

# ANALISIS PERBINCANGAN DALAM GRUP WHATSAPP DENGAN K-MEANS CLUSTERING

Maria Tri Elsa\*<sup>1</sup>), Sunneng Sandino Berutu<sup>2</sup>), Febe Maedjaja<sup>3</sup>)

1. Universitas Kristen Immanuel, Yogyakarta, Indonesia
2. Universitas Kristen Immanuel, Yogyakarta, Indonesia
3. Universitas Kristen Immanuel, Yogyakarta, Indonesia

## Article Info

**Kata Kunci:** Metode Elbow; K-Means; *Text Clustering*; Whatsapp

**Keywords:** Elbow Method; K-Means; *Text Clustering*; Whatsapp,

## Article history:

Received 29 September 2024

Revised 13 Oktober 2024

Accepted 4 November 2024

Available online 4 December 2024

## DOI :

<https://doi.org/10.29100/jipi.v9i4.5554>

\* Corresponding author.

Maria Tri Elsa

E-mail address:

[mariatrielsa@gmail.com](mailto:mariatrielsa@gmail.com)

## ABSTRAK

Whatsapp merupakan sebuah aplikasi yang mempermudah pengguna untuk mengirim pesan teks, foto, video, melakukan panggilan suara, dan panggilan video secara gratis dengan koneksi internet. “Grup Whatsapp Jual Beli Area UKRIM” memiliki sekitar 26.000 arsip pesan, yang membuktikan tingginya aktivitas dan interaksi anggota dalam transaksi jual beli. Anggota grup akan terbantu membuat strategi promosi barang jika mereka dapat memanfaatkan informasi mengenai kata dan barang yang paling sering muncul dalam *chat group*. Untuk mendapatkan informasi mengenai kata dan barang yang paling sering muncul dalam *chat group* diperlukan *text clustering*. Penelitian ini menggunakan Metode K-Means dalam melakukan *text clustering* untuk memperoleh kata-kata yang sering muncul tersebut. Setelah melalui prosedur pra-pemrosesan teks dan penerapan Metode Elbow, jumlah data diperkecil menjadi 4.732 data dan ditentukan 10 *cluster* yang optimal dalam “Grup Whatsapp Jual Beli Area UKRIM”. Hasil penerapan K-Means memperlihatkan kata yang paling sering muncul dalam *cluster 0* adalah kata “jual”, dalam *cluster 1* kata “info”, dalam *cluster 2* kata “beli”, dalam *cluster 3* kata “japri”, dalam *cluster 4* kata “gas”, dalam *cluster 5* kata “motor”, dalam *cluster 6* kata “info”, “kost”, dalam *cluster 7* kata “hp”, dalam *cluster 8* kata “kucing”, dan dalam *cluster 9* kata “rak”. Secara menyeluruh kata yang paling banyak muncul adalah kata “info” dengan jumlah 910 dan kata yang paling sedikit muncul adalah kata “iphone” dan “hewan”. Barang dagangan yang sering muncul adalah “meja”, “kipas”, “lemari”, “kasur”, dan “laptop”.

## ABSTRACT

Whatsapp is an application that makes it easy for users to send text messages, photos, videos, make voice calls, and video calls for free with an internet connection. “Group Whatsapp Jual Beli Area UKRIM” has around 26,000 message archives, which proves the high activity and interaction of members in buying and selling transactions. Group members will be helped to create a strategy for promoting goods if they can utilize information about the words and goods that appear most frequently in the *group chat*. To get information about the words and items that appear most often in *chat group*, *text clustering* is needed. This research uses the K-Means Method in conducting *text clustering* to obtain the words that often appear. Through the text pre-processing process and the application of Elbow Method, the amount of data was reduced to 4,732, and 10 optimal *clusters* were obtained in the UKRIM Area Buying and Selling Whatsapp Group. The results of applying K-Means showed the most used word in *cluster 0* was the word “jual”, in *cluster 1* the word “info”, in *cluster 2* the word “beli”, in *cluster 3* the word “japri”, in *cluster 4* the word “gas”, in *cluster 5* the word “motor”, in *cluster 6* the word “info”, “kost”, in *cluster 7* the word “hp”, in *cluster 8* the word “kucing”, and in *cluster 9* the word “rak”. Overall, the word that appeared the most was the word “info” with a total of 910 and the words that appeared the least were the word “iphone” and “hewan”. Merchandise items that frequently appeared were “meja”, “kipas”, “lemari”, “kasur”, and “laptop”.

## I. PENDAHULUAN

WhatsApp merupakan sebuah aplikasi yang mempermudah pengguna untuk mengirim pesan teks, foto, video, melakukan panggilan suara, dan panggilan video secara gratis dengan koneksi internet. Fitur grup WhatsApp menjadi sarana yang sangat bermanfaat dalam memfasilitasi promosi, khususnya dalam konteks grup jual beli, seperti “Grup Whatsapp Jual Beli Area UKRIM”. Grup ini dibentuk pada tanggal 18 Maret 2020 sebagai wadah untuk promosi produk di area Universitas Kristen Immanuel (UKRIM) dan sekitarnya. Dengan jumlah anggota 882 orang, grup ini telah menjadi tempat aktif untuk berbagai transaksi jual beli. Dari tahun 2022 hingga 2024, “Grup Whatsapp Jual Beli Area UKRIM” telah mengakumulasi pesan dengan jumlah 26.343 data. Statistik pertumbuhan “Grup Whatsapp Jual Beli Area UKRIM” menunjukkan tren yang positif, dimana setiap hari terjadi penambahan anggota baru dan aktivitas promosi dari para penjual yang konsisten, setidaknya 30 pesan dari para penjual yang mempromosikan barang dagangan mereka setiap hari. Hal ini mencerminkan minat yang terus berkembang dari mahasiswa dan masyarakat sekitar UKRIM. Partisipasi yang aktif dari penjual dalam memanfaatkan *platform* ini untuk meningkatkan penjualan membuktikan tingkat aktivitas dan interaksi yang tinggi antar anggota dalam berbagai transaksi jual beli. Penelitian ini menganalisis kata dan barang yang paling banyak diperbincangkan, diharapkan dapat memberikan informasi kepada anggota grup untuk mempermudah mengambil keputusan dalam strategi promosi di “Grup Whatsapp Jual Beli Area UKRIM”.

Untuk memahami kata-kata yang sering diperbincangkan dalam grup dan barang apa yang paling banyak dipromosikan, penelitian ini menerapkan *Text Clustering* menggunakan Metode K-Means. *Clustering* adalah teknik untuk mengelompokkan data yang sejenis ke dalam *cluster* atau kelompok sehingga data yang ada didalam *cluster* tersebut serupa [1]. Metode yang diterapkan pada penelitian ini adalah K-Means, Algoritma K-Means adalah algoritma pengelompokan data berdasarkan kemiripan, yang didasarkan pada penentuan k titik awal sebagai pusat *cluster* dan mengelompokkan data berdasarkan jarak Euclidean antara setiap titik data dan centroid terdekat [2]. Metode K-Means mengelompokkan data ke dalam *cluster* dengan cara membaginya berdasarkan kesamaan karakteristik[3]. Data yang memiliki kemiripan yang tinggi dikelompokkan ke dalam *cluster* yang sama, dan data yang memiliki perbedaan karakteristik yang signifikan ditempatkan ke dalam *cluster* yang berbeda. Metode K-Means dikenal karena kecepatan dan kemampuannya dalam mengelompokkan dokumen dalam, menganalisis citra dan berbagai bidang lain. Alasan utama penggunaan K-Means adalah kemampuan K-Means dalam menangani data berukuran besar dengan tingkat ketelitian yang cukup tinggi terhadap objek dan K-Means tidak terpengaruh dengan urutan objek. K-Means mampu mengelompokkan objek besar secara cepat sehingga proses pengelompokkan lebih cepat [4]. Dengan demikian K-Means menjadi pilihan tepat untuk analisis “Grup Whatsapp Jual Beli Area UKRIM” yang memiliki banyak kumpulan data *chat*.

