

**ANALISIS METODE *CLUSTERING* DENGAN ALGORITMA  
*DENSITY – BASED SPATIAL CLUSTERING OF APPLICATION  
WITH NOISE (DBSCAN)* UNTUK PENGELOMPOKAN  
*CYBERBULLYING* DI INSTAGRAM**

**SKRIPSI**

Diajukan Untuk Memenuhi Sebagian Persyaratan Dalam  
Memperoleh Gelar Sarjana Teknik Informatika Universitas Medan Area

Diajukan Oleh :

**Budiman Sujatmiko**

**198160018**



**PROGRAMSTUDISARJANATEKNIKINFORMATIKA**

**FAKULTASTEKNIK**

**UNIVERSITASMEDANAREA**

**2023**

i

**UNIVERSITAS MEDAN AREA**

© Hak Cipta Di Lindungi Undang-Undang

1. Dilarang Mengutip sebagian atau seluruh dokumen ini tanpa mencantumkan sumber
2. Pengutipan hanya untuk keperluan pendidikan, penelitian dan penulisan karya ilmiah
3. Dilarang memperbanyak sebagian atau seluruh karya ini dalam bentuk apapun tanpa izin Universitas Medan Area

Document Accepted 17/1/24

Access From (repository.uma.ac.id)17/1/24

## LEMBAR PENGESAHAN

Judul Skripsi : Analisis Metode *Clustering* Dengan Algoritma *Density- Based Spatial Of Clustering Application With Noise (DBSCAN)* Pengelompokan *Cyberbullying* Di Instagram

Nama : Budiman Sujatniko

NPM : 198160018

Fakultas : Teknik

Prodi : Teknik Informatika

Disetujui Oleh

Pembimbing

Andre Hasudungan Lubis, S.Ti, Msc

Diketahui Oleh :

Dekan :

Kaprodi:



Dr. Eng. Supriatno, ST, MT  
NIDN: 0105058804



Rizki Muliono, S.Kom, M.Kom  
NIDN: 0109038902

Tanggal Lulus : 26 September 2023

## HALAMAN PERNYATAAN

Saya menyatakan bahwa skripsi yang saya susun, sebagai syarat memperoleh gelar sarjana merupakan hasil karya tulis saya sendiri. Adapun bagian – bagian tertentu dalam penulisan skripsi ini yang saya kutip dari hasil karya orang lain telah dituliskan sumbernya secara jelas sesuai dengan norma, kaidah, dan etika penulisan ilmiah.

Saya bersedia menerima sanksi pencabutan gelar akademik yang saya peroleh dan sanksi – sanksi lainnya dengan peraturan yang berlaku, apabila di kemudian hari ditemukan adanya plagiat dalam skripsi ini.



Medan, 26 September 2023

Budiman Sujatmiko

19816001



## HALAMAN PERNYATAAN PERSETUJUAN PUBLIKASI

### TUGAS AKHIR/SKRIPSI/TESIS UNTUK KEPENTINGAN AKADEMIS

Sebagai sivitas akademik Universitas Medan Area, Saya bertanda tangan dibawah ini:

Nama : Budiman Sujatmiko

NPM : 198160018

Program Studi : Teknik Informatika

Fakultas : Teknik

Jenis Karya : Skripsi

Demi pengembangan ilmu pengetahuan, menyetujui untuk memberikan kepada Universitas Medan Area **Hak Bebas Royalti Noneklusif (Non-exclusive Royalty-Free Right)** atas karya ilmiah saya yang berjudul : **Analisis Metode Clustering Dengan Algoritma Density- Based Spatial Of Clustering Application With Noise (DBSCAN) Pengelompokan Cyberbullying Di Instagram.**

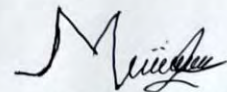
Beserta perangkat yang ada (jika diperlukan). Dengan Hak Bebas Royalti Noneklusif ini Universitas Medan Area berhak menyimpan, mengalih media/format-kan, mengelola dalam bentuk pangkalan data (*Database*), merawat, dan memublikasikan skripsi saya selama tetap mencantumkan nama saya sebagai penulis/pencipta dan sebagai pemilik Hak Cipta.

Demikian pernyataan ini saya buat dengan sebenarnya.

Dibuat di : Medan

Pada Tanggal : 12 Juli 2023

Yang Menyatakan



(Budiman Sujatmiko)

## ABSTRAK

*Cyberbullying* adalah bentuk dari sikap intimidasi yang dilakukan terhadap pengguna media sosial yang terjadi di dunia maya seperti Twitter, Facebook, Instagram. Sehingga, diperlukan adanya pengelompokan untuk mengatasi tindakan *cyberbullying* untuk membantu pihak – pihak berwenang seperti kominfo (Kementerian Komunikasi dan Informatika) untuk mengatasi dan meminimalisir terjadinya tindakan *cyberbullying* yang terjadi di media sosial khususnya Instagram dan memberikan informasi edukasi tentang *cyberbullying*.

Maka dari itu perlu adanya pengelompokan tindakan tersebut seperti penggunaan data mining untuk mengolah data yang besar dan membantu untuk mengambil keputusan. Pengelompokan pada data mining memiliki berbagai jenis algoritma, salah satunya adalah algoritma DBSCAN (*Density-based spatial clustering of applications with noise*). Algoritma DBSCAN adalah suatu metode pengelompokan berbasis kepadatan untuk melakukan clustering yang membangun area berdasarkan densitas yang terkoneksi (Setiawan dkk., 2021). Pada penelitian ini, memiliki jumlah data 649 komentar yang berisi kalimat *bullying* dan *nonbullying*. Peneliti menggunakan algoritma DBSCAN sebagai metode untuk clustering kalimat *cyberbullying* dan menggunakan *Davies Bouldin Index* untuk mengevaluasi hasil pengelompokan. Dari hasil perhitungan manual dilakukan dengan menentukan jarak Euclidean dan menghitung dengan algoritma DBSCAN dengan menentukan parameter epsilon adalah 0,4 dan nilai MinPts adalah 2 menghasilkan 3 kluster yaitu C1 {d1,d2,d3,d4,d5,d6}, C2 {d7,d8}, dan C3 {d9,d10}. Dan hasil nilai dari perhitungan *Davies Bouldin Index* untuk clustering ini. Semakin rendah nilai DBI, maka semakin baik clusteringnya. Dalam perhitungan ini mendapatkan hasil sekitar 2,1826.

**Kata kunci :** *Cyberbullying, Clustering, Data Mining, DBSCAN, DBI*

## **ABSTRACT**

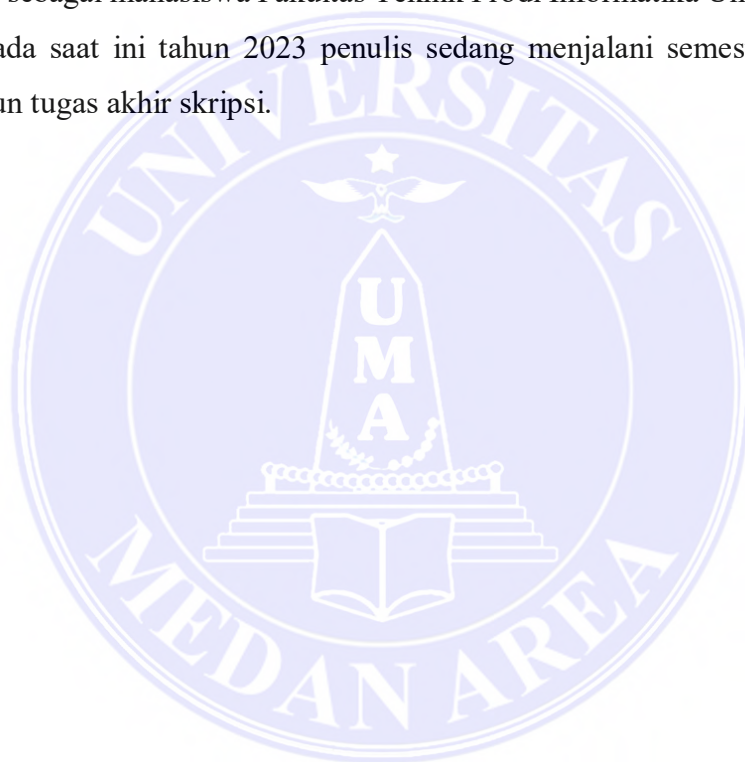
*Cyberbullying is a form of intimidation carried out against social media users that occurs in cyberspace such as Twitter, Facebook, Instagram. So, there is a need for a grouping to overcome cyberbullying acts to help authorities such as Kominfo (Ministry of Communication and Information Technology) to overcome and minimize the occurrence of cyberbullying acts that occur on social media, especially Instagram and provide educational information about cyberbullying. Therefore, it is necessary to group these actions, such as the use of data mining to process large data and help make decisions. Grouping in data mining has various types of algorithms, one of which is the DBSCAN (Density-based spatial clustering of applications with noise) algorithm. The DBSCAN algorithm is a density-based grouping method for clustering that builds areas based on connected density (Setiawan et al., 2021). In this research, the total data was 649 comments containing bullying and non-bullying sentences. Researchers used the DBSCAN algorithm as a method for clustering cyberbullying sentences and used the Davies Bouldin Index to evaluate the clustering results. From the results of manual calculations carried out by determining the Euclidean distance and calculating with the DBSCAN algorithm by determining the epsilon parameter is 0.4 and the MinPts value is 2 resulting in 3 clusters namely C1 {d1, d2, d3, d4, d5, d6}, C2 {d7, d8}, and C3{d9,d10}. And the value results from the Davies Bouldin Index calculation for this clustering. The lower the DBI value, the better the clustering. In this calculation the result is around 2.1826.*

*Keywords: Cyberbullying, Clustering, Data Mining, DBSCAN, DBI*

## Daftar Riwayat Hidup

Penulis lahir di Pir Trans Sosa 1A pada tanggal 09 Januari 2001 dari Bapak Misman dan Ibu Eka Susilawati. Penulis pertama kali mengenyam pendidikan dibangku SD Negeri 1713 Pir Trans Sosa 1A pada tahun 2007-2013, meneruskan pendidikan di Sekolah Menengah Pertama Swasta Makarti diselesaikan pada tahun 2013-2016, Meneruskan Pendidikan Sekolah Menengah Kejuruan Negeri 1 pada tahun 2016 – 2019.

Pada tahun 2019 penulis lulus dari SMKN 1 Huta Raja Tinggi dan pada 2019 terdaftar sebagai mahasiswa Fakultas Teknik Prodi Informatika Universitas Medan Area. Pada saat ini tahun 2023 penulis sedang menjalani semester 8 (Delapan) menyusun tugas akhir skripsi.





## KATA PENGANTAR

Puji syukur peneliti ucapkan kepada Tuhan Yang Maha Esa oleh karena kasih karunia, berkat kemurahan dan pertolongan-Nya sehingga peneliti dapat menyelesaikan skripsi yang berjudul “ ANALISIS METODE CLUSTERING DENGAN ALGORITMA DBSCAN ( DENSITY – BASED SPATIAL CLUSTERING OF APPLICATION WITH NOISE)”. Adapun penulisan skripsi ini sebagai syarat untuk menyelesaikan Pendidikan Strata 1 (S1) Program Studi Teknik Informatika Universitas Medan Area.

Peneliti menyadari bahwa keberhasilan dalam menyelesaikan skripsi ini tidak lepas dari bimbingan, arahan, bantuan serta dukungan yang baik dari berbagai pihak. Oleh karena itu, pada kesempatan ini dengan kerendahan hati peneliti mengucapkan terima kasih sebanyak-banyaknya kepada pihak yang terlibat dalam pengerjaan skripsi ini kepada:

1. Allah Swt. atas segala rahmat dan kasih karunia-nya yang telah memberikan kekuatan bagi penulis dalam menyelesaikan skripsi ini.
2. Yang teristimewa dan yang terkasih kedua orang tua saya, Bapak Misman dan Ibu Eka Susilawati yang selalu memberikan semangat dan dukungan yang sangat luar biasa, nasehat, motivasi dan kasih sayang yang tak terhingga. Terimakasih atas setiap doa yang telah kalian panjatkan kepada Tuhan Yang Maha Esa.
3. Yayasan H.Agus Salim Universitas Medan Area.
4. Bapak Prof. Dr. Dadan Ramdan, M. Eng, M. Sc selaku Rektor Universitas Medan Area.
5. Bapak Dr. Rahmad Syah, S.Kom, M.Kom selaku Dekan Fakultas Teknik Universitas Medan Area.
6. Bapak Rizki Muliono, S.Kom, M.Kom selaku Kepala Bagian Prodi Teknik Informatika Universitas Medan Area sekaligus sebagai Sekretaris dalam seminar proposal sampai sidan skripsi saya.
7. Bapak Andre Hasudungan Lubis, S.Ti, M.Sc selaku dosen pembimbing yang telah banyak memberikan bimbingan, arahan, dengan penuh kesabaran dan



selalu bersedia meluangkan waktu dan tenaga dalam membimbing dari awal pembuatan proposal hingga selesai penyusunan skripsi ini.

8. Bapak Dr. Dian Noviandri, ST, M.Kom sebagai ketua dalam seminar hasil saya sampai sidang skripsi yang telah memberikan saran dan masukan dalam skripsi saya.
9. Ibu Susilawati, S.Kom, M.Kom sebagai penguji dalam seminar proposal saya sampai sidang skripsi yang telah memberikan saran dan masukan dalam skripsi saya.
10. Kepada seluruh keluarga besar yang selalu memberikan semangat dan doa dalam mengerjakan skripsi ini.
11. Saya juga berterima kasih kepada teman-teman yang telah mengingatkan dan menjadi tempat berbagi pikiran dalam mengerjakan skripsi ini.

Akhir kata, peneliti mengucapkan terima kasih kepada semua pihak yang membantu dalam proses penelitian dalam penyelesaian skripsi ini dan semoga bermanfaat untuk peneliti selanjutnya.

Penulis,

Budiman Sujatmiko

## Daftar Isi

LEMBAR PENGESAHAN.....	<b>Error! Bookmark not defined.</b>
HALAMAN PERNYATAAN.....	<b>Error! Bookmark not defined.</b>
ABSTRAK .....	iv
Daftar Riwayat Hidup .....	ivii
KATA PENGANTAR .....	vii
Daftar Isi .....	vix
Daftar Tabel.....	xii
Daftar Gambar .....	xiii
BAB I PENDAHULUAN .....	1
1.1 Latar Belakang .....	1
1.2 Rumusan Masalah .....	3
1.3 Batasan Masalah.....	3
1.4 Tujuan Penelitian.....	4
1.5 Manfaat Penelitian.....	4
1.6 Sistematika Penulisan.....	5
BAB II TINJAUAN PUSTAKA .....	7
2.1 Instagram.....	7
2.2 Definisi <i>Cyberbullying</i> .....	8
2.2.1 Teks yang berkaitan dengan <i>cyberbullying</i> .....	10
2.3 Data Mining .....	10
2.4 <i>Text Mining</i> .....	11

2.4.1 <i>Text Preprocessing</i> .....	12
2.4.2 <i>Term Weighting</i> .....	13
2.5 Metode <i>Clustering</i> .....	14
2.3.1 Jenis – jenis metode dalam <i>clustering</i> .....	15
2.6 Algoritma DBSCAN.....	15
2.7 <i>Davis Boulden Index (DBI)</i> .....	17
2.7 Penelitian Terdahulu.....	19
<b>BAB III METODE PENELITIAN</b> .....	<b>22</b>
3.1 Prosedur Penelitian.....	22
3.2 Identifikasi masalah.....	22
3.3 Studi Literatur .....	23
3.4 Pengumpulan Data.....	23
3.5 Pengolahan Data.....	24
3.6 Pengujian Data .....	35
<b>BAB IV HASIL DAN PEMBAHASAN</b> .....	<b>45</b>
4.1 Hasil .....	45
4.2 Pembahasan .....	45
4.2.1 Pengambilan Data .....	45
4.2.2 Tahapan <i>Text Preprocessing</i> .....	47
4.2.4 <i>Term Weighting</i> .....	50
4.2.4 Perhitungan Algoritma DBSCAN .....	51
4.2.5 <i>Davies Boulding Indexs (DBI)</i> .....	55

