

Algoritme *K-Means* dalam Pengelompokan Kantor Cabang untuk Optimalisasi Manajemen Perbankan

Angga Ardianto^{#1}, Bowo Relawanto^{#2}, Arief Wibowo^{#3}

[#]Magister Ilmu Komputer, Fakultas Teknologi Informasi, Universitas Budi Luhur
Jl. Ciledug Raya, Petukangan Utara, Jakarta Selatan, Indonesia

¹ardhian21@gmail.com

²bwrelawan85@gmail.com

³arief.wibowo@budiluhur.ac.id

Abstract— Branch segmentation is needed by companies to facilitate management in planning and managing these branches. Mistakes in determining the segmentation or grouping of branches will have an impact on planning or management, such as the efficiency of operational costs, calculation of branch performance, operational supervision, and optimization of company resources. As a first step in optimizing branch management, optimal branch grouping, or according to branch similarities, can be adjusted to the size or size of a branch. In the supervisory and monitoring functions, this grouping is also a consideration and prioritization of supervision, where large branches will of course use different plans with smaller branches. In setting the budget to avoid potential fraud, it is best if the operating budget is adjusted according to the size of the branch. This study uses the *K-Means* algorithm to classify branch offices based on transactions per month and the number of types of transactions according to the required segmentation. Branches can be grouped into large, medium, and small groups. The results show that the *K-Means* algorithm can produce bank branch groupings based on the number of types of transactions and the average transaction per month, which is divided into three clusters. The three clusters are the large branch cluster or cluster 1, the intermediate branch cluster or cluster 2, and the small branch cluster or cluster 3. The test uses the Davies Bouldin Index of 0.5.

Keywords - Data mining, *K-Means* algorithm, banking, branch segmentation, optimizing management.

Abstrak— Segmentasi cabang diperlukan perusahaan untuk mempermudah manajemen dalam membuat perencanaan dan pengelolaan cabang-cabang tersebut. Kesalahan penentuan segmentasi atau pengelompokan cabang akan berdampak pada perencanaan atau pengelolaan, seperti pada efisiensi biaya operasional, penghitungan kinerja cabang, pengawasan operasional, dan optimalisasi sumber daya perusahaan. Sebagai langkah awal dalam optimalisasi pengelolaan cabang, pengelompokan cabang yang optimal, atau sesuai dengan kemiripan cabang, bisa disesuaikan dengan besar atau kecilnya sebuah cabang. Dalam fungsi pengawasan dan pemanataan, pengelompokan ini juga menjadi pertimbangan dan prioritas pengawasan, di mana cabang besar tentunya akan menggunakan perencanaan yang berbeda dengan cabang yang lebih kecil. Dalam pengaturan anggaran untuk menghindari potensi *fraud*, sebaiknya anggaran operasional disesuaikan dengan besarnya cabang tersebut. Penelitian ini menggunakan algoritme *K-Means* untuk mengelompokkan kantor cabang berdasarkan transaksi per bulan dan jumlah jenis transaksi sesuai dengan segmentasi yang dibutuhkan. Cabang dapat dikelompokkan menjadi kelompok besar, sedang, dan kecil. Hasil penelitian menunjukkan

bahwa algoritme *K-Means* ini dapat menghasilkan pengelompokan cabang-cabang bank berdasarkan jumlah jenis transaksi dan rata-rata transaksi per bulan yang dibagi menjadi tiga *cluster*. Ketiga *cluster* itu adalah *cluster* cabang besar atau *cluster* 1, *cluster* cabang menengah atau *cluster* 2, dan *cluster* cabang kecil atau *cluster* 3. Pengujian menggunakan Davies Bouldin Index sebesar 0,5.

Kata Kunci - Data mining, algoritme *K-Means*, perbankan, segmentasi cabang, optimalisasi manajemen.

