

Analisi Sentimen Kepercayaan Publik Terhadap Donald Trump pada Aplikasi X dengan Metode K-Means Clustering

SKRIPSI SARJANA INFORMATIKA

Karya ilmiah sebagai salah satu syarat untuk memperoleh gelar Sarjana Informatika dari Fakultas Teknologi Komunikasi dan Informatika

Oleh :

Wahyu Setiya Ramadan

217064516130



**PROGRAM STUDI INFORMATIKA
FAKULTAS TEKNOLOGI KOMUNIKASI DAN
INFORMATIKA UNIVERSITAS NASIONAL
2025**

Analisi Sentimen Kepercayaan Publik Terhadap Donald Trump pada Aplikasi X dengan Metode K-Means Clustering

SKRIPSI SARJANA INFORMATIKA

Karya ilmiah sebagai salah satu syarat untuk memperoleh gelar Sarjana Informatika dari Fakultas Teknologi Komunikasi dan Informatika

Oleh :

Wahyu Setiya Ramadan

217064516130



**PROGRAM STUDI INFORMATIKA
FAKULTAS TEKNOLOGI KOMUNIKASI DAN
INFORMATIKA UNIVERSITAS NASIONAL
2025**

HALAMAN PENGESAHAN

TUGAS AKHIR

**ANALISIS SENTIMEN KEPERCAYAAN PUBLIK TERHADAP
DONALD TRUMP PADA APLIKASI X DENGAN METODE K-
MEANS CLUSTERING**



Wahyu Setiya Ramadan

217064516130

Dosen Pembimbing

A handwritten signature in black ink that reads 'Rima'.

Rima Tamara Aldisa, S.Kom., M.Kom

NIDN. 0318019401

PERNYATAAN KEASLIAN TUGAS AKHIR

Saya menyatakan dengan sesungguhnya bahwa Tugas Akhir dengan judul :

ANALISIS SENTIMEN KEPERCAYAAN PUBLIK TERHADAP DONALD TRUMP PADA APLIKASI X DENGAN METODE K- MEANS CLUSTERING

Yang dibuat untuk melengkapi salah satu persyaratan menjadi Sarjana Komputer pada Program Studi Informatika, Fakultas Teknologi Komunikasi dan Informatika Universitas Nasional, sebagaimana yang saya ketahui adalah bukan merupakan tiruan atau publikasi dari Tugas Akhir yang pernah diajukan atau dipakai untuk mendapatkan gelar di lingkungan Universitas Nasional maupun perguruan tinggi atau instansi lainnya, kecuali pada bagian – bagian tertentu yang menjadi sumber informasi atau acuan yang dicantumkan sebagaimana mestinya.

Jakarta, 28 Februari 2025



Wahyu Setiya Ramadan
217064516130

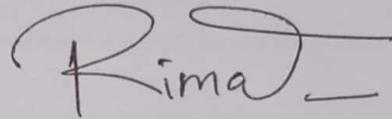
LEMBAR PERSETUJUAN REVIEW AKHIR

Tugas Akhir dengan judul :

ANALISIS SENTIMEN KEPERCAYAAN PUBLIK TERHADAP DONALD TRUMP PADA APLIKASI X DENGAN METODE K- MEANS CLUSTERING

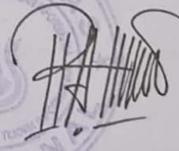
Dibuat untuk melengkapi salah satu persyaratan menjadi Sarjana Komputer pada Program Studi Informatika, Fakultas Teknologi Komunikasi dan Informatika Universitas Nasional. Tugas Akhir ini diujikan pada Sidang Review Akhir Semester Ganjil 2024-2025 pada tanggal 25 Februari Tahun 2025

Dosen Pembimbing 1



Rima Tamara Aldisa, S.Kom., M.Kom
NIDN.0318019401

Ketua Program Studi



Ratih Titi Komala Sari, S.T., M.M., MMSI
NIDN.0301038302

LEMBAR PERSETUJUAN JUDUL YANG TIDAK ATAU YANG DIREVISI

Nama : Wahyu Setiya Ramadan
NPM : 217064516130
Fakultas/Akademi : Fakultas Teknologi Komunikasi dan Informatika
Program Studi : Informatika
Tanggal Sidang : 25 Februari 2025

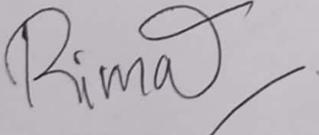
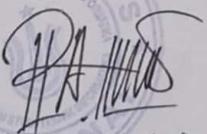
JUDUL DALAM BAHASA INDONESIA :

ANALISIS SENTIMEN KEPERCAYAAN PUBLIK TERHADAP DONALD TRUMP
PADA APLIKASI X DENGAN METODE K-MEANS CLUSTERING

JUDUL DALAM BAHASA INGGRIS :

ANALYSIS OF PUBLIC CONFIDENCE SENTIMENT TOWARDS DONALD
TRUMP ON X APPLICATION USING THE K-MEANS CLUSTERING
METHOD

TANDA TANGAN DAN TANGGAL

Pembimbing 1	Ka. Prodi	Mahasiswa
TGL : 28 Februari 2025	TGL : 28 Februari 2025	TGL : 28 Februari 2025
	 RATIH TITI K.S.	

KATA PENGANTAR

Segala puji bagi Allah SWT atas rahmat dan karunia-Nya, yang memungkinkan penulis menyelesaikan skripsi berjudul "**Analisis Sentimen Kepercayaan Publik terhadap Donald Trump pada Aplikasi X dengan Metode K-Means Clustering**". Skripsi ini disusun untuk memenuhi syarat meraih gelar Sarjana Komputer di Program Studi Informatika, Universitas Nasional. Penulis mengucapkan terima kasih kepada semua pihak yang memberikan dukungan, bimbingan, dan motivasi selama proses penulisan. Tanpa bantuan mereka, skripsi ini tidak akan terselesaikan. Kepada Allah SWT, yang telah memberikan berkah, petunjuk, dan anugerah-Nya dalam perjalanan kami menyelesaikan skripsi ini.

1. Orang Tua kami, Bapak Sudiadi dan Ibu Raminah serta Paman dan pacar saya yang selalu support saya yaitu Sarwono dan Dias Indriyani yang senantiasa memberikan cinta, doa, dukungan, dan pengorbanan tanpa henti selama kami menempuh perjalanan ini. Terima kasih atas kesabaran, motivasi, dan dorongan yang tak pernah berhenti dari kalian.
2. Ibu Dosen Pembimbing, Rima Tamara Aldisa. S.Kom., M.Kom. atas bimbingan, arahan, dan masukan berharga yang telah diberikan selama proses penyusunan skripsi ini. Ibu telah meluangkan waktu dan tenaga untuk membimbing kami dengan penuh kesabaran dan bijaksana.
3. Teman-Teman, Irfan, Fiqih, Alham, Rahman, Bagas yang selalu memberikan dukungan, semangat, dan motivasi selama perjalanan kami menyelesaikan skripsi ini.

Jakarta, 18 Februari 2024

Wahyu Setiya Ramadan

DAFTAR ISI

Bab I	10
Pendahuluan	10
1.1 Latar Belakang.....	10
1.2 Rumusan Masalah	12
1.3 Tujuan Penelitian.....	12
1.4 Batasan Masalah	12
1.5 Kontribusi	13
BAB II	14
TINJAUAN PUSTAKA	14
2.1 Studi Literatur	14
2.2 Teori Analisis Sentimen	18
2.3 Machine Learning	19
2.4 Media Sosial dan Pengaruhnya	19
2.5 K-Means Clustering	20
2.6 Testing Data & Training Data	22
2.6.1 Testing Data	22
2.6.2 Training Data	22
BAB III	24
METODE PENELITIAN	24
3.1 Lokasi Penelitian	24
3.2 Waktu Penelitian.....	25
3.3 Design Sistem	26
3.4 Alat dan Bahan	29
3.5 Dataset.....	30
3.6 Preprocessing.....	31
BAB IV	32
Hasil dan Pembahasan	32
4.1 Hasil Processing.....	32
4.2 Hasil Pembuatan Model	45
BAB V	53

PENUTUP..... 53
DAFTAR PUSTAKA 55

DAFTAR TABEL

Tabel 2. 1 Studi Literatur	15
Tabel 2. 2 Data Point	21
Table 3. 1 Waktu Penelitian.....	24

DAFTAR GAMBAR

Gambar 3. 1 Flowchart Blok Diagram Desain Sistem Model	26
Gambar 3. 2 Insult per-tahun	30
Gambar 4. 1 Tweet per Cluster.....	32
Gambar 4. 2 Analisis Sentimen Cluster 0 Twitter Sentimen.....	33
Gambar 4. 3 Analisis Sentimen Cluster 1 Twitter Sentimen.....	35
Gambar 4. 4 Analisis Sentimen Cluster 2 Twitter Sentimen.....	36
Gambar 4. 5 Analisis Sentimen Cluster 3 Twitter Sentimen.....	37
Gambar 4. 6 Analisis Sentimen Cluster 4 Twitter Sentimen.....	38
Gambar 4. 7 Rata-rata Panjang Tweet.....	39
Gambar 4. 8 Distribusi Tweet	40
Gambar 4. 9 Tweet per Tahun	41
Gambar 4. 10 Tweet Sentimen	42
Gambar 4. 11 Hasil F1 Score	45
Gambar 4. 12 Kurva ROC.....	46

Abstrak

Penelitian ini bertujuan untuk menganalisis sentimen kepercayaan publik terhadap Donald Trump melalui aplikasi X dengan menggunakan metode K-Means Clustering. Dengan semakin populernya media sosial sebagai platform komunikasi politik, penting untuk memahami bagaimana sentimen publik terbentuk dan diekspresikan. Data yang digunakan dalam penelitian ini lebih dari 5000 tweet yang diambil dari Twitter, yang mencerminkan opini masyarakat terhadap Donald Trump.

