

KLASTERISASI PENGUNJUNG MALL MENGGUNAKAN ALGORITMA K-MEANS BERDASARKAN PENDAPATAN DAN PENGELUARAN

Bima Yudha Putra ^{1*}, Fathya Yuanita Azzahra², Ira Alfrina Erlanda³

^{1,2,3} Institut Teknologi Telkom Purwokerto; Jl. DI Panjaitan No.128, Karangreja, Purwokerto Kidul, Kec. Purwokerto Sel., Kabupaten Banyumas, Jawa Tengah 53147; (0281) 641629

Riwayat artikel:

Received: 21 Juli 2023

Accepted: 19 Agustus 2023

Published: 11 September 2023

Keywords:

Clustering, Algoritma, K-Means, Mall, Jupyter Notebook.

Correspondent Email:

20103014@ittelkom-pwt.ac.id

Abstrak. Pada penelitian ini, dilakukan analisis clustering terhadap data pelanggan mall dengan tujuan untuk membantu manajer dan tim pemasaran dalam merumuskan strategi guna meningkatkan pendapatan mall. Langkah-langkah dalam penelitian ini meliputi studi literatur, analisis dataset, identifikasi masalah, dan menentukan karakteristik clustering. Data mining digunakan sebagai teknik untuk mendapatkan informasi yang bernilai dari data pelanggan mall yang memiliki volume besar dan bersifat tidak teratur. Dengan pemahaman yang mendalam tentang karakteristik pengunjung mall, strategi yang lebih efektif dapat dirancang untuk menarik pengunjung baru, meningkatkan retensi pengunjung yang sudah ada, dan menciptakan pengalaman belanja yang lebih menarik. Dataset yang digunakan dalam penelitian ini adalah dataset pelanggan mall yang diperoleh dari Kaggle. Dalam dataset ini, karakteristik pelanggan dianalisis berdasarkan tingkat pendapatan tahunan, usia, jenis kelamin, dan frekuensi kunjungan. Hasil clustering ini dapat membantu manajer dan tim pemasaran dalam mengidentifikasi kelompok pelanggan yang lebih disukai dan kurang disukai, serta mengevaluasi efisiensi dan efektivitas kegiatan bisnis perusahaan dalam memberikan layanan yang sesuai untuk berbagai jenis pelanggan dengan harga yang tepat.

Abstract. In this study, a clustering analysis was conducted on mall customer data with the aim of assisting managers and the marketing team in formulating strategies to increase mall revenue. The steps in this research include literature study, dataset analysis, problem identification, and determining clustering characteristics. Data mining is used as a technique to obtain valuable information from mall customer data which has a large volume and is irregular in nature. With a deep understanding of the characteristics of mall visitors, more effective strategies can be devised to attract new visitors, increase retention of existing visitors, and create a more engaging shopping experience. The dataset used in this study is the mall customer dataset obtained from Kaggle. In this dataset, customer characteristics are analyzed based on annual income level, age, gender, and frequency of visits. The results of this clustering can assist managers and marketing teams in identifying preferred and less preferred customer groups, as well as evaluating the efficiency and effectiveness of a company's business activities in providing appropriate services for various types of customers at the right price..

1. PENDAHULUAN

Di era saat ini yang sangat kompetitif, persaingan di pasar semakin ketat. Oleh karena itu, sangat penting bagi suatu perusahaan untuk memiliki keunggulan bersaing yang berkelanjutan guna mempertahankan atau meningkatkan posisinya serta mengoptimalkan peluang dan pertumbuhan laba yang signifikan. Ada banyak jenis keunggulan yang bisa dimiliki, tetapi salah satu yang paling efektif adalah mampu menarik dan mempertahankan klien atau pelanggan dengan baik. Dalam upaya ini, perusahaan harus fokus pada pelayanan prima, inovasi produk yang berkelanjutan dan membangun hubungan pelanggan jangka panjang yang kuat. Dengan memperhatikan faktor-faktor ini, perusahaan dapat menciptakan pengalaman pelanggan yang memuaskan, membangun loyalitas, dan meningkatkan citra merek mereka di pasar yang semakin kompetitif. Untuk mencapai hal tersebut, perusahaan harus memiliki pemahaman yang baik tentang para pelanggannya sehingga mereka dapat menyesuaikan strategi bisnis mereka untuk memberikan manfaat maksimal kepada pelanggan dan memastikan tingkat kepuasan yang tinggi. Kehadiran pelanggan secara menyeluruh memiliki nilai yang sangat berarti bagi perusahaan sebagai aset yang berharga. Oleh karena itu, perusahaan harus melakukan upaya khusus untuk memelihara hubungan jangka panjang dengan mereka. Namun demikian, tidak semua pelanggan akan memberikan dampak positif bagi perusahaan. Beberapa dari mereka mungkin tidak menguntungkan dalam jangka waktu tertentu. Dalam hal ini, perlu adanya evaluasi terhadap nilai ekonomi dari setiap pelanggan guna mengidentifikasi bagaimana cara terbaik untuk mengelola hubungan dengan mereka secara optimal [1].

