

## Pengelompokan Toko *E-commerce* Shopee berdasarkan Reputasi Toko menggunakan Metode *Clustering K-Medoids*

Felicia Marvela Evanita<sup>1</sup>, Imam Cholissodin<sup>2</sup>, Sigit Adinugroho<sup>3</sup>

Program Studi Teknik Informatika, Fakultas Ilmu Komputer, Universitas Brawijaya  
Email: <sup>1</sup>fmarvela@student.ub.ac.id, <sup>2</sup>imamcs@ub.ac.id, <sup>3</sup>sigit.adinu@ub.ac.id

### Abstrak

Perkembangan internet mendorong terbentuknya *e-commerce* atau *electronic commerce*. Salah satu *e-commerce* yang dengan pengunjung paling banyak di Indonesia adalah Shopee dengan jumlah pengunjung lebih dari 72 juta tiap bulannya pada akhir tahun 2019. Walaupun *e-commerce* membawa dampak baik, pengguna tetap dihadapkan dengan risiko dari transaksi jual beli melalui *e-commerce*. Pengguna harus lebih berhati-hati dalam memilih toko yang akan dipercaya agar terhindar dari risiko tersebut. Pengguna sendiri dihadapkan oleh banyak pilihan pada saat mencari produk dan pengguna harus mempertimbangkan toko yang akan dipilih. Pengelompokan toko pada *e-commerce* Shopee berdasarkan reputasi menggunakan metode *clustering K-Medoids* dapat dilakukan untuk menyelesaikan permasalahan tersebut. Data pada penelitian ini diambil dengan melakukan web scraping pada 100 akun toko pada *e-commerce* Shopee. Langkah-langkah pada penelitian ini adalah preprocessing, normalisasi, mencari distance tiap data, clustering dengan *K-Medoids*, dan evaluasi dengan *Silhouette Coefficient*. Pada penelitian ini dilakukan pengujian pada banyak cluster dan banyak data yang digunakan. Dari kedua pengujian tersebut didapatkan bahwa nilai rata-rata *Silhouette Coefficient* terbaik sebesar 0,317681 didapatkan pada penggunaan cluster sebanyak 2 dan data sebanyak 100.

**Kata kunci:** pengelompokan toko, clustering, *K-Medoids*, *Silhouette Coefficient*

### Abstract

*The growth of internet encouraged the creation of e-commerce or electronic commerce. E-commerce with the most visitors in Indonesia is Shopee with more than 72 million visitors each month at the end of 2019. Although e-commerce has a lot of good impact, users are still faced with the risks from using e-commerce. Users must be more careful in choosing a store to trust in order to avoid these risks. Users are faced with many choices while looking for products and users must consider which store should they choose. Store clustering on Shopee e-commerce based on store reputation with K-Medoids clustering could solve this problem. The data that used in this study were taken from 100 store in Shopee e-commerce by web scraping. Steps that taken were preprocessing the data, normalization, finding the distance for each data, clustering with K-Medoids, and evaluate using Silhouette Coefficient. In this study, the number of cluster and data were tested. From these tests, it was found that the best Silhouette Coefficient average was 0,317681 while using 2 clusters and 100 data.*

**Keywords:** store clustering, clustering, *K-Medoids*, *Silhouette Coefficient*

## 1. PENDAHULUAN

Perkembangan teknologi internet mendorong terbentuknya *e-commerce* atau *electronic commerce*. Transaksi jual beli melalui *e-commerce* telah membentuk *digital market* yang baru. Pada *e-commerce*, harga lebih transparan dan pemasaran dapat dilakukan secara global (Laudon & Traver, 2016). *E-commerce* berkembang di berbagai negara, salah

satunya di Indonesia.

