

Klusterisasi Data Review Pengguna Aplikasi Marketplace Blibli.Com dengan Algoritma K-Means dan K-Medoids

Muhammad Rafi Haidar Arsyad¹, Sulastr²

^{1,2}Universitas Stikubank Semarang; jl. Tri Lomba Juang, Mugassari, Kec. Semarang Selatan., Kota Semarang, Fax (024) 8443240
e-mail: *¹rafihaidar@edu.unisbank.ac.id, ²sulastr²@edu.unisbank.ac.id

Abstrak

Marketplace merupakan sarana bagi konsumen untuk melakukan belanja online, seiring dengan meningkatnya pengguna *marketplace*, peran *Review* aplikasi merupakan hal yang semakin penting. Ulasan yang disampaikan oleh pengguna aplikasi merupakan sumber informasi penting untuk menilai kepuasan pelanggan terhadap aplikasi. Penelitian ini bertujuan untuk mengelompokkan data *review* dari aplikasi *marketplace* Blibli dengan menggunakan metode *K-Means Clustering* dan *K-Medoids Clustering*. Dataset yang digunakan merupakan data ulasan pelanggan dari aplikasi Blibli dari tahun 2022 hingga 2023 sebanyak 17.255 *review*. Hasil penelitian yang didapatkan dari 2 metode tersebut 4 *cluster* optimal yang terdapat pada data ulasan. Kata yang sering muncul “aplikasi” “barang” “belanja” ditampilkan bentuk *wordcloud* dan hasil klusterisasi dalam bentuk *cluster plot*. Hasil yang didapatkan untuk memperbaiki pelayanan aplikasi Blibli.

Kata kunci : *Review, K-Means, K-Medoids, Blibli*

Abstract

The marketplace serves as a platform for consumers to engage in online shopping. With the increasing use of marketplaces, the role of application reviews becomes increasingly crucial. Reviews provided by application users are a significant source of information for assessing customer satisfaction with the application. This research aims to categorize review data from the Blibli marketplace application using the K-Means Clustering and K-Medoids Clustering methods. The dataset used consists of customer reviews from the Blibli application spanning from 2022 to 2023, totaling 17,255 reviews. The research results indicate that both methods yield four optimal clusters in the review data. Frequently occurring words such as 'application,' 'goods,' and 'shopping' are visualized in a word cloud, and the clustering results are presented in a cluster plot. The obtained findings aim to enhance the service quality of the Blibli application.

Keywords : *Review, K-Means, K-Medoids, Blibli*

1. PENDAHULUAN

Review sangat penting sebagai sumber daya berharga untuk meningkatkan kualitas aplikasi. Analisis ulasan dapat dilakukan dengan berbagai metode, salah satunya adalah metode *clustering*. [1] *Text Mining* merupakan proses yang digunakan untuk pengambilan informasi berupa dokumen teks yang memiliki kualitas tinggi, dari data teks tersebut dapat mengetahui permasalahan dari sebuah topik tertentu. [2]

Metode *K-Means Clustering* digunakan untuk menganalisa data ulasan perjalanan yang di dapatkan dari situs Tripadvisor. Pengelompokan data yang di dapatkan dibagi menjadi 2 kluster berdasarkan peringkat tertinggi dan terendah sehingga dapat memberikan informasi mengenai pengelompokan data ulasan berdasarkan rating tertinggi wisatawan pada suatu objek tempat wisata. dengan menggunakan *tools* Rapidminer, *cluster* 0 untuk yang tertinggi dan *cluster* 1 untuk yang terendah. *cluster* 0 dengan peringkat tertinggi terdiri dari galeri seni dengan rating sebesar 7,118 sedangkan untuk *cluster* 1 dengan peringkat terendah terdiri dari restoran dengan data rating sebesar 10,394. [3]

Pengelompokan data lagu pop yang ada di aplikasi Spotify menggunakan Algoritma *K-Means* dan *K-Medoids*. Pengumpulan data *genre* lagu pop menggunakan teknik *scraping* data (web scrapper) mendapatkan data sebanyak 3.259 lagu pop. Dengan menggunakan metode *Silhouette Coefficient* mendapatkan 3 *cluster* optimal. Perbandingan antara metode *K-Means* dengan *K-Medoids*, untuk nilai

tertinggi dari setiap anggota yang terdapat pada algoritma *k-means cluster* ke 3 dengan jumlah total 1.774 anggota.[4]

