

**ANALISIS METODE *CLUSTERING* DENGAN ALGORITMA
SPECTRAL CLUSTERING DALAM PENGELOMPOKAN
TINGKAT TINDAK KRIMINALITAS**

SKRIPSI

EGI PRATAMA

198160027



PROGRAM STUDI TEKNIK INFORMATIKA

FAKULTAS TEKNIK

UNIVERSITAS MEDAN AREA

MEDAN

2023

UNIVERSITAS MEDAN AREA

© Hak Cipta Di Lindungi Undang-Undang

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

Document Accepted 5/6/24

Access From (repository.uma.ac.id)5/6/24

HALAMAN PENGESAHAN

HALAMAN PENGESAHAN

Judul Skripsi : Analisis Metode *Clustering* Dengan Algoritma *Spectral Clustering* Dalam Pengelompokan Tingkat Tindak Kriminalitas

Nama : Egi Pratama

NPM : 198160027

Fakultas : Teknik

Prodi : Teknik Informatika

Disetujui Oleh :

Pembimbing

Dr. Rahmad Syah, S.Kom., M.Kom

NIDN : 0105058804

Diketahui Oleh :

Dekan Fakultas Teknik



Eng. R. Priatno, ST., MT

NIDN : 0102027402

Ketua Prodi Teknik Informatika



Rizki Moliono, S.Kom., M.Kom

NIDN : 0109038902

Tanggal Lulus : 21 Desember 2023

HALAMAN PERNYATAAN

HALAMAN PERNYATAAN

Saya menyatakan bahwa skripsi yang saya susun, sebagai syarat memperoleh gelar sarjana merupakan hasil karya tulis saya sendiri. Adapun bagian-bagian tertentu dalam penulisan skripsi ini yang saya kutip dari hasil karya orang lain telah dituliskan sumbernya secara jelas sesuai dengan norma, kaidah, dan etika penulisan ilmiah. Saya bersedia menerima sanksi pencabutan gelar akademik yang saya peroleh dan sanksi-sanksi lainnya dengan peraturan yang berlaku, apabila dikemudian hari ditemukan adanya plagiat dalam skripsi ini.

Medan, 21/ Desember/ 2023



Egi Pratama

198160027

iii

HALAMAN PERNYATAAN PERSETUJUAN PUBLIKASI TUGAS AKHIR/SKRIPSI/TESIS UNTUK KEPENTINGAN AKADEMIS

HALAMAN PERNYATAAN PERSETUJUAN PUBLIKASI TUGAS AKHIR/SKRIPSI/TESIS UNTUK KEPENTINGAN AKADEMIS

Sebagai civitas akademika Universitas Medan Area saya yang bertanda tangan di bawah ini.

Nama : Egi Pratama

NPM : 198160027

Program Studi: Teknik Informatika

Fakultas : Teknik

Jenis Karya : Skripsi

Demi pengembangan ilmu pengetahuan, menyetujui untuk memberikan kepada Universitas Medan Area Hak Bebas *Royalti Noneksklusif (Non-exclusive Royalti Free Right)* atas karya ilmiah saya yang berjudul :

Analisis Metode *Clustering* dengan Algoritma *Spectral Clustering* dalam Pengelompokan Tingkat Tindak Kriminalitas

Berserta perangkat yang ada (jika diperlukan). Dengan Hak Bebas *Royalti Noneksklusif* ini Universitas Medan Area berhak menyimpan, mengalihmedia/format-kan, mengelola dalam bentuk pangkalan data (*database*), merawat, dan mempublikasikan tugas akhir/skripsi/tesis saya selama tetap mencantumkan nama saya sebagai penulis/pencipta dan sebagai pemilik Hak Cipta.

Demikian pertanyaan ini saya buat dengan sebenarnya.

Dibuat di : Medan

Pada tanggal : 21/ Desember/ 2023

Yang menyatakan



(Egi Pratama)

ABSTRAK

Kriminalitas merupakan tindakan yang mengganggu ketertiban dan kenyamanan umum, kriminalitas dianggap sebagai tindakan yang negatif karena merugikan korban baik secara fisik ataupun secara mental. Di Indonesia sendiri terdapat berbagai macam tindakan kriminalitas, mulai dari tindakan yang sangat berbahaya seperti pembunuhan, pencabulan. Kemudian tindakan yang dianggap berbahaya seperti pencurian, serta tindakan yang dianggap cukup berbahaya yaitu tindakan kejahatan seperti kerusuhan umum, demonstrasi. Tindakan kriminalitas yang meradang perlu untuk dianalisis untuk kemudian dapat dicegah sesuai dengan kapasitas hukum dan tindakan pencegahan. Penelitian ini menggunakan data dari masyarakat Indonesia yang mengeluh di *twitter* tentang tindakan kriminalitas yang marak terjadi, menggunakan 10.000 sampel data yang dikumpulkan pada tanggal 25 Juni 2023 menggunakan teknik data mining sebagai bahan analisis dan menggunakan metode *spectral clustering* sebagai algoritma analisis untuk mendapatkan hasil klasterisasi tindakan kejahatan. Data dianalisis menjadi tiga kluster yaitu kluster sangat berbahaya dengan presentasi 90.3% atau 9026 data, kemudian kluster berbahaya dengan presentasi 7% atau 697 data, dan kluster cukup berbahaya dengan presentasi 2.8% atau 277 data.

Kata Kunci: Kriminalitas, *Data Mining*, *Spectral Clustering*, *Clustering*

ABSTRACT

Crime is an act that disturbs public order and comfort, crime is considered a negative action because it harms the victim both physically and mentally. In Indonesia itself there are various kinds of criminal acts, ranging from very dangerous acts such as murder, obscenity. Then actions that are considered dangerous, such as theft. As well as actions that are considered quite dangerous, namely acts of crime such as public riots, demonstrations. Inflammatory criminal acts need to be analyzed so that they can be prevented in accordance with legal capacity and preventive measures. This study uses data from Indonesian people who complain on tweeters about criminal acts that are rife, uses 10,000 samples of data collected on June 25, 2023 using data mining techniques as material for analysis and uses the spectral clustering method as an analytical algorithm to get the results of clustering crimes. The data were analyzed into three clusters, namely the very dangerous cluster with a presentation of 90.3 % or 9026 data, then the dangerous cluster with a presentation of 7% or 697 data, and the moderately dangerous cluster with a presentation of 2.8% or 277 data.

Keyword: *Crime, Data Mining, Spectral Clustering, Clustering*

RIWAYAT HIDUP

Penulis lahir di Desa Silo Bonto pada tanggal 26 Januari 2001 dari Ayah Edy Suyanto dan Ibu Wagiyah. Penulis adalah anak pertama dari 3 (tiga) bersaudara

Penulis pertama kali mengenyam Pendidikan dibangku SD 016534 Lubuk Buaya pada tahun 2006-2012, meneruskan Pendidikan di Sekolah Menengah Pertama Negeri 2 Silau Laut diselesaikan pada tahun 2012-2015, Meneruskan Pendidikan Sekolah Menengah Kejuruan Negeri 1 Air Joman pada tahun 2015-2018.

Pada tahun 2018 penulis lulus dari SMKN 1 Air Joman dan pada 2019 terdaftar sebagai mahasiswa Fakultas Teknik Prodi Teknik Informatika Universitas Medan Area. Pada saat ini tahun 2023 penulis sedang menjalani semester 8 (Delapan) menyusun tugas akhir skripsi.

KATA PENGANTAR

Puji syukur penulis panjatkan kepada Tuhan Yang Maha Kuasa atas segala karuniaNya sehingga skripsi ini berhasil diselesaikan. Tema yang dipilih dalam penelitian ini dengan judul “Analisis Metode *Clustering* Dengan Algoritma *Spectral Clustering* Dalam Pengelompokan Tingkat Tindak Kriminalitas ”.

Skripsi ini merupakan salah satu syarat untuk menyelesaikan pendidikan untuk mencapai gelar sarjana di Program Studi Teknik Informatika Fakultas Teknik Universitas Medan Area. Pada kesempatan ini penulis mengucapkan terima kasih kepada :

1. Bapak Prof. Dr. Dadan Ramdan, M.Eng, M.Sc. selaku Rektor Universitas Medan Area.
2. Bapak Dr. Eng. Supriatno, ST., MT selaku Dekan Fakultas Teknik Universitas Medan Area.
3. Bapak Rizki Muliono, S.Kom., M.Kom selaku Kepala Program Studi Teknik Informatika Universitas Medan Area.
4. Bapak Dr. Rahmad Syah, S.Kom., M. Kom selaku Dosen pembimbing yang telah membantu penulis dari segi materi dan moril sehingga penulis dapat menyelesaikan skripsi ini.
5. Orang tua penulis yaitu Bapak Edy Suyanto dan Ibu Wagiyah yang telah mendoakan tiada henti dan memberikan semangat serta membantu penulis dalam segi materi dan moril sehingga penulis dapat menyelesaikan skripsi ini dengan sebaik baiknya.
6. Seluruh Dosen dan Staf Program Studi Teknik Informatika Universitas Medan Area.

7. Seluruh teman-teman yang sudah memberikan dukungannya selama penulisan proposal skripsi ini, khususnya teman-teman Teknik Informatika angkatan 2019.
8. Seluruh pihak yang tidak dapat disebutkan satu persatu yang membantu dalam penyelesaian skripsi ini.

Penulis menyadari bahwa penelitian ini masih memiliki kekurangan, oleh karena itu kritik dan saran yang bersifat membangun sangat penulis harapkan demi kesempurnaan penelitian ini. Penulis berharap tugas penelitian ini dapat bermanfaat baik kalangan pendidikan maupun masyarakat. Akhir kata penulis ucapkan terima kasih.

Medan, 21 Desember 2023

Penulis,

(Egi Pratama)
198160027

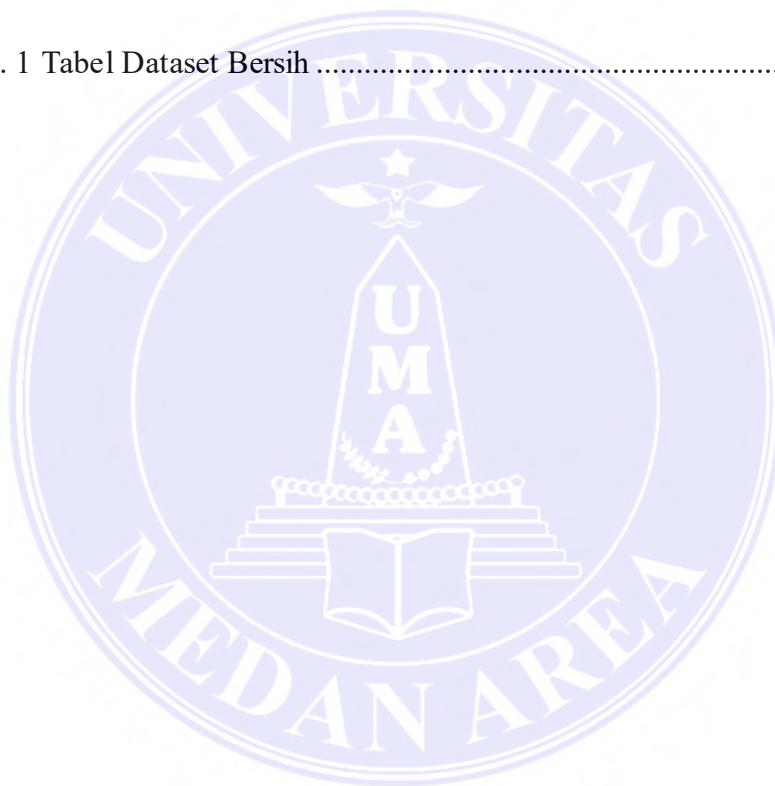
DAFTAR ISI

HALAMAN PENGESAHAN.....	ii
HALAMAN PERNYATAAN	iii
HALAMAN PERNYATAAN PERSETUJUAN PUBLIKASI TUGAS AKHIR/SKRIPSI/TESIS UNTUK KEPENTINGAN AKADEMIS	iv
ABSTRAK.....	v
<i>ABSTRACT</i>	vi
RIWAYAT HIDUP.....	vii
KATA PENGANTAR.....	viii
DAFTAR ISI.....	x
DAFTAR TABEL.....	xii
DAFTAR GAMBAR	xiii
BAB I PENDAHULUAN.....	1
1.1 Latar Belakang	1
1.2 Rumusan Masalah	4
1.3 Batasan Masalah.....	4
1.4 Tujuan Penelitian.....	5
1.5 Manfaat Penelitian.....	5
1.6 Sistematika Penulisan.....	6
BAB II TINJAUAN PUSTAKA.....	7
2.1 Sosial Media <i>Twitter</i>	7
2.2 <i>Data Mining Clustering</i>	7
2.2.1 Pengertian <i>Data Mining</i>	10
2.2.2 Tahapan <i>Data Mining</i>	11
2.2.3 <i>Scrapping Data</i>	15
2.2.4 Teknik <i>Data Mining Clustering</i>	16
2.2.5 <i>Spectral Clustering</i>	17
2.3 Kriminalitas	21
2.3.1 Pengertian Kriminalitas.....	21
2.3.2 Faktor Penyebab Kriminalitas.....	22
2.4 <i>Davies Boulden Index (DBI)</i>	23
2.5 Penelitian Terdahulu	25
BAB III METODE PENELITIAN.....	28

3.1	Sumber Data Penelitian	28
3.2	Bahan dan Alat.....	28
3.3	Populasi dan Sampel	29
3.4	Tingkat Tindak Kriminalitas	29
3.5	Tahapan Penelitian	32
3.6	Metode Pengumpulan Data.....	33
3.7	<i>Data Clean</i>	37
3.8	Metode Penelitian.....	38
3.8.1.	<i>Process Data</i>	38
3.8.2.	Evaluasi.....	39
BAB IV HASIL DAN PEMBAHASAN		41
4.1.	<i>Preprocess Data</i>	41
4.2.	<i>Process Data</i>	48
4.3.	Evaluasi	60
BAB V KESIMPULAN DAN SARAN.....		61
5.1	Kesimpulan.....	61
5.2	Saran.....	61
DAFTAR PUSTAKA		62
LAMPIRAN.....		67

DAFTAR TABEL

Tabel 2. 1 Penelitian Terdahulu	25
Tabel 3. 1 Kebutuhan Perangkat Keras	28
Tabel 3. 2 Kebutuhan Perangkat Lunak	29
Tabel 3. 3 Tabel <i>Query</i>	33
Tabel 3. 4 Tabel Dataset	33
Tabel 3. 5 Tabel Data Clean	37
Tabel 4. 1 Tabel Dataset Bersih	41



DAFTAR GAMBAR

Gambar 2. 1 Tahapan Data Mining	11
Gambar 3. 1 Tahapan Analisis Penelitian	32
Gambar 4. 1 Dataset Sebelum	48
Gambar 4. 2 Dataset Sesudah.....	48
Gambar 4. 3 Matriks Similaritas	48
Gambar 4. 4 <i>Similarity Graph</i>	49
Gambar 4. 5 Matriks Diagonal.....	50
Gambar 4. 6 <i>Diagonal Graph</i>	51
Gambar 4. 7 Nilai dan Vektor <i>Eigen</i>	52
Gambar 4. 8 <i>Eigen Value</i> dan <i>Eigen Vector</i>	53
Gambar 4. 9 Centroid	54
Gambar 4. 10 Jarak Data ke Pusat Kluster.....	55
Gambar 4. 11 Pusat Kluster.....	55
Gambar 4. 12 Distribusi Kluster	56
Gambar 4. 13 Distribusi Kluster	57
Gambar 4. 14 Jumlah Kejahatan Dalam Kluster.....	58
Gambar 4. 15 Jumlah Data Pada Kluster Terlabel.....	58
Gambar 4. 16 Tingkat Kejahatan Teratas Pada Kluster.....	59
Gambar 4. 17 Ukuran Kluster Terlabel	59

BAB I

PENDAHULUAN

1.1 Latar Belakang

Media sosial mengacu pada teknologi interaktif yang memungkinkan penciptaan dan berbagi informasi, ide, dan minat melalui komunitas dan jaringan virtual (Kamal, 2020). Platform media sosial seperti Facebook, Twitter, dan Instagram telah merevolusi cara orang terhubung dan berinteraksi satu sama lain. Twitter adalah platform media sosial yang memungkinkan penggunaannya untuk berbagi pesan singkat yang disebut *tweet* (Akbar & Sugiharto, 2023). Pada tahun 2023, Twitter memiliki sekitar 353,90 juta pengguna aktif bulanan di seluruh dunia. Twitter memiliki batasan 2.400 tweet per hari dan 500 pesan langsung yang dikirim per hari (Navisha et al., 2023). Manfaat twitter sangat banyak, termasuk konektivitas, hubungan global, pendidikan, dan merupakan sumber informasi dari suatu kejadian atau peristiwa seperti tindakan kriminalitas yang meresahkan masyarakat.