5.1 Kesimpulan .....	57
5.2 Saran .....	57
DAFTAR PUSTAKA .....	58





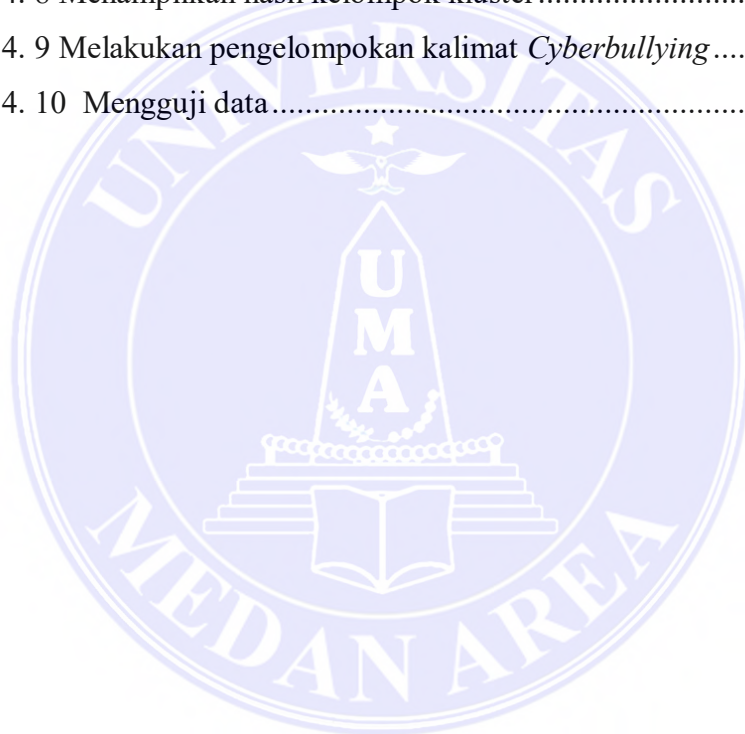
## Daftar Tabel

Tabel 2. 1 Contoh kalimat <i>Cyberbullying</i> di Instagram.....	10
Tabel 2. 2 Tabel Penelitian Terdahulu .....	19
Tabel 3. 1 Contoh kalimat <i>Cyberbullying</i> diberikan label .....	24
Tabel 3. 2 Proses <i>case folding</i> pada kalimat <i>Cyberbullying</i> .....	25
Tabel 4. 1 Data <i>Cyberbullying</i> .....	45
Tabel 4. 2 Tabel Pengujian Algoritma DBSCAN .....	55



## Daftar Gambar

Gambar 3. 1 Diagram dalam tahapan penelitian .....	22
Gambar 4.1 Tahapan <i>Case Folding</i> .....	47
Gambar 4.2 Tahapan <i>Tokenizing</i> .....	48
Gambar 4.3 Pemberian Nilai Pada Setiap Kata .....	49
Gambar 4.4 Tahapan <i>Filtering</i> .....	50
Gambar 4.5 Proses Pembobotan disetiap kata dalam <i>document</i> .....	50
Gambar 4.6 Proses Perkalian TF - IDF .....	51
Gambar 4. 7 Perhitungan Algoritma DBSCAN .....	52
Gambar 4. 8 Menampilkan hasil kelompok kluster .....	53
Gambar 4. 9 Melakukan pengelompokan kalimat <i>Cyberbullying</i> .....	54
Gambar 4. 10 Mengguji data .....	56



# BAB I

## PENDAHULUAN

### 1.1 Latar Belakang

Di era revolusi 4.0 yang telah masuk di Indonesia membuat semakin banyak teknologi yang semakin berkembang pesat. Tentunya dengan ini berjalan beriringan dengan tingkat perkembangan penggunaan media sosial, baik dari kalangan anak remaja, orang dewasa bahkan kalangan orang tuapun tidak ingin ketinggalan mengikuti zaman era revolusi 4.0. Di tahun 2022, penggunaan Internet dan media sosial mengalami peningkatan yang cukup signifikan sebanyak 12,6%. Pada tahun 2021, pengguna media sosial memiliki 170 juta pengguna, naik hingga 191,4 juta pengguna di tahun 2022 (Ramadhan dkk., 2022). Disisi lain, pada tahun 2021, pengguna media sosial Instagram tembus mencapai di angka 87 juta pengguna, sebanyak 52,5% didominasi oleh kaum perempuan (Mustaqimah & Rahmah, 2022). Melihat tren data tersebut, Jumlah pengguna media sosial yang ada di Indonesia mengalami peningkatan yang cukup meningkat pada setiap bulan dan tahunnya.

Tetapi, dibalik penggunaan media sosial yang semakin meningkat banyak yang masih menyalahgunakannya dan merugikan orang lain seperti kerugian materi, emosi, mental, bahkan sampai terjadi *cyberbullying* di media sosial, salah satunya pengguna media sosial Instagram banyak digunakan oleh kalangan remaja sampai dewasa. Instagram merupakan situs jejaring sosial atau *platform* yang digunakan untuk melihat ataupun mencari informasi dalam bentuk foto dan video (Lukiani dkk., 2021). Sehingga pada penggunaan media sosial menimbulkan pro dan kontra dikalangan masyarakat karena media sosial dapat menyebabkan kecanduan bagi individu dan menjadikan ketergantungan untuk menggunakannya bahkan bisa meningkatkan kemalasan tentunya menurunkan motivasi belajar bagi kalangan pelajar atau mahasiswa.

*Cyberbullying* adalah tindakan kegiatan yang berbahaya yang sering terjadi dipengguna media sosial dengan mengeluarkan kalimat – kalimat negatif yang dapat membahayakan kesehatan mental seseorang (Muzdalifah & Putri, 2019).

Sehingga, diperlukan adanya pengelompokan untuk mengatasi tindakan *cyberbullying* untuk membantu pihak – pihak berwenang seperti kominfo (Kementerian Komunikasi dan Informatika) untuk mengatasi dan meminimalisir terjadinya tindakan *cyberbullying* yang terjadi di media sosial khususnya Instagram dan memberikan informasi edukasi tentang *cyberbullying*.

Maka dari itu perlu adanya pengelompokan tindakan tersebut seperti penggunaan data mining untuk mengolah data yang besar dan membantu untuk mengambil keputusan. Data mining merupakan sebuah proses untuk menemukan pola atau informasi menarik dalam data terpilih dengan menggunakan metode atau teknik tertentu (Manurung & Hasugian, 2019). Data mining digunakan untuk mengklasifikasikan sebuah pola untuk mendapatkan data di dalamnya (Mustafa & Simpen, 2019). *Text Mining* merupakan bagian dari *data mining*. *Text Mining* adalah sebuah proses mencari informasi dimana seorang pengguna dapat melakukan interaksi dengan kumpulan – kumpulan dokumen menggunakan alat analisis merupakan komponen yang ada di dalam data mining (Fauziah dkk., 2019). *Text mining* dapat di pergunakan untuk menemukan pola – pola yang ada di teks dalam bahasa yang natural yang tidak tersusun seperti, artikel, halaman web, buku, email, dan lain-lainnya (Fitriani dkk., 2021).

Pengelompokan pada data mining memiliki berbagai jenis algoritma, salah satunya adalah algoritma DBSCAN (*Density-based spatial clustering of applications with noise*). Algoritma DBSCAN adalah suatu metode pengelompokan berbasis kepadatan untuk melakukan clustering yang membangun area berdasarkan densitas yang terkoneksi (Setiawan dkk., 2021). Kelebihan algoritma DBSCAN dibandingkan dengan algoritma lain menurut (Sari & Primajaya, 2019) diantaranya, seperti dalam melakukan sebuah proses kluster yang tidak terstruktur bisa lebih efisien, pada proses *clustering* yang dilakukan berdasarkan tingkat kerapatan atau kedekatan jarak antar objek, dapat melakukan pendektasian outlier/noise, dalam melakukan proses input tidak perlu mendapatkan awalan berupa jumlah kluster seperti pada algoritma *K-Means* dan *K-Medoids*, pada kluster yang sulit tidak beraturan dapat dikenalin.



Algoritma DBSCAN telah banyak digunakan pada penelitian yang berkaitan dengan *Text-Mining*. Pada penelitian terdahulu dengan judul pengelompokan *tweets* pada akun twitter tokopedia menggunakan Algoritma DBSCAN(Alamsyah dkk., 2022)dengan hasil penelitian yaitu bahwa konsumen dari perusahaan e-commerce Tokopedia lebih tertarik dengan program WIB atau Waktu Indonesia Belanja yang diadakan setiap bulan oleh Tokopedia. Hal ini sesuai dengan riset perusahaan konsultan marketing MarkPlus Inc. pada tahun 2020 yang menyatakan bahwa kampanye WIB dari Tokopedia menempati peringkat kedua kampanye e-commerce yang paling dikenal masyarakat Indonesia sebanyak 71%. Oleh karena itu, PT Tokopedia Indonesia diharapkan mampu memanfaatkan konten WIB semaksimal mungkin.

Sehingga, dapat disimpulkan bahwa semakin berkembangnya teknologi penggunaan media sosial juga semakin bertambah penggunaannya. Sejalan dengan itu, banyak terjadi penyalahgunaan media sosial yang dapat merugikan bagi berbagai pihak, seperti *cyberbullying*. *Cyberbullying* adalah suatu tindakan yang dapat merusak psikologis korbannya. Sehingga perlu dilakukannya sebuah tindakan untuk mengatasi permasalahan ini dengan cara mengelompokan teks-teks yang mengandung kalimat *cyberbullying*. Dengan dibantu teknik data mining untuk memecahkan masalah untuk mengatasi kumpulan data yang besar dan menggunakan algoritma DBSCAN. Sehingga, nantinya hasil pengelompokan ini dapat menjadi informasi edukatif bagi masyarakat. Maka Pada peneliti kali ini, Penelitian dapat melakukans sebuah penelitian dengan judul “**Analisis Metode Clustering dengan Algoritma DBSCAN Pengelompokan Cyberbullying di Instagram**”.

## 1.2 Rumusan Masalah

Berdasarkan latar belakang diatas, maka rumusan masalah pada penelitian ini adalah bagaimana menganalisis metode *clustering* dengan algoritma DBSCAN untuk pengelompokkan *cyberbullying* di Instagram dapat dilakukan.

## 1.3 Batasan Masalah

Batasan masalah dalam penelitian ini adalah:

1. Data yang digunakan dalam proses penelitian ini merupakan data sekunder yang didapatkan dari sumber dataset Kaggle.
2. Data yang digunakan merupakan data dalam bahasa Indonesia dengan jumlah data sebanyak 649 komentar di Instagram
3. Dalam proses melakukan penelitian ini dibantu dengan menggunakan Algoritma DBSCAN.
4. Dalam proses pengujian data menggunakan *Davies Boulding Index* (DBI).
5. Pada proses penelitian Analisis metode *clustering*, implementasinya menggunakan *Google Colab*.

#### 1.4 Tujuan Penelitian

Adapun tujuan dari penelitian ini dilakukan antara lain:

1. Untuk mengelompokkan data *cyberbullying* di Instagram dengan menggunakan Algoritma DBSCAN.
2. Mengkategorikan dan mengidentifikasi berupa kalimat-kalimat yang mencerminkan perilaku *cyberbullying* di Instagram.
3. Membantu dalam memahami karakteristik dan tren *cyberbullying* di platform Instagram, yang dapat digunakan sebagai bahan referensi untuk pengembangan kebijakan, langkah-langkah pencegahan, atau tindakan penegakan hukum.
4. Menguji Algoritma DBSCAN untuk dilakukan penelitian, apakah algoritma tersebut bisa untuk dilakukan sebuah *clustering* atau pengelompokan kalimat – kalimat *cyberbullying* di Instagram.

#### 1.5 Manfaat Penelitian

Adapun manfaat yang dapat di peroleh dari penelitian ini yaitu:

1. Untuk memberikan kemudahan dan mempercepat penyelesaian kasus dalam mengelompokkan data *Cyberbullying* di instagram dengan menggunakan Algoritma DBSCAN.
2. Untuk mengedukasi dalam memberikan informasi tentang *cyberbullying* yang termasuk kalimat – kalimat tindakan *cyberbullying*

3. Untuk membantu pihak – pihak berwenang seperti kominfo untuk mengatasi dan menimalisir terjadinya tindakan *cyberbullying* yang terjadi di media sosial khususnya instagram.
4. Untuk memberikan referensi khususnya kepada orang tua untuk mengawasi anaknya supaya tidak melukan tindakan *cyberbullying*.
5. Dapat mengurangi atau menimalisir terjadi tindakan *cyberbullying* di media sosial.
6. Dapat menggunakan media sosial dengan lebih bijak lagi.

## 1.6 Sistematika Penulisan

Sistematika penulisan pada penelitian ini ditunjukan kepada pembaca untuk lebih mudah memahami isi dari penelitian ini. Berikut sistematika penulisan penelitian ini secara garis besar, yaitu:

### **BAB I PENDAHULUAN**

Pada bagian pendahuluan membahas seputar latar belakang dari penelitian ini, dimana untuk mengetahui sebab penelitian ini dilakukan dan selanjutnya membahas rumusan masalah, batasan masalah, tujuan masalah dan manfaat penelitian.

### **BAB II TINJAUAN PUSTAKA**

Pada tinjauan pustaka membahas hal – hal yang mendasar berisi teori – teori yang berkaitan dengan analisis metode *clustering* dengan algoritma DBSCAN pengelompokan *cyberbullying* di Instagram dan hal- hal yang mendukung dalam melakukan penelitian analisis *clustering*.

### **BAB III METODE PENELITIAN**

Pada bagian ini membahas tentang sumber data dan tahapan yang dilakukan dalam melakukan penelitian ini serta gambaran umum sistem yang akan dikerjakan.

### **BAB IV PEMBAHASAN DAN HASIL**

Pada bagian ini dilakukan pembahasan dan menampilkan hasil dari penelitian yang telah dilakukan.

## **BAB V KESIMPULAN DAN SARAN**

Pada bagian ini peneliti menarik kesimpulan dari hasil penelitian yang telah dilakukan dan memberikan beberapa saran – saran yang nantinya, diharapkan penelitian selanjutnya dapat dikembangkan.





## BAB II

### TINJAUAN PUSTAKA

#### 2.1 Instagram

Instagram merupakan sebuah aplikasi sosial media yang digunakan untuk berbagi foto dan video yang memungkinkan seorang pengguna Instagram untuk mengambil foto dan video, pengguna bisa menerapkan fitur digital yang tersedia di Instagram dan bisa membagikannya ke semua pengguna jejaring sosial (Y. Fitriani, 2021). Perkembangan Instagram semenjak pertama kali dirilis pada 06 Oktober 2010 mengalami peningkatan yang sangat signifikan yang awal mula rilis hanya berjumlah 25.000 pengguna Instagram. Selain digunakan untuk media sosial, Instagram juga digunakan sebagai media informasi dan komunikasi bagi pengguna, namun untuk sekarang Instagram juga digunakan sebagai media untuk peluang bisnis online. Demi kenyamanan penggunaannya, Instagram terus menghadirkan fitur – fitur baru yang dapat digunakan untuk memanjakan pengguna Instagram, mulai dari fitur Instagram *story*, filter Instagram, *instagram reels*, live Instagram, IGTV, dan hal berbagai fitur lainnya yang terdapat pada Instagram.

Pada umumnya tujuan dibentuknya Instagram adalah untuk sebagai sarana kegemaran bagi pengguna Instagram yang memiliki keinginan untuk berbagi ataupun mempublikasikan kegiatan, barang, tempat photo, dan video diri sendiri kedalam Instagram (Febriani dkk., 2021). Dalam hal ini memberikan kebebasan berekspresi bagi pengguna Instagram itu sendiri. Pengguna Instagram dapat menggunakan Instagram juga sebagai media hiburan dan menyenangkan. Di dalam perkembangan teknologi yang semakin pesat, Instagram juga telah melakukan evolusi menjadi media yang mampu menyesuaikan perkembangan dari perubahan kehidupan sosial di masa sekarang, seperti sebagai media *entertainment*. Media *entertainment* adalah alat komunikasi yang digunakan sebagai tempat hiburan dan kesenangan. Selain Instagram digunakan sebagai media sosial dan media *entertainment*, Instagram merupakan salah satu alat yang bisa bermanfaat bagi penggunaannya untuk mencari dan menemukan berbagai macam ide kreatif dan inspirasi yang dibutuhkan dari melihat photo ataupun video dari pengguna Instagram yang lainnya (Apifah, 2021).

Dari berbagai macam manfaat dan tujuan penggunaan Instagram pasti ada dampak buruk ataupun dampak negatif yang di peroleh dari penggunaan instagram salah satunya yang saat ini terus berkembang adalah kasus *cyberbullying* yang terdapat dalam Instagram (Maladi, 2020)

## 2.2 Definisi *Cyberbullying*

*Cyberbullying* adalah bentuk dari sikap intimidasi yang dilakukan terhadap pengguna media media sosial yang terjadi di dunia maya seperti Twitter, Facebook, Instagram yang dapat menyebabkan ketidak nyamanan terhadap pengguna media sosial tersebut (Winoto, 2019). Terjadinya tindakan dari *cyberbullying* biasanya dilakukan oleh pengguna remaja karena menganggap bahwa media sosial yang digunakan merupakan dunia bagi mereka untuk melakukan berbagai macam bentuk ekspresi yang mereka inginkan, mengutarakan pendapatnya, menuangkan perasaan yang sedang di alami, dan mengungkapkan rasa emosi dan kekecewanya. Bentuk ekspresi yang biasa di temukan di media sosial seperti berupa foto, video, teks/kalimat sara, dan bentuk audio visual. Terjadi tindakan *cyberbullying* karena kurangnya edukasi yang diberikan oleh orang tua dan orang sekitar dan kurangnya pengawasan yang dilakukan sehingga pengguna media sosial seperti anak remaja tidak bisa mengontrol dirinya sendiri.