Pengelompokan data ulasan aplikasi pedulilindungi dengan menggunakan Algoritma *K-Medoids*. Dengan menggunakan data ulasan valid sebanyak 50.688 *review* melalui proses *Scraping* Data bantuan Google Play Scrapper pemrograman Python. Pengelompokan data teks ulasan menggunakan metode *K-Medoids* dan KDD (*Knowledge Discovery In Database*) dengan jumlah cluster optimal 2 terhadap data *review* yang didapatkan. Menghasilkan nilai DBI yang mendekati nol sebesar 2,93. kata yang sering muncul yang masuk ke dalam *cluster* pertama yaitu “Aplikasi” dan *cluster* kedua yaitu “vaksin”. [5]

Penerapan metode Algoritma K-Means Clustering dalam menganalisa data *review* pengguna aplikasi BCA Mobile di Google Playstore. Dengan menggunakan metode *Silhouette Score* mendapatkan hasil validitas optimal jumlah *cluster* 10. Pengujian 10 *cluster* yang di dapatkan mendapatkan nilai *silhouette score* sebesar 0,1027277 dan rata-rata *rating* pengguna aplikasi sebesar 2,65.[6]

Klusterisasi dataset *review* barang toko elektronik di *e-commerce* Lazada dari tahun 2019 hingga 2022 sebanyak 203.788 *review* barang data resmi yang di ambil dari *library dataset kaggle* bersifat *open source*. mendapatkan 2 *cluster* yang terdapat pada *review* dengan *frequency* kata “barang”, dalam penghitungan 2 *cluster* metode *silhouette score* untuk menentukan validitas hasil yang optimal. Nilai *silhouette score* yang di dapatkan dari 2 *cluster* tersebut mendapatkan rata – rata yang tertinggi 0,75. kata paling sering muncul di *cluster* 1 dengan hasil kata sebagai berikut “barang”, ”sesuai”, “cepat” sehingga bisa disimpulkan kata yang sering muncul tersebut dijadikan acuan kepada toko dalam mempertahankan atau meningkatkan kualitas pelayanannya.[7]

Penelitian ini menggunakan data *review* aplikasi Blibli yang ada di Google Play Store. Pengelompokan data *review* ke dalam beberapa kluster ini menggunakan 2 metode *K-Means* dan *K-Medoids*. Tujuan Penelitian adalah berupa informasi pengelompokan data ulasan berdasarkan jumlah *cluster maksimum* dengan menggunakan metode *Silhouette Score* dan kata yang sering muncul yang didapatkan dari stopwords di visualisasikan menggunakan *wordcloud*. Visualisasi hasil dari *K-Means* dan *K-Medoids* berupa *Cluster plot*.

2. METODE PENELITIAN

Penelitian yang dilakukan menggunakan Algoritma *K-Means Clustering* dan *K-Medoids Clustering* dengan tujuan melakukan pengelompokan kata yang ada dalam data *review* yang digunakan untuk melihat perbandingan hasil kedekatan antarkata dari 2 metode tersebut. *K-Means* perhitungan jarak di hitung dari data pada masing masing rata-rata (*Means*), sedangkan untuk *K-Medoids Clustering* data yang digunakan sebagai nilai tengah (*Medoids*) untuk perhitungan jarak dihitung berdasarkan jarak masing masing data, metode ini *sensitive* terhadap *noise*, *outlier*, dan meminimalkan jumlah perbedaan. Tahapan penelitian sebagai berikut:

1. Pengumpulan Data

Pada tahapan pengumpulan data, Data *review* pengguna aplikasi blibli yang digunakan berasal dari hasil *Scraping* dengan bahasa pemrograman python. Data ulasan yang peneliti dapatkan merupakan sebuah aktifitas ulasan pengguna yang ada di *google playstore* terhadap aplikasi blibli.com. Metode *scraping* mendapatkan data ulasan dari rentang tahun 2022 hingga 2023. Data *Review* ulasan aplikasi blibli.com mendapatkan sebanyak 17.255 *review* berbentuk csv.

