

## Clustering Review Pengguna Aplikasi Zenius pada Layanan Google Play Store Menggunakan Metode DBSCAN dan HDBSCAN

Fitri Dwi Handayani, Isnaini Rosyida

Prodi Matematika, Universitas Negeri Semarang, Kampus Sekaran Gunungpati, Kota Semarang 50229 Indonesia

\*Corresponding author: Fitridwhaan@students.unnes.ac.id



P-ISSN

E-ISSN: 2986-4178

### Riwayat Artikel

Dikirim: 29 Maret 2023

Direvisi: 5 Mei 2023

Diterima: 18 Mei 2023

### ABSTRAK

Kemajuan teknologi telah melahirkan suatu inovasi yaitu terciptanya aplikasi *E-learning* atau *Edutech* yang saat ini banyak diminati untuk membantu siswa dalam belajar mandiri di rumah. Pada *Google play store* beberapa *rating* dan *review* aplikasi memudahkan pengguna untuk mencari aplikasi dengan *review* terbaik. Adanya analisis *rating* dan *review* memungkinkan developer untuk menilai bagian mana dari aplikasi yang perlu difokuskan untuk perbaikan. Analisis ulasan terhadap topik ulasan dapat dilakukan dengan berbagai cara, salah satunya dapat menggunakan metode *clustering*. Penelitian ini bertujuan untuk mengetahui *review* dari sebuah aplikasi yang paling sering ditulis oleh pengguna aplikasi. Untuk mendapatkan hasil tersebut, dilakukan *review clustering* terhadap pengguna aplikasi *e-learning* di *Google Play Store* dengan menggunakan metode DBSCAN dan HDBSCAN. Selanjutnya untuk menilai metode mana yang terbaik untuk *clustering* dilakukan dengan membandingkan nilai rata-rata SC (*Silhouette Coefficient*) dari kedua metode tersebut. Hasil dari penelitian ini adalah rata-rata nilai SC metode HDBSCAN lebih baik dibandingkan dengan metode DBSCAN. HDBSCAN memiliki nilai SC rata-rata 0,2941, sedangkan metode DBSCAN hanya memiliki nilai SC rata-rata 0,1310.

**Kata Kunci:** *Review, Clustering, DBSCAN, HDBSCAN, Silhouette Coefficient*

### ABSTRACT

*Advances in technology have given birth to an innovation, namely the creation of E-learning or Edutech applications which are currently in great demand to assist students in independent learning at home. On the Google Play Store several application ratings and reviews make it easier for users to find applications with the best reviews. The existence of an analysis of a rating and review allows developers to assess which parts of the application need to focus on improvement. Analysis of reviews on review topics can be done in various ways, one of which can use the clustering method. This study is head to find out the reviews of an application that application users most often write. To obtain these results, a clustering review of e-learning application users was carried out on the Google Play Store using the DBSCAN and HDBSCAN methods. Furthermore, to assess which method is the best for clustering, it is done by comparing the average SC value (silhouette coefficient) of the two ways. The results of this study are that the average SC value of the HDBSCAN method is better than the DBSCAN method. HDBSCAN has an average SC value of 0.2941, while the DBSCAN method only has an average SC value of 0.1310.*

**Keywords:** *Review, Clustering, DBSCAN, HDBSCAN, Silhouette Coefficient*

## 1. Pendahuluan

Menurut Siahaan (2020) [1] pemanfaatan teknologi dalam proyek bagi siswa dapat meningkatkan kreativitas mereka dalam mengembangkan ilmu mereka. Hal ini sejalan dengan pendapat Setyosari (2007) [2] yang menyatakan bahwa pembelajaran melalui jaringan memiliki potensi, termasuk pembelajaran yang bermakna, kemudahan akses, dan peningkatan hasil belajar. Kemajuan teknologi juga menghasilkan inovasi dalam bentuk aplikasi *e-learning* atau *Edutech* yang saat ini populer sebagai alat bantu siswa dalam belajar mandiri di rumah. Aplikasi ini menawarkan kepraktisan, kemudahan, dan akses yang lebih merata bagi seluruh siswa di Indonesia.

PT Zona Edukasi Nusantara (*Zenius Education*) adalah perusahaan *Edutech* yang didirikan pada tahun 2007. Aplikasi *Zenius* telah diunduh lebih dari 5.000.000 kali dan termasuk dalam 10 besar aplikasi pendidikan terpopuler versi *Google Play Store* pada tahun 2023 *Google Play Store* (2023) [3].

*Play Store* adalah layanan digital yang disediakan oleh *Google*. Aplikasi *Play Store* memiliki fitur *rating* dan ulasan pengguna. *Rating* digunakan sebagai alat persaingan antara penyedia aplikasi atau pengembang. Selain itu, *rating* dan ulasan di *Play Store* membantu pengguna dalam mencari aplikasi yang terbaik karena ada banyak pilihan aplikasi di *Play Store*. *Rating* dan ulasan membantu pengguna dalam menemukan aplikasi dengan ulasan terbaik.