I. PENDAHULUAN

Dalam industri keuangan, bank mempunyai salah satu tujuan yaitu mempertahankan usahanya dan memperoleh keuntungan. Akan tetapi, dalam menjalankan usaha pasti memiliki banyak rintangan, terutama persaingan yang begitu banyak dan meningkatnya biaya operasional sehingga membuat perusahaan mengambil keputusan untuk melakukan penutupan cabang usaha atau kantor. Dalam melakukan keputusan tersebut tentunya perusahaan membutuhkan data dan perlu melakukan analisis supaya tidak terjadi kesalahan atau meminimalisir terjadinya kesalahan saat mengambil sebuah keputusan. Bank perlu menganalisis dalam segi anggaran, operasional, sumber daya, keuntungan yang didapat pada suatu cabang, serta membandingkan cabang yang lainnya dengan faktor - faktor yang mempengaruhi cabang saat ini.

Bank merupakan jenis usaha di bidang jasa keuangan dan investasi. Dalam kondisi yang kompetitif manajer operasional kantor pusat bank dituntut untuk melakukan analisis keputusan dalam pengurangan cabang. Selama ini manajer mengalami kesulitan dalam mengambil keputusan karena tidak adanya data yang menjadikan suatu kesimpulan yang bisa mengacu untuk mengambil suatu keputusan. Pertimbangan dalam pengambilan keputusan yaitu biaya operasional, sumber daya, asset, dan profit yang didapat pada suatu cabang. Faktor-faktor tersebut dapat dijadikan dasar dalam pengelompokan. Tujuan penelitian ini adalah melakukan pengelompokan kantor cabang dengan metode *clustering K-Means*.

Klasterisasi adalah salah metode *data mining* yang mampu melakukan pengelompokan objek-objek ke dalam *cluster-cluster*. *Cluster* merupakan sekelompok obyek data yang *similar* satu dengan lainnya, namun berada di dalam kelompok yang sama [1]. Teknik *clustering* telah banyak

digunakan untuk memecahkan masalah terkait pemisahan data. Teknik *clustering* dapat digunakan sebagai cara dalam mengelompokkan teks yang mempunyai kesamaan konten. Studi terdahulu menyatakan bahwa dengan metode *clustering*, kita dapat melakukan proses pengelompokan daerah yang padat, dan akhirnya menemukan pola-pola distribusi, serta mendapatkan pengetahuan tentang keterkaitan antar atribut data [2].

Dalam studi ini, hal yang ingin diselesaikan adalah cara mengelompokkan cabang bank berdasarkan produktivitasnya. Kantor pusat sangat kesulitan untuk memantau cabang-cabang yang produktif ataupun sebaliknya. Tanpa adanya pengelompokan cabang yang bersifat *profit* ataupun tidak *profit*, maka kantor pusat akan mengalami keterbatasan dalam mengambil keputusan, seperti saat ingin melakukan *branchless* atau pengurangan cabang yang terindikasi kurang produktif atau profit serta monitoring sumber daya manusia dan juga aset pada cabang yang bersangkutan. Salah satu metode yang dapat digunakan untuk menyelesaikan permasalahan tersebut adalah metode *K-Means*.

K-Means adalah salah satu algoritme *data mining* yang mampu melakukan pengelompokan secara partisi dan memisahkan data menjadi kelompok yang berbeda [3]. Apabila *K-Means* melakukan pengelompokan data ke dalam beberapa *cluster* dalam satu kelompok yang sama, maka akan memiliki karakteristik yang berbeda dengan kelompok lainnya [4]. Tujuan pengelompokan data tersebut agar fungsi objektif dapat diminimalkan. Dengan kata lain, variasi pada suatu kelompok akan diminimalkan dan variasi antar kelompok yang ada akan dimaksimalkan [5].

Berdasarkan studi sebelumnya oleh Rini dan Mujiati [6], disimpulkan bahwa beberapa cabang yang masuk dalam *cluster* tingkat rendah sebanyak 4 cabang, 15 cabang berada di *cluster* tingkat sedang, dan 11 cabang berada di *cluster* tingkat tinggi. Selanjutnya dilakukan perhitungan model regresi, sehingga hasilnya ada beberapa cabang yang tetap memiliki peluang tinggi masuk dalam kategori tingkat rendah. Oleh karena itu, perusahaan harus melakukan tindakan pada cabang-cabang yang masih berpeluang dalam tingkat rendah.