Kriminalitas adalah setiap dan semua perbuatan yang berdampak negatif secara ekonomi dan psikologis serta bertentangan dengan norma sosial dan agama serta hukum Indonesia. Tindak kriminalitas biasa berpengaruh terhadap keamanan masyarakat serta mengancam ketenangan lahir dan batinnya. Kesejahteraan dan ketentraman masyarakat kemungkinan akan terkena dampak negatif jika anggota masyarakat percaya bahwa keselamatan mereka terancam (Suriani, 2020). Baik pria maupun wanita biasanya terlibat dalam tindakan kriminal yang berpotensi menimbulkan kerugian bagi orang lain atau diri mereka sendiri, baik secara sadar

maupun tidak sadar. Menurut Premasundari dan Yamini (2019), berbagai macam kejahatan, termasuk pembunuhan, penyerangan, penipuan, dan lainnya, sedang meningkat. Kebutuhan mendasar meliputi kebutuhan sosial seperti peran sosial dan rasa aman, serta kebutuhan biologis seperti tidur, makan, dan minum. Oleh karena itu, rasa aman manusia berada pada risiko yang lebih besar seiring dengan meningkatnya tingkat kejahatan.

Teori menyatakan bahwa kejahatan terjadi ketika wilayah aktivitas korban atau sasaran tumpang tindih dengan wilayah aktivitas pelaku. Pola kejahatan memberikan pemahaman tentang mengapa tempat-tempat tertentu menjadi lokasi kejahatan (Ainsiyah et al., 2021). Manusia memilih dan melakukan kejahatan dengan kehendak bebas mereka yang selalu ingin mendapatkan kesenangan dan menghindari rasa sakit, berdasarkan tradisi klasik yang sudah ditinggalkan dalam kriminologi. Kejahatan adalah konsekuensi normal dari sifat manusia yang tidak terkontrol (Gottfredson and Hirschi, 2022). Ancaman utama terhadap ketenangan hidup manusia dalam menghadapi perubahan nilai-nilai sosial adalah kelanjutan tindak kriminal. Keamanan pribadi juga merupakan hak asasi manusia yang harus dijaga, oleh karena itu, kerja sama erat antara pemerintah dan masyarakat sangat penting dalam upaya mengurangi tingkat kejahatan di Indonesia (Sinaga & Anshori, 2022).

Tindakan kriminalitas dapat diartikan sebagai perilaku masyarakat yang menimbulkan ketidaknyamanan dan menghadirkan masalah sosial. Perbuatan kriminal ini tidak hanya dilakukan oleh orang dewasa, tetapi juga melibatkan remaja dan anak-anak. Faktor utama yang menyebabkan hal ini adalah masalah ekonomi, dimana masyarakat yang menghadapi keterbatasan dalam pemenuhan

kebutuhan atau mengalami pengangguran cenderung melakukan apa saja untuk memenuhi kebutuhan mereka. Selain itu, tindakan kriminal yang dilakukan oleh anak-anak atau remaja seringkali terjadi karena mereka terputus dari pendidikan akibat kesulitan biaya sekolah, yang menyebabkan mereka terlibat dalam kegiatan tersebut (Aprillia, 2022).

Tindak kriminalitas yang terjadi di Indonesia pada tahun 2021 adalah sejumlah 239.481 kejadian (Badan Pusat Statistik (BPS), 2022). Jumlah tersebut bukan jumlah yang kecil, ditambah tindak kriminalitas yang terjadi juga beragam. Tindak kejahatan terus meningkat, proses untuk clusterisasi tindak kejahatan menjadi kurang efisien karena kemungkinan kasus kriminal yang harus dihadapi semakin banyak.

Untuk mengatasi masalah ini, metode *spectral clustering* menjadi solusi yang dibutuhkan. Teknik pengelompokan yang disebut *spectral clustering* didasarkan pada kemiripan antar data dengan tahapan *spectral* yang kemudian dikombinasikan dengan metode *clustering* seperti menggunakan K-Means untuk mendapatkan pengelompokan data sesuai dengan kluster yang diinginkan. Hubungan antara berbagai data mengungkapkan kesamaan. Grafik akan dibuat dari data yang ada di *Spectral Clustering*. di mana masing-masing data membentuk simpul grafik. *Edge* terdiri dari hubungan antar data yang biasanya mewakili jarak antara dua *record* yang berhubungan (Li et al., 2020).

Dengan tujuan meningkatkan pemahaman tentang tingkat kejahatan di Indonesia melalui data dari *tweet* masyarakat Indonesia, penulis menggunakan metode *spectral clustering* untuk klasterisasi. Diharapkan bahwa penggunaan

metode ini akan membantu pihak Kepolisian dalam memperkuat langkah-langkah keamanan, pencegahan, dan penindakan kejahatan. Selain itu, kesadaran masyarakat tentang potensi kejahatan juga diharapkan semakin meningkat.

1.2 Rumusan Masalah

Berdasarkan dari latar belakang diatas, maka terdapat beberapa rumusan masalah yang dikaji dalam penelitian ini yaitu “Bagaimana Hasil Pengelompokan Tingkat Tindak Kriminalitas Menggunakan Metode *Spectral Clustering* di Indonesia”.

1.3 Batasan Masalah

Agar dalam penelitian dan penulisan tidak keluar dari tema permasalahan, maka dibutuhkan batasan-batasan masalah sebagai berikut:

1. Pengelompokan tingkat tindak kriminalitas di Indonesia.
2. Pengelompokkan menggunakan metode *spectral clustering* dan *K-Means* sebagai algoritma partisi.
3. Faktor yang diamati ditinjau dari tindakan kriminalitas berdasarkan *tweet* masyarakat Indonesia yang *discrapping* pada tanggal 25 Juni 2023.
4. Kluster dibagi berdasarkan tindakan kejahatan dari 3 level yaitu cukup berbahaya, berbahaya dan sangat berbahaya.
5. Implementasi pada penelitian ini menggunakan *Google Colaboratory* dan *Python* sebagai bahasa pemrograman.

1.4 Tujuan Penelitian

Berdasarkan dari latar belakang dan rumusan masalah diatas, maka tujuan dalam penyusunan tugas akhir ini yaitu “Bagaimana Hasil Pengelompokan Tingkat Tindak Kriminalitas Menggunakan Metode *Spectral Clustering* di Indonesia”.

1.5 Manfaat Penelitian

Adapun beberapa manfaat yang dapat diambil dari hasil penelitian ini yaitu sebagai berikut:

1. Manfaat Teoritis
 - a. Memberikan kontribusi terhadap ilmu pengetahuan, khususnya dalam bidang terkait untuk meningkatkan proses pembelajaran yang baik.
 - b. Hasil penelitian ini dapat menjadi acuan bagi penelitian lain yang fokus pada topik yang sama.
2. Manfaat Praktis
 - a. Bagi Peneliti dan Masyarakat Umum
Dapat menambah wawasan tentang analisis metode *clustering* dengan *spectral clustering* pengelompokan tingkat tindak kriminalitas.
 - b. Bagi Institusi
Melalui penulisan ini diharapkan nantinya dapat digunakan sebagai tambahan referensi kepustakaan terutama dalam upaya menganalisis metode *clustering* dengan *spectral clustering* pengelompokan tingkat tindak kriminalitas. Bagi pihak

kepolisian setempat dapat menjadi referensi guna meningkatkan aktivitas keamanan, pencegahan, dan penindakan kejahatan.

1.6 Sistematika Penulisan

Untuk memudahkan dalam penulisan, maka penyusunan tugas akhir ini ditulis dengan sistematika sebagai berikut:

BAB I : PENDAHULUAN

Berisi beberapa cakupan yaitu latar belakang, rumusan masalah, batasan masalah, tujuan penelitian, manfaat penelitian, dan sistematika penulisan.

BAB II : TINJAUAN PUSTAKA

Berisi tinjauan pustaka dan landasan teori yang bersangkutan dari masing-masing bagian penulisan.

BAB III : METODE PENELITIAN

Berisi tentang jenis penelitian, lokasi penelitian, sistematika penelitian, dan jadwal penelitian.

BAB IV : PEMBAHASAN

Berisi tentang pengolahan data yang diperoleh serta analisa dari masalah yang diangkat dengan menggunakan teori-teori yang berhubungan.

BAB V : PENUTUP

Berisi kesimpulan dan saran dari penelitian yang telah dilakukan.

BAB II

TINJAUAN PUSTAKA

2.1 Sosial Media *Twitter*

Data tweet merupakan salah satu contoh sumber data real-time yang dapat digunakan. Terdapat beberapa pendekatan yang dapat dilakukan untuk menganalisis data *tweet*, salah satunya adalah menggunakan metode *text mining* (Fridom Mailo et al., 2019).

Twitter adalah sebuah *platform online* yang dimiliki dan dijalankan oleh perusahaan *Twitter Inc.* Platform ini didirikan pada tahun 2006 oleh Jack Dorsey dan berpusat di San Bruno, California, dekat San Francisco, tempat situs ini pertama kali dibuat. Di Indonesia, *Twitter* telah menjadi salah satu media sosial yang populer, terutama dalam bentuk *microblogging*. *Twitter* memungkinkan pengguna untuk membagikan konten dalam format *microblogging* dengan batasan 280 karakter tulisan (Rezeki et al., 2020).

Berdasarkan informasi dari *Hootsuite*, *Twitter* memiliki sekitar 63,6% pengguna media sosial, menjadikannya peringkat ke-5 dalam kategori media sosial yang paling sering diakses pada tahun 2021. Rata-rata, pengguna menghabiskan waktu sekitar 8,1 jam per bulan untuk mengakses *Twitter*. Hal ini menunjukkan bahwa banyak orang menggunakan *Twitter* dalam jangka waktu yang cukup lama (Simbolon & Siahaan, 2021).

2.2 *Data Mining Clustering*

Data mining merupakan proses ekstraksi informasi dari kumpulan data besar dengan meraih data dari sumber yang belum terpahami. Dalam umumnya, metode

statistik, matematika, dan bahkan kecerdasan buatan digunakan dalam proses pertambahan data. Salah satu metodenya adalah klasterisasi, yang sering diadopsi dalam pertambahan data. Klasterisasi bertujuan untuk mengelompokkan data dengan ciri serupa dalam satu kelompok, sementara data yang memiliki karakteristik berbeda dikelompokkan secara terpisah. Keuntungan utama dari klasterisasi adalah pemisahan data yang berguna untuk prediksi dan analisis permasalahan bisnis, serta pengenalan entitas dalam berbagai ranah. (*computer vision dan image processing*) (Binus University, 2021). Adapun beberapa riset terhadap yang melakukan penelitian mengenai data mining, yaitu sebagai berikut:

Riset pertama merupakan riset yang dilakukan oleh Nurfahida dan Sembiring (2021) yang berjudul “Analisis *Text Clustering* Masyarakat Di *Twitter* Mengenai *McDonald's* Menggunakan *Orange Data Mining*”. Studi ini bertujuan untuk menganalisis pandangan masyarakat melalui komentar yang disampaikan di platform *Twitter*. Metode yang digunakan adalah analisis *Vader*, yang diikuti dengan pembentukan profil *tweet* serta visualisasi distribusinya. Penelitian ini menjalankan proses analisis menggunakan aplikasi *OrangeData Mining*. Langkah-langkah yang diambil mencakup persiapan teks seperti *transformasi*, *tokenisasi*, *normalisasi*, dan penyaringan, yang bertujuan untuk mempersiapkan teks agar dapat diolah secara analitis. Temuan dari penelitian ini mengindikasikan bahwa tanggapan masyarakat terhadap kolaborasi *McDonald's* *x**BTS* terdiri dari enam respon, dengan respon tertinggi ditemukan pada perasaan keterkejutan yang dirasakan oleh masyarakat.

Adapun riset kedua yang juga membahas mengenai data mining yaitu riset yang dilakukan oleh Kurniawan et al., (2020) yang berjudul “*Data Mining*

Menggunakan Metode *K-Means Clustering* Untuk Menentukan Besaran Uang Kuliah Tunggal”. Dataset yang digunakan dalam studi ini berisi informasi mengenai calon mahasiswa baru di Universitas Negeri Padang. Pengelompokan data ini bertujuan untuk mendukung penentuan biaya UKT (Uang Kuliah Tunggal) bagi calon mahasiswa baru di institusi ini. Hasil dari penelitian ini menghasilkan kelompok-kelompok UKT sebanyak lima, yang terdiri dari kategori UKT 1 sebesar Rp. 500.000, kategori UKT 2 sebesar Rp. 1.000.000, kategori UKT 3 sebesar Rp. 2.000.000, kategori UKT 4 sebesar Rp. 3.000.000, dan kategori UKT 5 sebesar Rp. 4.000.000.

Riset ketiga yang juga membahas mengenai data mining merupakan riset yang dilakukan oleh Budiman (2019) yang berjudul “Penerapan Data Mining Untuk Menentukan Lokasi Promosi Penerimaan Mahasiswa Baru Pada Universitas Banten Jaya (Metode *K-Means Clustering*)”. Tujuan dari penelitian ini adalah untuk mengidentifikasi lokasi atau daerah yang memiliki potensi untuk menarik mahasiswa baru ke perguruan tinggi tersebut. Dengan menggunakan metode *clustering* dalam *data mining*, objek penelitian akan dikelompokkan berdasarkan kesamaan karakteristiknya. Ini akan membantu dalam mengidentifikasi daerah-daerah yang memiliki potensi untuk menarik mahasiswa baru. Pendekatan ini dalam menentukan lokasi promosi penerimaan mahasiswa baru melalui metode data mining diharapkan akan memberikan dampak positif dan berkelanjutan dalam strategi promosi. Dengan demikian, diharapkan dapat meningkatkan jumlah mahasiswa baru yang mendaftar setiap tahunnya.

2.2.1 Pengertian *Data Mining*

Data mining adalah proses analisis data dengan tujuan untuk menemukan hubungan atau pola yang tidak terduga, serta merangkum data dalam bentuk yang berbeda dari sebelumnya, yang memiliki makna dan manfaat bagi pemilik data. *Data mining* merupakan disiplin lintas berbagai bidang ilmu, yang menggabungkan teknik-teknik dari pembelajaran mesin, pengenalan pola, statistika, basis data, dan visualisasi data untuk mengatasi tantangan dalam mengambil informasi berharga dari kumpulan data yang besar (Utomo dan Mesran, 2020).