Proses analisis dimulai dengan tahap preprocessing teks, termasuk normalisasi, tokenisasi, dan penghapusan stopwords. Setelah itu, algoritma K-Means diterapkan untuk mengelompokkan data berdasarkan sentimen, yang kemudian diuji dengan menggunakan Silhouette Coefficient untuk mengukur akurasi pengelompokan. Hasil analisis menunjukkan bahwa terdapat variasi signifikan dalam sentimen publik, dengan cluster yang berbeda mencerminkan pandangan positif, negative dan netral terhadap Donald Trump. Penelitian ini diharapkan dapat memberikan wawasan yang lebih mendalam tentang persepsi masyarakat terhadap Donald Trump serta menunjukkan efektivitas metode K-Means dalam analisis sentimen. Temuan ini juga menekankan pentingnya media sosial sebagai sumber informasi yang dapat mempengaruhi opini publik dan kebijakan.

Kata Kunci : K-Means Clustering, Social Media, Preprocessing, Silhouette Coefficient.

Abstract

This study aims to analyze the sentiment of public trust in Donald Trump through the X application using the K-Means Clustering method. With the growing popularity of social media as a political communication platform, it is important to understand how public sentiment is formed and expressed. The data used in the study were more than 5000 tweets taken from Twitter, which reflects public opinion of Donald Trump.

The analysis process begins with the text preprocessing stage, including normalization, tokenization, and removal of stopwords. After that, the K-Means algorithm was applied to group the data based on sentiment, which was then tested using Silhouette Coefficient to measure the accuracy of the grouping. The results of the analysis show that there is a significant variation in public sentiment, with different clusters reflecting positive, negative and neutral views on Donald Trump. This research is expected to provide deeper insights into public perception of Donald Trump as well as show the effectiveness of the K-Means method in sentiment analysis. The findings also emphasize the importance of social media as a source of information that can influence public opinion and policy.

Keywords : K-Means Clustering, Social Media, Preprocessing, Silhouette Coefficient.

Bab I

Pendahuluan

1.1 Latar Belakang

Penggunaan media sosial telah menjadi semakin populer dalam beberapa tahun terakhir, dan telah menjadi alat yang penting bagi politisi untuk terhubung dengan konstituen mereka dan untuk menyebarkan informasi. Namun, penggunaan media sosial oleh politisi juga telah memicu perdebatan tentang kebenaran dan keakuratan informasi yang disebar. (Tan et al., 2024) Metode Clustering merupakan metode analisa data yang dapat digunakan dalam memecahkan masalah dalam suatu pengelompokan data. Salah satu metode yang ada di dalam metode clustering adalah metode K-mens (Goni et al., 2021) Metode K-means merupakan suatu metode yang dapat melakukan pengelompokan data dalam jumlah yang cukup besar dengan perhitungan waktu yang relatif cepat dan efisien. Dalam artikel ini diuji penggunaan algoritma K-Means dalam mengelompokkan sentimen masyarakat terhadap eksistensi kepercayaan pada media sosial twitter. (Gustientiedina et al., 2019)

Penelitian ini menganalisis sentimen terhadap eksistensi kepercayaan pada media sosial twitter. Kinerja dari sebuah algoritma klasifikasi dipengaruhi dari jenis data dan fiturnya, maka dari itu data set berupa teks yang akan diolah harus melalui tahapan Text preprocessing seperti case folding, stemming, tokenizing, Text Normalization serta stopwords, lalu setelah itu data akan masuk tahapan selanjutnya yaitu tahapan klasifikasi menggunakan algoritma K-Means dan diuji dengan perhitungan Silhouette Coefficient untuk mendapatkan nilai akurasi yang sesuai dengan harapan sehingga dapat mengklasifikasikan data untuk mendapatkan hasil kesimpulan (Wirayasa & Santoso, 2022). Lebih dari 5000 dataset Tweet telah digunakan untuk memposting dan berbagi informasi tentang pengguna, dan juga

konten Tweet dapat mengekspresikan perasaan. Twitter merupakan situs web yang mempunyai layanan menyediakan kumpulan data opini dari orang-orang di seluruh dunia. Hasil dari penyaluran opini dan komentar, Tweet merupakan sumber informasi yang dapat digunakan untuk menganalisis opini publik terhadap institusi dan individu. Opini pada tweet ini dapat digunakan untuk melihat bagaimana sentimen berjalan (Irsyad & Pribadi, 2020).

Proses algoritma K-means dimulai dengan memilih jumlah kluster yang diinginkan. Setiap cluster direpresentasikan oleh sebuah centroid, yaitu titik pusat dari kluster tersebut. Selanjutnya, langkah-langkah K-means dilakukan secara berulang hingga konvergen, yaitu ketika tidak ada lagi perubahan dalam posisi centroid (Haviluddin et al., 2021). Menyatakan bahwa media sosial telah menjadi sumber informasi yang utama bagi masyarakat, dan telah mempengaruhi cara masyarakat memandang dan merespons kebijakan dan tindakan politik. (Şeker kaya, 2020) Oleh karena itu, analisis sentimen kepercayaan publik terhadap Donald Trump pada aplikasi X dengan metode K-Means Clustering dapat memberikan wawasan yang berharga tentang bagaimana pengguna aplikasi tersebut merespons terhadap berita dan informasi tentang Donald Trump. Dalam konteks aplikasi X, analisis sentimen dapat digunakan untuk memahami bagaimana pengguna aplikasi tersebut merespons terhadap berita dan informasi tentang Donald Trump. (Alexandridis et al., 2021) Selain itu, analisis sentimen kepercayaan publik terhadap Donald Trump pada aplikasi X dengan metode K-Means Clustering juga dapat membantu memahami bagaimana kecerdasan buatan dapat digunakan untuk meningkatkan transparansi dan akuntabilitas dalam pemerintahan. Menurut penelitian yang dilakukan oleh Singh et al. pada tahun 2019, penggunaan kecerdasan buatan dalam analisis sentimen dapat membantu meningkatkan transparansi dan akuntabilitas dalam pemerintahan (del Valle & de la Fuente, 2023) Selain itu, analisis kepercayaan publik terhadap Donald Trump pada aplikasi X dengan metode K-Means Clustering juga dapat membantu memahami bagaimana

kecerdasan buatan dapat digunakan untuk meningkatkan transparansi dan akuntabilitas dalam pemerintahan, penggunaan kecerdasan buatan dalam analisis sentimen dapat membantu meningkatkan transparansi dan akuntabilitas dalam pemerintahan (Metcalf et al., 2023). Dengan penelitian di atas, penulis akan membuat analisis kepercayaan publik dengan menggunakan algoritma K-means dengan tujuan agar masyarakat bisa melihat sentimen tersebut secara detail dengan akurasi yang baik.

1.2 Rumusan Masalah

- 1 Bagaimana sentimen publik terhadap Donald Trump pada aplikasi X dapat dianalisis menggunakan metode K-Means Clustering?
- 2 Apa saja karakteristik dari cluster sentimen yang terbentuk dalam analisis ini?
- 3 Seberapa akurat algoritma K-Means dalam mengelompokkan sentimen publik mengenai kepercayaan terhadap Donald Trump di media sosial?

1.3 Tujuan Penelitian

- 1 Menganalisis sentimen publik terhadap Donald Trump pada aplikasi X menggunakan metode K-Means Clustering.
- 2 Mengidentifikasi karakteristik dari cluster sentimen yang terbentuk dalam analisis ini, sehingga dapat memberikan wawasan mengenai persepsi publik.
- 3 Mengukur akurasi algoritma K-Means dalam mengelompokkan sentimen publik mengenai kepercayaan terhadap Donald Trump di media sosial.

1.4 Batasan Masalah

- 1 Penelitian ini hanya berfokus pada analisis sentimen publik terhadap Donald Trump di aplikasi X, tanpa membahas aspek politik atau sosial lainnya yang mungkin mempengaruhi sentimen tersebut.
- 2 Data yang digunakan dalam penelitian ini terbatas pada tweet yang diambil dari Twitter dalam periode tertentu. Tweet yang tidak relevan atau tidak sesuai dengan kata kunci yang ditentukan tidak akan dianalisis.
- 3 Penelitian ini hanya menggunakan metode K-Means Clustering untuk analisis sentimen.

1.5 Kontribusi

1. Penelitian ini memberikan pemahaman yang lebih baik tentang sentimen publik terhadap Donald Trump di aplikasi X, yang dapat membantu dalam memahami pandangan dan persepsi masyarakat.
2. Penelitian ini menunjukkan penerapan metode K-Means Clustering dalam analisis sentimen, yang dapat menjadi referensi bagi penelitian selanjutnya dalam bidang analisis data teks.
3. Dengan menganalisis data dari Twitter, penelitian ini menyoroti pentingnya media sosial sebagai sumber informasi dan opini publik, serta dampaknya terhadap politik dan kebijakan.