2. Data Text Preprocessing

Pada tahapan berikut ini merupakan langkah selanjutnya melakukan pemrosesan data untuk dilakukan seleksi data teks *review* yang dibutuhkan dalam penelitian menggunakan bantuan *tools* Google Colabs. Dataset *review* tersebut akan dibagi menjadi 3 tahap pemrosesan yaitu:

a) Cleaning

Tahapan *Cleaning* merupakan langkah dimana setiap kalimat akan dilakukan penyeleksian yang nantinya akan mendapatkan menghasilkan data *review* menjadi huruf kecil semua, proses *cleaning* akan melakukan pembersihan data menghilangkan angka, tanda baca, *emoticon*, spasi berlebihan. Proses pengolahan data *review* menggunakan bahasa pemrograman R dengan *tools* Google Colabs. Hasil Proses *Cleaning* bisa dilihat pada Tabel 1.

Tabel 1 Hasil Proses *Cleaning*

Review	Cleaning
pertama pakai langsung bagus Top Up Flazz	pertama pakai langsung bagus top up flazz

b) Stopword Removal

Pada proses ini akan dilakukan pembersihan kata yang kurang bermanfaat menggunakan *Stopword Removal*. Pada proses ini dilakukan pemilihan kata (*sortir*) yang kurang dibutuhkan akan dihilangkan. Seperti contoh doang, cuma, yang dan sebagainya. Proses ini menggunakan *corpus* yang di inputkan ke dalam pengolahan data ulasan. Hasil proses *stopword removal* dapat dilihat pada tabel 2.

Tabel 2 Hasil Proses *Stopword Removal*

Review	Stopword Removal
sekian ecommerce cuma BliBli doang, mudah digunakan buat belanja online, gratis ongkir pengiriman cepat	sekian ecommerce blibli mudah digunakan belanja online gratis ongkir pengiriman cepat

c) Tokenization

Tahapan pemrosesan *tokenization* ini bertujuan untuk pemisahan kalimat yang sudah dibersihkan dalam data teks ulasan aplikasi blibli.com menjadi sebuah kata yang terpisahkan antara satu dengan yang lainnya. Hasil proses *tokenization* dapat dilihat pada tabel 3.

Tabel 3 Hasil Proses *Tokenization*

Review	Tokenization
belanja online gratis ongkir pengiriman cepat	“belanja” “online” “gratis” “ongkir” “pengiriman” “cepat”

3. TF-IDF (*Term Frequency-Inverse Dense Frequency*)

Pada tahapan ini akan melakukan pemrosesan data teks *review* menggunakan metode TF-IDF. Proses ini merupakan salah satu cara untuk melakukan penghitungan bobot setiap kata yang umum digunakan dalam suatu dokumen teks, sehingga dapat mengetahui seberapa penting sebuah kata untuk di olah. Dalam pemrosesan ini akan mengetahui relevansi bobot dari suatu kata terhadap seluruh dokumen teks yang ada di dalam *corpus*. Setiap kata yang diberikan bobot (*weight*) akan dilakukan penghitungan frekuensi kemunculan dalam data teks tersebut.

$$W_{i,k} = TF_{i,k} \times IDF_k$$

Keterangan :

W_{ik} = bobot dari kata ke-j pada data teks *review* ke-i

Tf_{ik} = jumlah kata yang sering muncul ke-k pada *review* ke-i

IDF_k = *inverse document frequency* pada kata ke-k

$$IDF_k = \log \left(\frac{N}{DF_k} \right)$$

Keterangan :

IDF_k = *inverse document frequency* pada kata ke-k

N = jumlah keseluruhan data *review*

DF_k = banyaknya kalimat *review* yang mengandung kata-k

4. Metode yang digunakan

Dalam penelitian ini untuk melakukan pengelompokan data teks *review* aplikasi Blibli menggunakan metode Algoritma *K-Means* dan *K-Medoids*. Untuk mendapatkan jumlah *cluster* yang maksimal dalam klusterisasi data teks *review* menggunakan metode *Silhouette Score*. Selanjutnya data yang telah di proses akan di tampilkan berupa *wordcloud* dan *cluster plot*.

5. Pengujian

Proses pengolahan dan pengujian data *review* aplikasi Blibli menggunakan bahasa pemrograman R dengan bantuan *tools* Google Colabs. *Dataset* yang digunakan merupakan data asli yang diambil dengan metode *Scraping* Data pada Google Play Scraper dengan bahasa pemrograman Python.

Dataset yang digunakan sebanyak 17.255 *review* pengguna aplikasi Blibli dengan format csv. Kemudian *dataset review* dapat dilakukan pembersihan manual menggunakan excel.