Pemanfaatan Data Mining sebagai salah satu teknik untuk mendapatkan informasi yang bernilai telah menjadi umum dan luas digunakan oleh berbagai pihak. Pada masa kini, Data Mining juga menjadi sangat signifikan dalam berbagai sektor bisnis, termasuk industri retail. Data Mining merupakan suatu proses yang digunakan untuk mengidentifikasi pola berharga dari data yang berskala besar dan tidak terstruktur. Melalui proses ini, informasi yang sebelumnya tidak diketahui dapat digali dan

dipahami untuk memulai tindakan tindak lanjut yang tepat [2]. Pusat perbelanjaan modern, yang juga dikenal sebagai mall, merupakan tempat yang memiliki satu atau beberapa department store besar sebagai magnet bagi toko-toko kecil dan restoran. Mall memiliki karakteristik arsitektur tertentu, dimana toko-toko menghadap ke koridor utama atau area pejalan kaki yang menjadi pusat dari pusat perbelanjaan modern. Koridor ini berfungsi sebagai jalur lalu lintas dan juga ruang komunal yang memfasilitasi interaksi antara pengunjung dan penjual [3].

Segmentasi adalah suatu strategi yang dipergunakan untuk memecah kelompok konsumen menjadi bagian-bagian yang berbeda, dengan maksud untuk mengenal pasti ciri-ciri, perilaku, atau kebutuhan yang berbeda di antara mereka [4]. Maksud utama dari segmen ini adalah untuk mengidentifikasi target konsumen secara optimal. Segmen dapat dikerjakan dengan menggunakan metode k-means, seperti yang dicontohkan dalam studi yang dikerjakan oleh Dhanachandra dan rekannya. Dalam riset mereka, para peneliti menerapkan algoritma k-means untuk melaksanakan segmentasi. Algoritma k-means dipilih karena dianggap efisien dalam penggunaannya. Peneliti juga merekomendasikan penggunaan metode pencarian k-ideal, seperti elbow method dan metode lainnya, untuk mempercepat penentuan centroid dan meningkatkan efisiensi dalam proses segmentasi [5].

Pendapatan mall terus menurun seiring dengan perkembangan e-commerce yang pesat. Namun, dengan menganalisis beberapa kelompok pengunjung, kita dapat mengidentifikasi karakteristik unik dari setiap pengunjung mall tersebut. Informasi mengenai karakteristik ini akan kita gunakan untuk meningkatkan pendapatan mall melalui strategi yang sesuai dengan preferensi dan kebutuhan pengunjung. Dengan pemahaman yang mendalam tentang karakteristik pengunjung mall, kita dapat merancang strategi yang lebih efektif untuk menarik pengunjung baru, meningkatkan retensi pengunjung yang sudah ada, dan menciptakan pengalaman belanja yang lebih menarik. Dengan mengelompokkan pelanggan ke dalam cluster, kita dapat

memisahkan pelanggan utama dan kurang populer. Setelah segmentasi pelanggan ini, auditor dapat menilai efektivitas dan efisiensi bisnis perusahaan saat ini dalam memberikan layanan yang tepat kepada berbagai jenis pelanggan dengan harga yang tepat. Tujuan utamanya adalah untuk menilai apakah diskon telah ditawarkan dengan tepat untuk berbagai jenis pelanggan. Untuk mencapai tujuan tersebut, teknik pengelompokan yang disebut clustering sering digunakan untuk membagi pelanggan ke dalam kelompok-kelompok dengan karakteristik serupa. Setelah dataset melalui tahap preprocessing, data dapat dijalankan dalam proses Clustering.

