

**ANALISIS DAN IMPLEMENTASI DBSCAN DALAM
CLUSTERISASI POLA BULLYING MEDIA SOSIAL
TWITTER PADA AKUN GIBRAN RAKABUMING
RAKA**

SKRIPSI

DISUSUN OLEH

MUTIARA SANTI

2109010095



UMSU
Unggul | Cerdas | Terpercaya

**PROGRAM STUDI SISTEM INFORMASI
FAKULTAS ILMU KOMPUTER DAN TEKNOLOGI INFORMASI
UNIVERSITAS MUHAMMADIYAH SUMATERA UTARA
MEDAN**

2025

**ANALISIS DAN IMPLEMENTASI DBSCAN DALAM
CLUSTERISASI POLA BULLYING MEDIA SOSIAL TWITTER
PADA AKUN GIBRAN RAKABUMING RAKA**

SKRIPSI

**Diajukan sebagai salah satu syarat untuk memperoleh gelar Sarjana Komputer
(S.Kom) dalam Program Studi Sistem Informasi pada Fakultas Ilmu Komputer
dan Teknologi Informasi, Universitas Muhammadiyah Sumatera Utara**

MUTIARA SANTI

NPM. 2109010095

**PROGRAM STUDI SISTEM INFORMASI
FAKULTAS ILMU KOMPUTER DAN TEKNOLOGI INFORMASI
UNIVERSITAS MUHAMMADIYAH SUMATERA UTARA
MEDAN
2025**

LEMBAR PENGESAHAN

Judul Skripsi : ANALISIS DAN IMPLEMENTASI DBSCAN DALAM
CLUSTERISASI POLA BULLYING MEDIA SOSIAL
TWITTER PADA AKUN GIBRAN RAKABUMING RAKA

Nama Mahasiswa : MUTIARA SANTI

NPM : 2109010095

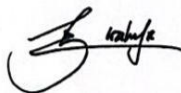
Program Studi : SISTEM INFORMASI

Menyetujui
Komisi Pembimbing



(Martiano S.Pd. S.Kom., M.Kom.)
NIDN. 0128029302

Ketua Program Studi



(Dr. Firahmi Rizky, S.Kom., M.Kom.)
NIDN. 0116079201

Dekan



(Dr. Al-Khowarizmi, S.Kom., M.Kom.)
NIDN. 0127099201

PERNYATAAN ORISINALITAS

ANALISIS DAN IMPLEMENTASI DBSCAN DALAM *CLUSTERISASI* POLA BULLYING MEDIA SOSIAL TWITTER PADA AKUN GIBRAN RAKABUMING RAKA

SKRIPSI

Saya menyatakan bahwa karya tulis ini adalah hasil karya sendiri, kecuali beberapa kutipan dan ringkasan yang masing-masing disebutkan sumbernya.

Medan, Mei 2025

Yang membuat pernyataan



Mutiara Santi

NPM. 2109010095

**PERNYATAAN PERSETUJUAN PUBLIKASI
KARYA ILMIAH UNTUK KEPENTINGAN
AKADEMIS**

Sebagai sivitas akademika Universitas Muhammadiyah Sumatera Utara, saya bertanda tangan dibawah ini:

Nama	: Mutiara Santi
NPM	: 2109010095
Program Studi	: Sistem Informasi
Karya Ilmiah	: Skripsi

Demi pengembangan ilmu pengetahuan, menyetujui untuk memberikan kepada Universitas Muhammadiyah Sumatera Utara Hak Bedas Royalti Non-Eksekutif (*Non-Exclusive Royalty free Right*) atas penelitian skripsi saya yang berjudul:

**ANALISIS DAN IMPLEMENTASI DBSCAN DALAM *CLUSTERISASI* POLA
BULLYING MEDIA SOSIAL TWITTER PADA AKUN GIBRAN
RAKABUMING RAKA**

Beserta perangkat yang ada (jika diperlukan). Dengan Hak Bebas Royalti Non-Eksekutif ini, Universitas Muhammadiyah Sumatera Utara berhak menyimpan, mengalih media, memformat, mengelola dalam bentuk database, merawat dan mempublikasikan Skripsi saya ini tanpa meminta izin dari saya selama tetap mencantumkan nama saya sebagai penulis dan sebagai pemegang dan atau sebagai pemilik hak cipta.

Demikian pernyataan ini dibuat dengan sebenarnya.

Medan, Mei 2025

Yang membuat pernyataan



Mutiara Santi

NPM. 2109010095

RIWAYAT HIDUP

DATA PRIBADI

Nama Lengkap : Mutiara Santi
Tempat dan Tanggal Lahir : Medan, 19 Oktober 2001
Alamat Rumah : Psr8 Helvetia JL.Sidomulio No.68a
Telepon/Faks/HP : 082174921004
E-mail : mutiaraasantii17@gmail.com
Instansi Tempat Kerja : -
Alamat Kantor : -

DATA PENDIDIKAN

SD	: 067255 Medan	TAMAT: 2008-2014
SMP	: 43 Medan	TAMAT: 2014-2017
SMK	: Tritech Informatika Medan	TAMAT: 2017-2020

KATA PENGANTAR



Puji syukur Alhamdulillah, Penulis panjatkan kehadiran Allah SWT., yang telah melimpahkan banyak rahmat dan karunia-Nya serta memberi kekuatan kepada Penulis untuk yang menuntaskan tugas akhir dalam meraih Strata 1 ini. Skripsi ini Penulis sajikan dalam bentuk buku yang sederhana. Judul Skripsi pada penelitian ini adalah sebagai berikut. **“ANALISIS DAN IMPLEMENTASI DBSCAN DALAM *CLUSTERISASI* POLA BULLYING MEDIA SOSIAL TWITTER PADA AKUN GIBRAN RAKABUMING RAKA”**. Adapun Tujuan penulisan skripsi ini dibuat sebagai salah satu syarat kelulusan Program Strata Satu (S1) Sistem Informasi Universitas Muhammadiyah Sumatera Utara. Sebagai bahan penulisan diambil berdasarkan hasil penelitian, observasi dan beberapa sumber literatur yang mendukung penulisan ini. Penulis menyadari bahwa tanpa bimbingan dan dorongan dari semua pihak.

Penulis tentunya berterima kasih kepada berbagai pihak dalam dukungan serta doa dalam penyelesaian skripsi. Penulis juga mengucapkan terima kasih kepada:

1. Bapak Prof. Dr. Agussani, M.AP., Rektor Universitas Muhammadiyah Sumatera Utara (UMSU)
2. Bapak Dr. Al-Khowarizmi, S.Kom., M.Kom. Dekan Fakultas Ilmu Komputer dan Teknologi Informasi (FIKTI) UMSU.
3. Bapak Halim Maulana, S.T., M.Kom, Wakil Dekan Fakultas Ilmu Komputer dan Teknologi Informasi (FIKTI) UMSU.
4. Bapak Martiano, S.Pd., S.Kom., M.Kom Ketua Prodi Sistem Informasi Fakultas Ilmu Komputer dan Teknologi Informasi (FIKTI) UMSU.

5. Ibu Yoshida Sary, S.E., S.Kom., M.Kom Sekretaris Program Studi Sistem Informasi Fakultas Ilmu Komputer dan Teknologi Informasi (FIKTI) UMSU.
6. Bapak Martiano, S.Pd., S.Kom., M.Kom Selaku Dosen Pembimbing
7. Semua pihak yang terlibat langsung ataupun tidak langsung yang tidak dapat penulis ucapkan satu-persatu yang telah membantu penyelesaian skripsi ini.
8. Kepada keluarga tercinta ayah, ibu, serta saudara-saudara penulis yang selalu menjadi sumber kekuatan, doa, dan semangat dalam setiap langkah. Tanpa cinta, dukungan moral, dan dorongan tanpa henti dari kalian, mungkin perjalanan ini tidak akan sampai pada tahap ini.
9. Kepada teman-teman terdekat yang senantiasa hadir dalam suka maupun duka. Terima kasih atas segala bantuan, diskusi, motivasi, serta kebersamaan yang tak ternilai selama proses penyusunan skripsi ini. Kalian adalah bagian penting dalam cerita perjalanan akademik ini.
10. Secara khusus penulis ingin menyampaikan terima kasih yang mendalam kepada seseorang yang sangat spesial, yang selalu hadir dengan pengertian, kesabaran, dan semangat yang tidak pernah surut. Terima kasih telah menjadi bagian penting dalam perjalanan ini, menjadi penyemangat di kala ragu, dan tetap ada

**ANALISIS DAN IMPLEMENTASI DBSCAN DALAM
CLUSTERISASI POLA BULLYING MEDIA SOSIAL TWITTER
PADA AKUN GIBRAN RAKABUMING RAKA**

ABSTRAK

Penelitian ini bertujuan untuk menganalisis pola bullying yang terjadi di media sosial Twitter terhadap akun Gibran Rakabuming menggunakan algoritma DBSCAN (Density-Based Spatial Clustering of Applications with Noise). Data dikumpulkan melalui teknik web scraping dengan kata kunci terkait seperti "Gibran" dan "putra Jokowi", kemudian dilakukan preprocessing untuk mengelompokkan tweet yang mengandung unsur bullying. Algoritma DBSCAN dipilih karena kemampuannya dalam memisahkan data penting (cluster) dari data noise secara otomatis dan akurat, sehingga menghasilkan analisis yang lebih mendalam dan relevan terhadap fenomena cyberbullying. Penelitian ini menggunakan parameter MinPts untuk menentukan jumlah cluster dan outlier yang optimal, serta mengevaluasi hasil dengan Silhouette Score. Hasil analisis diharapkan dapat memberikan wawasan mengenai karakteristik pola bullying di Twitter serta rekomendasi untuk tindakan pencegahan dan penanganan cyberbullying. Selain itu, penelitian ini berkontribusi pada pengembangan ilmu komputer dan teknologi informasi, khususnya penerapan algoritma clustering dalam analisis data teks di media sosial.

Kata Kunci: *DBSCAN; clustering; bullying; media sosial; Twitter; cyberbullying.*

**ANALYSIS AND IMPLEMENTATION OF DBSCAN IN CLUSTERIZATION
OF TWITTER SOCIAL MEDIA BULLYING PATTERNS ON GIBRAN
RAKABUMING RAKA'S ACCOUNT**

ABSTRACT

This study aims to analyze bullying patterns occurring on the Twitter social media platform targeting the account of Gibran Rakabuming by implementing the DBSCAN (Density-Based Spatial Clustering of Applications with Noise) algorithm. Data was collected using web scraping techniques with keywords such as "Gibran" and "putra Jokowi", followed by preprocessing to cluster tweets containing bullying elements. The DBSCAN algorithm was chosen for its ability to separate significant data clusters from noise automatically and accurately, resulting in a more comprehensive and relevant analysis of cyberbullying phenomena. This study utilizes the MinPts parameter to determine optimal clusters and outliers and evaluates the results using the Silhouette Score. The analysis is expected to provide insights into the characteristics of bullying patterns on Twitter and recommendations for cyberbullying prevention and management. Additionally, this research contributes to the field of computer science and information technology, especially in applying clustering algorithms for text data analysis in social media.

Keywords: DBSCAN; clusterin; bullying; social media; Twitte; cyberbullying.

DAFTAR ISI

LEMBAR PENGESAHAN	ii
PERNYATAAN ORISINALITAS	iii
PERNYATAAN PERSETUJUAN PUBLIKASI	iv
RIWAYAT HIDUP	v
KATA PENGANTAR	vi
ABSTRAK.....	viii
<i>ABSTRACT</i>	ix
DAFTAR ISI.....	x
DAFTAR GAMBAR	xiv
BAB I	1
PENDUHLUAN	1
1.1 Latar Belakang Masalah	1
1.2 Rumusan Masalah.....	3
1.3 Batasan Masalah	4
1.4 Tujuan Penelitian	4
1.5 Manfaat Penelitian	5
BAB II.....	7
LANDASAN TEORI	7
2.1 Media Sosial	7
2.2 Twitter	8
2.3 Akun Gibran Rakabuming Raka.....	8
2.4 Clustering	9
2.4.1 Metode <i>Clustering</i>	10
2.4.2 Tantangan Dalam <i>Clustering</i>	11
2.5 Data Mining.....	11
2.6 Algoritma DBSCAN	12
2.7 Text Mining.....	14
2.8 Phyton	15
2.9 Website	15
2.9.1 HTML (Hypertext Markup Language)	16
2.9.2 PHP (Hypertext Preprocessor).....	17
2.9.3 CSS (Cascading Style Sheets)	17
2.9.4 Mysql.....	18

2.9.5	UML (Unified Modelling Language)	18
2.9.6	Bootstrap.....	26
2.9.7	Database	27
2.10	Sistem	27
2.11	Flowchart	28
2.12	Xampp	29
2.13	Literatur Review.....	30
Tabel 2.6 Literatur <i>Review</i>		30
BAB III		33
METODOLOGI PENELITIAN.....		33
3.1	Objek Penelitian	33
3.2	Jenis Penelitian	33
3.3	Tempat dan Waktu Penelitian.....	34
3.3.1	Tempat Penelitian.....	34
3.4	Variable Penelitian	35
3.6	Teknik Pengumpulan Data	37
3.7	Tahapan Analisis Data	37
3.8	Analisis Sistem.....	38
3.9	Spesifikasi Alat yang Digunakan	38
3.10	Perancangan Database	39
3.11	Flowchart.....	41
3.12	Use Case Diagram	43
3.13	Activity Diagram	43
3.14	Sequence Diagram.....	44
3.15	Class Diagram	44
3.16	Metode Perancangan Software Interface	45
BAB IV		49
HASIL DAN PEMBAHASAN.....		49
4.1	Crawling Data.....	49
4.2	Implementasi Sistem.....	51
4.3	Analisis Pola CyberBullying	58
4.4	Uji Coba Program.....	62
4.5	Pengujian Login.....	63
4.6	Pengujian Kelola Data Twitter	64
4.7	Pengujian Titik Cluster.....	65
4.8	Pengujian Proses DBSCAN.....	66
4.9	Pengujian Hasil Laporan	66
4.10	Kelebihan dan Kekurangan Sistem.....	68

BAB V.....	71
PENUTUP.....	71
5.1 Kesimpulan.....	71
5.2 Saran.....	72
DAFTAR PUSTAKA	73
Lampiran	76

DAFTAR TABEL

Tabel 2.1 <i>Use Case</i> Diagram	20
Tabel 2.2 Activity Diagram	22
Tabel 2.3 Class Diagram.....	23
Tabel 2.4 Sequence Diagram	25
Tabel 2.5 Flowchart	28
Tabel 2.6 Literatur Review	30
Tabel 3.1 Waktu Penelitian	35
Tabel 3.2 Login	39
Tabel 3.3 Tabel Data Twitter	39
Tabel 3.4 Tabel Proses <i>Clustering</i>	40
Tabel 3.5 Tabel Hasil	40
Tabel 3.5 Login	39
Tabel 3.6 Tabel Data Twitter	39
Tabel 3.7 Tabel Proses Clustering	40
Tabel 3.8 Tabel Hasil	40
Tabel 4. 1 Pengujian Login	64
Tabel 4. 2 Pengujian Kelola Data Twitter.....	64
Tabel 4. 3 Pengujian Titik Cluster	65
Tabel 4. 4 Pengujian Proses DBSCAN.....	66
Tabel 4. 5 Pengujian Hasil Laporan.....	68

DAFTAR GAMBAR

Gambar 3.1 Alur Metode Penelitian	34
Gambar 3.2 <i>Flowchart</i> Sistem	41
Gambar 3.3 <i>Flowchart</i> Algoritma DBSCAN <i>Clustering</i>	42
Gambar 3.4 <i>Use Case</i> Diagram.....	43
Gambar 3.5 <i>Activity</i> Diagram.....	43
Gambar 3.6 <i>Sequence</i> Diagram.....	44
Gambar 3.7 <i>Class</i> Diagram	45
Gambar 3.8 Layar Login	46
Gambar 3.9 Tampilan Home.....	46
Gambar 3.10 Halaman Data Twitter	47
Gambar 3.11 Titik <i>Cluster</i>	47
Gambar 3.12 Halaman Proses <i>Clustering</i>	48
Gambar 3.13 Laporan Analisa	48
Gambar 4. 1 Proses Crawling Data	50
Gambar 4. 2 Hasil Crawling Data	51
Gambar 4. 3 Tampilan Menu Login.....	52
Gambar 4. 4 Halaman Dashboard	53
Gambar 4. 5 Menu Kelola Data Twitter	54
Gambar 4. 6 Menu Titik Cluster	55
Gambar 4. 7 Menu Proses Clustering DBSCAN	56
Gambar 4. 8 Menu Laporan Hasil DBSCAN.....	58
Gambar 4.9 Visualisasi Hasil Clustering	59
Gambar 4.10 Wordcloud.....	60
Gambar 4.11 Statistik PerCluster	61
Gambar 4.12 Silhouette Score	62

BAB I

PENDUHLUAN

1.1 Latar Belakang Masalah

Cyberbullying adalah tindakan negatif yang dilakukan oleh seseorang atau kelompok tertentu yang melibatkan pengiriman chat, foto, meme, atau video melalui akun seseorang dengan tujuan untuk menyindir, menghina, melecehkan, atau mendiskriminasi orang tersebut. Setiap orang memiliki kebebasan untuk berbagi dan berkomentar. Namun, saat ini banyak pengguna media sosial yang menggunakannya untuk menghina dan memfitnah reputasi mereka, hal tersebut merupakan salah satu bentuk *Cyberbullying*. Serangan daring yang disengaja, atau perilaku mengancam terhadap seseorang di ruang virtual dunia maya. Ini juga disebut sebagai ancaman verbal atau pelecehan berkelanjutan secara daring di media sosial. *Bullying* paling banyak terjadi di wilayah Asia Barat dengan prevalensi 45,1% dan di wilayah Afrika 43,5% , sedangkan Eropa memiliki prevalensi *bullying* terendah yaitu 8,4%. *Cyberbullying* dapat menyebabkan gejala mental yang maksimal pada korban *bullying*, yang berujung pada depresi, perubahan perilaku, pola tidur, dan nafsu makan pada korban *cyberbullying* (Syahira & Kurniawan, 2024).

Sebagai salah satu figur publik, kehadiran Gibran Rakabuming Raka yang terpilih menjadi Wakil Presiden RI, Gibran mempunyai kemampuan unik dalam menyampaikan pesan politik melalui humor yang nyeleneh. Pada akun media sosialnya @gibran_tweet, Gibran selalu menampilkan gaya komunikasi yang jenaka dan terkadang nyentrik (Rizqi Nandadita Pamungkas et al., 2024). Akan tetapi pengguna internet di media sosial, khususnya di X

(sebelumnya Twitter), masih antusias membicarakan isu terkait "Fufufafa". Ia terus menerus menjadi sasaran kritik, pelecehan, dan hinaan, dan Gibran Rakabuming Raka mengakui bahwa semua itu menyakitinya. Namun dia menanggapi pelecehan itu dengan tenang. Menurutnya, penghinaan dan perundungan tidaklah pantas (Rian Alfianto, 2024) Kebebasan dan luasnya jangkauan jaringan media sosial menciptakan peluang bagi individu untuk melakukan kejahatan. Di kolom komentar akun Twitter Gibran Rakabuming Raka, terdapat ujaran kebencian terhadap salah satu publik figur yakni Gibran Rakabuming Raka. Ujaran kebencian tersebut terdapat di kolom komentar akun Twitter berupa hinaan, komentar yang menggunakan bahasa kasar, vulgar, dan merendahkan martabat. Mencakup kata-kata kotor, bahasa yang menghina, bahasa tidak sopan dan dimaksudkan untuk mempermalukan atau melecehkan target di depan umum (Saddhono & Maret, 2024) .

