Dari hasil *clustering*, kita bisa melihat *centroid* dari setiap *cluster* pada bagian *ExampleSet* sebagai berikut:

Attribute	cluster_0	cluster_1	cluster_2
pc_1	-2060.935	19040.374	4878.675
pc_2	-479.073	-9582.249	2065.418

Gambar 9. Hasil *centroid* untuk tiap *cluster*

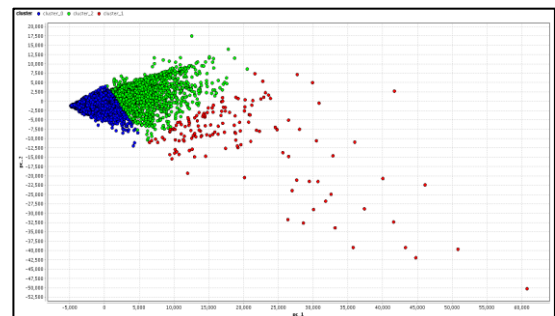
Kemudian dari hasil *centroid*-nya, kita bisa menganalisa bahwa hasil *cluster* dari 0 ke 2 pada dua dimensi yang telah direduksi ialah rendah ke tinggi. Sehingga kita bisa mengetahui segmentasi pelanggan berdasarkan hasilnya.

Setelah proses *running* juga akan tampilan contoh datanya yang telah ditambahkan atribut baru yaitu *cluster* seperti yang terlihat pada gambar di bawah ini.

R...	CUST_ID	cluster	pc_1	pc_2
1	C10001	cluster_0	-4326.384	-921.567
2	C10002	cluster_2	4118.917	2432.846
3	C10003	cluster_2	1497.908	1997.579
4	C10004	cluster_2	1394.549	1488.743
5	C10005	cluster_0	-3743.352	-757.343
6	C10006	cluster_0	-1910.333	-931.459
7	C10007	cluster_2	10923.837	-4046.348
8	C10008	cluster_0	-2497.208	-165.253
9	C10009	cluster_0	730.474	1211.650
10	C10010	cluster_2	3680.958	2054.986
11	C10011	cluster_0	-2815.249	-994.356
12	C10012	cluster_0	-2500.951	-1786.392
13	C10013	cluster_0	-779.138	-2297.271
14	C10014	cluster_2	1958.900	210.212
15	C10015	cluster_0	-1703.442	803.150

Gambar 10. Data hasil *clustering*

Terakhir, hasil *cluster* divisualisasikan dalam bentuk sebaran plot pada gambar berikut:



Gambar 11. Visualisasi hasil *clustering*

4.3 Analisis Hasil

Tabel 3. Rata-rata nilai tiap atribut terhadap *cluster* yang terbentuk

No	Nama Atribut	Nilai Rata-rata		
		C0	C1	C2
1	balance	845.6890055	4463.121548	3514.784675
2	balance_frequency	0.854890518	0.925782605	0.94077099
3	purchases	620.9310118	10887.2085	1485.792121

4	oneoff_purchases	323.473 0173	7718. 57435 4	920.556 2687
5	installments_purchases	297.753 6074	3168. 78381	565.559 6653
6	cash_advance	478.543 6758	4100. 36248 6	2263.04 0823
7	purchases_frequency	0.47219 5065	0.795 31023 8	0.52420 4856
8	oneoff_purchases_frequency	0.15844 8397	0.654 92832	0.30358 66
9	purchases_installments_frequency	0.35547 222	0.626 20815 6	0.37376 2431
10	cash_advance_frequency	0.10466 9167	0.182 65451 7	0.22284 5422
11	cash_advance_trx	2.20464 199	7.333 33333 3	6.09045 6807
12	purchases_trx	10.9208 1311	89.56 46258 5	21.0298 5075
13	credit_limit	2803.37 9394	12512 .2449	9002.46 7004
14	payments	997.698 2559	16485 .0000 6	2945.05 4529
15	minimum_payments	549.381 3641	2194. 92040 2	1688.58 0061
16	prc_full_payment	0.15966 1894	0.352 58240 8	0.12276 1324
17	tenure	11.4394 7209	11.83 67346 9	11.7281 773

Keterangan:

- : terendah
- : menengah
- : tertinggi

Berikut ini hasil analisis *cluster* pada dataset pemegang kartu kredit aktif selama 6 bulan terakhir dengan 17 atribut yang menggambarkan kebiasaan pengguna:

1. *Cluster 0*, nilai rata-rata untuk semua atribut memiliki nilai terendah dari ketiga *cluster* yang terbentuk. Hal ini menandakan bahwa pengguna pada *cluster* ini secara frekuensi penggunaan, *buying power* menggunakan kartu kredit dll. masih relatif rendah sehingga perlu perlakuan khusus misalnya memberikan penawaran atau promo seperti pengumpulan poin, *cashback* dll. untuk meningkatkan jumlah transaksi dan daya beli pengguna menggunakan kartu kredit. Segmen ini juga dapat dikatakan

relatif rentan untuk berhenti atau kurang loyal dalam menggunakan produk.

