

# Implementasi Algoritma *K-Means* untuk Pengelompokan *Customer Churn* dalam Menentukan Strategi Pemasaran Bank

## *Implementation of the K-Means Algorithm for Customer Churn Segmentation in Developing Bank Marketing Strategies*

<sup>1</sup>Reva Nur Rahmadiana, <sup>2</sup>Dinda Lestarini\*

<sup>1,2</sup>Program Studi Sistem Informasi, Fakultas Ilmu Komputer, Universitas Sriwijaya

<sup>1,2</sup>Jl. Palembang- Indralaya KM.33, Indralaya, Sumatra Selatan, Indonesia

\*e-mail: [revanrrhmdna11@gmail.com](mailto:revanrrhmdna11@gmail.com), [dinda.lestarini@gmail.com](mailto:dinda.lestarini@gmail.com)

(received: 26 May 2025, revised: 26 May 2025, accepted: 29 May 2025)

### Abstrak

*Customer churn* atau hilangnya nasabah merupakan tantangan utama dalam industri perbankan karena dapat menimbulkan kerugian yang signifikan. Penelitian ini bertujuan untuk mengelompokkan nasabah berdasarkan karakteristik yang memengaruhi risiko *churn* menggunakan algoritma *K-Means*. Data yang digunakan merupakan data sekunder berupa dataset *bank customer churn* sebanyak 9.763 data yang diperoleh dari platform *Kaggle*, dan diproses mengikuti kerangka kerja *CRISP-DM*. Proses *clustering* dilakukan menggunakan aplikasi *RapidMiner* dengan evaluasi performa menggunakan *Davies Bouldin Index* untuk menentukan nilai *K* terbaik. Hasil menunjukkan nilai optimal diperoleh saat  $K = 4$ . Analisis *centroid* menunjukkan saldo (*balance*) dan estimasi gaji (*estimated salary*) merupakan variabel utama yang membentuk *cluster*. *Cluster 1* dan *cluster 3* memiliki jumlah nasabah *churn* tertinggi, *cluster 1* berisi nasabah dengan saldo tinggi namun gaji rendah, sedangkan *cluster 3* mencakup nasabah dengan saldo dan gaji tinggi. Hal ini menunjukkan bahwa tingginya saldo tidak selalu menjamin loyalitas pelanggan, dan faktor seperti penghasilan turut berperan penting dalam mencegah *customer churn*. Berdasarkan hasil ini, strategi yang disarankan meliputi edukasi layanan dan program loyalitas untuk *cluster 1*, serta layanan eksklusif dan pendekatan personal untuk *cluster 3*. Penelitian ini membuktikan bahwa metode *K-Means* efektif untuk menghasilkan segmentasi nasabah yang relevan sebagai dasar dalam merancang strategi pemasaran yang lebih terarah dan efisien.

**Kata kunci:** *customer churn, k-means, clustering, bank, segmentasi pelanggan*

### Abstract

*Customer churn, or the loss of banking clients, represents a major challenge in the banking industry due to its potential to cause significant financial losses. This study aims to segment customers based on characteristics that influence their churn risk using the K-Means algorithm. The data used in this research is secondary data consisting of 9,763 customer records from a bank customer churn dataset obtained via the Kaggle platform. The data processing follows the CRISP-DM framework. Clustering was conducted using RapidMiner, and performance was evaluated using the Davies-Bouldin Index to determine the optimal number of clusters (K). The results indicate that the optimal number of clusters is  $K = 4$ . Centroid analysis revealed that balance and estimated salary are the primary variables contributing to cluster formation. Cluster 1 and Cluster 3 had the highest number of churned customers. Cluster 1 consisted of customers with high balances but low salaries, while Cluster 3 included customers with both high balances and high salaries. These findings suggest that a high balance does not necessarily guarantee customer loyalty, and that income level plays an important role in preventing churn. Based on the analysis, recommended strategies include providing financial education and loyalty programs for customers in Cluster 1, and offering exclusive services and personalized approaches for those in Cluster 3. This study demonstrates that the K-Means algorithm is effective in producing relevant customer segmentation, serving as a valuable foundation for developing more targeted and efficient marketing strategies.*

**Keywords:** *customer churn, k-means, clustering, bank, customer segmentation*

<http://sistemasi.ftik.unisi.ac.id>

## 1 Pendahuluan

Seiring berkembangnya zaman keuangan secara perlahan menjadi posisi yang sangat penting di bank-bank komersial. Sebagai sumber keuntungan bank, nasabah dianggap sebagai aset bank yang paling berharga. Oleh karena itu, banyak bank yang menghabiskan energi dalam mencari nasabah baru dengan harapan dapat meningkatkan jumlah nasabah guna memperluas keuntungan, tetapi pada kenyataannya bank masih sering kali lalai dalam mempertahankan nasabah yang sudah ada namun pada kenyataannya bahwa mempertahankan nasabah yang sudah ada akan lebih menguntungkan dari pada mendapatkan nasabah baru[1].

*Churn* nasabah terjadi ketika nasabah memutuskan untuk tidak menggunakan layanan bank tersebut lagi. Fenomena ini dikenal dengan istilah *churn* karena nasabah memutuskan untuk mengakhiri hubungan dengan bank tertentu dan mencari pelayanan di bank lain[2], [3]. Sangat penting bagi bank untuk menganalisis *customer churn* serta mengembangkan strategi untuk mempertahankan pelanggan. Oleh karena itu, kehilangan *customer (churn)* menjadi masalah yang serius bagi bank karena membuat pihak bank harus menelan kerugian[4]. Dalam hal ini, bank akan fokus mempertahankan nasabah lama serta menarik pelanggan baru. Nasabah lama akan cenderung lebih loyal terhadap jasa yang ditawarkan karena sudah percaya dan memahami jasa yang digunakannya[5].