*Review* adalah kegiatan di mana seseorang memberikan pendapat tentang sesuatu yang telah mereka coba. *Review* memiliki manfaat yang besar dalam berbagai bidang, termasuk pemasaran, pendidikan, dan seni. Menurut Agustina, *et al.* (2019) [4] informasi yang terdapat dalam ulasan produk atau penilaian pengguna dapat berdampak *positif* atau *negatif* bagi perusahaan. Sejalan dengan Kusumasondjaja, *et al.* (2012) [5] yang menyatakan bahwa ulasan produk atau penilaian pengguna di platform digital berfungsi sebagai informasi bagi pengguna lain dan juga sebagai rekomendasi. Menurut Sari, *et al.* (2019) [6] *user review* sangat penting sebagai sumber daya berharga untuk meningkatkan kualitas aplikasi. Analisis ulasan dapat dilakukan dengan berbagai metode, salah satunya adalah metode *clustering*

*Clustering* adalah proses mempartisi kumpulan objek data menjadi subset yang disebut *cluster*. Menurut Rai (2010) [7] *clustering* merupakan metode untuk mengelompokkan data ke dalam kelompok-kelompok serupa, yang disebut *cluster*. Setiap *cluster* terdiri dari objek-objek yang mirip satu sama lain dan berbeda dari objek-objek dalam kelompok lain. Algoritma *clustering* dapat diklasifikasikan berdasarkan model *cluster* yang digunakan, seperti *Clustering Berbasis Centroid*, *Clustering Berbasis Kepadatan*, *Clustering Berbasis Distribusi*, *Clustering Berbasis Hirarki*, atau *Clustering Berbasis Konektivitas*.

Pemilihan jarak merupakan salah satu hal yang harus dilakukan saat melakukan *clustering*. Tujuan dari *clustering* sendiri adalah untuk mengelompokkan objek-objek yang memiliki kemiripan ke dalam satu kelompok. Salah satu cara menentukan jarak pada *clustering* adalah menggunakan jarak *euclidean*.

$$\text{Jarak} = \sqrt{(x - x_p)^2 + (y - y_p)^2} \quad (1)$$

$x$ : koordinat sumbu  $x$  titik tujuan,  $y$ : koordinat sumbu  $y$  titik tujuan,  $x_p$ : koordinat pusat sumbu  $x$ ,  $y_p$ : koordinat pusat sumbu  $y$

*Text Mining* adalah proses penggalian informasi dari kumpulan dokumen dari waktu ke waktu menggunakan berbagai alat analisis untuk mengidentifikasi dan mengungkap pola data yang ada. Teknik ini melibatkan eksplorasi dan analisis data teks tidak terstruktur dengan bantuan perangkat lunak yang dapat mengidentifikasi konsep,

praktik, topik, dan atribut lainnya. Menurut Feinerer, *et al.* (2008) [8] *Text Mining* merupakan bidang ilmu interdisipliner yang melibatkan data *mining*, *linguistik*, statistik komputasi, dan ilmu komputer. Teknik yang umum digunakan dalam *Text Mining* yaitu klasifikasi teks, pengelompokan teks, pembuatan ontologi dan taksonomi, peringkasan dokumen, dan analisis korpus laten. Dalam penelitian ini, tahapan *Text Mining* terdiri dari *Text Preprocessing* (*preprocessing* teks) dan *Text Representation* (representasi teks). Tahap *Text Preprocessing* melibatkan pemilihan dan pengorganisasian data teks untuk struktur yang lebih teratur selanjutnya tahap *Text Representation* melibatkan transformasi data teks menjadi representasi yang lebih mudah untuk diproses secara matematis.

*Density-Based Spatial Clustering of Application with Noise* (DBSCAN) adalah algoritma yang dirancang khusus untuk menemukan *cluster* dan *noise* dalam database Ester, *et al.* (1996) [9]. Menurut Mumtaz & Duraiswamy (2010) [10] *Density-Based Spatial Clustering of Application with Noise* (DBSCAN) adalah algoritma di mana pengelompokan didasarkan pada kepadatan, pengelompokan dilakukan melalui pertumbuhan area dengan kepadatan tinggi dan dapat menemukan segala bentuk pengelompokan.

*Hierarchical Density-Based Spatial Clustering of Applications with Noise* (HDBSCAN) adalah algoritma analisis kluster yang banyak digunakan karena ketahanannya terhadap *noise* dalam kumpulan data Ghamarian & Marquis, (2019) [11]. Menurut Wahyuni *et al.* (2021) [12] *Hierarchical DBSCAN* atau HDBSCAN adalah sebuah konsep dan perbaikan algoritmik pada *Ordering points* untuk mengidentifikasi algoritma *clustering structure* (OPTICS) dengan parameter input tunggal *Minpts* (*Minimum Points*).

Penelitian Erfina, *et al.* (2020) [13] tentang analisis *review* pengguna aplikasi *E-learning* sebagai variabel tambahan untuk menentukan aplikasi *E-learning* terbaik di *Play Store*, hasilnya menunjukkan bahwa beberapa aplikasi memiliki peringkat terbaik dari versi *review* pengguna, dengan analisis sentimen aplikasi *E-Learning* di *Google Play Store* membuat predikat terbaik bagi aplikasi tersebut menjadi relevan.

Penelitian Damayanti (2021) [14] tentang pengelompokan *review* pengguna untuk aplikasi Zenius menggunakan algoritma *K-Means* dan SOM (*Self-organizing Map*) menunjukkan bahwa analisis *review* sangat vital dalam sebuah aplikasi. Analisis *review* memberikan gambaran masukan untuk perbaikan aplikasi yang perlu dilakukan. Analisis ulasan terhadap topik ulasan dapat dilakukan dengan berbagai cara, salah satunya dapat menggunakan metode *clustering*.