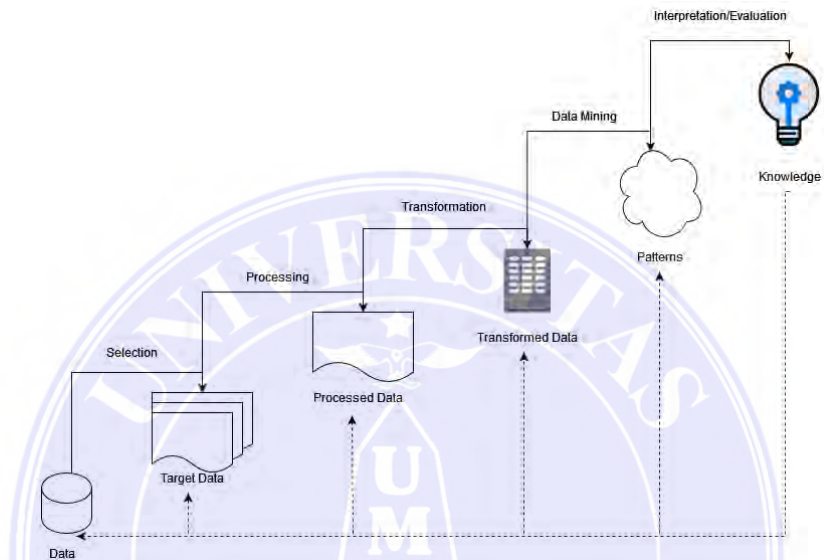
Data mining adalah suatu proses di mana pola-pola tersembunyi dalam data dianalisis dari berbagai sudut pandang untuk dikategorikan menjadi informasi yang bermanfaat. Proses ini melibatkan pengambilan data yang umumnya tersimpan dalam gudang data (*data warehouse*) dengan tujuan untuk menghasilkan informasi yang efisien. Penggunaan algoritma-algoritma khusus dalam data mining memfasilitasi pengambilan keputusan bisnis dan menghasilkan informasi yang berharga dari data yang ada (Arhami dan Nasir, 2020).

Menurut (Faid et al., 2019) *data mining* merupakan ilmu yang membahas tentang bagaimana menambang pengetahuan dari sebuah data. Adapun menurut (Purwadi et al., 2019) *Data mining* adalah proses analisis data yang bertujuan untuk menemukan pola atau hubungan yang tersembunyi dalam kumpulan data. Melalui teknik-teknik khusus, data mining mampu mengolah volume data yang besar dan mengubahnya menjadi informasi berupa pola yang memiliki signifikansi dan makna bagi mereka yang mengambil keputusan.

Sehingga dapat disimpulkan, data mining adalah proses pertambangan pada

data untuk mendapatkan pola serta trend yang bermanfaat bagi penambang yang kemudian bisa diolah menjadi informasi lebih terbarukan yang lebih terstruktur sesuai dengan kebutuhan.

2.2.2 Tahapan *Data Mining*



Gambar 2. 1 Tahapan Data Mining

Berikut tahapan data mining (Dwirohayati, 2020):

1. *Data selection*

Sebelum memulai tahap penggalian informasi dalam proses *Knowledge Discovery in Databases* (KDD), langkah penting yang perlu dilakukan adalah seleksi data dari kumpulan data operasional. Data yang telah terpilih untuk proses data mining akan disimpan dalam suatu berkas terpisah, yang berbeda dari basis data operasional. Hal ini memungkinkan untuk memisahkan data yang digunakan dalam proses analisis dan penemuan informasi dari data operasional yang digunakan

dalam aktivitas sehari-hari. Dengan demikian, pendekatan ini memungkinkan lebih efisien dan teratur dalam melaksanakan tahap-tahap KDD.

2. *Pre-processing/cleaning*

Sebelum melakukan proses penambangan data, diperlukan langkah pembersihan pada data yang menjadi fokus KDD. Tahap pembersihan melibatkan tindakan seperti menghilangkan data yang sama, memeriksa konsistensi data yang tak sejalan, serta memperbaiki kesalahan yang mungkin muncul dalam data.

3. *Transformation*

Koding adalah langkah transformasi terhadap data yang telah terpilih, dengan tujuan untuk menyusun data tersebut agar cocok untuk proses penambangan data. Tahap koding dalam proses KDD melibatkan pendekatan kreatif dan sangat bergantung pada jenis atau pola informasi yang diinginkan dari basis data..

4. *Data mining*

Data mining adalah proses analisis data yang bertujuan untuk menemukan pola atau hubungan yang tersembunyi dalam kumpulan data. Melalui teknik-teknik khusus, *data mining* mampu mengolah volume data yang besar dan mengubahnya menjadi informasi berupa pola yang memiliki signifikansi dan makna bagi mereka yang mengambil keputusan.

5. *Interpretation/evaluation*

Informasi yang dihasilkan dari proses penambangan data perlu diungkapkan dalam format yang dapat dengan mudah dimengerti oleh pihak-pihak yang berkepentingan. Tahap ini merupakan bagian dari proses KDD yang disebut interpretasi. Langkah ini meliputi pengecekan apakah pola atau informasi yang telah diidentifikasi sesuai dengan fakta atau hipotesis yang telah ada sebelumnya.

Data mining adalah suatu rangkaian proses maka dibagi menjadi beberapa tahap antara lain (Sulistiani dan Darwis, 2020):

1. Pembersihan data: untuk membuang data yang tidak konsisten dan noise.
2. Integrasi data: untuk menggabungkan data dari beberapa sumber.
3. Transformasi data : untuk mengubah data menjadi bentuk yang sesuai untuk di mining.
4. Aplikasi teknik data mining.
5. Evaluasi pola yang ditemukan : untuk menemukan informasi yang menarik atau pun bernilai.
6. Presentasi pengetahuan dengan teknik visualisasi.

Tahapan proses data mining ini terdiri dari (Yasin et al., 2022):

1. Pembersihan Data (*Data Cleaning*)

Penghapusan data adalah tahap di mana noise atau data yang tidak relevan serta data yang tidak konsisten dihilangkan untuk memastikan kelancaran

proses pemrosesan data. Secara umum, data yang diperoleh seringkali memiliki ketidaksempurnaan dan kompleksitas yang tidak sepenuhnya teratasi.

2. Integrasi Data (*Data Integration*)

Integrasi data adalah proses menggabungkan berbagai data menjadi satu dalam sebuah basis data. Penting untuk melaksanakan integrasi data dengan hati-hati, karena ketidakakuratan dalam integritas data bisa menghasilkan hasil yang tidak akurat bahkan menyesatkan..

3. Seleksi Data (*Data Selection*)

Seleksi data memiliki tujuan untuk mengambil hanya data yang relevan dalam proses penambangan data. Data seleksi merupakan data yang ada dalam basis data dan biasanya tidak seluruhnya digunakan. Hanya data yang memenuhi kriteria konsistensi dan validitas yang akan diambil dari basis data untuk dianalisis.

4. Transformasi Data (*Data Transformation*)

Transformasi data melibatkan perubahan atau penggabungan data ke dalam format yang cocok untuk diproses dalam penambangan data.

5. Proses *Mining*

Proses penambangan adalah tahapan inti ketika suatu metode diterapkan atau diimplementasikan untuk menemukan informasi dan pengetahuan yang tersembunyi dalam basis data.

6. Evaluasi, Interpretasi Dan Visualisasi

Tahap ini adalah proses untuk mengamati data yang telah diolah dan menilai apakah data tersebut memiliki kepentingan atau nilai menarik untuk diidentifikasi dan dimasukkan ke dalam Knowledge Based.

2.2.3 *Scrapping Data*

Scrapping data merupakan suatu metode pengumpulan data dari berbagai sumber di internet secara otomatis (Almaqbal et al., 2020). Proses ini menggunakan perangkat lunak khusus atau bot untuk mengakses situs web, mengekstraksi informasi yang relevan, dan menyimpannya dalam format yang dapat diolah lebih lanjut, seperti file sistem atau basis data. Tujuan utama dari *scrapping data* adalah untuk mengumpulkan jumlah data yang besar dengan cepat dan efisien, sehingga memungkinkan analisis lebih mendalam dan pemahaman yang lebih baik tentang berbagai aspek yang ingin dipelajari atau dijelajahi. Metode ini sering digunakan dalam berbagai bidang, seperti penelitian, analisis pasar, dan pengembangan produk, karena memungkinkan akses ke informasi yang mungkin sulit didapatkan secara manual. Penting untuk diingat bahwa saat melakukan *scrapping data*, perlu mematuhi aturan dan kebijakan etika dari situs-situs yang diakses untuk menghindari pelanggaran privasi atau hak cipta.

Web Scrapping, juga dikenal sebagai ekstraksi web, merujuk pada teknik untuk mengambil data dari *World Wide Web* (WWW) dan menyimpannya dalam file sistem atau basis data untuk analisis data (Kesuma & Iskandar, 2022). Pengambilan data dari web, yang dikenal sebagai *Web Scrapping*, dapat dilakukan secara manual oleh pengguna atau secara otomatis oleh bot atau *crawler web*. Proses *Web*

Scraping dari internet dapat dibagi menjadi dua langkah berurutan, yaitu pertama adalah mengakuisisi sumber daya web, dan langkah kedua adalah mengekstraksi informasi yang diinginkan dari data yang telah diunduh.

2.2.4 Teknik Data Mining Clustering

Clustering adalah langkah pengelompokan yang didasarkan pada prinsip kesamaan antar kelas dan mengurangi kesamaan antar kelompok. Berbagai algoritma klasterisasi telah dikembangkan untuk mencapai kinerja yang optimal (Indriyani dan Irfiani, 2019).

Clustering adalah salah satu teknik yang termasuk dalam fungsionalitas data mining. Algoritma klasterisasi digunakan untuk mengelompokkan sejumlah data menjadi kelompok-kelompok data yang memiliki kesamaan tertentu (*cluster*) (Tamba dan Kesuma, 2019). *Clustering* adalah aktivitas yang bertujuan untuk mengelompokkan data yang memiliki kesamaan antara satu data dengan data lainnya. (Aditya et al., 2020).

Clustering adalah tugas yang bertujuan mengelompokkan data yang memiliki kesamaan di antara satu data dengan data lainnya menjadi kluster atau kelompok. Maksudnya adalah agar data dalam satu kluster memiliki tingkat kemiripan (*similarity*) yang paling tinggi, dan data antar kluster memiliki kemiripan yang paling rendah. Klasterisasi juga bisa diartikan sebagai metode untuk membagi-bagi data yang diimplementasikan di berbagai bidang seperti pemasaran, analisis masalah bisnis, segmentasi pasar dan prediksi, pola dalam visi komputer, zonasi wilayah, serta identifikasi objek dan pemrosesan gambar. (Nishom, 2019).

Clustering tidak melibatkan variabel target, yang membedakannya dari

klasifikasi. Banyak metode klasterisasi yang dikembangkan oleh para ahli. Setiap metode memiliki karakteristik, kelebihan, dan kelemahan tertentu. Metode klasterisasi dapat dibagi menjadi dua berdasarkan strukturnya, yaitu pengelompokan hierarkis dan partisi. Pengelompokan hierarkis memiliki aturan di mana data tunggal dapat dianggap sebagai kelompok, beberapa kelompok kecil bisa digabung menjadi satu kelompok besar, dan ini berlanjut sampai semua data tergabung dalam satu kelompok. Di sisi lain, metode klasterisasi partisi mengelompokkan data ke dalam sejumlah kelompok yang tidak tumpang tindih antara satu kelompok dengan yang lain, sehingga setiap data hanya termasuk dalam satu kelompok (Salsabila, 2019).

2.2.5 Spectral Clustering

Menurut Meila dalam Millenia (2022) *Spectral Clustering* adalah suatu algoritma yang digunakan untuk mengidentifikasi k klaster dengan memanfaatkan eigenvector dari suatu matriks. Matriks ini dihasilkan dari matriks kemiripan antara setiap pasangan data yang hendak dikelompokkan.

Menurut Bach dan Jordan yang dikutip oleh Hatta et al., (2021) *Spectral Clustering* adalah algoritma pengelompokan yang berkembang yang telah berkinerja lebih baik daripada banyak algoritma pengelompokan tradisional dalam banyak kasus. *Spectral Clustering* membuat setiap titik data sebagai grafik-node dan kemudian mengubah masalah pengelompokan menjadi masalah partisi-grafik.

Spectral clustering adalah salah satu teknik pengelompokan yang sangat *user-friendly* dan relatif cepat, terutama untuk data yang memiliki dimensi tinggi, bahkan hingga data berjumlah ribuan. Dalam metode spectral clustering,

pendekatan ini menganggap pengelompokan data sebagai grafik dan melakukan partisi data tanpa asumsi tertentu mengenai bentuk kluster data. (Wulandari dan Novita, 2021).

Berikut merupakan langkah-langkah dalam melakukan *spectral clustering* menurut Yusuf dalam Wulandari (2020):

1. Membentuk graf kemiripan dari dataset melibatkan penggunaan vertex atau simpul yang mewakili objek dalam dataset. Bobot pada setiap sisi atau edge menggambarkan jarak antara satu simpul dengan simpul lainnya. Penghitungan jarak ini melibatkan fungsi kesamaan Gauss.:

$$s(x_i, x_j) = \exp\left(-\frac{\|x_i - x_j\|^2}{2\sigma^2}\right) \quad (2.1)$$

2. Graf G adalah sebuah grafik yang tidak memiliki arah, dengan matriks bobot W untuk setiap edge $(i, j) \in E$, diberikan bobot w_{ij} dengan nilai $w_{ij} = w_{ji} \geq 0$. Matriks kesamaan.

$$W = (W_{ij})_{i,j=1}^n \quad (2.2)$$

Dengan bobot edge (i, j) mewakili kesamaan antara data x_i dan data x_j . Hitung d_i yang merupakan derajat dari simpul $v_i \in V$ yang didefinisikan sebagai

$$d_i = \sum_j^n W_{ij} \quad (2.3)$$

Dari degree vertex tersebut dapat dibentuk matriks degree yang merupakan matriks diagonal yang berisi bobot setiap vertex.

3. Data dipartisi menjadi k kluster yang independen, di mana k mewakili

jumlah kluster, dengan langkah normalisasi menggunakan matriks Laplacian dan matriks bobot W serta matriks derajat D yang dihitung menggunakan persamaan

$$L_{sym} = D^{-\frac{1}{2}} L D^{-\frac{1}{2}} = I - D^{-\frac{1}{2}} W D^{-\frac{1}{2}} \quad (2.4)$$

Dengan menggunakan notasi $L = D - W$, di mana D adalah matriks diagonal yang elemennya berasal dari persamaan (2.4), yang merepresentasikan bobot dari semua simpul yang bertetangga dengan simpul i , dan W berasal dari persamaan (2.2).

4. Lakukan perhitungan terhadap k -eigen vektor pertama dari matriks Laplacian. Dengan demikian, terbentuklah matriks k -eigen yang terdiri dari vektor eigen pertama dari matriks Laplacian. Matriks k -eigen ini memiliki dimensi $n \times k$, di mana n adalah jumlah objek dalam dataset..
5. Lakukan normalisasi data menggunakan matriks k -eigen, sehingga tercipta k kolom yang mewakili nilai eigen yang telah dinormalisasi pada setiap kolomnya.
6. Tahap akhir adalah dengan melakukan pengelompokan dengan menggunakan algoritma *clustering* untuk membentuk partisi kelompok, salah satunya adalah menggunakan *K-Means Clustering*.