Dalam melakukan teknik *Clustering K-Means* pada kasus ini diharapkan bank pusat dapat mengambil keputusan dengan tepat serta menerima informasi cabang mana yang produktif ataupun sebaliknya supaya dapat meningkatkan efisiensi cabang.

II. METODOLOGI

Objek penelitian ini menggunakan data yang bersumber dari Bank XYZ. Pada kasus ini metode yang digunakan adalah KDD (*Knowledge Discovery in Database*).

A. *Knowledge Discovery in Database (KDD)*

KDD adalah metode untuk memperoleh pengetahuan baru dari data yang terkumpul, serta mencari hubungan yang terkait dalam tabel-tabel basis data yang saling berhubungan dan mendapatkan pengetahuan baru dari rangkaian proses penambangan untuk dasar pengambilan keputusan dan strategi.

Proses KDD memiliki beberapa tahap dimulai dari *data selection* hingga penemuan *knowledge* [7].

1) *Data Selection*

Proses seleksi data dari data-data yang terkumpul sangat perlu dilakukan agar bentuk pemodelan sesuai dengan metode yang digunakan. Proses seleksi data adalah *dataset* yang diolah dengan algoritme yang diusulkan, dan hasilnya dapat tersimpan dalam berkas terpisah dari basis data awal. Pada studi ini, data yang digunakan adalah data perbankan dari bulan Juli 2019 sampai bulan Juli 2020. Dari data tersebut terdapat beberapa variabel di antaranya, jumlah transaksi per cabang, jumlah nasabah per cabang, jumlah pinjaman per cabang, jumlah rekening per cabang, dll. Namun, data yang diseleksi untuk penerapan *data mining* yaitu jumlah jenis transaksi dan total transaksi. Berdasarkan hasil wawancara dengan manajemen kantor pusat, jumlah transaksi merupakan variabel yang paling tepat yang dapat mewakili besar atau kecilnya suatu cabang, begitu juga dengan jumlah jenis transaksi pada cabang.

2) *Preprocessing/Cleaning*

Proses *cleaning* dilakukan sebelum tahap *data mining* dilakukan. Perlunya proses *cleaning* tersebut bertujuan untuk mengolah data yang dijadikan fokus ke dalam KDD, serta mengeliminasi duplikasi data, menyelesaikan persoalan data yang bersifat inkonsisten, serta memperbaiki kesalahan pada data. Di sisi lain, *enrichment* juga perlu dilakukan karena berguna untuk mengolah informasi yang ada menjadi lebih relevan. Pada penelitian ini, data yang tidak diperlukan dan data yang kurang lengkap akan dilakukan proses *cleaning* supaya hasil pengelompokan dapat diperoleh secara maksimal.

3) *Transformation*

Proses transformasi pada data dilakukan agar data menjadi lebih sesuai untuk digunakan dalam proses *data mining*. Proses *coding* dilakukan sebagai respon untuk proses yang bergantung pada jenis informasi yang akan dicari dari basis data.

4) *Data Mining*

Data mining merupakan proses mencari pola atau pengetahuan yang menarik pada data-data yang sudah diseleksi sebelumnya. *Data mining* memiliki macam-macam teknik, metode, serta algoritme dalam melakukan prosesnya. Pemilihan metode atau algoritme *data mining* bergantung pada tujuan yang ingin dicapai serta proses pada KDD secara menyeluruh. Pada proses ini, penelitian dilakukan dengan menggunakan *RapidMiner* untuk mengelola data menggunakan teknik *Clustering K-Means*.

5) *Evaluation*

Tahap pengujian diperlukan untuk memastikan bahwa model yang terbentuk adalah yang terbaik. Tahap ini meliputi pemeriksaan apakah fakta yang dihasilkan dari proses pemodelan bertentangan dengan hipotesis atau fakta yang ada sebelumnya. Tahap ini akan menyertakan perbandingan data yang dihasilkan oleh Excel dan *RapidMiner*.