BAB II

TINJAUAN PUSTAKA

2.1 Studi Literatur

Penelitian ini meneliti pengaruh media sosial terhadap kepercayaan publik terhadap pemerintah. Hasil penelitian menunjukkan bahwa media sosial memiliki pengaruh yang signifikan terhadap kepercayaan publik terhadap pemerintah. (Mihartinah et al., 2023) telah melakukan penelitian tentang analisis sentimen opini publik terhadap kebijakan pemerintah menggunakan metode K-Means Clustering dan menemukan bahwa metode ini dapat digunakan untuk menganalisis sentimen opini publik dengan akurasi yang tinggi. Penelitian sebelumnya juga menunjukkan bahwa media sosial dan kepemimpinan memiliki pengaruh yang signifikan terhadap kepercayaan publik terhadap pemerintah. telah melakukan penelitian tentang pengaruh media sosial terhadap kepercayaan publik terhadap pemerintah dan menemukan bahwa media sosial memiliki pengaruh yang signifikan terhadap kepercayaan publik terhadap pemerintah. (Chaudhry et al., 2021) Penelitain kedua dengan metode yang sama yaitu Implementasi dari pendekatan metode K-Means untuk mengetahui kecenderungan opini masyarakat terhadap pemilu dalam media sosial twitter. Pada penelitian ini dilakukan untuk mengetahui kecenderungan opini masyarakat terhadap pemilu apakah termasuk kedalam sentimen positif atau negative (Kurniawan & Susanto, 2019).

Tabel 2. 1 Studi Literatur

No.	Nama penulis, tahun	Judul	Dataset	Metode	Hasil
1.	Ross, Robert M (2023)	Expressive Responding in Support of Donald Trump: An Extended Replication of Schaffner and Luks	Data survei pasca-inaugurasi Trump 2017	Analisis sentimen, klasifikasi	Responden pro-Trump cenderung lebih memilih jawaban yang mendukung kelompoknya meski bukti sebaliknya.
2.	Jacobson, Gary C. (2023)	Comparing the impact of Joe Biden and Donald Trump on popular attitudes toward their parties	Survei kepercayaan pemilih Trump pasca-pemilu 2020	Analisis survei, klasifikasi	Mayoritas pemilih Trump masih mempercayai klaim pemilu curang meski telah dibantah.
3.	Irma Surya Kumala (2023)	Analisis Sentimen Penggunaan Aplikasi Shopee Menggunakan	Data scraping dari ulasan Shopee	Support Vector Machine (SVM)	Algoritma SVM mengklasifikasikan sentimen positif dan negatif dengan akurasi 98%.

		Algoritma Support Vector Machine (SVM)			
4.	Insan , Moh Khoirul Hayati (2023)	Sentiment Analysis on Application BRIMO Reviews	Ulasan Google Play	Naive Bayes	Naive Bayes memberikan akurasi 84.52% dalam klasifikasi ulasan aplikasi BRIMO
5.	Syahril Dwi Prasetyo (2023)	Analisis Sentimen Relokasi Ibukota Nusantara	Data Twitter	Naive Bayes dan KNN	Naive Bayes: 82.27% akurasi; KNN: 88.12%, unggul dalam klasifikasi sentimen.
6.	Chen (2023)	K-Means Clustering-Based Kernel Canonical Correlation Analysis	Dataset SAVEE dan eINTERFACE'05	K-Means Clustering	K-means meningkatkan pengenalan emosi multimodal dengan akurasi yang lebih tinggi.
7.	Zubair, M.(2023)	A Study on Sentiment Analysis Techniques	Data Twitter	Analisis Sentimen, K-means	Membahas algoritma sentimen dan clustering pada data Twitter untuk

		of Twitter Data			mengidentifikasi opini publik.
8.	Zheng, W (2023)	Developing Multi-Labelled Corpus of Twitter Short Texts	Dataset Sentiment140 Twitter	Semi-automatic Multi-Label Annotation	Metode semi-otomatis meningkatkan akurasi anotasi multi-label pada teks singkat Twitter.
9.	Camacho-Collados, J. (2022)	TimeLMs: Diachronic Language Models from Twitter	Data Twitter historis	Model Bahasa, K-means	Model TimeLMs meningkatkan kapasitas model bahasa untuk menangani data Twitter yang terus berkembang.
10.	Törnberg, P. (2023)	ChatGPT-4 Outperforms Experts and Crowd Workers in Annotating Political Twitter Messages	Twitter politisi AS (Pemilu 2020)	Klasifikasi teks, analisis sentimen	lebih akurat dibandingkan anotator manusia dalam menganotasi pesan politik di Twitter.

Tabel tersebut bertujuan untuk memberikan ringkasan komprehensif mengenai penelitian-penelitian yang telah dilakukan terkait analisis sentimen dan klasifikasi teks, khususnya dalam konteks kepercayaan publik terhadap Donald Trump. Dengan mencantumkan informasi mengenai judul, dataset, metode, dan hasil dari setiap penelitian, tabel ini memungkinkan pembaca untuk memahami berbagai pendekatan yang telah diterapkan, termasuk penggunaan algoritma K-Means Clustering dalam analisis data media sosial. Selain itu, tabel ini menggambarkan hasil-hasil yang diperoleh dari penelitian sebelumnya, yang dapat memberikan wawasan penting bagi penelitian selanjutnya. Dengan demikian, tabel ini tidak hanya berfungsi sebagai referensi, tetapi juga sebagai sumber inspirasi untuk penelitian yang ingin mengeksplorasi lebih jauh analisis sentimen pada aplikasi dan platform media sosial.

2.2 Teori Analisis Sentimen

Analisis sentimen, atau penambangan opini, adalah proses yang bertujuan untuk mengekstrak dan memahami opini atau pendapat dari dokumen teks mengenai topik tertentu. Dalam konteks penelitian ini, analisis sentimen digunakan untuk mengevaluasi opini masyarakat terhadap pemilihan presiden (Pilpres) 2020 di Amerika, khususnya melalui platform media sosial seperti Twitter.

Tujuan analisis sentimen dalam penelitian ini adalah untuk menentukan kecenderungan opini, apakah cenderung positif, negatif, atau netral. Ini penting dalam konteks politik, di mana opini publik dapat mempengaruhi hasil pemilihan. Selanjutnya, data mentah perlu dibersihkan melalui tahapan pre-processing, yang meliputi normalisasi untuk menghapus elemen yang tidak diperlukan, case folding untuk konsistensi huruf, tokenizing untuk memecah teks menjadi kata-kata, stopword removal untuk menghapus kata-kata umum yang tidak memberikan

makna penting, dan stemming untuk mengubah kata menjadi bentuk dasarnya.

Setelah proses pre-processing, dilakukan klustering dengan menggunakan algoritma K-Means untuk mengelompokkan data menjadi dua kategori: positif dan negatif. Data beropini positif mengandung kata-kata seperti "baik" dan "jujur", sedangkan yang negatif mengandung kata-kata seperti "bohong" dan "fitnah".(Kurniawan & Susanto, 2020).

2.3 Machine Learning

Machine Learning (ML) digunakan untuk mengajarkan mesin cara menangani data lebih banyak efisien terkadang setelah melihat data. Dengan banyaknya kumpulan data tersedia, permintaan untuk pembelajaran mesin meningkat banyak industri menerapkan pembelajaran mesin untuk mengekstrak data yang relevan. Tujuan pembelajaran mesin adalah untuk belajar dari data penelitian telah dilakukan tentang cara membuat mesin belajar sendiri tanpa diprogram secara eksplisit. Banyak matematikawan dan pemrogram menerapkan beberapa pendekatan untuk menemukan solusi dari masalah ini yaitu memiliki kumpulan data yang sangat besar.(Mahesh, 2020)

2.4 Media Sosial dan Pengaruhnya

Media sosial merupakan platform yang memainkan peran penting dalam pembentukan opini publik, terutama selama momen-momen kritis seperti pemilihan umum. Dalam konteks Pemilihan Presiden Amerika 2020, media sosial, khususnya Twitter, menjadi arena bagi pengguna untuk mengekspresikan pendapat dan berbagi informasi mengenai calon presiden, Donald Trump dan Joe Biden. Melalui platform ini, calon dapat menyampaikan argumen, kebijakan, dan

pandangan mereka secara langsung kepada masyarakat, sementara pengguna dapat memberikan respons yang cepat dan luas. Pengaruh media sosial terhadap pemilih sangat signifikan. Informasi yang disebar di platform ini dapat memengaruhi keputusan pemilih, baik melalui penyampaian berita maupun melalui ekspresi emosional yang terkandung dalam tweet. Selain itu, media sosial sering kali menciptakan "echo chambers," di mana individu berinteraksi dengan orang-orang yang memiliki pandangan serupa, yang pada gilirannya memperkuat keyakinan mereka dan menciptakan ruang diskusi yang homogen. (Nugroho et al., 2021)

2.5 K-Means Clustering

Algoritma k-means sebagai metode pengelompokan yang paling dikenal dan digunakan secara luas. Meskipun merupakan pendekatan pembelajaran tanpa pengawasan, algoritma ini biasanya dipengaruhi oleh inisialisasi dan memerlukan jumlah kluster yang ditentukan sebelumnya. Penulis mengembangkan skema pembelajaran tanpa pengawasan untuk algoritma k-means yang dapat menemukan jumlah kluster optimal secara otomatis, tanpa memerlukan inisialisasi atau pemilihan parameter. Algoritma k-means klasik bekerja dengan cara mengelompokkan data berdasarkan kedekatan terhadap pusat kluster, tetapi sering kali hasilnya dipengaruhi oleh bagaimana pusat kluster tersebut diinisialisasi. Untuk mengatasi masalah ini, penulis menggunakan algoritma k-means clustering yang dapat secara otomatis menentukan jumlah kluster optimal. Algoritma ini mengintegrasikan konsep entropi sebagai penalti dalam fungsi objektif, sehingga meningkatkan informasi yang diperoleh dari data. (Sinaga & Yang, 2020)

$$J = \sum_{i=1}^k \sum_{x \in C_i} \|x - \mu_i\|^2 \quad (1)$$

Simbol:

- J : Total jarak kuadrat antara data dan pusat cluster.