Menurut riset dari iprice, salah satu *e-commerce* dengan pengunjung paling banyak di Indonesia adalah Shopee dengan jumlah pengunjung lebih dari 72 juta tiap bulannya pada akhir tahun 2019 (iprice.co.id, 2019). Shopee adalah platform yang dirancang untuk memberikan pengalaman berbelanja secara *online* dengan mudah, aman, dan cepat. Shopee juga didukung dengan logistik dan sistem

pembayaran yang kuat. Pada aplikasi Shopee, terdapat berbagai kategori produk, dimulai dari perlengkapan rumah tangga, elektronik, kesehatan, *fashion* hingga kecantikan.

Walaupun *e-commerce* membawa dampak baik, pengguna tetap dihadapkan pada risiko dari transaksi jual beli melalui *e-commerce*, tidak terkecuali pada *e-commerce* Shopee. Menurut Turban, terdapat 5 risiko yang dihadapi oleh pengguna. Risiko tersebut meliputi produk yang didapatkan tidak sesuai dengan gambar yang dijanjikan, barang yang didapatkan rusak, kesalahan barang yang diterima berupa warna, ukuran, ataupun jumlah, barang tidak terkirim karena hilang ataupun terlambat, dan penipuan (Turban, 2002).

Pengguna harus lebih berhati-hati dalam memilih toko yang akan dipercaya agar terhindar dari risiko yang dihadapi. Dalam memilih toko, salah satu hal yang perlu diperhatikan adalah reputasi toko. Reputasi toko sendiri terbagi menjadi 2 dimensi. Pertama, reputasi toko mencerminkan kemampuan penjual untuk memberikan pelayanan yang baik. Kedua, reputasi toko juga memberikan rasa percaya kepada calon pelanggan (Luthfihadi & Dhewanto, 2013). Reputasi toko dapat dilihat berdasarkan banyak pengikut atau *followers*, banyak penilaian terhadap toko dan kinerja *chat*. Menurut Pratomo, memiliki banyak pengikut atau *followers* dapat menambahkan daya saing sebuah toko dengan toko lainnya (Pratomo, 2020). Banyak pengikut juga dapat menunjukkan bahwa toko tersebut dipercayai oleh para pengikut. Fitur banyak penilaian sangat berpengaruh pada reputasi toko karena menurut Jaikumar, pembeli cenderung lebih memilih toko dengan jumlah penilaian yang banyak dibandingkan memilih toko dengan harga barang yang lebih murah (Jaikumar, 2018). Fitur kinerja *chat* juga berpengaruh pada reputasi toko karena pelayanan kepada pembeli sangat penting bagi sebuah toko (Utz, et al., 2012). Fitur-fitur yang dapat menentukan reputasi toko tersebut akan digunakan pada penelitian ini beserta tambahan beberapa fitur lain yang tersedia pada halaman toko pada *e-commerce* Shopee, yaitu fitur banyak produk, banyak mengikuti dan waktu bergabung. Berdasarkan permasalahan tersebut, dibutuhkan sistem yang dapat memberikan pengelompokan terhadap toko pada *e-commerce* Shopee.

Hingga sekarang belum ada sistem yang dapat melakukan pengelompokan terhadap toko-toko pada aplikasi Shopee berdasarkan reputasi

toko. Terdapat sebuah sistem serupa dengan sistem pengelompokan toko, yaitu sistem pengelompokan terhadap kinerja akademik mahasiswa di Universitas Brawijaya dengan melakukan clustering menggunakan algoritme K-Means (Aziz, et al., 2018). Clustering adalah salah satu metode dalam data mining dengan tujuan utama mencari pola dari data untuk kemudian mengelompokkan data tersebut berdasarkan pola yang didapatkan, data dengan pola yang sama akan masuk ke cluster yang sama (Adolfsson, et al., 2018). Algoritme K-Means adalah algoritme pada clustering yang melakukan pengelompokan dengan membandingkan jarak tiap data ke centroid tiap cluster dan nilai centroid didapatkan dari rata-rata tiap cluster (Chayangkoon & Srivihok, 2016).