Penelitian menggunakan K-Means Clustering telah dilakukan oleh Nirwana Hendrastuty(2024) penerapan data mining dalam implementasi Algoritma K-Means, untuk evaluasi hasil pembelajaran siswa sehingga didapatkan pola tersembunyi dalam hasil pembelajaran, mengelompokkan siswa sesuai dengan tingkat pencapaian, dan menghasilkan informasi yang untuk guru dan *stakeholder* pendidikan [5]. Penelitian oleh Allbila, Ibnu, dan Fitri (2023) melakukan penerapan Algoritma K-Means untuk mendapatkan *cluster* nilai ujian nasional, yang bertujuan untuk mendapatkan pola dalam hasil ujian dan membantu pemahaman lebih lanjut terkait karakteristik kelompok nilai yang berbeda [6]. Penelitian oleh Ramzi dan Iwan (2024) optimasi Algoritma K-Means untuk mendapatkan *cluster* terkait nilai ujian nasional yang bertujuan untuk memperjelas informasi dalam data nilai ujian nasional Sekolah Menengah Kejuruan (SMK) di Jawa Tengah tahun 2019 [7]. Penelitian “Analisis Perbincangan dalam GrupWhatsapp dengan K-Means Clustering” ini menggunakan objek *chat* “Grup Whatsapp Jual Beli Area UKRIM” yang sebelumnya belum pernah dianalisis secara mendalam.

## II. METODE PENELITIAN

Penelitian ini menerapkan metode K-Means sebagai pendekatan utama untuk analisis data. Proses penelitian dilakukan melalui 5 tahap yaitu: pengumpulan data, pra-pemrosesan teks, vektorisasi teks, menentukan jumlah *cluster*, dan implementasi Metode K-Means. Tahap penelitian ini dapat dilihat pada Gambar 1.



Gambar. 1 Tahap Penelitian

### A. Pengumpulan Data

Pengumpulan data yaitu langkah pengambilan informasi atau fakta yang penting untuk keperluan penelitian, analisis, atau pemahaman lebih lanjut tentang suatu topik masalah [8]. Penelitian ini menganalisis data yang berasal dari *chat* “Grup Whatsapp Jual Beli Area UKRIM”. Proses pengumpulan data dilakukan dengan mengestrak data dari grup Whatsapp, serta menghubungkan Smartphone ke aplikasi Pushbullet untuk mengakses API dari *chat group* tersebut. Data *chat* yang terkumpul dari periode tahun 2022 hingga 2024 berjumlah 26.343 data. Informasi dalam data *chat* Whatsapp mencakup tanggal, nama pengirim, dan isi pesan yang disampaikan.

### B. Pra-Pemrosesan Teks

Pra-pemrosesan teks (*text preprocessing*) adalah langkah yang dilakukan pada data teks sebelum data di proses dan penghapusan *noise* dalam data [9]. Tujuannya yaitu untuk membersihkan dan mempersiapkan data teks agar dapat digunakan dengan lebih efektif dalam berbagai jenis analisis teks [10]. Langkah pra-pemrosesan teks meliputi :

#### 1. *Cleaning*

Pada proses ini yang dilakukan adalah menghilangkan karakter khusus, seperti tanda baca yang tidak relevan, mengurangi *noise* dan mempersiapkannya agar dapat dianalisis lebih lanjut dengan akurat [11]. Pada proses ini meliputi: ubah huruf kapital menjadi huruf kecil, hapus angka, hapus emoji, hapus url, hapus tanda baca [12]. Proses ini bertujuan untuk meningkatkan kualitas kebersihan data teks.

#### 2. *Tokenization*

*Tokenization* adalah proses memecah teks menjadi bagian-bagian yang lebih kecil, disebut sebagai token [13]. Proses ini memecah teks menjadi beberapa kata atau unit-unit yang lebih kecil, menghapus tanda baca serta karakter khusus, dan mengonversi huruf menjadi huruf kecil [14]. Sebagai contoh pada kalimat: “Saya sedang membaca buku”, akan diubah menjadi kata: “saya”, “sedang”, “membaca”, “buku”. Tujuan utama dari *tokenization* yaitu untuk memecah kalimat menjadi kata-kata atau token agar dapat diproses lebih lanjut.

#### 3. *Stopwords Removal*

*Stopwords removal* yaitu proses untuk menghilangkan atau menghapus kata-kata yang sering muncul namun tidak memiliki arti [15]. Contoh stop word meliputi kata-kata seperti “dan”, “atau”, “yang”, “di”, “dari”, “ada” dan sebagainya. Dengan menghapus *stopwords*, teks yang diproses menjadi lebih fokus pada kata-kata kunci yang lebih penting untuk analisis.

#### 4. *Stemming*

*Stemming* adalah proses dalam pemrosesan teks yang mengubah kata-kata ke bentuk dasarnya atau menghilangkan imbuhan kata [16]. Proses ini dilakukan untuk mengurangi variasi kata yang memiliki akar kata yang sama ke bentuk yang seragam, sehingga meningkatkan konsistensi dan memudahkan analisis teks. Tujuan proses ini adalah untuk menyatukan variasi kata yang memiliki akar kata sama agar dapat dianalisis lebih mudah dan mengubah kata sesuai dengan aturan penulisan Bahasa Indonesia yang benar [14]. Pada proses *stemming* menggunakan library *sastrawi* pada Python.

### C. Vektorisasi Teks

Vektorisasi teks merupakan proses transformasi data teks menjadi format numerik yang dapat dipahami dan diproses oleh algoritma machine learning. Tujuannya adalah untuk mengekstrak informasi kontekstual dari teks

sambil mempertahankan struktur datanya. Vektorisasi teks ini memungkinkan mesin atau algoritma machine learning dalam memproses data teks dan memperoleh pemahaman dari pesan yang terdapat dalam *chat* “Grup Whatsapp Jual Beli Area UKRIM”. Penelitian ini menggunakan TF-IDF(*Term Frequency-Inverse Document Frequency*) dalam vektorisasi teks.

#### D. Menentukan Jumlah Cluster

Proses ini memiliki signifikansi yang besar karena merupakan langkah awal dalam menyusun kelompok teks yang memiliki kesamaan. Penelitian ini menggunakan Metode Elbow untuk menemukan jumlah *cluster* yang paling sesuai. Metode Elbow adalah suatu teknik dalam analisis *cluster* untuk menemukan jumlah *cluster* yang optimal dalam sebuah dataset. Grafik hasil dari Metode Elbow menunjukkan bahwa jumlah *cluster* yang optimal dapat diidentifikasi dengan jelas melalui penurunan yang paling besar. Pada suatu titik, akan terjadi penurunan drastis dalam grafik, yang ditandai dengan adanya sebuah lekukan yang disebut kriteria siku[17]. Metode elbow digunakan untuk menentukan jumlah *cluster* terbaik dengan menganalisis perbandingan antara jumlah *cluster* dan tingkat penurunan dalam grafik. Jika terdapat penurunan yang signifikan antara jumlah *cluster* yang berturut-turut, seperti dari  $K=2$  ke  $K=3$  dan dari  $K=3$  ke  $K=4$ , yang menghasilkan sudut pada grafik. Nilai *cluster* yang menciptakan penurunan paling tajam dianggap sebagai jumlah *cluster* yang optimal. Jika penurunan drastis pada titik  $K=3$  maka jumlah *cluster* yang ideal adalah  $K=3$  [7]. Nilai  $k$  yang terbaik kemudian digunakan sebagai dasar untuk melakukan proses *clustering* dengan K-Means.

#### E. Implementasi Metode K-Means

Implementasi Metode K-Means dalam analisis data melibatkan penggunaan Algoritma K-Means untuk mengelompokkan data berdasarkan tingkat kemiripan antar data. Metode ini membagi data ke dalam *cluster* sehingga data yang serupa akan dikelompokkan ke dalam satu *cluster* yang sama dan data yang mempunyai karakteristik yang berbeda dikelompokkan ke dalam *cluster* berbeda [18]. Proses K-Means dimulai dengan menentukan jumlah *cluster* yang diinginkan, lalu secara acak memilih titik awal sebagai pusat *cluster*. Data akan dikelompokkan dalam *cluster* berdasarkan jarak Euclidean antara setiap titik data dengan pusat *cluster* terdekat.