Penelitian Mustakim, *et al.* (2021) [15] tentang *Clustering of Public Opinion on Natural Disasters in Indonesia Using DBSCAN and K-Medoids Algorithms* memperoleh hasil Algoritma DBSCAN memiliki validitas *cluster Silhouette Index* (SI) tertinggi dengan rata-rata 0,9140 dan waktu eksekusi rata-rata 83,40 detik sedangkan algoritma *K-Medoids* hanya memiliki nilai SI sebesar 0,2258 dengan rata-rata waktu eksekusi 849,93 detik.

Penelitian Wibowo, *et al.* (2021) [16] tentang *Revealing Tourist Hotspots In Yogyakarta City Based On Social Media Data Clustering* memperoleh hasil bahwa Algoritma DBSCAN dan HDBSCAN untuk kedua sumber data menghasilkan jumlah *cluster* yang sama, dengan distribusi spasial yang serupa. penentuan *cluster* memakai prosedur pemecahan HDBSCAN mempunyai output yang sangat fleksibel & bisa diterapkan pada penelitian dari data titik 2 dimensi buat membentuk *cluster* yang baik.

Pada penelitian ini akan dilakukan *clustering review* pengguna aplikasi Zenius pada layanan *Google Play Store* menggunakan metode DBSCAN (*Density-Based Spatial Clustering of Application with Noise*) dan metode HDBSCAN (*Hierarchical Density-Based Spatial Clustering of Applications with Noise*) dan membandingkan kedua metode tersebut menggunakan *silhouette coefficient*. Adapun tujuan dari penelitian ini adalah untuk mengetahui *review* pengguna aplikasi Zenius pada layanan *Google Play Store*, memperoleh hasil *clustering review* menggunakan metode DBSCAN dan HDBSCAN, dan memperoleh

hasil perbandingan *clustering review* pengguna aplikasi Zenius pada layanan *Google Play Store* menggunakan metode DBSCAN dan HDBSCAN.

## 2. Metodologi Penelitian

Penelitian ini dilakukan di PT Zona Edukasi Nusantara (*Zenius Education*), perusahaan *Edutech* di Indonesia, dengan menggunakan data yang diperoleh dari aplikasi *Google Play Store*. Penelitian ini mengadopsi pendekatan kuantitatif dan menggunakan rumus statistik untuk menganalisis data dan fakta yang diperoleh. Menurut Hermawan (2019) [17] penelitian kuantitatif merupakan metode penelitian yang induktif, objektif, dan ilmiah, dengan data berupa angka atau pernyataan yang dinilai, serta dianalisis menggunakan analisis statistik. Pengumpulan data dalam penelitian ini menggunakan data sekunder, yaitu data yang diperoleh dari sumber yang sudah ada. Data yang digunakan adalah ulasan dan rating pengguna aplikasi Zenius pada layanan *Google Play Store*. Jumlah data *review* pengguna yang digunakan dalam penelitian ini adalah 25.008 data, yang mencakup periode tahun 2019 hingga 2022. Proses pengambilan data dilakukan melalui teknik *web scraping* menggunakan *Python* di *Google Colab*. Teknik analisis data dalam penelitian ini ditunjukkan sebagai berikut.

1. Mengambil data atau *scraping review* (*Rating*, *Ulasan*) pada aplikasi Zenius dari *Google Play Store* menggunakan *Python Google Collab*.
2. Melakukan *Text Preprocessing* menggunakan *Python Google Collab*.
  - a. *Case Folding*, mengkonversi keseluruhan teks dalam dokumen menjadi huruf kecil.
  - b. *Data Cleaning*, pada tahap ini *review* pengguna aplikasi Zenius akan dibersihkan dari hal-hal yang tidak diperlukan seperti karakter angka, tanda baca, *emoticon*, dan spasi berlebih.
  - c. *Tokenizing*, metode pemisahan kata dalam kalimat untuk tujuan analisis teks lebih lanjut.
  - d. *Stopwords Removal*, pembuangan kata-kata yang sering muncul namun tidak relevan dalam suatu dokumen.
  - e. *Slang Word Replacement*, *review* pengguna aplikasi Zenius akan diubah menjadi kata baku berdasarkan kamus *slang word* yang telah dikumpulkan.
3. Melakukan *Text Representation* menggunakan *R* pada *Rstudio*.
  - a. *N-Gram*, mengurutkan elemen yang berdekatan dari urutan teks atau kata-kata tertentu.
  - b. *TF-IDF (Term Frequency - Inverse Document Frequency)*, memberikan bobot hubungan suatu kata atau *term* terhadap suatu dokumen.
4. Melakukan eksplorasi data untuk mengetahui data *review* pengguna aplikasi Zenius.
5. Melakukan analisis *clustering* serta menentukan jumlah *cluster* terbaik dengan kriteria nilai *Silhouette* pada setiap *rating* menggunakan *R* pada *Rstudio*.
  - a. Analisis *clustering* menggunakan metode *Density Based Spatial Clustering Of Applications With Noise* (DBSCAN).
  - b. Analisis *clustering* menggunakan metode *Hierarchical Density-Based Spatial Clustering of Applications with Noise* (HDBSCAN).
6. Membandingkan hasil *clustering* DBSCAN dan HDBSCAN berdasarkan nilai *Silhouette*.
  - a. Koefisien *Silhouette* per objek ditampilkan dalam rumus.