*Clustering* seperti yang dilakukan dalam pengelompokan kinerja akademik mahasiswa dapat dilakukan untuk mempermudah pengolahan data-data *e-commerce* Shopee. *Clustering* dapat dilakukan untuk mengelompokkan toko-toko yang ada di Shopee berdasarkan reputasi yang didapatkan dari fitur-fitur masukan. Algoritme yang akan digunakan pada penelitian ini adalah algoritme K-Medoids.

Pemilihan algoritme K-Medoids dalam penelitian ini didasarkan pada hasil penelitian oleh Baihaqi, Indartono, dan Banat pada penelitian dengan judul Penerapan Teknik *Clustering* sebagai Strategi Pemasaran pada Penjualan Buku di Tokopedia dan Shopee. Dari penelitian tersebut didapatkan bahwa nilai evaluasi Davies Bouldin Index dari metode K-Medoids lebih baik dibandingkan nilai evaluasi Davies Bouldin Index pada algoritme K-Means. Nilai evaluasi Davies Bouldin Index (DBI) pada algoritme K-Medoids adalah sebesar -1,484 dan -1,216, sedangkan pada algoritme K-Means nilai evaluasi yang didapatkan adalah sebesar -1,017 dan -0,684. Selain itu, menurut Han, Kamber, dan Pei algoritme K-Medoids dapat mengatasi kelemahan pada algoritme K-Means, algoritme K-Means dinilai sensitif terhadap *outlier* karena jika data yang nilainya cukup jauh dari data-data lain dimasukkan ke sebuah *cluster* maka nilai rata-rata *cluster* tersebut dapat berubah secara drastis (Han, et al., 2011). Perbedaan antara algoritme K-Means dan algoritme K-Medoids adalah pada algoritme K-Means digunakan nilai rata-rata dari data sebagai titik pusat *cluster*, sedangkan pada K-Medoids digunakan salah satu data sebagai objek representatif (*medoids*) yang menjadi titik pusat dari *cluster* (Kaur, et al.,

2014).

## 2. DASAR TEORI

### 2.1. E-commerce Shopee

*E-commerce* atau yang dalam Bahasa Indonesia disebut dengan perdagangan secara elektronik adalah aktivitas pembelian, penjualan, dan pemasaran produk berupa barang ataupun jasa melalui sistem elektronik seperti internet, televisi ataupun jaringan komputer lainnya (idcloudhouse.com, 2020). Shopee adalah salah satu platform *e-commerce* yang dapat digunakan oleh masyarakat Indonesia. Shopee pertama kali diluncurkan pada tahun 2015 di Singapura, Indonesia, Malaysia, Taiwan, Thailand, Filipina, dan Vietnam. Pada tahun 2020, Shopee adalah platform *e-commerce* yang paling unggul di Asia Tenggara dan Taiwan (careers.shopee.co.id, 2020).

### 2.2. Web Scraping

*Web scraping* merupakan sebuah teknik untuk mengambil informasi secara otomatis dari sebuah *website*. *Web scraping* bertujuan untuk mencari informasi tertentu dari sebuah *website*. Penggunaan *web scraping* pada penelitian ini ditujukan agar informasi yang diambil lebih terfokus sehingga dapat memudahkan dalam melakukan *clustering* pada data yang berhasil dikumpulkan. Langkah dari *web scraping* adalah (Ayani, et al., 2019):

#### 1. Create Scarping Template

Pada langkah ini yang akan dilakukan adalah mempelajari HTML dari website yang akan diambil informasinya. Pada tahap ini juga akan dilakukan penentuan tag dari HTML yang akan diambil informasinya.

#### 2. Explore Site Navigation

Pada langkah ini yang akan dilakukan adalah mempelajari teknik navigasi pada website yang akan diambil informasinya.

#### 3. Automate Navigation and Extraction

Pada langkah ini akan dilakukan otomatisasi pada pengambilan informasi dari website yang akan diambil informasinya berdasarkan langkah 1 dan langkah 2.

#### 4. Extracted Data and Package History

Setelah informasi berhasil diambil dari website pada langkah 3, maka informasi tersebut akan disimpan dalam database.