### III. HASIL DAN PEMBAHASAN

Dalam proses pengambilan data dari *chat group*, langkah pertama dilakukan dengan export *chat* whatsapp. Selanjutnya, Smartphone dihubungkan ke Aplikasi Pushbullet untuk mengirimkan export riwayat *chat*. Dengan menggunakan aplikasi Pushbullet, API Key dari *chat* tersebut diperoleh yang kemudian dapat digunakan untuk mengolah data dengan bahasa pemrograman Python. Gambar 2 merupakan contoh *chat group*.



Gambar. 2 Screenshot chat

```
+ Code + Text
'13/05/22, 18:10 - +62 812-5540-326: Menarikk\n',
'13/05/22, 21:37 - +62 816-3216-2380: Ada yg jual tabung gas ijo?\n',
'13/05/22, 21:38 - +62 856-4368-7654: <Media omitted>\n',
'13/05/22, 21:38 - +62 856-4368-7654: Ada\n',
'13/05/22, 21:38 - +62 856-4368-7654: Kosong an 159 rb\n',
'16/05/22, 00:00 - +62 813-5407-2773: ada jual meja?\n',
'16/05/22, 08:05 - +62 813-6187-8963: <Media omitted>\n',
'16/05/22, 12:19 - +62 852-4387-4572: joined using this group's invite link\n',
'16/05/22, 14:27 - +62 822-7207-7215: joined using this group's invite link\n',
'16/05/22, 15:00 - +62 856-5106-7133: Ada yg jual spiker kah?\n',
'16/05/22, 15:25 - +62 883-2570-9600: This message was deleted\n',
'16/05/22, 15:25 - +62 853-4981-3202: <Media omitted>\n',
'16/05/22, 15:47 - Tory TS: Berapa bang?\n',
'16/05/22, 16:15 - +62 853-2946-7512: Kk kalau ngrim stiker agak sopan dikit yaa?\n',
'Ini group umum bukan group pribadi\n',
'16/05/22, 16:16 - +62 853-2946-7512: Jangan terlalu bar2 ngrim striker.\n',
'Sopan dikit\n',
'16/05/22, 16:53 - +62 823-2504-2113: <Media omitted>\n',
'17/05/22, 13:10 - +62 813-4407-7508: Omom tantatante\n',
'Ada yg jualan rescuers boleh di up\n',
'Terimakasih\n',
'17/05/22, 16:57 - +62 822-7203-6204: <Media omitted>\n',
'17/05/22, 16:57 - Anastasya TI: <Media omitted>\n',
'17/05/22, 19:48 - +62 813-9224-5959: Dijual harga brp ya?\n',
'17/05/22, 20:38 - +62 822-7203-6204: Pc kak\n',
'17/05/22, 22:51 - +62 823-1890-2155: joined using this group's invite link\n',
'17/05/22, 22:52 - +62 813-4882-2757: joined using this group's invite link\n',
'17/05/22, 22:54 - +62 895-6178-39969: joined using this group's invite link\n',
'18/05/22, 02:48 - +62 888-6444-282: joined using this group's invite link\n',
'18/05/22, 06:58 - Your security code with +62 813-4705-8670 changed. Tap to learn more.\n',
'18/05/22, 06:59 - Your security code with +62 813-4705-8670 changed. Tap to learn more.\n',
'18/05/22, 11:24 - +62 838-4323-3027: Info kos an putri budget 400 isian dong kak.\n',
'18/05/22, 11:26 - +62 857-4242-1453: Info kos an putri budget 400 isian dong kak.\n',
0s completed at 14:18
```

Gambar. 3 Hasil data chat

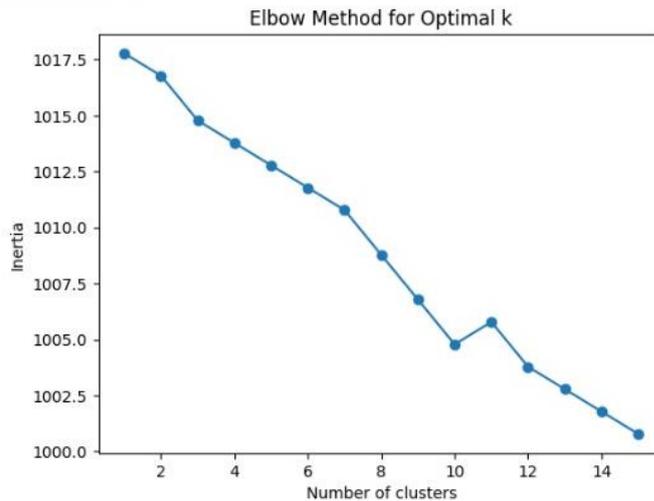
Gambar 3 menunjukkan bahwa pengambilan data *chat group* dengan 3 variabel yaitu: tanggal, pengirim, dan pesan. Data yang didapat dari “Grup Whatsapp Jual Beli Area UKRIM” yaitu 26.343 data dari tanggal 12 Mei 2022 sampai 7 Februari 2024. Data *chat group* masih memiliki banyak *noise*, simbol, tanda baca, *emoticon*, pesan kosong, pesan yang tidak bermakna. Sehingga perlu dilakukan pembersihan data, setelah melalui proses pembersihan data tersisa sejumlah 4.732 data. Berkurangnya jumlah data dikarenakan terdapat penghapusan kata yang tidak relevan seperti: “*this message was deleted*”, “*media omitted*”, pesan duplikat, *emoticon*, simbol, dan penghapusan kata yang tidak memiliki arti. Contoh hasil data *chat whatsapp* sebelum dan sesudah pra-pemrosesan ditunjukkan pada Tabel I.

TABEL I  
 SATUAN HASIL PRA-PEMROSESAN TEKS

Tahap pra-pemrosesan	Sebelum pra-pemrosesan	Sesudah pra pemrosesan
Cleaning	Ges, ada info yang jual kompor satu tungku ? Kalau ada PC ya. Tenkiu ðŸ™¼ðŸŒŸðŸŒŸ	Ges ada info yang jual kompor satu tungku Kalau ada PC ya Tenkiu
Tokenization	Ges ada info yang jual kompor satu tungku Kalau ada PC ya Tenkiu	['ges', 'ada', 'info', 'yang', 'jual', 'kompor', 'satu', 'tungku', 'kalau', 'ada', 'PC', 'ya', 'tenkiu']
Stop words removal	Ada yang punya info kosan cewek daerah kalasan	['info', 'kosan', 'cewek', 'daerah', 'kalasan']
Stemming	Ada yang jual catokan	jual catok

Pada proses vektorisasi teks, data teks akan diubah menjadi *vektor* karena menggunakan Metode K-Means data harus berupa *vektor*. Teks dalam bentuk aslinya berupa karakter atau kata-kata yang tidak mudah diproses oleh komputer secara langsung. Oleh karena itu, diperlukan konversi teks dalam bentuk numerik agar dapat digunakan untuk analisis tentang teks tersebut [19]. Vektorisasi teks pada *chat group* ditemukan jumlah dokumen 4.732 dan jumlah fitur 4051. Setiap teks dalam kumpulan data *chat* tersebut dipresentasikan sebagai vektor dalam ruang fitur dengan dimensi 4051, dengan jumlah total dokumen 4732. Jumlah fitur ini mencerminkan keragaman kata-kata yang ada dalam kumpulan data *chat group*. Setiap dokumen memiliki kombinasi kata-kata yang berbeda, yang menghasilkan *vektor* yang unik untuk setiap dokumen. Algoritma otomatis mengelompokkan teks-teks yang memiliki pola yang mirip.

Dalam menentukan jumlah *cluster* dengan Metode Elbow penulis menentukan jumlah k maksimal kemudian grafik hasilnya akan menunjukkan *cluster* yang terbaik. Pada penelitian ini didapat 10 *cluster* terbaik untuk pengelompokan data. Hasil dari Metode Elbow dapat dilihat pada Gambar 4.