6. Validasi dan Evaluasi Penelitian

Proses *Text Preprocessing* yang dilakukan dalam serangkaian tahapan *cleaning*, *stopword removal* dan *tokenizing* dalam proses validasi data teks *review* dan evaluasi penelitian akan mengubah beberapa objek yang digunakan sebagai bahan penelitian dalam pemrograman bahasa R dengan menggunakan *tools* Google Colabs. Untuk mendapatkan hasil yang terbaik pada fokus penelitian ini data yang telah dibersihkan dilanjutkan proses *preprocessing*, tahap ini dilakukan uji coba dengan menentukan jumlah *cluster* yang optimal terhadap data *review* aplikasi blibli dengan bantuan metode *silhouette score*, sehingga proses finalisasi data *review* klusterisasi menggunakan 2 metode *K-Means* dan *K-Medoids* dapat bekerja secara maksimal.

3. HASIL DAN PEMBAHASAN

Pembahasan penelitian yang dilakukan digunakan untuk pengelompokan kata yang ada pada *dataset review* aplikasi blibli, kata yang sering muncul akan di tampilkan ke dalam *wordcloud*. Selanjutnya akan dilakukan klusterisasi menggunakan 2 metode *K-Means* dan *K-Medoids* yang hasilnya akan dibandingkan berdasarkan *scatter plot*, Sehingga informasi yang didapatkan dapat mengetahui hasil klusterisasi data dari kedua metode tersebut.

Penelitian ini menggunakan dataset ulasan pengguna aplikasi Blibli sebanyak 17.255 *review* yang didapatkan melalui proses *Scraping Data* dengan Google Play Scrapper menggunakan bahasa pemrograman Python dalam bentuk file csv. Data *review* asli pengguna aplikasi Blibli yang ada di Google Playstore, data yang didapatkan dilakukan proses pengolahan dan pembersihan terlebih dahulu agar dapat di proses dengan 2 metode Algoritma *K-Means Clustering* dan *K-Medoids Clustering* menggunakan Google Colabs.

1. Data Cleaning

Tahap pembersihan data atau Data Cleaning dilakukan dengan menggunakan Microsoft Excel dan Google Colabs, pada tahap pembersihan data melalui excel berupa file csv dibersihkan dari data yang kosong dan data ganda, kemudian melakukan pemilihan objek yang dijadikan fokus penelitian. Data awal kosong dan ganda sebelum dibersihkan sebanyak 17.255 dan data *review* bersih yang akan digunakan untuk pengolahan klusterisasi data melalui Google Colabs sebanyak 1999. Program bahasa R untuk melakukan proses data *cleaning* dapat dilihat pada gambar 1.

```
# Baca file CSV sebagai teks
lines <- readLines("reviews_rafi.csv")

# Buat data frame dengan satu kolom "content"
data <- data.frame(content = lines, stringsAsFactors = FALSE)

# Tahap pembersihan data:
# 1. Mengubah teks menjadi huruf kecil
data$content <- tolower(data$content)

# 2. Menghilangkan karakter khusus atau tanda baca
data$content <- gsub("[[:punct:]]", "", data$content)

# 3. Menghapus spasi ekstra
data$content <- gsub("\\s+", " ", data$content)

# Membersihkan angka
data$content <- gsub("\\d", "", data$content)

# 4. Menghapus duplikat (jika ada)
data <- unique(data)

# 5. Menghapus baris dengan teks yang kosong
data <- data[data$content != "", ]

# Tampilkan beberapa baris pertama data
head(data)

# Data
write.csv(data, file = "data_bersih.csv", row.names = FALSE)
```