- k : Jumlah cluster.
- C_i : Himpunan data dalam cluster ke- i .
- x : Data point.
- μ_i : Pusat (mean) dari cluster ke- i .
- $\|x - \mu_i\|^2$: Jarak kuadrat antara data point x dan pusat cluster
- Misalkan kita memiliki sekumpulan data 2D sebagai berikut:

Data Point	(x, y)
1	(1, 2)
2	(1, 4)
3	(1, 0)
4	(10, 2)
5	(10, 4)
6	(10, 0)

Tabel 2. 2 Data Point

Kita ingin mengelompokkan data ini menjadi 2 cluster ($k=2$)

1. **Inisialisasi:** Pilih 2 pusat cluster secara acak:
 - $\mu_1 = (1,2)$
 - $\mu_2 = (10,2)$
2. **Pengulangan (Iterasi):**
 - Hitung jarak setiap data point ke setiap pusat cluster dan tetapkan ke cluster terdekat.

3. Perbarui Cluster:

- Hitung rata-rata posisi setiap cluster untuk memperbarui pusatnya.

4. **Iterasi Ulang:** Ulangi langkah 2 dan 3 hingga pusat cluster tidak berubah.

Hasil Akhir

Setelah beberapa iterasi, kita mungkin mendapatkan:

- **Cluster 1:** $\mu_1=(1,2)$ dengan anggota $(1,2),(1,4),(1,0)$
- **Cluster 2:** $\mu_2=(10,2)$ dengan anggota $(10,2),(10,4),(10,0)$

2.6 Testing Data & Training Data

2.6.1 Testing Data

Tujuan utama penggunaan data pengujian adalah untuk mengukur seberapa baik algoritma dapat mengelompokkan data baru yang tidak terlihat sebelumnya. Data pengujian harus representatif dari populasi data yang lebih besar agar hasil yang diperoleh dapat digeneralisasikan dengan baik. Setelah model k-means dilatih menggunakan data pelatihan, data pengujian dimasukkan ke dalam model untuk melihat kluster mana yang diidentifikasi oleh algoritma. Hasil pengelompokan kemudian dibandingkan dengan label yang diketahui, jika tersedia, untuk mengevaluasi akurasi pengelompokan.

2.6.2 Training Data

Konteks algoritma pengelompokan k-means adalah kumpulan data yang digunakan untuk melatih model agar dapat mengidentifikasi pola dan struktur dalam data. Proses pelatihan ini melibatkan penggunaan data yang telah diberi label atau tidak, di mana model akan belajar untuk mengelompokkan data berdasarkan kedekatan antar poin data. Selama tahap ini, algoritma k-means melakukan iterasi untuk menentukan pusat kluster (centroid) dan

mengelompokkan data ke dalam kluster berdasarkan jarak ke pusat tersebut. Penting bagi data pelatihan untuk mencakup berbagai variasi dan representasi dari populasi yang lebih besar agar model dapat belajar dengan efektif dan menghasilkan kluster yang akurat.

BAB III

METODE PENELITIAN

3.1 Lokasi Penelitian

Penelitian ini dilakukan terutama di dua lokasi: di rumah dan bekerja sama dengan teman. Di rumah, kami melakukan penelitian secara mandiri, dengan penekanan pada pengumpulan dan analisis data terperinci. Sebaliknya, kolaborasi dengan teman terjadi melalui diskusi dan penyelesaian tugas bersama, yang memungkinkan terjadinya pertukaran ide yang konstruktif dan saling mendukung untuk menyelesaikan penelitian dengan lebih efektif. Kombinasi kedua pendekatan ini berperan penting dalam mencapai hasil yang optimal ketika membahas topik akhir dan memudahkan pelaksanaan penelitian.

Table 3. 1 Waktu Penelitian

Kegiatan	Bulan					
	Sep	Okt	Nov	Des	Jan	Feb
Identifikasi Masalah						
Studi Literatur						
Pengumpulan data set						
Pembangunan Model						

Seminar Proposal						
Pengujian Kinerja						
Sidang Akhir						

Penelitian dilaksanakan dalam kurun waktu enam bulan, dimulai sejak September 2024 hingga Februari 2025. Proses penelitian dimulai dengan identifikasi masalah yang dilakukan pada bulan September, seperti yang ditunjukkan dalam Tabel 3.1. Setelah itu, kegiatan pengumpulan dataset dilakukan secara intensif selama dua bulan, yaitu September dan Oktober. Pada bulan Desember, penulis melakukan seminar proposal untuk mempresentasikan rencana penelitian.

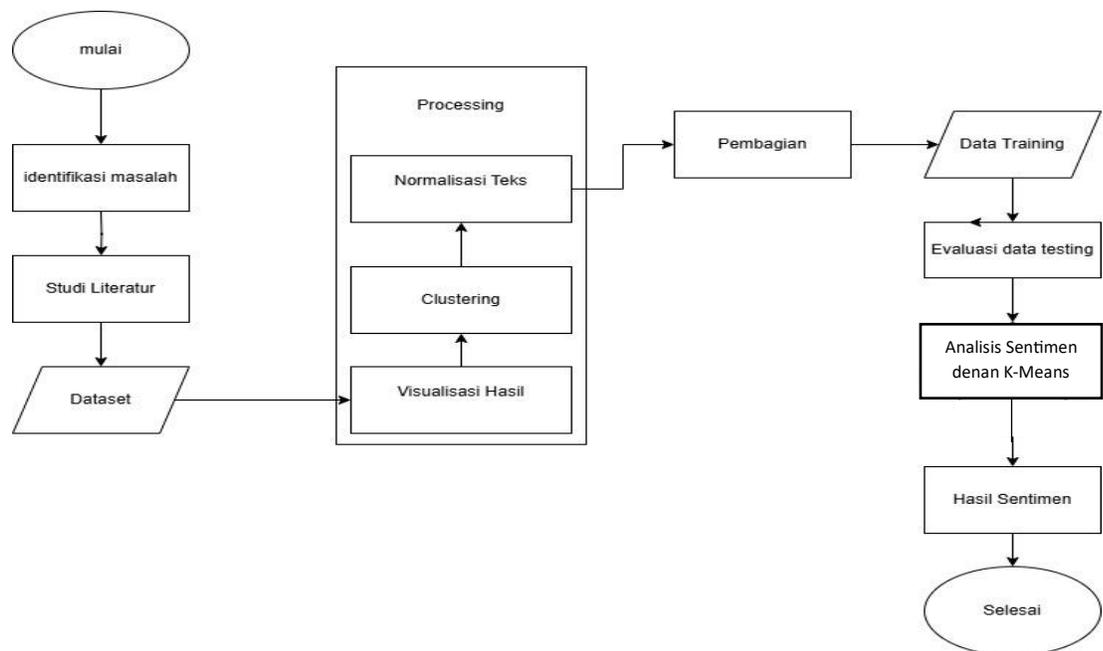
Tahap pembangunan model menjadi fokus utama penelitian selama periode September hingga Oktober. Proses pembangunan model ini berjalan paralel dengan studi literatur yang berlangsung sepanjang periode tersebut, dari September hingga November. Setelah model dianggap cukup baik, dilakukan pengujian kinerja pada bulan Januari. Tahap akhir penelitian meliputi penyusunan laporan skripsi dan proses bimbingan yang berlangsung sejak Desember hingga Februari. Proses bimbingan ini dilakukan secara berkala untuk memastikan kualitas laporan skripsi. Puncak dari seluruh rangkaian penelitian adalah sidang akhir yang dijadwalkan pada bulan Februari.

3.2 Waktu Penelitian

Penelitian dilakukan selama periode enam bulan mulai September 2024 hingga Februari 2025. Proses penelitian diawali dengan identifikasi masalah yang dilakukan pada bulan September, seperti terlihat pada Tabel 3.1. Kegiatan

pengumpulan data intensif kemudian dilakukan selama dua bulan pada bulan September dan Oktober. Pada bulan Desember, penulis mengadakan seminar proposal untuk mempresentasikan rencana penelitian. Tahap pembuatan model menjadi fokus penelitian dari bulan September hingga Oktober. Proses pemodelan berlangsung bersamaan dengan tinjauan literatur yang dilakukan pada bulan September hingga November. Setelah model ini terbukti cukup baik, pengujian kinerja dilakukan pada bulan Januari. Tahap akhir investigasi, yang mencakup persiapan laporan akhir dan proses konsultasi, berlangsung dari bulan Desember hingga Februari. Proses konsultasi ini dilakukan secara berkala untuk menjamin kualitas laporan akhir. Puncak dari keseluruhan rangkaian studi ini yang dijadwalkan pada bulan Februari.