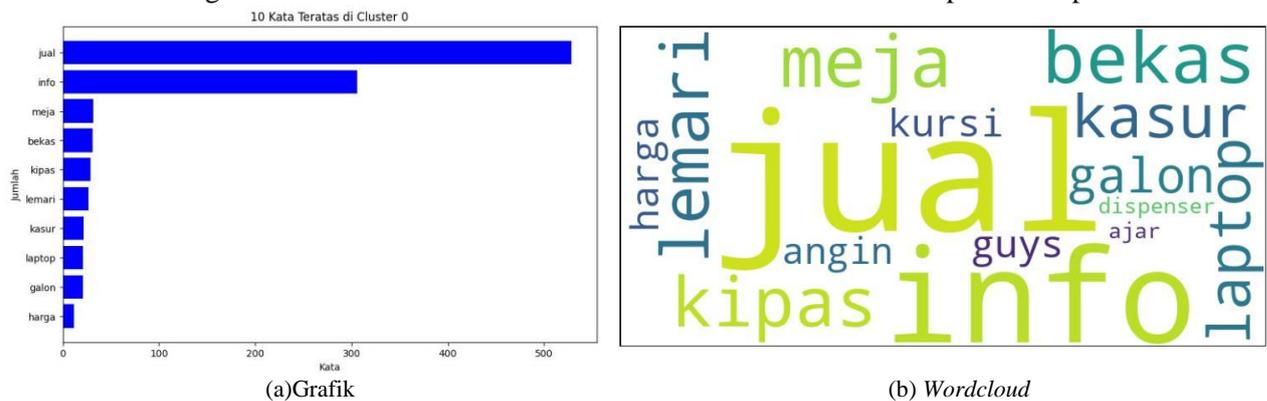


Gambar. 4 Hasil Jumlah *Cluster* dengan Metode Elbow

Penelitian ini menghasilkan identifikasi terhadap 10 kelompok *cluster* yang terbentuk dalam *chat group*, yang mencerminkan penggunaan kata-kata yang paling sering digunakan dalam *chat* “Grup Jual Beli Area UKRIM”. Setiap *cluster* memiliki kata-kata dominan.

### 1. Cluster 0

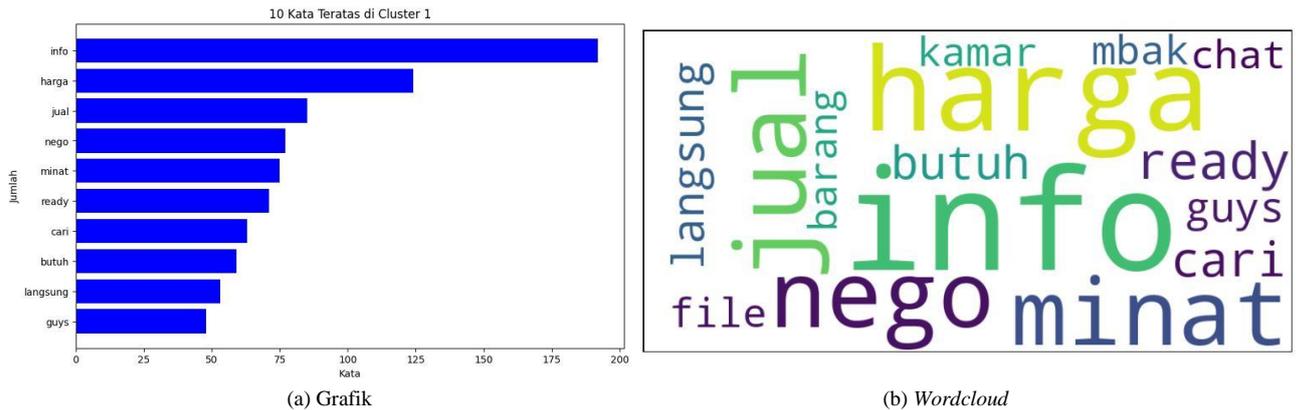
Pada *cluster* 0 ditemukan bahwa 3 kata teratas yang paling sering muncul adalah “jual” dengan jumlah 528, “info” dengan jumlah 306, dan “meja” dengan jumlah 32. Analisis menunjukkan bahwa, terdapat beberapa kata lain yang dominan muncul seperti “bekas”, “kipas”, “lemari”, “kasur”, “laptop”, dan “galon”. Kemunculan kata-kata ini mengindikasikan bahwa topik yang sering dibahas dalam obrolan grup ini berkaitan dengan penjualan, permintaan barang-barang bekas seperti perabotan rumah tangga(lemari, kasur), elektronik(laptop, kipas), dan kebutuhan sehari-hari (galon). Tren utama dalam *cluster* ini adalah aktivitas penjualan barang, permintaan informasi terkait dengan penjualan, dan fokus pada barang-barang bekas dan barang elektronik serta kebutuhan sehari-hari. Hasil dari *cluster* 0 dapat dilihat pada Gambar 5.



Gambar. 5 Hasil *cluster* 0 dengan kata paling sering muncul ditampilkan dalam grafik bar dan *wordcloud*

### 2. Cluster 1

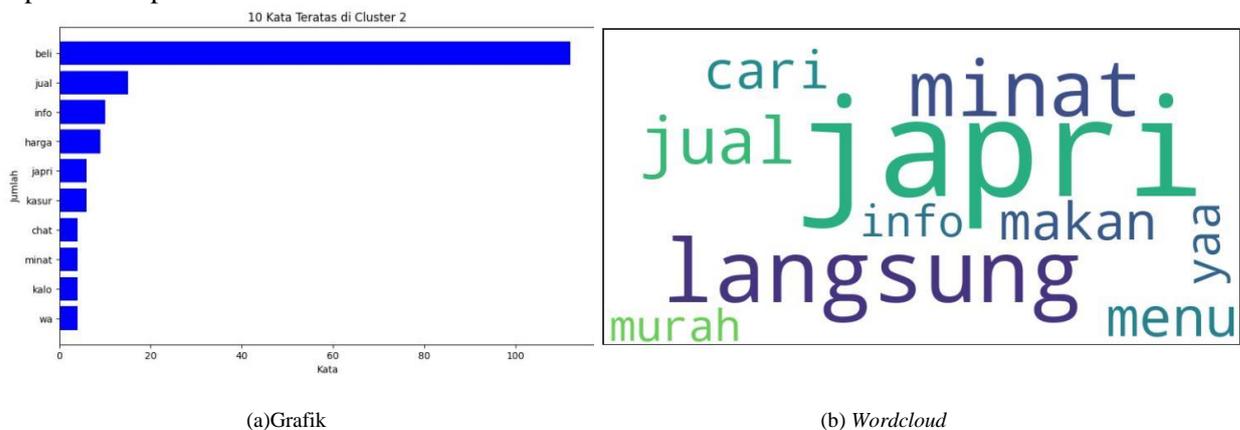
Pada *cluster* 1, terdapat kata “info” yang paling dominan dengan jumlah kemunculan sebanyak 192 kali, dan kata-kata lain seperti “harga”, “jual”, “nego”, “minat”, “ready”, “cari”, “butuh”, “langsung” dan “guys”. Kemunculan kata-kata ini mengindikasikan bahwa *cluster* ini didominasi oleh percakapan terkait pertukaran informasi harga, penjualan, negosiasi, minat pembelian, ketersediaan barang, pencarian barang. Pada *cluster* ini menunjukkan bahwa anggota grup berinteraksi secara aktif untuk bertukar informasi terkait dengan jual beli, serta untuk melakukan transaksi negosiasi. Tren utama dalam *cluster* ini adalah aktivitas yang tinggi dalam pertukaran informasi terkait jual beli, intensitas interaksi yang tinggi, transaksi negosiasi yang sering terjadi, dan responsif terhadap permintaan transaksi. Dari *cluster* ini, peserta grup dapat mengambil beberapa manfaat tentang informasi harga, negosiasi transaksi, ketersediaan barang dan peningkatan jaringan. Hasil penelitian pada *cluster* 1 dapat dilihat pada Gambar 6.