2. TINJAUAN PUSTAKA

2.1 Dataset

Dataset adalah kumpulan data yang disimpan dalam memori, yang memiliki semua karakteristik, fitur, dan fungsionalitas dari basis data standar. Dataset dapat terdiri dari beberapa tabel, dimana tabel-tabel tersebut dapat saling berhubungan (relationship). Tabel dalam kumpulan data dapat menggunakan kunci asing dan mempertahankan integritas referensial. Dataset merupakan objek yang mewakili data dan hubungannya dalam penyimpanan. Struktur mirip dengan data dalam database. Dataset berisi kumpulan tabel dan data [6]. Ada dua jenis dataset, yaitu:

1. Private Dataset: Merupakan kumpulan data tersebut berasal dari organisasi atau komunitas yang menjadi subjek penelitian. Contohnya termasuk agensi, rumah sakit, pabrik, perusahaan jasa, dll.
2. Public Dataset: Merupakan kumpulan data yang dapat diakses melalui repositori publik yang disetujui oleh para peneliti. Contohnya termasuk data yang tersedia dari UCI (University of California, Irvine) dan sumber.

2.2 Clustering

Clustering adalah metode untuk mengidentifikasi kelompok umum dalam kumpulan data. Unit di setiap grup biasanya lebih mirip satu sama lain daripada unit di grup lain. Sejak tahun 1970-an, segmentasi berbasis cluster telah menikmati popularitas besar dalam penelitian data, khususnya di bidang pemasaran. Meskipun pengelompokan

tidak termasuk dalam metode analisis data terstruktur, namun sangat fleksibel dan sangat bergantung pada jenis data atau sampel yang digunakan [7].

2.3 Metode Elbow

Metode Elbow digunakan untuk mendapatkan informasi dalam menentukan jumlah kluster yang optimal dengan membandingkan persentase hasil antara jumlah kluster yang akan membentuk sudut siku pada suatu titik [8].

Berikut adalah langkah-langkah algoritma metode Elbow dalam menentukan nilai k pada K-Means [8]:

1. Mulai dengan menginisialisasi nilai awal k.
2. Tingkatkan nilai k.
3. Hitung nilai sum of square error (SSE) untuk setiap nilai k.
4. Analisis hasil SSE untuk mencari penurunan yang signifikan.
5. Cari dan tentukan nilai k yang membentuk sudut siku pada grafik hasil SSE, karena nilai tersebut akan menjadi pilihan terbaik untuk jumlah kluster yang optimal.

2.4 K-Means

K-Means adalah salah satu metode data mining yang digunakan untuk mengelompokkan atau mengelompokkan data. Pada metode ini, data dibagi menjadi satu atau lebih cluster, dimana data yang memiliki karakteristik yang sama dikelompokkan dalam satu cluster yang sama sedangkan data dengan karakteristik yang berbeda dikelompokkan dalam cluster yang berbeda [9]. Sarwono menjelaskan algoritma K-Means secara rinci, yang dapat disimpulkan sebagai berikut [10]:

1. Pilih nilai k sebagai jumlah kluster yang diinginkan.
2. Inisialisasi pusat centroid secara acak sebanyak k, untuk menghitung jarak setiap data masukan terhadap setiap centroid menggunakan rumus jarak Euclidean.

$$d(x_i, \mu_j) = \sqrt{\sum (x_i, \mu_j)^2}$$

Keterangan :

x_i : Data Kriteria

μ_j : Centroid pada cluster ke -j

- Setiap data dipisahkan berdasarkan kedekatannya dengan centroid, yaitu dengan mencari jarak terpendek.
- Nilai centroid baru diupdate dengan mengambil rata-rata dari setiap kluster yang terkait menggunakan rumus sebagai berikut :

$$\mu_j(t + 1) = \frac{1}{N_{sj}} \sum_{j \in s_j} x_j$$

Keterangan :

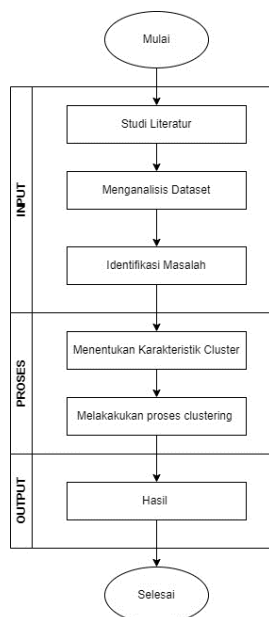
$\mu_j(t + 1)$: centroid baru pada iterasi (t+1)

N_{sj} : Data pada cluster S_j

- Jika data dalam setiap cluster masih mengalami perubahan, maka langkah 2 hingga 5 akan diulang berulang kali hingga tidak ada lagi perubahan anggota dalam setiap cluster.