Menurut (Jalal dkk., 2021) *cyberbullying* adalah bentuk tindakan *bullying* yang begitu rentan terjadi dikalangan anak remaja, bentuk tindakan intimidasi yang dilakukan untuk menyakiti orang lain melalui media sosial yang digunakan oleh pelakunya. Adapun negatif – negatif yang terjadi, yang dapat menyebabkan terjadinya tindakan *cyberbullying* yaitu (MAEMUNAH, 2020):

1. Memiliki sikap iseng di dunia maya.
2. Kurangnya mendapat edukasi dari orang tua.
3. Mencari perhatian di dunia maya
4. Pengguna menjadi korban *bully* di dunia nyata sehingga melampiaskan di dunia maya
5. Memiliki rasa kecewa terhadap pengguna instagram lainnya seperti artis, dll.

6. Melihat konten – konten negatif yang memicu terjadinya fenomena *cyberbullying*.

Menurut (Irfan, 2022) ada beberapa bentuk atau jenis – jenis *bullying* yang terjadi di media sosial seperti:

1. *Harassment* (gangguan) merupakan tindakan dengan melakukan mengirimkan pesan terhadap korban yang berisi gangguan yang dilakukan secara terus menerus pada sms, e-mail, maupun teks di jejaring sosial.
2. *Flaming* (terbakar) adalah tindakan dengan cara mengirimkan pesan teks yang memiliki isi berupa kata – kata yang penuh amarah dan frontal. Kata frontal adalah kalimat yang disampaikan secara berlebihan dan tidak wajar sehingga korban akan mengalami ketidak nyamanan.
3. *Denigration* (fitnah) merupakan tindakan yang dilakukan dengan cara mengumbar keburukan seseorang pengguna instagram dengan maksud untuk menjelekan pengguna instagram tersebut.
4. *Cyberstalking* (Penguntitan) yaitu tindakan untuk mengganggu dan mencemarkan nama baik seseorang secara terus menerus sehingga akan menyebabkan ketakutan terhadap orang tersebut.
5. *Outing* (menyebarkan rahasia pribadi) dan *Trickery* (penipuan) adalah tindakan *cyberbullying* yang dilakukan dengan cara membujuk atau melakukan tipuan seseorang untuk mengungkapkan rahasia pribadi dan akan menyebarkannya.
6. *Impersonation* merupakan tindakan *cyberbullying* dengan cara berpura – pura menjadi orang lain dan melakukan pengiriman pesan yang tidak baik.

Dari penjelasan diatas penyebab dari terjadinya *cyberbullying*, perlunya edukasi yang harus dilakukan oleh orang tua, orang sekitar bahkan pemerintah dan tentunya peran dari orang tua itu sendiri sangat penting untuk mengawasi perilaku anaknya di media sosial untuk bisa menggunakan media sosial dengan baik. Sehingga tindakan *cyberbullying* dapat di minimalisir dan akan memberikan dampak yang sangat positif yang akan di dapat oleh penggunanya.

### 2.2.1 Teks yang berkaitan dengan *cyberbullying*

Berikut merupakan beberapa sampel data kalimat *cyberbullying* berupa komentar yang terjadi di media sosial khususnya aplikasi instagram, yaitu:

Tabel 2. 1 Contoh kalimat *Cyberbullying* di Instagram

No.	Komentar Instagram	Kategori
1.	“Kaka tidur yaa, udah pagi, gabooleh capek2”	<i>Nonbullying</i>
2.	“Makan nasi padang aja begini badannya”	<i>Nonbullying</i>
3.	“Uda gila sekarang ini orang wkwk, udah jomblo jadi psikolog ga laku, ya jadilah seperti itu”	<i>Bullying</i>
4.	“Apaan sih gak jelas hidupnya”	<i>Bullying</i>
5.	“Kaku kek kanebo kering hahaha”	<i>Bullying</i>
6.	“Cantik banget kalau pakai baju kayak gitu”	<i>Nonbullying</i>
7.	“Itu tangan apa paha hahaha”	<i>Bullying</i>
8.	“Inimah bukan main alat musik lagi. Olahraga jari dan kaki ini mah”	<i>Nonbullying</i>
9.	“Sampe ga bisa berkata2 lagi buat isyana, sangat2 sempurna org ini”	<i>Nonbullying</i>
10.	“ORANG INI TERLIHAT SANGAT SANGAT STRES DAN GILA”	<i>Bullying</i>

### 2.3 Data Mining

Data mining merupakan suatu proses untuk menemukan pola, korelasi, dan tren terbaru yang sedang berkembang dan yang memiliki makna dengan memilah – memilah data dalam jumlah yang cukup besar yang tersimpan di dalam repository (Nabila dkk., 2021). Menurut (Mardi, 2017). Data Mining merupakan sebuah proses yang menggunakan teknik matematika, kecerdasan buatan, statistik, dan *machine learning* untuk mengidentifikasi informasi yang bermanfaat. Data mining sering digunakan untuk menemukan pengetahuan yang tersembunyi didalam sebuah database. Tujuan utama dari data mining yaitu untuk menggali, menambang, untuk menemukan pengetahuan dari data yang telah didapat. Data mining adalah salah satu langkah dalam (KDD) *Knowledge discovery in database* (Kharis & Zili, 2022).



*Knowledge discovery in database* (KDD) merupakan menidentifikasi pola yang ada di dalam data dan keseluruhan proses untuk mencari data, dimana pola yang di dapatkan memiliki sifat baru, sah, dapat dimengerti dan bermanfaat. Menurut (Kharis & Zili, 2022) proses *knowledge discovery in database* (KDD) secara garis besar diantaranya :

1. *Data Integration* (Integrasi data)
2. *Data Cleaning* (Pembersihan data)
3. *Data Transformation* (Transformasi data)
4. *Data Selection* (Pemilihan data)
5. *Data Mining* (Penggalian data)
6. *Pattern Evaluation* (Evaluasi pola)
7. *Knowledge presentation* (Penyajian pengetahuan)

Menurut (Utomo & Mesran, 2020) *data mining* serta KDD memiliki tujuan yang sama yakni dengan menggunakan data yang telah ada ataupun data yang sudah tersedia pada database kemudian dilakukannya pengolahan data untuk memperoleh suatu informasi baru yang dapat bermanfaat.

#### **2.4 Text Mining**

*Text mining* merupakan suatu langkah untuk analisis teks yang dapat digunakan secara otomatis oleh sebuah komputer untuk mendapatkan suatu informasi yang berkualitas dari sebuah rangkain yang telah dirangkum dalam bentuk dokumen (Putra & Wardani, 2020). Tujuan *text mining* adalah untuk menemukan sebuah pola pada data supaya dapat digunakan oleh manusia untuk memabntu pekerjaannya (Sholih'afif dkk., 2021). Sumber data yang digunakan pada *text mining* merupakan kumpulan – kumpulan teks yang memiliki format dengan tidak struktur.

*Text mining* adalah sebuah proses mencari informasi yang di dapatkan dari sekumpulan dokumen dari masa ke masa dengan serangkain alat analisis untuk dapat dilakukan identifikasi dan mengeksplorasi pada penggunaan pola data yang ada (Alamsyah dkk., 2022). *Text mining* dapat dipergunakan untuk menemukan pola – pola yang ada di teks dalam bahasa yang natural yang tidak tersusun seperti,



artikel, halaman web, buku, email, dll (Fitriani dkk., 2021). Menurut (Rachman dkk., 2021) langkah awal sebelum melakukan suatu data teks dianalisis menggunakan metode dalam *text mining* adalah melakukan *text preprocessing*, tujuan dilakukan *text mining* ini adalah untuk memperoleh suatu *set term index*.

#### 2.4.1 Text Preprocessing

*Text preprocessing* merupakan sebuah proses yang dilakukan untuk menyeleksi berupa data text, agar data tersebut bisa lebih terstruktur dengan melalui berbagai tahapan meliputi *case folding*, *tokenizing*, *filtering* dan tahapan *stemming* (Hermawan & Ismiati, 2020). Berikut merupakan tahapan dari *text preprocessing* (Rachman dkk., 2021) yaitu :

##### 1. Case Folding

*Case folding* adalah sebuah proses yang dilakukan untuk mengubah huruf atau text menjadi huruf yang lebih kecil (sederhana) (Jumeilah & others, 2017).

##### 2. Tokenizing

*Tokenizing* merupakan sebuah proses yang digunakan untuk pemotongan *string input* berdasarkan tiap kata penyusunannya (Khaira dkk., 2020). Pada tahap pemrosesan ini juga dilakukan sebuah tindakan untuk menghilangkan angka, tanda baca dan karakter lain selain huruf alphabet (Umar & Fatimah, 2018).

##### 3. Filtering

*Filtering* merupakan sebuah tahapan dengan melakukan pemilihan kata – kata yang penting saja yang diperoleh dari hasil *token*, yaitu kata – kata yang dapat digunakan untuk dijadikan sebuah wakilan isi dari suatu dokumen (Aziza dkk., 2022). Pada tahapan ini merupakan menghilangkan *stopword* dan mengubah kata – kata menjadi lebih terhadap kata yang berimbuan sederhana agar mudah dipahami. *Stopword* adalah kosakata yang bukan termasuk kata unik atau ciri pada sebuah dokumen atau tidak melakukan untuk menyampaikan pesan apapun pada kalimat atau teks secara signifikan (Purnajaya dkk., 2022). Penghilangan *stopword* ini dapat mengurangi ukuran index dan waktu pemrosesan.

#### 4. Stemming

*Stemming* merupakan suatu proses yang digunakan untuk pengubahan bentuk kata menjadi kata dasar (Dinata dkk., 2020). Pada tahap *stemming* dapat digunakan untuk melakukan proses pengelompokan kata yang memiliki kata dasar dan arti yang serupa namun memiliki bentuk yang berbeda karena mendapatkan imbuhan yang berbeda pula (Sanrilla dkk., 2022).

##### 2.4.2 Term Weighting

Untuk melakukan pembobotan dalam *Terim Weighting* seperti berikut :

##### 1. Terim Frequency

*Terim Frequency* merupakan sebuah frekuensi dari kemunculan suatu *term* (frasa/kata) dalam sebuah dokumen yang terkait (Pembrani & dkk., 2019). Semakin besar jumlah kemunculan sebuah *term* dalam dokumen, maka semakin besar pula bobotnya (Hidayah dkk., 2021).

##### 2. Term Weighting

Setelah jumlah *frequency* dari setiap *term*nya terdefinisi, maka langkah selanjutnya yaitu mencari bobot dari setiap *term* tersebut. Perhitungan bobot dari setiap *term* ini biasa disebut dengan *term weighting*. Untuk menghitung bobot dari TF (*TF weight*) dapat dihitung dengan Persamaan 2.1 Sebagai berikut :

$$W_{t,f} = \begin{cases} 1 + \log_{10} tf_{t,d}, & \text{if } tf_{t,d} > 0 \\ 0, & \text{otherwise} \end{cases} \quad (2.1)$$

Keterangan :

- $W_{t,f}$  = Bobot dari kemunculan *term* pada tiap dokumen
- $tf_{t,d}$  = Kemunculan *term* pada dokumen tertentu

##### 5. Inversi Document Frequency

*Inversi Document Frequency* (IDF) adalah suatu dokumen yang mengandung *term* atau token atau kata  $t$ . Untuk memperoleh dari nilai IDF, maka dapat digunakan Persamaan 2.2

$$idf_t = \log_{10} N/df_t \quad (2.2)$$

Keterangan :

- N = Jumlah total dokumen

df<sub>t</sub>= Banyak dokumen yang memuat t

## 6. TF – IDF (*Term Frequency-Inversi Document Frequency*)

*Weight Term Document* atau biasa disebut dengan TF-IDF dari sebuah *term* atau token atau kata yang merupakan hasil yang diperoleh dari perkalian antara *tf weight* dengan *idf*. Untuk memperoleh *weight term document*, maka dapat dihitung dengan menggunakan Persamaan 2.3

$$\begin{aligned} W_{t,d} &= W_{t,f} \times idf_t \\ &= W_{t,f} \times \log_{10} N/df_t \end{aligned} \quad (2.3)$$

Keterangan:

- N : Jumlah total dokumen
- $W_{t,f}$  : Bobot dari kemunculan *term* pada tiap dokumen
- $idf_t$ = Banyak dokumen yang memuat

## 2.5 Metode *Clustering*

Metode *clustering* data merupakan sebuah proses pengolahan data yang dilakukan untuk mengelompokkan sekumpulan data tanpa sebuah atribut kelas yang telah didefinisikan sebelumnya yang memiliki sifat *unsupervised* (tanpa arahan) yang memiliki arti bahwa karakteristik tiap kluster tidak ditentukan sebelumnya, melainkan berdasarkan tingkat kemiripan atribut – atribut dari suatu kelompok atau kluster(Santoso, 2017). Melakukan sebuah analisa terhadap data dengan menggunakan metode *clustering*, akan sangat membantu untuk membentuk setiap partisi – partisi yang tentunya berguna untuk mengolah data dengan jumlah yang besar menjadi komponen – komponen yang lebih kecil, sehingga data tersebut bisa lebih disederhanakan berdasarkan setiap objeknya. Objek didalam data *cluster* memiliki kemiripan karakteristik antara satu sama lainnya dan memiliki perbedaan dengan *cluster* yang lainnya.

Metode *clustering* memiliki manfaat dalam melakukan analisa dan prediksi. Metode *clustering* biasanya digunakan dalam berbagai macam hal seperti digunakan untuk *Business Intelligence*, pengenalan pola citra, bidang ilmu biologi, *web search*, dan untuk keaman web (*security*). Menurut (Defiyanti, 2017) *Clustering* atau klusterisasi merupakan salah satu alat bantu pada data mining yang memiliki tujuan untuk mengelompokkan objek – objek data ke dalam cluster berdasarkan tingkat kemiripan data satu dengan yang lainnya. Objek – objek yang dikelompokkan berdasarkan memiliki prinsip kemaksimalan kesamaan objek pada cluster yang sama dan memaksimalkan perbedaan terhadap cluster yang berbeda. Objek yang memiliki kesamaan biasanya dapat diperoleh dari nilai – nilai atribut atribut yang mendefinisikan objek data, sedangkan objek – objek data biasanya direpresentasikan sebagai sebuah titik dalam ruang multidimensi. Menurut (Reliovani dkk., 2021) ada beberapa manfaat menggunakan metode *clustering* diantaranya :

1. Metode *Clustering* merupakan metode segmentasi data yang sangat bermanfaat dalam melakukan prediksi dan melakukan sebuah analisa masalah bisnis tertentu seperti, *marketing*, segmentasi pasar dan melakukan sebuah pemetaan wilayah.
2. Metode *Clustering* dapat melakukan identifikasi objek dalam berbagai bidang contohnya seperti *image processing* dan *computer vision*.

### 2.3.1 Jenis – jenis metode dalam *clustering*

Terdapat beberapa jenis – jenis metode *clustering* yang dapat digunakan, seperti:

1. *Centroid – base clustering*
2. *Distribution – based clustering*
3. *Density – based clustering*
4. *Hierarchical clustering* atau *connectivity-based cluster*

### 2.6 Algoritma DBSCAN

Algoritma merupakan sebuah sistem kerja komputer yang memiliki susunan secara sistematis dan logis dengan tujuan untuk memecahkan sebuah masalah yang sedang terjadi atau untuk menemukan tujuan tertentu (Maulana dkk, 2017).

Algoritma pada umumnya digunakan untuk dapat melakukan inferensi otomatis, melakukan perhitungan, dan dapat untuk mengolah data pada komputer dengan menggunakan software tertentu sesuai kebutuhan. Sehingga tujuan algoritma ini merupakan untuk memecahkan masalah dengan langkah – langkah yang logis dan terurut.

Algoritma DBSCAN merupakan salah satu algoritma yang digunakan untuk pengelompokan data yang dimana pengelompokan data algoritma berbasis kepadatan (Virgatama dkk., 2019). Algoritma DBSCAN merupakan salah satu algoritma yang populer digunakan dan menjadi pilihan untuk di kutip dalam literature ilmiah. Algoritma DBSCAN adalah suatu metode pengelompokan berbasis kepadatan untuk melakukan clustering yang membangun area berdasarkan densitas yang terkoneksi (Setiawan dkk., 2021). Adapun kelebihan dan kekurangan algoritma DBSCAN(Herdiani dkk., 2019), yaitu :

1. Kelebihan algoritma DBSCAN diantaranya:
  - a. Dapat memperoleh *cluster* dengan akurat dari bentuk data yang tidak struktur jika dibandingkan dengan *partitional clustering*.
  - b. Mampu menangani *outliner/noise*.
  - c. Algoritma DBSCAN sangat baik untuk data dengan jumlah yang sangat besar.
2. Kekurangan algoritma DBSCAN diantaranya:
  - a. Pada *dataset* berdimensi tinggi hasilnya kurang maksimal.
  - b. Memiliki permasalahan pada saat identifikasi *cluster* dari kepadatan yang bervariasi.
  - c. Tidak dapat membentuk fungsi kepadatan yang sesungguhnya, tetapi lebih kearah poin – poin kepadatan yang saling berhubungan dan membentuk *graf*.