Tahapan *K-Means Clustering* meliputi (Wahyudi et al., 2021):

- a) Penentuan Jumlah Kluster (K)
- b) Penentuan tak pusat Kluster Secara Random (Centroid)

$$v = \frac{\sum_{i=1}^n x_i}{n}; i = 1, 2, 3, \dots, n \quad (2.5)$$

Dimana;

V = centroid pada kluster

x_i = objek ke-i

n = banyaknya objek/jumlah objek yang menjadi anggota kluster

c) Hitung jarak setiap data ke pusat kluster

Hitung jarak setiap objek ke masing-masing centroid dari masing-masing kluster. *Euclidian Distance* digunakan untuk menghitung jarak antar objek dengan centroid.

$$d(x, y) = |x - y| = \sqrt{\sum_{i=1}^n (X_i - Y_i)^2}; i = 1, 2, 3, \dots, n \quad (2.6)$$

Dimana;

X_i = objek x ke-i

Y_i = daya y ke-i

N = banyaknya objek

d) Kelompokkan data ke dalam kluster dengan jarak yang paling pendek

e) Hitung Pusat Kluster

$$C_{kj} = \frac{\sum_{i=1}^p x_{ij}}{p}; i = 1, 2, 3, \dots, n \quad (2.7)$$

Dimana:

X_{ij} = ϵ Kluster ke-k

P = banyaknya anggota kluster ke-k

2.3 Kriminalitas

Kriminalitas merujuk pada perbuatan-perbuatan melanggar hukum, undang-undang, norma, dan nilai-nilai yang berlaku dalam masyarakat. Fenomena ini umum terjadi dan merupakan isu yang sering dihadapi oleh berbagai lapisan masyarakat di seluruh dunia. Kriminalitas dapat menimbulkan dampak negatif baik secara ekonomis maupun psikologis, sekaligus melanggar aturan hukum yang berlaku di negara serta norma-norma sosial hingga nilai-nilai agama. Oleh karena itu, wajar jika setiap tindakan kriminalitas mendapat penentangan dari semua anggota masyarakat (Merdeka.com, 2023).

2.3.1 Pengertian Kriminalitas

Kriminalitas adalah istilah yang mengacu pada pelanggaran hukum atau tindakan yang bertentangan dengan aturan-aturan yang berlaku. Berbagai jenis kejahatan seperti pembunuhan, penyerangan, pemerkosaan, dan lainnya, semakin berkembang secara luas dalam waktu yang sekarang ini (Yamini, 2019). Kriminalitas adalah perilaku yang melanggar norma-norma hukum dan dapat berdampak merusak terhadap kehidupan sosial dalam masyarakat. Kriminalitas merupakan masalah yang memiliki urgensi yang tinggi dan perlu mendapat perhatian serius dari pemerintah (Yenni dan Utnasari, 2019).

Kriminalitas merujuk pada tindakan yang menyalahi norma, merugikan individu lain, dan melanggar hukum, nilai-nilai agama, serta norma-norma sosial

(Dewi et al., 2019). Kriminalitas adalah sebuah isu yang sering muncul dalam lingkungan masyarakat dan perlu diberikan perhatian karena dapat merugikan berbagai kepentingan serta membawa dampak negatif kepada masyarakat. Dampak-dampak tersebut meliputi perasaan cemas, ketidakamanan, kepanikan, dan ketakutan (Fahmi et al., 2021).

Kriminalitas adalah perbuatan yang melanggar norma dan aturan yang berlaku, sehingga menghasilkan ketidakselarasan dalam kehidupan. (Sari & Azhar, 2019). Menurut Kartono dalam (Putra et al., 2021) Kriminalitas mencakup berbagai bentuk tindakan dan perbuatan yang merugikan secara ekonomi dan psikologis, yang bertentangan dengan hukum yang berlaku di Indonesia serta norma-norma sosial dan agama. Dengan kata lain, tindakan kriminalitas merujuk pada segala jenis perbuatan yang melanggar hukum dan norma-norma sosial, sehingga mendapat penolakan dari masyarakat.

2.3.2 Faktor Penyebab Kriminalitas

Faktor penyebab kriminalitas menurut (Putra et al., 2021) yaitu diantaranya faktor internal yang meliputi:

1. Faktor kebutuhan ekonomi yang mendesak
2. Faktor ketenagakerjaan (pengangguran atau memiliki pekerjaan)
3. Faktor taraf kesejahteraan.

Faktor eksternal meliputi:

1. Faktor pendidikan
2. Faktor pergaulan atau pengaruh lingkungan.

Kejahatan memiliki penyebab-penyebab yang dapat dikelompokkan menjadi faktor biologis, faktor sosiologis, dan faktor ekonomi. Perubahan dalam faktor-faktor ini dapat mendorong individu untuk melakukan tindakan kriminal. Rentang tindakan tersebut meliputi dari kejahatan yang rendah hingga kejahatan yang lebih serius (Fachrurrozi et al., 2021).

2.4 *Davies Boulden Index (DBI)*

David Boulden Index (DBI) bisa digunakan untuk menentukan cluster dengan menghitung jarak antar cluster satu dengan jarak cluster lainnya dalam satu waktu yang sama (Badruttamam & Marrudani, 2020). *Clustering* yang baik adalah *clustering* dengan nilai rendah (Kartikasari, 2021). Tahapan dalam evaluasi dengan *Davies Boulden Index (DBI)* (Septiani et al., 2022):

1. Menghitung *Sum Of Square Within Cluster (SSW)*

Menghitung keterikatan anggota cluster atau menghitung seberapa mirip anggota satu dengan lainnya, semakin mirip maka semakin baik. SSW dihitung untuk melihat matriks.

$$SSW_i = \frac{1}{m_i} \sum_j^{m_i} d(x_j, c_i) \quad (2.8)$$

Keterangan:

m_i = Jumlah data dalam *cluster* ke – i

x = Data dalam *cluster*

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

x_j = Data pada *cluster* tersebut

c_j = centroid *cluster* ke- i

2. Menghitung *Sum Of Square Between Cluster* (SSB)

Tujuannya adalah untuk menghitung heterogenitas/separasi pada cluster satu dengan lainnya.

$$SSB_{i,j} = (c_i, c_j) \quad (2.9)$$

Keterangan:

C_i = Cluster 1

C_j = Cluster lainnya

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

3. Menghitung rasio

Tujuannya adalah untuk mengetahui seberapa bagus perbandingan nilai cluster satu dengan lainnya.

$$DBI = \frac{1}{K} \sum_{i=1}^k \max_{i \neq j} (R_{i,j}) \quad (2.10)$$

Keterangan:

K = Kluster yang ada

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

Max = Dicari rasio antara kluster yang terbesar

2.5 Penelitian Terdahulu

Tabel 2. 1 Penelitian Terdahulu

No	Penulis (Tahun)	Topik	Hasil
1	Resti Noor Fahmi, Mohamad Jajuli, Nina Sulistiyowati (2021)	Analisis Pemetaan Tingkat Kriminalitas di Kabupaten Karawang menggunakan Algoritma K-Means	hasil penelitian, terungkap bahwa pengelompokan daerah rawan kriminalitas di Karawang pada tahun 2019 menghasilkan 23 kecamatan dalam cluster tidak rawan, 3 kecamatan dalam cluster rawan, dan 4 kecamatan dalam cluster sangat rawan. Pada tahun 2020, pengelompokan menghasilkan 22 kecamatan dalam cluster tidak rawan, 4 kecamatan dalam cluster rawan, dan 4 kecamatan dalam cluster sangat rawan. Dengan Hasil evaluasi koefisien silhouette pada tahun 2019 sebesar 0,52 dan pada tahun 2020 sebesar 0,54.
2	Lilis Suriani (2020)	Pengelompokan Data Kriminal Pada Poldasu Menentukan	Penelitian ini adalah metode algoritma k-means dengan menggunakan software rapidminer

		<p>Pola Daerah Rawan Tindak Kriminal Menggunakan Data Mining Algoritma K-Means Clustering</p>	<p>7.3. Temuan dari penelitian ini mengindikasikan bahwa pembagian pengelompokan daerah rawan oleh Poldasu belum terbukti efektif dan efisien. Pembagian kategori dilakukan berdasarkan data Jumlah Tindak Pidana (JTP) dan Jumlah Pelaku Tindak Pidana (JPTP), yang kemudian digunakan untuk mengkategorikan daerah menjadi rawan, sedang, atau tidak rawan.</p>
3	<p>Danang Sutejo, Yosep Agus Pronoto, Hani Zullfia Zahro (2020)</p>	<p>Sistem Informasi Geografis Pengelompokan Tingkat Kriminalitas Kota Malang Menggunakan Metode K-Means</p>	<p>pengelompokan menghasilkan tiga kluster yang dinilai menggunakan indeks Davies Bouldin. Hasil terendah dari indeks ini adalah sebesar 2,401. Dalam pembagian wilayah, kluster 1 (C1) menunjukkan intensitas tinggi keamanan dan meliputi Kecamatan Kedungkandang. Kluster 2 (C2) dianggap cukup rawan dan mencakup Kecamatan Blimbing. Kluster 3 (C3) menunjukkan intensitas rawan dan melibatkan Kecamatan Klojen dan Sukun.</p>

			Sementara kluster 4 (C4) dikategorikan sebagai sangat rawan dan mencakup Kecamatan Lowokwaru.
4	Millenia Winadya Putri, Indah Manfaati Nur, Rochdi Wasono (2022)	Implementasi <i>Spectral Clustering Algorithm</i> Untuk Pengelompokan Sasaran Vaksinasi Covid-19 Di Indonesia	Masyarakat dikelompokkan menjadi tiga kluster, yaitu kluster dengan tingkat vaksinasi tinggi, sedang, dan rendah. Evaluasi hasil kluster menggunakan <i>Davies Boulden index</i> (DBI) menunjukkan nilai yang hampir sama sehingga dapat terlihat bahwa anggota-anggota pada suatu kluster di masing-masing kategori tidak jauh berbeda.

BAB III

METODE PENELITIAN

3.1 Sumber Data Penelitian

Penelitian ini menggunakan data sekunder yang diperoleh dari Twitter mengenai data kriminalitas yang terjadi di Indonesia. Data tersebut diambil berdasarkan *query* pencarian tentang kriminalitas yang terjadi di Indonesia.

3.2 Bahan dan Alat

Dalam penelitian ini, bahan yang digunakan adalah data sekunder yang diperoleh dari *Twitter* mengenai data kriminalitas yang terjadi di Indonesia. Alat-alat yang digunakan dalam penelitian ini meliputi:

3.2.1. Perangkat Keras

Alat yang digunakan dalam penelitian ini adalah satu buah unit laptop yang dilengkapi dengan perangkat keras (*hardware*) dan perangkat lunak (*software*). Perangkat keras komputer yang digunakan adalah perangkat keras yang dapat mendukung perangkat lunak yang memiliki kemampuan atau tampilan grafis yang cukup baik. Spesifikasi yang digunakan dalam penelitian ini ditunjukkan pada Tabel 3.1

Tabel 3. 1 Kebutuhan Perangkat Keras

No	Nama Perangkat	Deskripsi
1	Laptop	Asus X441Ba
2	Processor	AMD A9-9425
3	RAM/Penyimpanan	8GB/1TB HDD

3.2.2. Perangkat Lunak

Tabel 3. 2 Kebutuhan Perangkat Lunak

No	Nama	Deskripsi
1	Windows 10 64-bit	Operating System
2	Python 3.10.11 dan Google Colab	Tools untuk membangun dan melatih (training) model
3	Pandas	Pustaka Python untuk memanipulasi dan menganalisis data.
4	Scikit-learn	Pustaka Python yang menyediakan algoritma dan fungsi untuk analisis data, termasuk pengelompokan data (<i>clustering</i>).
5	Matplotlib dan Seaborn	Pustaka Python untuk visualisasi data.
6	Microsoft Office 2016	Tools untuk membuat laporan penelitian

3.3 Populasi dan Sampel

Populasi dalam penelitian ini adalah seluruh tweet yang mengandung informasi tentang kriminalitas di Indonesia di *platform* Twitter. Sampel yang digunakan adalah sebanyak 10.000 *tweet* yang berhasil dikumpulkan melalui metode *scraping* data. Sampel tersebut dianggap mewakili populasi dan digunakan untuk analisis data lebih lanjut.

3.4 Tingkat Tindak Kriminalitas

Tingkat tindak kriminalitas dalam penelitian ini dibagi menjadi 3 tingkatan tindak, yaitu sebagai berikut:

1. Cukup Berbahaya

Kejahatan dapat dikatakan cukup berbahaya ketika kejahatan tersebut tidak berkaitan dengan nyawa atau fisik orang lain. Contohnya, kejahatan terhadap ketertiban umum, dimana ketertiban umum merupakan kewajiban bagi warga negara (Hartati, 2020). Maka, tindak kejahatan yang masuk kedalam kriteria tersebut yang kemudian dirangkum sebagai kamus untuk klasifikasi Dalam konteks Kitab Undang-Undang Hukum Pidana (KUHP) Republik Indonesia serta *The International Classification of Crime for Statistical Purposes* (ICCS), terdapat klasifikasi kejahatan yang dapat dijabarkan sebagai berikut:

- a. Narkotika dan Psikotropika
- b. Penipuan/Perbuatan Curang
- c. Penggelapan
- d. Korupsi
- e. Terhadap Ketertiban Umum

2. Sangat Berbahaya

Kejahatan dapat dikatakan sangat berbahaya apabila kejahatan tersebut berkaitan dengan perampasan nyawa orang lain. Kejahatan terhadap nyawa (*homicide*) memang merupakan tindak kejahatan yang paling tinggi hierarkinya dalam klasifikasi kejahatan (Abd & Musahib, 2022). Kemudian kejahatan terhadap fisik, dan atau perampasan kemerdekaan. Maka, tindak kejahatan yang masuk kedalam kriteria tersebut yang kemudian dirangkum sebagai kamus untuk klasifikasi dalam konteks Kitab Undang-Undang

Hukum Pidana (KUHP) Republik Indonesia serta *The International Classification of Crime for Statistical Purposes* (ICCS) adalah sebagai berikut:

- a. Pembunuhan
 - b. Penganiayaan Berat
 - c. Penganiayaan Ringan
 - d. Kekerasan dalam Rumah Tangga
 - e. Perkosaan
 - f. Pencabulan
 - g. Penculikan
3. Berbahaya