3. METODE PENELITIAN

Penelitian ini terdiri dari tiga tahap utama, yakni tahap masukan, tahap proses, dan tahap keluaran. Setiap tahapan memiliki urutan kegiatan pada Gambar 3.1 di bawah ini menunjukkan rangkaian langkah-langkah dalam penelitian.



Gambar 1. Alur Penelitian

3.1 Studi Literatur

Studi literatur pada penelitian ini bertujuan untuk memperkuat pemilihan metode penelitian dan tema yang dipilih dengan mencari referensi literatur yang berkaitan dengan metode K-Means dengan bahan referensi yang bersumber dari jurnal, website, dan internet, serta sumber yang terkait sehingga diharapkan dapat membantu dalam penyusunan penelitian ini.

3.2 Menganalisis Dataset

Pada langkah ini, peneliti melakukan analisis terhadap dataset yang diperoleh dari Kaggle sebagai materi penelitian. Dataset ini berbentuk file CSV yang akan diproses pada tahap berikutnya. Dataset yang telah dianalisis sesuai dengan tema penelitian, yaitu tentang pelanggan mall.

3.3 Identifikasi Masalah

Proses identifikasi masalah dimulai dengan pemahaman terhadap kebutuhan objek penelitian yang akan dilakukan, serta memahami hasil dari permasalahan yang ada pada objek tersebut melalui pelaksanaan penelitian. Tahap ini peneliti mulai mencari referensi dari penelitian sebelumnya terkait dengan penggunaan metode penelitian yang sama.

3.4 Menentukan Karakteristik Clustering

- Karakteristik pelanggan akan dievaluasi berdasarkan kriteria-kriteria berikut: Tingkat pendapatan tahunan yang diperoleh (gaji)
- Jumlah pengeluaran yang dihabiskan untuk berbelanja di mall

3.5 Melakukan Proses Clustering

Proses pengklusteran dilakukan menggunakan Jupyter Notebook dengan menggunakan bahasa pemrograman Python. Langkah pertama dalam pengklusteran adalah mencari nilai K yang optimal menggunakan metode Elbow.

3.6 Hasil

Hasil dari proses clustering akhir akan digunakan sebagai acuan untuk memahami karakteristik pelanggan mall. Karakteristik tersebut akan dimanfaatkan untuk menyediakan layanan kepada pelanggan, seperti memberikan diskon, dengan harapan dapat meningkatkan penjualan dan mendapatkan keuntungan tambahan bagi mall tersebut.

4. HASIL DAN PEMBAHASAN

4.1. Clustering Pengunjung Mall

4.1.1 Dataset

Terdapat sebuah dataset pengunjung mall dalam format csv yang diambil dari Kaggle. Dataset ini terdiri dari 200 data dengan 5 kolom, yaitu ID_Pelanggan, Gender, Usia, Pendapatan_Tahunan_Ribuan_USD, dan Pengeluaran_USD. Dalam proses pengklusteran, fokus akan diberikan pada kolom Pendapatan_Tahunan dan Pengeluaran. Untuk melihat dataset secara lengkap, dapat dilihat dalam gambar berikut:

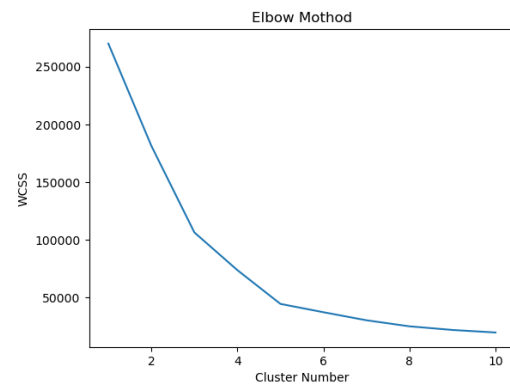
ID_Pelanggan	Gender	Usia	Pendapatan_Tahunan_Ribuan_USD	Pengeluaran_USD
0	1	Pria	19	15
1	2	Pria	21	15
2	3	Wanita	20	16
3	4	Wanita	23	16
4	5	Wanita	31	17
...
195	196	Wanita	35	120
196	197	Wanita	45	126
197	198	Pria	32	126
198	199	Pria	32	137
199	200	Pria	30	137