Metode yang digunakan adalah algoritma *Density- Based Spatial Clustering of Application with noise* (DBSCAN) karena memiliki keunggulan dalam mendeteksi data yang tidak termasuk dalam *cluster* dan dianggap sebagai data *noise*. Untuk data yang *noise*, *clustering* lebih optimal karena data yang terdapat pada *cluster* merupakan data yang bersih dan datanya lebih kompatibel. Penelitian ini mengumpulkan data, memprosesnya terlebih dahulu, memasukkannya kedalam algoritma DBSCAN untuk melihat *cluster* mana yang terbentuk, mengevaluasi program dengan menghitung skor siluet, dan melakukan analisis *cluster* untuk menentukan jenis konten. Pengelompokan DBSCAN memiliki dua parameter: *MinPts* dan *Epsilons*. Fungsi *MinPts* menentukan data minimum yang dapat digunakan sebagai data

awal untuk menentukan batas *cluster* dan menggunakan *Epsilon* untuk menentukan jarak antar data dalam *cluster*. DBSCAN menggunakan fungsi jarak *Euclidean* untuk menentukan jarak antar elemen. Proses *clustering* menggunakan algoritma *clustering* DBSCAN melakukan proses *clustering* dengan menggunakan parameter yang berbeda-beda untuk mencapai hasil terbaik (Wikantari et al., 2023).

Penerapan metode *clustering* DBSCAN memiliki 2 parameter yaitu *MinPts* dan *Epsilon*. *MinPts* digunakan untuk menentukan data minimum yang dapat digunakan sebagai data awal untuk menentukan batas *cluster*. DBSCAN menggunakan fungsi jarak *Euclidean* untuk menentukan jarak antar elemen. Kumpulan data yang telah dikumpulkan dimasukkan ke dalam proses pra-pemrosesan untuk membantu membersihkan data. Hal ini memungkinkan input data pada algoritma *clustering* nantinya menghasilkan *cluster* yang lebih optimal .

Penelitian menggunakan DBSCAN, dikarenakan DBSCAN memberikan nilai akurasi pada kasus-kasus sebelumnya berdasarkan studi literatur, Sehingga penelitian menggunakan DBSCAN dapat mendeteksi dalam mengklasifikasi pola *cyberbullying* pada akun media sosial Gibran Rakabuming Raka.

1.2 Rumusan Masalah

Berdasarkan uraian latar belakang yang telah disampaikan, permasalahan yang dapat diidentifikasi adalah :

- a. Bagaimana menganalisis pola *cyberbullying* pada akun twitter

Gibran Rakabuming Raka

- b. Bagaimana merancang algoritma DBSCAN dalam mengenal pola *bullying*
- c. Bagaimana mengimplementasikan algoritma DBSCAN dalam mengenal pola pengelompokan *cyberbullying*

1.3 Batasan Masalah

Untuk menghindari pembahasan diluar permasalahan, maka dibutuhkan batasan masalah yang diuraikan sebagai berikut:

- a. Penelitian ini hanya akan menggunakan data tweet yang diambil dari akun Twitter Gibran Rakabuming Raka selama periode tertentu, misalnya dari tahun 2023 hingga 2024.
- b. Penelitian ini hanya akan menerapkan algoritma DBSCAN untuk *clustering* tanpa membandingkannya dengan algoritma lain seperti *K-Means*.
- c. Fokus analisis adalah pada pola bullying yang teridentifikasi dalam *cluster*, bukan pada analisis sentimen atau aspek lain dari tweet.
- d. Aspek hukum atau etika terkait bullying di media sosial tidak akan menjadi fokus utama dalam penelitian ini.

1.4 Tujuan Penelitian

Adapun tujuan yang ingin dicapai pada penulisan proposal ini adalah sebagai berikut:

- a. Menganalisis dan mengidentifikasi pola *bullying* yang muncul dalam tweet yang ditujukan kepada akun Gibran Rakabuming

Raka di media sosial Twitter.

- b. Mengimplementasikan algoritma DBSCAN untuk melakukan *clustering* terhadap data tweet yang telah dikumpulkan, guna mengelompokkan tweet-tweet yang memiliki kesamaan karakteristik *bullying*.
- c. Memberikan wawasan mengenai dampak *bullying* di media sosial terhadap individu, dengan fokus pada akun Gibran Rakabuming Raka, serta memahami konteks dan jenis *bullying* yang terjadi.
- d. Menyusun rekomendasi berdasarkan hasil analisis untuk tindakan pencegahan atau penanganan *bullying* di media sosial, baik bagi pengguna Twitter maupun bagi pihak-pihak terkait seperti pengelola platform media sosial.
- e. Memberikan kontribusi terhadap literatur yang ada mengenai analisis *bullying* di media sosial dengan menggunakan pendekatan *clustering*, serta memperkaya pemahaman tentang pola interaksi negatif di *platform* tersebut.

1.5 Manfaat Penelitian

Adapun manfaat penelitian ini, yaitu :

- a. Penelitian ini dapat memberikan wawasan yang lebih mendalam mengenai pola-pola *bullying* yang terjadi di media sosial, khususnya di Twitter, sehingga dapat membantu dalam memahami interaksi negatif di *platform* tersebut.
- b. Dengan mempublikasikan hasil penelitian ini, diharapkan dapat meningkatkan kesadaran masyarakat tentang isu *cyberbullying*

dan pentingnya menjaga etika berinteraksi di media sosial, serta mendorong tindakan preventif dari pengguna itu sendiri.

- c. Penelitian ini akan memberikan kontribusi pada bidang ilmu komputer dan teknologi informasi, khususnya dalam penerapan algoritma *clustering* seperti DBSCAN untuk analisis data teks, serta memperkaya literatur yang ada mengenai analisis *bullying* di media sosial.

BAB II

LANDASAN TEORI

2.1 Media Sosial

Media sosial adalah sebuah media online, dengan para penggunanya bisa dengan mudah berpartisipasi, berbagi, dan menciptakan isi meliputi blog, jejaring sosial, wiki, forum dan dunia virtual. Blog, jejaring sosial dan wiki merupakan bentuk media sosial yang paling umum digunakan oleh masyarakat di seluruh dunia. Dampak positif dari media sosial adalah memudahkan kita untuk berinteraksi dengan banyak orang, memperluas pergaulan, jarak dan waktu bukan lagi masalah, lebih mudah dalam mengekspresikan diri, penyebaran informasi dapat berlangsung secara cepat, biaya lebih murah. Sedangkan dampak negatif dari media sosial adalah menjauhkan orang-orang yang sudah dekat dan sebaliknya, interaksi secara tatap muka cenderung menurun, membuat orang-orang menjadi kecanduan terhadap internet, menimbulkan konflik, masalah privasi, rentan terhadap pengaruh buruk orang lain. Adanya media sosial telah mempengaruhi kehidupan sosial dalam masyarakat. Perubahan-perubahan dalam hubungan sosial (*social relationships*) atau sebagai perubahan terhadap keseimbangan (*equilibrium*) hubungan sosial dan segala bentuk perubahan-perubahan pada lembaga-lembaga kemasyarakatan didalam suatu masyarakat, yang mempengaruhi sistem sosialnya, termasuk didalamnya nilai-nilai, sikap dan pola perilaku diantara kelompok-kelompok dalam masyarakat. Perubahan sosial positif seperti kemudahan memperoleh dan menyampaikan informasi,

memperoleh keuntungan secara sosial dan ekonomi. Sedangkan perubahan sosial yang cenderung negatif seperti munculnya kelompok – kelompok sosial yang mengatasnamakan agama, suku dan pola perilaku tertentu yang terkadang menyimpang dari norma – norma yang ada (Yusuf & Khasanah, 2021).

2.2 Twitter

Twitter adalah platform media sosial populer di Indonesia yang digunakan orang untuk berteman, mengekspresikan emosi, dan berbagi informasi. Berdasarkan data, Indonesia merupakan pengguna Twitter terbesar di Asia dan menempati peringkat ke-6 dunia. Di Indonesia, lebih dari 237.556.363 orang terdaftar sebagai pengguna aktif Twitter atau mewakili sekitar 2,41% dari total populasi. Twitter sangat mudah diakses di seluruh dunia dan juga berbeda secara internasional dengan media sosial lain seperti Instagram yang hanya fokus pada foto dan video sedangkan Twitter fokus pada kata-kata berbeda. Karena informasi menyebar jauh lebih cepat di Twitter dibandingkan di media sosial lain seperti Facebook atau Instagram, status pribadi dan tweet sering kali menjadi topik yang sedang tren (Reski Amalia Salam et al., 2024).

2.3 Akun Gibran Rakabuming Raka

Salah satu tokoh politik di Indonesia, Gibran Rakabuming Raka sebagai seorang Wakil Presiden Republik Indonesia melalui aplikasi media sosial khususnya Platform X (sebelumnya dikenal sebagai Twitter) pemilik akun @gibran_tweet, mempunyai kemampuan unik dalam menyampaikan pesan politik melalui humor yang nyentrik. Fenomena ini

menarik perhatian berbagai kalangan, terutama generasi milenial dan gen Z yang cenderung menganut gaya ekspresi dan komunikasi yang tidak lazim (Rizqi Nandadita Pamungkas et al., 2024) Gaya komunikasi humor yang biasa digunakan Gibran Rakabuming Raka menunjukkan bahwa seluruh komentar dalam media sosial Gibran Rakabuming Raka menggunakan teori tindak tutur yang dapat melanggar UU ITE. Komentar yang ditulis oleh pengguna media sosial memiliki banyak bentuk yang dapat mencakup, ejekan, hinaan, pencemaran nama baik, fitnah, dan lelucon. Pernyataan tersebut dimaksudkan untuk mengejek Gibran sebagai orang yang tidak layak dan tidak pantas menyandang atau melakukan sesuatu. Hal ini dapat membuat perasaan seseorang menjadi tidak nyaman atau sakit hati. Oleh karena itu, komentar tersebut mungkin masuk dalam kategori penghinaan atau pencemaran nama baik pemilik akun (Saddhono & Maret, 2024).

2.4 Clustering

Clustering adalah metode dalam data mining yang digunakan untuk mengelompokkan sekumpulan data atau objek ke dalam beberapa kelompok (*cluster*) berdasarkan kesamaan karakteristik di antara mereka. Tujuan utama *clustering* adalah agar objek dalam satu *cluster* memiliki kemiripan yang tinggi, sementara objek di *cluster* yang berbeda memiliki perbedaan yang signifikan. Dengan demikian, *clustering* membantu dalam memahami struktur data dan menemukan pola tersembunyi di dalamnya (Hayat et al., 2021).

Salah satu algoritma yang sering digunakan dalam *clustering* adalah *K-Means*. Algoritma ini berusaha mempartisi data ke dalam sejumlah *cluster* dengan meminimalkan variasi dalam satu *cluster* dan memaksimalkan variasi antar *cluster*. Prosesnya dimulai dengan menentukan jumlah *cluster* yang diinginkan, kemudian secara iteratif menyesuaikan pusat *cluster* (*centroid*) hingga mencapai *konvergensi* (Mutiah et al., 2024).

2.4.1 Metode *Clustering*

Terdapat berbagai metode *clustering* yang digunakan dalam berbagai konteks, antara lain:

a. *Partitional Clustering*

Metode ini mempartisi dataset menjadi sejumlah *cluster* yang ditentukan sebelumnya. Algoritma *K-Means* adalah contoh paling populer. Kelebihannya adalah kecepatan, tetapi memiliki kelemahan dalam menangani data non-linear atau bentuk *cluster* kompleks.

b. *Hierarchical Clustering*

Metode ini membangun hierarki data, baik melalui pendekatan *agglomerative* (menggabungkan data secara bertahap) maupun *divisive* (memisahkan data secara bertahap). Metode ini cocok untuk visualisasi hubungan antar data tetapi memiliki kelemahan dalam skala besar.

c. *Density-Based Clustering*

Metode ini mengidentifikasi *cluster* berdasarkan

kepadatan data, seperti DBSCAN. *Cluster* terbentuk ketika titik-titik dalam data memiliki cukup banyak tetangga dalam radius tertentu.

2.4.2 Tantangan Dalam *Clustering*

Beberapa tantangan dalam *clustering* meliputi:

- a. Definisi *Cluster*: Berbagai algoritma memiliki definisi *cluster* yang berbeda, sehingga hasilnya bervariasi.
- b. Skalabilitas: Kemampuan algoritma untuk menangani dataset besar.
- c. *Noise* dan *Outliers*: Data yang tidak relevan atau jauh dari *cluster* dapat memengaruhi hasil.

2.5 *Data Mining*

Data mining adalah kumpulan proses yang membantu mengeksplorasi dan mengambil nilai dalam bentuk informasi dan hubungan kompleks yang disimpan dalam *database*. Hal ini dilakukan dengan mengekstraksi pola informasi dari data yang membantu mengubahnya menjadi informasi baru dan lebih berguna yang diperoleh melalui ekstraksi, dan juga dengan mengenali pola berharga atau menarik yang diperoleh dari data dalam *database*. Data mining digunakan untuk mengelola data dalam jumlah besar dan mendukung proses penyimpanan data transaksional dan pengolahan data *warehouse* sehingga pengguna dapat mengambil informasi yang di perlukan (Faridzi et al., 2024).

2.6 Algoritma DBSCAN

Density-based method merupakan metode yang dikembangkan berdasarkan *density* (kepadatan) tertentu. Metode ini membentuk *cluster* berdasarkan kepadatan yang cukup tinggi ke dalam *cluster-cluster* dan untuk objek yang tidak masuk ke *cluster* manapun dianggap sebagai *noise*. Sedangkan Salah satu metode berbasis kepadatan (*density-based method*) adalah metode *Density-Based Spatial Clustering Algorithm with Noise* (DBSCAN). Metode ini merupakan sebuah metode *clustering* yang membangun area berdasarkan kepadatan yang terkoneksi (*density-connected*). Setiap objek dari sebuah radius area (*cluster*) harus mengandung setidaknya sejumlah minimum data yang disimbolkan dengan (radius) dan *MinPts*(titik minimum) (Alamsyah et al., 2022). Indikator kata-kata yang mengandung perundungan (bullying), dengan merujuk pada kaidah baku bahasa dari KBBI, yang kemudian akan digunakan dalam proses pemodelan atau pengelompokan data menggunakan algoritma DBSCAN.

Adapun penjelasan rinci terkait kelebihan dan kekurangan algoritma DBSCAN (Fadhlillah et al., 2024) keunggulan algoritma DBSCAN :

- a. Menggunakan konsep berbasis kepadatan, sehingga *outlier/noise* dapat dideteksi.
- b. Tidak seperti *k-means* dan *k-medoid*, jumlah *cluster* (k) tidak perlu ditentukan terlebih dahulu.
- c. Dapat mendeteksi bentuk *cluster* tidak beraturan yang sulit dideteksi dengan beberapa algoritma *clustering* umum lainnya.

- d. Hasil pengelompokan sangat bergantung pada nilai *MinPts* dan *Epsilon* yang ditentukan awal.

Algoritma DBSCAN membutuhkan parameter penting, jumlah minimum poin buat membentuk kelompok (*MinPts*) dan (*Epsilon*) (Isnarwaty & Irhamah, 2020).

Algoritma dari DBSCAN adalah sebagai berikut :

- a. Menentukan parameter *MinPts* dan *Eps*.
- b. Pilih kalimat *cyberbullying* p secara acak.
- c. Menghitung jumlah kalimat *cyberbullying*, Apabila jumlahnya mencukupi (lebih dari atau sama), data akan ditandai menjadi inti (*core point*).
- d. Menghitung jarak titik *core point* menggunakan *point* yg lain memakai jarak *Euclidean*.

Berikut merupakan rumus jarak *Euclidean* digunakan untuk menghitung jarak antar titik dalam konteks pengelompokan DBSCAN.

$$d_{(P,Q)} = \sqrt{\sum_{i=1}^n (p_i - q_i)^2}$$

keterangan :

- $P = (p_1, p_2, \dots, p_n)$: Koordinat titik pertama dalam ruang berdimensi n
- $Q = (q_1, q_2, \dots, q_n)$: Koordinat titik kedua dalam ruang berdimensi n
- $(p_i - q_i)^2$: Selisih antara masing-masing komponen dari Q yang kemudian dikuadratkan.

- $\sqrt{\sum_{i=1}^n}$: Menjumlahkan seluruh hasil kuadrat selisih komponen dari indeks 1 hingga n .
 - $\sqrt{}$: Mengambil akar kuadrat dari jumlah tersebut untuk jarak Euclidean.
- e. Buat *cluster* baru menggunakan menambahkan kalimat *cyberbullying* p ke dalam *cluster*.
 - f. Melakukan identifikasi dalam data yang ditandai sebagai *core point*
 - g. Lanjutkan proses hingga seluruh *point* sudah diproses.
 - h. Apabila terdapat kalimat *cyberbullying* yang tidak masuk ke pada *cluster* manapun akan ditandai menjadi *noise*.

2.7 Text Mining

Text Mining adalah penemuan pengetahuan dalam bentuk tekstual dalam database disingkat (*Knowledge Discovery in Textual Database*) atau KDTD, dan ekstraksi data dalam bentuk tekstual, termasuk minat pada pengetahuan yang baru dibuat dan ditafsirkan, disebut juga pencarian data teks (sebagai bagian dari proses mengekstraksi atau mengambil data teks yang sebelumnya tidak diketahui sehingga dapat di pahami bahwa memiliki pola atau wawasan yang potensial dan dapat ditindaklanjuti dari kumpulan teks yang diambil)

Penggunaan bahasa lisan tertulis atau lengkap dan ringkas, serta tidak terstruktur. Karena 80% dari seluruh perusahaan memiliki dokumen informasi dalam format teks, *text mining* sebagai salah satu cabang data mining dianggap memiliki nilai komersial yang jauh lebih besar

dibandingkan data mining itu sendiri. Namun, data teks yang lebih kompleks yang berisi pola teks tidak terstruktur selalu sulit untuk diekstraksi dan dicari. *Teks Mining* adalah bidang penelitian luas yang mencakup hampir setiap lini kehidupan kita (Firdaus et al., 2021).

2.8 *Python*

Python adalah bahasa pemrograman yang banyak digunakan oleh perusahaan dan pengembang besar untuk mengembangkan berbagai jenis aplikasi desktop, web, dan seluler. *Python* telah berkembang menjadi bahasa pemrograman yang banyak digunakan di industri dan pendidikan karena sintaksisnya yang sederhana, ringkas, intuitif. *Python* adalah bahasa pemrograman (dibuat di Belanda) oleh Guido van Rossum yang namanya diambil dari acara televisi favorit *Guido, Monty Python's Flying Circus*. Van Rossum mengembangkan *Python* sebagai hobi. *Python* telah berkembang menjadi bahasa pemrograman yang banyak digunakan di industri dan pendidikan. Popularitas *Python* menjadikannya bahasa pemrograman yang banyak dipelajari oleh mahasiswa, terutama dari kampus-kampus berbasis IT (Pamulang, 2024).