6) *Knowledge*

Pada tahap inilah pola atau struktur informasi dipresentasikan kepada pengguna. Pada tahap ini pengetahuan yang didapatkan bisa dimengerti oleh semua orang dan dapat dijadikan dalam acuan pengambilan keputusan. Data yang telah diolah akan menghasilkan suatu pengetahuan atau *knowledge* yang kemudian akan divisualisasikan menggunakan gambar.

B. *Data Visualization*

Proses penyajian data agar lebih mudah diterima dan dipahami adalah dengan melakukan visualisasi. Kondisi ini berkembang sejak visual berbentuk simbol-simbol mampu merepresentasikan suatu makna dari sesuatu yang disajikan. Visualisasi mengubah data menjadi informasi yang bisa dimengerti secara universal [8].

C. *Langkah Algoritme K-Means*

Algoritme *K-Means* adalah salah satu algoritme pengelompokan iteratif yang bekerja dengan cara mempartisi *dataset* ke sejumlah *K cluster* yang diinginkan [9]. Metode tersebut akan membagi data ke dalam beberapa kelompok di mana kelompok tersebut memiliki sifat atau karakteristik yang sama. Tujuan dari pengelompokan tersebut adalah meminimalisir keragaman dalam suatu kelompok dan memaksimalkan jenis dalam kelompok [10]. Adapun metode kerja algoritme *K-Means* adalah sebagai berikut:

- a. menetapkan nilai *k* sebagai jumlah *cluster* yang ingin dihasilkan,
- b. menetapkan pusat *cluster*,
- c. menghitung jarak setiap data ke pusat *cluster* menggunakan persamaan Euclidean,

$$d_{ik} = \sqrt{\sum_j^m (C_{ij} - C_{kj})^2} \tag{1}$$

- d. mengelompokkan data ke dalam *cluster* dengan jarak yang paling pendek menggunakan persamaan:

$$\min \sum_{k=1}^k d_{ik} = \sqrt{\sum_j^m (C_{ij} - C_{kj})^2} \tag{2}$$

- e. menghitung pusat *cluster* baru menggunakan persamaan:

$$C_{kj} = \frac{\sum_{i=1}^p X_{ij}}{P} \tag{3}$$

Di mana $X_{ij} \in$ *cluster* ke-*k* dan *P* adalah banyaknya anggota *cluster* ke-*k*.

- f. melakukan iterasi dari mulai langkah ke-2 hingga ke-4 sampai tidak ada lagi perpindahan anggota *cluster* yang terbentuk.

III. HASIL DAN PEMBAHASAN

A. *Implementasi Algoritme K-Means*

Pada tahap awal studi ini dilakukan simulasi manual untuk mengimplementasikan algoritme *K-Means*. Jumlah *cluster* 3, jumlah data 416 (cabang), dan jumlah atribut 2 (total transaksi 1 tahun dan jumlah jenis transaksi). *Dataset* penelitian dapat dilihat pada Tabel I.

1) *Penentuan pusat awal cluster*

Untuk penentuan awal dilakukan dengan mengambil data acak sebanyak 3 data, seperti diperlihatkan pada Tabel II:

2) *Perhitungan jarak pusat cluster*

Proses perhitungan jarak pusat *cluster* dengan setiap data menggunakan rumus *Euclidian Distance* dan menghasilkan matriks jarak dengan formula berikut:

$$d_{ik} = \sqrt{\sum_j^m (C_{ij} - C_{kj})^2}$$

C_{ij} adalah pusat kluster dan C_{kj} adalah data.