*Data mining* yaitu proses pengelolaan data dengan menggunakan kecerdasan buatan, matematika, teknik statistik dan *machine learning* untuk mengidentifikasi informasi yang relevan dan berguna dari *database* besar[6] [7]. *Machine learning* adalah penerapan kecerdasan buatan (AI) memberikan sistem kemampuan yang secara otomatis tanpa diprogram secara eksplisit untuk mempelajari dan meningkatkan kemampuannya berdasarkan pengalaman[8]. *Data mining* merupakan bagian dari proses penemuan pengetahuan dalam basis data dan mencakup beberapa langkah seperti pemilihan data, prapemrosesan, transformasi, penambangan data, dan evaluasi hasil[9]. Metode *data mining* yang dimanfaatkan dalam pengelompokan data *customer churn* tersebut ialah dengan teknik *clustering* yang merupakan pengelompokan data *customer churn* berdasarkan kemiripan data[1], [10]. *Clustering* yaitu teknik *unsupervised* yang tidak menunjukkan label pada dataset untuk membentuk sebuah kelompok, sehingga perlu dilakukan evaluasi terhadap data untuk mendapatkan sebuah hasil *clustering* [11]. Metode *K-Means* bertujuan untuk membagi sekumpulan N objek menjadi k kluster, di mana setiap gugus diwakili oleh nilai rata-rata objeknya[12]. Algoritma ini sederhana dan menyatu ke minima lokal dengan cepat[13] [14]. Untuk mempersiapkan data yang akan digunakan, dilakukan *data preparation* yang merupakan proses penyiapan data, mengatur data sehingga dapat digunakan pada *machine learning* yang digunakan.

Penelitian ini bertujuan untuk mengelompokkan lebih lanjut *bank customer churn* menggunakan algoritma *K-Means* berdasarkan kemiripan data guna membantu pengembangan segmentasi pelanggan dan strategi pemasaran yang lebih efektif. Dengan pendekatan kuantitatif serta metode analisis data sekunder, penelitian ini memanfaatkan teknik data mining dengan mengadopsi kerangka kerja *CRISP-DM* yang mencakup tahapan *business understanding* hingga *deployment*. Algoritma *K-Means* digunakan untuk mengeksplorasi dan mengelola data pelanggan secara efektif, sehingga hasil penelitian ini diharapkan memberikan kontribusi akademis dan praktis dalam pengembangan ilmu data mining serta solusi aplikatif bagi industri perbankan dalam menghadapi masalah *customer churn*.

## 2 Tinjauan Literatur

Penelitian mengenai pengelompokan *customer churn* menggunakan algoritma *K-Means* telah mengalami perkembangan pesat dalam beberapa tahun terakhir. Rahmati dan Wijayanto[15] menunjukkan bahwa algoritma *K-Means* mampu menghasilkan pembagian *cluster* yang lebih konsisten dan terstruktur dibandingkan dengan metode *clustering* lainnya seperti *Fuzzy C-means* dan *hierarchical clustering*. Melalui analisis indeks pembangunan manusia, mereka membuktikan bahwa *K-Means* efektif dalam membentuk kelompok homogen yang memudahkan pengambilan keputusan berbasis data. Relevansi temuan ini dalam konteks analisis *churn* sangat besar, mengingat *churn* sering kali dipengaruhi oleh variabel-variabel perilaku pelanggan yang tidak terdistribusi secara linear.

Dalam bidang perbankan, Zhang[1] mengembangkan pendekatan hibrida dengan menggabungkan *K-Means* dan *XGBoost* untuk meningkatkan akurasi prediksi *churn* pelanggan.

<http://sistemasi.ftik.unisi.ac.id>

Penelitian ini menunjukkan bahwa pengelompokan pelanggan sebelum tahap klasifikasi menghasilkan performa prediktif yang jauh lebih baik, karena pelanggan dengan profil risiko yang mirip dapat dianalisis dalam kelompok yang lebih homogen. Temuan ini menggarisbawahi pentingnya segmentasi berbasis perilaku dalam mengoptimalkan strategi retensi, terutama dalam industri yang memiliki basis pelanggan yang sangat heterogen seperti sektor keuangan.

Wicaksono[4] menegaskan temuan serupa dalam studi komparatif teknik klasifikasi untuk prediksi *churn*. Ia menemukan bahwa model-model klasifikasi yang diawali dengan tahap segmentasi pelanggan menunjukkan tingkat akurasi yang lebih tinggi dibandingkan model yang langsung mengklasifikasikan seluruh basis pelanggan secara seragam. Segmentasi berbasis *clustering* membantu mengungkapkan pola dalam perilaku pelanggan, seperti frekuensi transaksi, penggunaan produk tambahan, atau keterlibatan digital, yang sangat berpengaruh terhadap kecenderungan *churn*.

Sementara itu, Novianto[16] menerapkan *K-Means* dalam industri asuransi untuk mengelompokkan nasabah berdasarkan karakteristik bisnis dan risiko mereka. Dengan pendekatan ini, strategi pemasaran yang lebih personal dapat dirancang, meningkatkan efektivitas retensi dan akuisisi pelanggan. Hasil penelitian ini memperkuat pandangan bahwa pendekatan satu strategi untuk semua pelanggan tidak lagi efektif dalam dunia bisnis modern, termasuk dalam pengelolaan *churn*.

Penelitian Damanik dan Jambak[5] dalam industri telekomunikasi menggunakan pendekatan klasifikasi C4.5 untuk mengidentifikasi faktor *churn*, namun mereka juga menekankan pentingnya pemahaman segmentasi pelanggan untuk meningkatkan akurasi klasifikasi dan efektivitas retensi. Studi ini memperjelas bahwa tanpa pemetaan segmen pelanggan yang rinci, strategi intervensi *churn* menjadi kurang terarah dan kurang efektif.

Melengkapi temuan-temuan sebelumnya, Wibowo dan Rusindiyanto[2] secara khusus mengkaji *churn* nasabah bank dengan pendekatan *machine learning* dan *clustering*. Mereka menemukan bahwa dengan mengelompokkan nasabah berdasarkan perilaku transaksi, keterlibatan terhadap produk, dan riwayat interaksi layanan, bank dapat merancang program pemasaran dan retensi yang lebih presisi, menghasilkan peningkatan loyalitas nasabah yang nyata. Penelitian ini memberikan bukti kuat bahwa dalam konteks industri perbankan, teknik *clustering* seperti *K-Means* bukan hanya sekadar alat analitik, tetapi juga menjadi dasar penting untuk pengembangan strategi bisnis yang berbasis data.

Berdasarkan keseluruhan analisis literatur tersebut, dapat disimpulkan bahwa penerapan algoritma *K-Means* dalam mengelompokkan nasabah berdasarkan perilaku *churn* mampu mengungkapkan heterogenitas tersembunyi dalam populasi pelanggan. Pengelompokan ini memberikan dasar yang lebih kuat untuk merancang strategi pemasaran yang tersegmentasi dan personal, berbeda dari pendekatan konvensional yang cenderung seragam. Namun, masih terdapat celah dalam penelitian sebelumnya, yakni kurangnya eksplorasi mengenai bagaimana karakteristik spesifik dari masing-masing *cluster* dapat dijadikan dasar untuk pengembangan strategi pemasaran yang benar-benar *customized* dan berbasis kebutuhan nyata pelanggan di industri perbankan.