Menurut (Isnarwaty & Irhamah, 2020) algoritma DBSCAN membutuhkan parameter yang sangat penting, yaitu parameter radius (*Eps*) dan jumlah minimum point yang digunakan untuk membentuk kelompok (*MinPts*), berikut tahapan algoritma DBSCAN yaitu:



1. Menentukan jumlah parameter *MinPts* dan *Eps*.
2. Pilih kalimat *cyberbullying*  $p$  secara acak
3. Menghitung jumlah kalimat *cyberbullying* yang ditentukan oleh parameter radius (*Eps*). Jika jumlahnya mencukupi (lebih dari atau sama dengan  $\epsilon$ ), data akan ditandai dengan sebagai inti (*core point*).
4. Menghitung jarak titik *core point* dengan *point* yang lain dengan jarak *Euclidean*. Berikut ini merupakan rumus jarak *Euclidean* yang ditunjukkan pada Persamaan (2.4).

$$d_{ip} = \sqrt{\sum_{j=1}^m (x_{ji} - y_{jp})^2}, \quad (2.4)$$

Keterangan :

$d_{ip}$  = Jarak *Euclidean* dari kalimat *cyberbullying* ke -  $i$  ke pusat *cluster* ke -  $k$

$x_{ji}$  = Frekuensi kemunculan kata ke- $j$  pada kalimat *cyberbullying* ke- $i$

$y_{jp}$  = Frekuensi kemunculan kata ke- $j$  pada titik pusat ke- $p$

$m$  = Banyak kata

5. Buat *cluster* baru dengan menambahkan kalimat *cyberbullying*  $p$  ke dalam *cluster*.
6. Melakukan identifikasi pada data yang ditandai sebagai *core point*.
7. Lanjutkan proses sampai semua *point* telah diproses.
8. Jika ada kalimat *cyberbullying* yang tidak termasuk ke dalam *cluster* manapun akan ditandai sebagai *noise*.

## 2.7 Davis Boulden Index (DBI)

*Davies Bouldin Index* dapat digunakan untuk menentukan jumlah kluster dengan nilai terbaik dengan memiliki sebuah tujuan untuk memaksimalkan jarak antar kluster satu dengan jarak kluster yang lainnya dengan waktu yang sama untuk meminimalkan jarak antar kluster (Badruttamam dkk., 2020). Untuk mendapatkan hasil *clustering* yang baik, jarak *inter-cluster* harus tinggi dan jarak *intra-cluster* harus rendah, sehingga dapat ditemukan hasil kluster terbaik karena nilai DBI yang lebih rendah (Kartikasari, 2021). Ada empat tahapan dalam menghitung DBI (Septiani dkk., 2022) diantaranya, yaitu:

1. Untuk menghitung  $SSW$  (*Sum Of Square Within Cluster*) ialah dengan keterikatan anggota satu *cluster* atau seberapa mirip antara anggota satu dan dua dan semakin kecil, maka semakin bagus dikarenakan semakin mirip.  $SSW$  dihitung untuk mengetahui matrik/ kohesi/ homogenitas. Kohesi merupakan keterikatan anggota *cluster* dalam satu *cluster*.

$$SSW_i = \frac{1}{m_i} \sum_j^{m_i} d(x_j, c_i) \quad (2.5)$$

Keterangan :

$m_i$  = Jumlah data dalam *cluster* ke -  $i$

$x$  = Data dalam *cluster*

$d(x, c)$  = Jarak data ke *centroid*

$x_j$  = Data pada *cluster* tersebut

$c_j$  = *centroid cluster* ke- $i$

2. Untuk menghitung  $SSB$  (*Sum Of Square Between Cluster*) adalah jarak antar *cluster* cukup besar sehingga terpisah ke dalam kelompok lain. Tujuan  $SSB$  adalah untuk mengetahui heterogenitas/separasi. Separasi adalah perbedaan antara satu *cluster* dengan *cluster* lainnya.

$$SSB_{i,j} = (c_i, c_j) \quad 2.6$$

Keterangan :

$C_i$  = *Cluster* 1

$C_j$  = *Cluster* lainnya

$d(c_i, c_j)$  = Jarak antar *centroid* satu dengan lainnya

3. Untuk menghitung rasio yang berfungsi untuk dapat mengetahui seberapa bagus nilai perbandingan *cluster* satu dengan yang lainnya. Jumlah kohesi harus kecil sedangkan jumlah separasi harus lebih besar.

$$R_{i,j} = \frac{SSW_i + SSW_j}{SSB_{i,j}} \quad 2.7$$

Keterangan :

$R_{i,j}$  = Rasio antar *cluster*

$SSW_i$  = *Cluster* 1

$SSW_j$  = *Cluster* 2

$SSB_{i,j}$  = Separasi dari *cluster* 1 dan 2

#### 4. Untuk menghitung DBI (*Davis Bouldin Index*)

Faktanya jika hasil dari perhitungan DBI yang di dapatkan semakin kecil mendekati nol akan tetapi tidak negative (non – negatif  $\geq 0$ ), maka nilai hasil *clustering* semakin baik.

$$DBI = \frac{1}{K} \sum_{i=1}^k \max_{i \neq j} (R_{i,j}) \quad 2.8$$

Keterangan :

K = Kluster yang ada

$R_{i,j}$  = Rasio antara kluster i dan j

Max = Dicari rasio antara kluster yang terbesar

### 2.7 Penelitian Terdahulu

Berikut merupakan beberapa penelitian terdahulu yang menjadi referensi pada penelitian analisis metode *clustering* dengan algoritma DBSCAN pengelompokan *cyberbullyig* di instagram, yaitu:

Tabel 2. 2 Tabel Penelitian Terdahulu

No.	Penulis (Tahun)	Topik	Hasil
1.	Firdi Setiawan, Fairuz Azmi, Casi Setianingsih (2021)	<i>Clustering</i> Pada Data Sentimen Transportasi Online Menggunakan Algoritma DBSCAN.	Dari ketiga dataset positif, negatif, dan netral telah diuji menghasilkan nilai <i>silhouette coefficient</i> yang berbeda – beda.
2.	Deanira Qinanty Alamsyah, Sudarno, Puspita Kartikasari(2022)	Pengelompokan Tweets Pada Akun Twitter Tokopedia Menggunakan Algoritma <i>Density Based</i>	Berdasarkan hasil yang didapatkan terlihat bahwa konsumen dari

		<i>Spatial Clustering Of Applications With Noise.</i>	perusahaan e-commerce Tokopedia lebih tertarik dengan program WIB atau Waktu Indonesia Belanja yang diadakan setiap bulan oleh Tokopedia..
3.	Miranti Alysha Zulia Larasati, Nurul Anisa Sri Winarsih, Muhammad Syaifur Rohman, Galuh Wilujeng Saraswati (2022)	Penerapan Metode <i>K-Means Clustering</i> dalam Menganalisa Sentimen Masyarakat Terhadap K-Popers Pada Twitter	Metode <i>K-Means</i> baik dalam melakukan klasifikasi sentiment masyarakat pada sosial twitter dengan menghasilkan nilai <i>Silhouette index</i> 0.687974 yang artinya dataset tersebut digolongkan pada struktur yang baik
4.	Damaris Lalang, dan Mariam Lanmay(2022)	Aplikasi Metode Fuzzy Clustering Means Untuk Data Trending Kasus Vaksin Corona Pada Jejaring Sosial Twitter	Hasil pengolahan data vaksin covid-19 dengan <i>fuzzy clustering means</i> menghasilkan akhir konsergen dengan

			menampilkan 2 <i>cluster</i> dari 180 term.
5.	Wahyu Hadikristanto(2022)	Implementasi <i>Content Moderation</i> Dalam Sosial Media Instagram Untuk Deteksi <i>Cyberbullying</i> Dengan <i>Machine Learning</i> Berbasis <i>Cloud</i>	<i>Content Moderation</i> dengan <i>Machine Learning</i> Berbasis <i>Cloud</i> cukup efisien, akrena nilai akurasi yang di dapatkan sebesar 85%. Nilai akurasi yang tinggi menunjukkan hasil <i>Content Moderation</i> yang dilakukan dengan <i>Machine Learning</i> Berbasis <i>Cloud</i> cukup baik dan dapat diandalkan dalam mengelola konten di media sosial.

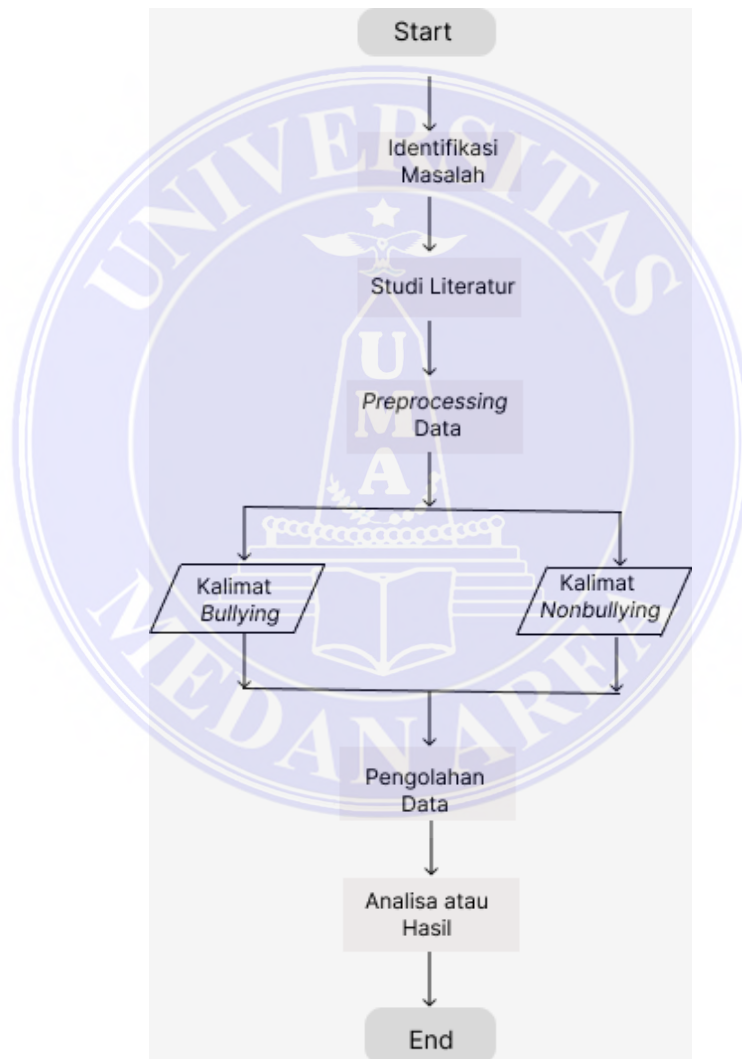


## BAB III

### METODE PENELITIAN

#### 3.1 Prosedur Penelitian

Pada penelitian ini menggunakan prosedur kerja yang meliputi beberapa prosedur yang di mulai dengan melakukan persiapan penelitian, pengumpulan data, pengolahan data serta pengambilan keputusan data (analisa atau hasil). Untuk lebih jelas, berikut tahapan penelitian yang dilakukan dapat dilihat pada Gambar 3.1.



Gambar 3. 1 Diagram dalam tahapan penelitian

#### 3.2 Identifikasi masalah

Pada tahapan penelitian ini, penulis mengidentifikasi permasalahan yang timbul dikalangan masyarakat, khususnya keresahan yang terjadi pada penggunaan

media sosial. Permasalahan yang terjadi merupakan tindakan *cyberbullying* yang terjadi di media sosial. Permasalahan timbul akibat tumbuhnya perkembangan teknologi dan maraknya penggunaan media sosial tanpa adanya pengetahuan dari penggunanya. Sehingga penulis berfikir kritis untuk menyelesaikan permasalahan yang ada dengan perkembangan teknologi salah satunya analisis metode *clustering* dengan algoritma DBSCAN pengelompokan *cyberbullying* di Instagram. Penelitian ini dilakukan untuk membantuk pihak – pihak berwenang seperti Kementerian Komunikasi dan Informatika (kominfo) untuk memberikan edukasi kepada pengguna dan memberikan informasi kepada orang tua untuk mengawasi anaknya dalam penggunaan media sosial, sehingga tindakan *cyberbullying* di media sosial khususnya Instagram dapat di atasi dan diminimalisirkan tindakan tersebut.

### 3.3 Studi Literatur

Studi literatur pada penelitian ini dilakukan untuk mencari referensi dan rujukan yang berkaitan dengan penelitian ini, gunanya agar untuk mengembangkan dalam penelitian ini. Sehingga dalam penelitian ini dapat memberikan kontribusi yang baik dalam berkelanjutannya objek pembahasan dalam penelitian ini. Adapun sumber-sumber yang diambil untuk dijadikan literatur pada penelitian ini adalah yang berkaitan dengan Instagram, *cyberbullyig*, *data mining*, *text-mining*, metode *clustering*, algoritma DBSCAN, *Davies Bouldin Index*(DBI), dan penelitian terdahulu.

### 3.4 Pengumpulan Data

Penelitian ini menggunakan data sekunder sebagai sumber data. Data sekunder adalah data yang diperoleh yang dimana data tersebut telah dikumpulkan atau sudah tersedia untuk dilakukannya penelitian dari data pihak lain (Trilaksono & Handayani, 2020). Pada penelitian ini, peneliti akan mengambil dataset dari sebuah situs website resmi yang telah mengeluarkan banyak dataset, yang dimana data tersebut bertujuan untuk dilakukan sebuah penelitian. Website yang digunakan peneliti adalah [www.kaggle.com](http://www.kaggle.com).

Sampel data yang akan digunakan pada penelitian penerapan metode *clustering* dengan algoritma DBSCAN pengelompokan *cyberbullying* di Instagram meliputi termasuk kalimat *bullying* dan *nonbullying*. Dengan jumlah sample data

yang digunakan dalam penelitian ini adalah sebanyak 650 data kalimat *cyberbullying*. Dimana sebanyak 322 sample data kalimat *nonbullying* dan sebanyak 328 sample data kalimat *bullying*. Data yang digunakan adalah kalimat *cyberbullying* yang termasuk kalimat *bullying* dan *nonbullying*.

### 3.5 Pengolahan Data

Pada tahap ini, penulis akan melakukan pengolahan data yang telah diperoleh dengan menggunakan Google Colab menggunakan tahapan pada algoritma DBSCAN. Berikut penjelasan sederhana mengenai pengolahan data yang akan dilakukan, yaitu :

1. Data yang diinputkan akan melakukan proses perhitungan dengan algoritma DBSCAN pada prosedur lengkap di Persamaan (2.4).
2. Selanjutnya data akan menghitung jumlah kluster untuk menentukan nilai terbaik dengan menggunakan *Davies Bouldin Index*, pada prosedur di Persamaan (2.5).
3. Setelah data di proses dengan melakukan perhitungan sesuai dengan prosedur pada Persamaan (2.4) dan (2.5) maka data dapat di implementasikan dengan Google Colab untuk menentukan algoritma DBSCAN dapat digunakan untuk proses pengelompokan data kalimat *cyberbullying* di Instagram.

Berikut merupakan contoh kalimat dan diberikan label bernilai(0 & 1) yang akan di jadikan sebagai contoh untuk perhitungan manual pada dataset *cyberbullying* dengan algoritma DBSCAN , yaitu :

Tabel 3. 1 Contoh kalimat *Cyberbullying* diberikan label

No.	Komentar	Label
1.	“Kesini cmn mo liat komen aja”	0
2.	“Ka yaya kalo ragu tetap secantik itu yah:)”	0
3.	“Raisa pake baju merah sumpah cantik banget”	0
4.	“Jelek aja tukang selingkuh lu buriq”	1

5.	“Dah jelek semangattt pula WKWK”	1
6.	“Dikasih bidadari malah milih sampah”	1
7.	“Muka lu keriput kek aki aki”	1
8.	“Dulu bahagia sekarang banyak utang”	1
9.	“Berdiri di kaki sendiri lebih baik ”	0
10.	“G bole melet dosa rik”	0

Berdasarkan table 3.1 proses pelabelan pada kalimat *cyberbullying* dilakukan untuk mempermudah dalam proses klasifikasi, pada label yang diberikan nilai 0 maka label tersebut untuk menandakan bahwa nilai 0 merupakan kalimat *nonbullying*, sedangkan label yang diberikan nilai 1 merupakan kalimat *bullying*. Berikut merupakan contoh proses perhitungan *text preprocessing*

### 1. Case Folding

Pada tahap *case folding* ini, dokumen yang berisi setiap karakter akan dilakukan perubahan menjadi huruf kecil. Pada setiap tahap *preprocessing* dilakukan proses penghilangan spasi yang berlebihan. Pada tahap ini, semua perubahan terjadinya proses pada dokumen menjadi huruf kecil, seperti huruf P pada “Pengendalian”, “Perbandingan”, dan “Peramalan” akan diubah menjadi “pengendalian”, “perbandingan” dan “peramalan”.