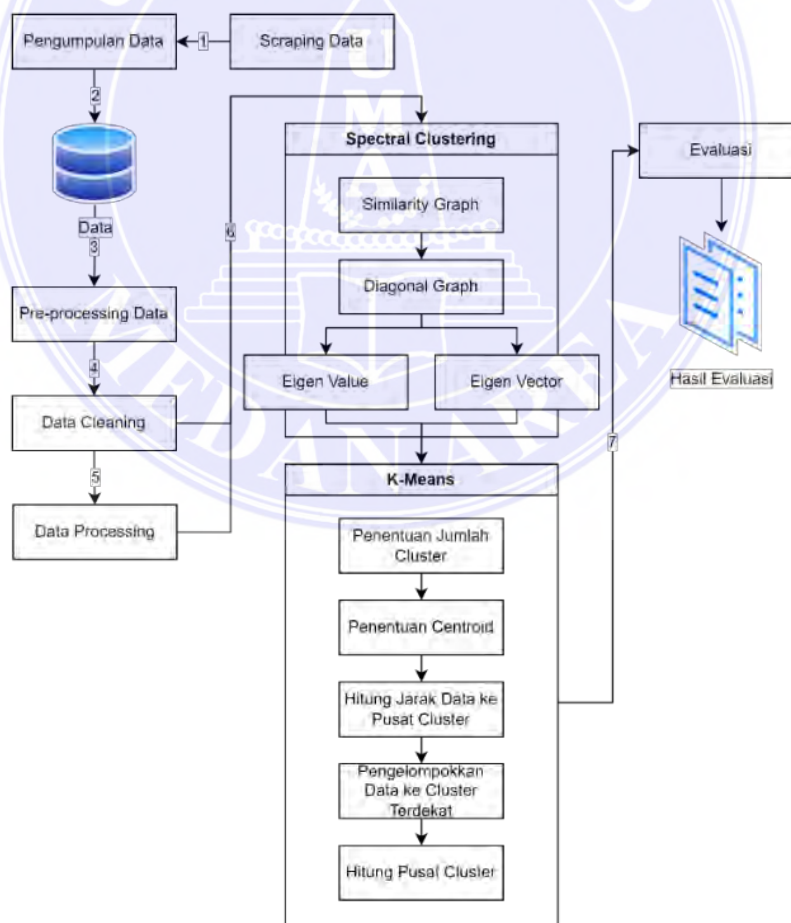
Kejahatan dapat dikatakan berbahaya apabila kejahatan tersebut tidak termasuk kedalam poin satu dan poin dua sebelumnya. Maka, tindak kejahatan yang masuk kedalam kriteria tersebut yang kemudian dirangkum sebagai kamus untuk klasifikasi dalam konteks Kitab Undang-Undang Hukum Pidana (KUHP) Republik Indonesia serta *The International Classification of Crime for Statistical Purposes* (ICCS) adalah sebagai berikut:

- a. Pencurian dengan Kekerasan
- b. Pencurian dengan Kekerasan Menggunakan Senjata Api
- c. Pencurian dengan Kekerasan Menggunakan Senjata Tajam

- d. Pencurian
- e. Pencurian dengan Pemberatan
- f. Pencurian Kendaraan Bermotor
- g. Pengrusakan/Penghancuran Barang
- h. Pembakaran dengan Sengaja
- i. Penadahan

3.5 Tahapan Penelitian

Berikut ini adalah garis besar dari kerangka kerja yang akan diterapkan:



Gambar 3. 1 Tahapan Analisis Penelitian

3.6 Metode Pengumpulan Data

Data dikumpulkan dengan menggunakan metode *scraping data* dari Twitter. Untuk melakukan *scraping*, *query-query* tertentu digunakan untuk mencari tweet-tweet yang relevan dengan kriminalitas di Indonesia. Setelah itu, tweet-tweet tersebut diambil berdasarkan jumlah tweet yang ditentukan sebelumnya, dalam hal ini sebanyak 10.000 tweet.

Berikut adalah daftar *query* untuk proses *scraping data* tersebut:

Tabel 3. 3 Tabel Query

No.	Query
1	'kejahatan', 'pembunuhan', 'pencurian', 'perampokan', 'narkoba', 'pemeriksaan', 'penipuan', 'curat', 'perdagangan manusia', 'penggelapan', 'pemalsuan', 'penculikan', 'penyerangan', 'pemerasan', 'kekerasan', 'pelanggaran', 'terorisme', 'pemeriksaan', 'korupsi', 'penyalahgunaan narkoba', 'pencabulan', 'pemeriksaan anak', 'penadahan', 'kriminalitas jalanan', 'perjudian', 'pencemaran nama baik'.

Kemudian hasil dari *scraping* menghasilkan dataset dengan sampel sebagai berikut:

Tabel 3. 4 Tabel Dataset

Query	Tweet	Tanggal	Lokasi
kejahatan	@dearmauv iya kalo di taxi driver kan dia jahat banget tuh nah kalo di	2023-06-25 11:52:22+00:00	ã...^ã...^ã...Ž ã,±ã,,·ã...‡

	crash course kejahatan dia ngapain ke guru Choi?		
kejahatan	ya Allah apa sih yg membuat manusia bisa kejam hatinya? kejahatan ada di mana2 banget ini....	2023-06-25 11:48:51+00:00	Tidak tersedia
kejahatan	UKM Anggota Polsek Banguntapan stanby di Simpang 4 Blok O, Antisipasi gangguan Kamtibmas maupun tindak kejahatan #PolsekBanguntapan #PolriPresisi #PolresBantul https://t.co/09WJWF820a	2023-06-25 11:45:50+00:00	Tidak tersedia
kejahatan	Patroli ikon kamtibmas Polsek Banguntapan harkamtibmas cegah kejahatan jalanan, pencurian dan tawuran. Wilayah aman kondusif #PatroliJogja #harkamtibmas #PolsekBanguntapan #PolriPresisi https://t.co/eELdti5EaP	2023-06-25 11:43:53+00:00	Tidak tersedia
kejahatan	Operasi Kejahatan Jalanan Cegah Pelaku Tawuran, Geng Motor dan Pelaku Begal, Mnggu (25/6). https://t.co/YWYKbn0GDs	2023-06-25 11:42:17+00:00	Tidak tersedia

kejahatan	dunia teknologi semakin maju, dan kejahatan semakin banyak di mana-mana, stay safe y'all!	2023-06-25 11:41:01+00:00	Tidak tersedia
kejahatan	Apel Operasi Kejahatan Jalanan Berikan Rasa Aman Bagi Warga yang Beraktifitas di Malam Hari, Minggu (25/6). https://t.co/7sth6djtHH	2023-06-25 11:39:46+00:00	Tidak tersedia
kejahatan	Personil Polsek Mlati Melaksanakan patroli dialogis dan mencegah kejahatan maupun gangguan Kamtibmas, dipimpin Aiptu Ihsanudin, strong point di Mbako 88 Cebongan Sumberadi Mlati Sleman. #polripresisi #polsekmlati #polrestasleman #poldajogja #PatroliJogja https://t.co/ogou5RrEZd	2023-06-25 11:32:27+00:00	Mlati, Indonesia
kejahatan	Personil Polsek Mlati Melaksanakan patroli dialogis dan mencegah kejahatan maupun gangguan Kamtibmas, dipimpin	2023-06-25 11:28:40+00:00	Mlati, Indonesia

	<p>Aiptu Ihsanudin, strong point di Mustofa Barkas Tlogoadi Mlati Sleman.</p> <p>#polripresisi #polsekmlati #polrestasleman #poldajogja #PatroliJogja https://t.co/mjyYLv8kID</p>		
kejahatan	<p>Patroli antisipasi kejahatan (Curat, Curas, Curanmor dan Tindak Pidana Lainnya) dipimpin Aiptu Ihsanudin, Strong point' di Toko Elektronik Cebongan Tlogoadi Mlati Sleman. Kegiatan berlangsung aman dan kondusif.</p> <p>#polripresisi #polsekmlati #polrestasleman #poldajogja #PatroliJogja https://t.co/gZYP9xvcJL</p>	<p>2023-06-25 11:23:55+00:00</p>	<p>Mlati, Indonesia</p>

3.7 Data Clean

Data yang dikumpulkan dengan menggunakan *scrapping data* kemudian dilakukan tahap *processing data* untuk mendapatkan data yang optimal, tahapan *processing data* pada penelitian ini menggunakan tahapan *cleaning data* untuk membersihkan data dari *noise*, *outlier*, dan *duplicate data*.

Berikut adalah *sample* dataset dari data yang telah berhasil dilakukan proses *cleaning*:

Tabel 3. 5 Tabel Data Clean

text_clean	token	normal	stop	stemmed	final
dearmauv	['dearmauv',	['dearmauv',	['dearmauv',	['dearmauv',	iya kalau
iya kalo di	'iya', 'kalo',	'iya', 'kalau',	'iya', 'kalau',	'iya', 'kalau',	kan jahat
taxi driver	'di', 'taxi',	'di', 'taxi',	'taxi',	'taxi',	bangetkalau
kan dia jahat	'driver',	'driver',	'driver',	'driver',	crash jahat
banget tuh	'kan', 'dia',	'kan', 'dia',	'kan', 'jahat',	'kan', 'jahat',	ngapain
nah kalo di	'jahat',	'jahat',	'banget',	'banget',	guru
crash course	'banget',	'banget',	'tuh', 'nah',	'tuh', 'nah',	
kejahatan	'tuh', 'nah',	'tuh', 'nah',	'kalau',	'kalau',	
dia ngapain	'kalo', 'di',	'kalau', 'di',	'crash',	'crash',	
ke guru choi	'crash',	'crash',	'course',	'course',	
	'course',	'course',	'kejahatan',	'jahat',	
	'kejahatan',	'kejahatan',	'ngapain',	'ngapain',	
	'dia',	'dia',	'guru',	'guru',	
	'ngapain',	'ngapain',	'choi']	'choi']	
	'ke', 'guru',	'ke', 'guru',			

	'choi']	'choi']			
--	---------	---------	--	--	--

3.8 Metode Penelitian

Penelitian ini menggunakan metode *data mining*, menggunakan metode *spectral clustering*, untuk menganalisis data kriminalitas yang diperoleh dari Twitter. Metode ini melibatkan beberapa tahapan utama, yaitu pengumpulan data, pemrosesan data yang mencakup *spectral clustering*. Tahapan pemrosesan data dapat dijabarkan sebagai berikut:

3.8.1. Process Data

1. Penentuan Jumlah Kluster

Dalam penentuan jumlah kluster untuk membagi data menjadi 3 tingkat tindak kriminalitas maka ditentukan 3 jumlah K (Kluster).

2. Penentuan Titik Pusat Kluster

$$v = \frac{\sum_{i=1}^n Xi}{n}; i = 1, 2, \dots, n \quad (1)$$

$$v = \frac{(538 + 266 + 9196)}{3}$$

$$v = 3333.$$

Dimana;

V = *centroid* pada kluster

xi = objek ke-i

n = banyaknya objek/jumlah objek yang menjadi anggota kluster

3. Hitung Jarak Setiap Data ke Pusat Kluster

$$d(x, y) |x - y| = \sqrt{\sum_{i=1}^n (X_i - Y_i)^2}; i = 1, 2, \dots, n \quad (2)$$

$$\begin{aligned} d(x, y) |x - y| &= \sqrt{(33333 - 3333.3)^2} \\ &= \mathbf{29999.7} \end{aligned}$$

Dimana;

X_i = objek x ke-i

X_i = daya y ke-i

N = banyaknya objek

4. Penghitungan Pusat Kluster

Untuk penghitungan pusat kluster akan digunakan data kluster ke-1 dengan mengambil 3 data pertama sebagai sample, sebagai berikut:

$$C_{kj} = \frac{\sum_{i=1}^p X_{ij}}{p}; i = 1, 2, 3, \dots, n \quad (3)$$

$$\begin{aligned} C_{kj} &= \frac{(4 + 13 + 5)}{3} \\ &= \mathbf{22} \end{aligned}$$

Dimana:

X_{ij} = € Kluster ke-k

P = banyaknya anggota kluster ke-k

3.8.2. Evaluasi

Evaluasi *Davies-Bouldin Index (DBI)* dengan perhitungan sebagai berikut:

$$DBI = \frac{1}{k} \sum_{i=1}^k \max_{i \neq j} (R_{i,j}, \dots k) \quad (4)$$

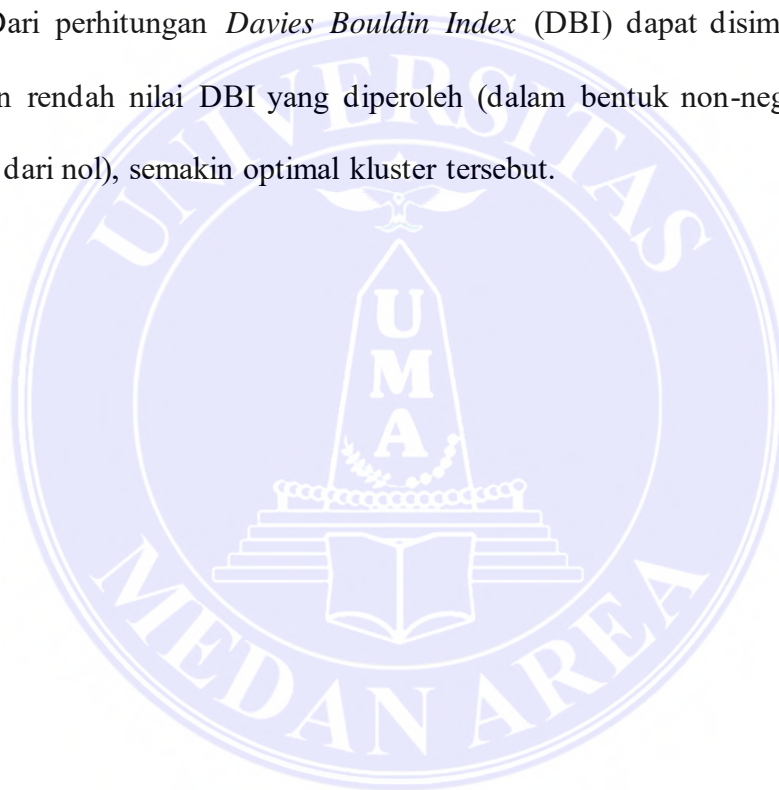
$$DBI = 2.7614929229864593$$

Dimana:

$R_{i,j}$ = Ratio dari nilai SSW dan SSB

k = Jumlah *kluster*.

Dari perhitungan *Davies Bouldin Index* (DBI) dapat disimpulkan bahwa semakin rendah nilai DBI yang diperoleh (dalam bentuk non-negatif dan tidak kurang dari nol), semakin optimal kluster tersebut.



BAB V

KESIMPULAN DAN SARAN

5.1 Kesimpulan

Berdasarkan analisis yang telah dilakukan terkait pengelompokan tingkat tindak kriminalitas di Indonesia menggunakan metode spectral clustering dengan data *tweeter* sebanyak 10.000 data, diperoleh kesimpulan bahwa metode *spectral clustering* mampu mengelompokkan data dari *tweet* masyarakat Indonesia untuk mempartisi data tingkat tindak kriminalitas di Indonesia menjadi tiga kategori, yaitu Cukup Berbahaya dengan presentasi 2,8% atau sebanyak 277 data, Berbahaya dengan presentasi 7.0% atau sebanyak 697 data, dan Sangat Berbahaya dengan presentasi 90,3% atau sebanyak 9026 data, kemudian menghasilkan skor evaluasi *Davies-Bouldin Index* (DBI) sebesar 2.7614929229864593.

5.2 Saran

Berikut adalah beberapa saran untuk pengembangan lebih lanjut:

1. Memperluas kamus kata kunci yang digunakan dalam proses labelisasi kluster untuk mencakup lebih banyak jenis kejahatan yang relevan.
2. Melakukan validasi dan verifikasi lebih lanjut terhadap hasil pengelompokan menggunakan metode spectral clustering dengan membandingkannya dengan metode pengelompokan lainnya.
3. Menggabungkan informasi tambahan seperti data demografi atau geografis untuk memperkaya analisis pengelompokan tingkat tindak kriminalitas.
4. Melakukan analisis lebih lanjut terhadap kluster untuk mengidentifikasi pola atau karakteristik khusus yang terkait dengan kejahatan.