3.3 Design Sistem



Gambar 3. 1 Flowchart Blok Diagram Desain Sistem Model

➤ **Masukan (Mulai):**

- Proses dimulai dengan data masukan yang berasal dari aplikasi X. Data ini berkaitan dengan kepercayaan publik terhadap Donald Trump.

➤ **Identifikasi Masalah:**

- Masalah yang akan dianalisis adalah mengenai sentimen publik terhadap Donald Trump.

➤ **Studi Literatur:**

- Dilakukan kajian literatur terkait dengan analisis sentimen dan metode clustering yang relevan.

➤ **Dataset:**

- Data yang akan digunakan dalam analisis dikumpulkan dari aplikasi X.

➤ **Processing:**

- Tahap pemrosesan data meliputi:
 - **Visualisasi Hasil:** Hasil clustering divisualisasikan untuk memudahkan pemahaman.
 - **Clustering:** Algoritma clustering, dalam hal ini K-Means, diterapkan untuk mengelompokkan data berdasarkan fitur sentimen.
 - **Normalisasi Teks:** Teks data dinormalisasi, misalnya mengubah semua huruf menjadi kecil.

➤ **Pembagian Data:**

- Data yang telah diproses dibagi menjadi data pelatihan dan data pengujian.

➤ **Data Pelatihan:**

- Data pelatihan digunakan untuk melatih model analisis sentimen.

➤ **Evaluasi Data Pengujian:**

- Data pengujian digunakan untuk mengevaluasi kinerja model analisis sentimen.

➤ **Analisis Sentimen dengan K-Means:**

- Hasil analisis sentimen dari data dikelompokkan menggunakan algoritma K-Means.

➤ **Hasil Sentimen:**

- Hasil akhir dari analisis sentimen dengan K-Means diperoleh.

➤ **Selesai:**

- Proses analisis sentimen kepercayaan publik terhadap Donald Trump selesai.

Secara keseluruhan, flowchart ini menggambarkan alur proses untuk menganalisis sentimen publik terhadap Donald Trump dengan memanfaatkan data dari aplikasi X. Metode K-Means Clustering digunakan untuk mengelompokkan data berdasarkan tingkat sentimen positif, negatif, atau netral.

3.4 Alat dan Bahan

Dalam melakukan analisis sentimen dan clustering pada data dari aplikasi X, diperlukan alat dan bahan sebagai berikut:

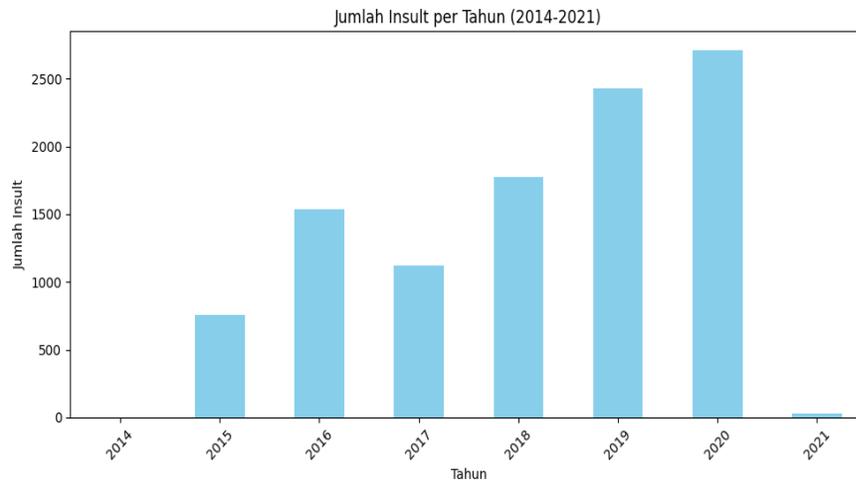
1. Perangkat Keras (Hardware):

- Komputer/Laptop dengan spesifikasi yang memadai untuk menjalankan proses analisis data dan komputasi.
- Koneksi internet yang stabil untuk mengakses data dari aplikasi X.

2. Perangkat Lunak (Software):

- Bahasa pemrograman Python: Python menjadi pilihan yang populer dalam analisis data, machine learning, dan natural language processing.
- Librari/Paket Python:
 - **Pandas:** Untuk manipulasi dan analisis data.
 - **NLTK (Natural Language Toolkit):** Untuk pemrosesan bahasa alami, termasuk analisis sentimen.
 - **Matplotlib dan Seaborn:** Untuk visualisasi data.

3.5 Dataset



Gambar 3. 2 Insult per-tahun

Berdasarkan grafik "Jumlah Insult per Tahun (2014-2021)", dapat dilihat bahwa terjadi tren peningkatan yang signifikan dalam jumlah insult yang dilontarkan selama periode 2014 hingga 2020. Pada tahun 2014, jumlah insult tercatat sekitar 0, namun terus meningkat setiap tahunnya. Pada tahun 2015, insult meningkat menjadi sekitar 750, lalu kembali naik di tahun 2016 menjadi sekitar 1.500. Puncak tertinggi terjadi pada tahun 2020, di mana jumlah insult mencapai sekitar 2.600.

Meskipun terjadi penurunan di tahun 2021 menjadi sekitar 200 insult, angka tersebut masih jauh lebih tinggi dibandingkan tahun-tahun sebelum 2020. Kenaikan insult yang terus-menerus dari 2014 hingga 2020 menunjukkan adanya eskalasi yang signifikan dalam intensitas penggunaan bahasa yang bersifat menghina dan merendahkan orang lain selama rentang waktu tersebut.

3.6 Preprocessing

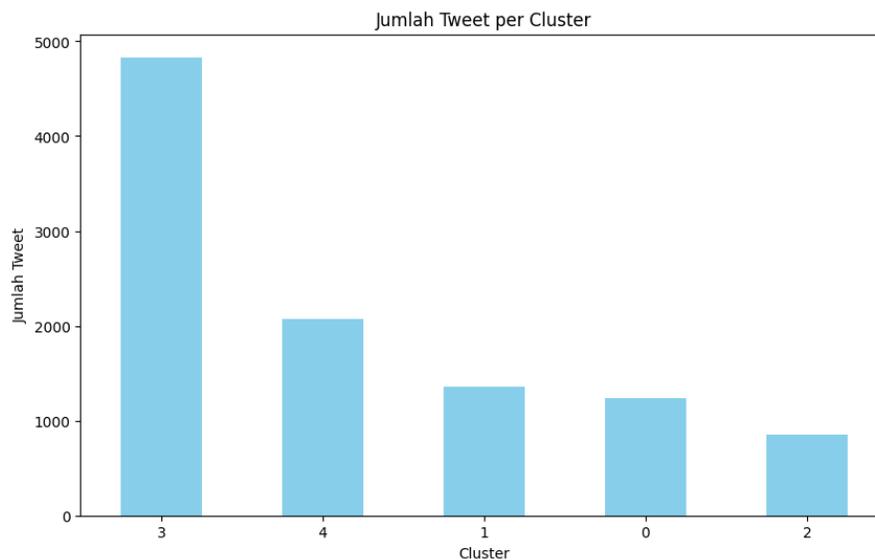
Proses preprocessing untuk dataset “Dataset_new_real donaldtrump_insult_tweets_2014_to_2021.csv” dimulai dengan pembacaan data menggunakan library pandas, di mana file CSV dimuat ke dalam DataFrame. Setelah itu, struktur data diperiksa dengan menampilkan beberapa baris awal dan akhir, serta menentukan tipe data masing-masing kolom. Selanjutnya, langkah penting adalah menangani missing values; dalam hal ini, setelah pemeriksaan, tidak ditemukan missing values yang memerlukan penanganan. Kemudian, kolom 'date' diubah menjadi tipe datetime untuk memungkinkan analisis berbasis waktu, dan tipe data kolom lain seperti 'insult' dan 'tweet' dikonfirmasi sebagai string. Setelah itu, fitur tambahan ditambahkan dengan membuat kolom 'year' yang mengekstrak tahun dari kolom 'date'. Penggunaan teknik NLP dipertimbangkan untuk menambah nilai analisis, misalnya, dengan menambahkan kolom 'sentiment' yang menunjukkan sentimen dari kolom 'insult' atau 'tweet', menggunakan library seperti TextBlob atau VADER. Pembersihan teks dilakukan dengan menghilangkan tanda baca, mengonversi teks ke huruf kecil, dan menghapus karakter khusus pada kolom 'insult' dan 'tweet', serta menggunakan library seperti NLTK atau spaCy untuk tokenisasi, penghapusan stopwords, dan lemmatisasi. Terakhir, pemeriksaan outlier dilakukan pada kolom 'insult'. Jika ditemukan outlier, keputusan dibuat untuk menghapusnya atau melakukan transformasi data untuk mengurangi dampaknya. Setelah langkah-langkah preprocessing ini selesai, dataset siap untuk analisis lebih lanjut, seperti analisis tren atau klasifikasi sentimen.