Tabel 3. 2 Proses *case folding* pada kalimat *Cyberbullying*

No.	Text Asli	Hasil <i>Case Folding</i>
1.	“Kesini cmn mo liat komen aja”	kesini cmn mo liat komen aja
2.	“Ka yaya kalo ragu tetap secantik itu yah:)”	ka yaya kalo ragu tetap secantik itu yah

3.	“Raisa pake baju merah sumpah cantik banget”	raisa pake baju merah sumpah cantik banget
4.	“Jelek aja tukang selingkuh lu buriq”	jelek aja tukang selingkuh lu buriq
5.	“Dah jelek semangattt pula WKWK”	dah jelek semangattt pula wkwk
6.	“Dikasih bidadari malah milih sampah”	dikasih bidadari malah milih sampah
7.	“Muka lu keriput kek aki aki”	muka lu keriput kek aki aki
8.	“Dulu bahagia sekarang banyak utang”	dulu bahagia sekarang banyak utang
9.	“Berdiri di kaki sendiri lebih baik”	berdiri di kaki sendiri lebih baik
10.	“G bole melet dosa rik”	g bole melet dosa rik

## 2. Tokonezing

Pada proses tahap *tokonezing* ini, akan dilakukan sebuah proses penghilangan angka, tanda baca dan karakter lain kecuali huruf alphabet, karena pada huruf alphabet karakternya merupakan bukan karakter yang memiliki pengaruh terhadap pemrosesan teks.

Tabel 3.3 Proses *tokonezing* pada kalimat *cyberbullying*

No.	Teks Awal	Hasil <i>Tokenezing</i>
1	Kesini cmn mo liat komen aja	Kesini cmn mo liat komen aja



2	Ka yaya kalo ragu tetap secantik itu yah:)	Ka Yaya kalo ragu tetap secantik itu yah:)
3	Raisa pake baju merah sumpah cantik banget	Raisa pake baju merah sumpah cantik banget
4	Jelek aja tukang selingkuh lu buriq	Jelek aja tukang selingkuh lu buriq
5	Dah jelek semangattt pula WKWK	Dah jelek semangattt pula WKWK
6	Dikasih bidadari malah milih sampah	Dikasih bidadari malah milih

		sampah
7	Muka lu keriput kek aki aki	Muka lu keriput kek aki aki
8	Dulu bahagia sekarang banyak utang	Dulu bahagia sekarang banyak utang
9	Berdiri di kaki sendiri lebih baik	Berdiri di kaki sendiri lebih baik
10	G bole melet dosa rik	G bole melet dosa rik

### 3. *Filtering*

Pada proses *filtering*, tahap ini dilakukan suatu proses untuk menghilangkan *stopword*. *Stopword* adalah kosakata yang tidak termasuk kata unik atau memiliki ciri di dalam isi dokumen atau secara signifikan tidak menyampaikan pesan. Seperti pada kata “menggunakan”, “analisis”, “dan”, “untuk”, “kasus” dan lain – lain.

Tabel 3.4 Proses *filtering* pada kalimat *cyberbullying*

No	Hasil <i>Tokenizing</i>	Hasil <i>Filtering</i>
1	Kesini cmn mo liat komen aja	Kesini liat komen
2	Ka Yaya kalo ragu tetap secantik itu yah:)	Yaya ragu tetap secantik
3	Raisa pake baju merah sumpah cantik banget	Raisa baju merah cantik
4	Jelek aja tukang selingkuh lu buriq	Jelek tukang selingkuh buriq
5	Dah jelek	Jelek semangat

	semangatt pula WKWK	
6	Dikasih bidadari malah milih sampah	Dikasih Bidadari milih sampah
7	Muka lu keriput kek aki aki	Muka Keriput Aki
8	Dulu bahagia sekarang banyak utang	Bahagia banyak utang
9	Berdiri di kaki sendiri lebih baik	Berdiri kaki sendiri baik
10	G bole melet dosa rik	Melet Dosa

#### 4. Stemming

Pada proses *stemming*, tahap ini dilakukan suatu proses untuk melakukan perubahan kata menjadi kata dasar dari setiap kata hasil yang di dapat pada proses *filtering* dengan cara menghilangkan kata yang bersifat imbuhan, agar database tidak merasa kesulitan untuk melakukan proses *stemming*.

Tabel 3.5 Proses *stemming* pada kalimat *cyberbullying*

No	Hasil <i>Filtering</i>	Hasil <i>Stemming</i>
1	Kesini liat komen	Sini liat komen
2	Yaya ragu tetap secantik	Yaya ragu tetap cantik
3	Raisa baju merah cantik	Raisa baju merah cantik
4	Jelek tukang selingkuh buriq	Jelek tukang selingkuh buriq
5	Jelek semangat	Jelek semangat
6	Dikasih Bidadari milih sampah	Kasih Bidadari milih sampah
7	Muka	Muka



	Keriput Aki	Keriput Aki
8	Bahagia banyak utang	Bahagia banyak utang
9	Berdiri kaki sendiri baik	Diri kaki sendiri baik
10	Melet Dosa	Melet Dosa

Berikut merupakan proses pembobotan, nilai pada setiap dokumen dengan menggunakan perhitungan Term Frequency:

Tabel 3.6 Tabel Perhitungan *Term Frequency*

Text Processing	TF										df	d/df
	D1	D2	D3	D4	D5	D6	D7	D8	D9	D10		
Sini	1	0	0	0	0	0	0	0	0	0	1	10
Liat	1	0	0	0	0	0	0	0	0	0	1	10
Komen	1	0	0	0	0	0	0	0	0	0	1	10
Yaya	0	1	0	0	0	0	0	0	0	0	1	10
Ragu	0	1	0	0	0	0	0	0	0	0	1	10
Tetap	0	1	0	0	0	0	0	0	0	0	1	10
Cantik	0	1	1	0	0	0	0	0	0	0	2	5
Raisa	0	0	1	0	0	0	0	0	0	0	1	10

Baju	0	0	1	0	0	0	0	0	0	0	1	10
Merah	0	0	1	0	0	0	0	0	0	0	1	10
Jelek	0	0	0	1	1	0	0	0	0	0	2	5
Tukang	0	0	0	1	0	0	0	0	0	0	1	10
Selingkuh	0	0	0	1	0	0	0	0	0	0	1	10
Buriq	0	0	0	1	0	0	0	0	0	0	1	10
Semangat	0	0	0	0	1	0	0	0	0	0	1	10
Kasih	0	0	0	0	0	1	0	0	0	0	1	10
Bidadari	0	0	0	0	0	1	0	0	0	0	1	10
Milih	0	0	0	0	0	1	0	0	0	0	1	10
Samapah	0	0	0	0	0	1	0	0	0	0	1	10
Muka	0	0	0	0	0	0	1	0	0	0	1	10
Keriput	0	0	0	0	0	0	1	0	0	0	1	10
Aki	0	0	0	0	0	0	1	0	0	0	1	10
Bahagia	0	0	0	0	0	0	0	1	0	0	1	10
Banyak	0	0	0	0	0	0	0	1	0	0	1	10
Utang	0	0	0	0	0	0	0	1	0	0	1	10
Diri	0	0	0	0	0	0	0	0	1	0	1	10
Kaki	0	0	0	0	0	0	0	0	1	0	1	10
Sendiri	0	0	0	0	0	0	0	0	1	0	1	10
Baik	0	0	0	0	0	0	0	0	1	0	1	10

Melet	0	0	0	0	0	0	0	0	0	0	1	1	10
Dosa	0	0	0	0	0	0	0	0	0	0	1	1	10

Berikut merupakan proses pembobotan, nilai pada setiap dokumen dengan menggunakan perhitungan TF- IDF (*Term Frequency – Invers Document Frequency*):

Tabel 3.7 Tabel Perhitungan TF- IDF

Text Processing	IDF log10 (D/DF)	W ( TF X IDF )										
		D1	D2	D3	D4	D5	D6	D7	D8	D9	D10	
Sini	1	1	0	0	0	0	0	0	0	0	0	0
Liat	1	1	0	0	0	0	0	0	0	0	0	0
Komen	1	1	0	0	0	0	0	0	0	0	0	0
Yaya	1	0	1	0	0	0	0	0	0	0	0	0
Ragu	1	0	1	0	0	0	0	0	0	0	0	0
Tetap	1	0	1	0	0	0	0	0	0	0	0	0
Cantik	0,698	0	0,698	0,698	0	0	0	0	0	0	0	0
Raisa	1	0	0	1	0	0	0	0	0	0	0	0
Baju	1	0	0	1	0	0	0	0	0	0	0	0
Merah	1	0	0	1	0	0	0	0	0	0	0	0
Jelek	0,698	0	0	0	0,698	0,698	0	0	0	0	0	0
Tukang	1	0	0	0	1	0	0	0	0	0	0	0
Selingkuh	1	0	0	0	1	0	0	0	0	0	0	0

Buriq	1	0	0	0	1	0	0	0	0	0	0
Semangat	1	0	0	0	0	1	0	0	0	0	0
Kasih	1	0	0	0	0	0	1	0	0	0	0
Bidadari	1	0	0	0	0	0	1	0	0	0	0
Milih	1	0	0	0	0	0	1	0	0	0	0
Samapah	1	0	0	0	0	0	1	0	0	0	0
Muka	1	0	0	0	0	0	0	1	0	0	0
Keriput	1	0	0	0	0	0	0	1	0	0	0
Aki	1	0	0	0	0	0	0	1	0	0	0
Bahagia	1	0	0	0	0	0	0	0	1	0	0
Banyak	1	0	0	0	0	0	0	0	1	0	0
Utang	1	0	0	0	0	0	0	0	1	0	0
Diri	1	0	0	0	0	0	0	0	0	1	0
Kaki	1	0	0	0	0	0	0	0	0	1	0
Sendiri	1	0	0	0	0	0	0	0	0	1	0
Baik	1	0	0	0	0	0	0	0	0	1	0
Melet	1	0	0	0	0	0	0	0	0	0	1
Dosa	1	0	0	0	0	0	0	0	0	0	1

### 3.6 Pengujian Data

Setelah data diolah dengan melakukan serangkaian tahapan text preprocessing dan melakukan pembobotan nilai pada setiap kata yang ada didalam document. Maka, proses selanjutnya adalah data akan dilakukan sebuah proses

perhitungan algoritma dengan menentukan nilai epsilon dengan nilai minimum berdasarkan kata yang ada dilama document. Setelah mendapatkan hasil dari perhitungan algoritma DBSCAN, data akan dilakukan pengujian untuk mengetahui hasil data yang telah dilakukan pengolahan. Dalam pengujian data ini peneliti akan menggunakan DBI untuk menguji tingkat pengelompokan yang paling baik dengan preposisi nilai 0 dan 1. Semakin dekata kata yang memiliki nilai 0, maka kata tersebut merupakan kalimat *Nonbullying*, sedangkan kata yang mendekati nilai 1, maka kata tersebut merupakan kalimat *bullying*. Berikut merupakan perhitungan algoritma DBSCAN:

Menghitung jarak titik *core point* dengan *point* yang lain dengan jarak *Euclidean*. Berikut ini merupakan rumus jarak *Euclidean* yang ditunjukkan pada Persamaan (2.4).

$$\text{Jarak Euclidean} = \sqrt{\sum_{j=1}^m (x_{ji} - y_{jp})^2},$$

1. Jarak antara d1 dan d2:

$$X = (1,1,1)$$

$$Y = (1,1,1,0.698)$$

$$\text{Jarak Euclidean} (d1,d2) = \sqrt{(1-1)^2} + \sqrt{(1-1)^2} + \sqrt{(1-1)^2} + \sqrt{(0-0.698)^2}$$

$$\text{Jarak Euclidean} (d1,d2) = 0.698$$

2. Jarak antara d2 dan d3:

$$X = (1,1,1,0.698)$$

$$Y = (0.698,1,1,1)$$

$$\text{Jarak Euclidean} (d2,d3) = \sqrt{(1-0.698)^2} + \sqrt{(1-1)^2} + \sqrt{(1-1)^2} + \sqrt{(1-1)^2}$$

$$\text{Jarak Euclidean} (d2,d3) = 0.602$$

3. Jarak antara d3 dan d4:



$$X = (0.698, 1, 1, 1)$$

$$Y = (0.698, 1, 1, 1)$$

$$\text{Jarak Euclidean (d3,d4)} = \sqrt{(0.698 - 0.698)^2} + \sqrt{(1 - 1)^2} + \sqrt{(1 - 1)^2} + \sqrt{(1 - 1)^2}$$

$$\text{Jarak Euclidean (d3,d4)} = 0$$

4. Jarak antara d4 dan d5:

$$X = (0.698, 1, 1, 1)$$

$$Y = (0.698, 1)$$

$$\text{Jarak Euclidean (d4,d5)} = \sqrt{(0.698 - 0.698)^2} + \sqrt{(1 - 1)^2}$$

$$\text{Jarak Euclidean (d4,d5)} = 0$$

5. Jarak antara d5 dan d6:

$$X = (0.698, 1)$$

$$Y = (1, 1, 1, 1)$$

$$\text{Jarak Euclidean (d5,d6)} = \sqrt{(0.698 - 1)^2} + \sqrt{(1 - 1)^2}$$

$$\text{Jarak Euclidean (d5,d6)} = 0.302$$

6. Jarak antara d6 dan d7:

$$X = (1, 1, 1, 1)$$

$$Y = (1, 1, 1)$$

$$\text{Jarak Euclidean (d6,d7)} = \sqrt{(1 - 1)^2} + \sqrt{(1 - 1)^2} + \sqrt{(1 - 1)^2}$$

$$\text{Jarak Euclidean (d6,d7)} = 0$$

7. Jarak antara d7 dan d8:

$$X = (1, 1, 1)$$

$$Y = (1,1,1)$$

$$\text{Jarak Euclidean (d7,d8)} = \sqrt{(1-1)^2} + \sqrt{(1-1)^2} + \sqrt{(1-1)^2}$$

$$\text{Jarak Euclidean (d7,d8)} = 0$$

8. Jarak antara d8 dan d9:

$$X = (1,1,1)$$

$$Y = (1,1,1,1)$$

$$\text{Jarak Euclidean (d8,d9)} = \sqrt{(1-1)^2} + \sqrt{(1-1)^2} + \sqrt{(1-1)^2} + \sqrt{(1-1)^2}$$

$$\text{Jarak Euclidean (d8,d9)} = 0$$

9. Jarak antara d9 dan d10:

$$X = (1,1,1)$$

$$Y = (1,1)$$

$$\text{Jarak Euclidean (d9,d10)} = \sqrt{(1-1)^2} + \sqrt{(1-1)^2} + \sqrt{(1-1)^2}$$

$$\text{Jarak Euclidean (d9,d10)} = 0$$

Setelah melakukan perhitungan dengan menentukan jarak *Euclidean*. Maka dapat dilakukan perhitungan kluster algoritma DBSCAN. Untuk mengelompokkan data dengan algoritma DBSCAN, peneliti menggunakan hasil dari perhitungan jarak *Euclidean* yang telah dilakukan sebelumnya. Selanjutnya, menentukan titik inti (core points) dan membentuk kluster.

Dengan ketentuan nilai epsilon ( $\epsilon$ ) = 0,4 da nilai minum points(MinPts) = 2

- Jumlah tetangga d1:1 (d2)
- Jumlah tetangga d2:2(d1,d3)
- Jumlah tetangga d3:3(d2,d4,d5)
- Jumlah tetangga d4:3(d3,d5,d6)
- Jumlah tetangga d5:3(d3,d4,d6)

- Jumlah tetangga d6:2(d4,d5)
- Jumlah tetangga d7:1(d8)
- Jumlah tetangga d8:1(d7)
- Jumlah tetangga d9:1(d10)
- Jumlah tetangga d10:1(d9)

Kluster dapat terbentuk dari titik inti yang saling terhubung dalam jarak  $\epsilon$ . Sehingga hasil klustering berdasarkan nilai parameter  $\epsilon = 0,4$  dan  $\text{MinPts} = 2$ , menghasilkan 3 kluster dari titik inti.

- C1 : {d1,d2,d3,d4,d5,d6}
- C2 : {d7,d8}
- C3 : {d9,d10}

Dari 3 hasil kluster yang diperoleh, tidak ditemukan titik noise. Karena titik noise merupakan titik yang bukan inti dan tidak termasuk dalam kluster. Semua titik yang tidak termasuk dalam kluster dan tidak menjadi inti adalah titik noise.