### 2.3. Min-max Normalization

Min-max Normalization adalah salah satu cara untuk melakukan normalisasi data dengan cara memetakan nilai  $v$  dari atribut  $A$  menjadi  $v'$  ke dalam skala  $[new\_min_A, new\_max_A]$  dengan Persamaan (1) (Han, et al., 2011):

$$v'_i = \frac{v_i - \min_A}{\max_A - \min_A} (new\_max_A - new\_min_A) + new\_min_A \quad (1)$$

Keterangan:

$v'_i$  = nilai hasil normalisasi

$v_i$  = nilai awal

$\min_A$  = nilai terkecil pada fitur  $A$

$\max_A$  = nilai terbesar pada fitur  $A$

$new\_min_A$  = nilai terkecil dari skala yang baru

$new\_max_A$  = nilai terbesar dari skala yang baru

### 2.4. K-Medoids Clustering

Algoritme K-Medoids adalah sebuah metode dalam *clustering* yang membagi data menjadi sebanyak  $k$  *cluster* dan menggunakan objek sebagai titik pusat atau yang biasa disebut dengan *medoid* pada tiap *cluster*-nya (Kaur, et al., 2014). Langkah-langkah *clustering* dengan menggunakan algoritme K-Medoids adalah (Han, et al., 2011):

a. Inisialisasi objek sebagai pusat *cluster* (*medoid*) sesuai dengan nilai  $k$  yang merupakan banyak *cluster*.

b. Tempatkan tiap data ke *cluster* terdekat dengan mencari nilai jarak dari tiap data ke tiap *medoid cluster* yang sebelumnya telah ditentukan secara acak. Hal ini dilakukan untuk menentukan kedekatan antara data tersebut dengan tiap *cluster*. Nilai jarak dapat dicari dengan jarak Euclidean pada Persamaan (2):

$$D_{ij} = \sqrt{(x_{1i} - x_{1j})^2 + (x_{2i} - x_{2j})^2 + \dots + (x_{ni} - x_{nj})^2} \quad (2)$$

Keterangan:

$D_{ij}$  = jarak data ke  $i$  ke pusat *cluster*  $j$

$x_{ni}$  = data ke  $i$  pada atribut data ke  $n$

$x_{nj}$  = data pusat ke  $j$  pada atribut ke  $n$

c. Pilih objek baru pada tiap *cluster* sebagai calon pengganti *medoid* lama secara acak.

d. Tempatkan kembali tiap data ke *cluster* berdasarkan *medoid* yang baru.

e. Hitung nilai simpangan ( $S$ ) dengan mengurangi total jarak baru dengan total jarak lama. Jika nilai simpangan kurang dari 0, maka tukar kembali *medoid* dengan objek yang baru.

f. Lakukan langkah ke 3-5 hingga *medoid* tidak mengalami perubahan. Maka akan didapatkan pusat *cluster* dan anggota masing-masing *cluster*.

### 2.5. Silhouette Coefficient

Metode Silhouette Coefficient dapat digunakan untuk pengujian kualitas dari *cluster* yang dihasilkan (Aziz, et al., 2018). Tahapan dalam mencari Silhouette Coefficient (Rousseeuw, 1986):

- a. Pada tiap objek, cari nilai rata-rata objek satu dengan objek lainnya pada *cluster* yang sama. Nilai tersebut adalah nilai  $a_i$ .
- b. Pada tiap objek, cari nilai rata-rata objek satu dengan semua objek yang berada pada *cluster* yang berbeda. Nilai tersebut adalah nilai  $b_i$ .
- c. Setelah semua nilai diketahui, tentukan nilai Silhouette Coefficient dengan Persamaan (3):

$$S_i = \frac{b_i - a_i}{\max(a_i, b_i)} \tag{3}$$

Keterangan:

$S_i$  = nilai Silhouette Coefficient

$a_i$  = rata-rata jarak antara objek  $i$  dengan objek lain pada *cluster* yang sama