$$SC_1(p) = \frac{b(p)-a(p)}{\max\{a(p),b(p)\}}, p = 1,2, \dots \quad (2)$$

- b. Nilai  $SC$  suatu cluster ( $SC_2(i)$ ) diperoleh dengan menghitung nilai rata-rata ( $SC_1(p)$ ) dari seluruh data yang bergabung dengan cluster tersebut menggunakan persamaan berikut.

$$SC_2(i) = \frac{1}{n_i} \sum_{x_p \in C_i} SC_1(p) \quad (3)$$

- c. Setelah itu nilai  $SC$  global diperoleh dengan menghitung nilai rata-rata ( $SC_2(i)$ ) dari semua cluster dengan menggunakan kesepakatan sebagai berikut.

$$SC = \frac{\sum_{i=0}^c n_i SC_2(i)}{\sum_{i=0}^c n_i} \quad (4)$$

$a_p$ : rata-rata jarak data ke- $p$  dengan semua data pada satu cluster yang sama,  $b_p$ : rata-rata jarak data ke- $p$  dengan semua data pada cluster yang berbeda,  $SC_1(p)$ : Nilai *Silhouette Coefficient* pada data ke- $p$ ,  $SC_2(i)$ : nilai *Silhouette Coefficient* pada cluster ke- $i$ ,  $SC$ : nilai *Silhouette Coefficient* global,  $x_p$ : data pengamatan ke- $p$ ,  $C_i$ : Cluster ke- $i$ ,  $n_i$ : jumlah data pada cluster ke- $i$ ,  $c$ : jumlah cluster.

Terdapat beberapa kriteria subjektif untuk mengukur pengelompokan berdasarkan nilai *Silhouette Coefficient*. Interpretasi nilai *Silhouette Coefficient* ditunjukkan pada Tabel 1 [18].

**Tabel 1.** Interpretasi Nilai *Silhouette Coefficient*

| Silhouette Coefficient | Interpretasi                   |
|------------------------|--------------------------------|
| 0.71 – 1.00            | Struktur yang dihasilkan kuat  |
| 0.51 – 0.70            | Struktur yang dihasilkan baik  |
| 0.26 – 0.50            | Struktur yang dihasilkan lemah |
| $\leq 0.25$            | Tidak Terstruktur              |

7. Memberikan beberapa rekomendasi berdasarkan informasi yang didapatkan dari hasil *clustering*.
8. Membuat kesimpulan dan saran.

### 3. Hasil dan Pembahasan

Hasil *scraping* data *review* pengguna aplikasi Zenius pada layanan *Google Play Store* menghasilkan 25.008 *review* pengguna dari tanggal 18 Maret 2019 hingga 3 Oktober 2022. Data yang diperoleh terdiri dari data *Rating*, *Ulasan*, dan *Tanggal*. Selanjutnya, data *review* tersebut akan dikelompokkan menjadi 5 kategori *rating* berdasarkan *rating* yang diberikan oleh pengguna, yaitu dari *rating* 1 hingga *rating* 5. Ilustrasi data *rating* dan *review* pengguna aplikasi Zenius dapat dilihat pada Gambar 1.



**Gambar 1.** Total *Review* Pengguna Aplikasi Zenius

Sebelum menganalisis data ulasan pengguna aplikasi Zenius di *Play Store*, data akan dilakukan dalam beberapa tahap, yaitu *text preprocessing* dan *text representation* untuk mengubah data teks menjadi matriks angka yang digunakan selama proses analisis data. Pada tahap *Text Preprocessing* menghasilkan sejumlah data baru dari data sebelumnya, karena pada beberapa proses *Text Preprocessing*, kata-kata yang tidak diperlukan dalam

review akan dihilangkan, berikut adalah data baru untuk review pengguna setelah dilakukan *Text Preprocessing* dapat dilihat pada Tabel 2.

**Tabel 2.** Data Sebelum dan Sesudah Preprocessing

| Rating | Sebelum | Sesudah | Selisih |
|--------|---------|---------|---------|
| 5      | 18782   | 14236   | 4546    |
| 4      | 2137    | 1790    | 347     |
| 3      | 1300    | 1200    | 100     |
| 2      | 753     | 725     | 28      |
| 1      | 2036    | 1936    | 100     |

Setelah proses *Text Preprocessing*, dilanjutkan dengan tahap *Text Representation*, yaitu proses mengubah data teks menjadi representasi yang lebih mudah diproses, dengan tujuan merepresentasikan dokumen teks yang tidak terstruktur secara numerik agar dapat dihitung secara matematis. Data diubah menjadi bentuk *bigram* (padanan dua kata) yang dilanjutkan proses *Document Term Matrix* dengan menggunakan pembobotan TF-IDF (*Term Frequency - Inverse Document Frequency*) menggunakan program R di *Rstudio*. Dalam proses ini data yang digunakan adalah review data dalam bentuk bigram atau padanan dua kata dapat dilihat pada Tabel 3.

**Tabel 3.** Contoh TF-IDF

| kata                | $x_1$    | ... | $x_{25008}$ |
|---------------------|----------|-----|-------------|
| <b>Video_gratis</b> | 596,5... | :   | 0           |
| <b>Hemat_kuota</b>  | 0        | :   | 0           |
| <b>Zenius_keren</b> | 0        | :   | 5,738,...   |

Ulasan pengguna aplikasi Zenius pada berbagai *rating* menunjukkan karakteristik yang berbeda. Ulasan dengan *rating* tinggi cenderung bernilai *positif*. Analisis *Word Cloud Bigram* dari seluruh data review menunjukkan dominasi kata-kata *positif* seperti "terima kasih", "mudah dimengerti", "bantu belajar", dan "terima kasih zenius". Tidak ditemukan ulasan dengan kata-kata *negatif* terkait aplikasi ini, menunjukkan lebih banyak ulasan *positif* daripada *negatif*.