2.9 *Website*

Website atau web merupakan suatu dokumen berupa kumpulan halaman- halaman yang memuat berbagai informasi dalam bentuk digital. Informasi ini didistribusikan melalui Internet dalam bentuk teks, gambar, animasi, video, atau kombinasi keduanya, dan dapat diakses oleh banyak orang di seluruh dunia selama memiliki koneksi Internet. Situs web adalah kumpulan halaman situs yang dikelompokkan ke dalam domain atau

subdomain di *World Wide Web* (WWW) di internet. Halaman *web* adalah dokumen yang biasanya ditulis dalam format HTML (*Hypertext Markup Language*) dan selalu dapat diakses melalui HTTP. HTTP adalah protokol yang mentransfer informasi dari server situs web dan menampilkannya kepada pengguna melalui *browser web*. Situs *web* biasanya dibuat agar dapat diakses dari aplikasi browser menggunakan *Uniform Resource Locator* (URL). Sebuah situs web terdiri dari serangkaian halaman web yang saling terkait, biasanya disimpan di server web yang dapat diakses melalui Internet atau jaringan area lokal (LAN) (Maria Marlince Weripang et al., 2024).

2.9.1 HTML (*Hypertext Markup Language*)

HTML adalah singkatan dari *Hypertext MarkupLanguage*. HTML mengacu pada bahasa Mark yang dimaksudkan untuk menentukan format atau *style* dari teks yang ditandai. Dokumen HTML adalah dasar dari seluruh konten yang ditampilkan di *World Wide Web* (WWW) dan terdiri dari dua bagian penting. Konten, informasi, dan serangkaian instruksi menampilkan konten tersebut. Setiap dokumen HTML harus ada minimal empat elemen :

`<html> ... </html>`

`<head> ... </head>`

`<title> ... </title>`

`<body> ... </body>`

Elemen-elemen yang menentukan bagian-bagian penting dari sebuah dokumen HTML adalah dokumen, bagian judul, dan isi.

Keempat elemen harus disertakan meskipun tidak ada konten yang disertakan. Setiap elemen HTML ditentukan oleh satu atau dua tag (biasanya tag awal dan akhir). Tag selalu diapit tanda kurung sudut: <...>. Tag akhir dimulai dengan garis miring (/) (Mira Orisa et al., 2023).

2.9.2 PHP (*Hypertext Preprocessor*)

PHP, kependekan dari Hypertext Preprocessor, adalah bahasa pemrograman yang berjalan di sisi server. Bahasa pemrograman PHP pertama kali diperkenalkan oleh Rasmus Lerdorf sebagai aplikasi open source. Pengembang kemudian mulai mengembangkan bahasa pemrograman ini pada tahun 1997, dan lebih dari 500.000 situs web di seluruh dunia menggunakan PHP pada sistem mereka untuk terhubung ke database, menampilkan konten dinamis, dan banyak lagi (Mira Orisa et al., 2023)

2.9.3 CSS (*Cascading Style Sheets*)

Cascading Style Sheets (CSS) merupakan bahasa pemrograman markup (HTML) yang sering digunakan untuk membuat halaman web menjadi lebih menarik. *Cascading Style Sheets* (CSS) mengontrol format tampilan halaman HTML dan XHTML, dan juga dapat diterapkan ke semua dokumen XML, SVG, dan XUL. CSS digunakan untuk memisahkan konteks utama dari tampilan dokumen, seperti layout, warna, font, dll. CSS merupakan rekomendasi dari World Wide Web Consortium (W3C). Stylesheet adalah file teks sederhana yang berekstensi CSS (Mira Orisa et al., 2023).

2.9.4 Mysql

MySQL adalah perangkat lunak sistem manajemen database untuk menyimpan dan memproses data. MySQL merupakan server database SQL multi- user dan multi-threaded. Database ini dibuat untuk tujuan sistem database dan sangat bagus. MySQL untuk sistem operasi Unix bersifat freeware, dan terdapat shareware versi untuk sistem operasi Windows (Azizah & Saptono, 2020).

Menurut (Parlaungan S. & Wisnu, 2020) MySQL adalah jenis server database yang sangat terkenal. Popularitasnya disebabkan oleh fakta bahwa MySQL menggunakan SQL sebagai bahasa dasar untuk akses ke database. MySQL adalah RDBMS (Sistem Manajemen Basis Data Relasional). Di MySQL, database berisi satu atau lebih tabel. Sebuah tabel terdiri dari beberapa baris, setiap baris berisi satu atau lebih kolom. Ada beberapa cara untuk mengelola database MySQL, termasuk prompt DOS (tool command line).

2.9.5 UML (Unified Modelling Language)

Unified Modeling Language (UML) adalah bahasa berbasis gambar untuk memvisualisasikan, menentukan, menulis, dan mendokumentasikan sistem pengembangan perangkat lunak berbasis objek. Meskipun Unified Modeling Language (UML) bukan bahasa pemrograman, model yang dibuatnya secara langsung mereferensikan berbagai bahasa pemrograman, sehingga dapat membuat pemetaan langsung dari model yang dibuat menggunakan Unified Modeling Language dapat dilakukan menggunakan objek -berorientasi pada

bahasa pemrograman seperti Java (UML).UML terdiri dari sejumlah elemen grafis yang membentuk sembilan diagram,dan diagram sequence (Aurin et al., 2021).

Unified Modeling Language (UML) adalah bahasa spesifikasi standar yang digunakan untuk mendokumentasikan, menentukan, dan membuat perangkat lunak. UML adalah metodologi untuk mengembangkan sistem berorientasi objek, seperti alat yang mendukung sistem yang menginformasikan pengambilan keputusan eksekutif dan manajemen operasi bisnis, di mana sistem adalah kombinasi manusia, teknologi informasi, dan prosedur yang terorganisasi (Sandfreni et al., 2021).

Maka dari itu, dapat menyimpulkan bahwa UML atau “Unified Modeling Language” adalah teknik pemodelan visual yang berfungsi sebagai sarana untuk perancangan sistem berorientasi objek. Ada berbagai jenis UML yang masing- masing memiliki karakteristik tersendiri dan dapat digunakan tergantung tujuannya.

a. *Use Case Diagram*

Use Case Diagram ini menunjukkan fungsionalitas sistem atau kelas dan cara berinteraksi dengan dunia luar, *use case diagram* membantu mengidentifikasi persyaratan sistem dan memahami cara kerja sistem. Komponen yang termasuk dalam use case diagram terdiri dari:

- *Actor*: Pengguna perangkat lunak aplikasi, ini bisa berupa manusia, perangkat keras, atau sistem

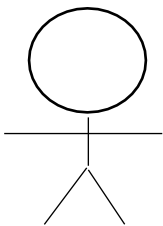
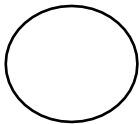
informasi lainnya. *Actor* dapat memasukkan informasi ke dalam sistem , menerima informasi dari sistem , atau keduanya.

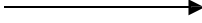
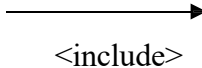
- *Use Case* : Suatu tindakan, atau sesuatu yang dilakukan oleh pengguna sistem aplikasi, termasuk interaksi antara aktor dan perangkat lunak aplikasi.

(Ramdany, 2024).

Berikut adalah simbol-simbol yang ada pada *Use Case* Diagram :

Tabel 2.1 Use Case Diagram

NO	Bentuk simbol	Nama simbol	Fungsi simbol
1		Actor	Menjelaskan sesuatu yang sedang berinteraksi dengan sistem
2		UseCase	Suatu perilaku dari sistem tanpa menggunakan struktur
3		Assosiation	Jalan komunikasi antar aktor dengan use case yang saling Berpartisipasi dalam use case

4		UseCase Generalization	Hubungan antara use case umum dengan use case yang lebih spesifik, yang mewarisis dan menambah fitur terhadapnya.
5		Include	Penambahan perilaku ke dalam use case dasar yang secara eksplisit menjelaskan penambahannya

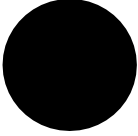
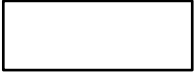
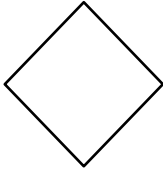
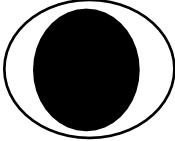
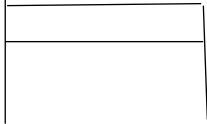
b. *Activity Diagram*

Activity Diagram menunjukkan berbagai aliran aktivitas dalam sistem yang di rancang, bagaimana setiap aliran dimulai, keputusan apa yang dibuat, dan bagaimana aktivitas berakhir.

Activity Diagram juga dapat mewakili proses paralel yang mungkin terjadi dalam beberapa aktivitas. Aktivitas dapat direalisasikan oleh satu *use case* atau lebih. Aktivitas menjelaskan proses yang sedang berlangsung, sementara *use case* menjelaskan bagaimana aktor menggunakan sistem untuk melakukan aktivitas (Ramdany, 2024) Ini adalah simbol-

simbol yang digunakan pada *activity* diagram :

Tabel 2.2 Activity Diagram

No	Simbol	Nama Simbol	Fungsi Simbol
1		Initial	Memberi petunjuk dimana aliran kerja dimulai.
2		Action	Memulai dalam sebuah activity
3		Decision	Memberi petunjuk dimana keputusan akan dibuat
4		Final	Menunjukkan dimana keputusan akan berakhir
5		Swimlane	Menyatukan activity berdasarkan aktor

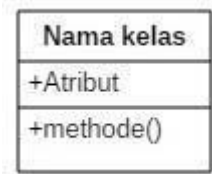
c. Class Diagram

Class Diagram adalah visualisasi kelas suatu sistem dan merupakan jenis diagram yang paling umum digunakan. Diagram ini menunjukkan hubungan antar kelas dan penjelasan rinci setiap kelas dalam model desain (dalam tampilan logis) sistem. Kelas memiliki tiga area utama : nama, atribut, dan

operasi. Nama digunakan untuk memberikan ID pada kelas, fungsi atribut digunakan untuk menentukan properti data yang dimiliki objek kelas, dan operasi fungsi digunakan untuk menetapkan suatu fungsi pada objek (Ramdany, 2024).

Berikut merupakan simbol-simbol dalam *class* diagram :

Tabel 2.3 Class Diagram

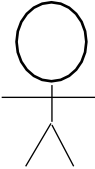
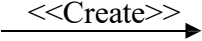
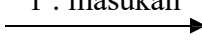
No	Nama Simbol	Nama Simbol	Fungsi Simbol
1		<i>Class</i>	Menggambarkan sebuah kelas pada sistem yang terbagi menjadi 3 bagian. Bagian atas adalah nama kelas. Bagian tengah adalah atribut kelas. Bagian bawah adalah <i>methode</i> dari kelas.
2		<i>Association</i>	Hubungan statis antar kelas. menggambarkan kelas yang memiliki atribut berupa kelas lain atau kelas yang harus mengetahui eksistensi kelas lain.

3		<i>Agregation</i>	Hubungan yang menyatakan bahwa suatu kelas menjadi atribut bagi kelas lain.
4		<i>Composition</i>	Bentuk khusus dari <i>agregation</i> dimana kelas yang menjadi bagian diciptakan setelah kelas <i>whole</i> dibuat.
5		<i>Generalization</i>	Relasi antar kelas dengan makna generalisasi-spesialisasi (umum-khusus).

d. *Sequence Diagram*

Sequence Diagram mewakili interaksi antar objek dalam bentuk pesan selama periode waktu. *Sequence Diagram* terdiri dari dimensi vertikal (waktu) dan dimensi horizontal (objek terkait). Pesan direpresentasikan sebagai garis panah dari satu objek ke objek lainnya. Pada tahap desain berikutnya, pesan dipetakan ke operasi/metode kelas. *Activation bar* menunjukkan waktu eksekusi proses (Ramdany, 2024).

Tabel 2.4 Sequence Diagram

No	Bentuk Simbol	Nama Simbol	Fungsi Simbol
1		Actor	Representasi entitas eksternal yang berinteraksi dengan sistem
2		Waktu aktif	Menyatakan bahwa objek tersebut sedang aktif
3		Pesan tipe create	Menyatakan objek menciptakan objek lain dan panah menunjukkan objek yang dibikin
4		Pesan tipe Send	Objek yang mengirimkan input/data/informasi ke objek lain melalui panah yang mengarah ke objek penerima

5	1 : keluaran →	Pesan tipe return	Objek mengembalikan kontrol ke objek tujuan setelah operasi, dengan panah ke objek penerima.
---	-------------------	----------------------	--

2.9.6 Bootstrap

Bootstrap adalah bagian dari kerangka bahasa CSS dan digunakan khusus untuk membangun halaman web front-end . Kerangka kerja ini awalnya disebut Twitter Blueprint. Framework bootstrap ini biasanya digunakan oleh para developer untuk membuat pengembangan website menjadi lebih mudah dan cepat. Bootstrap sendiri terdiri dari banyak file, dan file dalam framework ini terdiri dari beberapa kumpulan kode CSS dan JavaScript dalam bentuk kelas (Arista & Firmansyah, 2022).

Berdasarkan informasi dari para ahli, dapat disimpulkan bahwa Bootstrap merupakan salah satu framework yang khusus digunakan untuk membuat website front-end. Bootstrap juga membantu mengembangkan banyak fitur yang diperlukan untuk aplikasi dan halaman web, termasuk tipografi, tombol, navigasi , dan komponen lainnya.

2.9.7 Database

Basis data adalah kumpulan data atau informasi yang disimpan dan disusun secara sistematis dalam komputer dan dapat ditinjau, diproses, atau dimanipulasi dengan menggunakan program komputer untuk memperoleh informasi dari basis data. Ada dua jenis database yaitu database relasional (SQL) dan database nonrelasional (NoSQL). Basis data relasional menggunakan Structured Query Language (SQL) untuk mengakses data, sedangkan basis data non-relasional (NoSQL) dirancang khusus untuk model data tertentu dan berguna untuk membangun aplikasi modern dan memberikan fleksibilitas. Basis data NoSQL dikenal karena kemudahan pengembangan, fleksibilitas, kinerja yang unggul dari skala apa pun, fungsionalitas, dan skalabilitas tinggi (Haryanto & Haryanto, 2024).

2.10 Sistem

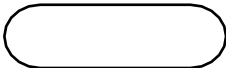
Sistem adalah kumpulan beberapa elemen yang terus berinteraksi dan berintegrasi dengan lingkungan serta dipengaruhi oleh faktor internal dan eksternal untuk mencapai tujuan tertentu. Sistem juga merupakan kumpulan komponen- komponen yang saling berkaitan dan bekerja secara harmonis menuju suatu tujuan yang diinginkan. Sistem adalah jaringan proses yang saling berhubungan yang dikelompokkan bersama untuk melakukan suatu aktivitas atau mencapai tujuan tertentu (Qifli Ilhamdi et al., 2024). Dapat di simpulkan bahwa pengertian sistem adalah gabungan dari beberapa komponen yang terhubung untuk mencapai suatu tujuan.

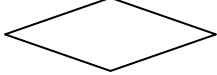
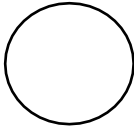
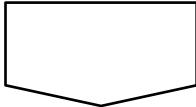
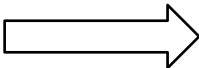
2.11 Flowchart

Flowchart adalah gambar dalam bentuk grafik dengan langkah-langkah dan urutan prosedur program. *flowchart* membantu mengoperasionalkan proses analisis, desain, dan pengkodean dengan memecah masalah menjadi bagian yang lebih kecil. *flowchart* biasanya memudahkan penyelesaian masalah pada saat evaluasi lebih lanjut. Definisi lain dari *flowchart* adalah diagram yang berisi simbol-simbol grafik. Mewakili alur suatu proses dan mempunyai ciri-ciri tertentu yang menunjukkan sejumlah langkah yang disimbolkan atau dapat diartikan sebagai representasi grafis dari langkah-langkah atau urutan suatu prosedur program. Fitur *flowchart* digunakan untuk memberikan gambaran suatu proses produksi dengan cara yang mudah dipahami dan mudah dikenali melalui serangkaian langkah dari satu proses ke proses lainnya. Selain itu, ini menyederhanakan proses dan mempermudah pengguna untuk memahami informasi apa yang diperlukan (Malabay, 2021).

Berikut merupakan tabel flowchart :

Tabel 2.5 Flowchart

NO	Simbol Flowchart	Nama	Arti Simbol
1.		Terminator	Awalan atau akhir konsep (prosedur)
2.		Process	Penggambaran bentuk proses operasional

3.		Document	Dokumen/laporan berbentuk print out
4.		Decision	Keputusan atau sub-pont.
5.		Data	Proses input-output tanpa tergantung dengan jenis peralatannya
6.		On-Page Reference/Connector	Penghubung alur dalam page yang sama
7.		Off-Page Reference /Off-Page Connector	Penghubung alur dalam different page
8.		Flow	Arah alur dalam sebuah prosedur

2.12 Xampp

XAMPP adalah perangkat lunak yang mendukung banyak sistem operasi dan merupakan kumpulan beberapa program. Fungsinya adalah server mandiri (*localhost*) yang terdiri dari program server *HTTP Apache*, basis data *MySQL*, dan penerjemah bahasa yang ditulis dalam bahasa pemrograman *PHP* dan *Perl*. Nama *XAMPP* merupakan singkatan dari *X* (untuk keempat sistem operasi), *Apache*, *MySQL*, *PHP*, dan *Perl*. *Xampp* merupakan server web yang mudah digunakan

yang dapat menyajikan tampilan halaman web yang dinamis (Parlaungan S. & Wisnu, 2020).