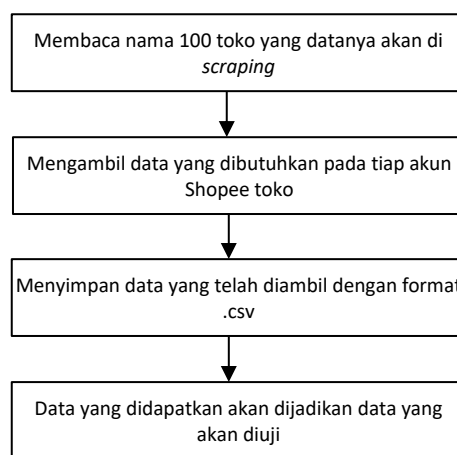
$b_i$  = minimal rata-rata jarak antara objek  $i$  dengan semua objek pada *cluster* yang berbeda

Hasil perhitungan dari nilai Silhouette Coefficient terdapat pada rentang -1 hingga 1. Hasil Silhouette Coefficient dapat dikatakan sudah baik apabila nilai yang didapatkan bernilai positif, yang berarti tiap data sudah berada pada *cluster* yang benar. Jika hasil Silhouette Coefficient bernilai negatif artinya terjadi *overlapping*, yaitu titik berada di antara dua *cluster*.

## 3. METODOLOGI

### 3.1. Pengumpulan Data

Pada penelitian ini data dikumpulkan dengan melakukan *web scraping* untuk mengambil data dari web *e-commerce* Shopee. Setelah melakukan *scraping*, data yang diperoleh akan dimasukkan ke dalam *file* *dokuji.csv* yang nantinya akan digunakan sebagai data yang diuji. Pada Gambar 1 ditunjukkan diagram alir dari proses pengambilan data:



Gambar 1. Diagram Alir Pengambilan Data

Pada penelitian ini data yang digunakan adalah 100 data toko yang dipilih secara acak dari Shopee dengan fitur banyak produk, banyak pengikut, banyak mengikuti, banyak penilai, kinerja *chat*, dan waktu bergabung. Pengambilan data dilakukan pada tanggal 18 September 2020.

### 3.2. Implementasi Algoritme

Tahapan yang dilakukan untuk melakukan implementasi algoritme *clustering* dengan K-Medoids pada penelitian ini adalah:

#### 1. *Preprocessing* data

Pada penelitian ini data yang digunakan adalah data yang didapatkan dari proses *web scraping* pada web *www.shopee.com*. Data yang didapatkan dari *web scraping* tersebut akan disimpan dengan format *csv*. Pada penelitian ini, *preprocessing* data dilakukan dengan menyamakan satuan dari tiap fitur atau standarisasi satuan serta melakukan normalisasi data. Pada tiap fitur, satuan yang digunakan akan disamakan agar memudahkan dalam melakukan *clustering*. Contohnya pada fitur waktu bergabung akan dilakukan menyamakan satuan dengan menggunakan satuan hari. Pada penelitian ini, normalisasi dilakukan dengan Min-max Normalization.

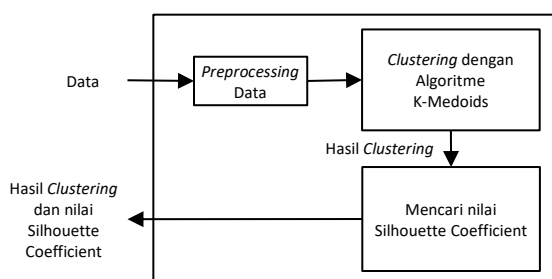
#### 2. *Clustering* dengan algoritme K-Medoids

Pada tahap ini dilakukan *clustering* dengan algoritme K-Medoids. *Clustering* dengan metode ini dilakukan dengan membagi data menjadi sebanyak  $k$  *cluster* dan menggunakan objek sebagai titik pusat atau yang biasa disebut dengan *medoid* pada tiap *cluster*-nya. Setelah menjalankan tahap *clustering* dengan algoritme K-Medoids maka akan didapatkan hasil berupa

keanggotaan dari tiap *cluster*. Akan dilakukan pengujian dengan mengubah nilai *k* dan banyak data. Dari hasil *clustering* ini akan terlihat karakteristik dari fitur pada tiap *cluster* yang terbentuk.