Berikut ini adalah simulasi perhitungan jarak dari data ke-1 terhadap pusat *cluster*.

$$C_1 = \sqrt{(1.778 - 4.226)^2}$$

$$C_1 = 2.448$$

$$C_2 = \sqrt{(1.995 - 22.973)^2}$$

$$C_2 = 21.195$$

$$C_3 = \sqrt{(2.280 - 71.341)^2}$$

$$C_3 = 69.563$$

TABEL I
CONTOH DATA TOTAL TRANSAKSI CABANG

Kode Cabang	Total Transaksi	Jumlah Jenis Transksi
7586	1.045	5
7608	1.072	5
7611	1.126	5
0800	1.245	4
7388	1.281	5
7630	1.281	4
7501	1.302	3
7400	1.303	5
7458	1.326	5
7538	1.479	3

TABEL II
PUSAT AWAL CLUSTER

Data	Total Transaksi	Jumlah Jenis Transksi
Data ke-100	4.226	10
Data ke-226	22.973	11
Data ke-356	71.341	11

Perhitungan dilanjutkan hingga data ke-*N* (416). Setelah itu, didapatkan matriks jarak terdekat sesuai contoh pada Tabel III.

Langkah selanjutnya adalah untuk menentukan jarak terdekat pada setiap *cluster* dengan menghitung selisih antar ketiga *cluster* yang ditunjukkan pada Tabel IV. Setiap data pada kolom C1, C2, dan C3 pada matrik menunjukkan nilai jarak data terhadap *centroid*. C1 merupakan nilai jarak data terhadap *centroid* pertama dan seterusnya.

1) *Pengelompokan data*

Dengan melakukan proses iterasi di atas, maka akan menghasilkan kelompok data seperti contoh pada Tabel V.

TABEL III
MARIKS JARAK TERDEKAT

Data	Total	Jenis	C ₁	C ₂	C ₃
7586	1.045	5	3.184	21.928	70.296
7608	1.072	5	3.156	21.901	70.269
7611	1.126	5	3.103	21.847	70.215
0800	1.245	4	2.981	21.728	70.096
7388	1.281	5	2.948	21.692	70.060
7630	1.281	4	2.948	21.692	70.060
7501	1.302	3	2.927	21.671	70.039
7400	1.303	5	2.926	21.670	70.038
7458	1.326	5	2.903	21.647	70.015
7538	1.479	3	2.750	21.494	69.862

TABEL IV
JARAK TERDEKAT PADA SETIAP CLUSTER

Data	Total	Jenis	C ₁	C ₂	C ₃	Jarak Terdekat
7586	1.045	5	3.184	21.928	70.296	3.184
7466	1.072	5	3.156	21.901	70.269	3.156
7608	1.072	5	3.103	21.847	70.215	3.103
7611	1.126	4	2.981	21.728	70.096	2.981
7251	1.170	5	2.948	21.692	70.060	2.948
7591	1.171	4	2.948	21.692	70.060	2.948
0800	1.245	3	2.927	21.671	70.039	2.927
7388	1.281	5	2.926	21.670	70.038	2.926
7630	1.281	5	2.903	21.647	70.015	2.903
7405	1.286	3	2.750	21.494	69.862	2.750

TABEL V
PENGELOMPOKAN DATA

Data	Total	Jenis	C ₁	C ₂	C ₃	Jarak Terdekat	Cluster
7586	1.045	5	3.184	21.928	70.296	3.184	1
7466	1.072	5	3.156	21.901	70.269	3.156	1
7608	1.072	5	3.103	21.847	70.215	3.103	1
7611	1.126	4	2.981	21.728	70.096	2.981	1
7251	1.170	5	2.948	21.692	70.060	2.948	1
7591	1.171	4	2.948	21.692	70.060	2.948	1
0800	1.245	3	2.927	21.671	70.039	2.927	1
7388	1.281	5	2.926	21.670	70.038	2.926	1
7630	1.281	5	2.903	21.647	70.015	2.903	1
7405	1.286	3	2.750	21.494	69.862	2.750	1

Berdasarkan contoh data matriks tersebut dihasilkan rangkuman data seperti pada Tabel VI.