BAB IV

Hasil dan Pembahasan

4.1 Hasil Processing

a). Jumlah Tweet Cluster

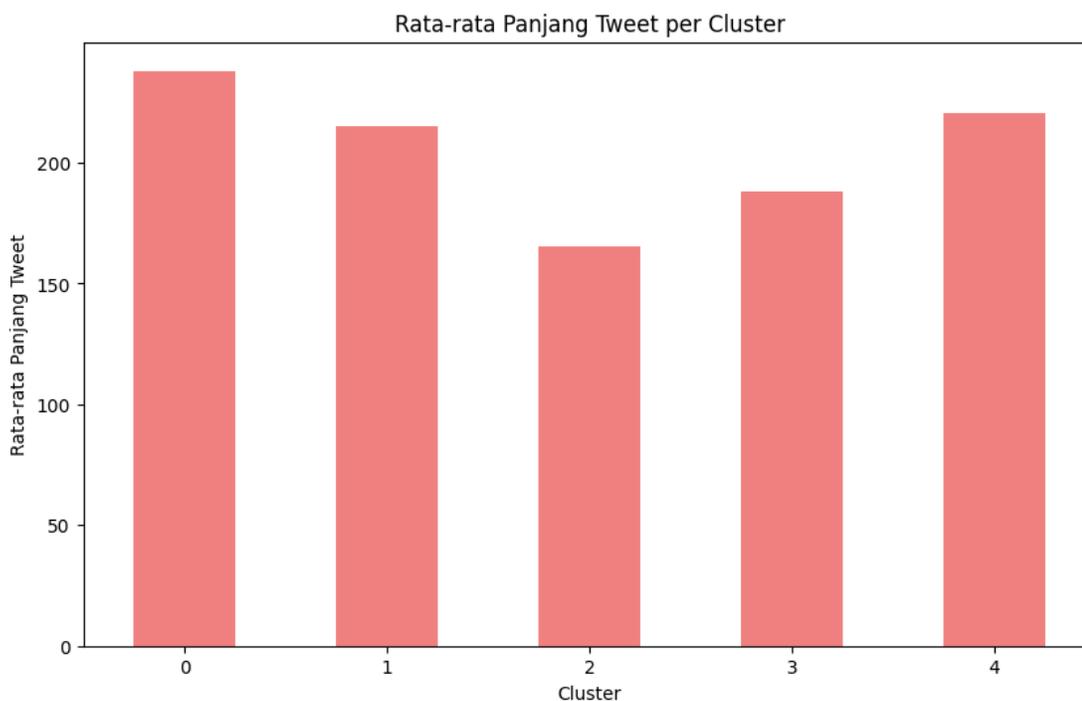


Gambar 4. 1 Tweet per Cluster

Dengan jelas terlihat bahwa cluster 3 memiliki jumlah tweet yang sangat dominan, mencapai sekitar 4.800 tweet. Hal ini mengindikasikan bahwa cluster 3 merupakan kelompok terbesar dan paling aktif dalam dataset yang dianalisis. Jumlah tweet di cluster 3 jauh melebihi cluster-cluster lainnya, menunjukkan adanya topik atau pembahasan yang sangat menarik perhatian pengguna pada cluster tersebut. Sementara itu, cluster 4 juga menunjukkan jumlah tweet yang cukup tinggi, sekitar 2.000 tweet. Walaupun tidak sebanyak cluster 3, angka ini masih cukup signifikan dan menunjukkan cluster 4 juga berperan penting dalam aktivitas percakapan yang terekam dalam dataset. Namun, untuk cluster-cluster lainnya, jumlah tweet terlihat jauh lebih rendah.

terpolarisasi, dengan istilah-istilah seperti "Obstruction," "Fraud," "Rigged," dan "Bad" yang mencerminkan pandangan yang sangat kritis terhadap Demokrasi dan tindakan-tindakannya. Sebaliknya, kata-kata seperti "Country," "People," dan "Collusion" menyiratkan perdebatan dan perselisihan yang sedang berlangsung tentang kebijakan dan perilaku partai tersebut. Kehadiran istilah-istilah seperti "Report," "Crime," "President," dan "FBI" menyoroti konteks politik yang lebih luas, termasuk penyelidikan, tuduhan, dan peran lembaga pemerintah dalam diskusi-diskusi ini. Kata "Nothing" menonjol sebagai fitur yang menonjol, yang mungkin menyiratkan sikap yang meremehkan atau skeptis terhadap klaim atau narasi tertentu. Secara keseluruhan, awan kata ini menyajikan sekilas wacana politik yang penuh dengan kontroversi yang mengelilingi Demokrasi, dengan campuran bahasa yang mendukung, kritis, dan berpotensi konspirasi. Hal ini mencerminkan sifat diskursus politik kontemporer di Amerika Serikat yang penuh kontroversi dan sering kali terpecah-pecah.

g). Rata-rata Tweet per Cluster

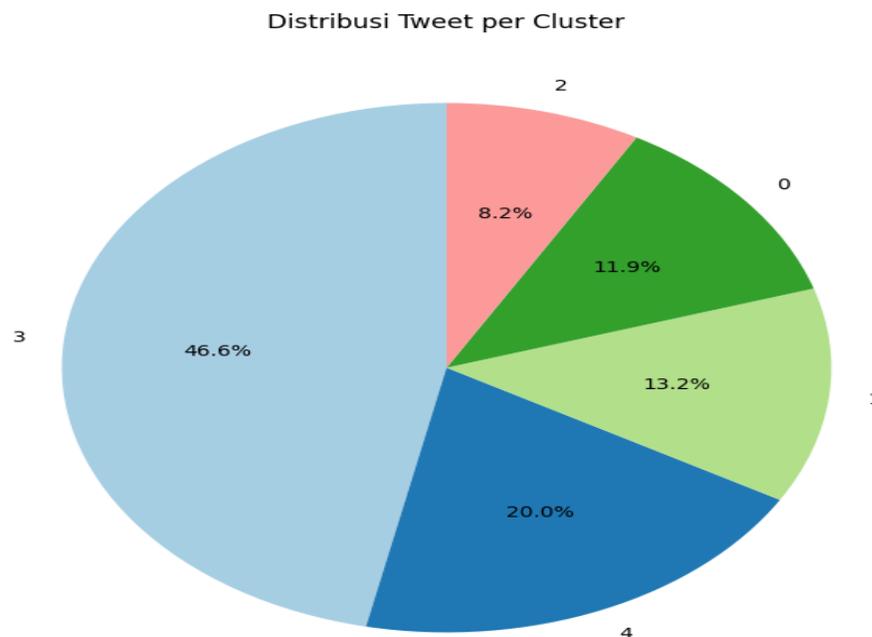


Gambar 4. 7 Rata-rata Panjang Tweet

Gambar tersebut adalah sebuah diagram batang yang menunjukkan rata-rata panjang tweet per cluster. Pada sumbu horizontal, terdapat label untuk masing-masing cluster yang bernomor dari 0 hingga 4, sementara sumbu vertikal menggambarkan rata-rata panjang tweet dalam satuan yang sesuai. Setiap batang berwarna merah muda memberikan representasi visual yang jelas mengenai panjang tweet pada setiap cluster. Dari tampilan diagram, terlihat bahwa cluster 4 memiliki rata-rata panjang tweet tertinggi, diikuti oleh cluster 1 dan 3, yang menunjukkan bahwa konten dalam cluster-cluster tersebut mungkin lebih kompleks atau mendalam dibandingkan dengan cluster lainnya. Sebaliknya, cluster 0 dan 2 memiliki rata-rata panjang tweet yang lebih rendah, menunjukkan kemungkinan bahwa konten di dalamnya lebih singkat atau langsung. Diagram ini memberikan wawasan mengenai variasi panjang tweet

berdasarkan pengelompokan, yang dapat berguna untuk analisis lebih lanjut mengenai karakteristik komunikasi di media sosial.

h). Distribusi Tweet per Cluster

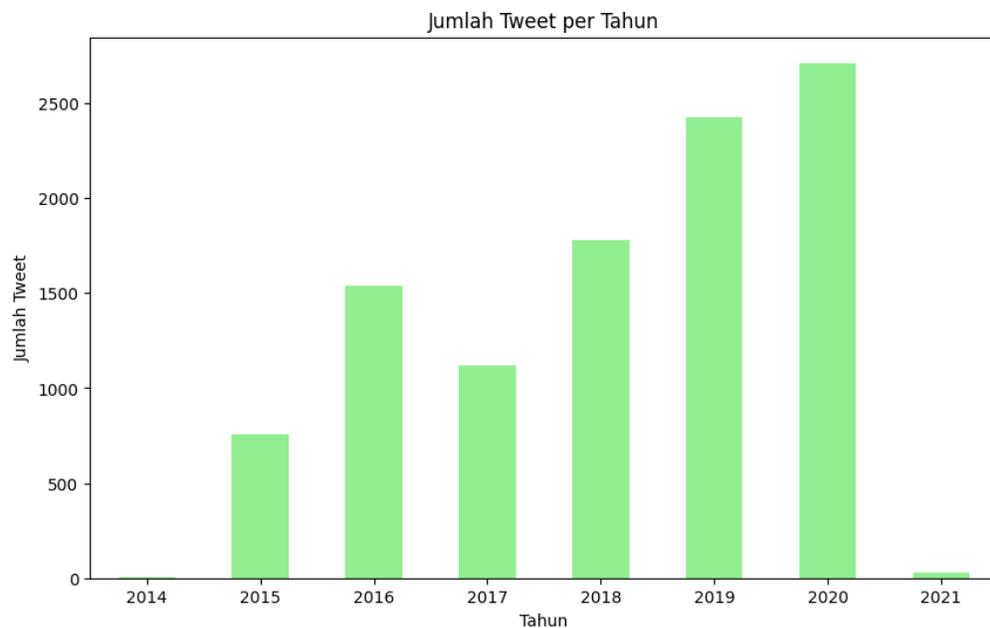


Gambar 4. 8 Distribusi Tweet

Gambar tersebut merupakan sebuah diagram pai yang menggambarkan distribusi tweet berdasarkan cluster. Setiap sektor pada diagram menunjukkan persentase kontribusi masing-masing cluster terhadap total tweet. Cluster 3 menonjol dengan persentase terbesar, yaitu 46,6%, yang menunjukkan bahwa hampir setengah dari total tweet termasuk dalam kategori ini. Diikuti oleh cluster 4 dengan 20%, cluster 1 dengan 13,2%, dan cluster 0 serta cluster 2 masing-masing dengan 11,9% dan 8,2%. Warna-warna yang berbeda untuk setiap sektor tidak hanya memudahkan visualisasi, tetapi juga mengindikasikan