2.13 *Literatur Review*

Penelitian ini didukung oleh beberapa referensi yang dimana dapat memperkuat penelitian dalam Menyusun proposal skripsi ini. Beberapa referensi mengenai DBSCAN sebagai berikut:

Tabel 2.6 *Literatur Review*

No	Nama	Judul penelitian	Hasil penelitian
1	(Novieta & Fitriani, 2021)	Penerapan Density- Based Spatial Clustering Of Applications With Noise (DBSCAN) Sebagai Text Clustering Pada Akun Twitter Layanan Provider Indosat	Penerapan metode DBSCAN untuk mengelompokkan pendapat pelanggan di Twitter yang berkaitan dengan layanan provider Indosat. Penelitian menunjukkan bahwa nilai parameter Eps yang lebih besar dapat mengurangi noise dalam clustering namun

			dapat mengakibatkan homogenitas cluster yang menurun. Hasil evaluasi dilakukan menggunakan silhouette coefficient untuk menentukan kualitas cluster
2	(Alamsyah et al., 2022)	Pengelompokan Tweets Pada Akun Twitter Tokopedia Menggunakan Algoritma Density Based Spatial <i>Clustering</i> Of Applications With Noise	Penerapan algoritma DBSCAN pada akun Twitter @tokopedia menghasilkan 2 <i>cluster</i> dengan nilai Silhouette Coefficient 0,575. Konten kampanye WIB (Waktu Indonesia Belanja) memiliki rata- rata retweet dan like

			tertinggi, sedangkan noise didominasi oleh informasi giveaway dengan engagement terendah.
3	(Isnarwaty & Irhamah, 2020)	Text <i>Clustering</i> pada Akun TWITTER Layanan Ekspedisi JNE, J&T, dan Pos Indonesia Menggunakan Metode Density- Based Spatial <i>Clustering of</i>	Analisis tweet pada layanan ekspedisi menunjukkan bahwa kata kunci utama adalah "cek" dan "dm" untuk JNE, "paket" untuk J&T, dan "kirim" untuk Pos Indonesia. Metode DBSCAN

BAB III

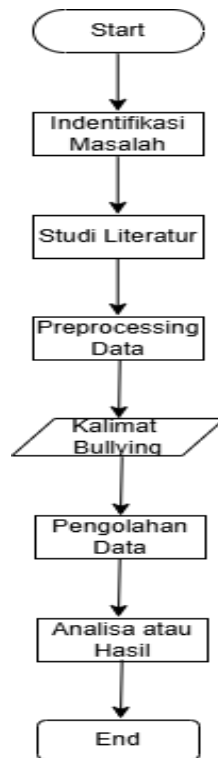
METODOLOGI PENELITIAN

3.1 Objek Penelitian

Penelitian ini berfokus pada analisis pola *bullying* terhadap akun Gibran Rakabuming di jejaring sosial Twitter dengan menggunakan algoritma DBSCAN. Subjek penelitian adalah data pertukaran tweet yang berisi kata-kata dan ekspresi yang diduga merupakan *bullying*. Data ini diperoleh langsung dari Twitter melalui teknik *webscraping* dengan menggunakan parameter *minpts* seperti kata kunci “Gibran”, “putra Jokowi”, dan ekspresi terkait lainnya. Akun Gibran Rakabuming dipilih sebagai fokus investigasi karena ia merupakan tokoh media sosial ternama dan karenanya kerap menjadi bahan perbincangan, baik kritikan, dukungan, maupun serangan pribadi.

3.2 Jenis Penelitian

Jenis penelitian yang digunakan adalah penelitian kuantitatif deskriptif dengan pendekatan eksperimen algoritmik. Penelitian kuantitatif adalah analisis data dalam bentuk numerik yang dihasilkan dari proses ekstraksi fitur teks dan pengelompokan menggunakan algoritma DBSCAN. Untuk menjelaskan pola perundungan, pendekatan deskriptif berdasarkan hasil pengelompokan digunakan, termasuk mengidentifikasi kelompok utama, mendistribusikan pola interaksi, dan mendeteksi anomali.



Gambar 3.1 Alur Metode Penelitian

Dalam penelitian ini, melakukan eksperimen algoritmik untuk mengimplementasikan dan mengevaluasi algoritma DBSCAN pada data yang diproses. Tujuan dari penelitian ini adalah untuk mengukur efektivitas parameter algoritmik *minpts* dalam menghasilkan klaster yang bermakna untuk memberikan gambaran yang lebih akurat mengenai pola perundungan di jejaring sosial pada twitter.

3.3 Tempat dan Waktu Penelitian

3.3.1 Tempat Penelitian

Survei dilakukan secara daring, dengan platform media sosial Twitter sebagai sumber utama pengumpulan data. Seluruh proses penelitian, termasuk pengumpulan data, analisis, dan interpretasi, akan dilakukan secara virtual dan di lokasi studi yang fleksibel tergantung pada akses internet perangkat peneliti.

3.3.2 Waktu Penelitian

Tabel 3.1 Waktu Penelitian

No	Kegiatan Penelitian	Waktu Penelitian					
		December	Januari	Februari	Maret	April	Mei
1.	Pengajuan Judul						
2.	Riset Awal						
3.	Pembuatan Proposal						
4.	Bimbingan Proposal						
5.	Seminar Proposal						
6.	Riset						
7.	Penyusunan Skripsi						
8.	Bimbingan Skripsi						
9.	Sidang Meja Hijau						

Pada tabel 3.1 menunjukkan waktu penelitian yang di lakukan peneliti dari awal proses pengajuan judul sampai meja hijau.

3.4 Variable Penelitian

Dalam penelitian ini, ada dua jenis variabel yaitu, variabel bebas yang terdiri dari nilai Minimum Points (*MinPts*) yang menjadi parameter penting dalam algoritma DBSCAN, Nilainya akan menentukan bagaimana data dikelompokkan ke dalam *cluster* atau dianggap sebagai *outlier*. dan variabel terikat yang terdiri dari jumlah klaster (*n_cluster*), jumlah outlier (*n_outlier*), dan nilai *Silhouette Score*, digunakan untuk mengevaluasi hasil clusterisasi berdasarkan nilai *MinPts*. Dengan mencoba berbagai nilai *MinPts*, peneliti dapat menemukan pengaturan terbaik yang menghasilkan jumlah *cluster* optimal, meminimalkan *outlier*, dan meningkatkan kualitas clusterisasi (dengan nilai *Silhouette Score* yang tinggi).

3.5 Algoritma DBSCAN

Untuk menentukan parameter *MinPts* dalam algoritma DBSCAN (*Density-Based Spatial Clustering of Applications with Noise*), menggunakan rumus

heuristik, yaitu :

Menentukan nilai *MinPts*:

1. Heuristik Logaritmik

$$MinPts \geq \ln(N)$$

- N = jumlah total data dalam dataset
- Digunakan ketika jumlah data sangat besar

2. Dua Kali Jumlah Dimensi (d)

$$MinPts \geq 2 \times d$$

- d = jumlah fitur dalam dataset

Jika dataset memiliki 5 fitur, maka *MinPts* bisa ditetapkan sebagai 10

Misalnya, dataset berisi 500 tweet, dan fitur yang digunakan adalah jumlah kemunculan kata-kata negatif (sentimen *bullying*), jumlah likes, jumlah retweet, dll.. Jika dataset memiliki 5 fitur, maka:

$$MinPts \geq 2 \times 5 = 10$$

Atau

$$MinPts \geq \ln(500) \approx 6.21 \rightarrow 7$$

bisa ditetapkan *MinPts* = 10 sebagai titik awal, dan dilakukan eksperimen dengan nilai sekitar 5–10 untuk menemukan hasil optimal.

3. Evaluasi Kualitas Kluster

Metrik untuk mengukur kualitas kluster yang dihasilkan dari DBSCAN:

- Silhouette Score: Mengukur seberapa baik setiap titik diklasifikasikan ke dalam kluster. Rumus: $s(i) = \frac{b(i) - a(i)}{\max\{a(i), b(i)\}}$

$a(i)$ adalah jarak rata-rata dari titik i ke semua poin lain dalam kluster yang sama

$b(i)$ adalah jarak rata-rata dari titik i ke semua poin dalam kluster terdekat.

3.6 Teknik Pengumpulan Data

Dalam melakukan sebuah penelitian, diperlukannya sebuah data data yang valid untuk mendukung analisis suatu penelitian, Dan Teknik pengumpulan data yang digunakan penulis dalam penelitian ini ialah:

- a. Pengumpulan data dilakukan menggunakan *crawling* untuk memperoleh tweet yang relevan dari akun Gibran Rakabuming (@gibran_tweet) serta interaksi lainnya yang menyebut akun tersebut.
- b. Kriteria pengumpulan data tweet dalam periode 2020 hingga 2024 dengan meta data yang diperlukan yaitu konten tweet, waktu posting, dan hastag yang digunakan.
- c. Pengolahan dan penyimpanan data tweet yang telah dikumpulkan dapat dianalisis secara efektif untuk mendeteksi pola *bullying* di media sosial pengguna twitter.

3.7 Tahapan Analisis Data

Metode analisis data yang digunakan dalam penelitian ini adalah clustering menggunakan metode DBSCAN. Tahap analisis yang dilakukan adalah sebagai berikut:

- a. Melakukan pengumpulan dan *preprocessing* data.
- b. Melakukan ekstrasi fitur dalam proses nalisis *clustering*.
- c. Kemudian melakukan *clustering* penerapan DBSCAN menggunakan parameter *Minpts* dan *Eps*.
- d. Validasi *cluster* menggunakan koefisien *silhouette* pada kelompok yang telah dibentuk menggunakan metode DBSCAN.

- e. Setelah itu dilakukan analisis pembahasan dari hasil pengelompokkan kedua metode tersebut.

3.8 Analisis Sistem

Analisis sistem adalah penguraian keseluruhan sistem informasi menjadi komponen atau bagian-bagiannya dengan tujuan mengidentifikasi dan mengevaluasi masalah, peluang, dan hambatan yang muncul dalam sistem juga mengusulkan perbaikan. Tujuan dari analisis sistem ini adalah untuk mengidentifikasi dan mengevaluasi masalah yang ada dalam sistem dan memberikan gambaran yang jelas tentang sistem yang akan dikembangkan. Analisis sistem juga membantu memperjelas konsep desain dengan elemen-elemen yang akan disertakan dalam sistem dari perspektif administrator sistem (admin), pengguna (user), dan perangkat yang akan digunakan untuk membuat sistem nantinya.

3.9 Spesifikasi Alat yang Digunakan

Spesifikasi alat yang digunakan untuk menentukan kebutuhan yang mendukung pembuatan system. Antara lain:

- a. Kebutuhan *hardware*, Meliputi :
 - Laptop dengan *processor* intel Core i5
 - *Memory* minimal 8 GB
- b. Kebutuhan *software*, Meliputi :
 - Sistem operasi *windows 10 Enterprise* 64-bit
 - Bahasa pemrograman *Python* untuk mengimplementasikan algoritma dan *crawling*
- c. Data yang di butuhkan, Meliputi :

- Data dengan pola *bullying* di akun **@gibran_tweet**, pada periode tahun 2023 hingga 2024.

3.10 Perancangan *Database*

Dalam perancangan *database* perencanaan data record tersimpan dalam beberapa file dengan arsitektur data sebagai berikut :

a. Tabel Login

Tabel login berisi atribut-atribut mengenai *username password* yang harus diisi agar dapat masuk dalam sistem. Berikut tampilan tabel login di bawah ini:

Tabel 3.2 Login

No	Field Name	Type Field	Field Size
1	<i>Username</i>	<i>Varchar</i>	10
2	<i>Password</i>	<i>Varchar</i>	10

b. Tabel Data Twitter

Berikut rancangan struktur data yang dapat dilihat di bawah ini:

Tabel 3.3 Tabel Data Twitter

No	Field Name	Type Field	Field Size	Keterangan
1	ID_Tweet	<i>Varchar</i>	20	<i>Primary Key</i>
2	<i>Username</i>	<i>Varchar</i>	50	<i>Username Twitter</i>
3	Isi_Tweet	<i>Text</i>	-	Konten tweet
4	Tanggal_Tweet	<i>DateTime</i>	-	Waktu tweet
5	Sentimen	<i>Varchar</i>	20	Positif, Negatif
6	<i>Cluster_ID</i>	<i>Int</i>	-	<i>Cluster hasil DBSCAN</i>

c. Tabel Proses *Clustering*

Tabel proses *clustering* digunakan untuk menampung record data nilai

clustering yang kemudian akan disimpan sebagai *database*. Berikut tampilan rancangan struktur yang berisi atribut-atribut data di bawah ini:

Tabel 3.4 Tabel Proses *Clustering*

No	Field Name	Type Field	Field Size	Keterangan
1	<i>Cluster_ID</i>	<i>Int</i>	-	<i>Primary Key</i> , ID untuk <i>cluster</i>
2	ID_Tweet	<i>Varchar</i>	20	<i>Foreign Key</i> ke Data_Twitter
3	Sentimen	<i>Varchar</i>	20	Hasil analisis sentimen
4	Jarak	<i>Float</i>	-	Jarak antar data poin
5	<i>Noise</i>	<i>Boolean</i>	-	<i>True</i> jika <i>noise</i> , <i>False</i> jika tidak

d. Tabel Hasil

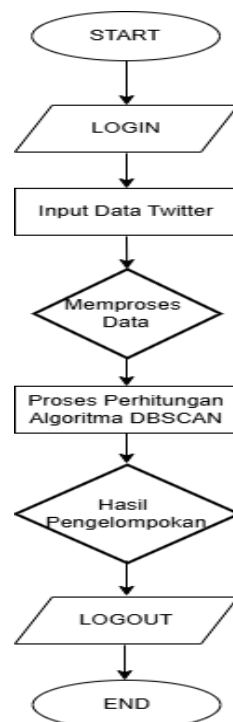
Tabel hasil digunakan untuk menampung record data hasil *clustering* keseluruhan yang kemudian akan disimpan sebagai *Database*. Berikut tampilan rancangan struktur tabel yang berisi atribut-atribut data di bawah ini:

Tabel 3.5 Tabel Hasil

No	Field Name	Type Field	Field Size	Keterangan
1	<i>Cluster_ID</i>	<i>Int</i>	-	<i>Primary Key</i> , ID untuk <i>cluster</i>
2	Jumlah_Anggota	<i>Int</i>	-	Total jumlah data dalam <i>cluster</i>

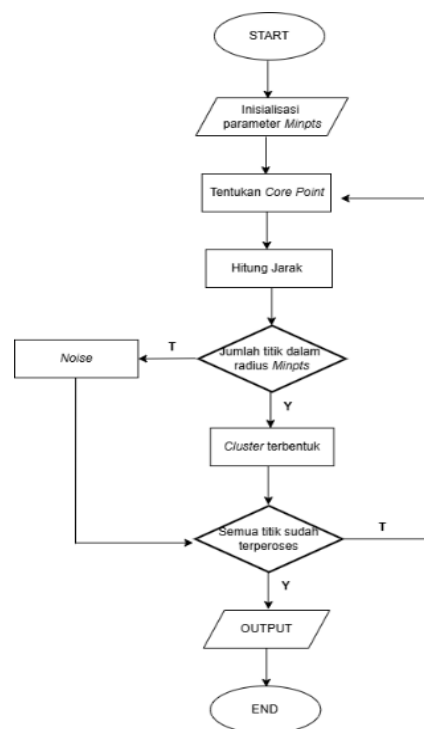
3	Deskripsi_Cluster	Text	-	Karakteristik cluster
4	Persentase_Noise	Float	-	Persentase data noise dalam cluster
5	Jumlah_Sentimen_Pos	Int	-	Jumlah tweet dengan sentimen positif
6	Jumlah_Sentimen_Neg	Int	-	Jumlah tweet dengan sentimen negatif

3.11 Flowchart



Gambar 3.2 Flowchart Sistem

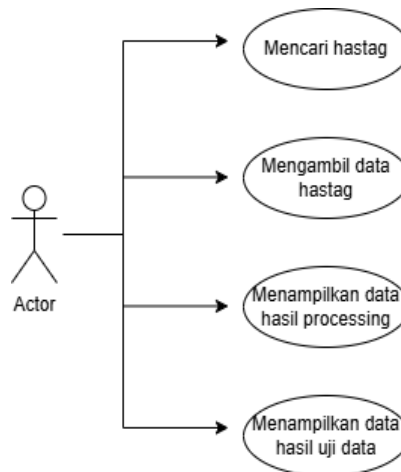
Flowchart sistem diatas merupakan sebuah alur atau proses untuk masuk ke dalam sistem yang akan dibangun untuk sistem dan memberikan gambaran alur kerja yang terstruktur, mulai dari input data kemudian sistem akan memproses perhitungan Algoritma DBSCAN dan hasil dari perhitungan tersebut dikelompokkan.



Gambar 3.3 Flowchart Algoritma DBSCAN Clustering

Flowchart diatas merupakan *flowchart* dari algoritma DBSCAN *clustering* yang dimana melakukan metode yang tepat untuk menganalisis pola *bullying* di media sosial karena mampu memisahkan data penting (*cluster*) dari data tidak relevan (*noise*) secara otomatis dan akurat. Hal ini memungkinkan hasil analisis yang lebih mendalam dan relevan untuk mendukung penelitian terkait fenomena *cyberbullying*.

3.12 Use Case Diagram

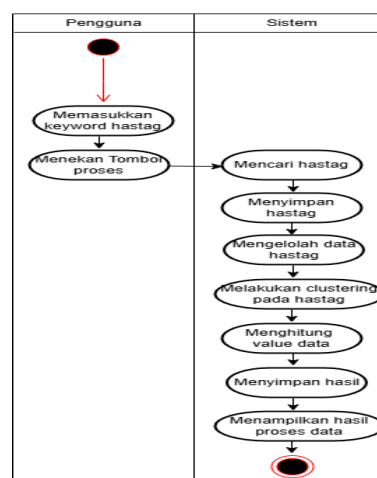


Gambar 3.4 Use Case Diagram

Dari Gambar 3.4 dapat dilihat bahwa actor dapat melakukan dua hal, yaitu mencari hashtag sesuai keyword yang dimasukkan, dan mengambil data hashtag tersebut.

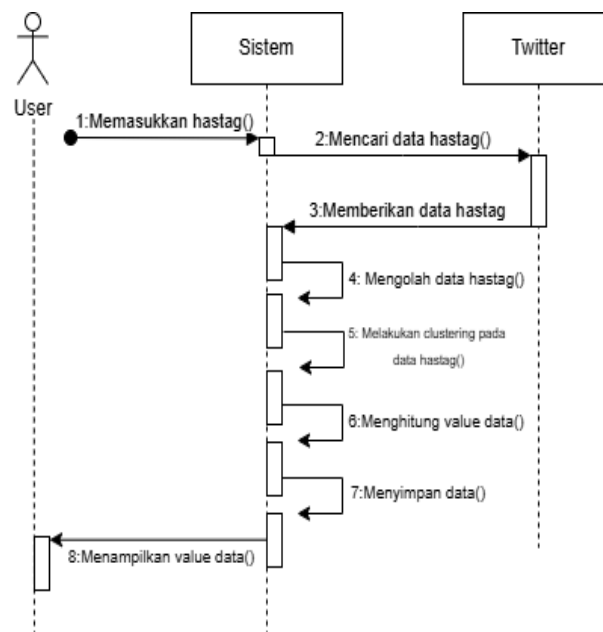
3.13 Activity Diagram

Activity diagram berfungsi untuk menjelaskan runutan kejadian dari suatu kegiatan. Dalam activity diagram dijelaskan awal dari alur suatu kegiatan, macam-macam kemungkinan yang dapat terjadi, dan keadaan akhir dari kegiatan tersebut. Alur activity diagram dapat dilihat pada Gambar 3.5



Gambar 3.5 Activity Diagram

3.14 Sequence Diagram

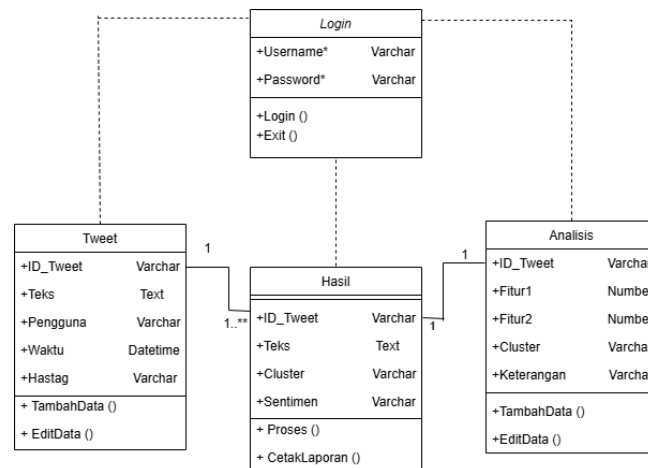


Gambar 3.6 Sequence Diagram

Pada Gambar 3.6, setelah mengambil data hashtag dilakukan beberapa tahap untuk mengolah data, yaitu mengubah format data menjadi *lowercase*, menghilangkan *stopword* atau kata yang umum digunakan, menghilangkan tanda baca yang tidak diperlukan, dan memilah data antara data yang memiliki kata yang memiliki makna *bullying* dan bukan. Kemudian data tersebut dikelompokkan menjadi beberapa *cluster* sesuai dengan parameter yang diberikan dan ditampilkan dalam bentuk grafik.