200 rows x 5 columns

Gambar 2. Dataset Awal

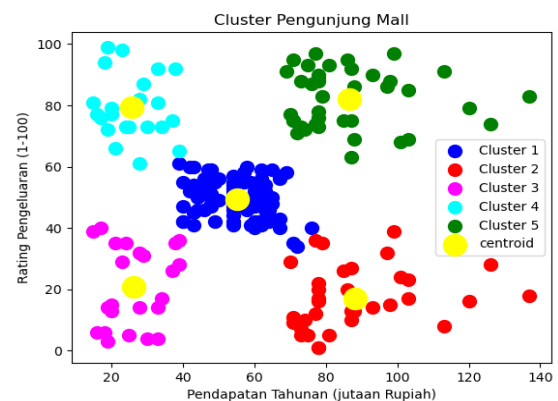
4.1.2 Proses Clustering

Proses pengklusteran dilakukan menggunakan Jupyter Notebook dengan menggunakan bahasa pemrograman Python. Langkah pertama dalam pengklusteran adalah mencari nilai K yang optimal menggunakan metode Elbow, dengan Mengatur random state menjadi 42. Metode elbow adalah teknik yang digunakan dalam pengklusteran untuk menentukan jumlah kluster yang optimal dalam suatu dataset. [11]. Metode ini berdasarkan pada grafik visual yang menunjukkan penurunan kecepatan penjelasan variasi dalam kluster seiring dengan penambahan jumlah kluster yang digunakan. Grafik tersebut biasanya berbentuk seperti lengkungan siku yang menyerupai siku tangan manusia, mirip dengan siku pada tangan yang membentuk sudut elbow. Di titik tersebut, penurunan penjelasan variasi menjadi lebih lambat dan tidak signifikan dengan adanya penambahan kluster lebih lanjut. Oleh karena itu, nilai K yang berada di titik elbow ini dianggap sebagai jumlah kluster yang optimal untuk dataset yang sedang dianalisis.



Gambar 3. Grafik K Optimal

Nilai K yang optimal ditentukan ketika dalam metode Elbow, grafik eksperimen menunjukkan penurunan yang signifikan dan membentuk sudut yang tajam. Dalam konteks ini, setelah memeriksa grafik percobaan (Gambar 2), ditemukan bahwa jumlah cluster yang ideal adalah 5. Untuk proses pengklusteran, metode K-Means digunakan, dan hasil akhirnya adalah pembagian data menjadi 5 cluster yang masing-masing ditandai dengan warna merah, biru, hijau, cyan, dan magenta. Gambar 3 menampilkan hasil pengklusteran tersebut.



Gambar 4. Hasil Clustering

Setelah proses optimasi clustering, telah diketahui titik centroid untuk setiap cluster. Titik centroidnya adalah sebagai berikut:

- Cluster 1 : [55.2962963 49.51851852]
- Cluster 2 : [88.2 17.11428571]
- Cluster 3 : [26.30434783 20.91304348]
- Cluster 4 : [25.72727273 79.36363636]
- Cluster 5 : [86.53846154 82.12820513]

5. KESIMPULAN

Berdasarkan hasil clustering yang telah diperoleh, pelanggan mall dapat

dibedakan menjadi 5 karakteristik yang berbeda-beda.

1. Karakteristik pertama adalah kelompok berwarna biru, yang menunjukkan bahwa anggota kelompok ini memiliki pendapatan tinggi dan skor pengeluaran yang tinggi pula.
2. Karakteristik kedua adalah kelompok berwarna merah, yang menunjukkan bahwa anggota kelompok ini memiliki pendapatan tinggi dan skor pengeluaran yang rendah.
3. Karakteristik ketiga adalah kelompok berwarna magenta, yang menunjukkan bahwa anggota kelompok ini memiliki pendapatan rendah dan skor pengeluaran yang rendah.
4. Karakteristik keempat adalah kelompok berwarna cyan, yang menunjukkan bahwa anggota kelompok ini memiliki pendapatan tinggi dan skor pengeluaran yang rendah.
5. Karakteristik kelima adalah kelompok berwarna hijau, yang menunjukkan bahwa anggota kelompok ini memiliki pendapatan tinggi dan skor pengeluaran yang juga tinggi.