**Gambar 2.** Word Cloud Seluruh Data Review

Visualisasi dan eksplorasi isi review pada Gambar 2 hanya memberikan informasi umum tentang kata-kata yang sering muncul dalam review. Untuk mendapatkan informasi lebih lanjut, analisis *clustering* digunakan dalam penelitian ini. Metode DBSCAN dan HDBSCAN digunakan dalam studi ini untuk mengelompokkan data ulasan berdasarkan *rating*. Hasil *clustering* dari kedua metode tersebut dievaluasi berdasarkan beberapa kriteria, dan pendekatan terbaik dan jumlah *cluster* dipilih untuk masing-masing *rating*. Pada penelitian ini jarak antar objek dihitung menggunakan jarak *Euclidean*. Hasil perhitungan jarak antar objek dapat dilihat pada Tabel 4.

**Tabel 4.** Jarak Euclidean

| Jarak Euclidean |        |
|-----------------|--------|
| (i,j)           | d(i,j) |
| (1,1)           | 0.0    |

|             |     |
|-------------|-----|
| (1,2)       | 11  |
| ⋮           | ⋮   |
| (2035,2035) | 0.0 |

### 3.1 Clustering DBSCAN

Diperoleh berdasarkan hasil *clustering* pada Tabel 5 dengan menggunakan metode DBSCAN pada data *rating 1*, *rating 2*, *rating 3*, *rating 4*, dan *rating 5* adalah sebagai berikut.

**Tabel 5.** Clustering DBSCAN

| Rating | Eps | MinPts | Cluster | Noise | SC     |
|--------|-----|--------|---------|-------|--------|
| 1      | 17  | 2      | 1       | 18    | 0.1047 |
| 2      | 12  | 2      | 1       | 10    | 0.0994 |
| 3      | 13  | 2      | 1       | 41    | 0.0901 |
| 4      | 15  | 2      | 1       | 67    | 0.0853 |
| 5      | 32  | 2      | 1       | 379   | 0,2753 |

#### a. Rating 1



**Gambar 3.** Word Cloud Clustering DBSCAN Rating 1 Cluster 1



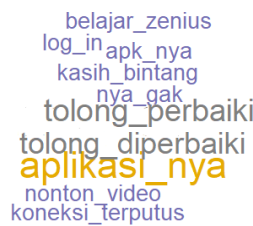
**Gambar 4.** Word Cloud Clustering DBSCAN Rating 1 Noise Points

*Clustering* pada *rating 1* pada Gambar 3 dan Gambar 4 memberikan informasi bahwa banyak keluhan pengguna terkait masalah aplikasi Zenius. Keluhan mencakup kesulitan masuk, menjalankan video, membuka aplikasi, mengunduh materi, dan registrasi. pengguna juga mengeluhkan adanya biaya tambahan untuk membuka beberapa video pembelajaran dan terlalu banyak promosi yang dilakukan oleh Zenius.

#### b. Rating 2



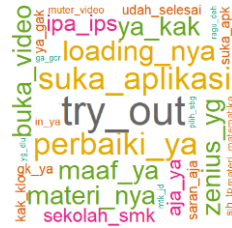
**Gambar 5.** Word Cloud Clustering DBSCAN Rating 2 Cluster 1



**Gambar 6.** Word Cloud Clustering DBSCAN Rating 2 Noise Points

*Clustering* pada *rating 2* pada Gambar 5 dan Gambar 6 memberikan informasi bahwa banyak keluhan dari pengguna terkait masalah saat menggunakan aplikasi seperti kesulitan penggunaan aplikasi, *bug*, kinerja lambat, kesulitan membuka video pembelajaran, dan paket berlangganan dianggap tidak sesuai.

c. Rating 3



Gambar 7. Word Cloud Clustering  
DBSCAN Rating 3 Cluster 1



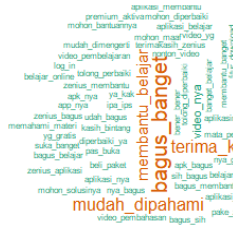
Gambar 8. Word Cloud Clustering  
DBSCAN Rating 3 Noise Points

Clustering pada rating 3 pada Gambar 7 dan Gambar 8 memberikan informasi bahwa banyak keluhan dari pengguna terkait masalah saat menggunakan aplikasi seperti fitur *try out* yang tidak berfungsi, tampilan pertanyaan yang tidak nyaman, *bug*, kinerja lambat, dan kesulitan mengakses materi.

d. Rating 4



Gambar 9. Word Cloud Clustering  
DBSCAN Rating 4 Cluster 1



Gambar 10. Word Cloud Clustering  
DBSCAN Rating 4 Noise Points

Clustering pada rating 4 pada Gambar 9 dan Gambar 10 memberikan informasi bahwa banyak keluhan dari pengguna terkait masalah saat menggunakan aplikasi, seperti kesulitan membuka aplikasi dan video pembelajaran, *error* setelah *update*, kesulitan membuka soal latihan, ketidaksesuaian fasilitas pengguna premium, kurikulum yang tidak sesuai dengan siswa SMK. Selain keluhan, pengguna juga merasa puas dengan aplikasi Zenius karena dianggap membantu dalam belajar dengan penjelasan yang mudah dipahami.

e. Rating 5



Gambar 11. Word Cloud Clustering  
DBSCAN Rating 5 Cluster 1



Gambar 12. Word Cloud Clustering  
DBSCAN Rating 5 Noise Points

Clustering pada rating 5 pada Gambar 11 dan Gambar 12 memberikan informasi bahwa sebagian besar pengguna merasa terbantu dalam proses pembelajaran dengan menggunakan aplikasi Zenius. Pembelajaran mudah dipahami, fitur pengajaran lengkap, tampilan aplikasi bagus, dan pembahasan soal yang baik.