Gambar 1. Program Data Cleaning

Tujuan dari program data cleaning yaitu untuk pembersihan data teks *review* aplikasi blibli yang memiliki karakteristik, tanda baca, angka, emoji dan membuat semua kalimat ulasan yang ada menjadi huruf kecil (*lower case*), selanjutnya kata yang memiliki makna dapat di proses ke tahap berikutnya *stopword removal*. Hasil data *cleaning* dapat dilihat pada gambar 2.

```
content
<chr>
1 pertama paket langsung bagus top up fitez proses cepet ngaku nunggu lama ment estimasi sudah masuk nunggu saldo dana seneng terima cashback banyak aplikasi keren top up cepet masuknya voucher gratis ongkir mudah aplikasi bagus terpacaya
2 sekian ecommerce cuma bibi doang mudah digunakan buat belanja online gratis ongkir pengiriman cepet
3 kalo beli item bibi store nya ribet bangetka chat ukur sendal apa live chattpi cs nya ga tau barang yg tanyakpndi pencet chat nya desripai barang yg plihdntanya link nya diiya laudttempat keren gampang
4 tp login tp yg pake akun yg beda akun bilang indikasi pelanggan ga okeout voucher promo ga gila keren ongkirnya mahal banget ya uda klo ga belanja bibi ga bisor keren akun bibi biar hapus aplikasi bibinya shopee lazada belanja online moga tetap lancar ga kaya market place
5 pasipasti cepet si biru tua si oren tampil konten bagus gerak cepet bagus keluh konsumen yg masuk 2 foto kit bukalapak fotopelebarmentah 1 bibi belanja online
6 level diamondtiap belanja batal sistem karna langgar syarat ketentuanpausdntal belanja wkerenkerenberanja bufor anak diaga pake angsapuan prihat kapada ngg tibatama sakumama bilang langgar syarat ketentuanbelanja bayar tunai kreditkoreasa pokok customer net buki
```

Gambar 2. Hasil Data Cleaning

$$d_{Euclidean}(x, y) = \sqrt{\sum_{i=1}^n (X_i - Y_i)^2}$$

Keterangan

$x = x_1, x_2, \dots, x_n$, dan $y = y_1, y_2, \dots, y_n$ merupakan banyaknya n atribut(kolom) antara 2 record.

Selanjutnya melakukan perhitungan nilai SSE(*Sum Of Square Error*) pada setiap data dengan nilai *cluster* yang berbeda. Prinsip yang ada pada perhitungan nilai SSE adalah proses *trial and error* akan diulang sampai nilai *cluster* mencapai nilai minimum.[8]

$$SSE = \sum_{k=1}^k \sum_{X_{ci}} \|X_i - C_i\|_2^2$$

Keterangan:

K = jumlah *cluster*

X_i = atribut data

X_{ci} = atribut *centroid*

C_i = *centroid*

Untuk mendapatkan *cluster* yang optimal dari suatu data *review* menggunakan metode *silhouette score*, dengan cara *silhouette* memungkinkan data *review* menghasilkan hingga 10 *cluster*. Dalam penelitian ini mendapatkan 4 *Cluster* yang optimal.[9]

$$s(i) = \frac{b(i) - a(i)}{\max(a(i), b(i))}$$

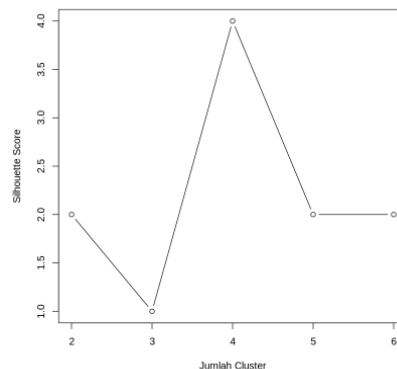
Keterangan :

Hitung $s(i)$: merupakan *koefisien silhouette* yang akan dicari atau titik ke i

Hitung $a(i)$: merupakan penghitungan jarak rata-rata dari suatu titik dengan titik yang lain dalam kelompok (*cluster*) yang sama

Hitung $b(i)$: merupakan penghitungan jarak rata-rata dari suatu titik dengan semua titik dalam *clusternya* dengan *cluster* terdekat.

Hasil *Cluster Silhouette Score* yang optimal untuk data *review* blibli bisa dilihat pada gambar 7.