Hasil clustering tersebut dianalisis dan menghasilkan 5 karakteristik yang dapat digunakan untuk membantu manajer dan tim pemasaran dalam merumuskan strategi guna meningkatkan pendapatan mall. Data yang telah diperoleh tidak menutup kemungkinan untuk mengalami perubahan seiring berkembangnya zaman. Terlebih lagi, manusia memiliki sifat yang dapat berubah-ubah sesuai dengan lingkungan dan kondisi tertentu.

UCAPAN TERIMA KASIH

Kami bersyukur dan mengucapkan terima kasih kepada Tuhan Yang Maha Esa atas berkah, rahmat, dan petunjuk-Nya dalam penelitian ini. Kami juga ingin menyampaikan rasa terima kasih kepada Kaggle, sebuah website yang telah menyediakan data yang berharga untuk penelitian kami. Ketersediaan data dari Kaggle sangat membantu dalam mengumpulkan informasi yang diperlukan.

DAFTAR PUSTAKA

- [1] 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.

- [2] G. Schuh *Et Al.*, "Data Mining Definitions And Applications For The Management Of Production Complexity," In *Procedia Cirp*, Elsevier B.V., 2019, Pp. 874–879. Doi: 10.1016/J.Procir.2019.03.217.
- [3] R. Savitri, "Pusat Perbelanjaan Modern (Mall) Dengan Penekanan Ruang Terbuka Publik," 2018.
- [4] G. Febrina Wulandari, "Segmentasi Pelanggan Menggunakan Algoritma K-Means Untuk Customer Relationship Management (Crm) Pada Hijab Miulan," 2020.
- [5] N. Dhanachandra, K. Mangle, and Y. J. Chanu, "Image Segmentation Using K-means Clustering Algorithm and Subtractive Clustering Algorithm," in *Procedia Computer Science*, Elsevier, 2015, pp. 764–771. doi: 10.1016/j.procs.2015.06.090.
- [6] M. Honda Lombok Timur Nusa Tenggara Barat Yahya, "Penggunaan Algoritma K-Means Untuk Menganalisis Pelanggan Potensial Pada Dealer SPS," *Jurnal Informatika dan Teknologi*, vol. 2, no. 2, pp. 109–118, 2019.
- [7] O. Dogan, J. L. Bayo-Monton, C. Fernandez-Llata, and B. Oztaysi, "Analyzing of gender behaviors from paths using process mining: A shopping mall application," *Sensors (Switzerland)*, vol. 19, no. 3, Feb. 2019, doi: 10.3390/s19030557.
- [8] N. Putu, E. Merliana, And A. J. Santoso, *Prosiding Seminar Nasional Multi Disiplin Ilmu & Call For Papers Unisbank (Sendi_U) Kajian Multi Disiplin Ilmu Untuk Mewujudkan Poros Maritim Dalam Pembangunan Ekonomi Berbasis Kesejahteraan Rakyat Analisa Penentuan Jumlah Cluster Terbaik Pada Metode K-Means Clustering*. 2015.
- [9] M. Benri, H. Metisen, And S. Latipa, "Analisis Clustering Menggunakan Metode K-Means Dalam

- Pengelompokkan Penjualan Produk Pada Swalayan Fadhila,” 2015.
- [10] N. W. Rohmawati, S. Defiyanti, M. Jajuli, And T. H. Informatika Fakultas Ilmu Komputer Universitas Singaperbangsa Karawang Jl Ronggowaluyo Telukjambe Timur Karawang, “Implementasi Algoritma K-Means Dalam Pengklasteran Mahasiswa Pelamar Beasiswa,” 2015.
- [11] V. Annisa Ekasetya And A. Jananto, “Klusterisasi Optimal Dengan Elbow Method Untuk Pengelompokan Data Kecelakaan Lalu Lintas Di Kota Semarang,” *Dinamika Informatika*, Vol. 12, No. 1, Pp. 20–28, 2020.