3. Mencari nilai Silhouette Coefficient  
Setelah dilakukan *clustering* dengan algoritme K-Medoids akan dilakukan evaluasi dengan cara mencari nilai Silhouette Coefficient. Hal ini dilakukan untuk memastikan bahwa *clustering* yang dilakukan sudah cukup baik.

Pada Gambar 2 akan ditunjukkan diagram blok dari implementasi algoritme K-Medoids pada penelitian ini:

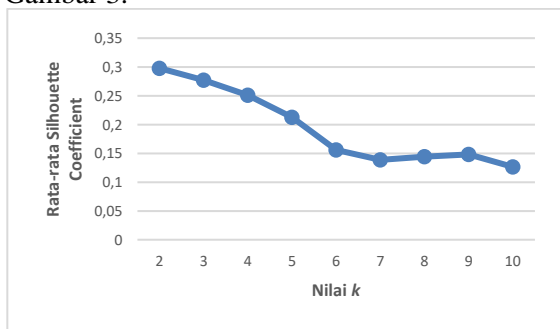


Gambar 2. Diagram Blok Implementasi Algoritme K-Medoids

#### 4. PENGUJIAN DAN ANALISIS

##### 4.1. Pengujian Banyak Cluster atau Nilai *k*

Pada pengujian banyak *cluster* atau nilai *k* dalam algoritme K-Medoids dilakukan pengujian pada nilai *k* sebesar 2 hingga 10. Pada pengujian banyak *cluster* ini digunakan data sebanyak 100. Pada setiap nilai *k* akan dilakukan 10 kali pengujian dan akan diambil nilai rata-rata dari Silhouette Coefficient yang didapatkan dari 10 pengujian tersebut. Diagram garis dari pengujian banyak *cluster* ditunjukkan pada Gambar 3:



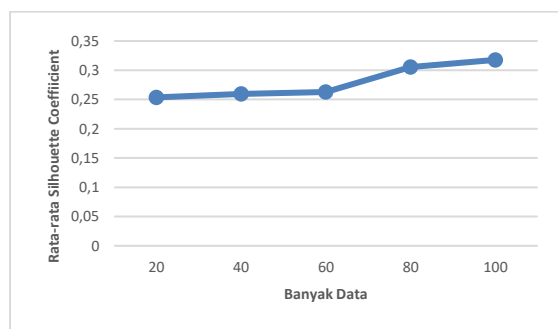
Gambar 3. Diagram Garis Pengujian Banyak Cluster atau Nilai *k*

Dari pengujian tersebut dapat disimpulkan bahwa semakin sedikit *cluster* yang akan

dibentuk atau semakin kecil nilai *k* yang digunakan maka nilai evaluasi dengan Silhouette Coefficient yang didapatkan lebih baik. Nilai *k* yang sebaiknya digunakan pada *clustering* untuk mendapatkan nilai evaluasi yang paling baik adalah nilai *k* sama dengan 2 yang mana nilai evaluasi yang didapatkan adalah sebesar 0,297795.

##### 4.2. Pengujian Banyak Data

Pada pengujian banyak data yang digunakan pada algoritme K-Medoids dilakukan pengujian pada banyak data sebesar 20, 40, 60, 80, dan 100. Nilai *k* yang digunakan pada pengujian ini adalah 2. Pada tiap banyak data akan dilakukan 10 kali pengujian dan nilai rata-rata Silhouette Coefficient yang didapatkan dari kali 10 percobaan akan diambil. Diagram garis dari pengujian banyak data dapat dilihat pada Gambar 4:



Gambar 4. Diagram Garis Pengujian Banyak Data

Dari pengujian tersebut dapat disimpulkan bahwa semakin banyak data yang digunakan maka nilai evaluasi dengan Silhouette Coefficient yang didapatkan akan lebih baik. Banyak data yang sebaiknya digunakan pada penelitian ini untuk mendapatkan nilai evaluasi yang paling baik adalah 100 data atau keseluruhan data.