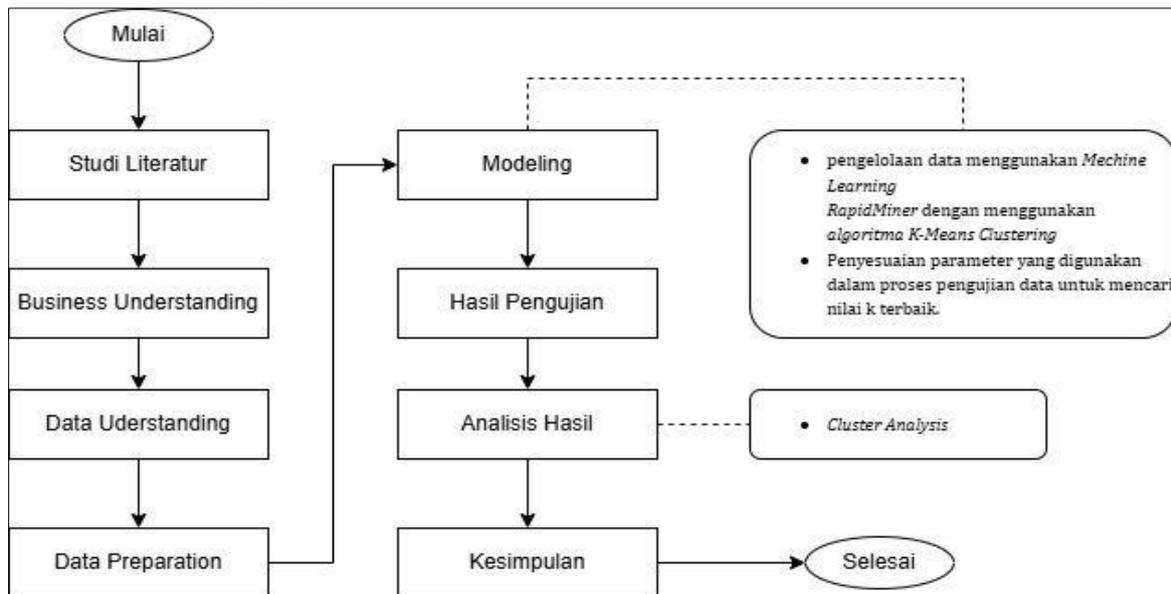
Oleh karena itu, penelitian ini mengembangkan hipotesis bahwa penerapan algoritma *K-Means* untuk pengelompokan *customer churn* akan menghasilkan segmentasi nasabah yang memiliki perbedaan karakteristik *churn risk* yang bermakna, dan bahwa pemanfaatan segmentasi ini dalam perumusan strategi pemasaran akan meningkatkan efektivitas program retensi nasabah. Secara lebih spesifik, penelitian ini berhipotesis bahwa setiap *cluster* yang terbentuk tidak hanya akan berbeda dalam tingkat kecenderungan *churn*, tetapi juga dalam variabel perilaku seperti frekuensi transaksi, penggunaan layanan tambahan, dan loyalitas terhadap produk bank. Dengan memahami secara mendalam profil masing-masing *cluster*, strategi pemasaran dapat disesuaikan secara dinamis, seperti penyusunan program loyalitas khusus, penawaran produk tambahan, atau peningkatan kualitas layanan bagi segmen berisiko tinggi. Dengan demikian, strategi pemasaran berbasis segmentasi diharapkan mampu tidak hanya menurunkan tingkat *churn* secara keseluruhan, tetapi juga meningkatkan *customer lifetime value* (CLV) secara berkelanjutan.

### 3 Metode Penelitian

Bagian ini menjelaskan metode yang digunakan dalam penelitian, yang mencakup pendekatan kuantitatif berbasis data sekunder dan penerapan kerangka kerja *CRISP-DM* yang dirancang untuk mendukung proses analisis *bank customer churn* dengan algoritma *K-Means* secara sistematis dan terstruktur. Pendekatan ini memungkinkan proses pengolahan data dilakukan secara menyeluruh,

mulai dari identifikasi permasalahan bisnis hingga pemanfaatan hasil analisis untuk mendukung pengambilan keputusan strategis dalam segmentasi pelanggan.

### 3.1 Desain Penelitian



Gambar 1. Tahap penelitian

Pada penelitian ini penulis menggunakan pendekatan kuantitatif dengan metode analisis data sekunder. Data yang digunakan merupakan data *bank customer churn* yang diperoleh dari *website kaggle* yang merupakan sebuah platform *data science* dan *machine learning*. Algoritma *K-Means* digunakan sebagai model utama dalam mengelompokkan *customer churn* pada bank berdasarkan kemiripan atau ketidakmiripan data yang ada. Berdasarkan Gambar 1 tahap penelitian ini mengadopsi kerangka kerja *CRISP-DM* karena kerangka kerja tersebut ditujukan untuk melakukan proses analisis dari suatu industri sebagai strategi pemecahan masalah dari bisnis atau unit penelitian, penelitian ini menggunakan kerangka kerja *CRISP-DM* tidak hanya pada tahap penelitian tetapi juga diterapkan sebagai standar proses data *mining*. Kerangka kerja *CRISP-DM* sendiri memiliki enam alur yaitu, *Business Understanding*, *Data Understanding*, *Data Preparation*, *Modeling*, *Evaluation* dan *Deployment*[17].

### 3.2 Data Understanding

Pada tahap *data understanding* ini, peneliti mengidentifikasi dan mengumpulkan data pelanggan dari sumber publik, yaitu dataset *bank customer churn* yang diperoleh dari platform *Kaggle*. Peneliti kemudian mengevaluasi dan memahami struktur data dengan menelaah setiap atribut, baik dari sisi relevansi maupun kelengkapan. Proses ini menjadi dasar dalam menentukan fitur-fitur penting untuk pemodelan menggunakan algoritma *K-Means*, sehingga pengelompokan yang dihasilkan dapat merepresentasikan perilaku nyata pelanggan dan mendukung strategi pemasaran yang lebih terarah.