3.15 Class Diagram

Class diagram pada bagian ini digunakan untuk menjelaskan hubungan antara class yang ada pada sistem dan bagaimana masing-masing class berhubungan untuk mencapai suatu tujuan. Penggambaran class diagram pada aplikasi data mining sebagai berikut.



Gambar 3.7 Class Diagram

3.16 Metode Perancangan *Software Interface*

Perancangan *Software Interface* merupakan sebuah tahapan dalam proses mengimplementasikan metode atau algoritma yang telah dirancang ke dalam suatu bahasa pemrograman. Pada tahapan ini akan menghasilkan aplikasi yang digunakan sebagai sarana untuk mempresentasikan hasil dari metode yang menjadi tujuan dalam penelitian ini. Berikut perancangan *software interface* yang akan digunakan dalam analisis dan implementasi DBSCAN dalam clusterisasi pola *bullying* media sosial twitter pada akun Gibran Rakabuming dengan menggunakan algoritma *clustering*.

a. Halaman Login

Pada gambar Layar login merupakan tampilan dari rancangan halaman login untuk user masuk ke menu dashboard user harus mengisi username dan password.

LOGIN SISTEM

Username

Password

Reset Login

Gambar 3.8 Layar Login

b. Halaman Dashboard

Halaman ini merupakan halaman dashboard yang dimana terdapat beberapa fitur menu yang dapat diakses.

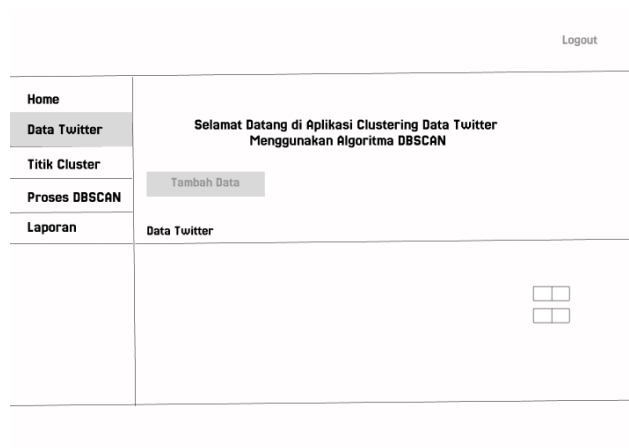
Logout

Home	Selamat Datang di Aplikasi Clustering Data Twitter Menggunakan Algoritma DBSCAN Background
Data Twitter	
Titik Cluster	
Proses DBSCAN	
Laporan	

Gambar 3.9 Tampilan Home

c. Halaman Kelola Data Twitter

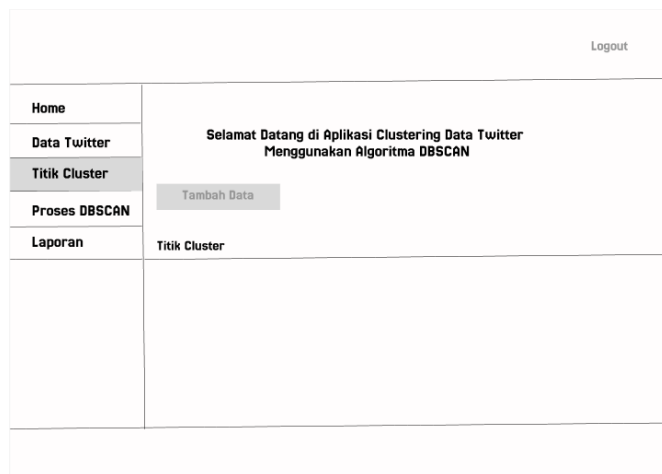
Halaman ini merupakan halaman untuk mengelola data twitter, halaman ini admin dapat menginput data sesuai kolom yang ada.



Gambar 3.10 Halaman Data Twitter

d. Rancangan Layar Titik *Cluster*

Merupakan tampilan dari rancangan halaman titik *cluster*. Setelah melakukan login sebelumnya yang dapat diakses oleh admin seperti melihat data twitter serta mengelola titik *cluster* untuk perhitungan algoritma nantinya.



Gambar 3.11 Titik *Cluster*

e. Rancangan Layar Proses *Clustering*

Merupakan tampilan dari rancangan halaman proses DBSCAN. Pada halaman ini admin dapat melihat data twitter yang hasil

pengelompokkan dengan algoritma DBSCAN beserta nilai jarak pada setiap *cluster* yang ada.

Logout						
Home	Selamat Datang di Aplikasi Clustering Data Twitter Menggunakan Algoritma DBSCAN					
Data Twitter						
Titik Cluster						
Proses DBSCAN						
Laporan						
	Hasil Clustering					
	ID_Twitter	Nama Akun	Jarak Cluster1	Jarak Cluster2	Cluster Terpilih	Keterangan
	xxx	xxx	999	999	999	xxx
	xxx	xxx	999	999	999	xxx

Gambar 3.12 Halaman Proses *Clustering*

f. Rancangan Layar Laporan

hasil perhitungan DBSCAN merupakan tampilan dari halaman laporan hasil perhitungan algoritma DBSCAN. Pada halaman ini akan terlihat jelas hasil *clustering* twitter *bullying* dan *non-bullying*

Logout					
Clustering DBSCAN Hasil Pola Bullying di Twitter Gibran Rakabuming					
ID_Twitter	Nama Akun	Jarak Cluster1	Jarak Cluster2	Cluster Terpilih	Keterangan
xx	xxx	999	999	999	999
xx	xxx	999	999	999	999
xx	xxx	999	999	999	999
Dari proses clustering dapat diambil kesimpulan bahwa : xxxxxx					
Diketahui Oleh :					
(xxxxxxxxxxxxxxxx)					

Gambar 3.13 Laporan Analisa

BAB IV

HASIL DAN PEMBAHASAN

4.1 Crawling Data

Pada tahapan ini, peneliti melakukan proses crawling data sebagai langkah awal dalam mengumpulkan informasi yang relevan untuk keperluan analisis. Teknik crawling digunakan untuk memperoleh data dari platform media sosial Twitter, khususnya yang mengandung hashtag #fufufafa, yang diasumsikan memiliki keterkaitan dengan isu atau diskusi yang sedang berkembang selama periode waktu tertentu. Proses pengumpulan data dilakukan dalam rentang waktu mulai dari tanggal 08 Agustus 2024 hingga 01 April 2025, dengan tujuan untuk menangkap dinamika percakapan publik secara aktual dan kontekstual selama kurun waktu tersebut.

Melalui proses crawling ini, berhasil dikumpulkan sebanyak 500 data atau entri, yang masing-masing memuat informasi penting seperti 'title', 'author', 'publish_date', 'article_text', 'url', 'main_image', dan 'tag'. Setiap elemen data ini diharapkan dapat memberikan kontribusi terhadap pemahaman yang lebih mendalam mengenai pola penyebaran informasi, opini publik, atau potensi indikasi terhadap suatu fenomena sosial tertentu. Setelah data terkumpul, langkah selanjutnya adalah melakukan proses preprocessing, yaitu tahap pembersihan dan penyaringan data, di mana hanya kolom atau atribut yang dianggap relevan dan sesuai dengan fokus penelitian yang akan digunakan dalam tahap analisis lebih lanjut.



```

[1] # Install Node.js (because tweet-harvest built using Node.js)
!sudo apt-get update
!sudo apt-get install -y ca-certificates curl gnupg
!sudo mkdir -p /etc/apt/keyrings
!curl -fsSL https://deb.nodesource.com/gpgkey/nodesource-repo.gpg.key | sudo gpg --dearmor -o /etc/apt/keyrings/nodesource.gpg

!NODE_MAJOR=20 && echo "deb [signed-by=/etc/apt/keyrings/nodesource.gpg] https://deb.nodesource.com/node_${NODE_MAJOR}.x nodistro main" | sudo tee /etc/apt/sources.list.d/nodesource.list

!sudo apt-get update
!sudo apt-get install nodejs -y

!node -v

[ ] # Crawl Data

filename = 'fufufafacrawling.csv'
search_keyword = 'fufufafa since:2024-08-08 until:2025-04-01 lang:id'
limit = 500

!npx -y tweet-harvest@2.6.1 -o "{filename}" -s "{search_keyword}" --tab "LATEST" -l {limit} --token {twitter_auth_token}

```

Gambar 4. 1 Proses Crawling Data

Setelah proses crawling data berhasil diselesaikan, seluruh data yang telah dikumpulkan kemudian ditampilkan dalam bentuk tabel sebagai representasi visual awal dari hasil ekstraksi. Gambar berikut menyajikan cuplikan dari data mentah yang diperoleh melalui crawling, yang secara umum terdiri atas berbagai atribut penting seperti judul, penulis, tanggal publikasi, isi teks, tautan artikel, gambar utama, serta tag yang terkait. Visualisasi ini dimaksudkan untuk memberikan gambaran awal mengenai struktur dan isi data sebelum memasuki tahap pengolahan lebih lanjut.

data yang telah diperoleh diekspor ke dalam format Excel (.csv) untuk memudahkan proses pengelolaan, pemfilteran, serta analisis lanjutan pada tahapan berikutnya. File Excel ini tidak hanya memudahkan dalam segmentasi data berdasarkan atribut tertentu, tetapi juga berfungsi sebagai dokumentasi awal terhadap hasil crawling yang telah dilakukan, yang kemudian akan menjadi dasar dalam proses preprocessing dan analisis data yang bersifat lebih mendalam.

Gambar 4. 2 Hasil Crawling Data

Setelah proses crawling data berhasil dilakukan dan data berhasil dikumpulkan, langkah berikutnya adalah mengimplementasikan data tersebut ke dalam sistem untuk menjalani proses preprocessing. Tahapan ini mencakup pembersihan data, normalisasi, serta pemilahan atribut yang relevan agar data menjadi lebih terstruktur dan siap digunakan dalam proses analisis lebih lanjut secara optimal.

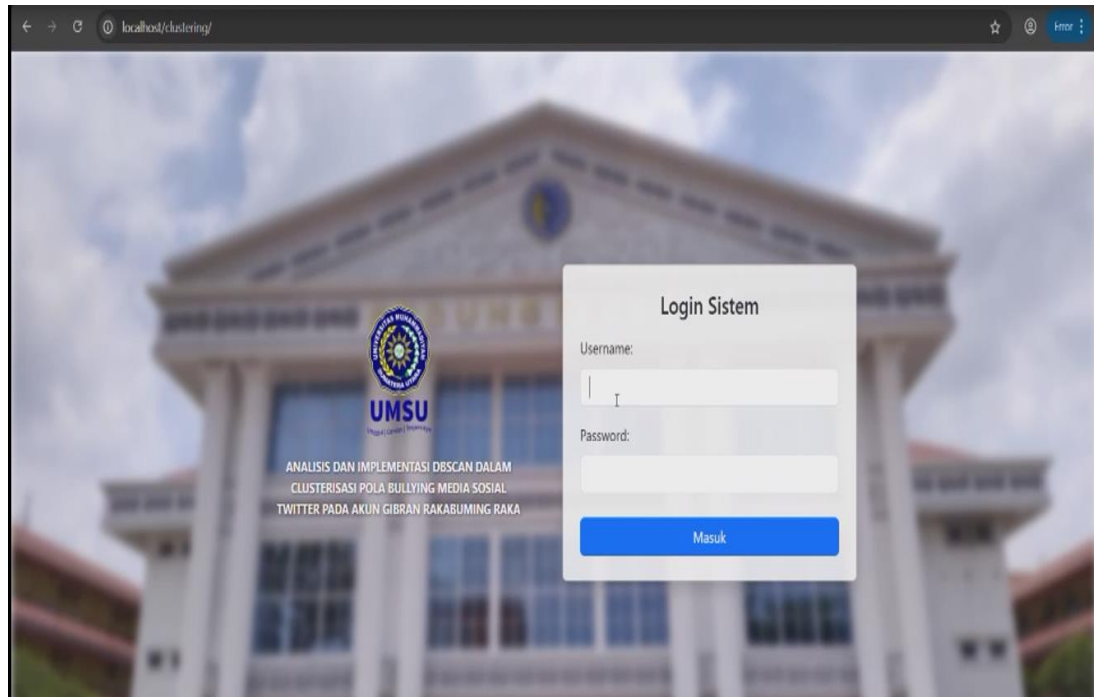
4.2 Implementasi Sistem

Berikut hasil dari implementasi sistem yang nantinya akan dilakukan untuk menganalisis dbscan,seluruh tampilan sistem ialah sebagai berikut :

a. Tampilan login

Tampilan login dengan username dan password dibuat untuk membatasi akses hanya kepada pengguna yang memiliki hak masuk ke

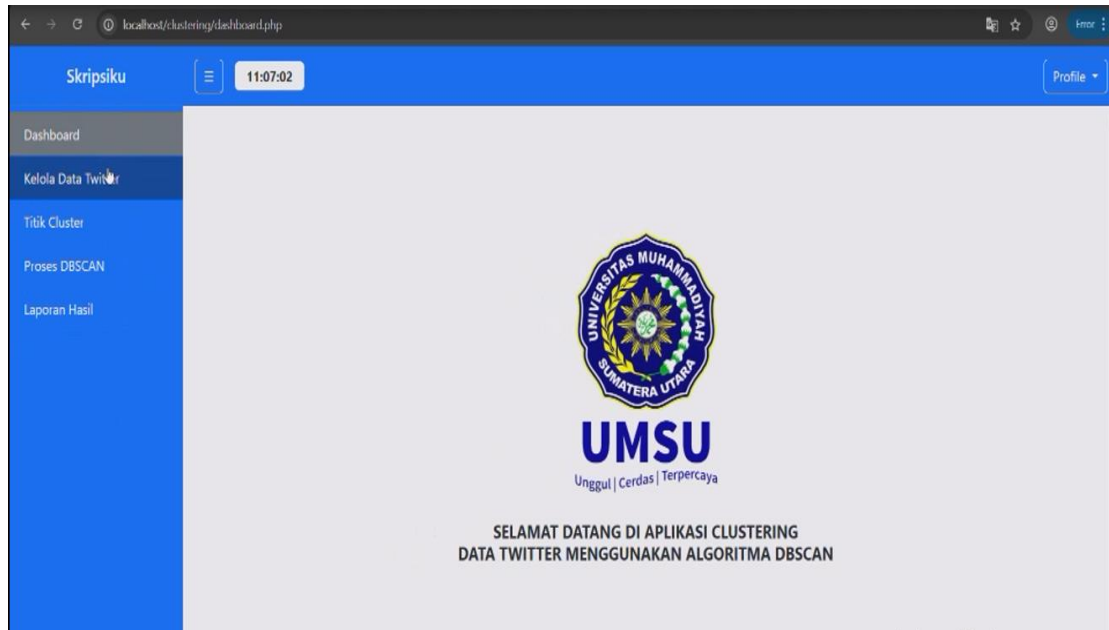
dalam sistem. Dengan cara ini, keamanan data dapat lebih terjaga dan setiap aktivitas pengguna bisa tercatat dengan baik, sehingga sistem menjadi lebih aman dan tertata.



Gambar 4. 3 Tampilan Menu Login

b. Tampilan Dashboard

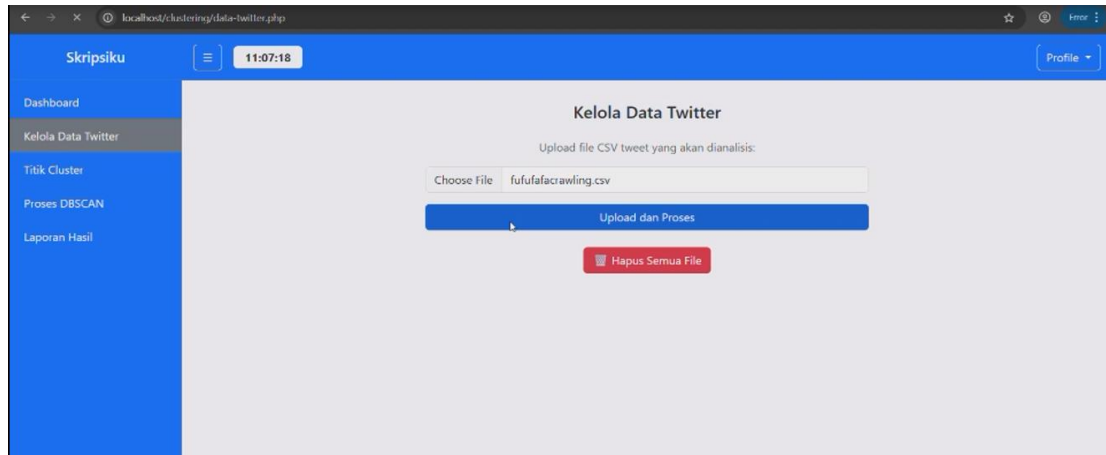
Menu dashboard dibuat sebagai tampilan utama setelah pengguna berhasil login ke dalam sistem. Tujuannya adalah untuk menyajikan ringkasan informasi penting secara cepat, terstruktur, dan mudah dipahami. Melalui dashboard, pengguna dapat melihat data inti, grafik, notifikasi, dan navigasi menuju fitur-fitur utama lainnya tanpa harus membuka satu per satu. Dengan adanya dashboard, sistem menjadi lebih interaktif, efisien, dan memudahkan pengguna dalam mengambil keputusan atau melakukan tindakan selanjutnya.



Gambar 4. 4 Halaman Dasboard

c. Menu Kelola Data Twitter

Menu Kelola Data Twitter digunakan untuk mengunggah file CSV berisi data tweet yang akan dianalisis. Pengguna dapat memilih file hasil crawling, lalu menekan tombol "Upload dan Proses" untuk menyimpan dan memproses data ke dalam sistem. Tersedia juga tombol "Hapus Semua File" untuk menghapus seluruh data yang telah diunggah sebelumnya, sehingga proses pengelolaan data lebih rapi dan terkontrol.