DAFTAR PUSTAKA

- Abd, O., & Musahib, R. (2022). Pertanggung Jawaban Pidana Terhadap Hilangnya Nyawa Orang Yang Dilakukan Secara Bersama-Sama. 2(9).
- Aditya, A., Jovian, I., & Sari, B. N. (2020). Implementasi K-Means Clustering Ujian Nasional Sekolah Menengah Pertama Di Indonesia Tahun 2018/2019. *Jurnal Media Informatika Budidarma*, 4(1), 51-58.
- Ainsiyah, E. D., Saragih, K. W., Purba, P. M., Ppkn, D. P., & Usi, F. (2021). Fenomena Meningkatnya Kriminalitas Dan Kekerasan Di Indonesia. *MORALITA*, 2(1), 50–61. <https://doi.org/10.36985/Jma.V2i1.284>
- Akbar, Y., & Sugiharto, T. (2023). Analisis Sentimen Pengguna Twitter Di Indonesia Terhadap Chatgpt Menggunakan Algoritma C4.5 Dan Naïve Bayes. *Jurnal Sains Dan Teknologi*, 5(1), 115–122. <https://doi.org/10.55338/Saintek.V4i3.1368>
- Almaqbal, I. S. H., Al Khufairi, F. M. A., Khan, M. S., Bhat, A. Z., & Ahmed, I. (2020). Web Scrapping: Data Extraction From Websites. *Journal Of Student Research*. <https://doi.org/10.47611/Jsr.Vi.942>
- Aprillia, B. (2022). Konstruksi Sosial Masyarakat Tentang Tindakan Kriminalitas Di Desa Parengan. *Prosiding Seminar Nasional Ilmu Ilmu Sosial (SNIIS) Jurusan Ilmu Ilmu Sosial FISH Universitas Negeri Surabaya*, 1, 348–354.
- Arhami, M., Kom, M., & Muhammad Nasir, S. T. (2020). *Data Mining-Algoritma Dan Implementasi*. Penerbit Andi.
- Badan Pusat Statistik (BPS). (2022). *Statistik Kriminal 2022*.
- Badruttamam, A., & Marrudani, D. A. I. (2020). Penerapan Analisis Cluster K-Modes Dengan Validasi Davies Bouldin Index Dalam Menentukan Karakteristik Kanal Youtube Di Indonesia (Studi Kasus: 250 Kanal Youtube Indonesia Teratas Menurut Socialblade). *JURNAL GAUSSIAN*, 9(3), 263–272.
- Binus University. (2021). Metode Data Mining Clustering. <https://sis.binus.ac.id/2021/10/29/Metode-Data-Mining-Clustering/>. Diakses Pada 3 April 2023, 11.48.
- Budiman, R. (2019). Penerapan Data Mining Untuk Menentukan Lokasi Promosi Penerimaan Mahasiswa Baru Pada Universitas Banten Jaya (Metode K-Means Clustering). *Protekinfo (Pengembangan Riset Dan Observasi Teknik Informatika)*, 6, 6-14.
- Dewi, S. M., Windarto, A. P., Damanik, I. S., & Satria, H. (2019). Analisa Metode K-Means Pada Pengelompokan Kriminalitas Menurut Wilayah. *Seminar Nasional Sains & Teknologi Informasi (SENSASI)*, 620–625.
- Fachrurrozi, K., Fahmiwati, F., Hakim, L., Aswadi, A., & Lidiana, L. (2021). Pengaruh Kemiskinan Dan Pengangguran Terhadap Kriminalitas Di

Indonesia Di Tahun 2019. *Jurnal Real Riset*, 3(2), 173-178.

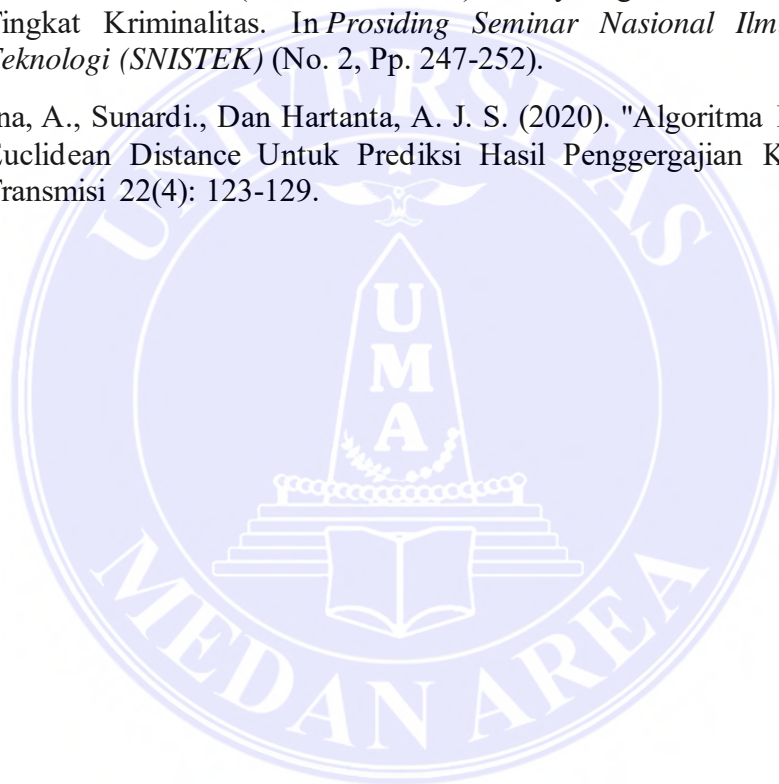
- Fahmi, R. N., Jajuli, M., & Sulistiyowati, N. (2021). Analisis Pemetaan Tingkat Kriminalitas Di Kabupaten Karawang Menggunakan Algoritma K-Means. *INTECOMS: Journal Of Information Technology And Computer Science*, 4(1), 67-79.
- Faid, M., Jasri, M., & Rahmawati, T. (2019). Perbandingan Kinerja Tool Data Mining Weka Dan Rapidminer Dalam Algoritma Klasifikasi. *Teknika*, 8(1), 11-16.
- Fridom Mailo, F., Lazuardi, L., Manajemen Dan Kebijakan Kesehatan Fakultas Kedokteran, D., Masyarakat Dan Keperawatan Universitas Gadjah Mada, K., Sistem Informasi Manajemen Kesehatan Fakultas Kedokteran, D., Masyarakat Dan Keperawatan, K., & Gadjah Mada, U. (2019). Analisis Sentimen Data Twitter Menggunakan Metode Text Mining Tentang Masalah Obesitas Di Indonesia. In *Jurnal Sistem Informasi Kesehatan Masyarakat Journal Of Information Systems For Public Health* (Vol. 4, Issue 1).
- Hartati, S. (2020). Penegakan Hukum Terhadap Tindak Pidana Ketertiban Umum Oleh Pemerintah Kabupaten Boyolali Studi Kasus Putusan Nomor: 19/Pid.B/2016/PN.Byl. *Dinamika Hukum*, 11(2), 219–229.
- Hatta, M. S., Azmi, F., & Setianingsih, C. (2021). Clustering Pada Data Sentimen Penggunaan Transportasi Online Menggunakan Algoritma Spectral Clustering. *Eproceedings Of Engineering*, 8(6).
- Hidayati, Q. R., & Surono, S. (2021). Implementasi Algoritma Spectral Clustering Untuk Analisis Sentimen. *Delta: Jurnal Ilmiah Pendidikan Matematika*, 9(1), 27-36.
- Indriyani, F., & Irfiani, E. (2019). Clustering Data Penjualan Pada Toko Perlengkapan Outdoor Menggunakan Metode K-Means. *JUITA: Jurnal Informatika*, 7(2), 109-113.
- Kamal, M. (2020). Media Sosial Sebagai Budaya Baru Pembelajaran Di SD Muhammadiyah 9 Malang. *Jurnal Komunikasi Nusantara*, 2(1), 17–27. <https://doi.org/10.33366/jkn.v2i1.31>
- Kartikasari, M. D. (2021). Self-Organizing Map Menggunakan Davies-Bouldin Index Dalam Pengelompokan Wilayah Indonesia Berdasarkan Konsumsi Pangan. *Jambura Journal Of Mathematics*, 3(2), 187–196. <https://doi.org/10.34312/jjom.v3i2.10942>
- Kesuma, M. E.-K., & Iskandar, R. (2022). Analisis Toko Dan Asal Toko Fashion Pria Di Shopee Menggunakan Data Scrapping Dan Exploratory Data Analysis. *Majalah Ilmiah Teknologi Elektro*, 21(1). <https://doi.org/10.24843/mite.2022.v21i01.p17>
- Kurniawan, H., & Defit, S. (2020). Data Mining Menggunakan Metode K-Means Clustering Untuk Menentukan Besaran Uang Kuliah Tunggal. *Journal Of Applied Computer Science And Technology*, 1(2), 80-89.

- Li, X., Zhao, X., Chu, D., & Zhou, Z. (2020). An Autoencoder-Based Spectral Clustering Algorithm. *Soft Computing*, 24(3). <https://doi.org/10.1007/S00500-019-03994-5>
- M. Premasundari And C. Yamini. 2019. "A Violent Crime Analysis Using Fuzzy C-Means Clustering Approach". *Ictact J. Soft Comput.* Vol. 6956, No. 4, Page 3, Doi: 10.21917/Ijsc.2019.0270.
- M. R. Gottfredson And T. Hirschi. 2022. A General Theory Of Crime.
- Merdeka.Com. 2023. Kriminalitas Adalah Tindak Kejahatan Yang Melanggar Hukum, Ini Penjelasan Lengkapnya. <https://www.merdeka.com/jatim/kriminalitas-adalah-tindak-kejahatan-yang-melanggar-hukum-ini-penjelasan-nya-klm.html>. Diakses Pada 3 April 2023, 12:28.
- Millenia, W. P. (2022). *Pengelompokan Vaksinasi Covid-19 Di Indonesia Menggunakan Algoritma Spectral Clustering Dengan Un-Normalized Dan Normalized Laplacian* (Doctoral Dissertation, Universitas Muhammadiyah Semarang).
- Navisha, A. N., Ambarwati, R., & Hariasih, M. (2023). Twitter Social Network Interaction As Customer Engagement In Competition For E-Commerce E-Health Performance In Indonesia. *Jurnal Manajerial*, 10(2), 303–324. <https://doi.org/10.30587/Manajerial.V10i02.5279>
- Nishom, M. (2019). Perbandingan Akurasi Euclidean Distance, Minkowski Distance, Dan Manhattan Distance Pada Algoritma K-Means Clustering Berbasis Chi-Square. *Jurnal Informatika*, 4(01), 20-24.
- Nurhafida, S. I., & Sembiring, F. (2021, September). Analisis Text Clustering Masyarakat Di Twiter Mengenai Mcdonald’Sxbts Menggunakan Orange Data Mining. In *Seminar Nasional Sistem Informasi Dan Manajemen Informatika Universitas Nusa Putra* (Vol. 1, No. 01, Pp. 28-35).
- Purwadi, P., Ramadhan, P. S., & Safitri, N. (2019). Penerapan Data Mining Untuk Mengestimasi Laju Pertumbuhan Penduduk Menggunakan Metode Regresi Linier Berganda Pada BPS Deli Serdang. *J. SAINTIKOM (Jurnal Sains Manaj. Inform. Dan Komputer)*, 18(1), 55.
- Putra, A. D., Martha, G. S., Fikram, M., & Yuhan, R. J. (2021). Faktor-Faktor Yang Memengaruhi Tingkat Kriminalitas Di Indonesia Tahun 2018. *Indonesian Journal Of Applied Statistics*, 3(2), 123-131.
- Putri, M. W., Nur, I. M., & Wasono, R. (2022). Implementasi Spectral Clustering Algorithm Untuk Pengelompokan Sasaran Vaksinasi Covid-19 Di Indonesia. *Jurnal Statistika Universitas Muhammadiyah Semarang*, 10(1), 26-31.
- Putri, M. W., Nur, I. M., & Wasono, R. (2022). Implementasi Spectral Clustering Algorithm Untuk Pengelompokan Sasaran Vaksinasi Covid-19 Di Indonesia. *Jurnal Statistika*, 10(1), 26-31.

- Rezeki, S. R. I., Restiviani, Y., & Zahara, R. (2020). Penggunaan Sosial Media Twitter Dalam Komunikasi Organisasi (Studi Kasus Pemerintah Provinsi DKI Jakarta Dalam Penanganan Covid-19). *JOURNAL OF ISLAMIC AND LAW STUDIES*, 4(2), 63–78.
- Rohmah, Q., Dan Sugiyarto. (2021). "Implementasi Algoritma Spectral Clustering Untuk Analisis Sentimen". *Jurnal Ilmiah Pendidikan Matematika*, 9(1), 27-36.
- Salsabila, N. (2019). "Klasifikasi Barang Menggunakan Metode Clustering Kmeans Dalam Penentuan Prediksi Stok Barang". Skripsi. Universitas Islam Negeri Maulana Malik Ibrahim. Malang
- Sari, N. C. & Azhar, Z. (2019). Analisis Kausalitas Kriminalitas, Pendidikan, Dan Kemiskinan Di Indoneisa. *Jurnal Kajian Ekonomi Dan Pembangunan*, 1(2), 635-644.
- Septiani, I. W., Fauzan, Abd. C., & Huda, M. M. (2022). Implementasi Algoritma K-Medoids Dengan Evaluasi Davies-Bouldin-Index Untuk Clusterisasi Harapan Hidup Pasca Operasi Pada Pasien Penderita Kanker Paru-Paru. *Jurnal Sistem Komputer Dan Informatika (JSON)*, 3(4), 556–566. <https://doi.org/10.30865/Json.V3i4.4055>
- Simbolon, C. A. D., & Siahaan, C. (2021). Penggunaan Komunikasi Media Sosial Twitter Di Kalangan Remaja Di Kecamatan Cibinong, Kabupaten Bogor. *JISIP: Jurnal Ilmu Sosial Dan Ilmu Politik*, 10(3), 219–226.
- Sinaga, Y. Y., & Anshori, A. M. (2022). Faktor Penyebab Tingginya Kenakalan Dan Kriminalitas Remaja Dalam Masyarakat. *Jurnal Dakwatul Islam*, 7(1). <https://doi.org/10.46781/dakwatulislam.v7i1.582>
- Sulistiani, H., & Darwis, D. (2020). Penerapan Metode Agile Untuk Pengembangan Online Analytical Processing (OLAP) Pada Data Penjualan (Studi Kasus: CV Adilia Lestari). *J. Coreit*, 6(1), 50-56.
- Suriani, L. (2020). Pengelompokan Data Kriminal Pada Poldasu Menentukan Pola Daerah Rawan Tindak Kriminal Menggunakan Data Mining Algoritma K-Means Clustering. *Jurnal Sistem Komputer Dan Informatika (JSON)*, 1(2), 151-157.
- Tamba, S. P., & Kesuma, F. T. (2019). Penerapan Data Mining Untuk Menentukan Penjualan Sparepart Toyota Dengan Metode K-Means Clustering: Data Mining; K-Means-Clustering. *Jurnal Sistem Informasi Dan Ilmu Komputer Prima (JUSIKOM PRIMA)*, 2(2), 67-72.
- Utomo, D. P., & Mesran, M. (2020). Analisis Komparasi Metode Klasifikasi Data Mining Dan Reduksi Atribut Pada Data Set Penyakit Jantung. *Jurnal Media Informatika Budidarma*, 4(2), 437-444.
- Wahyudi, I., Burhanis Sulthan, M., & Suhartini, L. (2021). Analisa Penentuan Cluster Terbaik Pada Metode K-Means Menggunakan Elbow Terhadap Sentra Industri Produksi Di Pamekasan. In *Jurnal Aplikasi Teknologi*

Informasi Dan Manajemen (JATIM) (Vol. 2, Issue 2).