2) *Penentuan pusat cluster baru*

Tahap berikutnya adalah menghitung pusat *cluster* baru berdasarkan data anggota setiap *cluster* yang telah didapatkan menggunakan formula berikut:

$$C_1 = \frac{2.448 + 2.231 + \dots + \text{data ke-}N}{416}$$

$$C_1 = 4.749$$

$$C_2 = \frac{21.195 + 20.978 + \dots + \text{data ke-}N}{416}$$

$$C_2 = 28.387$$

$$C_3 = \frac{69.563 + 69.346 + \dots + \text{data ke-}N}{416}$$

$$C_3 = 90.923$$

Dari perhitungan tersebut, maka didapatkan pusat *cluster* baru seperti yang ditunjukkan pada Tabel VII.

Iterasi selanjutnya akan dilakukan dengan cara yang sama sehingga tidak ada perubahan data yang terjadi pada suatu *cluster*.

3) *Implementasi RapidMiner*

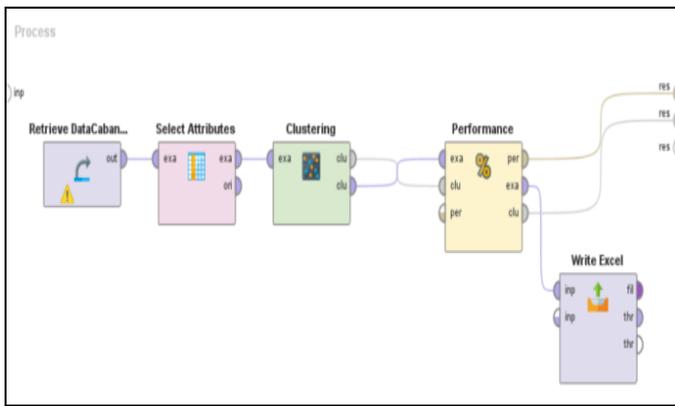
Pengelolaan data dengan menggunakan *K-Means* pada *RapidMiner* seperti yang diperlihatkan pada Gambar 1. Pada gambar tersebut digunakan pemodelan *Clustering K-Means* sehingga didapatkan inisialisasi jumlah *cluster* sebanyak 3 *cluster* sesuai dengan hasil pengolahan yang diperlihatkan pada Tabel VIII dan Tabel IX.

TABEL VI
JUMLAH CABANG PADA SETIAP CLUSTER

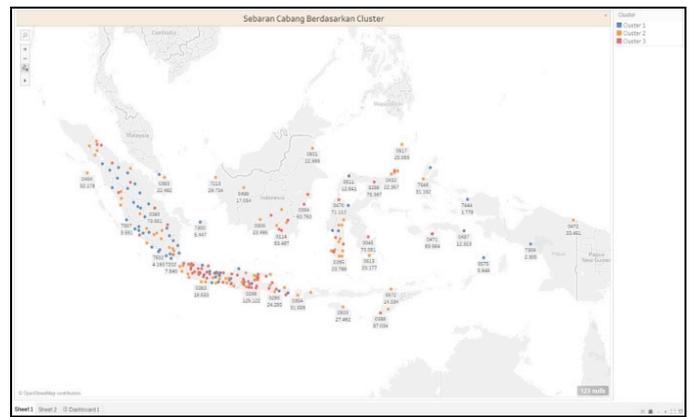
Cluster	Jumlah Cabang
1	173
2	139
3	104

TABEL VII
PUSAT CLUSTER BARU

Cluster	Jumlah Cabang	Total Transaksi	Jumlah Jenis Transaksi
1	173	4.749	6
2	139	28.387	9
3	104	90.923	10