Gambar 7 Hasil *cluster Silhouette Score*

Hasil dari perhitungan *K-Means Clustering* menggunakan bahasa pemrograman R terhadap data *review* pengguna aplikasi Blibli ditampilkan pada gambar 8.

```

K-means clustering with 4 clusters of sizes 1, 5, 1865, 128

Cluster means:
  aplikasi      bagus      banyak      cashback      cepet      dana
1  0.00000000  0.00000000  0.00000000  0.00000000  0.00000000  0.00000000
2  0.00000000  0.00000000  0.00000000  0.00000000  0.00000000  0.00000000
3  0.00346100  0.01356100  0.00429441  0.00927842  0.00285206  0.01796834
4  0.10650110  0.01709949  0.00702358  0.00725574  0.00148463  0.00148463

  estimasi      flazz      grati      keren      lama      langsung
1  0.00000000  0.00000000  0.00000000  0.00000000  0.00000000  0.00000000
2  0.00000000  0.00000000  0.00000000  0.00000000  0.00000000  0.00000000
3  0.00583158  0.0001729232  0.01567839  0.02467526  0.002409691  0.01563414
4  0.00000000  0.00000000  0.00000000  0.02776852  0.00000000  0.01174689

  hasil
1  0.00000000  0.00000000  0.00000000  0.00000000  0.00000000  0.00000000
2  0.00000000  0.00000000  0.00000000  0.00000000  0.00000000  0.00000000
3  0.19369397  0.0001729232  0.004293842  0.001652155  0.009165346  0.01251164
4  0.02467124  0.00000000  0.007874277  0.00000000  0.002444277  0.00000000

  ongkir      pakai      pertama      prose      saldo      senang
1  0.00000000  0.00000000  0.00000000  0.00000000  0.00000000  0.00000000
2  0.00000000  0.00000000  0.00000000  0.00000000  0.00000000  0.00000000
3  0.02599223  0.01914118  0.000309156  0.02086955  0.01669339  0.001729232
4  0.00000000  0.01912955  0.00000000  0.00000000  0.00000000  0.00000000

  sudah      terima      terpercaya      top      voucher      belanja
1  0.00000000  0.00000000  0.00000000  0.00000000  0.00000000  0.00000000

```

Gambar 8 Hasil Hitung *K-Means Clustering*

7. K-Medoids Clustering

Metode K-Medoids merupakan klusterisasi data *non hierarki* untuk mengolah data nilai tengah (Pusat *Cluster*) disebut dengan *medoid*, perhitungan jarak dokumen teks ini di hitung berdasarkan jarak antara masing masing data. Dalam Pemrograman bahasa R *K-Medoids* menggunakan fungsi *pam*. [10]

[illegible]

Gambar 9 Program dan Hasil *K-Medoids Clustering*

[illegible]

Gambar 10 *Clustering Vector* Hasil Program *K-Medoids Clustering*

Dengan menggunakan bantuan mesin pemrograman bahasa R dan menggunakan metode TF-IDF dan *Euclidean Distance* mendapatkan hasil *K-Medoids Clustering*. Hasil *K-Medoids* dan *clustering vector* dapat dilihat pada gambar 9 dan 10.

Tahapan Metode *K-Medoids Clustering* sebagai berikut:

1. Tentukan k (jumlah *cluster*) yang diinginkan.
2. Pilih secara acak medoid awal sebanyak k dari n data.
3. Hitung jarak masing-masing objek ke *medoid* sementara, kemudian tandai jarak terdekat objek ke *medoid* dan hitung totalnya.
4. Lakukan iterasi *medoid*, yaitu pilih secara acak objek pada masing-masing cluster sebagai kandidat medoid baru (*non-medoid*) dan hitung total simpangan (S) antara jarak terdekat objek

ke medoid awal dan medoid baru. Jika $S < 0$, maka tukar objek dengan *medoids* untuk membentuk sekumpulan k baru sebagai *medoid*.

5. Ulangi langkah 3 sampai 4 dan hentikan jika sudah tidak terjadi perubahan anggota *medoid*.

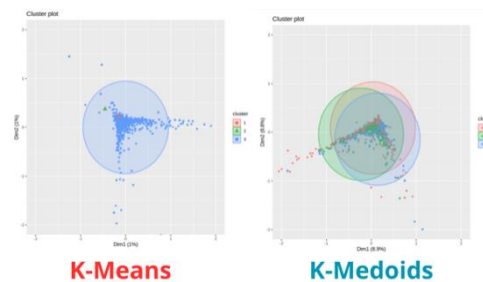
Selanjutnya proses penentuan kalimat teks ulasan berdasarkan *cluster medoids* menggunakan bantuan mesin pemrograman R fungsi pam di Google Colabs. Pada gambar 11 dapat dilihat kalimat *review* telah ditentukan secara otomatis menggunakan fungsi *pam.review k-medoids*.