### 3.2 Clustering HDBSCAN

Diperoleh berdasarkan hasil *clustering* pada Tabel 6 dengan menggunakan metode HDBSCAN adalah sebagai berikut.

**Tabel 6. Clustering HDBSCAN**

| Rating | MinPts | Cluster | Noise | SC     |
|--------|--------|---------|-------|--------|
| 1      | 6      | 2       | 950   | 0,2305 |
| 2      | 3      | 3       | 54    | 0,3846 |
| 3      | 3      | 3       | 299   | 0,3057 |
| 4      | 5      | 11      | 1211  | 0,3490 |
| 5      | 2      | 2       | 4     | 0,2011 |

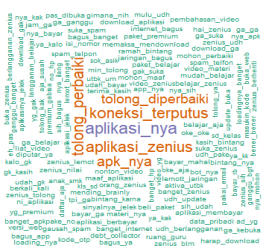
a. Rating 1



Gambar 13. Word Cloud Clustering HDBSCAN Rating 1 Cluster 1



Gambar 14. Word Cloud Clustering HDBSCAN Rating 1 Cluster 2



Gambar 15. Word Cloud Clustering HDBSCAN Rating 1 Noise Points

*Clustering* pada *rating* 1 pada Gambar 13, Gambar 14, dan Gambar 15 memberikan informasi bahwa banyak keluhan dari pengguna terkait masalah saat menggunakan aplikasi yaitu pengguna mengalami kesulitan dalam membuka aplikasi, masalah login, penggunaan kuota internet yang dianggap boros, permintaan *update* yang terlalu sering, kesulitan mengakses video pembelajaran, kesalahan aplikasi, koneksi terputus, adanya *bug* dalam aplikasi, terbatasnya materi pembelajaran untuk pengguna *non*-langganan, promosi yang berlebihan, ketidaksesuaian pertanyaan dengan materi, dan kualitas video yang kurang baik.

b. Rating 2



Gambar 16. Word Cloud Clustering HDBSCAN Rating 2 Cluster 1



Gambar 17. Word Cloud Clustering HDBSCAN Rating 2 Cluster 2



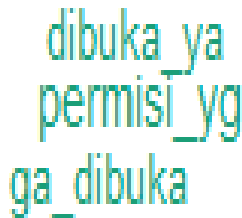
**Gambar 18.** Word Cloud Clustering  
HDBSCAN Rating 2 Cluster 3



**Gambar 19.** Word Cloud Clustering  
HDBSCAN Rating 2 Noise Points

Clustering pada rating 2 pada Gambar 16, Gambar 17, Gambar 18, dan Gambar 19 memberikan informasi bahwa banyak keluhan dari pengguna terkait masalah saat menggunakan aplikasi yaitu Pengguna mengeluhkan *loading* aplikasi yang lambat, sulit membuka beberapa video, kesulitan saat membuka aplikasi dan *login*, penurunan kinerja setelah *update*, adanya *bug*, dan *error* dalam penggunaan aplikasi, fitur langganan yang kurang baik, akses terbatas terhadap materi, dan soal-soal latihan pendahuluan pada beberapa materi pelajaran.

c. Rating 3



**Gambar 20.** Word Cloud Clustering  
HDBSCAN Rating 3 Cluster 1



**Gambar 21.** Word Cloud Clustering  
HDBSCAN Rating 3 Cluster 2



**Gambar 22.** Word Cloud Clustering  
HDBSCAN Rating 3 Cluster 3



**Gambar 23.** Word Cloud Clustering  
HDBSCAN Rating 3 Noise Points

Clustering pada rating 3 pada Gambar 20, Gambar 21, Gambar 22, dan Gambar 23 memberikan informasi bahwa banyak keluhan dari pengguna seperti kesulitan saat membuka aplikasi, video pembelajaran sulit dibuka, aplikasi sering mengalami gangguan, masalah dalam pengunduhan, terlalu sering mengupdate, kualitas video dinilai kurang, soal latihan kurang lengkap, dan beberapa soal sulit dibuka. Selain masalah teknis, pengguna juga mengeluhkan keterbatasan materi untuk siswa SMK (Sekolah Menengah Kejuruan).

d. Rating 4



**Gambar 24.** Word Cloud Clustering  
HDBSCAN Rating 4 Cluster 1



**Gambar 25.** Word Cloud Clustering  
HDBSCAN Rating 4 Cluster 2



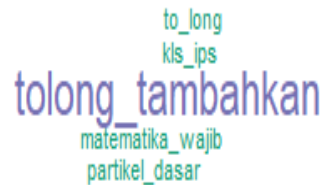
**Gambar 26.** Word Cloud Clustering  
HDBSCAN Rating 4 Cluster 3