proporsi relatif dari setiap cluster. Diagram ini memberikan gambaran yang jelas tentang bagaimana tweet tersebar di antara berbagai cluster, mengindikasikan bahwa cluster 3 mungkin memiliki tema atau konten yang paling dominan dalam dataset yang dianalisis.

i). Jumlah Tweet per Tahun

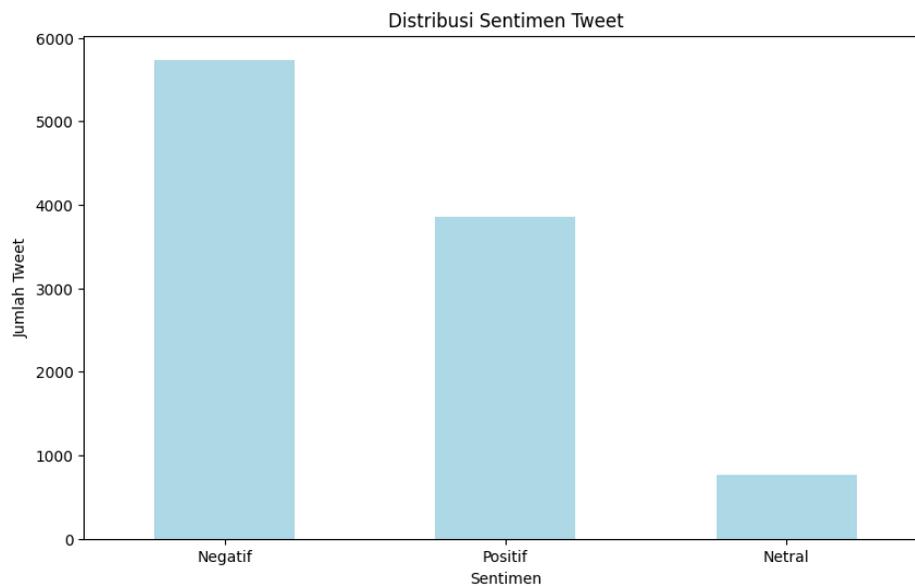


Gambar 4. 9 Tweet per Tahun

Grafik ini menunjukkan tren kenaikan yang cukup signifikan dalam jumlah total tweet per tahun selama periode 2014 hingga 2021. Pada tahun 2014, jumlah tweet tercatat di bawah 500.000. Namun, angka ini terus meningkat secara konsisten setiap tahunnya, hingga mencapai puncaknya pada tahun 2021 dengan lebih dari 2,5 juta tweet. Menariknya, laju pertumbuhan jumlah tweet tidak selalu linear. Terdapat beberapa titik di mana kenaikannya melambat, seperti pada tahun 2016 dan 2017 yang hanya meningkat sedikit. Hal ini kemungkinan terkait dengan dinamika topik-topik diskusi yang sedang

berkembang pada masa tersebut. Namun, grafik menunjukkan lonjakan signifikan pada tahun 2019 dan 2020, dengan peningkatan masing-masing sekitar 1,5 juta dan 2 juta tweet. Fenomena ini dapat disebabkan oleh peristiwa-peristiwa penting, seperti pemilihan umum atau isu-isu aktual lainnya yang menarik perhatian publik secara lebih luas pada waktu itu. Secara keseluruhan, tren kenaikan jumlah tweet per tahun ini mencerminkan semakin tingginya aktivitas dan keterlibatan publik dalam platform digital, khususnya terkait dengan topik-topik politik dan isu sosial yang sedang berkembang. Analisis lebih mendalam terhadap konteks dan faktor-faktor yang memengaruhi dinamika ini dapat memberikan wawasan yang lebih komprehensif.

j). Distribusi Sentimen Variabel



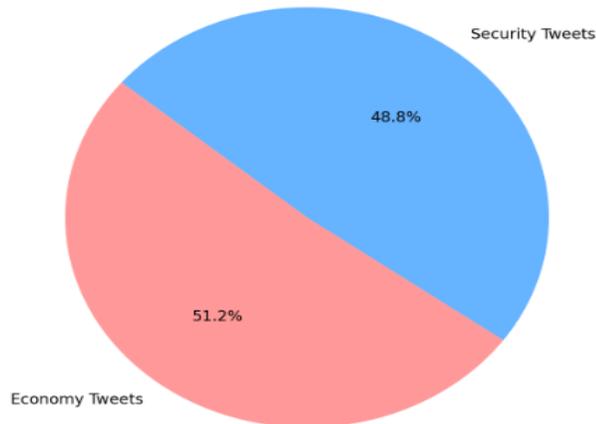
Gambar 4. 10 Tweet Sentimen

Dari visualisasi, sangat jelas terlihat bahwa sentimen negatif mendominasi dengan jumlah tweet sekitar 5.400. Hal ini menunjukkan bahwa sebagian besar tweet dalam dataset yang dianalisis memiliki kecenderungan

sentimen yang negatif. Kemungkinan besar, tweet-tweet ini memuat ungkapan-ungkapan, emosi, atau opini yang cenderung kritis, pesimistis, atau bahkan menyalahkan. Di sisi lain, sentimen positif tercatat sebanyak sekitar 3.600 tweet. Meskipun jumlahnya lebih rendah dibandingkan sentimen negatif, angka ini masih cukup signifikan. Tweet-tweet dengan sentimen positif ini mungkin mengandung ekspresi dukungan, apresiasi, atau optimisme terkait topik-topik tertentu. Yang menarik, sentimen netral hanya terdapat sekitar 900 tweet. Jumlah ini jauh lebih sedikit dibandingkan dua kategori sentimen lainnya. Hal ini mengindikasikan bahwa tweet-tweet dalam dataset ini cenderung berada pada spektrum sentimen yang lebih ekstrem, baik positif maupun negatif, dan kurang bersifat netral atau objektif. Secara keseluruhan, distribusi sentimen tweet ini mencerminkan adanya polarisasi opini yang kuat dalam dataset yang dianalisis. Lebih jauh lagi, analisis terhadap konteks dan topik-topik spesifik yang mendasari pola sentimen ini dapat memberikan pemahaman yang lebih komprehensif mengenai dinamika percakapan publik yang terekam.

k). Distribusi Tweet Ekonomi dan Keamanan

Distribution of Relevant Tweets



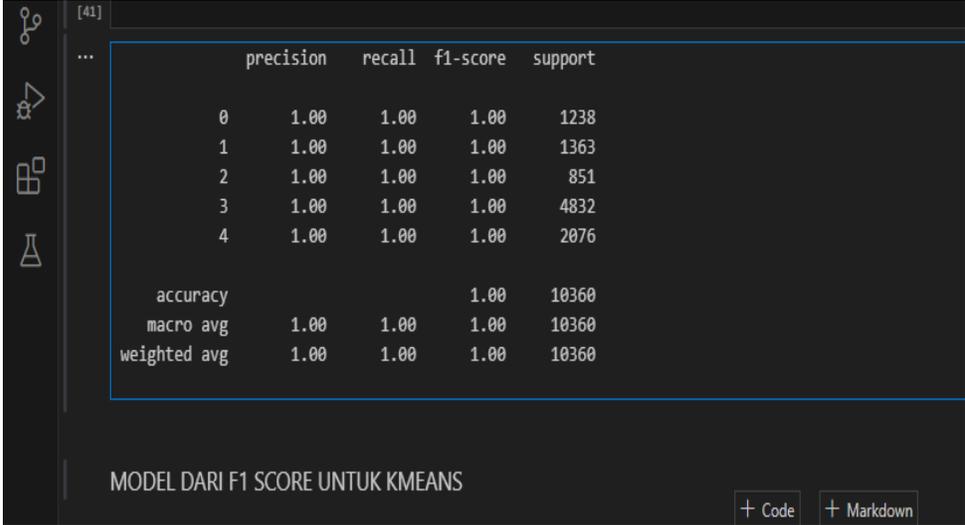
Gambar 4. 11 Twet Ekonomi dan Keamanan

Diagram lingkaran pada gambar tersebut menunjukkan distribusi tweet yang relevan terkait dua kategori, yaitu tweet tentang ekonomi dan tweet tentang keamanan. Dari data yang disajikan, terlihat bahwa 51,2% dari total tweet berasal dari kategori ekonomi, sementara 48,8% berasal dari kategori keamanan. Meskipun keduanya hampir seimbang, tweet tentang ekonomi sedikit lebih mendominasi. Diagram ini memberikan gambaran yang jelas tentang fokus percakapan di media sosial berkaitan dengan dua isu penting ini. Penting untuk dicatat bahwa meskipun rasio antara kedua kategori ini cukup dekat, dominasi tweet ekonomi menunjukkan bahwa isu-isu terkait dengan perekonomian mungkin lebih menjadi perhatian publik pada saat ini. Hal ini dapat mencerminkan tren atau peristiwa terkini yang mempengaruhi masyarakat, seperti perubahan kebijakan ekonomi, fluktuasi pasar, atau situasi finansial yang kritis. Dengan demikian, diagram ini tidak hanya menyajikan angka, tetapi juga mengindikasikan dinamika sosial dan perhatian masyarakat terhadap isu-isu yang sedang hangat diperbincangkan. Melalui analisis

semacam ini, kita dapat memahami lebih dalam tentang apa yang menjadi fokus diskusi publik dan bagaimana kedua isu ini saling berinteraksi dalam konteks yang lebih luas.