[illegible]

Gambar 11. Hasil *Cluster Medoids Review*

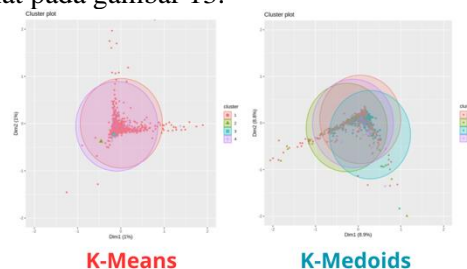
- ## 8. Hasil Perbandingan 2 Metode

Pada proses ini akan melihat hasil perhitungan jarak dengan metode *Euclidean Distance* pada dataset dan *centroid*. Dengan melakukan uji coba kluster 3,4, dan 5 pada *K-Means* dan *K-Medoids*. Dalam perhitungan jarak melalui proses *trial and error* yang dilakukan secara berulang hingga menghasilkan tidak adanya perpindahan objek data teks ke *cluster* lain. Nilai *Cluster* awal yang akan digunakan adalah K-3. Hasil *Cluster Plot K-Means* dan *K-Medoids* dengan K-3 dapat dilihat pada gambar 12.



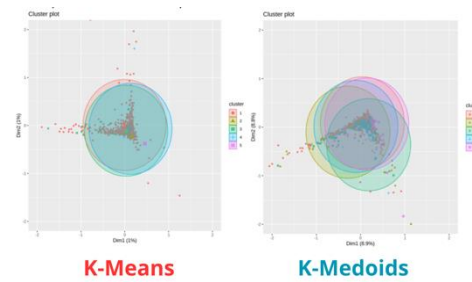
Gambar 12 Hasil *Cluster Plot k-means* dan *k-medoids* Dengan 3 cluster

Berikutnya akan melakukan proses perhitungan jarak dengan cara yang sama menggunakan metode *Euclidean Distance*, menggunakan nilai kluster K-4. Hasil *Cluster Plot K-Means* dan *K-Medoids* dengan K-4 dapat dilihat pada gambar 13.



Gambar 13 Hasil *Cluster Plot k-means* dan *k-medoids* Dengan 4 cluster

Berikutnya akan melakukan proses perhitungan jarak dengan cara yang sama menggunakan metode *Euclidean Distance*, menggunakan nilai kluster K-5. Hasil *Cluster Plot K-Means* dan *K-Medoids* dengan K-5 dapat dilihat pada gambar 14.

Gambar 14. Hasil *Cluster Plot k-means* dan *k-medoids* Dengan 5 KlusterTabel 4. Hasil Algoritma *K-Means* Menggunakan 3 Kluster (3,4 dan 5)

Kluster	Nilai SSE	K1	K2	K3	K4	K5
3	1,3%	Cashback	Aplikasi	Blibli	-	-
4	1,8%	Dana	Cashback	Blibli	Aplikasi	-
5	2,1%	Dana	Blibli	Aplikasi	Cashback	Bagus

Tabel 5. Hasil Algoritma *K-Medoids* Menggunakan 3 Kluster (3,4 dan 5)

Kluster	Nilai Objective Function	K1	K2	K3	K4	K5
3	1,594	Cashback	Aplikasi	Blibli	-	-
4	1,592	Aplikasi	Blibli	Dana	Cashback	-
5	1,590	Aplikasi	Cashback	Bagus	Dana	Blibli

Berdasarkan tabel 4 dapat dilihat bahwa nilai SSE terkecil pada algoritma *K-Means Clustering* adalah di jumlah kluster = 3. Kata yang sering muncul pada kluster 3 adalah *cashback*, aplikasi dan blibli. Kata yang sering muncul pada kluster 4 adalah dana, *cashback*, blibli dan aplikasi. Dan kata yang sering muncul pada kluster 5 adalah dana, blibli, aplikasi, *cashback* dan bagus.

Berdasarkan tabel 5 dapat dilihat bahwa nilai *Objective Function minimum* pada algoritma *K-Medoids Clustering* adalah di jumlah kluster = 5. Kata yang sering muncul pada kluster 3 adalah *cashback*, aplikasi, blibli. Kata yang sering muncul pada kluster 4 adalah aplikasi, blibli, dana dan *cashback*. Untuk kluster 5 kata yang sering muncul adalah aplikasi, *cashback*, bagus, dana dan blibli.

Berdasarkan tabel 4 dan 5 dapat dilihat bahwa dengan menggunakan algoritma *K-Means* dan algoritma *K-Medoids* menghasilkan kata yang muncul sama pada jumlah kluster = 3 yaitu *cashback*, aplikasi dan blibli.