### 3.3 Data Preparation

Data yang digunakan oleh penulis dalam proses penelitian ini diambil dari situs publik yaitu *kaggle* yang merupakan dataset *bank customer churn* dengan record data sebanyak 10.000 data yang terdiri dari 17 atribut [Bank Customer Churn \(kaggle.com\)](http://www.kaggle.com/datasets/bankcustomerchurn). Tabel 1 dan Tabel 2 menunjukkan informasi dataset yang akan digunakan penulis dalam proses penelitian.

Tabel 1. Informasi dataset

No	Atribut	Keterangan
1	<i>CustomerId</i>	Id pelanggan
2	<i>Surname</i>	Nama keluarga nasabah
3	<i>CreditScore</i>	Skor kredit nasabah
4	<i>Geography</i>	Lokasi negara nasabah

5	<i>Gender</i>	Jenis Kelamin
6	<i>Age</i>	Usia
7	<i>Tenure</i>	Berapa lama penggunaan kartu kredit
8	<i>Balance</i>	Saldo rata-rata nasabah
09	<i>NumOfProducts</i>	Jumlah produk bank yang digunakan nasabah
10	<i>HasCrCard</i>	Apakah nasabah memiliki kartu kredit atau tidak
11	<i>IsActiveMember</i>	Apakah nasabah aktif atau tidak
12	<i>EstimatedSalary</i>	Perkiraan gaji nasabah
13	<i>Exited</i>	Apakah nasabah
14	<i>Complain</i>	keluhan
15	<i>Satisfaction Score</i>	Kepuasan nasabah
16	<i>Card Type</i>	Jenis kartu
17	<i>Point Earned</i>	Poin yang diperoleh nasabah

**Tabel 2. Dataset bank customer churn**

<i>CustomerId</i>	<i>Surname</i>	<i>CreditScore</i>	<i>Geography</i>	<i>Gender</i>	...	<i>Point Earned</i>
15634602	<i>Hargrave</i>	619	<i>France</i>	<i>Female</i>	...	464
15647311	<i>Hill</i>	608	<i>Spain</i>	<i>Female</i>	...	456
15619304	<i>Onio</i>	502	<i>France</i>	<i>Female</i>	...	377
15701354	<i>Boni</i>	699	<i>France</i>	<i>Female</i>	...	350
15737888	<i>Mitchell</i>	850	<i>Spain</i>	<i>Female</i>	...	425
15574012	<i>Chu</i>	645	<i>Spain</i>	<i>Male</i>	...	484
...	...	...	...	...	...	...

Data yang digunakan penulis pada penelitian ini tidak dapat langsung digunakan dataset tersebut perlu dipersiapkan dahulu. Tujuan dari tahap ini yaitu untuk mempersiapkan data yang akan diolah menggunakan algoritma serta *tools* yang digunakan[18]. Selanjutnya dilakukan *data preparation* merupakan proses penyiapan data, mengatur data sehingga dapat digunakan pada *machine learning* yang digunakan. Pada penelitian ini penulis melakukan tahap data *preparation* dengan melakukan proses *cleaning data* yang sebelumnya *record* data sebanyak 10.000 data menjadi 9.763 data setelah melakukan proses *cleaning data*, serta penulis melakukan drop kolom yang tidak dibutuhkan dari 17 kolom terdapat 9 kolom yang tidak digunakan yaitu kolom *CustomerId*, *Surname*, *Geography*, *Gender*, *Exited*, *Complain*, *Satisfaction Score*, *Card Type*, dan *Point Earned*. Tabel 3 menunjukkan dataset setelah *preparation*.

**Tabel 3. Dataset setelah pereparation**

<i>Kredit Score</i>	<i>Age</i>	<i>Tenure</i>	<i>Balance</i>	<i>NumOfProducts</i>	...	<i>EstimatedSalary</i>
619	42	2	0	1	...	101348.9
608	41	1	83807.86	1	...	112542.6
502	42	8	159660.8	3	...	113931.6
699	39	1	0	2	...	93826.63
850	43	2	125510.8	1	...	79084.1
...	...	...	...	...	...	...

Setelah proses persiapan data telah dilakukan, hasil data *preparation* tersebut dapat digunakan dalam tahap pengujian menggunakan aplikasi *Rapidminer*.

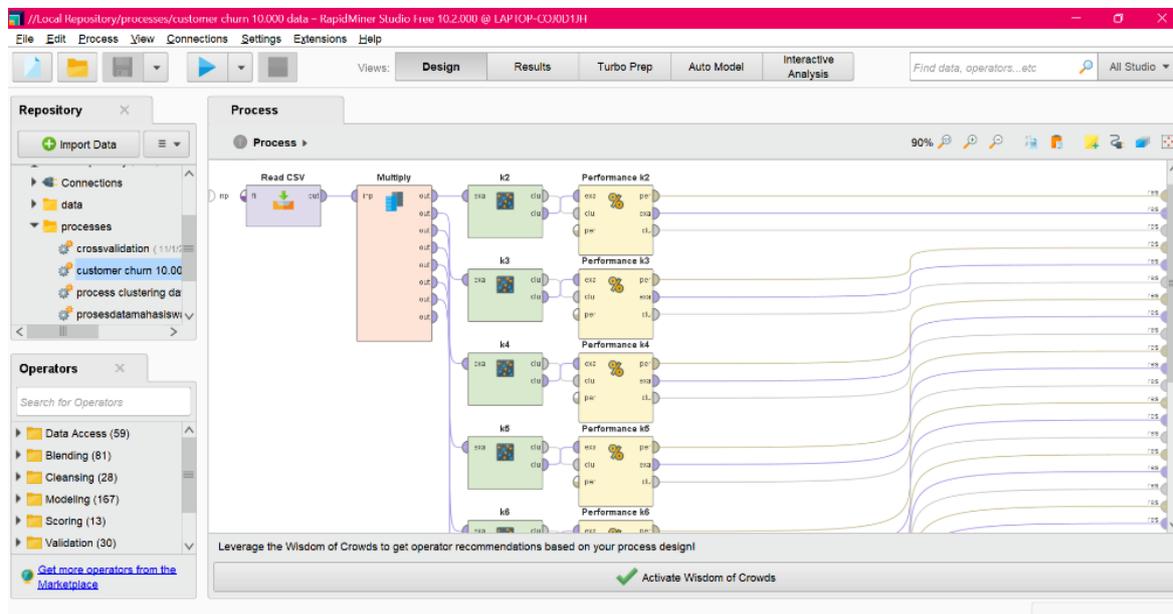
### 3.4 Modeling

Bagian ini menguraikan tahap pemodelan yang dilakukan menggunakan algoritma *K-Means* melalui perangkat lunak *RapidMiner*. Proses ini melibatkan pemilihan atribut penting, penentuan jumlah *cluster* optimal, serta pengaturan parameter untuk membentuk model segmentasi pelanggan yang representatif dan mendukung pengambilan keputusan strategis.