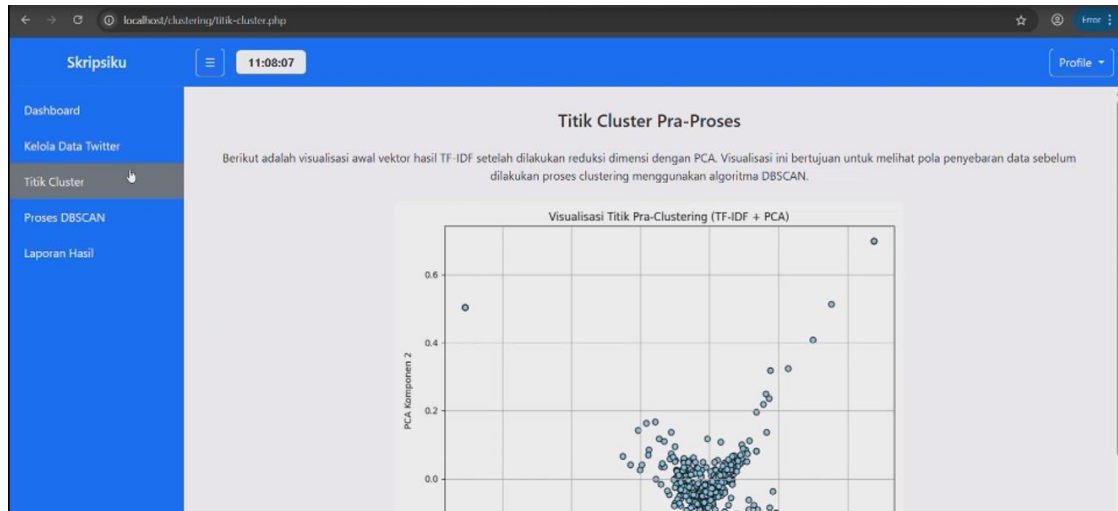
Gambar 4. 5 Menu Kelola Data Twitter

d. Menu Titik Cluster

Menu titik cluster ini digunakan untuk menampilkan visualisasi awal data sebelum dilakukan proses *clustering DBSCAN*.

Menu ini menyajikan grafik scatter plot hasil reduksi dimensi vektor TF-IDF menggunakan PCA, yang berguna untuk:

- Melihat penyebaran data dalam ruang 2 dimensi,
- Mengamati potensi pola atau kelompok (cluster) yang bisa terbentuk,
- Menjadi tahap analisis awal sebelum masuk ke proses clustering sesungguhnya.



Gambar 4. 6 Menu Titik Cluster

e. Menu Proses DBSCAN

Halaman "Proses Clustering DBSCAN" berfungsi sebagai antarmuka untuk menjalankan proses clustering (pengelompokan data) secara otomatis menggunakan algoritma *DBSCAN* (*Density-Based Spatial Clustering of Applications with Noise*) terhadap data yang telah dikumpulkan dalam konteks ini adalah data dari Twitter.

Adapun fitur input Minpts berfungsi sebagai Berfungsi untuk menentukan jumlah minimum titik tetangga yang diperlukan agar sebuah titik dapat dianggap sebagai core point. Parameter ini mempengaruhi sensitivitas pembentukan cluster. Semakin besar nilainya, semakin sulit titik membentuk cluster.

Contoh menggunakan minpts pada dbscan sesuai dengan jumlah

- < 100 data → Gunakan MinPts = 3–5
- 100–500 data → Gunakan MinPts = 4–6
- 500 data → Gunakan MinPts = 5–10

Pada kasus ini peneliti menggunakan 500 data twitter, maka peneliti menggunakan MinPts di angka 4-5 agar DBSCAN dapat secara optimal mendeteksi pola dan noise pada data sosial media.

Gambar 4. 7 Menu Proses Clustering DBSCAN

f. Menu Laporan Hasil

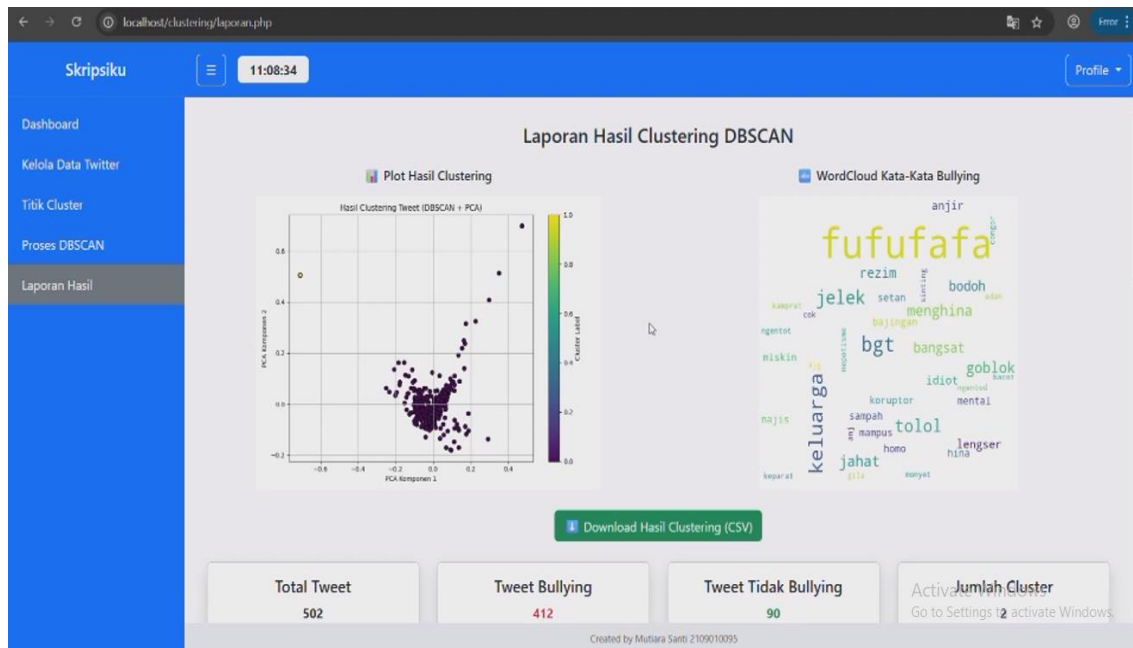
Halaman Laporan Hasil Clustering DBSCAN menyajikan rangkuman komprehensif dari proses analisis dan pengelompokan tweet menggunakan kombinasi algoritma DBSCAN (Density-Based Spatial Clustering of Applications with Noise) dan teknik reduksi dimensi PCA (Principal Component Analysis). Pada sisi kiri tampilan, pengguna dapat melihat visualisasi dua dimensi dari hasil clustering. Setiap titik pada grafik mewakili satu tweet, dengan posisi dan warna yang menunjukkan afiliasi tweet tersebut terhadap cluster tertentu. Titik-titik yang tersebar di luar pola utama dianggap sebagai outlier atau noise, yang berarti tweet tersebut tidak tergolong dalam kelompok manapun berdasarkan parameter clustering yang

digunakan.

Sementara itu, sisi kanan halaman menampilkan WordCloud yang secara visual merepresentasikan frekuensi kata-kata yang muncul dalam tweet yang terdeteksi sebagai mengandung unsur bullying. Semakin besar ukuran teks pada sebuah kata, semakin tinggi tingkat kemunculannya di dalam dataset. Kata-kata seperti “tolol”, “goblok”, “jelek”, dan “anjir” menjadi indikator kuat yang menandai keberadaan ujaran kebencian atau konten bermuatan negatif di media sosial.

Pada bagian bawah halaman, terdapat ringkasan statistik hasil clustering, yang mencakup total 502 tweet yang dianalisis. Dari jumlah tersebut, sebanyak 412 tweet dikategorikan sebagai bullying, sementara 90 tweet lainnya tergolong tidak bullying. Informasi ini diperkuat dengan jumlah cluster yang berhasil dihasilkan oleh proses DBSCAN, serta adanya fitur tambahan berupa tombol unduh data dalam format CSV, memudahkan pengguna dalam menyimpan dan mengolah hasil untuk kebutuhan dokumentasi atau analisis lanjutan.

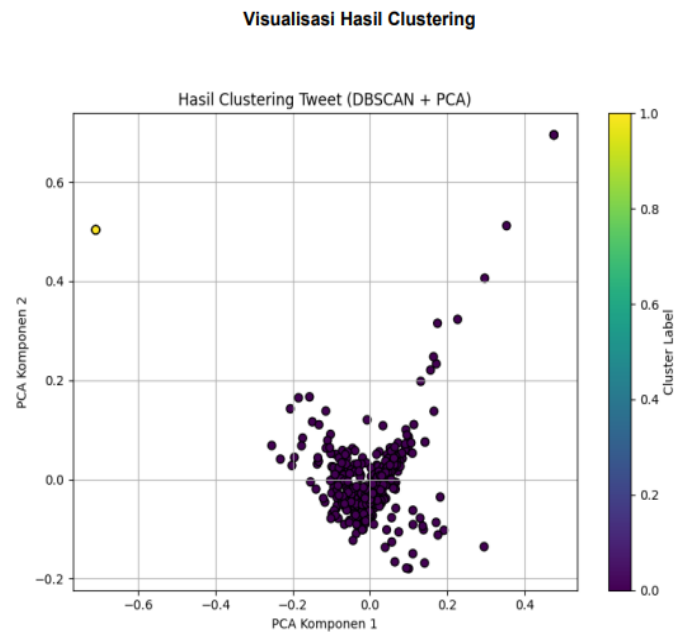
Secara keseluruhan, halaman ini dirancang untuk memberikan pemahaman mendalam mengenai distribusi tweet berdasarkan pola kemiripan konten, mengidentifikasi kosakata dominan dalam kasus cyberbullying, dan menyediakan laporan terstruktur yang dapat digunakan untuk keperluan riset, pelaporan, atau pengambilan keputusan berbasis data.



Gambar 4. 8 Menu Laporan Hasil DBSCAN

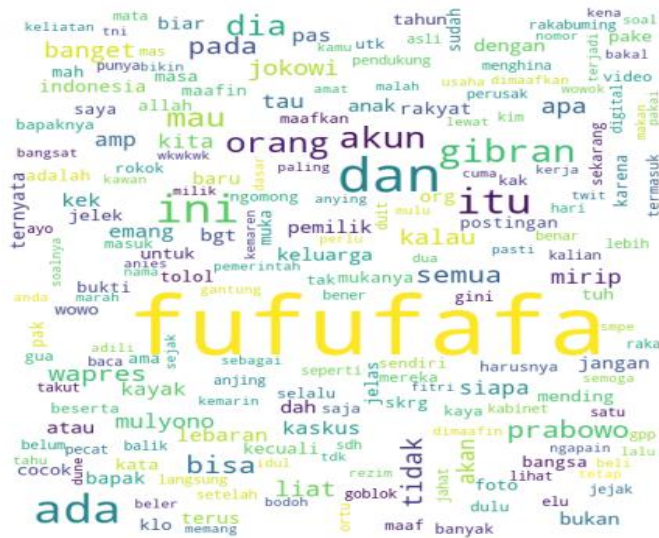
4.3 Analisis Pola CyberBullying

Berdasarkan hasil visualisasi clustering menggunakan algoritma DBSCAN yang dikombinasikan dengan PCA, terlihat adanya pemusatan titik data pada satu area tertentu, yang merepresentasikan kelompok tweet dengan karakteristik serupa. Cluster utama memiliki kepadatan tinggi, sementara beberapa titik yang berada di luar area pusat terdeteksi sebagai outlier, yang kemungkinan besar mengandung konten dengan intensitas bullying yang sangat tinggi atau menggunakan kata-kata yang berbeda dari mayoritas.



Gambar 4.9 Visualisasi Hasil Clustering

Selanjutnya, wordcloud yang dihasilkan menampilkan dominasi kata-kata kasar, ofensif, dan menghina seperti "goblok", "tolol", "anjir", "bangsat", "jelek", "jahat", hingga istilah bermuatan rasis atau seksual. Kemunculan kata-kata tersebut dalam ukuran besar menandakan frekuensi penggunaannya yang tinggi dalam tweet yang diklasifikasikan sebagai bullying. Ini mengindikasikan bahwa pola cyberbullying pada data yang dianalisis cenderung eksplisit dan langsung menyerang secara verbal.



Berdasarkan hasil clustering menggunakan algoritma DBSCAN terhadap 502 tweet, diperoleh dua cluster utama, yaitu Cluster 0 dan Cluster 1. Analisis terhadap distribusi konten bullying menunjukkan bahwa:

- Cluster 0 mencakup 497 tweet, di mana sebanyak 408 tweet (82,09%) dikategorikan sebagai bullying, sedangkan 89 tweet (17,91%) bersifat non-bullying. Hal ini menunjukkan bahwa sebagian besar tweet dalam cluster ini mengandung unsur ujaran kebencian atau serangan verbal yang relevan dengan topik "fufufafa".
- Cluster 1 mencakup 4 tweet, dan 100% dari tweet tersebut mengandung unsur bullying. Meskipun jumlah tweet pada cluster ini relatif kecil, tingkat bullying yang sangat tinggi mengindikasikan bahwa konten dalam cluster ini bersifat sangat toksik atau

mengandung kata-kata kasar secara eksplisit.

dari 502 tweet yang dianalisis, terdapat 412 tweet yang mengandung unsur bullying, yang berarti 82,07% dari total data yang dikumpulkan adalah tweet dengan indikasi cyberbullying. Temuan ini memperkuat indikasi bahwa kata kunci "fufufafa" menjadi pusat percakapan yang kerap disertai dengan ekspresi negatif atau penghinaan di media sosial.

Statistik per Cluster

Cluster	Total Tweet	Tweet Bullying	Non-Bullying	% Bullying
0	497	408	89	82.09%
1	4	4	0	100%

Gambar 4.11 Statistik PerCluster

Pada nilai diperoleh nilai Silhouette Score sebesar 0.8750, yang menunjukkan bahwa hasil clustering yang dilakukan tergolong sangat baik. Nilai ini mendekati angka 1, yang secara teoritis menandakan bahwa setiap data dalam satu cluster memiliki kemiripan tinggi terhadap data sekelompoknya, namun memiliki perbedaan yang cukup signifikan terhadap data di cluster lainnya. Dengan kata lain, pemodelan yang dilakukan berhasil mengelompokkan tweet bullying dan non-bullying secara jelas dan terpisah. Hal ini juga sejalan dengan visualisasi jarak antar cluster sebelumnya, di mana terlihat bahwa titik-titik data membentuk dua kelompok yang terpisah secara tegas tanpa adanya tumpang tindih antar cluster. Maka dapat disimpulkan bahwa model DBSCAN yang digunakan

efektif dalam mengidentifikasi dan mengelompokkan pola bullying pada data tweet yang dianalisis.



Gambar 4.12 Silhouette Score

Secara keseluruhan, pola cyberbullying yang teridentifikasi dalam kluster menunjukkan adanya konsistensi dalam penggunaan bahasa kasar, penghinaan terhadap fisik dan mental, serta adanya intensi untuk merendahkan atau menyerang individu atau kelompok tertentu. Hal ini menunjukkan bahwa algoritma DBSCAN berhasil mengelompokkan tweet secara efektif, sekaligus memberikan gambaran nyata tentang bentuk-bentuk ujaran kebencian yang terjadi di media sosial.

4.4 Uji Coba Program

Pengujian sistem dilakukan untuk memastikan bahwa seluruh fitur dan fungsional berjalan dengan baik sesuai yang diharapkan sebelum sistem diimplementasikan secara penuh. Adapun spesifikasi perangkat keras dan perangkat lunak yang diperlukan disiapkan guna menjamin bahwa sistem dapat dijalankan secara optimal dan tanpa kendala.

1. Satu unit laptop dengan spesifikasi :
 - a. Laptop dengan processor intel Core i5
 - b. RAM minimal 8 GB

- c. Hard Disk minimal 128 GB
2. Komponen Deskripsi :
- a. Visual Studio Code Editor kode selama pengembangan sistem
 - b. Laragon Web server dan database lokal (Apache, MySQL, PHP, dan lainnya) Bootstrap 5 Framework frontend untuk UI responsif
 - c. Python 3.11+ Bahasa pemrograman utama untuk preprocessing, clustering, dan visualisasi
 - d. PHP 8 Digunakan untuk backend web dan integrasi antarmuka pengguna
 - e. MySQL / MariaDB Database relasional untuk menyimpan tweet, hasil preprocessing, dan clustering

Berikut pengujian yang dilakukan setiap sistem:

4.5 Pengujian Login

Tabel di bawah ini menyajikan hasil pengujian pada fitur login aplikasi. Pengujian ini bertujuan untuk memastikan bahwa pengguna dapat mengakses sistem menggunakan kombinasi username dan password yang valid, serta sistem mampu menolak akses apabila data yang dimasukkan tidak sesuai.

Data Masukan	Hasil yang Diharapkan	Pengamatan	Kesimpulan
Username dan Password Benar	Login Ketampilan Menu dashboard	Login berhasil	[✓] valid [] Invalid

Username dan Password Salah	Menampilkan login gagal dan ulang masukan username dan password yang benar	Login Gagal	[✓] valid [] Invalid
-----------------------------	--	-------------	--------------------------

Tabel 4. 1 Pengujian Login

4.6 Pengujian Kelola Data Twitter

Pengujian ini dilakukan untuk memastikan bahwa semua sistem berjalan dengan baik termasuk dapat mengupload data file csv yang akan untuk dianalisis dan dapat menghapus semua data yang di upload untuk membuat analisis yang baru.

Data Masukan	Hasil yang Diharapkan	Pengamatan	Kesimpulan
Impor file data yang mau dianalisa	Memproses data dan mengupload kedalam sistem	File berhasil di proses dan siap di upload	[✓] valid [] Invalid
Hapus semua file data yang dianalisis	Menghapus semua data yang di upload	Dapat menghapus semua file data	[✓] valid [] Invalid

Tabel 4. 2 Pengujian Kelola Data Twitter

4.7 Pengujian Titik Cluster

Pada tahap ini dilakukan pengujian visualisasi terhadap sebaran data hasil ekstraksi fitur menggunakan metode TF-IDF, yang kemudian direduksi dimensinya dengan menggunakan Principal Component Analysis (PCA). Visualisasi ini ditampilkan dalam bentuk plot dua dimensi dengan sumbu PCA Komponen 1 dan PCA Komponen 2.

Data Masukan	Hasil yang Diharapkan	Pengamatan	Kesimpulan
Data hasil ekstraksi fitur TF-IDF dari dokumen/tweet	Sistem menampilkan sebaran data dalam bentuk visualisasi 2D hasil reduksi PCA	Sistem berhasil memvisualisasikan titik-titik yang merepresentasikan dokumen dalam ruang PCA	☒ valid ☐ Invalid
Lanjut ke Proses Clustering	Sistem melanjutkan ke tahapan clustering	Sistem berhasil berpindah ke halaman proses clustering	☒ valid ☐ Invalid

Tabel 4. 3 Pengujian Titik Cluster

4.8 Pengujian Proses DBSCAN

Pada halaman ini pengujian dbscan untuk memastikan bahwa semua berfungsi dengan baik, tabel berikut ini merangkum hasil dari pengujian tersebut ,termasuk memasukan minpts dan epsilon.