Gambar. 6 Hasil *cluster* 1 dengan kata paling sering muncul ditampilkan dalam grafik bar dan *wordcloud*

### 3. Cluster 2

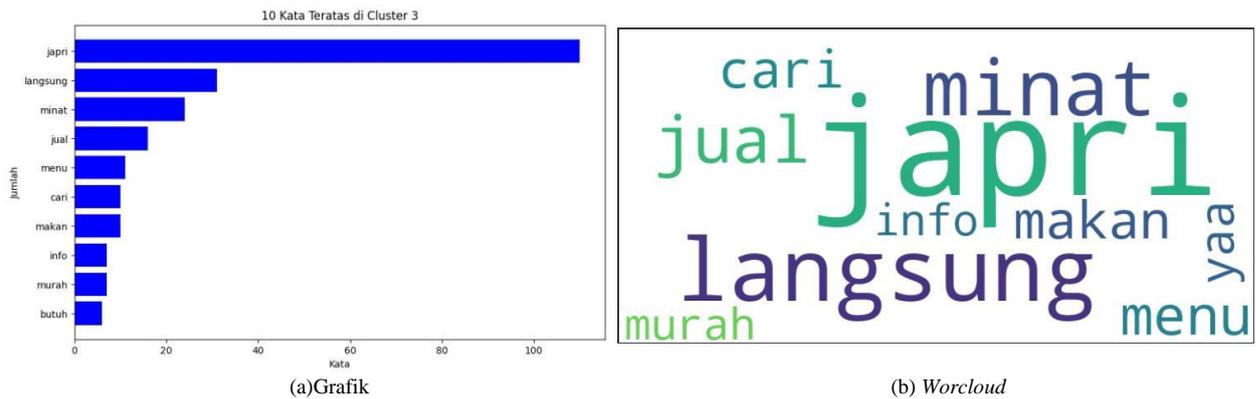
Pada *cluster* 2 peneliti menemukan kata “beli” menjadi kata yang paling sering muncul dengan jumlah 112, membuktikan bahwa fokus utama percakapan grup whatsapp dalam *cluster* ini adalah terkait dengan proses pembelian atau permintaan barang. Kata-kata seperti “jual”, “info”, “harga”, “japri”, “kasur”, “chat”, “minat”, “kalo” dan “wa” juga dominan muncul. Kemunculan kata-kata ini mencerminkan bahwa peserta grup ini aktif dalam pertukaran informasi tentang barang yang dijual atau dibeli hingga proses komunikasi antar anggota grup untuk menegosiasikan harga, menyatakan minat maupun berkomunikasi secara lebih pribadi dari kata “japri”. Tren utama dalam *cluster* ini adalah fokus pada pembelian atau permintaan barang, pertukaran informasi terkait dengan penjualan dan pembelian, serta adanya komunikasi yang lebih pribadi antara anggota grup. Untuk peserta grup informasi ini dapat digunakan untuk memperoleh informasi tentang barang yang dijual, berpartisipasi dalam proses negosiasi harga, memperluas jaringan komunikasi dengan anggota lain untuk tujuan bisnis. Hasil penelitian dari *cluster* 2 dapat dilihat pada Gambar 7.



Gambar. 7 Hasil *cluster* 2 dengan kata paling sering muncul ditampilkan dalam grafik bar dan *wordcloud*

### 4. Cluster 3

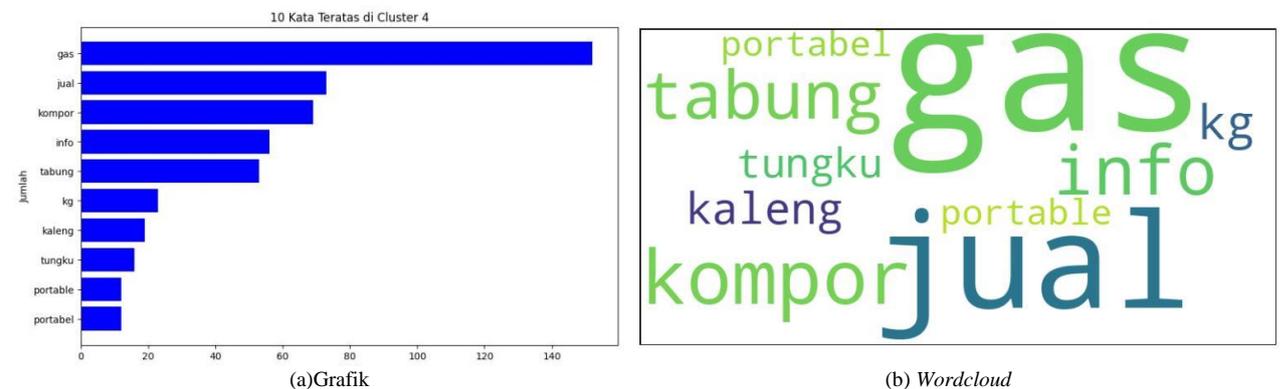
Pada *cluster* 3 peneliti menemukan kata yang paling dominan adalah “japri”, yang membuktikan adanya preferensi untuk berkomunikasi secara pribadi dengan jumlah kemunculan sejumlah 110. Terdapat juga kata-kata seperti “langsung”, “minat”, “jual”, “menu”, “cari”, “makan”, “info”, “murah”, dan “butuh”. Kemunculan kata-kata ini membuktikan beragamnya topik pembicaraan termasuk permintaan informasi, minat terhadap barang yang ditawarkan, pertukaran informasi mengenai menu atau makanan dan kebutuhan akan barang murah. Anggota grup berkomunikasi secara intensif, dalam melakukan transaksi dan negosiasi secara pribadi dari kata “japri”. Tren utama dalam *cluster* ini adalah presensi untuk berkomunikasi secara pribadi, beragamnya topik pembicaraan, intensitas komunikasi yang tinggi, dan fokus pada transaksi dan negosiasi yang lebih personal. Peserta grup dapat mengambil manfaat dari partisipasi aktif dalam *cluster* ini, termasuk privasi dan keamanan dalam berkomunikasi, kemampuan untuk melakukan negosiasi yang lebih efektif. Hasil dari penelitian ini dapat dilihat pada Gambar 8.



Gambar. 8 Hasil *cluster* 3 dengan kata paling sering muncul ditampilkan dalam grafik bar dan *wordcloud*

### 5. Cluster 4

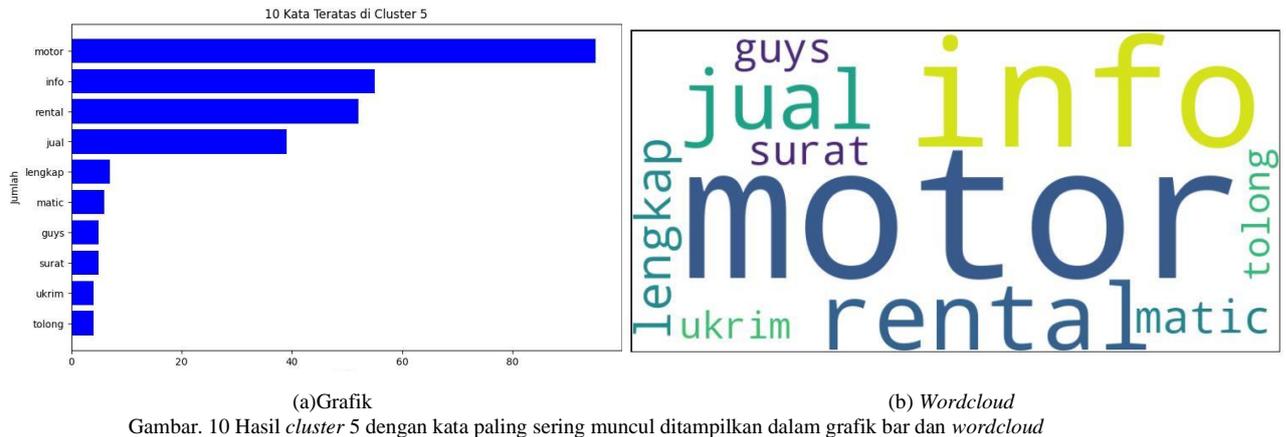
Pada *cluster* 4 peneliti menemukan bahwa kata yang paling dominan adalah “gas” dengan jumlah kemunculan 152 kali. Selain itu terdapat juga kata-kata “jual”, “kompor”, “info”, “tabung”, “kg”, “kaleng”, “tungku”, “portable”, dan “portabel”. Kemunculan kata-kata ini mengindikasikan bahwa pembahasan dalam *cluster* ini secara khusus berkaitan dengan topik seputar kompor dan gas. *Cluster* ini membuktikan bahwa grup ini menjadi tempat yang penting untuk bertukar informasi bagi anggota grup, memperoleh dan memberikan rekomendasi, serta melakukan transaksi terkait pemasaran produk-produk gas dan peralatan masak. Tren dalam *cluster* ini topik seputar kompor dan gas, pertukaran informasi dan rekomendasi, transaksi terkait pemasaran produk. Peserta grup mendapatkan informasi mendetail, rekomendasi produk, peluang transaksi, dan pengembangan pengetahuan terkait dengan produk gas dan peralatan masak. Hasil dari penelitian ini dapat dilihat pada Gambar 9.