### 3.4.1 Rancangan Model Algoritma K-Means

Pada penelitian ini proses pengelolaan data akan menggunakan *machine learning rapidminer*. Untuk memproses data tersebut kita harus melakukan *import* data terlebih dahulu setelah itu memilih atribut apa saja yang akan diproses menggunakan *rapidminer*. Setelah proses *import* dataset selesai, selanjutnya proses modeling menggunakan algoritma *K-Means clustering* dapat dilihat pada Gambar 2, yang merupakan tampilan proses analisis data pada perangkat lunak *RapidMiner*, yang digunakan dalam penelitian pengelompokan *customer churn* menggunakan algoritma *K-Means Clustering* dengan beberapa tahapan. *Read CSV*, operator ini digunakan untuk mengimpor dataset *customer churn* dari file CSV. Dataset ini berisi informasi pelanggan bank yang mencakup berbagai atribut seperti *kredit score*, *age*, *tenure*, *balance*, *numofproduct*, *hascard*, *isactivemember*, dan *estimatedsalary*. *Multiply*, operator ini menggandakan dataset menjadi beberapa salinan untuk memungkinkan proses analisis paralel dengan berbagai jumlah *cluster* (*k*). Ini penting untuk membandingkan performa model *K-Means* dengan nilai *k* yang berbeda, guna menentukan jumlah *cluster* terbaik. *Operator K-Means clustering* dengan nilai *k* berbeda-beda, yaitu dari *k2* hingga *k8*, setiap operator ini mencoba membentuk *cluster* pelanggan berdasarkan kemiripan atribut-atribut yang ada, untuk mengidentifikasi pola atau kelompok yang berisiko *churn*. Setiap model *clustering* kemudian dievaluasi menggunakan operator *performance*, yang membantu dalam mengukur seberapa baik data dikelompokkan dan sejauh mana pemisahan antar *cluster* terjadi.

Tujuan dari proses ini adalah untuk mengeksplorasi nilai *k* yang paling optimal untuk segmentasi pelanggan, sehingga bank dapat memahami karakteristik masing-masing segmen dan merancang strategi retensi pelanggan yang lebih tepat sasaran. Proses ini juga mencerminkan tahapan *Modeling* dan *Evaluation* dalam kerangka kerja *CRISP-DM* yang digunakan dalam penelitian.



Gambar 2. Rancangan model algoritma *k-means*

### 3.2.1 Penyesuaian Parameter

Berikut adalah penyesuaian parameter yang akan digunakan penulis dalam proses pengujian data untuk mencari nilai *k* terbaik dapat dilihat pada Tabel 4.

Tabel 4. Parameter

Parameter	Measure Types	Divergence
1	MixedMeasure	MixedEcludienDistance
2	NumericalMeasure	DynamicTimeWarpingDistance
3	NumericalMeasure	CorrelationSimilarity
4	NumericalMeasure	ChebychevDistance
5	NumericalMeasure	JaccardSimilarity

Tabel 4 merupakan penyesuaian berbagai jenis parameter yang digunakan dalam proses pengujian untuk menentukan nilai  $k$  terbaik dalam algoritma  $K$ -Means. Masing-masing parameter menggunakan tipe pengukuran (*Measure Types*) dan metode pengukuran jarak atau kemiripan (*Divergence*) yang berbeda-beda, untuk menilai sejauh mana tiap metode dapat menghasilkan pengelompokan (*clustering*) yang optimal. Penggunaan berbagai metode ini bertujuan untuk mengevaluasi sejauh mana jenis ukuran dan metode *divergence* dapat memengaruhi kualitas hasil klasterisasi dalam konteks analisis *customer churn*.

#### 4 Hasil dan Pembahasan

Bagian ini menyajikan hasil dari implementasi model  $K$ -Means terhadap data nasabah bank, serta pembahasan mendalam mengenai pola pengelompokan yang terbentuk. Analisis dilakukan untuk mengidentifikasi karakteristik setiap cluster yang berkaitan dengan potensi *customer churn*, guna mendukung strategi pemasaran yang lebih terarah dan berbasis data.

##### 4.1 Pengujian Mencari Nilai K Terbaik

Pada penelitian ini penulis akan melakukan pengelompokan data bank *customer churn* dengan menggunakan algoritma  $K$ -Means.  $K$ -Means yaitu metode pengelompokan yang cepat dan sederhana serta memiliki kemampuan pengelompokan data yang terbilang cukup besar dalam waktu komputasi yang cepat dan efisien [19]. Pada pengujian ini dilakukan proses pengujian mulai dari nilai  $K=2$  sampai dengan  $K=8$ . Semakin rendah nilai *Davies Bouldin Index* (DBI) yang dihasilkan maka akan semakin baik pula pengelompokannya [20]. Dari hasil pengujian mencari nilai  $k$  yang paling baik terletak di  $k=4$  dengan hasil nilai DBI  $k=4$  yaitu 0.642 dengan parameter *MixedMeasure* dan *MixedEcludienDistance*. Hasil pencarian nilai  $K$  terbaik dapat dilihat pada Tabel 5.

Tabel 5. Hasil pengujian pencarian nilai K

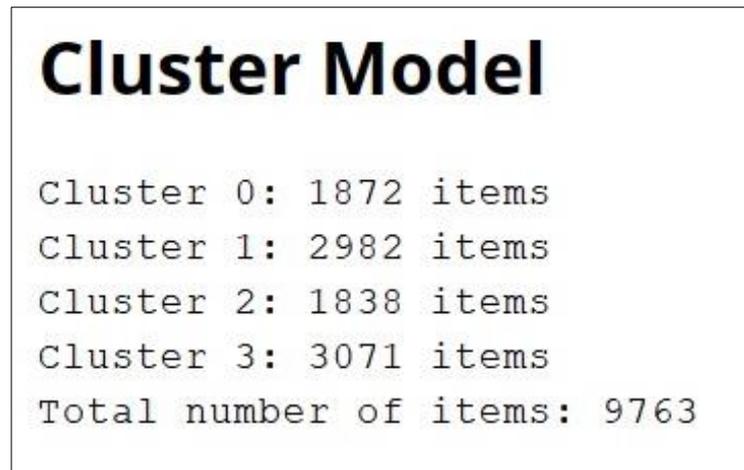
Parameter	Banyaknya K						
	K2	K3	K4	K5	K6	K7	K8
1	0.912	0.718	0.642	0.769	0.737	0.743	0.812
2	0.912	1.899	1.02	0.911	1.083	1.224	1.341
3	0.993	0.804	0.727	0.908	0.842	1.035	0.973
4	0.95	0.721	0.653	0.777	0.739	0.768	0.819
5	0.905	0.771	0.721	0.721	1.048	1.046	1.046