### DBI

Untuk mencari nilai terbaik, peneliti melakukan perhitungan pusat (centroid) dari setiap kluster yang didapatkan dari perhitungan jarak *Euclidean* dan algoritma DBSCAN. Pusat kluster digunakan dalam perhitungan SSW dan SSB.

- Pusat Kluster 1 (C1) : {d1,d2,d3,d4,d5,d6}
- Pusat Kluster 2 (C2) : {d7,d8}
- Pusat Kluster 3 (C3) : {d9,d10}

Menghitung jarak *Euclidean* antara setiap titik data dan pusat kluster terdekat, proses ini dilakukan untuk mempermudah dalam menghitung SSW.

### Untuk kluster 1(C1):

- Jarak antara d1 dan C1 :  $\sqrt{(1 - 0.833)^2} + \sqrt{(1 - 0.833)^2} + \sqrt{(1 - 0.833)^2} = 0.288$
- Jarak antara d2 dan C1 :  $\sqrt{(1 - 0.833)^2} + \sqrt{(1 - 0.833)^2} + \sqrt{(1 - 0.833)^2} = 0.288$

- c) Jarak antara d3 dan C1:  $\sqrt{(0.833 - 0.833)^2} + \sqrt{(1 - 0.833)^2} + \sqrt{(1 - 0.833)^2} = 0.167$
- d) Jarak antara d4 dan C1:  $\sqrt{(0.833 - 0.833)^2} + \sqrt{(1 - 0.833)^2} + \sqrt{(1 - 0.833)^2} = 0.167$
- e) Jarak antara d5 dan C1:  $\sqrt{(0.833 - 0.833)^2} + \sqrt{(1 - 0.833)^2} + \sqrt{(1 - 0.833)^2} = 0.167$
- f) Jarak antara d6 dan C1 :  $\sqrt{(1 - 0.833)^2} + \sqrt{(1 - 0.833)^2} + \sqrt{(1 - 0.833)^2} = 0.288$

**Untuk kluster 2(C2):**

- a) Jarak antara d7 dan C2 :  $\sqrt{(1 - 0.5)^2} + \sqrt{(1 - 0.5)^2} = 0.707$
- b) Jarak antara d8 dan C2 :  $\sqrt{(1 - 0.5)^2} + \sqrt{(1 - 0.5)^2} = 0.707$

**Untuk kluster 3 (C3):**

- a) Jarak antara d9 dan C3 :  $\sqrt{(1 - 1)^2} + \sqrt{(1 - 1)^2} = 0$
- b) Jarak antara d10 dan C3 :  $\sqrt{(1 - 1)^2} + \sqrt{(1 - 1)^2} = 0$

**Menghitung SSW**

Rumus SSW dapat dilihat pada persamaan 2.5. Berikut merupakan perhitungannya:

**Kluster 1 (C1) :**

Sebelumnya telah menghitung jarak antara titik data dalam kluster 1(d1,d2,d3,d4,d5,d6) dan pusat kluster :

- a) Jarak antara d1 dan C1 : 0.288
- b) Jarak antara d2 dan C1 : 0.288
- c) Jarak antara d3 dan C1 : 0.167
- d) Jarak antara d4 dan C1 : 0.167
- e) Jarak antara d5 dan C1 : 0.167
- f) Jarak antara d6 dan C1 : 0.288

Untuk menghitung SSW jarak – jarak kluster akan di kuadratkan, dijumlahkan, dan kemudian dibagi dengan jumlah titik data dalam Kluster 1(m1=6):

$$SSW_1 = \frac{1}{6} (0.288^2 + 0.288^2 + 0.167^2 + 0.167^2 + 0.167^2 + 0.288^2)$$

$$SSW_1 = \frac{1}{6} (0.082944 + 0.082944 + 0.027889 + 0.027889 + 0.027889 + 0.082944)$$

$$SSW_1 = \frac{1}{6} \times 0.332614$$

$$SSW_1 = 0.055436$$

Kluster 2(C2):

Selanjutnya menghitung jarak antara data dalam kluster 2 (d7,d8) dan pusat kluster C2 sebelumnya:

a) Jarak antara d7 dan C2 : 0.707

b) Jarak antara d8 dan C2 : 0.707

Untuk menghitung SSW jarak – jarak kluster akan di kuadratkan, dijumlahkan, dan kemudian dibagi dengan jumlah titik data dalam Kluster 2(m2 = 2):

$$SSW_2 = \frac{1}{2} (0.707^2 + 0.707^2)$$

$$SSW_2 = \frac{1}{2} (0.499849 + 0.499849)$$

$$SSW_2 = \frac{1}{2} \times 0.999698$$

$$SSW_2 = 0.499849$$

Kluster 3 (C3) :

Kluster 3 hanya memiliki dua titik data (d9,d10) yang memiliki jarak nol dengan pusat kluster C3, sehingga SSW3 untuk kluster 3 adalah 0.

Maka telah selesai melakukan perhitungan SSW untuk setiap kluster:

a) SSW untuk kluster 1 (SSW<sub>1</sub>)= 0.055436

b) SSW untuk kluster 2 (SSW<sub>2</sub>)= 0.499849

c) SSW untuk kluster 3 (SSW<sub>3</sub>)= 0

### Menghitung SSB



Untuk melakukan perhitungan  $SSB$  dapat dilihat pada rumus persamaan 2.6:

Kluster 1 (C1) dan Kluster 2 (C2):

Pada proses ini akan menghitung jarak antara pusat kluster C1 dan C2 :

$$SSB_{1,2} = (C1 - C2)^2$$

$$SSB_{1,2} = (0.833 - 0.5)^2$$

$$SSB_{1,2} = 0.110889$$

Kluster 1 (C1) dan Kluster 3 (C3):

Pada proses ini akan menghitung jarak antara pusat kluster C1 dan C3 :

$$SSB_{1,3} = (C1 - C3)^2$$

$$SSB_{1,3} = (0.833 - 1)^2$$

$$SSB_{1,3} = 0.027889$$

Kluster 1 (C1) dan Kluster 2 (C2):

Pada proses ini akan menghitung jarak antara pusat kluster C2 dan C3 :

$$SSB_{2,3} = (C2 - C3)^2$$

$$SSB_{2,3} = (0.5 - 1)^2$$

$$SSB_{2,3} = 0.25$$

Menghitung  $SSB$  antara kluster telah selesai dilakukan.  $SSB$  dapat dihitung untuk pasangan kluster yang memiliki lebih dari satu titik data, seperti dalam kasus kluster 1 dan kluster 2. Hasil nilai dari  $SSB$  ini dapat digunakan untuk perhitungan selanjutnya dalam melakukan perhitungan *Davies Bouldin Index* (DBI). Setelah  $SSW$  dan  $SSB$  selesai dihitung, maka proses selanjutnya adalah menghitung rasio  $R_{i,j}$  antara setiap pasangan kluster. Untuk melakukan perhitungan dapat dilihat rumus pada persamaan 2.7 :

1. Rasio pada  $R_{1,2}$  antara kluster 1 (C1) dan Kluster 2 (C2):

$$R_{1,2} = \frac{SSW_1 + SSW_2}{SSB_{12}}$$

$$R_{1,2} = \frac{0.055436 + 0.499849}{0.110889}$$

$$R_{1,2} = 4,551$$

2. Rasio pada  $R_{1,3}$  antara kluster 1 (C1) dan Kluster 3 (C3):

$$R_{1,3} = \frac{SSW_1 + SSW_3}{SSB_{13}}$$

$$R_{1,3} = \frac{0.055436 + 0}{0.027889}$$

$$R_{1,3} = 1,989$$

3. Rasio pada  $R_{2,3}$  antara kluster 2 (2) dan Kluster 3 (C3):

$$R_{2,3} = \frac{SSW_2 + SSW_3}{SSB_{23}}$$

$$R_{2,3} = \frac{0,499849 + 0}{0,25}$$

$$R_{2,3} = 1,9994$$

Sekarang telah menghitung rasio  $R_{i,j}$  untuk setiap pasangan kluster. Rasio – rasio ini akan digunakan dalam perhitungan *Davies Bouldin Index*(DBI). Untuk melakukan perhitungan DBI, dapat dilihat pada rumus persamaan 2.8:

Dari perhitungan diatas, memiliki 3 hasil kluster yaitu C1, C2, C3.

1. DBI untuk kluster 1 (C1) :

$$DBI_{c1} = \frac{1}{2} (\max(R_{1,2}, R_{1,3}))$$

$$DBI_{c1} = \frac{1}{2} (\max(4.551, 1.989))$$

$$DBI_{c1} = 3,27$$

2. DBI untuk kluster 2 (C2) :

$$DBI_{c2} = \frac{1}{2} (\max(R_{1,2}, R_{1,3}))$$

$$DBI_{c2} = \frac{1}{2} (\max(4.551, 1.994))$$

$$DBI_{c2} = 3,2757$$

3. DBI untuk kluster 3 (C3) :

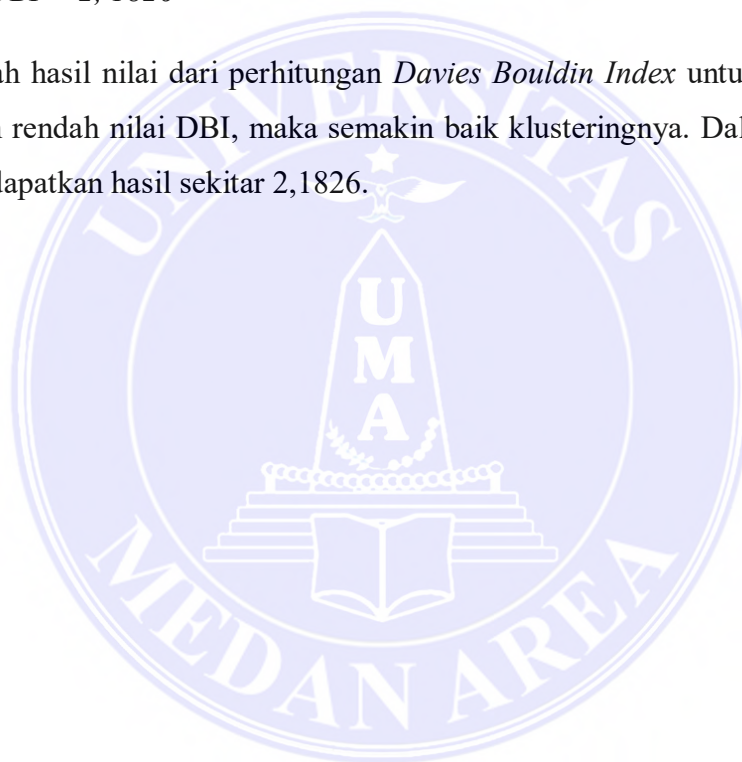
Karena kluster 3 hanya memiliki satu titik data, maka DBI untuk kluster 3 adalah 0. Setelah menghitung DBI untuk setiap kluster. Selanjutnya, akan menghitung nilai rata – rata DBI untuk semua kluster:

$$DBI = \frac{1}{3} (DBI_{c1} + DBI_{c2} + DBI_{c3})$$

$$DBI = \frac{1}{3} (2,27 + 3,2757 + 0)$$

$$DBI = 2,1826$$

Ini adalah hasil nilai dari perhitungan *Davies Bouldin Index* untuk klustering ini. Semakin rendah nilai DBI, maka semakin baik klusteringnya. Dalam perhitungan ini mendapatkan hasil sekitar 2,1826.



## BAB V KESIMPULAN DAN SARAN

### 5.1 Kesimpulan

Berdasarkan hasil dari pembahasan dan uji coba yang telah dilakukan, maka penelitian ini dapat disimpulkan :

1. Dengan menggunakan algoritma DBSCAN telah berhasil untuk mengelompokkan kalimat *Cyberbullying* di Instagram.
2. Proses Text Preprocessing dapat dilakukan dengan memperoleh hasil *case folding, tokenizing, filtering*, serta menghitung jumlah frekuensi kata yang sering muncul didalam document.
3. Pembobotan nilai pada setiap kata didalam dokument berhasil dilakukan dengan proses *term weightend* atau proses TF/IDF.
4. Dengan menggunakan algoritma DBSCAN dengan menentukan nilai  $\epsilon=0,5$  dan nilai  $\text{minpointnya} = 3$  maka mendapatkan nilai hasil -1 cluster.
5. Pengujian pada algoritma DBSCAN menggunakan *Davies Boilding Index* memperoleh nilai 1.

### 5.2 Saran

Saran untuk pengembangan pada penelitian ini dimasa yang akan datang sebagai berikut:

1. Diharapkan penelitian berikutnya bisa menggunakan data primer untuk melakukan penelitian
2. Dapat menambah kategori yang digunakan seperti tingkat *Bullyingnya*.
3. Sebaiknya penelitian berikutnya dapat menggunakan perbandingan algoritma, karena diharapkan dapat untuk mencari hasil yang terbaik.

## DAFTAR PUSTAKA

- Alamsyah, D. Q., Sudarno, S., & Kartikasari, P. (2022). Pengelompokan Tweets Pada Akun Twitter Tokopedia Menggunakan *Algoritma Density Based Spatial Clustering Of Applications With Noise*. *Jurnal Gaussian*, 11(1), 118–129.
- Apifah, U. (2021). Vidgram Sebagai Sarana Media Dakwah (Study Deskriptif Pada Akun Instagram@ Xkwavers). UIN RADEN INTAN LAMPUNG.
- Aziza, A., Juairiah, J., & Hasanah, R. (2022). *Deteksi Muatan Paham Radikalisme untuk Klasifikasi Konten Radikal pada Website Dakwah Islam Indonesia Menggunakan Metode Web Scrawler dan Text Mining*.
- Badruttamam, A., Sudarno, S., & Di Asih, I. M. (2020). Penerapan Analisis Kluster K-Modes Dengan Validasi Davies Bouldin Index Dalam Menentukan Karakteristik Kanal Youtube di Indonesia (Studi Kasus: 250 Kanal Youtube Indonesia Teratas Menurut Socialblade). *Jurnal Gaussian*, 9(3), 263–272.
- Defiyanti, S. (2017). Integrasi Metode Clustering dan Klasifikasi untuk Data Numerik. *Citee, No. July*, 256–261.
- Dinata, R. K., Safwandi, S., Hasdyna, N., & Mahendra, R. (2020). Kombinasi Algoritma Brute Force dan Stemming pada Sistem Pencarian Mashdar. *CESS (Journal of Computer Engineering, System and Science)*, 5(2), 273–278.
- Fauziah, S., Sulistyowati, D. N., & Asra, T. (2019). Optimasi Algoritma Vector Space Model Dengan Algoritma K-Nearest Neighbour Pada Pencarian Judul Artikel Jurnal. *Jurnal Pilar Nusa Mandiri*, 15(1), 21–26.
- Febriani, S., Candra, I., & Nastasia, K. (2021). Hubungan Antara Intimate Friendship Dengan Self Disclosure Pada Siswa Kelas XI SMA N 4 Kota Padang Pengguna Media Sosial Instagram. *Psyche 165 Journal*, 130–138.
- Fitriani, K., Isbandi, I., & Amaliyah, A. (2021). PERANCANGAN SISTEM MANAJEMEN DOKUMEN DENGAN MENGGUNAKAN METODE TEXT MINING DI KANTOR KELURAHAN SEKEJATI. *Telematika*, 3(1), 45–59.



- Fitriani, Y. (2021). Pemanfaatan media sosial sebagai media penyajian konten edukasi atau pembelajaran digital. *JISAMAR (Journal of Information System, Applied, Management, Accounting and Research)*, 5(4), 1006–1013.
- Herdiani, A., Asror, I., & others. (2019). Klasterisasi Tweet Terkait Dengan Pemilihan Presiden 2019 Menggunakan Ontology-based Concept Weighting Dan Dbscan. *EProceedings of Engineering*, 6(2).
- Hermawan, L., & Ismiati, M. B. (2020). Pembelajaran text preprocessing berbasis simulator untuk mata kuliah information retrieval. *Jurnal Transformatika*, 17(2), 188–199.
- Hidayah, N., Sahibu, S., & others. (2021). Algoritma Multinomial Navie Bayes Untuk Klasifikasi Sentimen Pemerintah Terhadap Penanganan Covid-19 Menggunakan Data Twitter. *Jurnal RESTI (Rekayasa Sistem Dan Teknologi Informasi)*, 5(4), 820–826.
- Irfan, I. (2022). *Cyberbullying: Varian Perundungan Terhadap Anak Berbasis Digital*. *Jurnal Litigasi Amsir*, 10(1), 112–117.
- Isnarwaty, D. P., & Irhamah, I. (2020). *Text Clustering* pada Akun TWITTER Layanan Ekspedisi JNE, J&T, dan Pos Indonesia Menggunakan Metode *Density-Based Spatial Clustering of Applications with Noise* (DBSCAN) dan K-Means. *Jurnal Sains Dan Seni ITS*, 8(2), D137--D144.
- Jalal, N. M., Idris, M., & Muliana, M. (2021). Faktor-faktor *cyberbullying* pada remaja. *IKRA-ITH HUMANIORA: Jurnal Sosial Dan Humaniora*, 5(2), 1–9.
- Jumeilah, F. S., & others. (2017). Penerapan Support Vector Machine (SVM) untuk Pengkategorian Penelitian. *Jurnal RESTI (Rekayasa Sistem Dan Teknologi Informasi)*, 1(1), 19–25.
- Kartikasari, M. D. (2021). *Self-Organizing Map* Menggunakan *Davies-Bouldin Index* dalam Pengelompokan Wilayah Indonesia Berdasarkan Konsumsi Pangan. *Jambura Journal of Mathematics*, 3(2), 187–196.
- Khaira, U., Johanda, R., Utomo, P. E. P., & Suratno, T. (2020). Sentiment analysis

of cyberbullying on twitter using SentiStrength. *Indonesian Journal of Artificial Intelligence and Data Mining*, 3(1), 21–27.