- Wulandari, S. (2020). Clustering Microarray Adenoma Menggunakan Spectral Clustering Dengan Algoritma Partitioning Around Medoid (PAM). In *SINASIS (Seminar Nasional Sains)* (Vol. 1, No. 1).
- Wulandari, S., & Novita, D. (2021). Analisis Clustering Virus MERS-Cov Menggunakan Metode Spectral Clustering Dan Algoritma K-Means. *STRING (Satuan Tulisan Riset Dan Inovasi Teknologi)*, 5(3), 315-323.
- Yamini, M. P. C. (2019). A Violent Crime Analysis Using Fuzzy C-Means Clustering Approach. *ICTACT Journal On Soft Computing*, 9(3), 1939-1944.
- Yenni, Y., & Utnasari, I. (2019, November). Fuzzy Logic Mamdani Memprediksi Tingkat Kriminalitas. In *Prosiding Seminar Nasional Ilmu Sosial Dan Teknologi (SNISTEK)* (No. 2, Pp. 247-252).
- Yudhana, A., Sunardi., Dan Hartanta, A. J. S. (2020). "Algoritma K-NN Dengan Euclidean Distance Untuk Prediksi Hasil Penggajian Kayu Sengon". *Transmisi* 22(4): 123-129.



LAMPIRAN

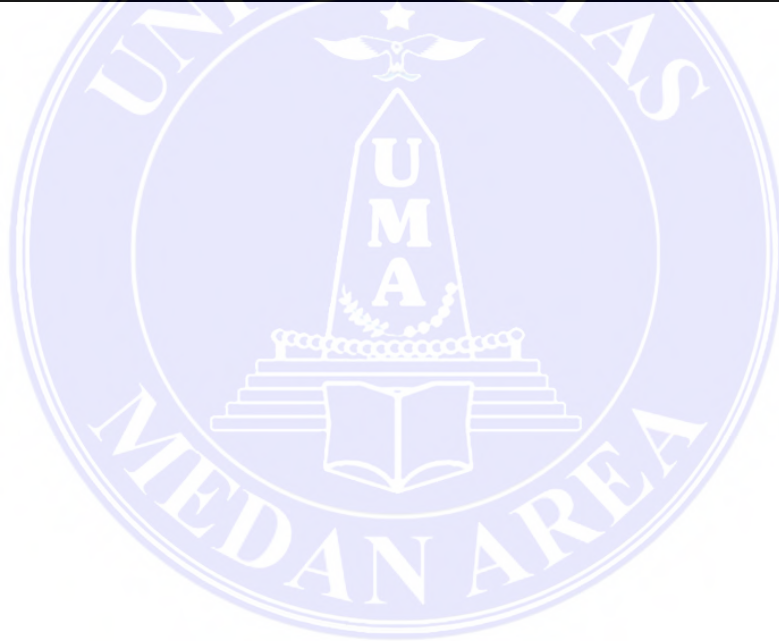
1. SAMPEL DATASET SEBELUM DIOLAH

Query	Tweet	Tanggal	Lokasi
kejahatan	@dearmauv iya kalo di taxi driver kan dia jahat banget tuh nah kalo di crash course kejahatan dia ngapain ke guru Choi?	2023-06-25 11:52:22+00:00	ã...^ã...^ã...Ž ã,,±ã,,·ã...‡
kejahatan	ya Allah apa sih yg membuat manusia bisa kejam hatinya? kejahatan ada di mana2 banget ini....	2023-06-25 11:48:51+00:00	Tidak tersedia
kejahatan	UKM Anggota Polsek Banguntapan stanby di Simpang 4 Blok O, Antisipasi gangguan Kamtibmas maupun tindak kejahatan #PolsekBanguntapan #PolriPresisi #PolresBantul https://t.co/09WJWF820a	2023-06-25 11:45:50+00:00	Tidak tersedia
kejahatan	Patroli ikon kamtibmas Polsek Banguntapan harkamtibmas cegah kejahatan jalanan, pencurian dan tawuran. Wilayah aman kondusif #PatroliJogja #harkamtibmas #PolsekBanguntapan #PolriPresisi https://t.co/eELdti5EaP	2023-06-25 11:43:53+00:00	Tidak tersedia
kejahatan	Operasi Kejahatan Jalanan Cegah Pelaku Tawuran, Geng Motor dan Pelaku Begal, Mnggu (25/6). https://t.co/YWYKbn0GDs	2023-06-25 11:42:17+00:00	Tidak tersedia
kejahatan	dunia teknologi semakin maju, dan kejahatan semakin banyak di mana-mana, stay safe y'all!	2023-06-25 11:41:01+00:00	Tidak tersedia
kejahatan	Apel Operasi Kejahatan Jalanan Berikan Rasa Aman Bagi Warga yang Beraktifitas di Malam Hari, Minggu (25/6). https://t.co/7sth6djtHH	2023-06-25 11:39:46+00:00	Tidak tersedia
kejahatan	Personil Polsek Mlati Melaksanakan patroli dialogis dan mencegah kejahatan maupun gangguan Kamtibmas, dipimpin Aiptu Ihsanudin, strong point di Mbako 88 Cebongan Sumberadi Mlati Sleman. #polripresisi #polsekmlati	2023-06-25 11:32:27+00:00	Mlati, Indonesia

	#polrestasleman #poldajogja #PatroliJogja https://t.co/ogou5RrEZd		
kejahatan	Personil Polsek Mlati Melaksanakan patroli dialogis dan mencegah kejahatan maupun gangguan Kamtibmas, dipimpin Aiptu Ihsanudin, strong point di Mustofa Barkas Tlogoadi Mlati Sleman. #polripresisi #polsekmlati #polrestasleman #poldajogja #PatroliJogja https://t.co/mjyYLV8kID	2023-06-25 11:28:40+00:00	Mlati, Indonesia
kejahatan	Patroli antisipasi kejahatan (Curat, Curas, Curanmor dan Tindak Pidana Lainnya) dipimpin Aiptu Ihsanudin, Strong point' di Toko Elektronik Cebongan Tlogoadi Mlati Sleman. Kegiatan berlangsung aman dan kondusif. #polripresisi #polsekmlati #polrestasleman #poldajogja #PatroliJogja https://t.co/gZYP9xvcJL	2023-06-25 11:23:55+00:00	Mlati, Indonesia

2. SAMPEL DATASET TELAH DIOLAH

text_clean	token	normal	stop	stemmed	final
dearmauv iya kalo di taxi driver kan dia jahat banget tuh nah kalo di crash course kejahatan dia ngapain ke guru choi	['dearmauv', 'iya', 'kalo', 'di', 'taxi', 'driver', 'kan', 'dia', 'jahat', 'banget', 'tuh', 'nah', 'kalo', 'di', 'crash', 'course', 'kejahatan', 'dia', 'ngapain', 'ke', 'guru', 'choi']	['dearmauv', 'iya', 'kalau', 'di', 'taxi', 'driver', 'kan', 'dia', 'jahat', 'banget', 'tuh', 'nah', 'kalau', 'kalo', 'di', 'crash', 'course', 'kejahatan', 'dia', 'ngapain', 'ke', 'guru', 'choi']	['dearmauv', 'iya', 'kalau', 'taxi', 'driver', 'kan', 'jahat', 'banget', 'tuh', 'nah', 'kalau', 'crash', 'course', 'kejahatan', 'ngapain', 'guru', 'choi']	['dearmauv', 'iya', 'kalau', 'taxi', 'driver', 'kan', 'jahat', 'banget', 'tuh', 'nah', 'kalau', 'crash', 'course', 'jahat', 'ngapain', 'guru', 'choi']	iya kalau kan jahat banget kalau crash jahat ngapain guru



3. Kode Program

a. Pengumpulan Data (*Crawling Data*)

```

!pip install pandas
!pip install
git+https://github.com/JustAnotherArchivist/snsrape.git

import pandas as pd
import snsrape.modules.twitter as sntwitter
import itertools

# Daftar query yang terkait dengan kejahatan dan ketertiban umum
queries = ['kejahatan', 'pembunuhan', 'pencurian',
           'perampokan', 'narkoba', 'pemeriksaan', 'penipuan', 'curat',
           'perdagangan manusia',
           'penggelapan', 'pemalsuan', 'penculikan',
           'penyerangan', 'pemeriksaan', 'kekerasan', 'pelanggaran',
           'terorisme', 'pemeriksaan',
           'korupsi', 'penyalahgunaan narkoba', 'pencabulan',
           'pemeriksaan anak', 'penadahan', 'kriminalitas jalanan',
           'perjudian', 'pencemaran nama baik']

# Jumlah tweet yang ingin Anda scrape untuk setiap query
tweet_count = 10000

# Membuat list kosong untuk menyimpan hasil scraping
tweets = []

# Melakukan scraping data dari Twitter untuk setiap query
for query in queries:
    # Melakukan scraping data dari Twitter
    for i, tweet in
enumerate(sntwitter.TwitterSearchScrapper(query + '
lang:id').get_items()):
        tweet_text = tweet.content
        tweet_date = tweet.date
        tweet_location = tweet.user.location if
tweet.user.location else 'Tidak tersedia'

        # Menambahkan data tweet ke list
        tweets.append([query, tweet_text, tweet_date,
tweet_location])

        if i+1 == tweet_count:
            break

```



```

# Membuat dataframe dari list tweets
df_tweets = pd.DataFrame(tweets, columns=['Query', 'Tweet',
'Tanggal', 'Lokasi'])

df_tweets.shape

# Menyimpan dataframe ke file CSV
df_tweets.to_csv('data_twitter.csv', index=False)

```

b. Processing Data

```

!pip install nltk
!pip install Sastrawi

import pandas as pd
import re

import string

def preprocess_text(kalimat):
    # Mengubah kalimat menjadi huruf kecil
    lower_case = kalimat.lower()

    # Menghapus angka dari kalimat
    hasil = re.sub(r"\d+", "", lower_case)

    # Menghapus tanda baca dari kalimat
    hasil =
hasil.translate(str.maketrans("", "", string.punctuation))

    # Menghapus spasi pada awal dan akhir kalimat
    hasil = hasil.strip()

    # Menghapus Username Twitter
    hasil = re.sub('@[^\s]+', '', hasil)

    # Menghapus https dan http
    hasil = re.sub(r"(?:\@|http?:\/\/|https?:\/\/|www)\S+", "",
hasil)

    # Menghilangkan Tanda Baca
    hasil = hasil.translate(str.maketrans('', '',
string.punctuation))

    # Mengganti karakter HTML dengan tanda petik
    hasil = re.sub('<.*?>', ' ', hasil)

```

```

# Mempertimbangkan huruf dan angka
hasil = re.sub('[^a-zA-Z0-9]', ' ', hasil)

# Mengganti line baru dengan spasi
hasil = re.sub("\n", " ", hasil)

# Menghapus single char
hasil = re.sub(r"\b[a-zA-Z]\b", " ", hasil)

# Menghilangkan emoji
emoji_pattern = re.compile("[
                                u"\U0001F600-\U0001F64F" #
emotikon wajah
                                u"\U0001F300-\U0001F5FF" #
simbol & benda
                                u"\U0001F680-\U0001F6FF" #
transportasi & simbol peralatan
                                u"\U0001F1E0-\U0001F1FF" #
bendera negara
                                u"\U00002500-\U00002BEF" #
karakter CJK (Chinese, Japanese, Korean)
                                u"\U00002702-\U000027B0" #
simbol & tanda
                                u"\U00002702-\U000027B0"
                                u"\U000024C2-\U0001F251"
                                u"\U0001f926-\U0001f937"
                                u"\U00010000-\U0010ffff"
                                u"\u200d"
                                u"\u2640-\u2642"
                                u"\u2600-\u2B55"
                                u"\u23cf"
                                u"\u23e9"
                                u"\u231a"
                                u"\u3030"
                                u"\ufe0f"
                                "]" +, flags=re.UNICODE)
hasil = emoji_pattern.sub(r'', hasil)

# Memisahkan dan menggabungkan kata
hasil = ' '.join(hasil.split())

return hasil

kamus = pd.read_csv('kamusalay.csv', header=None)
kamus.columns = ['kata_asli', 'kata_normal']

```

```

def normalisasi_teks(teks, kamus):
    kata_asli = teks.split()
    kata_normal = []

    for kata in kata_asli:
        normal = kamus[kamus['kata_asli'] ==
kata]['kata_normal'].values
        if normal:
            kata_normal.append(normal[0])
        else:
            kata_normal.append(kata)

    return ' '.join(kata_normal)

import nltk
nltk.download('all')
from nltk.tokenize import word_tokenize

def tokenize_text(kalimat):
    tokens = nltk.tokenize.word_tokenize(kalimat)
    return tokens

#Normalisasi
normalized_word = pd.read_csv("Normalisasi.csv",
encoding='latin1')

normalized_word_dict={}
for index, row in normalized_word.iterrows():
    if row[0] not in normalized_word_dict:
        normalized_word_dict[row[0]] = row[1]

#Filtering (Stopword Removal)
from Sastrawi.StopWordRemover.StopWordRemoverFactory import
StopWordRemoverFactory
factory = StopWordRemoverFactory()
stopwords = factory.get_stop_words()
# stopwords.extend([
#     'diri',
#     'akhir',
#     'akhiri',
#     'asalkan',
#     'belum',
#     'berakhir',
#     'bila',
#     'biasanya',
#     'bukan',
#     'diakhiri',

```

```

#     'diakhirinya',
#     'guna',
#     'lagi',
#     'kurang'
# ])
print(stopwords)

def stopwords_text(tokens):
    cleaned_tokens = []
    for token in tokens:
        if token not in stopwords:
            cleaned_tokens.append(token)
    return cleaned_tokens

#Stemming
from Sastrawi.Stemmer.StemmerFactory import StemmerFactory
stem_factory = StemmerFactory()
stemmer = stem_factory.create_stemmer()

def stemming_text(tokens):
    hasil = [stemmer.stem(token) for token in tokens]
    return hasil

#Normalisasi ke-2
with open('indonesian-words.txt', 'r', encoding='utf-8') as file:
    kamus_bahasa = set(file.read().splitlines())

def normalisasi_teks_part2(tokens, kamus, kamus_bahasa):
    kata_normal = []
    for kata in tokens:
        normal = kamus[kamus['kata_asli'] ==
kata]['kata_normal'].values
        if normal:
            kata_normal.append(normal[0])
        else:
            kata_normal.append(kata)
    kata_normal = [kata for kata in kata_normal if kata in
kamus_bahasa]
    return kata_normal

df = pd.read_csv('data_twitter.csv')
df.head()

df['text_clean'] = df['Tweet'].apply(preprocess_text)
df['normal'] = df['text_clean'].apply(normalisasi_teks,
kamus=kamus)

```

```

df['token'] = df['normal'].apply(tokenize_text)
df['normal_1'] = df['token'].apply(normalized_term)
df['stop'] = df['normal_1'].apply(stopword_text)
df['stemmed'] = df['stop'].apply(stemming_text)
df['normal_2'] = df['stemmed'].apply(normalisasi_teks_part2,
kamus=kamus, kamus_bahasa=kamus_bahasa)
df['final'] = df['normal_2'].apply(lambda x: ' '.join(x))

df = df[df['normal_2'].map(lambda x: len(x) > 0)]
df.shape

# Menyimpan DataFrame ke file CSV
df.to_csv('data_clean.csv', index=False)
print(f"Data pada file telah di-preprocess dan disimpan
kembali.")