Gambar. 9 Hasil *cluster* 4 dengan kata paling sering muncul ditampilkan dalam grafik bar dan *wordcloud*

### 6. Cluster 5

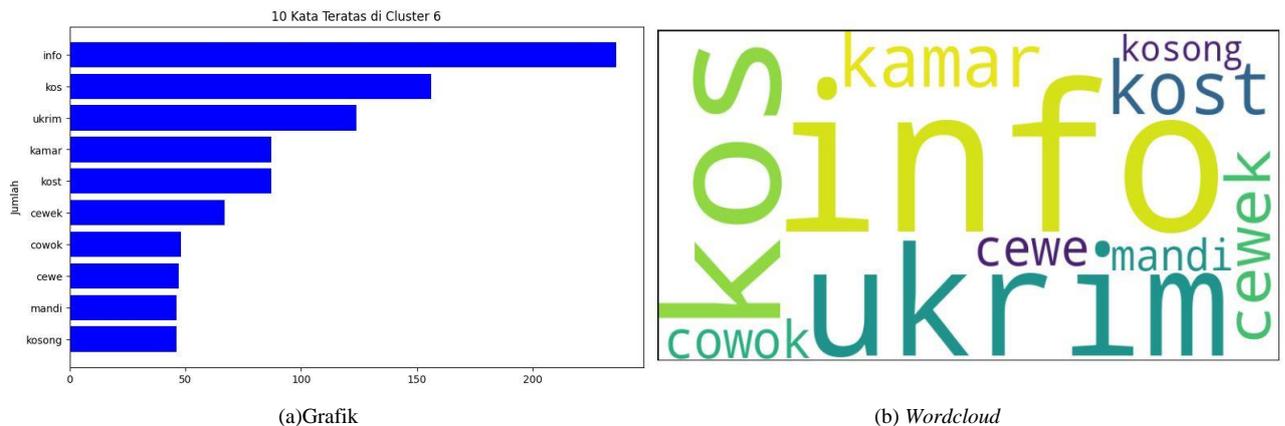
Pada *cluster* 5 ditemukan kata yang paling dominan muncul adalah “motor” dengan jumlah 95 kali. Terdapat kata-kata lain seperti “info”, “rental”, “jual”, “lengkap”, “matic”, “guys”, “surat”, “ukrim”, dan “tolong”. Diskusi dalam *cluster* ini secara umum pada topik penjualan dan rekomendasi mengenai kendaraan, khususnya motor. Anggota grup berbagi informasi mengenai penawaran motor untuk dijual, layanan rental motor, dan diskusi mengenai jenis dan spesifikasi motor. Kata-kata seperti “lengkap” dan “surat” membuktikan adanya pembicaraan mengenai kelengkapan dokumen dan persyaratan yang terkait dengan transaksi penjualan atau sewa motor. Tren utama dalam *cluster* ini adalah fokus pada penjualan, rental dan rekomendasi motor, pembicaraan mengenai kelengkapan dokumen dan persyaratan transaksi. Dengan demikian, *cluster* ini dapat memberikan manfaat yang berharga bagi peserta grup dalam mendapatkan informasi mendetail tentang penjualan dan rental motor, rekomendasi motor, pemahaman tentang persyaratan transaksi. Hasil penelitian pada *cluster* 5 dapat dilihat pada Gambar 10.



Gambar. 10 Hasil *cluster* 5 dengan kata paling sering muncul ditampilkan dalam grafik bar dan *wordcloud*

### 7. Cluster 6

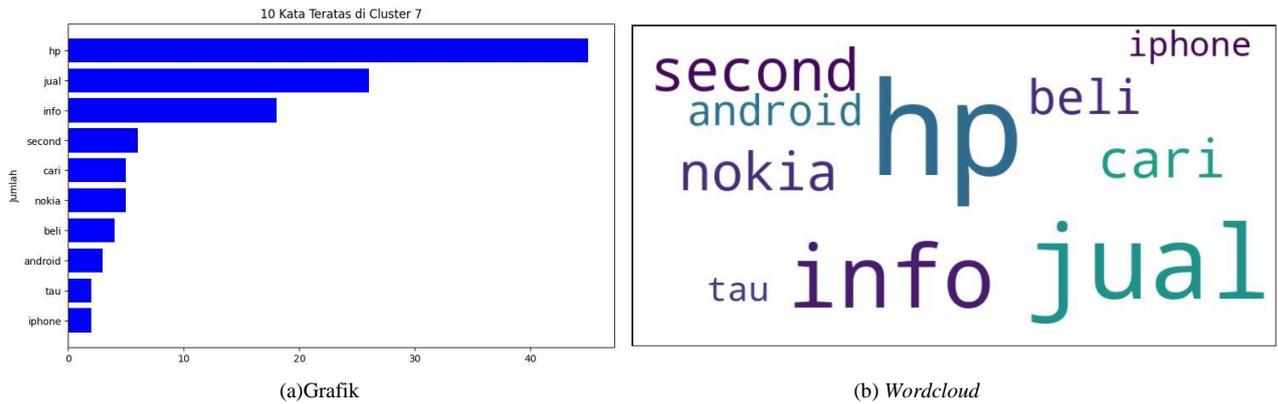
Pada *cluster* 6 ditemukan kata yang paling sering muncul adalah “info” dengan jumlah 236 kali. Serta kata “kos”, “ukrim”, “kamar”, “kost”, “cewek”, “cowok”, “cewe”, “mandi”, dan “kosong”. Kata-kata yang paling muncul ini membuktikan bahwa *cluster* 6 membahas tentang informasi seputar kos-kosan di sekitar area UKRIM. Anggota grup berbagi informasi mengenai ketersediaan kamar kost, fasilitas yang ditawarkan, harga sewa, serta informasi tambahan terkait dengan kos-kosan. Kata-kata seperti “cewek”, “cowok”, dan “mandi” mengindikasikan adanya obrolan mengenai jenis kos-kosan yang tersedia, termasuk kamar untuk perempuan dan laki-laki serta fasilitas mandi yang disediakan. Tren utama *cluster* ini adalah informasi tentang kos-kosan, pembahasan tentang jenis kos dan fasilitas. Manfaat dari *cluster* ini dapat memudahkan peserta grup yang didominasi oleh mahasiswa dan masyarakat sekitar UKRIM untuk menemukan kos yang sesuai. Hasil penelitian pada *cluster* 6 dapat dilihat pada Gambar 11.