**Gambar 27.** Word Cloud Clustering  
HDBSCAN Rating 4 Cluster 4



**Gambar 28.** Word Cloud Clustering  
HDBSCAN Rating 4 Cluster 5



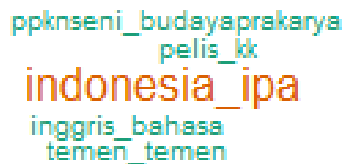
**Gambar 29.** Word Cloud Clustering  
HDBSCAN Rating 4 Cluster 6



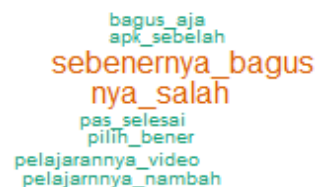
**Gambar 30.** Word Cloud Clustering  
HDBSCAN Rating 4 Cluster 7



**Gambar 31.** Word Cloud Clustering  
HDBSCAN Rating 4 Cluster 8



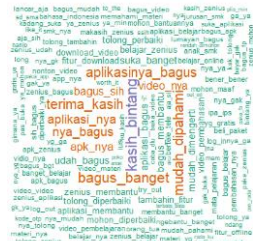
**Gambar 32.** Word Cloud Clustering  
HDBSCAN Rating 4 Cluster 9



**Gambar 33.** Word Cloud Clustering  
HDBSCAN Rating 4 Cluster 10



Gambar 34. Word Cloud Clustering  
HDBSCAN Rating 4 Cluster 11



Gambar 35. Word Cloud Clustering  
HDBSCAN Rating 5 Noise Points

Clustering pada rating 4 pada Gambar 24 sampai dengan Gambar 35 memberikan informasi bahwa banyak keluhan dari pengguna seperti kesulitan mengakses video pembelajaran, aplikasi lambat saat digunakan, masalah audio pada beberapa video, sulit membuka soal latihan, gangguan pada fitur *try out*, kesulitan membuka aplikasi Zenius, keterbatasan paket belajar bagi pengguna *non*-langganan, promosi yang dianggap berlebihan dan mengganggu, penggunaan kuota yang terlalu besar, dan pembahasan soal yang tidak akurat. *Rating* 4 juga menggambarkan kepuasan pengguna karena aplikasi Zenius dianggap membantu dalam proses belajar di masa pandemi, membantu pemahaman materi, dan memiliki video pembelajaran yang baik dan mudah dipahami

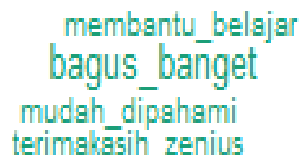
e. *Rating* 5



Gambar 36. Word Cloud Clustering  
HDBSCAN Rating 5 Cluster 1



Gambar 37. Word Cloud Clustering  
HDBSCAN Rating 5 Cluster 2



Gambar 38. Word Cloud Clustering  
HDBSCAN Rating 5 Noise Points

Clustering pada rating 5 pada Gambar 36, Gambar 37, dan Gambar 38 memberikan informasi bahwa pengguna memberikan ulasan *positif* terhadap aplikasi Zenius. Mereka mengatakan bahwa aplikasi ini membantu siswa dalam proses pembelajaran, memudahkan pemahaman materi, mudah digunakan, memiliki pesona yang menyenangkan, dan memberikan banyak manfaat. Pengguna juga merasa puas setelah menggunakan aplikasi Zenius. Selain itu, ada juga saran untuk menambahkan materi pembelajaran khususnya untuk jenjang SMK.

### 3.3 Perbandingan Nilai *Silhouette Coefficient*

Setelah diperoleh hasil *cluster* terbaik pada metode DBSCAN dan HDBSCAN, selanjutnya dilakukan perbandingan pada kedua metode tersebut berdasarkan kriteria nilai *Silhouette Coefficient*. Perbandingan hasil *clustering* berdasarkan nilai *Silhouette Coefficient* ditunjukkan pada Tabel 7 dibawah ini.

**Tabel 7.** Hasil Perbandingan *Silhouette Coefficient*

| Rating    | <i>Silhouette Coefficient</i> |               |
|-----------|-------------------------------|---------------|
|           | DBSCAN                        | HDBSCAN       |
| Rating 1  | 0,1047                        | 0,2305        |
| Rating 2  | 0,0994                        | 0,3846        |
| Rating 3  | 0,0901                        | 0,3057        |
| Rating 4  | 0,0853                        | 0,3490        |
| Rating 5  | 0,2753                        | 0,2011        |
| Rata-Rata | <b>0,1310</b>                 | <b>0,2941</b> |

Berdasarkan hasil Tabel 7 dapat disimpulkan bahwa kriteria nilai *Silhouette clustering* dengan metode HDBSCAN memberikan rata-rata hasil yang lebih baik dibandingkan *clustering* menggunakan metode DBSCAN pada keseluruhan *rating*.

## 4. Kesimpulan

Berdasarkan data *review* pengguna yang terdapat pada aplikasi Zenius di layanan *Google Play Store*, *review* dengan *rating* 5 memiliki proporsi tertinggi dibandingkan *review* dengan *rating* lainnya, artinya banyak pengguna yang merasa sangat puas dengan aplikasi Zenius. Semakin tinggi *ratingnya*, maka semakin banyak *review* pada *rating* tersebut yang mengandung hal-hal *positif* atau kelebihan dari aplikasi tersebut. Keluhan umum pengguna meliputi masalah akses, login, performa, dan video pembelajaran. Hasil evaluasi *Silhouette Coefficient* menunjukkan bahwa *clustering* dengan metode HDBSCAN lebih baik dibandingkan dengan metode DBSCAN. *Clustering* yang dilakukan dalam penelitian ini terbukti mampu menggali informasi lebih dalam dari keseluruhan *review* secara cepat dan efisien, sehingga berdasarkan hasil penelitian Zenius disarankan untuk menerapkan *text clustering* secara berkala untuk mendapatkan informasi lebih cepat dan efisien sebagai bahan evaluasi. tanpa harus membaca ulasan keseluruhan.