4. KESIMPULAN

Dari hasil penelitian yang telah dilakukan dengan menggunakan Data Review aplikasi *ecommerce* Blibli mengimplementasikan 2 metode klusterisasi data yaitu *K-Means* dan *K-Medoids* dengan menggunakan *Silhouette Score* didapat jumlah *cluster* optimal adalah 4 kluster. Kata yang paling sering muncul dalam data teks *review* melalui *wordcloud* adalah Aplikasi, Blibli, Barang dan Belanja. Pada penelitian ini telah dilakukan percobaan 3 kali yaitu jumlah kluster sama dengan 3,4, dan 5. Pada percobaan dengan menggunakan algoritma *K-Means* didapatkan nilai SSE terkecil di jumlah kluster = 3. Pada percobaan dengan menggunakan algoritma *K-Medoids* didapatkan nilai *Objective Function minimum* di jumlah kluster = 5. Hasil *running* data dengan menggunakan algoritma *K-Means* dan algoritma *K-Medoids* menghasilkan kata yang sama pada jumlah kluster = 3 yaitu *cashback*, aplikasi dan blibli.

DAFTAR PUSTAKA

- [1] F. Dwi Handayani And I. Rosyida, "Clustering Review Pengguna Aplikasi Zenius Pada Layanan Google Play Store Menggunakan Metode Dbscan Dan Hdbscan," *Emerging Statistics And Data Science Journal*, Vol. 1, No. 2, 2023.

- [2] M. Afdal, L. Rahma Elita, P. Studi Sistem Informasi, F. H. Sains Dan Teknologi Uin Suska Riau Jl Soebrantas Km, And P. Pekanbaru -Riau, "Penerapan Text Mining Pada Aplikasi Tokopedia Menggunakan Algoritma K-Nearest Neighbor," *Jurnal Ilmiah Rekayasa Dan Manajemen Sistem Informasi*, Vol. 8, No. 1, 2022.
- [3] M. Nurjanah And T. Arifin, "Penerapan Algoritma K-Means Untuk Analisis Data Ulasan Di Situs Tripadvisor," *Jurnal Responsif*, Vol. 3, No. 1, Pp. 75–82, 2021, [Online]. Available: [Http://Ejurnal.Ars.Ac.Id/Index.Php/Jti](http://ejurnal.ars.ac.id/index.php/jti)
- [4] J. Mantik And N. Ayu Privandhani, "(2022) 1542-1550 Accredited," 2022.
- [5] A. Habib Husaini, R. Mayasari, And U. Singaperbangsa Karawang, "Pengelompokan Ulasan Aplikasi Pedulilindungi Dengan Algoritma K-Medoids Pedulilindungi Application Review Grouping With The K-Medoids Algorithm," *Journal Of Information Technology And Computer Science (Intecoms)*, Vol. 5, No. 2, 2022.
- [6] G. S. Sunarko,) Wasino, And T. Sutrisno, "Jurnal Ilmu Komputer Dan Sistem Informasi Klasterisasi Sentimen Ulasan Pengguna Aplikasi Bca Mobile Pada Platform Google Play Store Dengan Algoritma K-Means Clustering."
- [7] M. D. Pamungkas And H. Februriyanti, "Penerapan Algoritma K-Means Clustering Untuk Mengelompokkan Data Review Barang Pada E-Commerce Lazada," *Semantik*, Vol. 8, No. 2, P. 99, Dec. 2022, Doi: 10.55679/Semantik.V8i2.29058.
- [8] L. Petra Refialy, H. Maitimu, And M. Soyano Pesulima, "Perbaikan Kinerja Clustering K-Means Pada Data Ekonomi Nelayan Dengan Perhitungan Sum Of Square Error (Sse) Dan Optimasi Nilai K Cluster," 2021.
- [9] H. Lailatul Ramadhania, L. Zakaria, And Dan Nusyirwan, "Aplikasi Metode Sillhouette Coefficient, Metode Elbow Dan Metode Gap Statistic Dalam Menentukan K Optimal Pada Analisis K-Medoids," 2023.
- [10] H. Dame Tampubolon, M. Safii, And D. Suhendro, "Penerapan Algoritma K-Means Dan K-Medoids Clustering Untuk Mengelompokkan Tindak Kriminalitas Berdasarkan Provinsi," Vol. 2, No. 2, Pp. 6–12, 2021, [Online]. Available: [Http://Creativecommons.Org/Licences/By/4.0/](http://creativecommons.org/licenses/by/4.0/)