2. *Cluster 1*, nilai rata-rata untuk hampir semua atribut memiliki nilai tertinggi dan berbeda cukup signifikan dibandingkan 2 *cluster* lainnya, hanya pada atribut *balance_frequency* dan *cash_advance_frequency* yang lebih rendah dari cluster 2. Hal ini menandakan bahwa pengguna pada *cluster* ini secara frekuensi dan nilai transaksi sudah terbilang tinggi, penggunaannya pun dapat dikategorikan sebagai pengguna loyal sehingga tidak butuh banyak penanganan khusus, cukup dipertahankan agar tidak menurun.
3. *Cluster 2*, nilai rata-rata untuk hampir semua atribut memiliki nilai menengah, lebih tinggi dari cluster 0 namun lebih rendah dari cluster 1, hanya pada atribut *balance_frequency* dan *cash_advance_frequency* yang lebih tinggi dari cluster 1. Cluster ini juga dianggap perlu mendapatkan perlakuan khusus agar secara jumlah dan besaran transaksinya dapat meningkat dan loyalitas pengguna tetap dapat dipertahankan bahkan meningkat.

5. KESIMPULAN

Berdasarkan hasil pengujian dan analisis *clustering* terhadap pelanggan kartu kredit, dapat disimpulkan bahwa:

1. Segmentasi pengguna kartu kredit terbukti dapat dilakukan menggunakan pendekatan analisis *clustering*.
2. Algoritma K-Means dianggap bekerja relatif baik dpada kasus ini dengan nilai K=3 dengan Davies Bouldin score sebesar – 0.778.
3. Algoritma PCA digunakan untuk memudahkan visualisasi hasil cluster agar lebih mudah dipahami.
4. Terdapat 3 cluster yang terbentuk, cluster 0 yang berisi 6592 pengguna termasuk dalam segmen rendah, cluster 1 yang berisi 147 pengguna termasuk dalam segmen tinggi, dan cluster 2 yang berisi 2211 pengguna termasuk dalam segmen menengah.

6. DAFTAR PUSTAKA

- [1] A. Nur Khormarudin, “Teknik Data Mining: Algoritma K-Means Clustering,” *J. Ilmu Komput.*, pp. 1–12, 2016, [Online]. Available: <https://ilmukomputer.org/category/data-mining/>
- [2] Aprilla Dennis, “Belajar Data Mining dengan RapidMiner,” *Innov. Knowl. Manag. Bus. Glob. Theory Pract. Vols 1 2*, vol. 5, no. 4, pp. 1–5, 2013, [Online]. Available: http://esjournals.org/journaloftechnology/archive/vol1no6/vol1no6_6.pdf%5Cnhttp://www.airccse.org/journal/nsa/5413nsa02.pdf
- [3] Azevedo, A. and Santos, M.F. (2008) KDD, SEMMA and CRISP-DM: A Parallel Overview. *Proceedings of the IADIS European Conference Data Mining*, Amsterdam, 2426 July 2008, 182-185.
- [4] C. Zai, “Implementasi Data Mining Sebagai Pengolahan Data,” *Portal Data*, vol. 2, no. 3, pp. 1–12, 2022.
- [5] Hilmi, M., N., 2015. Pemetaan Preferensi Mahasiswa Baru dalam Memilih Jurusan Menggunakan Artificial Neural Network (ANN) dengan Algoritma Self Organizing Maps (SOM). *Jurnal Gaussian Vol. 4*, No. 1: Hal. 53-60
- [6] Permatadevi, M. A., Hendrawan, R. A., & Hafidz, I., 2013. Karakteristik Pelanggan Telepon Kabel Menggunakan Clustering SOM dan K-Means untuk Mengurangi Kesalahan Klasifikasi Pelanggan Perusahaan Telekomunikasi. *Jurnal Teknik Pomits Vol. 1*, No. 1 Hal. 1-6.
- [7] Shearer, C. (2000). The CRISP-DM Model: The New Blueprint for Data Mining. *Journal of Data Warehousing*, 5, 13-22.
- [8] T. Saepurohman and E. Putro, B, “Analisis Principal Component Analysis (PCA) Untuk Mereduksi Faktor-Faktor yang Mempengaruhi Kualitas Kulit Kikil Sapi,” *Pros. Semin. Konf. Nas. IDEC*, no. 2008, p. C01.1-C01.10, 2019, [Online]. Available: <https://idec.ft.uns.ac.id/prosiding2019>
- [9] W. A. Wahyuni and S. Saepudin, “Penerapan Data Mining Clustering Untuk Mengelompokkan Berbagai Jenis Merk Mesin Cuci,” *Semin. Nas. Sist. ...*, pp. 306–313, 2021, [Online]. Available: <https://sismatik.nusaputra.ac.id/index.php/sisimatik/article/view/35%0Ahttps://sismatik.nusaputra.ac.id/index.php/sisimatik/article/download/35/31>