Kharis, S. A. A., & Zili, A. H. A. (2022). Learning Analytics dan Educational Data Mining pada Data Pendidikan. *JURNAL RISET PEMBELAJARAN MATEMATIKA SEKOLAH*, 6(1), 12–20.

Lukiani, E. R. M. L., Rizka, A. N., Afandi, T. Y., Arifin, Z., Surindra, B., Irmayanti, E., & Prastyaningtyas, E. W. (2021). Peran Instagram dalam Membentuk Perilaku Konsumsi Remaja. *Jurnal ABDINUS: Jurnal Pengabdian Nusantara*, 5(1), 97–104.

MAEMUNAH, M. (2020). *Program Pengabdian Kepada Masyarakat-Pelatihan Pencegahan Bullying Bagi Organisasi Kemahasiswaan Di Universitas Muhammadiyah Mataram*.

Maladi, M. A. (2020). Analisis Isi Cyberbullying Kasus Penganiayaan Anak Dibawah Umur Pada Followers Instagram Di Kota Makassar. Universitas Hasanuddin.

Manurung, E., & Hasugian, P. S. (2019). Data Mining Tingkat Pesanan Inventaris Kantor Menggunakan Algoritma Apriori Pada Kepolisian Daerah Sumatera Utara. *Journal Of Informatic Pelita Nusantara*, 4(2), 8–13.

Mardi, Y. (2017). Data Mining: Klasifikasi Menggunakan Algoritma C4. 5. *Jurnal Edik Informatika Penelitian Bidang Komputer Sains Dan Pendidikan Informatika*, 2(2), 213–219.

Maulana, G. G., & others. (2017). Pembelajaran Dasar Algoritma Dan Pemrograman Menggunakan El-Goritma Berbasis Web. *J. Tek. Mesin*, 6(2), 69–73.

Mustafa, M. S., & Simpen, I. W. (2019). Implementasi Algoritma *K-Nearest Neighbor* (KNN) Untuk Memprediksi Pasien Terkena Penyakit Diabetes Pada Puskesmas Manyampa Kabupaten Bulukumba. *SISITI: Seminar Ilmiah Sistem Informasi Dan Teknologi Informasi*, 8(1).

- Mustaqimah, N., & Rahmah, A. H. (2022). Peran Digital Influencer dalam Memberikan Edukasi Pemilihan Produk Kecantikan pada Akun Instagram@dr. richard\ lee. *Journal of Communication and Society*, 1(01), 1–13.
- Muzdalifah, F., & Putri, T. T. (2019). Pengaruh keterlibatan ayah terhadap cyberbullying remaja pengguna instagram. *Jurnal Psikogenesis*, 7(1), 1–12.
- Nabila, Z., Isnain, A. R., Permata, P., & Abidin, Z. (2021). Analisis Data Mining Untuk Clustering Kasus Covid-19 Di Provinsi Lampung Dengan Algoritma K-Means. *Jurnal Teknologi Dan Sistem Informasi*, 2(2), 100–108.
- Pembrani, E. C., & others. (2019). *Pembelajaran Relasi Non-Taksonomi Dari Dokumen Web*.
- Purnajaya, A. R., Lieputra, V., Tayanto, V., & Salim, J. G. (2022). Implementasi *Text Mining* untuk Mengetahui Opini Masyarakat Tentang Climate Change. *Journal of Information System and Technology*, 3(3), 36–44.
- Putra, M. P. R., & Wardani, K. R. N. (2020). Penerapan *Text Mining* Dalam Menganalisis Kepribadian Pengguna Media Sosial. *JUTIM (Jurnal Tek. Inform. Musirawas)*, 5(1), 63–71.
- Rachman, D. A. C., Goejantoro, R., & Amijaya, F. D. T. (2021). Implementasi *Text Mining* Pengelompokkan Dokumen Skripsi Menggunakan Metode *K-Means Clustering*. *EKSPONENSIAL*, 11(2), 167–174.
- Ramadhan, G. A., Fitriana, A., & Nurkinan, N. (2022). Pengaruh Komunikasi Pemerintahan Ridwan Kamil Dalam Membangun Personal Branding di Social Media Tiktok dan Instagram. *Jurnal Ilmiah Wahana Pendidikan*, 8(15), 255–263.
- Reliovani, R., Husein, N. N. S., Abdurrafi, K. Z., Al Husni, C. R., & Khowarizmi, M. A. (2021). *Mean Shift Algorithm to Determine Customer Segmentation in Online Store Sales*. *Gunung Djati Conference Series*, 3, 92–98.
- Sanrilla, S., Ransi, N., Tenriawaru, A. T. A., & others. (2022). Analisis Sentimen Masyarakat Terhadap Toko Online Aplikasi Shopee Menggunakan Metode

- Multinomial Naïve Bayes. *Jurnal Matematika Komputasi Dan Statistika*, 2(2), 68–75.
- Santoso, B. (2017). Perancangan aplikasi olap (*online analytical processing*) penjualan buku pada toko buku gramedia lubuklinggau dengan metode *clustering*. *Jurnal Teknologi Informasi MURA*, 9(2), 98–107.
- Sari, B. N., & Primajaya, A. (2019). Penerapan clustering DBSCAN untuk pertanian padi di kabupaten Karawang. *JIKO (Jurnal Inform. Dan Komputer) STMIK AKAKOM*, 4(1), 28–34.
- Septiani, I. W., Fauzan, A. C., & Huda, M. M. (2022). Implementasi Algoritma K-Medoids Dengan Evaluasi *Davies-Bouldin-Index* Untuk Klasterisasi Harapan Hidup Pasca Operasi Pada Pasien Penderita Kanker Paru-Paru. *Jurnal Sistem Komputer Dan Informatika (JSON)*, 3(4), 556–566.
- Setiawan, F., Setianingsih, C., & others. (2021). Clustering Pada Data Sentimen Transportasi Online Menggunakan Algoritma Dbscan. *EProceedings of Engineering*, 8(5).
- Sholih'afif, M., Muzakir, M., Al, M. I., & Al Awalaien, G. (2021). Text Mining Untuk Mengklasifikasi Judul Berita Online Studi Kasus Radar Banjarmasin Menggunakan Metode *Navie Bayes*. *Kumpul. J. Ilmu Komput*, 8(2), 199–208.
- Syahid, A., Sudana, D., & Bachari, A. D. (2022). Perundungan Siber (Cyberbullying) Bermuatan Penistaan Agama Di Media Sosial Yang Berdampak Hukum: Kajian Linguistik Forensik. *Semantik*, 11(1), 17–32.
- Umar, E., & Fatimah, T. (2018). Text Recognition Dengan Klasifikasi *Neural Network* Dan *Text-To-Speech* Pada Huruf Alphabet. *Telematika MKOM*, 9(3), 119–124.
- Utomo, D. P., & Mesran, M. (2020). Analisis Komparasi Metode Klasifikasi Data Mining dan Reduksi Atribut Pada Data Set Penyakit Jantung. *Jurnal Media Informatika Budidarma*, 4(2), 437–444.
- Virgatama, R., Suprayogi, A., & Firdaus, H. S. (2019). Identifikasi Pengaruh Sistem

Keamanan Lingkungan Terhadap Tingkat Kejahatan Pencurian Di Kota Surakarta Dengan Metode Sistem Informasi Geografis. *Jurnal Geodesi Undip*, 8(1), 398–407.

Winoto, Y. (2019). Remaja Dan Pandangannya Terhadap *Cyberbullying* Pada Media Facebook: Sebuah Kajian Tentang literasi Informasi. *Commed: Jurnal Komunikasi Dan Media*, 3(2), 121–132.





## Lampiran Hasil Plagiat

**turnitin** Similarity Report ID: oid:29477:43730956

PAPER NAME	AUTHOR
<b>cek sidang.docx</b>	<b>Budiman Sujatmiko</b>

---

WORD COUNT	CHARACTER COUNT
<b>7938 Words</b>	<b>46033 Characters</b>

PAGE COUNT	FILE SIZE
<b>56 Pages</b>	<b>474.1KB</b>

SUBMISSION DATE	REPORT DATE
<b>Sep 22, 2023 9:09 AM GMT+7</b>	<b>Sep 22, 2023 9:10 AM GMT+7</b>

---

**25% Overall Similarity**  
The combined total of all matches, including overlapping sources, for each database.

- 22% Internet database
- Crossref database
- 15% Submitted Works database
- 4% Publications database
- Crossref Posted Content database

**Excluded from Similarity Report**

- Small Matches (Less than 10 words)

Summary

## Lampiran Surat keterangan pembimbing



 **UNIVERSITAS MEDAN AREA**  
**FAKULTAS TEKNIK**

Kampus I : Jalan Kolan Nomor 1 Medan Estate/Jalan PBGI Nomor 1 ☎ (061) 7366878, 7360168, 7364348, 7366781, Fax. (061) 7366998 Medan 20223  
Kampus II : Jalan Setia Budi Nomor 79 / Jalan Gel Serayu Nomor 70 A, ☎ (061) 8225602, Fax. (061) 8226331 Medan 20122  
Website: [www.teknik.uma.ac.id](http://www.teknik.uma.ac.id) E-mail: [univ\\_medanama@uma.ac.id](mailto:univ_medanama@uma.ac.id)

---

Nomor : 422/FT.6/01.10/VI/2023 6 Juni 2023  
Lamp : -  
Hal : Perpanjang SK Pembimbing Tugas Akhir

Yth. Pembimbing Tugas Akhir  
**Andre Hasudungan Lubis, S. Ti, MSc**  
di  
Tempat

Dengan hormat,  
Schubungan telah berakhirnya waktu masa berlaku SK pembimbing nomor 27/FT.6/01.10/I/2023 tertanggal 16 Januari 2023 maka perlu diterbitkan kembali SK Pembimbing Skripsi baru atas nama mahasiswa berikut :

N a m a : Budiman Sujatmiko  
N P M : 198160018  
Jurusan : Teknik informatika

Oleh karena itu kami mengharapkan kesediaan saudara :


**Andre Hasudungan Lubis, S. Ti, MSc ( Sebagai Pembimbing )**

Adapun Tugas Akhir Skripsi berjudul :

***“Analisis Metode Clustering dengan Algoritma Density-Based Spatial Clustering of Applications with Noise (DBSCAN) Pengelompokan Cyberbullying di Instagram”***

SK Pembimbing ini berlaku selama enam bulan terhitung sejak SK ini diterbitkan. Jika proses pembimbing melebihi batas waktu yang telah ditetapkan, SK ini dapat ditinjau ulang.

Demikian kami sampaikan, atas kesediaan saudara diucapkan terima kasih.

  
**Dr. Rahmat Syah, S. Kom, M. Kom**

Lampiran surat pengantar riset



 **UNIVERSITAS MEDAN AREA**  
Kampus I : Jalan Kolam Nomor 1 Medan Estate ☎ (061) 7360168, 7366878, 7364348 📠 (061) 7368012 Medan 20223  
Kampus II : Jalan Setiabudi Nomor 79 / Jalan Sei Serayu Nomor 70 A ☎ (061) 8225602 📠 (061) 8226331 Medan 20122  
Website: www.uma.ac.id E-Mail: univ\_medanarea@uma.ac.id

---

Nomor : 628 /UMA/B/01.7/IV/2023 06 April 2023.  
Lamp. : 1 (satu) lembar.  
Hal : Penelitian dan Pengambilan Data Tugas Akhir

Kepada Yth.  
**Dekan Fakultas Teknik**  
Universitas Medan Area  
di - M e d a n

Dengan hormat, sesuai dengan surat dari Dekan Fakultas Teknik Universitas Medan Area nomor 250/FT.6/01.10/III/2023 tertanggal 31 Maret 2023, perihal Izin Penelitian dan Pengambilan Data Tugas Akhir di Laboratorium Komputer Program Studi Teknik Informatika Universitas Medan Area oleh mahasiswa sebagai berikut :

**Nama** : Budiman Sujatmiko  
**No. Pokok Mahasiswa** : 198160018  
**Program Studi** : Teknik Informatika  
**Fakultas** : Teknik

Pada prinsipnya disetujui yang bersangkutan melaksanakan pengambilan data tersebut, untuk penunjang tugas akhir dengan judul Skripsi “*Analisis Metode Clustering dengan Algoritma DBSCAN (Density – Based Spatial Clustering Application With Noise) Pengelompokan Cyberbullying di Instagram*”.

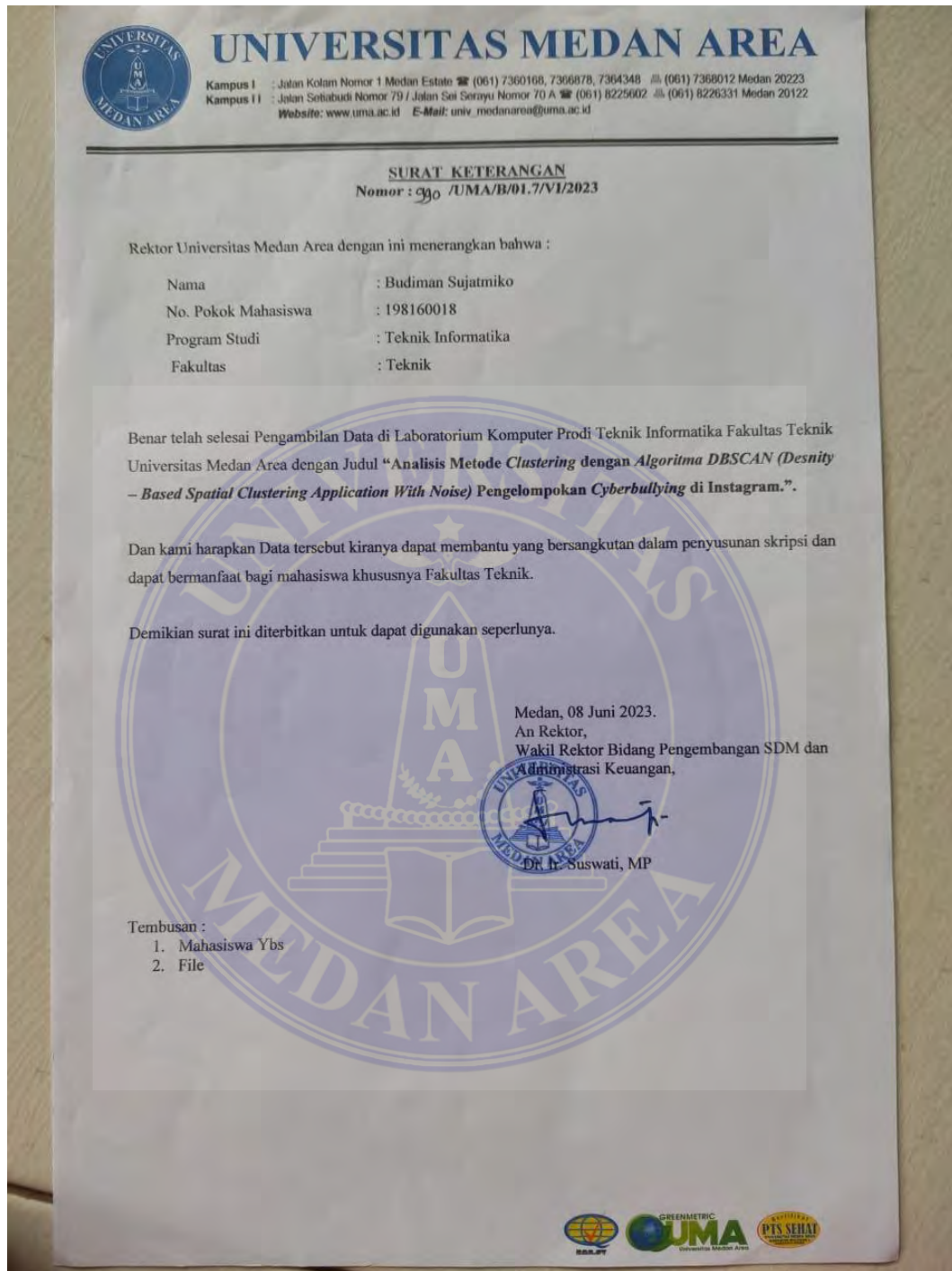
Wakil Rektor Bidang Pengembangan  
SDM dan Administrasi Keuangan,  
  