Gambar. 11 Hasil *cluster* 6 dengan kata paling sering muncul ditampilkan dalam grafik bar dan *wordcloud*

### 8. Cluster 7

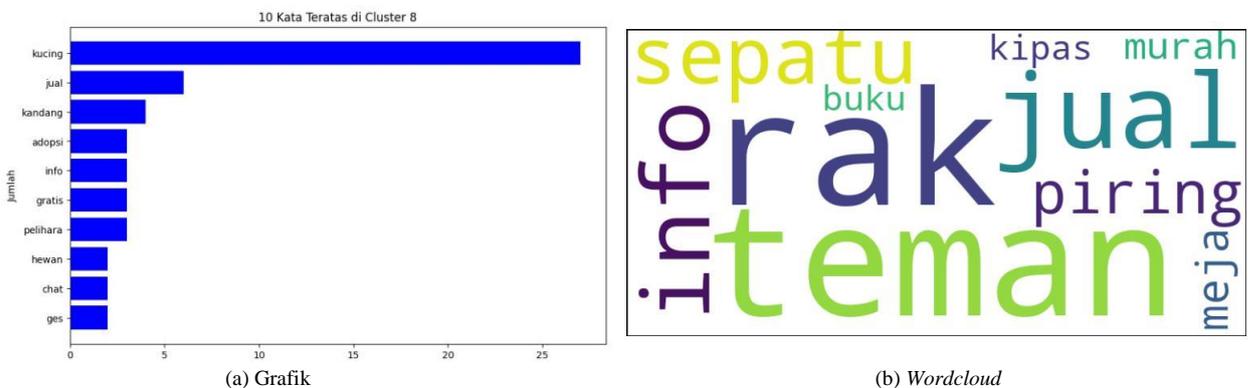
Pada *cluster* 7 ditemukan kata yang paling sering muncul adalah “hp” dengan jumlah 45 kali. Terdapat juga kata-kata seperti “jual”, “info”, “second”, “cari”, “nokia”, “beli”, “android”, “tau” dan “iphone”. Pada *cluster* 7 peserta grup banyak membahas tentang penjualan dan informasi terkait dengan handphone atau smartphone. Hasil *cluster* 7 membuktikan bahwa anggota grup saling berbagi informasi mengenai penawaran ponsel bekas (second), mencari ponsel tertentu dari kata “cari” maupun bertukar informasi mengenai berbagai merek ponsel seperti Nokia, Android, dan iPhone. Tren utama *cluster* ini yaitu penjualan dan informasi *handphone* bekas, ini dapat membantu peserta grup untuk mendapatkan informasi terkait *handphone* yang di cari. Hasil penelitian pada *cluster* 7 dapat dilihat pada Gambar 12.



Gambar. 12 Hasil *cluster* 7 dengan kata paling sering muncul ditampilkan dalam grafik bar dan *wordcloud*

### 9. Cluster 8

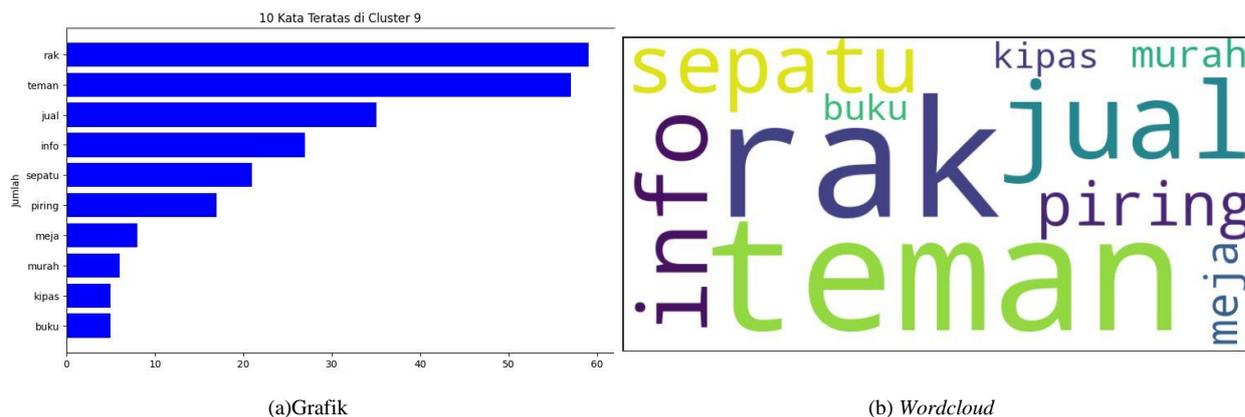
Pada *cluster* 8 ditemukan kata yang dominan adalah “kucing” dengan jumlah kemunculan 27 kali. Selain itu terdapat juga kata-kata seperti “jual”, “kandang”, “adopsi”, “info”, “gratis”, “peliharaan”, “hewan”, “chat”, dan “ges”. Analisis membuktikan bahwa *cluster* ini berkaitan dengan topik seputar kucing, penjualan, adopsi, serta informasi tentang hewan peliharaan. Adanya obrolan anggota grup tentang kucing yang ingin dijual, dicari, atau diadopsi, serta bertukar tips dan saran terkait perawatan dan pengelolaan hewan peliharaan. Kata “gratis” yang ditemukan mengindikasikan adanya penawaran hewan secara cuma-cuma. Hasil *cluster* ini membuktikan bahwa anggota “Grup Whatsapp Jual Beli Area UKRIM” terdapat pencinta hewan dan saling berbagi informasi mengenai hewan peliharaan. Tren dalam *cluster* adalah seputar kucing dan hewan peliharaan yang dapat membantu para pencinta hewan di dalam grup untuk mencari informasi terkait adopsi, penjualan dan perawatan tentang hewan peliharaan. Hasil penelitian pada *cluster* 8 dapat dilihat pada Gambar 13.



Gambar. 13 Hasil *cluster* 8 dengan kata paling sering muncul ditampilkan dalam grafik bar dan *wordcloud*

### 10. Cluster 9

Pada *cluster* 9 terdapat kata “rak” yang paling dominan muncul dengan jumlah 59 kali. Terdapat juga kata-kata seperti “teman”, “jual”, “info”, “sepatu”, “piring”, “meja”, “murah”, “kipas”, dan “buku”. Analisis ini mengindikasikan bahwa *cluster* ini berkaitan dengan topik terkait barang-barang rumah tangga dan barang untuk kebutuhan sehari-hari. Kata-kata seperti “rak”, “sepatu”, “piring”, “meja” dan “kipas” mengarah pada adanya diskusi dan pertukaran informasi terkait barang tersebut. Kata “murah” membuktikan adanya penawaran barang-barang dengan harga yang terjangkau. Pada *cluster* ini membuktikan bahwa grup Whatsapp Jual Beli Area UKRIM menyediakan informasi bagi anggota grup untuk berbagi informasi, menawarkan dan mencari barang rumah tangga dengan beberapa kategori dan harga yang beragam. Tren utama *cluster* ini adalah barang rumah tangga dan kebutuhan sehari-hari, serta pembahasan tentang penawaran barang dengan harga terjangkau. Informasi ini dapat membantu para peserta grup terutama mahasiswa dalam mencari perabotan rumah tangga untuk kebutuhan di kost dengan harga terjangkau. Hasil penelitian dari *cluster* 9 dapat dilihat pada Gambar 14.



Gambar. 14 Hasil *cluster* 9 dengan kata paling sering muncul ditampilkan dalam grafik bar dan *wordcloud*

Hasil temuan dari *clustering* data menunjukkan bahwa terdapat beberapa kategori barang yang sering muncul dalam 10 *cluster* yang terbentuk antar lain: “meja”, “kipas”, “lemari”, “kasur”, “laptop”, “galon”, “gas”, “kompor”, “tabung”, “tungku”, “kaleng”, “motor”, “kost”, “hp”, “kucing”, “rak”, “sepatu”, “piring”, “meja”, “kipas”, “buku”, “nokia”, “android”, “iphone”, dan “kamar”. Informasi ini memiliki relevansi yang signifikan dengan tujuan dan konteks penelitian, untuk memberikan informasi kepada anggota grup agar mempermudah pengambilan keputusan dalam strategi promosi. Dengan memahami kategori barang yang sering muncul, anggota grup dapat mengidentifikasi tren penjualan yang dominan dan menyesuaikan strategi promosi.

Pengujian terkait validitas hasil K-Means menggunakan *silhouette coefficient* menunjukkan klasterisasi data dengan menggunakan  $K=10$  menghasilkan nilai *silhouette coefficient* sebesar 0.05. Nilai ini mengindikasikan bahwa terdapat pemisahan yang relatif baik antara *cluster-cluster* yang dihasilkan [20]. Meskipun nilai *silhouette coefficient* tidak mendekati nilai maksimum 1, yang akan menunjukkan pemisahan yang sangat baik, nilai 0.05 masih menunjukkan bahwa klasterisasi memiliki tingkat kehomogenan yang layak dan cukup memisahkan antar *cluster*. Hal ini menandakan bahwa penggunaan metode K-Means dengan jumlah *cluster* 10 memberikan hasil yang dapat diterima dalam pemisahan dan pengelompokan data.