Gambar 1 *Clustering K-Means* pada *RapidMiner*



Gambar 2 Sebaran data cabang

TABEL VIII
HASIL DATA *CLUSTER K-MEANS* PADA IMPLEMENTASI *RAPIDMINER*

Cluster	Jumlah Cabang
0	104
1	147
2	165

TABEL IX
HASIL DATA *CLUSTER K-MEANS* PADA IMPLEMENTASI *RAPIDMINER*

Attribute	Cluster 0	Cluster 1	Cluster 2
Total	90.922,75	3.516,04	2.5760,64

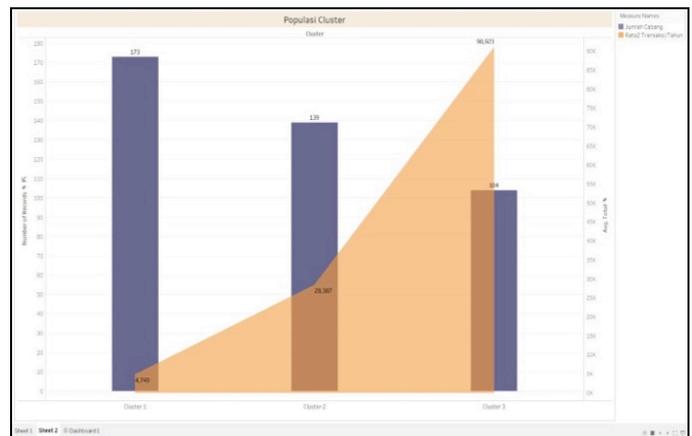
Proses *clustering* yang dilakukan dengan *RapidMiner* ini membutuhkan waktu 4 detik untuk data sebanyak 416 baris, dengan DBI atau Davies Bouldin Index 0,5.

B. Data Visualisasi

Data visualisasi yang digunakan dalam penelitian ini adalah menggunakan aplikasi *Tableau Business Intelligence*. Tujuan visualisasi ini adalah untuk memberikan gambaran secara *helicopter view* berdasarkan peta sebaran cabang di seluruh Indonesia. Pengguna informasi dapat melihat hasil pengelompokan yang dilakukan berdasarkan jumlah transaksi rata-rata per bulan dan berdasarkan jumlah jenis transaksi. Terdapat *cluster 1* atau kelompok cabang kecil dengan total cabang 173, *cluster 2* atau kelompok cabang sedang dengan total cabang 139, dan *cluster 3* atau kelompok cabang besar dengan total cabang 104.

Visualisasi yang pertama yaitu menggambarkan sebaran data cabang berdasarkan *geo tagging*. Data ditambahkan posisi *longitude* dan *latitude* pada setiap cabang dengan warna menyesuaikan *cluster* sehingga didapatkan hasil seperti pada Gambar 2.

Populasi cabang pada masing-masing cluster dibandingkan dengan rata-rata transaksi setiap bulan. *Cluster 1* terdapat 173 cabang dengan rata-rata transaksi sebesar 4.749 dan jumlah jenis transaksi 6, *cluster 2* terdapat 139 cabang dengan rata-rata transaksi sebesar 29.367 dan jumlah jenis transaksi 9, dan *cluster 3* terdapat 104 cabang dengan rata-rata transaksi sebesar 90.923 dan jumlah jenis transaksi 10, seperti diperlihatkan pada Gambar 3.



Gambar 3 Populasi cluster dengan rata-rata transaksi

IV. SIMPULAN

Hasil penelitian pada proses pengelompokan cabang yang dilakukan, baik secara manual maupun yang dilakukan menggunakan aplikasi *RapidMiner*, menghasilkan tiga kelompok dengan jumlah cabang setiap *cluster* tidak berbeda jauh. Adapun dari hasil proses pengelompokan tersebut menghasilkan kelompok cabang besar, kelompok cabang menengah, dan kelompok cabang kecil. Pengelompokan cabang berdasarkan jumlah transaksi per bulan dapat mempermudah manajemen dalam memberikan anggaran biaya operasional yang disesuaikan dengan kelompok cabang. Selain itu, juga dapat memberikan kemudahan dalam menentukan prioritas dalam hal pengawasan atau pemeriksaan dan memberikan keadilan dalam proses perhitungan kinerja karena perbandingan kinerjanya hanya dilakukan dalam satu kelompok. *Clustering* menggunakan metodologi *K-Means* terbukti dapat membantu dalam proses pengelompokan cabang sehingga kantor pusat atau manajemen berikutnya dapat merencanakan pengelolaan sesuai dengan jenis cabang berdasarkan pengelompokan.