## 5. Daftar Pustaka

- [1] M. Siahaan, "Dampak Pandemi Covid-19 Terhadap Dunia Pendidikan," *J. Kaji. Ilm.*, vol. 1, no. 1, pp. 1–3, 2020.
- [2] P. Setyosari, "Pembelajaran Sistem Online: Tantangan dan Rangsangan. Majalah Ilmiah Pembelajaran," *Maj. Ilm. Pembelajaran*, vol. 3, no. 2, 2007.
- [3] G. P. Store, "Aplikasi Populer Pendidikan," *play.google.com*, 2023. [Online]. Available: <https://play.google.com/store/apps/category/EDUCATION?hl=id>. [Accessed: 05-May-2023].
- [4] L. Agustina, A. O. Fayardi, and Irwansyah, "Online Review : Indikator Penilaian Kredibilitas Online dalam Platform E-commerce," *J. Ilmu Komun.*, no. 4, pp. 141–154, 2019.
- [5] S. Kusumasondjaja, T. Shanka, and C. Marchegiani, "Credibility of online reviews and initial trust : The roles of reviewer ' s identity and review valence," *J. Vacat. Mark.*, vol. 18, no. 3, pp. 185–195, 2012.
- [6] A. E. Sari, S. Widowati, and K. M. Lhaksana, "Klasifikasi Ulasan Pengguna

- Aplikasi Mandiri Online di Google Play Store dengan Menggunakan Metode Information Gain dan Naive Bayes Classifier,” *e-Proceedings Eng.*, vol. 6, no. 2, p. 9143, 2019.
- [7] P. Rai, “A Survey of Clustering Techniques,” *Int. J. Comput. Appl.*, vol. 7, no. 12, pp. 1–5, 2010.
- [8] I. Feinerer, K. Hornik, and D. Meyer, “Text mining infrastructure in R,” *J. Stat. Softw.*, vol. 25, no. 5, pp. 1–54, 2008.
- [9] M. Ester, H.-P. Kriegel, J. Sander, and X. Xu, “A Density-Based Algorithm for Discovering Clusters A Density-Based Algorithm for Discovering Clusters in Large Spatial Databases with Noise,” *Knowl. Discov. Database*, vol. 96, no. 34, pp. 226–231, 1996.
- [10] K. Mumtaz and K. D. Duraiswamy, “An Analysis on Density Based Clustering of Multi Dimensional Spatial Data,” *Indian J. Comput. Sci. Eng.*, vol. 1, no. 1, pp. 8–12, 2010.
- [11] I. Ghamarian and E. . Marquis, “Hierarchical Density-Based Cluster Analysis Framework for Atom Probe Tomography Data,” *Elsevier*, vol. 200, p. 28038, 2019.
- [12] N. A. Wahyuni, M. N. Hayati, and A. N. Rizki, “Metode Hierarchical Density-Based Spatial Clustering of Application with Noise ( HDBSCAN ) Pada Wilayah Desa / Kelurahan Tertinggal di Kabupaten Kutai Kartanegara ( Studi Kasus : Data Hasil Pendataan Potensi Desa ( PODES ) Tahun 2018 ) Hierarchical Densit,” *J. EKSPONENSIAL*, vol. 12, no. 1, pp. 47–52, 2021.
- [13] A. Erfina, E. S. Basryah, A. Saepulrohman, and D. Lestari, “Analisis Sentimen Aplikasi Pembelajaran Online Di Play Store Pada Masa Pandemi Covid-19 Menggunakan Algoritma Support Vector Machine (SVM),” *Semin. Nas. Informatika*, vol. Vol. 1, no. No. 1, pp. 145–152, 2020.
- [14] N. A. Damayanti, “Pengelompokan Ulasan Aplikasi Belajar Online Zenius Menggunakan Metode Self Organizing Maps (SOM) dan K-Means,” (*Doctoral Diss. Inst. Teknol. Sepuluh Nopember*)., 2021.
- [15] Mustakim, M. Z. Fauzi, Mustafa, A. Abdullah, and Rohayati, “Clustering of Public Opinion on Natural Disasters in Indonesia Using DBSCAN and K-Medoids Algorithms,” *J. Phys. Conf. Ser.*, vol. 1783, no. 1, p. 012016, 2021.
- [16] T. W. Wibowo, S. H. M. B. Santosa, B. Susilo, and T. H. Purwanto, “Revealing tourist hotspots in yogyakarta city based on social media data clustering,” *Geoj. Tour. Geosites*, vol. 34, no. 1, pp. 218–225, 2021.
- [17] I. Hermawan, *Metodologi Penelitian Pendidikan (Kualitatif, Kuantitatif dan Mixed Method)*. 2019.
- [18] A. Struyf, M. Hubert, and P. J. Rousseeuw, “Clustering in an object-oriented environment,” *J. Stat. Softw.*, vol. 1, pp. 1–30, 1996.