##### 4.3. Analisis Global

Analisis global ini dilakukan untuk mengetahui karakteristik dari anggota *cluster* dari hasil *clustering* dengan metode K-Medoids pada penelitian ini. Parameter yang digunakan pada analisis global ini adalah nilai *k* sebesar 2 dan nilai banyak data sebesar 100. Dari hasil *clustering* tersebut didapatkan nilai minimal, maksimal, dan *mean* tiap fitur. Dari nilai-nilai tersebut didapatkan karakteristik tiap fitur pada tiap *cluster*:

- a. Nilai fitur mengikuti, pengikut, penilaian,

dan bergabung pada *cluster* 1 cenderung lebih besar dibandingkan pada *cluster* 2.

- b. Nilai fitur produk pada *cluster* 1 cenderung lebih kecil dibandingkan pada *cluster* 2.
- c. Nilai fitur kinerja *chat* pada *cluster* 1 dan 2 tidak berbeda jauh.

## 5. PENUTUP

### 5.1. Kesimpulan

Beberapa kesimpulan yang didapatkan berdasarkan hasil dari pengujian pada penelitian ini adalah:

- a. Pada pengujian parameter banyak *cluster* atau nilai  $k$ , didapatkan hasil bahwa nilai evaluasi paling baik didapatkan pada saat penggunaan banyak *cluster* atau nilai  $k$  sebesar 2 pada penggunaan 100 data. Nilai evaluasi dengan Silhouette Coefficient yang didapatkan pada *cluster* sebanyak 2 adalah 0,297795. Dapat disimpulkan pula bahwa nilai Silhouette Coefficient yang didapatkan akan semakin kecil jika nilai  $k$  yang digunakan semakin besar.
- b. Pada pengujian parameter banyak data dapat dilihat pengaruh banyak data yang digunakan dengan nilai evaluasi yang didapatkan. Dari pengujian tersebut didapatkan bahwa nilai Silhouette Coefficient yang didapatkan akan semakin besar jika banyak data yang digunakan semakin banyak. Nilai Silhouette Coefficient terbaik didapatkan pada *clustering* menggunakan 100 data yang mana nilai Silhouette Coefficient yang didapatkan adalah sebesar 0,317681 pada nilai  $k$  sebesar 2.
- c. Pada analisis global didapatkan karakteristik tiap fitur pada tiap *cluster* yang terbentuk. Nilai fitur produk pada *cluster* 1 cenderung lebih kecil dibandingkan dengan pada *cluster* 2. Nilai fitur mengikuti, pengikut, penilaian dan bergabung pada *cluster* 1 cenderung lebih besar dibandingkan dengan pada *cluster* 2. Nilai fitur kinerja *chat* pada *cluster* 1 dan *cluster* 2 tidak berbeda jauh.

### 5.2. Saran

Beberapa saran yang dapat dilakukan pada penelitian selanjutnya berdasarkan hasil dari pengujian yang dilakukan adalah:

- a. Banyak data pada penelitian ini adalah 100 data sehingga dapat dikatakan bahwa nilai

evaluasi yang didapatkan belum optimal. Data yang digunakan pada penelitian selanjutnya dapat ditambahkan sehingga nilai evaluasi yang didapatkan akan lebih baik.