DAFTAR REFERENSI

[1] W. Nurul Rohmawati, dkk, "Implementasi algoritme *K-Means* dalam pengklasteran mahasiswa pelamar beasiswa," *Jitter 2015*, vol. 1, no. 2, hlm. 62–68, 2015.

- [2] B. M. Metisen dan H. L. Sari, "Analisis *clustering* menggunakan metode *K-Means* dalam pengelompokan penjualan produk pada Swalayan Fadhila," *J. Media Infotama*, vol. 11, no. 2, hlm. 110–118, 2015.
- [3] M. G. Sadewo, A. P. Windarto, dan A. Wanto, "Penerapan algoritme *clustering* dalam mengelompokkan banyaknya desa/kelurahan menurut upaya antisipasi/mitigasi bencana alam menurut provinsi dengan *K-Means*," dalam *Komik (Konferensi Nas. Teknol. Inf. dan Komputer)*, vol. 2, no. 1, hlm. 311–319, 2018.
- [4] F. Nasari dan S. Darma, "Penerapan *K-Means Clustering* pada data penerimaan mahasiswa baru (studi kasus: Universitas Potensi Utama)," dalam *Semnasteknomedia Online*, vol. 3, no. 1, hlm. 6–8, 2015.
- [5] M. G. Sadewo, A. P. Windarto, dan D. Hartama, "Penerapan *data mining* pada populasi daging ayam ras pedaging di Indonesia berdasarkan provinsi menggunakan *K-Means Clustering*," *Infotekjar (Jurnal Nas. Inform. dan Teknol. Jaringan)*, vol. 2, no. 1, hlm. 60–67, 2017.
- [6] R. A. Riadi dan M. D. Kartikasari, "Implementasi *K-Means clustering* dan regresi logistik ordinal terhadap kinerja cabang PT X," dalam *Prosiding Sendika*, vol. 6, no. 1, hlm. 47–57, 2020.
- [7] Y. Mardi, "*Data mining*: klasifikasi menggunakan algoritme C4.5," *J. Edik Inform.*, vol. 2, no. 2, hlm. 213–219, 2017.
- [8] N. Ahmad Syaripul dan A. Mukharil Bachtiar, "Visualisasi data interaktif data terbuka Pemerintah Provinsi DKI Jakarta: topik ekonomi dan keuangan daerah," *J. Sist. Inf.*, vol. 12, no. 2, hlm. 15–29, 2016.
- [9] I. Parlina, A. P. Windarto, A. Wanto, dan M. R. Lubis "Memanfaatkan algoritme *K-Means* dalam menentukan pegawai yang layak mengikuti *assessment center* untuk *clustering* program SDP," *Computer Engineering, Science, and System Journal*, vol. 3, no. 1, hlm. 87–93, 2018.
- [10] F. E. M. Agustin, "Implementasi algoritme *K-Means* untuk menentukan kelompok pengayaan materi mata pelajaran ujian nasional (studi kasus: SMP Negeri 101 Jakarta)," *J. Tek. Inform.*, vol. 8, no. 1, hlm. 73–78, 2015.

Angga Ardhianto, kelahiran Jakarta, meraih Sarjana Komputer (S.Kom) di Universitas Budi Luhur. Saat ini sebagai karyawan swasta yang fokus dalam dunia perbankan sejak tahun 2003.

Bowo Relawanto, meraih Sarjana Komputer (S.Kom) di Universitas Mercu Buana, Karyawan swasta sebagai *programmer* dan *analyst* yang fokus pada pengembangan perangkat lunak pada aplikasi *website* dan aplikasi *mobile*.

Arief Wibowo, dosen senior (*Associate Professor*) di Fakultas Teknologi Informasi Universitas Budi Luhur - Jakarta Indonesia. Meraih gelar Doktor Ilmu Komputer dari Universitas Gadjah Mada, Indonesia. Bidang penelitiannya adalah *data mining/text mining, knowledge management, user behavior, dan acceptance of information technology*.