Data Masukan		Hasil yang Diharapkan	Pengamatan	Kesimpulan
Input Nilai Mintpts Dan Epsilon		Sistem mengeksekusi proses clustering DBSCAN dan menghasilkan label cluster	Sistem berhasil menjalankan proses clustering dan memproses data sesuai parameter yang dimasukkan	[✓] valid [] Invalid
Jalankan DBSCAN		Sistem memproses dan menampilkan hasil clustering	Tombol dapat diklik dan proses clustering berjalan tanpa error	[✓] valid [] Invalid

Tabel 4. 4 Pengujian Proses DBSCAN

4.9 Pengujian Hasil Laporan

Pada halaman ini pengujian dapat dilakukan untuk menampilkan hasil seluruh analisa pada dbscan yang sudah dijalankan ,seperti tampilan titik cluster,wordcloud,download hasil laporan,jumlah tweet yang terdeteksi bullying dan lain-lainnya dapat dilihat pada tabel berikut:

Data Masukan	Hasil yang Diharapkan	Pengamatan	Kesimpulan
Dataset tweet hasil clustering telah tersedia	Menampilkan plot hasil clustering (DBSCAN + PCA)	Scatter plot berhasil muncul sesuai label cluster, dengan skala warna label cluster	☒ valid ☐ Invalid
Dataset hasil DBSCAN	Menampilkan WordCloud kata-kata bullying dari hasil tweet	WordCloud berhasil divisualisasikan dengan berbagai kata yang terdeteksi sebagai bullying	☒ valid ☐ Invalid
Klik tombol Download Hasil Clustering (CSV)	File CSV hasil clustering dapat diunduh	Tombol dapat diklik dan file berhasil diunduh	☒ valid ☐ Invalid
Jumlah tweet terdeteksi	Menampilkan total seluruh tweet dari	Angka total sesuai yang	☒ valid ☐ Invalid

	yang terdeteksi bullying dan tidak bullying	ditampilkan dan sesuai ekspektasi output dari proses DBSCAN	
--	---	---	--

Tabel 4. 5 Pengujian Hasil Laporan

4.10 Kelebihan dan Kekurangan Sistem

a. Kelebihan Sistem

1. Integrasi Langsung dengan Basis Data

Sistem ini tidak hanya memproses data dari file CSV, tetapi juga memungkinkan pemrosesan langsung dari basis data MySQL. Hal ini memberikan keunggulan dalam manajemen data, efisiensi penyimpanan, dan kemudahan pemantauan histori proses.

2. Otomatisasi Proses *Preprocessing* dan *Clustering*

Sistem secara otomatis melakukan preprocessing teks (seperti penghapusan URL, simbol, dan normalisasi huruf kecil), dilanjutkan dengan ekstraksi fitur menggunakan TF-IDF dan algoritma DBSCAN untuk clustering. Otomatisasi ini mempercepat proses dan meminimalkan kesalahan manual.

3. Pendekatan *Non-Supervised (Unsupervised Learning)*

Penggunaan DBSCAN sebagai algoritma clustering memungkinkan sistem untuk mendeteksi pola kelompok tweet tanpa membutuhkan label sebelumnya (ground truth), yang ideal untuk data real-time atau data sosial media yang belum diklasifikasi.

4. Kemampuan Deteksi Pola Bullying

Dengan memasukkan kamus kata-kata kasar atau ofensif dari file `bully_keywords.txt`, sistem dapat secara langsung menandai tweet yang berpotensi mengandung unsur bullying. Hal ini meningkatkan nilai praktis sistem dalam konteks forensik digital dan kajian sosial media.

5. Visualisasi Hasil Clustering

Sistem menyertakan visualisasi berbasis PCA dua dimensi dan wordcloud, yang mendukung analisis eksploratif serta memudahkan interpretasi hasil oleh pengguna atau peneliti.

6. Skalabilitas dan Portabilitas

Karena sistem dibangun menggunakan Python, PHP, dan MySQL, maka mudah diimplementasikan di berbagai platform dan dapat diskalakan lebih lanjut untuk mendukung analisis real-time maupun batch-processing yang lebih besar.

b. Kekurangan Sistem

1. Ketergantungan pada Kata Kunci Tetap

Pendeteksian tweet bullying saat ini sangat bergantung pada daftar kata kunci statis (`bully_keywords.txt`). Hal ini dapat menyebabkan sistem melewatkan variasi bentuk ujaran kasar yang tidak tercantum atau menghasilkan false positive jika konteksnya salah.

2. Belum Memanfaatkan Model NLP Lanjutan

Sistem tidak memanfaatkan pendekatan NLP modern seperti word embeddings, LSTM, atau transformer-based models (misalnya BERT), yang dapat meningkatkan akurasi dalam memahami konteks teks secara semantik.

3. Sensitivitas Parameter DBSCAN

Keakuratan hasil clustering sangat tergantung pada nilai parameter ϵ dan minPts . Tidak adanya metode optimasi otomatis (seperti k-distance graph) membuat pemilihan parameter bersifat subjektif dan bisa tidak optimal.

4. Kualitas *Preprocessing* Masih Dasar

Preprocessing terbatas pada pembersihan simbol dan lowercase, tanpa melakukan tokenisasi lanjutan, stopwords removal yang disesuaikan, ataupun stemming/lemmatization, sehingga fitur yang diekstraksi mungkin masih mengandung noise.

5. Evaluasi Terbatas pada *Silhouette Score*

Sistem hanya mengevaluasi kualitas clustering berdasarkan Silhouette Score, yang bisa bias terhadap bentuk cluster tertentu. Evaluasi yang lebih lengkap seharusnya juga mempertimbangkan metrik lain seperti Davies-Bouldin Index atau bahkan validasi manual jika ground truth tersedia.

BAB V

PENUTUP

5.1 Kesimpulan

Berdasarkan hasil penelitian dan uji coba yang dilakukan dalam menggunakan metode DBSCAN dalam mengelompokkan pola bullying, berikut kesimpulan yang bisa diambil :

1. Berdasarkan hasil clustering menggunakan algoritma DBSCAN, sistem berhasil mengelompokkan tweet ke dalam beberapa cluster berdasarkan kemiripan isi. Nilai Silhouette Score sebesar 0.8698 menunjukkan bahwa pemisahan antar cluster memiliki tingkat kohesi yang cukup baik. Selain itu, terdeteksi beberapa data sebagai noise, yang menandakan keberhasilan algoritma dalam mengidentifikasi outlier atau anomali.
2. Distribusi tweet bullying juga menunjukkan variasi antar cluster, yang memperkuat efektivitas representasi teks dalam membedakan karakteristik tweet. Secara keseluruhan, metode ini terbukti mampu melakukan deteksi awal terhadap potensi konten negatif di media sosial dengan pendekatan unsupervised.
3. Tahapan pra-pemrosesan data seperti pembersihan teks, normalisasi, dan pengambilan fitur berkontribusi signifikan terhadap akurasi clustering. Langkah ini menjamin bahwa algoritma berfungsi dengan data yang lebih bersih dan mencerminkan dengan baik.
4. Penetapan parameter eps dan MinPts memiliki dampak besar terhadap hasil pengelompokan. Dengan melakukan percobaan menggunakan

berbagai kombinasi nilai parameter, terungkap bahwa pemilihan parameter yang sesuai dapat meningkatkan kualitas cluster yang dihasilkan.

5. Dari segi implementasi, metode ini cocok untuk digunakan pada sistem pemantauan media sosial secara real-time, karena mampu secara otomatis mengelompokkan dan memfilter konten berdasarkan pola kemiripan dan kepadatan data.

5.2 Saran

Berdasarkan hasil pengujian menggunakan algoritma DBSCAN, terdapat beberapa saran yang dapat diimplementasikan untuk meningkatkan efektivitas program. Setelah diperoleh kesimpulan maka beberapa saran yang bisa di jadikan masukan yaitu :

1. Meski nilai Silhouette Score sudah cukup memuaskan, eksperimen lebih lanjut dengan variasi nilai eps dan MinPts yang lebih beragam bisa dilakukan untuk menemukan pengaturan optimal yang mungkin menghasilkan cluster yang lebih stabil dan bermakna
2. Hasil pengelompokan DBSCAN dapat dimanfaatkan sebagai komponen dari sistem pemantauan otomatis untuk mengidentifikasi konten negatif secara real time. Dianjurkan untuk membangun sistem pipeline yang terintegrasi sehingga hasil analisis dapat segera dan efisien ditindaklanjuti.
3. Gunakan DBSCAN pada data tweet yang dikelompokkan berdasarkan waktu (seperti harian atau mingguan) untuk mengamati tren atau peningkatan kasus bullying dari waktu ke waktu.

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. <https://doi.org/10.14710/j.gauss.v11i1.33992>
- Arista, A., & Firmansyah, B. (2022). Perancangan Sistem Informasi Pemesanan Menu Makanan dan Minuman Berbasis WEB. *Junif: Jurnal Nasional Informatika*, 3(1), 36–41.
- Aurin, K., Fajar, M., & Munir, A. (2021). Pemodelan Jaringan Sensor Nirkabel Menggunakan System Modeling Language. *Jtriste*, 8(1), 9–20.
- Azizah, N., & Saptono, H. (2020). Analisa dan perbandingan performa RDBMS MySQL dan Hive-Hadoop. *Jurnal Informatika Terpadu*, 6(1), 20–28. <https://doi.org/10.54914/jit.v6i1.216>
- Fadhlillah, R., Dwian, B., Zamzami, F., Karim, G. A., A, K. L., Irawati, F. D., & Permana, R. D. (2024). Analisis Klasterisasi Kepadatan Penduduk Kabupaten Muara Enim Menggunakan Algoritma DBSCAN. 2024(Senada), 982–992.
- Faridzi, S. Al, Faza Shafa Azizah, Mustafa, F., Nindya Putri, A., Ramadhika, G., Rizky Aditya, F., Sherli Fadilah, R., Habibi, Y., Sutrisno, M., Jumail, J., Dewi Risanty, R., & Rosanti, N. (2024). Pengolahan Data: Pemahaman Gempa Bumi Di Indonesia Melalui Pendekatan Data Mining. *Jurnal Pengabdian Kolaborasi Dan Inovasi IPTEKS*, 2(1), 262–270. <https://doi.org/10.59407/jpki2.v2i1.506>
- Firdaus, A., Firdaus, W. I., Studi, P., Informatika, T., Digital, M., & Sriwijaya, P. N. (2021). *Text Mining*. 13(1), 66–78.
- Haryanto, B., & Haryanto, B. (2024). Pengenalan Database NOSQL dan Perbandinganya dengan Database Relasional. *Insan Pembangunan Sistem Informasi Dan Komputer (IPSIKOM)*, 12(1), 1–7. <https://doi.org/10.58217/ipsikom.v12i1.272>
- Hayat, M. A., Subroto, P., Kurniawan, M. I., & ... (2021). KONTRIBUSI KOMUNIKASI MASSA INSTANSI PEMERINTAH MENGHADAPI PANDEMI COVID-19 DI KABUPATEN KAPUAS. In *Journal of Syntax ...*. researchgate.net. https://www.researchgate.net/profile/Popo-Subroto-2/publication/360064987_Kontribusi_Komunikasi_Massa_Instance_Pemerintah_Menghadapi_Pandemi_Covid-19_Di_Kabupaten_Kapuas/links/625fb8d9709c5c2adb899f0b/Kontribusi-Komunikasi-Massa-Instansi-Pemerintah-Mengha
- 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), 2–9.

<https://doi.org/10.12962/j23373520.v8i2.49094>

- Malabay. (2021). Pemanfaatan Flowchart Untuk Kebutuhan Deskripsi Proses Bisnis. *Jurnal Ilmu Komputer*, 12(1), 21–26. <https://digilib.esaunggul.ac.id/pemanfaatan-flowchart-untuk-kebutuhan-deskripsi-proses-bisnis-9347.html>
- Maria Marlince Weripang, Indri Anugrah Rahmadhani, & Matahari Matahari. (2024). Rancang Bangun Website Berbasis CMS Wordpress Versi 6.3 di Kampung Kalamanuk Distrik Fakfak Timur Tengah. *Neptunus: Jurnal Ilmu Komputer Dan Teknologi Informasi*, 2(2), 102–113. <https://doi.org/10.61132/neptunus.v2i2.113>
- Mira Orisa, Ahmad Faisol, & Mochammad Ibrahim Ashari. (2023). Perancangan Website Company Profile Menggunakan Design Science Research Methodology (Dsrn). *Jurnal Informatika Teknologi Dan Sains*, 5(1), 160–164. <https://doi.org/10.51401/jinteks.v5i1.2576>
- Mutiah, S., Hasnataeni, Y., Fitrianto, A., & Jumansyah, L. M. R. D. (2024). *Perbandingan Metode Klastering K-Means dan DBSCAN dalam Identifikasi Kelompok Rumah Tangga Berdasarkan Fasilitas Sosial Ekonomi di Jawa Barat Dalam era digital saat ini , jumlah data yang tersedia dari berbagai bidang , termasuk sosial dan ekonomi , terus. 09(September)*, 247–260.
- Novieta, L., & Fitriani, F. (2021). *PENERAPAN DENSITY-BASED SPATIAL CLUSTERING OF APPLICATIONS WITH NOISE (DBSCAN) SEBAGAI TEXT CLUSTERING PADA AKUN TWITTER LAYANAN PROVIDER*.
- Pamulang, U. (2024). *Praxis : Jurnal Pengabdian Kepada Masyarakat*. 4(3), 36–43.
- Parlaungan S., T. F., & Wisnu, D. (2020). Rancang Bangun Sistem Pengidentifikasi Travel Bag Pada Kelompok Biro Perjalanan Umroh/Haji Berbasis Web. *Jurnal Teknologi Dan Komunikasi STMIK Subang*, 13(1), 26–40. <https://doi.org/10.47561/a.v13i1.167>
- Qifli Ilhamdi, J., Julkarnain, M., & Yuliadi, Y. (2024). Sistem Informasi Administrasi Ukm Racana Olat Maras – Ai Renung Universitas Teknologi Sumbawa Berbasis Website. *JATI (Jurnal Mahasiswa Teknik Informatika)*, 8(3), 4309–4315. <https://doi.org/10.36040/jati.v8i3.9905>
- Ramdany, S. (2024). Penerapan UML Class Diagram dalam Perancangan Sistem Informasi Perpustakaan Berbasis Web. *Journal of Industrial and Engineering System*, 5(1). <https://doi.org/10.31599/2e9afp31>
- Reski Amalia Salam, Tasrif Akib, & Muhammad Dahlan. (2024). Analisis Wacana Kritis terhadap Sarkasme dalam Twitter Sejak Bulan September-November 2023. *Jurnal Onoma: Pendidikan, Bahasa, Dan Sastra*, 10(2), 1855–1863. <https://doi.org/10.30605/onoma.v10i2.3621>
- Rian Alfianto. (2024). *Trending Nomor 1 di X atau Twitter, Netizen Ungkap Bukti-bukti Baru Gibran Rakabuming Merupakan Pemilik Akun Kaskus Fufufafa*. 17 September. <https://www.jawapos.com/oto-dan-tekho/015096965/trending->

nomor-1-di-x-atau-twitter-netizen-ungkap-bukti-bukti-baru-gibran-rakabuming-merupakan-pemilik-akun-kaskus-fufufafa

- Rizqi Nandadita Pamungkas, Permadi, D., & Florina, I. D. (2024). Strategi Humor Gibran Rakabuming dalam Komunikasi Politik di Media Sosial X (Twitter). *Jurnal Pemerintahan Dan Politik*, 9(3), 175–182. <https://doi.org/10.36982/jpg.v9i3.4057>
- Saddhono, K., & Maret, U. S. (2024). *Potensi pelanggaran uu ite pada komentar wagranel di akun gibran rakabuming raka: kajian linguistik forensik*. 19, 310–321.
- Sandfreni, S., Ulum, M. B., & Azizah, A. H. (2021). Analisis Perancangan Sistem Informasi Pusat Studi Pada Fakultas Ilmu Komputer Universitas Esa Unggul. *Sebatik*, 25(2), 345–356. <https://doi.org/10.46984/sebatik.v25i2.1587>
- Syahira, M. A., & Kurniawan, R. (2024). Analisis Sentimen Cyberbullying Pada Media Sosial X Menggunakan Metode Support Vector Machine. *Jurnal Media Informatika Budidarma*, 8(3), 1724. <https://doi.org/10.30865/mib.v8i3.7926>
- Wikantari, M. M., Sibaroni, Y., & Ihsan, A. F. (2023). Clustering Content Types and User Motivation Using DBSCAN on Twitter. *Journal of Computer System and Informatics (JoSYC)*, 4(4), 741–748. <https://doi.org/10.47065/josyc.v4i4.3750>
- Yusuf, S. A., & Khasanah, U. (2021). Kajian literatur dan teori sosial dalam penelitian. In *Metode penelitian ekonomi syariah*. researchgate.net. https://www.researchgate.net/profile/Ismail-Wekke/publication/344211215_Metode_Penelitian_Ekonomi_Syariah/links/5f5c12bc4585154dbbcb2f12/Metode-Penelitian-Ekonomi-Syariah.pdf#page=89

Lampiran

1. Surat Penetapan Dosen Pembimbing



MAJELIS PENDIDIKAN TINGGI PENELITIAN & PENGEMBANGAN PIMPINAN PUSAT MUHAMMADIYAH
UNIVERSITAS MUHAMMADIYAH SUMATERA UTARA
FAKULTAS ILMU KOMPUTER DAN TEKNOLOGI INFORMASI

UMSU Terakreditasi A Berdasarkan Keputusan Badan Akreditasi Nasional Perguruan Tinggi No. 89/SK/BAN-PT/Akred/PT/III/2019
Pusat Administrasi: Jalan Mukhtar Basri No. 3 Medan 20238 Telp. (061) 6622400 - 66224567 Fax. (061) 6625474 - 6631003
<https://fiki.umsu.ac.id> fiki@umsu.ac.id [f umsumedan](#) [ig umsumedan](#) [t umsumedan](#) [u umsumedan](#)

PENETAPAN DOSEN PEMBIMBING
PROPOSAL/SKRIPSI MAHASISWA
NOMOR : 907/IL.3-AU/UMSU-09/F/2024

Assalamu'alaikum Warahmatullahi Wabarakatuh

Dekan Fakultas Ilmu Komputer dan Teknologi Informasi Universitas Muhammadiyah Sumatera Utara, berdasarkan Persetujuan permohonan judul penelitian Proposal / Skripsi dari Ketua / Sekretaris.

Program Studi : Sistem Informasi
Pada tanggal : 20 November 2024

Dengan ini menetapkan Dosen Pembimbing Proposal / Skripsi Mahasiswa.

Nama : Mutiara Santi
NPM : 2109010095
Semester : VII (Tujuh)
Program studi : Sistem Informasi
Judul Proposal / Skripsi : Analisis Dan Implementasi DB Scam Dalam Clusterisasi Pola Media Sosial

Dosen Pembimbing : Martiano, S.Pd., S.Kom., M.Kom.

Dengan demikian di izinkan menulis Proposal / Skripsi dengan ketentuan

1. Penulisan berpedoman pada buku panduan penulisan Proposal / Skripsi Fakultas Ilmu Komputer dan Teknologi Informasi UMSU
2. Pelaksanaan Sidang Skripsi harus berjarak 3 bulan setelah dikeluarkannya Surat Penetapan Dosen Pembimbing Skripsi.
3. **Proyek Proposal / Skripsi** dinyatakan " **BATAL** " bila tidak selesai sebelum Masa Kadaluarsa tanggal : 20 November 2025
4. Revisi judul.....

Wassalamu'alaikum Warahmatullahi Wabarakatuh.