##### 4.2 Analisis Hasil Pengujian

Subbagian ini menyajikan analisis mendalam terhadap hasil pengujian model  $K$ -Means *clustering* yang telah dilakukan. Evaluasi difokuskan pada karakteristik setiap *cluster* yang terbentuk berdasarkan atribut-atribut utama. Analisis ini bertujuan untuk mengidentifikasi pola perilaku nasabah yang berisiko *churn* serta memberikan dasar yang kuat dalam perumusan strategi retensi pelanggan yang lebih efektif dan tepat sasaran.

##### 4.2.1 Hasil Pengujian Cluster Model

Hasil dari proses pengelompokan data menggunakan model *clustering K-Means* dengan menggunakan parameter *MixedMeasure* dan *MixedEcludienDistance* dapat dilihat pada Gambar 3.

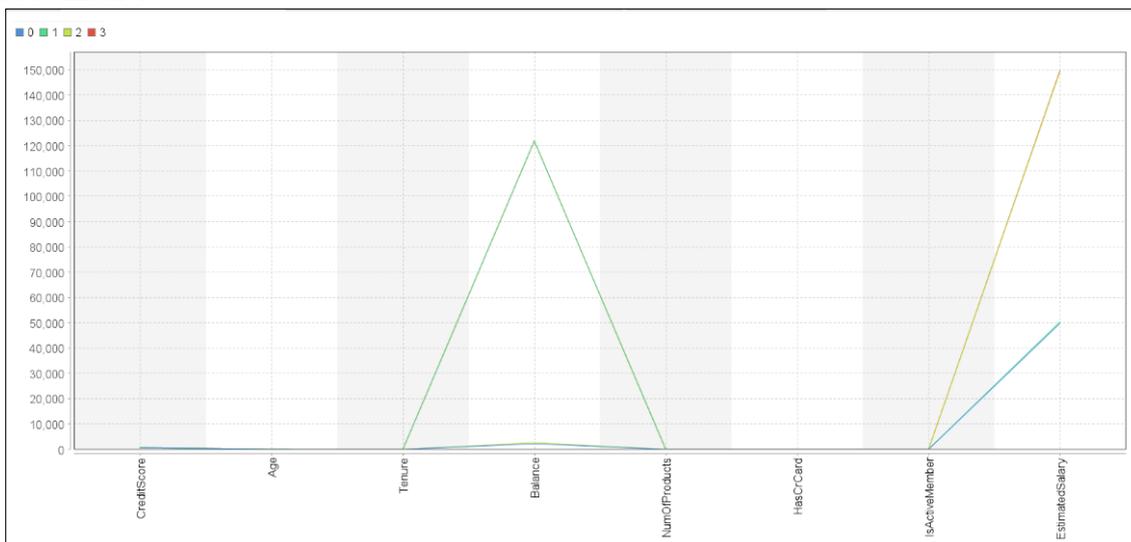


**Gambar 3. Hasil pengujian *cluster* model**

Dari hasil *cluster*, terbentuk empat *cluster* dengan jumlah item yang berbeda-beda. *Cluster* 0 terdiri dari 1.872 nasabah, *cluster* 1 memiliki 2.982 nasabah, *cluster* 2 sebanyak 1.838 nasabah, dan *cluster* 3 terdiri dari 3.071 nasabah. Total keseluruhan data yang berhasil dikelompokkan adalah sebanyak 9.763 item. Distribusi jumlah nasabah yang tidak merata pada tiap *cluster* mengindikasikan adanya karakteristik yang berbeda pada masing-masing kelompok. Hal ini yang akan dijadikan dasar untuk analisis lebih lanjut, seperti identifikasi perilaku pelanggan, segmentasi pasar, serta evaluasi risiko *customer churn* berdasarkan atribut-atribut yang dimiliki oleh setiap *cluster*.

#### 4.2.2 Plot Hasil Pengujian

Pada gambar plot hasil pengujian dapat dilihat bahwa *balance* dan *EstimatedSalary* yang mempengaruhi terbentuknya *cluster customer churn*, dapat dilihat pada grafik plot hasil pengujian pada Gambar 4.



**Gambar 4. Plot Hasil pengujian**

Diagram pengelompokan tersebut tidak menunjukkan secara rinci bagaimana pola karakteristik yang dihasilkan dari setiap kelompok. Oleh karena itu, untuk melihat secara rinci bagaimana pola karakteristik yang terbentuk pada setiap kelompok dapat dilihat pada hasil centeroid pada Gambar 5.

Attribute	cluster_0	cluster_1	cluster_2	cluster_3
CreditScore	649.485	651.678	648.436	650.946
Age	38.506	39.395	38.366	38.996
Tenure	5.069	4.964	5.075	5.009
Balance	2212.220	121916.541	2660.984	121857.506
NumOfProducts	1.771	1.369	1.764	1.401
HasCrCard	0.730	0.699	0.710	0.694
IsActiveMember	0.525	0.514	0.511	0.514
EstimatedSalary	49804.781	50404.249	149232.624	149692.774