#### IV. KESIMPULAN

Pada penelitian ini, *Text Clustering* dengan Metode K-Means berhasil menganalisis dan mengelompokkan data *chat* “Grup Whatsapp Jual Beli Area UKRIM”. Data *chat* yang didapatkan adalah 26.343 data dan setelah dilakukan pra-pemrosesan teks jumlahnya menjadi 4.732 data. *Clustering* dengan Metode K-Means pada *chat* “Grup Whatsapp Jual Beli Area UKRIM” menghasilkan sejumlah 10 *cluster* yang berbeda. Analisis terhadap setiap *cluster* memiliki masing-masing kata dominan. Kata yang memiliki jumlah kemunculan paling banyak adalah “info” dengan total 910 dan kata yang paling sedikit adalah kata “iphone” dan “hewan” dengan total masing-masing 2. Barang yang paling banyak muncul dalam *chat* adalah “kost” dengan jumlah 239 dan “gas” dengan jumlah 152. Hasil penelitian ini memberikan wawasan tentang dinamika komunikasi dalam komunitas jual beli online.

#### DAFTAR PUSTAKA

- [1] M. Nishom, “Perbandingan Akurasi Euclidean Distance, Minkowski Distance, dan Manhattan Distance pada Algoritma K-Means Clustering berbasis Chi-Square,” *J. Inform. J. Pengemb. IT*, vol. 4, no. 1, pp. 20–24, 2019, doi: 10.30591/jpit.v4i1.1253.
- [2] E. Fernando Ade Pratama, K. Khairil, and J. Jumadi, “Implementasi Metode K-Means Clustering Pada Segmentasi Citra Digital,” *J. Media Infotama*, vol. 18, no. 2, pp. 291–301, 2022.
- [3] A. Sulistiyawati and E. Supriyanto, “Implementasi Algoritma K-means Clustering dalam Penentuan Siswa Kelas Unggulan,” *J. Tekno Kompak*, vol. 15, no. 2, p. 25, 2021, doi: 10.33365/jtk.v15i2.1162.
- [4] A. Supriyadi, A. Triayudi, and I. D. Sholihati, “Perbandingan Algoritma K-Means Dengan K-Medoids Pada Pengelompokan Armada Kendaraan Truk Berdasarkan Produktivitas,” *JIPI (Jurnal Ilm. Penelit. dan Pembelajaran Inform.)*, vol. 6, no. 2, pp. 229–240, 2021, doi: 10.29100/jipi.v6i2.2008.
- [5] N. Hendrastuty, “Penerapan Data Mining Menggunakan Algoritma K-Means Clustering Dalam Evaluasi Hasil Pembelajaran Siswa,” *J. Ilm. Inform. Dan Ilmu Komput.*, vol. 3, no. 1, pp. 46–56, 2024, [Online]. Available: <https://doi.org/10.58602/jima-ilkom.v3i1.26>
- [6] A. R. Lashiyanti, I. Rasyid Munthe, F. A. Nasution, and E. P. Korespondensi, “Optimisasi Klasterisasi Nilai Ujian Nasional dengan Pendekatan Algoritma K-Means, Elbow, dan Silhouette,” *J. Ilmu Komput. dan Sist. Inf. (JIKOMSI)*, vol. 6, no. 1, pp. 14–20, 2023.
- [7] J. Ilmu and K. Ruru, “OPTIMIZING K-MEANS ALGORITHM WITH ELBOW AND SILHOUETTE,” vol. 1, pp. 1–6, 2024.
- [8] S. N. Br Sembiring, H. Winata, and S. Kusnarsi, “Pengelompokan Prestasi Siswa Menggunakan Algoritma K-Means,” *J. Sist. Inf. Triguna Dharma (JURSI TGD)*, vol. 1, no. 1, p. 31, 2022, doi: 10.53513/jursi.v1i1.4784.
- [9] M. W. Indriyanto, D. Hariyadi, and M. Habibi, “Investigasi Dan Analisis Forensik Digital Pada Percakapan Grup Whatsapp Menggunakan Nist Sp 800-86 Dan Support Vector Machine Digital Forensics Investigation and Analysis on Whatsapp Group Chats Using Nist Sp 800-86 and Support Vector Machine,” *Cyber Secur. dan Forensik Digit.*, vol. 3, no. 2, pp. 34–38, 2020.
- [10] M. Cendana and S. D. H. Permana, “Pra-Pemrosesan Teks Pada Grup Whatsapp Untuk Pemodelan Topik,” *Jurnal Mantik Penusa*, vol. 3, no. 3, pp.

- 107–116, 2019.
- [11] D. Musfiroh, U. Khaira, P. E. P. Utomo, and T. Suratno, “Analisis Sentimen terhadap Perkuliahan Daring di Indonesia dari Twitter Dataset Menggunakan InSet Lexicon,” *MALCOM Indones. J. Mach. Learn. Comput. Sci.*, vol. 1, no. 1, pp. 24–33, 2021, doi: 10.57152/malcom.v1i1.20.
- [12] M. Syarifuddin, “Analisis Sentimen Opini Publik Terhadap Efek Psbb Pada Twitter Dengan Algoritma Decision Tree, Knn, Dan Naïve Bayes,” *INTI Nusa Mandiri*, vol. 15, no. 1, pp. 87–94, 2020, doi: 10.33480/inti.v15i1.1433.
- [13] S. W. Ritonga, . Y., M. Fikry, and E. P. Cynthia, “Klasifikasi Sentimen Masyarakat di Twitter terhadap Ganjar Pranowo dengan Metode Naïve Bayes Classifier,” *Buuld. Informatics. Technol. Sci.*, vol. 5, no. 1, 2023, doi: 10.47065/bits.v5i1.3535.
- [14] R. Rasenda, H. Lubis, and R. Ridwan, “Implementasi K-NN Dalam Analisa Sentimen Riba Pada Bunga Bank Berdasarkan Data Twitter,” *J. Media Inform. Budidarma*, vol. 4, no. 2, p. 369, 2020, doi: 10.30865/mib.v4i2.2051.
- [15] N. L. P. C. Savitri, R. A. Rahman, R. Venyutzky, and N. A. Rakhmawati, “Analisis Klasifikasi Sentimen Terhadap Sekolah Daring pada Twitter Menggunakan Supervised Machine Learning,” *J. Tek. Inform. dan Sist. Inf.*, vol. 7, no. 1, pp. 47–58, 2021, doi: 10.28932/jutisi.v7i1.3216.
- [16] R. Slamet, W. Gata, A. Novtariany, K. Hilyati, and F. A. Jariyah, “Analisis Sentimen Twitter Terhadap Penggunaan Artis Korea Selatan Sebagai Brand Ambassador Produk Kecantikan Lokal,” *INTECOMS J. Inf. Technol. Comput. Sci.*, vol. 5, no. 1, pp. 145–153, 2022, doi: 10.31539/intecomsv5i1.3933.
- [17] A. Aditya, I. Jovian, and B. N. Sari, “Implementasi K-Means Clustering Ujian Nasional Sekolah Menengah Pertama di Indonesia Tahun 2018 / 2019,” vol. 4, pp. 51–58, 2020, doi: 10.30865/mib.v4i1.1784.
- [18] D. Adhe, C. Rachman, R. Goejantoro, F. Deny, and T. Amijaya, “Implementasi Text Mining Pengelompokan Dokumen Skripsi Menggunakan Metode K-Means Clustering Implementation Of Text Mining For Grouping Thesis Documents Using K-Means Clustering,” *J. EKSPONENSIAL*, vol. 11, no. 2, pp. 167–174, 2020.
- [19] Miquel Yosafat and J. Jatmika, “Implementasi Text Clustering Terkait Pilpres 2024 Menggunakan Metode K-Means,” *J. Sains Dan Komput.*, vol. 8, no. 01, pp. 6–12, 2024, doi: 10.61179/jurnalinfact.v8i01.496.
- [20] A. Yudhistira and R. Andika, “Pengelompokan Data Nilai Siswa Menggunakan Metode K-Means Clustering,” *J. Artif. Intell. Technol. Inf.*, vol. 1, no. 1, pp. 20–28, 2023, doi: 10.58602/jaiti.v1i1.22.