Ditetapkan di : Medan
Pada Tanggal : 18 Jumadil Awwal 1446 H
20 November 2024 M



Dekan
Dr. At-Rohwarizmi, S.Kom., M.Kom.
NIDN : 0127099201

Cc. File



2. Surat Persetujuan Topik Judul Penelitian



UMSU

Unggul | Cerdas | Terpercaya
Bila menjabar surat ini agar disebutkan nomor dan tanggalnya

MAJELIS PENDIDIKAN TINGGI PENELITIAN & PENGEMBANGAN PIMPINAN PUSAT MUHAMMADIYAH
UNIVERSITAS MUHAMMADIYAH SUMATERA UTARA
FAKULTAS ILMU KOMPUTER DAN TEKNOLOGI INFORMASI

UMSU Terakreditasi A Berdasarkan Keputusan Badan Akreditasi Nasional Perguruan Tinggi No. 89/SK/BAN-PT/Akred/PT/III/2019
Pusat Administrasi: Jalan Mukhtar Basri No. 3 Medan 20238 Telp. (061) 6622400 - 66224567 Fax. (061) 6625474 - 6631003
<https://fiki.umsu.ac.id> fiki@umsu.ac.id [f umsumedan](#) [ig umsumedan](#) [t umsumedan](#) [v umsumedan](#)

PERSETUJUAN TOPIK/JUDUL PENELITIAN

Nomor Agenda :
Nama : MUTIARA SANTI
NPM : 2109010095
Tanggal Persetujuan :
Topik Yang Disetujui Program Studi :
Nama Dosen Pembimbing : MARTIANU, S.Pd, S.kom., M.kom
Judul Yang Disetujui Dosen Pembimbing : ANALISIS DAN IMPLEMENTASI DISCOW DALAM
CLUSTERISASI POLA BULLYING MEDIA SOSIAL TWITTER
PADA AKUN GIBRAN RAKABUMING

Medan 18/2/2025

Disahkan oleh
Ketua Program Studi
Sistem Informasi

(.....)

Persetujuan
Dosen Pembimbing

(.....)



3. Cek Turnitin

Skripsimutiara.pdf

ORIGINALITY REPORT

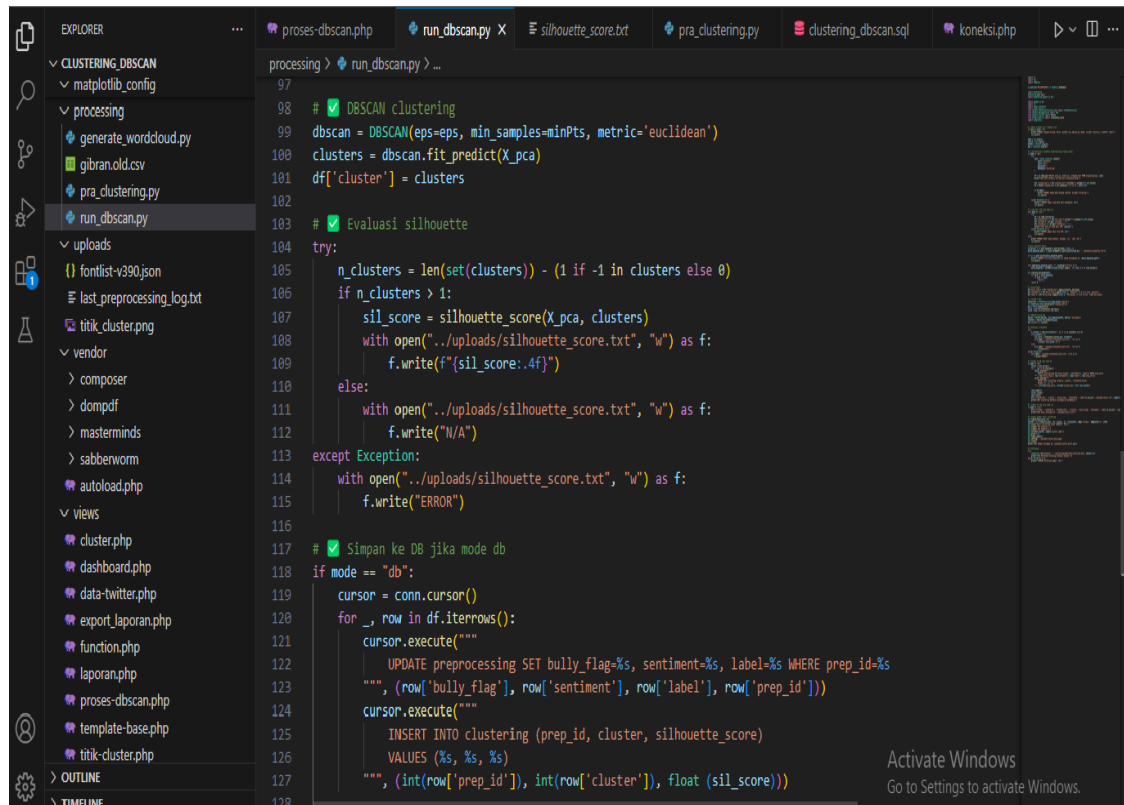
28%	24%	12%	18%
SIMILARITY INDEX	INTERNET SOURCES	PUBLICATIONS	STUDENT PAPERS

PRIMARY SOURCES

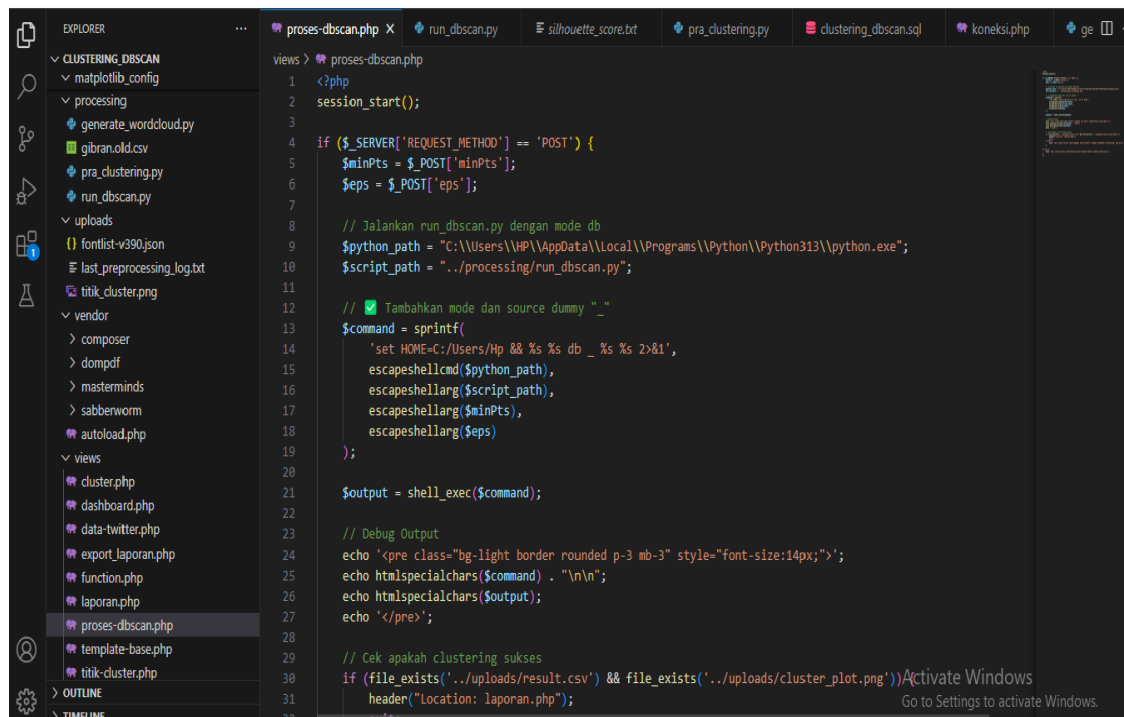
1	repository.umsu.ac.id Internet Source	3%
2	digilib.unila.ac.id Internet Source	2%
3	prosiding.unipma.ac.id Internet Source	1%
4	Submitted to Universitas Muria Kudus Student Paper	1%
5	repository.universitasbumigora.ac.id Internet Source	1%
6	smart.stmikplk.ac.id Internet Source	1%
7	repository.bsi.ac.id Internet Source	1%
8	journals.upi-yai.ac.id Internet Source	1%
9	www.scribd.com Internet Source	1%

10	journals.upi-yai.ac.id Student Paper	1%
11	ejurnal.dipaneegara.ac. Internet Source	1%
12	repository.bsi.ac.id Internet Source	1%
13	repository.uts.ac.id Internet Source	1%
14	Submitted to Universitas Muslim Indonesia Internet Source	1%
15	ejournal.unesa.ac.id Internet Source	1%
16	repository.dinamika.ac.id Student Paper	1%
17	Submitted to unimal Student Paper	1%
18	repository.sari-mutiara.ac.id Internet Source	1%
19	www.majid.web.id Internet Source	1%
20	text-id.123dok.com Internet Source	<1%
21	Submitted to Universitas Putera Batam Student Paper	<1%

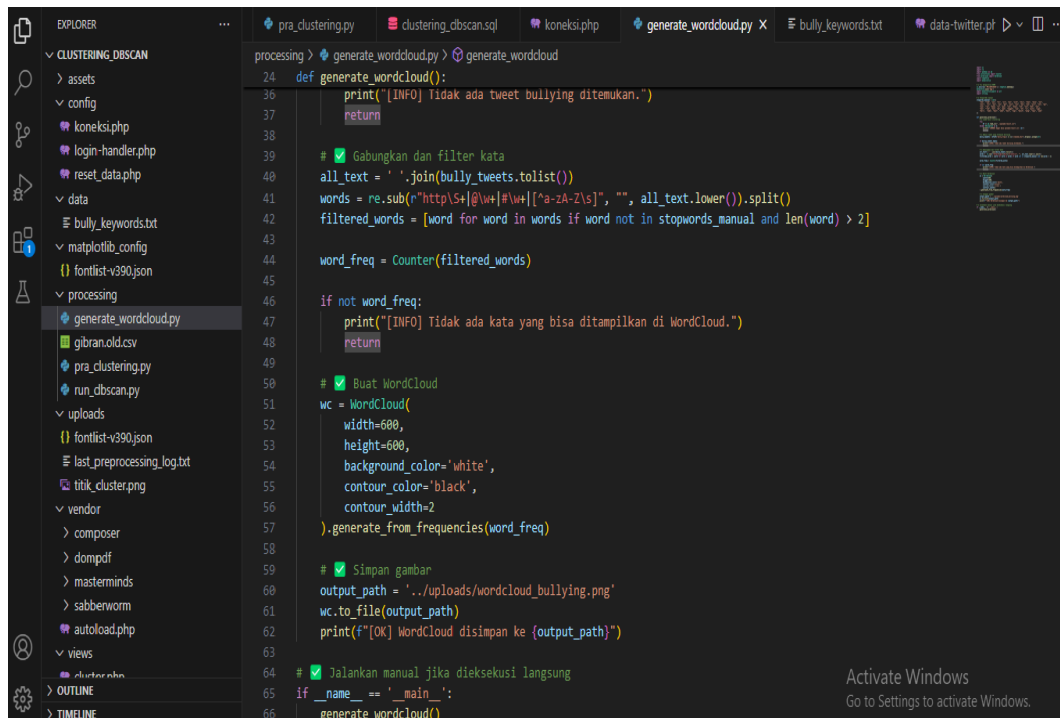
4. Codingan



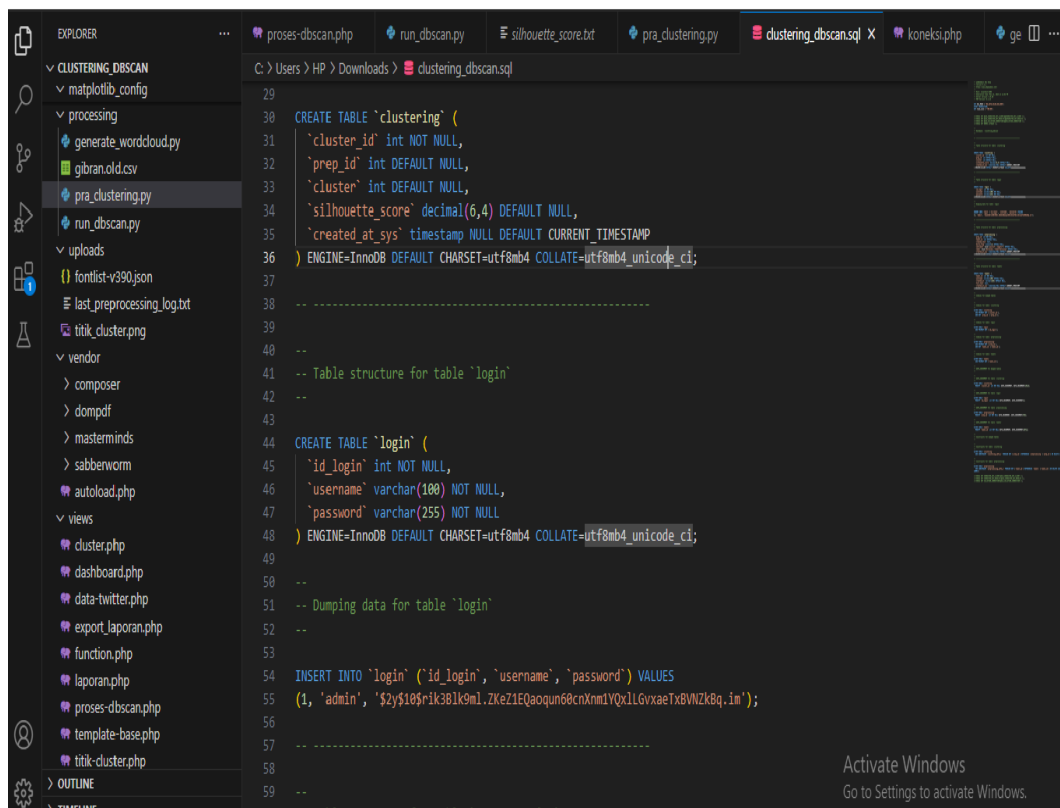
```
processing > run_dbscan.py > ...
97
98 # DBSCAN clustering
99 dbscan = DBSCAN(eps=eps, min_samples=minPts, metric='euclidean')
100 clusters = dbscan.fit_predict(X_pca)
101 df['cluster'] = clusters
102
103 # Evaluasi silhouette
104 try:
105     n_clusters = len(set(clusters)) - (1 if -1 in clusters else 0)
106     if n_clusters > 1:
107         sil_score = silhouette_score(X_pca, clusters)
108         with open("../uploads/silhouette_score.txt", "w") as f:
109             f.write(f"{sil_score:.4f}")
110     else:
111         with open("../uploads/silhouette_score.txt", "w") as f:
112             f.write("N/A")
113 except Exception:
114     with open("../uploads/silhouette_score.txt", "w") as f:
115         f.write("ERROR")
116
117 # Simpan ke DB jika mode db
118 if mode == "db":
119     cursor = conn.cursor()
120     for _, row in df.iterrows():
121         cursor.execute("""
122             UPDATE preprocessing SET bully_flag=%s, sentiment=%s, label=%s WHERE prep_id=%s
123             """, (row['bully_flag'], row['sentiment'], row['label'], row['prep_id']))
124         cursor.execute("""
125             INSERT INTO clustering (prep_id, cluster, silhouette_score)
126             VALUES (%s, %s, %s)
127             """, (int(row['prep_id']), int(row['cluster']), float(sil_score)))
128
```



```
views > proses-dbscan.php
1 <?php
2 session_start();
3
4 if ($_SERVER['REQUEST_METHOD'] == 'POST') {
5     $minPts = $_POST['minPts'];
6     $eps = $_POST['eps'];
7
8     // Jalankan run_dbscan.py dengan mode db
9     $python_path = "C:\\Users\\VHP\\AppData\\Local\\Programs\\Python\\Python313\\python.exe";
10    $script_path = "../processing/run_dbscan.py";
11
12    // Tambahkan mode dan source dummy ""
13    $command = sprintf(
14        'set HOME=C:/Users/Hp && %s %s db _ %s %s 2>&1',
15        escapeshellcmd($python_path),
16        escapeshellarg($script_path),
17        escapeshellarg($minPts),
18        escapeshellarg($eps)
19    );
20
21    $output = shell_exec($command);
22
23    // Debug Output
24    echo '<pre class="bg-light border rounded p-3 mb-3" style="font-size:14px;">';
25    echo htmlspecialchars($command) . "\n\n";
26    echo htmlspecialchars($output);
27    echo '</pre>';
28
29    // Cek apakah clustering sukses
30    if (file_exists('../uploads/result.csv') && file_exists('../uploads/cluster_plot.png')){
31        header("Location: laporan.php");
32        exit;
33    }
34}
```



```
processing > generate_wordcloud.py > generate_wordcloud
24 def generate_wordcloud():
36     print("[INFO] Tidak ada tweet bullying ditemukan.")
37     return
38
39     # Gabungkan dan filter kata
40     all_text = ' '.join(bully_tweets.tolist())
41     words = re.sub(r"http\S+|@\w+|#\w+|['\"a-zA-Z\s]", "", all_text.lower()).split()
42     filtered_words = [word for word in words if word not in stopwords_manual and len(word) > 2]
43
44     word_freq = Counter(filtered_words)
45
46     if not word_freq:
47         print("[INFO] Tidak ada kata yang bisa ditampilkan di WordCloud.")
48         return
49
50     # Buat WordCloud
51     wc = WordCloud(
52         width=600,
53         height=600,
54         background_color='white',
55         contour_color='black',
56         contour_width=2
57     ).generate_from_frequencies(word_freq)
58
59     # Simpan gambar
60     output_path = '../uploads/wordcloud_bullying.png'
61     wc.to_file(output_path)
62     print(f"[OK] WordCloud disimpan ke {output_path}")
63
64     # Jalankan manual jika dieksekusi langsung
65 if __name__ == '__main__':
66     generate_wordcloud()
```



```
C:\Users\HP > Downloads > clustering_dbscan.sql
29
30 CREATE TABLE `clustering` (
31   `cluster_id` int NOT NULL,
32   `prep_id` int DEFAULT NULL,
33   `cluster` int DEFAULT NULL,
34   `silhouette_score` decimal(6,4) DEFAULT NULL,
35   `created_at_sys` timestamp NULL DEFAULT CURRENT_TIMESTAMP
36 ) ENGINE=InnoDB DEFAULT CHARSET=utf8mb4 COLLATE=utf8mb4_unicode_ci;
37
38 -----
39
40 --
41 -- Table structure for table `login`
42 --
43
44 CREATE TABLE `login` (
45   `id_login` int NOT NULL,
46   `username` varchar(100) NOT NULL,
47   `password` varchar(255) NOT NULL
48 ) ENGINE=InnoDB DEFAULT CHARSET=utf8mb4 COLLATE=utf8mb4_unicode_ci;
49
50 --
51 -- Dumping data for table `login`
52 --
53
54 INSERT INTO `login` (`id_login`, `username`, `password`) VALUES
55 (1, 'admin', '$2y$10$frk3B1k9ml.ZKeZIEQaoqun60cnXnm1VQx1LGvxaeTx8VNzk8q.im');
56
57 -----
58
59 --
60 -- Table structure for table `proses_dbscan`
61 --
```