**Gambar 5. Centroid hasil pengujian**

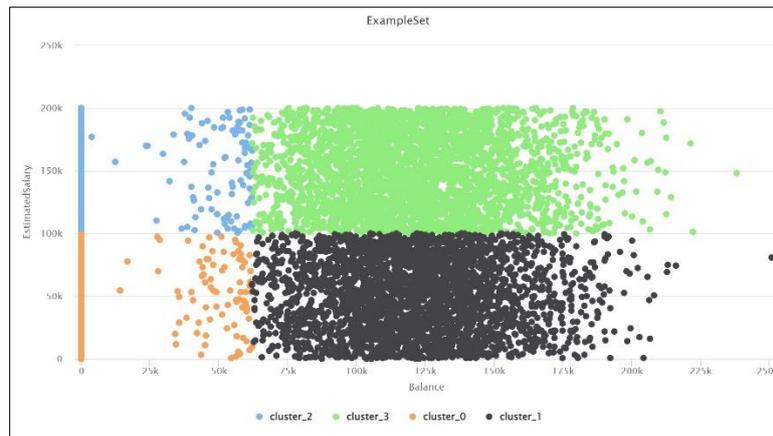
Setelah penerapan algoritma *K-Means*, *customer churn* terbagi menjadi 4 kelompok yang ditunjukkan pada Gambar 4 dan 6. *Cluster 0* terdiri dari pelanggan dengan saldo yang sangat rendah, serta estimasi gaji yang rendah. Hal ini menunjukkan bahwa pelanggan dalam *cluster* ini kemungkinan memiliki aktivitas keuangan yang rendah atau baru menggunakan layanan, dan bisa termasuk dalam kelompok yang berisiko *churn* karena kurangnya keterlibatan finansial. *Cluster 1* terdiri dari pelanggan dengan saldo rata-rata yang tinggi, tetapi memiliki estimasi gaji yang rendah. Pola ini dapat menunjukkan bahwa meskipun pelanggan memiliki simpanan besar, mereka mungkin kurang stabil secara penghasilan, atau merupakan pensiunan/investor. Pelanggan seperti ini bisa loyal, namun juga berisiko *churn* jika tidak diberikan layanan bernilai tambah. *Cluster 2* adalah kelompok pelanggan yang secara finansial paling kuat, dengan saldo dan estimasi gaji sama-sama tinggi. Mereka merupakan pelanggan bernilai tinggi (*high-value customers*), dan kemungkinan rendah untuk *churn* jika diberikan layanan yang sesuai. Kelompok ini ideal untuk diprioritaskan dalam program loyalitas. *Cluster 3* memiliki karakteristik serupa dengan *Cluster 2* dari sisi *EstimatedSalary*, namun memiliki sedikit lebih rendah saldo. Mereka tetap merupakan pelanggan bernilai tinggi, dan karena memiliki gaji tinggi, bisa jadi baru menabung atau memiliki arus kas aktif. Mereka cenderung stabil dan loyal, namun tetap penting untuk dimonitor keterlibatannya.

Berdasarkan delapan atribut yang digunakan dalam analisis, faktor saldo (*balance*) dan estimasi gaji (*estimated salary*) merupakan dua variabel yang paling signifikan dalam menentukan karakteristik nasabah. Nasabah dengan saldo besar cenderung memberikan kontribusi keuntungan yang lebih tinggi bagi bank, mengingat sumber pendapatan utama bank berasal dari suku bunga dan biaya layanan. Oleh karena itu, nasabah dengan potensi finansial yang tinggi perlu mendapatkan perhatian khusus dari pihak bank.

Salah satu *cluster* yang memerlukan penanganan strategis adalah *cluster 1*. Kelompok ini terdiri dari 2.982 nasabah dengan saldo rata-rata tinggi namun estimasi gaji yang rendah. Meskipun memiliki simpanan besar, rendahnya pendapatan dapat menunjukkan ketidakstabilan finansial, misalnya karena status sebagai pensiunan atau investor pasif. Tingginya jumlah nasabah dalam *cluster* ini yang mengalami *churn* menunjukkan adanya kebutuhan untuk meningkatkan keterlibatan mereka melalui penawaran layanan bernilai tambah yang disesuaikan dengan kebutuhan dan profil risiko masing-masing nasabah. Di sisi lain, *cluster 3* juga tidak boleh diabaikan. Kelompok ini terdiri dari 3071 nasabah dengan saldo dan estimasi gaji yang sama-sama tinggi. Karakteristik ini mengindikasikan bahwa nasabah pada *cluster* ini memiliki potensi untuk memberikan keuntungan jangka panjang bagi bank. Dengan demikian, baik *Cluster 1* maupun *Cluster 3* merupakan kelompok strategis yang dapat berkontribusi signifikan terhadap profitabilitas bank apabila dikelola dengan pendekatan yang tepat.

#### 4.2.3 Visualisasi Hasil Pengujian

Untuk memvisualisasikan hasil penelitian, penulis menerapkan *plot type scatter* yang dibuat menggunakan aplikasi *RapidMiner*, yang menyediakan fitur interaktif untuk menampilkan hasil klusterisasi secara grafis, sehingga membantu dalam interpretasi dan analisis data secara lebih intuitif. Visualisasi *scatter* ini menunjukkan sebaran data yang dihasilkan oleh setiap *cluster* mulai dari *cluster 0* hingga *cluster 3*, sehingga memudahkan dalam mengamati pola pengelompokan serta perbedaan antar *cluster* yang terbentuk dapat dilihat pada gambar 7.



Gambar 6. Visualisasi hasil pengujian

## 5 Kesimpulan

Algoritma *K-Means* digunakan dalam pengelompokan *customer churn* dengan nilai  $k$  terbaik yaitu  $k = 4$ , menggunakan parameter *MixedMeasure* dan *MixedEuclideanDistance*. Berdasarkan hasil pengujian dan analisis *centroid*, *cluster 1* dan *cluster 3* memiliki jumlah nasabah *churn* terbanyak, masing-masing 2.982 dan 3.071 nasabah. *Cluster 1* terdiri dari nasabah dengan saldo tinggi namun estimasi gaji rendah, mencerminkan nasabah pasif secara finansial seperti pensiunan, yang cenderung merasa tidak mendapatkan manfaat layanan. *Cluster 3* berisi nasabah dengan saldo dan estimasi gaji tinggi, namun lebih sensitif terhadap kualitas layanan dan ekspektasi. Hal ini menunjukkan bahwa saldo tinggi tidak menjamin loyalitas nasabah, karena faktor seperti penghasilan, keterlibatan, dan kepuasan juga berperan. Strategi untuk mengurangi *churn* pada *cluster 1* dan *3* dapat berupa edukasi dan peningkatan *engagement*, serta layanan program loyalitas eksklusif. Penelitian ini memiliki keterbatasan pada variabel yang digunakan, karena belum mempertimbangkan aspek kualitatif seperti pengalaman dan kepuasan pelanggan. Untuk penelitian selanjutnya, disarankan mengeksplorasi metode klusterisasi lain seperti DBSCAN serta mengintegrasikan model prediktif *machine learning* guna menghasilkan strategi retensi yang lebih akurat dan efektif.