- b. Pada penelitian ini data diambil dengan menggunakan *web scraping*. Pada penelitian selanjutnya sebaiknya menggunakan *web crawling* karena dengan menggunakan *web crawling* data yang didapatkan merupakan data terbaru.
- c. Pada penelitian selanjutnya dapat dilakukan penambahan fitur jumlah produk yang terjual atau fitur penilaian tiap barang sebagai pertimbangan reputasi toko.
- d. Pada penelitian selanjutnya dapat digunakan metode *clustering* lain untuk meningkatkan nilai evaluasi dari sistem.
- e. Pada penelitian selanjutnya, dapat dilakukan penambahan teknik autoencoder untuk mengidentifikasi *outlier* data serta untuk melakukan visualisasi *cluster* dari dimensi tinggi ke dimensi 2 atau 3. Dengan ditambahkan teknik autoencoder diharapkan nilai evaluasi Silhouette Coefficient yang didapatkan lebih optimal.

## 6. DAFTAR PUSTAKA

- Adolfsson, A., Ackerman, M. & Brownstein, N., 2018. To Cluster, or Not To Cluster: An Analysis of Clusterability Methods. *Pattern Recognition*, Issue 88, pp. 13-26.
- Ayani, D., Pratiwi, H. & Muhardi, H., 2019. Implementasi Web Scraping untuk Pengambilan Data pada Situs Marketplace. *Jurnal Sistem dan Teknologi Informasi*, 7(4), pp. 257-262.
- Aziz, F., Setiawan, B. & Arwani, I., 2018. Implementasi Algoritme K-means untuk Klasterisasi Kinerja Akademik Mahasiswa. *Jurnal Pengembangan Teknologi Informasi dan Ilmu Komputer*, 2(6), pp. 2243-2251.
- Baihaqi, W., Indatono, K. & Banat, S., 2019. Penerapan Teknik Clustering sebagai Strategi Pemasaran pada Penjualan Buku di Tokopedia dan Shopee. *Paradigma-Jurnal Komputer dan Informatika*, 21(2), pp. 243-248.
- careers.shopee.co.id, 2020. About. [Online] Tersedia di: <<https://careers.shopee.co.id/about/>> [Diakses 20 September 2020].

- Chayangkoon, N. & Srivihok, A., 2016. Two Step Clustering Model for K-means Algorithm. s.l., International Conference on Network, Communication and Computing, pp. 213-217.
- Han, J., Pei, J. & Kamber, M., 2011. Data Mining: Concepts and Techniques. s.l.:Elsevier.
- idcloudhouse.com, 2020. Pengertian E-Commerce dan Contohnya, Komponen, Jenis, dan Manfaat E-Commerce. [Online] Tersedia di: <<https://idcloudhost.com/pengertian-e-commerce-dan-contohnya-komponen-jenis-dan-manfaat-e-commerce/>> [Diakses 12 Agustus 2020].
- iprice.co.id, 2019. Peta E-Commerce Indonesia. [Online] Tersedia di: <<https://iprice.co.id/insights/mapofecommerce/>> [Accessed 30 Maret 2020].
- Jaikumar, S., 2018. How Do Consumers Choose Sellers in E-marketplace. Journal of Advertising Research, pp. 1-10.
- Kaur, N., Kaur, U. & Singh, D., 2014. K-Medoid Clustering Algorithm. International Journal of Computer Application and Technology (IJCAT), 1(1), pp. 42-45.
- Laudon, K. & Traver, C., 2016. E-commerce: business, technology, society. 12th ed. s.l.:Pearson.
- Luthfihadi, M. & Dhewanto, W., 2013. Technology Acceptance of E-commerce in Indonesia. International Journal of Engineering Innovation and Management, pp. 9-18.
- Pratomo, Y., 2020. Product Diversification And Advertising Cost Strategy Of "Glamgoddess" Store In The "Shopee" Place Market. Humanities, Management and Science Proceedings, 1(1), pp. 530-535.
- Rousseeuw, P., 1986. Silhouettes: A Graphical Aid to The Interpretation and Validation of Cluster Analysis. Journal of Computational and Applied Mathematics, Volume 20, pp. 53-65.
- Turban, E., 2002. Electronic Commerce: A managerial perspective 2002. s.l.:Prentice Hall.