Dr. Ir. Suswati, MP

**Tembusan :**  
1. Mahasiswa Ybs  
2. File





Lampiran surat selesai riset



### Lampiran codingan program

```
from google.colab import drive
drive.mount ('/content/drive')

import pandas as pd
import numpy as np

CYBER_DATA = '/content/drive/MyDrive/Dataset/cyber.xlsx'
CYBER_DATA=pd.read_excel(CYBER_DATA)
CYBER_DATA.head()

# ----- Case Folding -----
# gunakan fungsi Series.str.lower() pada Pandas
CYBER_DATA['cyber'] = CYBER_DATA['cyber'].str.lower()

print('Case Folding Result : \n')
print(CYBER_DATA['cyber'].head(5))
print('\n\n\n')

pip install nltk

import nltk
nltk.download('punkt')

import string
import re #regex library

# import word_tokenize & FreqDist from NLTK
from nltk.tokenize import word_tokenize
from nltk.probability import FreqDist
```



```

# ----- Tokenizing -----

def remove_tweet_special(text):
    # remove tab, new line, and back slice
    text = text.replace('\t', " ").replace('\n', " ").replace('\u', " ").replace('\\', "")
    # remove non ASCII (emoticon, chinese word, .etc)
    text = text.encode('ascii', 'replace').decode('ascii')
    # remove mention, link, hashtag
    text = ' '.join(re.sub("([@#][A-Za-z0-9]+)|(\w+:\w+\S+)", " ", text).split())
    # remove incomplete URL
    return text.replace("http://", " ").replace("https://", " ")

CYBER_DATA['cyber'] = CYBER_DATA['cyber'].apply(remove_tweet_special)

#remove number
def remove_number(text):
    return re.sub(r"d+", "", text)

CYBER_DATA['cyber'] = CYBER_DATA['cyber'].apply(remove_number)

#remove punctuation
def remove_punctuation(text):
    return text.translate(str.maketrans("", "", string.punctuation))

CYBER_DATA['cyber'] = CYBER_DATA['cyber'].apply(remove_punctuation)

#remove whitespace leading & trailing
def remove_whitespace_LT(text):
    return text.strip()

CYBER_DATA['cyber'] =
CYBER_DATA['cyber'].apply(remove_whitespace_LT)

```

```
#remove multiple whitespace into single whitespace
def remove_whitespace_multiple(text):
    return re.sub('\s+', ' ',text)

CYBER_DATA['cyber'] =
CYBER_DATA['cyber'].apply(remove_whitespace_multiple)

# remove single char
def remove_singl_char(text):
    return re.sub(r"\b[a-zA-Z]\b", "", text)

CYBER_DATA['cyber'] = CYBER_DATA['cyber'].apply(remove_singl_char)

# NLTK word rokenize
def word_tokenize_wrapper(text):
    return word_tokenize(text)

CYBER_DATA['cyber_tokens'] =
CYBER_DATA['cyber'].apply(word_tokenize_wrapper)

print("Tokenizing Result : \n")
print(CYBER_DATA['cyber_tokens'].head())
print("\n\n\n")

# NLTK calc frequency distribution
def freqDist_wrapper(text):
    return FreqDist(text)

CYBER_DATA['cyber_tokens_fdist'] =
CYBER_DATA['cyber_tokens'].apply(freqDist_wrapper)
```

```

print('Frequency Tokens : \n')
print(CYBER_DATA['cyber_tokens_fdist'].head().apply(lambda x :
x.most_common()))

pip install sastrawi

nltk.download('stopwords')

from nltk.corpus import stopwords
nltk.download("stopwords")
#----- get stopwords from NLTK stopwords-----
# get stopwords indonesia
list_stopwords = stopwords.words ('indonesian')

#----- manually add stopwords -----
# append additional stopwords
list_stopwords.extend(["yg", "dg", "rt", "dgn", "ny", "d", 'klo', 'kalo',
'amp', 'biar', 'bikin', 'bilang', 'gak', 'ga', 'krn',
'nya', 'nih', 'sih', 'si', 'tau', 'tdk', 'tuh', 'utk', 'ya',
'jd', 'jgn', 'sdh', 'aja', 'n', 't', 'nyg', 'hehe', 'pen',
'u', 'nan', 'loh', 'rt', '&', 'yah', 'sing', 'ana', 'wis', 'ora',
'liya', 'ing', 'yen', 'karo', 'aja', 'aku', 'ala', 'amarga', 'amargi',
'ta', 'opo', 'koe', 'sampean', 'ngamputen', 'byk', 'tlg', 'hp', 'sy', 'yth',
'pak', 'pa', 'dr', 'lg', 'trims', 'ga', 'tksh', 'jgn'])

#----- add stopwords from txt file -----
#read txt stopwords using pandas
#txt_stopwords = pd_read_excel("stopwords.txt", names= ["stopwords"], header =
None)

#convert stopwords string to list & append additional stopwords
#list_stopwords.extend(txt_stopword["stopwords"] [0].split(' '))

```

```
#-----  
#convert list to dictionary  
list_stopwords = set(list_stopwords)  
  
#remove stopword pada list token  
def stopwords_removal(words):  
    return [word for word in words if word not in list_stopwords]  
  
CYBER_DATA['cyber_tokens_WSW'] = CYBER_DATA  
['cyber_tokens'].apply(stopwords_removal)  
  
print(CYBER_DATA['cyber_tokens_WSW'].head())  
  
import re  
import string  
  
# Fungsi untuk melakukan normalisasi teks  
def text_normalization(text):  
    # Konversi teks menjadi huruf kecil  
    text = text.lower()  
  
    # Menghapus tanda baca  
    text = re.sub(r'[{}]' .format(string.punctuation), '', text)  
  
    # Menghapus angka  
    text = re.sub(r'\d+', '', text)  
  
    # Menghapus karakter-karakter khusus  
    text = re.sub(r'^a-zA-Z0-9\s', '', text)  
  
    # Menghapus spasi berlebih
```

```

text = re.sub(r'\s+', ' ', text).strip()

return text

# Melakukan normalisasi pada kolom 'cyber_tokens_WSW'
CYBER_DATA['cyber_normalized'] =
CYBER_DATA['cyber_tokens_WSW'].apply(lambda tokens: ' '.join(tokens))
CYBER_DATA['cyber_normalized'] =
CYBER_DATA['cyber_normalized'].apply(text_normalization)

print(CYBER_DATA['cyber_normalized'].head())

import nltk
nltk.download('wordnet')

import nltk
from nltk.stem import WordNetLemmatizer
from nltk.corpus import wordnet

# Unduh model POS tagger jika belum ada
nltk.download('averaged_perceptron_tagger')

# Fungsi untuk mendapatkan pos tag (bagian ucapan) dalam bentuk yang sesuai
untuk lemmatisasi
def get_wordnet_pos(word):
    tag = nltk.pos_tag([word])[0][1][0].upper()
    tag_dict = {"J": wordnet.ADJ,
               "N": wordnet.NOUN,
               "V": wordnet.VERB,
               "R": wordnet.ADV}
    return tag_dict.get(tag, wordnet.NOUN)

```



```
# Fungsi untuk melakukan lemmatisasi teks
def lemmatization(text):
    lemmatizer = WordNetLemmatizer()
    words = nltk.word_tokenize(text)
    lemmatized_words = [lemmatizer.lemmatize(word, get_wordnet_pos(word))
for word in words]
    return lemmatized_words

# Melakukan lemmatisasi pada kolom 'cyber_normalized'
CYBER_DATA['cyber_lemmatized'] =
CYBER_DATA['cyber_normalized'].apply(lemmatization)
CYBER_DATA['cyber_lemmatized'] =
CYBER_DATA['cyber_lemmatized'].apply(lambda tokens: ' '.join(tokens))

print(CYBER_DATA['cyber_lemmatized'].head())

!pip3 install swifter
!pip3 install pySastrawi

CYBER_DATA.head()

from nltk.stem import PorterStemmer

# Fungsi untuk melakukan stemming teks
def stemming(text):
    stemmer = PorterStemmer()
    words = nltk.word_tokenize(text)
    stemmed_words = [stemmer.stem(word) for word in words]
    return stemmed_words

# Melakukan stemming pada kolom 'cyber_normalized'
```

```
CYBER_DATA['cyber_stemmed'] =
CYBER_DATA['cyber_normalized'].apply(stemming)
CYBER_DATA['cyber_stemmed'] =
CYBER_DATA['cyber_stemmed'].apply(lambda tokens: ' '.join(tokens))

print(CYBER_DATA['cyber_stemmed'].head())

CYBER_DATA.to_excel("Text_Preprocessing3_Latihan.xlsx")
pip install -u scikit-learn

import pandas as pd
import numpy as np

CYBER_DATA2 = pd.read_excel("/content/Text_Preprocessing3_Latihan.xlsx")
CYBER_DATA2 = CYBER_DATA2[['cyber', 'cyber_tokens']]
CYBER_DATA2.columns = ['text', 'label']

print(CYBER_DATA2.head())

import pandas as pd
from google.colab import drive

# Mount Google Drive
drive.mount('/content/drive')

# Path file Excel di Google Drive
file_path = '/content/drive/MyDrive/Dataset/Text_Preprocessing3_Latihan.xlsx'

# Membaca file Excel
df = pd.read_excel(file_path)

# Mengubah tipe data kolom 'cyber' dari string menjadi list
```

```
df['cyber_list'] = df['cyber'].apply(lambda x: x.split())

# Menyimpan tweet_list dalam Pandas Series
cyber_list = df['cyber_list']

# Menampilkan hasil
print(cyber_list)

import pandas as pd
import numpy as np
from google.colab import drive

# Mount Google Drive
drive.mount('/content/drive')

# Path file Excel di Google Drive
file_path = '/content/drive/MyDrive/Dataset/Text_Preprocessing3_Latihan.xlsx'

# Membaca file Excel
df = pd.read_excel(file_path)

# Mengubah tipe data kolom 'cyber' dari string menjadi list
df['cyber_list'] = df['cyber'].apply(lambda x: x.split())

# Menghitung Term Frequency (TF) untuk setiap dokumen
def calc_TF(document):
    # Menghitung jumlah kemunculan kata dalam dokumen
    TF_dict = {}
    total_terms = len(document)
    for term in document:
        if term in TF_dict:
            TF_dict[term] += 1
```

```

else:
    TF_dict[term] = 1

# Menghitung nilai TF untuk setiap kata
for term in TF_dict:
    TF_dict[term] = TF_dict[term] / total_terms

return TF_dict

# Menghitung Inverse Document Frequency (IDF)
def calc_IDF(corpus, term):
    total_docs = len(corpus)
    doc_count = sum(1 for doc in corpus if term in doc)

    # Menghitung nilai IDF
    IDF = np.log10(total_docs / (doc_count + 1))

    return IDF

# Menghitung IDF untuk setiap term dalam corpus
def calc_TF_IDF(TF_dict, corpus):
    TF_IDF_dict = {}
    for term in TF_dict:
        TF_IDF_dict[term] = TF_dict[term] * calc_IDF(corpus, term)

    return TF_IDF_dict

# Menambahkan kolom baru 'TF_dict' yang berisi Term Frequency (TF)
df['TF_dict'] = df['cyber_list'].apply(calc_TF)

# Menghitung IDF dan menambahkan kolom baru 'TF_IDF_dict' yang berisi TF-
IDF

```

```
corpus = df['cyber_list'].tolist()
df['TF_IDF_dict'] = df['TF_dict'].apply(lambda x: calc_TF_IDF(x, corpus))

# Menampilkan hasil
print(df['TF_IDF_dict'])

import pandas as pd
import numpy as np
from sklearn.cluster import DBSCAN

from sklearn.cluster import DBSCAN
from sklearn.feature_extraction.text import TfidfVectorizer

# Membuat TF-IDF dari kolom 'cyber'
vectorizer = TfidfVectorizer()
X = vectorizer.fit_transform(df['cyber'])

# Menghitung algoritma DBSCAN
eps = 0.7 # Nilai jarak maksimum antara dua sampel
min_samples = 5 # Jumlah minimum sampel di sekitar suatu titik agar titik
tersebut dianggap inti
dbscan = DBSCAN(eps=eps, min_samples=min_samples)
dbscan.fit(X)

# Menambahkan kolom 'cluster' ke dalam DataFrame
df['cluster'] = dbscan.labels_

# Menampilkan hasil
print(df[['cyber', 'cluster']])

import pandas as pd
import numpy as np
```



```
from sklearn.cluster import DBSCAN

# Mengubah kolom 'TF_IDF_dict' menjadi DataFrame
df_tfidf = pd.DataFrame(df['TF_IDF_dict'].tolist())

# Menggantikan nilai 'nan' dengan 0
df_tfidf.fillna(0, inplace=True)

# Konversi DataFrame ke dalam bentuk matriks numerik
tfidf_matrix = df_tfidf.to_numpy()

# Inisialisasi model DBSCAN
dbscan = DBSCAN(eps=0.7, min_samples=2) # Sesuaikan nilai epsilon (eps) dan
min_samples sesuai kebutuhan

# Melakukan fitting dengan model DBSCAN
clusters = dbscan.fit_predict(tfidf_matrix)

# Menambahkan kolom 'Cluster' yang berisi hasil dari DBSCAN
df['Cluster'] = clusters

# Menampilkan hasil kelompok (klaster) yang terbentuk
print(df[['cyber', 'Cluster']])

# Menentukan indeks klaster untuk kalimat bullying dan non-bullying
bullying_cluster = df[df['Cluster'] == -1].index
non_bullying_cluster = df[df['Cluster'] == 0].index

# Menandai kalimat bullying dan non-bullying dalam kolom baru 'Label'
df.loc[bullying_cluster, 'Label'] = 'bullying'
df.loc[non_bullying_cluster, 'Label'] = 'nonbullying'
```

```

# Menampilkan hasil dengan kolom 'cyber' dan 'Label'
print(df[['cyber', 'Label']])

# Menandai kalimat bullying dan non-bullying dalam kolom baru 'Label'
df.loc[bullying_cluster, 'Label'] = 'bullying'
df.loc[non_bullying_cluster, 'Label'] = 'nonbullying'

# Menghitung jumlah kalimat bullying dan non-bullying
jumlah_bullying = len(df[df['Label'] == 'bullying'])
jumlah_nonbullying = len(df[df['Label'] == 'nonbullying'])

# Menampilkan hasil dengan kolom 'cyber', 'Label', dan jumlah kalimat
print(df[['cyber', 'Label']])
print(f'Jumlah Kalimat Bullying: {jumlah_bullying}')
print(f'Jumlah Kalimat Non-Bullying: {jumlah_nonbullying}')

# Menyimpan DataFrame 'df' ke dalam file Excel
output_file_path = '/content/drive/MyDrive/Dataset/hasil_latihan_File.xlsx' #
Sesuaikan dengan path file Excel tujuan
df.to_excel(output_file_path, index=False)
print("DataFrame telah disimpan ke dalam file Excel.")

# Menyimpan hasil ke dalam file Excel
output_file =
'/content/drive/MyDrive/Dataset/Text_Preprocessing3_hasil_latihan.xlsx'
df.to_excel(output_file, index=False)

df[['cyber', 'bullying']].to_excel(output_file, index=False)

pip install wordcloud

```

```
from wordcloud import WordCloud

wordcloud = WordCloud(width=800, height=400, background_color='white',
colormap='viridis', max_words=100)

import pandas as pd

# Mount Google Drive (jika belum dilakukan)
from google.colab import drive
drive.mount('/content/drive')

# Path file Excel di Google Drive
file_path = '/content/drive/MyDrive/Dataset/cyber.xlsx'

# Membaca data dari file Excel
df = pd.read_excel(file_path)

# Mengambil kolom komentar sebagai data teks
data_teks = df['cyber'].astype(str).str.cat(sep=' ')

# Impor pustaka yang diperlukan
from wordcloud import WordCloud
import matplotlib.pyplot as plt

# Menggabungkan semua teks dari DataFrame menjadi satu teks panjang
all_text = ''.join(df['cyber']) # Gantilah 'column_name' dengan nama kolom yang
berisi teks

# Membuat objek Word Cloud
wordcloud = WordCloud(width=800, height=400, background_color='white',
colormap='viridis', max_words=100)
```

```
# Generate Word Cloud dari teks
wordcloud.generate(all_text)

# Menampilkan Word Cloud
plt.figure(figsize=(10, 5))
plt.imshow(wordcloud, interpolation='bilinear')
plt.axis('off') # Hilangkan sumbu
plt.show()

wordcloud.to_file('cyber1.png')
```