## Referensi

- [1] T. Zhang, "Prediction and Clustering of Bank Customer Churn Based on XGBoost and K-means," 2022.
- [2] A.S. Wibow and Rusindiyanto, "Analisis Churn Nasabah Bank dengan Pendekatan Machine Learning dan Pengelompokan Profil Nasabah dengan Pendekatan Clustering," *Konstruksi: Publikasi Ilmu Teknik, Perencanaan Tata Ruang dan Teknik Sipil*, Vol. 2, No. 1, pp. 30–41, Jan. 2024, doi: 10.61132/konstruksi.v2i1.43.
- [3] N. Agian, S. Dinata, G. Abdurrahman, and N. Q. Fitriyah, "Perbandingan Optimasi Algoritma Random Forest menggunakan Teknik Boosting terhadap Kasus Klasifikasi Churn Pelanggan di Industri Telekomunikasi," *Jurnal Aplikasi Sistem Informasi dan Elektronika*, Vol. 5, No. 1, 2023.
- [4] A. Wicaksono, "Bianglala Informatika Uji Performa Teknik Klasifikasi untuk Memprediksi Customer Churn," Vol. 9, No. 1, p. 2021.
- [5] S. D. Damanik and M. I. Jambak, "Klasifikasi Customer Churn pada Telekomunikasi Industri untuk Retensi Pelanggan menggunakan Algoritma C4.5," *Media Online*, Vol. 3, No. 6, pp. 1303–1309, 2023, doi: 10.30865/klik.v3i6.829.
- [6] D. F. Pramesti, M. T. Furqon, and C. Dewi, "Implementasi Metode K-Medoids Clustering untuk Pengelompokan Data," 2017. [Online]. Available: <http://j-ptiik.ub.ac.id>
- [7] I. F. Tarigan, D. Hartama, and S. Saragih, "KLIK: Kajian Ilmiah Informatika dan Komputer Penerapan Data Mining pada Prediksi Kelayakan Pemohon Kredit Mobil dengan K-Medoids Clustering," *Media Online*, Vol. 1, No. 4, pp. 170–179, 2021, [Online]. Available: <https://djournals.com/klik>

- [8] T. M. Dista and F. F. Abdulloh, "Clustering Pengunjung Mall menggunakan Metode *K-Means* dan *Particle Swarm Optimization*," *JURNAL MEDIA INFORMATIKA BUDIDARMA*, Vol. 6, No. 3, p. 1339, Jul. 2022, doi: 10.30865/mib.v6i3.4172.
- [9] C. Zai and T. Komputer, "Implementasi Data Mining sebagai Pengolahan Data," 2022.
- [10] A. Ayu, D. Sulistyawati, and M. Sadikin, "Sistemasi: Jurnal Sistem Informasi Penerapan *Algoritma K-Medoids* untuk Menentukan Segmentasi Pelanggan Penerapan *Algoritma K-Medoids* untuk menentukan Segmentasi Pelanggan," 2021. [Online]. Available: <http://sistemasi.ftik.unisi.ac.id>
- [11] S. Paembonan, H. Abduh, and K. Kunci, "Penerapan Metode *Silhouette Coefficient* untuk Evaluasi *Clustering Obat Clustering*; *K-means*; *Silhouette coefficient*," 2021. [Online]. Available: <https://ojs.unanda.ac.id/index.php/jiit/index>
- [12] A. Febri, N. Ningsih, J. Lemantara, J. Raya Kedung Baruk No, and K. Rungkut, "Sistemasi: Jurnal Sistem Informasi Aplikasi Analisis Segmentasi Pelanggan untuk menentukan Strategi Pemasaran menggunakan Kombinasi Metode *K-Means* dan Model RFM," 2020. [Online]. Available: <http://sistemasi.ftik.unisi.ac.id>
- [13] R. Siagian, P. Sirait, and A. Halim, "Sistemasi: Jurnal Sistem Informasi Penerapan *Algoritma K-Means* dan *K-Medoids* untuk Segmentasi Pelanggan pada Data Transaksi *E-Commerce the Implementation of K-Means and K-Medoids Algorithm for Customer Segmentation on E-commerce Data Transactions*," 2022. [Online]. Available: <http://sistemasi.ftik.unisi.ac.id>
- [14] A. Fahim, "*K and Starting Means for K-Means Algorithm*," *J Comput SCI*, Vol. 55, 2021, doi: 10.1016/j.jocs.2021.101445.
- [15] R. Rizqina and W. A. Wahyu, "Analisis *Cluster* dengan *Algoritma K-Means*, *Fuzzy C-Means* dan *Hierarchical Clustering* (Studi Kasus: Indeks Pembangunan Manusia tahun 2019)," *JIKO (Jurnal Informatika dan Komputer)*, 2021.
- [16] R. Novianto, "Penerapan Data Mining menggunakan *Algoritma K-Means Clustering* untuk menganalisa Bisnis Perusahaan Asuransi," *JATISI (Jurnal Teknik Informatika dan Sistem Informasi)*, Vol. 6, No. 1, 2019, doi: 10.35957/jatisi.v6i1.150.
- [17] G. Radiana and A. Nugroho, "Analisis Sentimen berbasis Aspek pada Ulasan Aplikasi KAI Access menggunakan Metode *Support Vector Machine*," *Jurnal Pendidikan Teknologi Informasi (JUKANTI)*, Vol. 6, No. 1, 2023, doi: 10.37792/jukanti.v6i1.836.
- [18] F. Indriaharti Harida and N. Khazizah, "Analisis Cuaca di Kota Jakarta Bulan Januari Tahun 2018 menggunakan *Algoritma Decision Tree*," Vol. 14, No. 1, pp. 33–37, 2022, [Online]. Available: <https://www.kaggle.com/datasets/msf1203/pr>
- [19] S. Khairunnisa and M. I. Jambak, "Pengelompokan Cuaca Kota Palembang menggunakan *Algoritma K-Means Clustering* untuk mengetahui Pola Karakteristik Cuaca," *Jurnal Media Informatika Budidarma*, Vol. 6, No. 4, p. 2352, Oct. 2022, doi: 10.30865/mib.v6i4.4810.
- [20] P. S Ananda et al., "*KMeans Clustering* menggunakan *RapidMiner* dalam Segmentasi Pelanggan dengan Evaluasi *Davies Bouldin Index* untuk menentukan Jumlah *Cluster* paling Optimal," *Jurnal BATIRSI*, Vol. 6, No. 2, 2023.