```

c. Implementasi Metode *Spectral Clustering*

```

import pandas as pd
import numpy as np
from sklearn.feature_extraction.text import TfidfVectorizer
from sklearn.cluster import SpectralClustering
from sklearn.metrics import davies_bouldin_score
from scipy.sparse import linalg
import matplotlib.pyplot as plt

df = pd.read_csv('data_clean.csv')
df.head()

# Mengambil 20.000 data pertama
df = df.head(10000)

# 3. Feature Extraction
tfidf = TfidfVectorizer()
tfidf_matrix = tfidf.fit_transform(df['final'])

print(tfidf_matrix)

# 4. Similarity Graph
similarity_matrix = tfidf_matrix * tfidf_matrix.T

# Membuat DataFrame dari matriks similaritas
similarity_df = pd.DataFrame(similarity_matrix.toarray(),
index=df.index, columns=df.index)
similarity_df.head()

```



```

# Visualisasi Similarity Graph
plt.figure(figsize=(8, 6))
plt.imshow(similarity_matrix.toarray(), cmap='hot',
            interpolation='nearest')
plt.title('Similarity Graph')
plt.colorbar()
plt.show()

# 5. Diagonal Graph
diagonal_matrix =
np.diag(similarity_matrix.toarray().sum(axis=1))

# Membuat DataFrame dari matriks diagonal
diagonal_df = pd.DataFrame(diagonal_matrix, index=df.index,
                           columns=df.index)
diagonal_df.head()

# Visualisasi Diagonal Graph
plt.figure(figsize=(6, 6))
plt.imshow(diagonal_matrix, cmap='hot',
            interpolation='nearest')
plt.title('Diagonal Graph')
plt.colorbar()
plt.show()

# 6. Eigen Value dan Eigen Vector
laplacian_matrix = diagonal_matrix - similarity_matrix
eigen_values, eigen_vectors = linalg.eigsh(laplacian_matrix)

# Membuat DataFrame dari eigenvalues dan eigenvectors
eigen_df = pd.DataFrame(eigen_vectors, columns=[f'Eigen_{i+1}'
for i in range(eigen_vectors.shape[1])])

eigen_df.head()

# Visualisasi Eigen Values
plt.plot(range(1, len(eigen_values) + 1), eigen_values,
         marker='o')
plt.xlabel('Eigen Index')
plt.ylabel('Eigen Value')
plt.title('Eigen Values')
plt.show()

from sklearn.cluster import KMeans

# 7. Penentuan Jumlah K
k = 3 # Jumlah cluster yang diinginkan

```

```

# 8. K-Means Clustering
model = KMeans(n_clusters=k, random_state=0)
clusters = model.fit_predict(tfidf_matrix)

# Menampilkan hasil centroid
centroid_df = pd.DataFrame(model.cluster_centers_,
columns=tfidf.get_feature_names_out())
print("Centroid:")
print(centroid_df)

# Menghitung jarak data ke pusat cluster
distances_to_centers = model.transform(tfidf_matrix)
distances_df = pd.DataFrame(distances_to_centers,
columns=[f'Distance_to_Center_{i+1}' for i in range(k)])
print("\nJarak data ke pusat cluster:")
print(distances_df)

# Menampilkan pusat cluster
cluster_centers_df = pd.DataFrame(model.cluster_centers_,
columns=tfidf.get_feature_names_out())
print("\nPusat Cluster:")
print(cluster_centers_df)

from sklearn.decomposition import PCA
import matplotlib.pyplot as plt

# Menerapkan PCA untuk mereduksi dimensi data
pca = PCA(n_components=2)
reduced_features = pca.fit_transform(tfidf_matrix.toarray())

# Visualisasi distribusi cluster
plt.scatter(reduced_features[:, 0], reduced_features[:, 1],
c=clusters, cmap='viridis')
plt.title("Distribusi Cluster")
plt.xlabel("PC1")
plt.ylabel("PC2")
plt.show()

# 9. Evaluasi DBI
dbi_score = davies_bouldin_score(tfidf_matrix.toarray(),
clusters)
print("DBI Score:", dbi_score)

# 10. Visualisasi
cluster_counts = pd.Series(clusters).value_counts()
labels = ['Cluster 2', 'Cluster 1', 'Cluster 0']

```

```

cluster_labels = [labels[i] for i in clusters]
plt.pie(cluster_counts, labels=labels, autopct='%1.1f%%')
plt.title("Distribusi Tingkat Kejahatan")
plt.show()

# Menghitung jumlah kejahatan dalam setiap cluster
jumlah_kejahatan =
pd.Series(clusters).value_counts().sort_index()

# Membuat DataFrame distribusi kejahatan
df_distribusi_kejahatan = pd.DataFrame({'Cluster':
jumlah_kejahatan.index, 'Jumlah Kejahatan':
jumlah_kejahatan.values})

# Menampilkan tabel distribusi kejahatan
df_distribusi_kejahatan.head()

# Menggabungkan hasil clustering dengan data asli
df_clustered = pd.concat([df, pd.Series(clusters,
name='Cluster')], axis=1)
df_clustered.head()

df_clustered.shape

# Menampilkan isi dari masing-masing cluster
for i in range(k):
    cluster_data = df_clustered[df_clustered['Cluster'] == i]
    crimes = cluster_data['final'].unique()
    print(f"Cluster {i} - Jumlah Kejahatan: {len(crimes)}")
    print(crimes)
    print()

# Baca isi file "berbahaya.txt" dan simpan dalam list
with open('berbahaya.txt', 'r') as f:
    berbahaya_keywords = f.read().splitlines()

# Baca isi file "cukup_berbahaya.txt" dan simpan dalam list
with open('cukup_berbahaya.txt', 'r') as f:
    cukup_berbahaya_keywords = f.read().splitlines()

# Baca isi file "sangat_berbahaya.txt" dan simpan dalam list
with open('sangat_berbahaya.txt', 'r') as f:
    sangat_berbahaya_keywords = f.read().splitlines()

# Fungsi untuk melakukan labelisasi
def label_tingkat_kejahatan(cluster_id, cluster_data):
    if cluster_id == 0:

```

```

        return 'Cukup Berbahaya'
    elif cluster_id == 1:
        return 'Berbahaya'
    elif cluster_id == 2:
        return 'Sangat Berbahaya'

# Mengelompokkan data berdasarkan cluster
grouped_clusters = df_clustered.groupby('Cluster')

# Inisialisasi kolom label
df_clustered['Tingkat Kejahatan'] = None

# Melabeli setiap cluster
for cluster_id, cluster_data in grouped_clusters:
    label = label_tingkat_kejahatan(cluster_id, cluster_data)
    df_clustered.loc[cluster_data.index, 'Tingkat Kejahatan'] =
label

# Menghapus baris dengan label "Tidak terlabel"
df_clustered = df_clustered.dropna(subset=['Tingkat
Kejahatan'])

# Menampilkan data dengan kolom label
print(df_clustered[['final', 'Cluster', 'Tingkat Kejahatan']])

# Menghitung jumlah masing-masing tindak kejahatan dalam setiap
cluster
cluster_counts = df_clustered['Tingkat
Kejahatan'].value_counts()

# Menampilkan hasil
print("Jumlah masing-masing tindak kejahatan dalam setiap
cluster:")
print(cluster_counts)

df_clustered = df_clustered.dropna()

df_clustered[['final', 'Cluster', 'Tingkat
Kejahatan']].sample(5)

# Menampilkan jenis kejahatan teratas pada masing-masing
cluster
cluster_labels = df_clustered['Tingkat Kejahatan'].unique()
for label in cluster_labels:
    cluster_data = df_clustered[df_clustered['Tingkat
Kejahatan'] == label]
    top_crimes = cluster_data['final'].value_counts().head(5)

```

```

print(f"Cluster {label}:")
print(top_crimes)
print()

# Menampilkan jenis kejahatan teratas pada masing-masing
cluster
cluster_labels = df_clustered['Tingkat Kejahatan'].unique()
for label in cluster_labels:
    cluster_data = df_clustered[df_clustered['Tingkat
Kejahatan'] == label]
    top_crimes = cluster_data['final'].value_counts().sample(5)
    print(f"Cluster {label}:")
    print(top_crimes)
    print()

from matplotlib_venn import venn3
import matplotlib.pyplot as plt

# Menghitung jumlah data dalam masing-masing cluster
cluster_counts = df_clustered['Tingkat
Kejahatan'].value_counts()

# Mendapatkan data jumlah data yang ada di setiap cluster
cluster_0 = cluster_counts['Cukup Berbahaya']
cluster_1 = cluster_counts['Berbahaya']
cluster_2 = cluster_counts['Sangat Berbahaya']

# Membuat diagram Venn
venn_labels = {'100': cluster_0, '010': cluster_1, '001':
cluster_2}
venn = venn3(subsets=venn_labels, set_labels=('Cukup
Berbahaya', 'Berbahaya', 'Sangat Berbahaya'))

# Menampilkan diagram Venn
plt.title('Hubungan Cluster Tingkat Kejahatan')
plt.show()

import matplotlib.pyplot as plt

# Menghitung skor DBI untuk setiap nilai k
k_values = range(2, 10) # Ubah range sesuai dengan jumlah k
yang ingin dievaluasi
dbi_scores = []

for k in k_values:

```




```
model = SpectralClustering(n_clusters=k, affinity='linear')
clusters = model.fit_predict(tfidf_matrix.toarray())
dbi_score = davies_bouldin_score(tfidf_matrix.toarray(),
clusters)
dbi_scores.append(dbi_score)

# Plot skor DBI
plt.figure(figsize=(8, 6))
plt.plot(k_values, dbi_scores, marker='o')
plt.xlabel('Jumlah Cluster (k)')
plt.ylabel('Skor DBI')
plt.title('Evaluasi Davies-Bouldin Index (DBI)')
plt.grid(True)
plt.show()
```



4. SK Pembimbing



UNIVERSITAS MEDAN AREA
FAKULTAS TEKNIK

Kampus I : Jalan Kolang Nomor 1 Medan Estate/Jalan PBSI Nomor 1 ☎(061) 7366878, 7360168, 7364348, 7366781, Fax.(061) 7366998 Medan 20223
Kampus II : Jalan Setiabudi Nomor 79 / Jalan Sel Serayu Nomor 70 A, ☎ (061) 8225602, Fax. (061) 8226331 Medan 20122
Website: www.teknik.uma.ac.id E-mail: univ_medanarea@uma.ac.id

Nomor : 730/FT.6/01.10/X/2023
Lamp : -
Hal : **Perubahan Judul Tugas Akhir**

16 Oktober 2023

Yth, Pembimbing Tugas Akhir
Dr. Rahmadsyah, S.Kom, M. Kom
di
Tempat

Dengan hormat, Sehubungan dengan adanya perubahan judul tugas akhir maka perlu diterbitkan kembali SK Pembimbing Skripsi baru atas nama mahasiswa tersebut :

N a m a : Egi Pratama
N P M : 198160027
Jurusan : Teknik Informatika

Maka dengan hormat kami mengharapkan kesediaan saudara :


Dr. Rahmadsyah, S.Kom, M. Kom (Sebagai Pembimbing)

Adapun Tugas Akhir Skripsi berjudul :


“Analisis Metode Clustering dengan Algoritma Spectral Clustering dalam Pengelompokan Tingkat Tindak Kriminalitas”.

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

Demikian kami sampaikan, atas kesediaan saudara diucapkan terima kasih.


Dekan
Dr. Rahmad Syah, S. Kom, M. Kom

5. SK Riset

**UNIVERSITAS MEDAN AREA**
FAKULTAS TEKNIK

Kampus I : Jalan Kolam Nomor 1 Medan Estate/Jalan PBSI Nomor 1 ☎ (061) 7366878, 7360168, 7364348, 7366781, Fax. (061) 7366998 Medan 20223
Kampus II : Jalan Setiabudi Nomor 79 / Jalan Sei Serayu Nomor 70 A, ☎ (061) 8225602, Fax. (061) 8226331 Medan 20122
Website: www.teknik.uma.ac.id E-mail: univ_medanarea@uma.ac.id

Nomor : 397 /FT.6/01.10/V/2023 29 Mei 2023
Lamp : -
Hal : **Penelitian Dan Pengambilan Data Tugas Akhir**

Yth. Kepala Kepolisian Resort Asahan
Jln. Jend. A. Yani No.110, Kisaran Naga
Di
Asahan

Dengan hormat,
Kami mohon kesediaan Bapak/Ibu berkenan untuk memberikan izin dan kesempatan kepada mahasiswa kami tersebut dibawah ini :



NO	NAMA	NPM	PRODI
1	Egi Pratama	198160027	Teknik Informatika

Untuk melaksanakan Penelitian dan Pengambilan Data Tugas Akhir pada perusahaan/Instansi yang Bapak/Ibu Pimpin.

Perlu kami jelaskan bahwa Pengambilan Data tersebut adalah semata-mata untuk tujuan ilmiah dan Skripsi yang merupakan salah satu syarat bagi mahasiswa tersebut untuk mengikuti ujian sarjana pada Fakultas Teknik Universitas Medan Area dan tidak untuk dipublikasikan, dengan judul penelitian :

Analisis Metode Clustering dengan Algoritma Spectral Clustering Pengelompokan Wilayah Rawan Kriminalitas di Asahan

Atas perhatian dan kerja sama yang baik diucapkan terima kasih.



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

Tembusan :
1. Ka. BAMAI
2. Mahasiswa
3. File

6. SK Selesai Riset

KEPOLISIAN NEGARA REPUBLIK INDONESIA
DAERAH SUMATERA UTARA
RESOR ASAHAN
Jl. Jend. A. Yani No. 110 Kisaran - 21224



SURAT KETERANGAN
Nomor : SK / 289 / VI / 2023 / Reskrim

Yang bertanda tangan di bawah ini, menerangkan bahwa :

Nama : EGI PRATAMA
NPM : 198160027
Program Studi : Teknik Informatika
Universitas : Universitas Medan Area Fakultas Teknik Medan
Alamat Tempat Tinggal : Dsn IV Ds. Silo Bonto Kec. Silau Laut Kab. Asahan
NIK : 1209282601010001

Adalah benar telah selesai Survey / Research bahan penulisan Skripsi Mahasiswa tersebut
Berjudul : " Analisis Metode Clustering Dengan Algoritma Spectral Clustering Pengelompokan Tingkat
Tindak Kriminalitas"


Demikian Surat Keterangan ini dibuat dengan sebenarnya untuk dapat dipergunakan seperlu
nya dan apabila dikemudian hari terdapat kekeliruan akan diadakan perbaikan.

Dikeluarkan di : Kisaran
pada tanggal : 26 Juni 2023


a.n. KEPALA KEPOLISIAN RESOR ASAHAN
KASAT RESKRIM

MUHAMMAD SAID HUSEN, S.I.K.
AJUN KOMISARIS POLISI NRP 93030354

7. Hasil Turnitin


Similarity Report ID: oid:29477:50793014

23% Overall Similarity

Top sources found in the following databases:

- 18% Internet database
- 8% Publications database
- Crossref database
- Crossref Posted Content database
- 15% Submitted Works database

TOP SOURCES

The sources with the highest number of matches within the submission. Overlapping sources will not be displayed.

1	etd.umy.ac.id Internet	2%
2	proceeding.unindra.ac.id Internet	1%
3	docplayer.info Internet	1%
4	media.neliti.com Internet	<1%
5	jurnal.unimus.ac.id Internet	<1%
6	researchgate.net Internet	<1%
7	Universitas Raharja on 2023-04-10 Submitted works	<1%
8	Universitas Brawijaya on 2017-07-12 Submitted works	<1%

Sources overview