4.2 Hasil Pembuatan Model

a). F1 Score Model K-Means



```
[41]
...
      precision    recall  f1-score   support

0         1.00      1.00      1.00     1238
1         1.00      1.00      1.00     1363
2         1.00      1.00      1.00      851
3         1.00      1.00      1.00     4832
4         1.00      1.00      1.00     2076

accuracy          1.00      10360
macro avg         1.00      1.00      1.00     10360
weighted avg      1.00      1.00      1.00     10360
```

MODEL DARI F1 SCORE UNTUK KMEANS

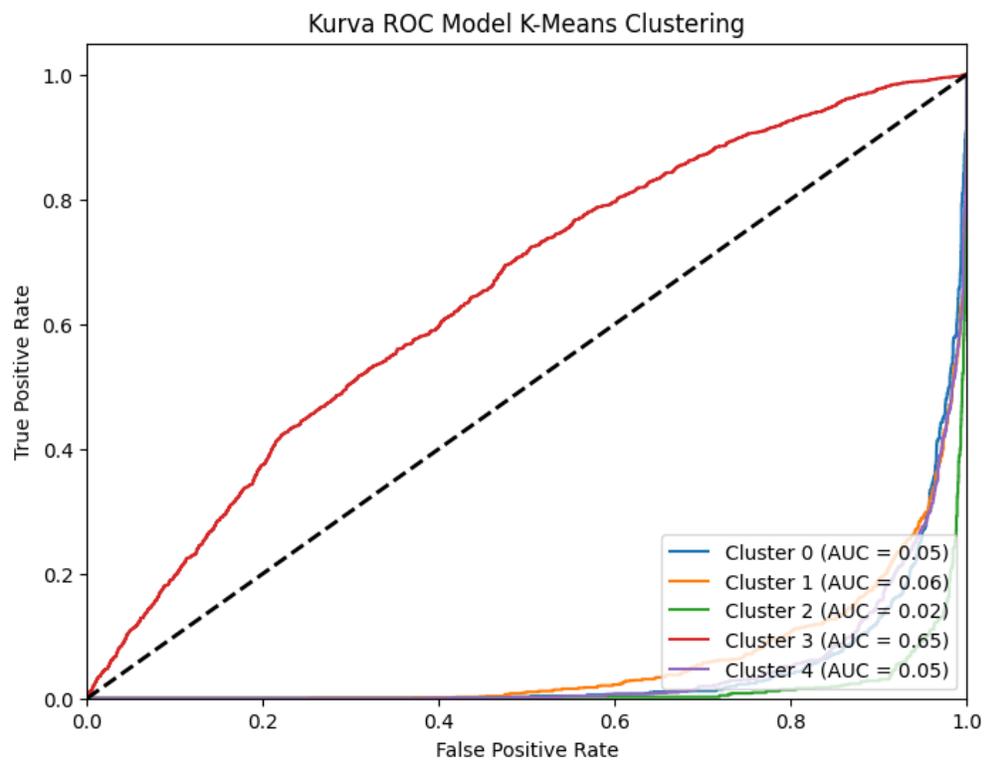
+ Code + Markdown

Gambar 4. 11 Hasil F1 Score

Gambar tersebut menunjukkan output dari evaluasi model menggunakan metrik klasifikasi, khususnya untuk algoritma K-Means, Semua nilai precision, recall, dan f1-score untuk setiap kelas adalah 1.00, yang menunjukkan bahwa model berhasil mengklasifikasikan semua instance dengan benar tanpa kesalahan. Di bagian bawah tabel, terdapat metrik agregat yang menunjukkan bahwa akurasi model secara keseluruhan juga 1.00, dengan nilai macro average dan weighted average untuk

precision, recall, dan f1-score semuanya 1.00. Hal ini menegaskan bahwa model K-Means yang digunakan dalam analisis ini sangat efektif, dengan kinerja sempurna dalam klasifikasi semua kelas yang ada.

b). ROC Model K-Means Clustering



Gambar 4. 12 Kurva ROC

Gambar tersebut menampilkan kurva ROC (Receiver Operating Characteristic) untuk model K-Means Clustering, yang digunakan untuk mengevaluasi kinerja klasifikasi. Pada grafik ini, sumbu X menunjukkan tingkat positif palsu (False Positive Rate), sedangkan sumbu Y menunjukkan tingkat positif benar (True Positive Rate). Terdapat beberapa garis yang mewakili masing-masing cluster, yaitu Cluster 0, Cluster 1, Cluster 2, Cluster

3, dan Cluster 4, dengan warna dan gaya garis yang berbeda. Garis putus-putus di tengah grafik berfungsi sebagai garis acuan yang menunjukkan kinerja model acak. Metrik AUC (Area Under the Curve) tertera untuk masing-masing cluster, di mana nilai AUC untuk Cluster 0 adalah 0.05, untuk Cluster 1 adalah 0.06, dan seterusnya. Nilai AUC yang rendah menunjukkan bahwa model ini kurang efektif dalam membedakan antara kelas positif dan negatif, mengindikasikan tantangan dalam klasifikasi yang dihasilkan oleh algoritma K-Means. Secara keseluruhan, grafik ini memberikan gambaran visual yang jelas tentang bagaimana model K-Means Clustering berperforma dalam hal sensitivitas dan spesifisitas, serta menunjukkan bahwa ada ruang untuk perbaikan dalam akurasi klasifikasi model tersebut.

Perhitungan Manual Model

Accuracy:

$$\text{Accuracy} = \frac{\text{Total TP}}{\text{Total Data}}$$

$$\text{Total TP} = 1238 + 1363 + 851 + 4832 + 2076$$

$$\text{Total TP} = 10360$$

$$\text{Total Data} = 10360$$

$$\text{Accuracy} = \frac{10360}{10360} = 1.00$$

Precision:

$$\text{Precision} = \frac{TP}{TP + FP}$$

Class 0

$$TP_0 = 1238, \quad FP_0 = 0$$

$$\text{Precision}_0 = \frac{1238}{1238 + 0} = 1.00$$

Class 1

$$TP_1 = 1363, \quad FP_1 = 0$$

$$\text{Precision}_1 = \frac{1363}{1363 + 0} = 1.00$$

Class 2

$$TP_2 = 851, \quad FP_2 = 0$$

$$\text{Precision}_2 = \frac{851}{851 + 0} = 1.00$$

Class 3

$$TP_3 = 4832, \quad FP_3 = 0$$

$$\text{Precision}_3 = \frac{4832}{4832 + 0} = 1.00$$

Class 4

$$TP_4 = 2076, \quad FP_4 = 0$$

$$\text{Precision}_4 = \frac{2076}{2076 + 0} = 1.00$$

Recall:

$$\text{Recall} = \frac{TP}{TP + FN}$$

Class 0

$$TP_0 = 1238, \quad FN_0 = 0$$

$$\text{Recall}_0 = \frac{1238}{1238 + 0} = 1.00$$

Class 1

$$TP_1 = 1363, \quad FN_1 = 0$$

$$\text{Recall}_1 = \frac{1363}{1363 + 0} = 1.00$$

Class 2

$$TP_2 = 851, \quad FN_2 = 0$$

$$\text{Recall}_2 = \frac{851}{851 + 0} = 1.00$$

Class 3

$$TP_3 = 4832, \quad FN_3 = 0$$

$$\text{Recall}_3 = \frac{4832}{4832 + 0} = 1.00$$

Class 4

$$TP_4 = 2076, \quad FN_4 = 0$$

$$\text{Recall}_4 = \frac{2076}{2076 + 0} = 1.00$$

F1-Score:

$$F1\text{-Score} = \frac{2 \cdot \text{Precision} \cdot \text{Recall}}{\text{Precision} + \text{Recall}}$$

Class 0

$$F1\text{-Score}_0 = \frac{2 \cdot 1.00 \cdot 1.00}{1.00 + 1.00}$$

$$F1\text{-Score}_0 = \frac{2 \cdot 1.00}{2.00} = 1.00$$

Class 1

$$F1\text{-Score}_1 = \frac{2 \cdot 1.00 \cdot 1.00}{1.00 + 1.00}$$

$$F1\text{-Score}_1 = \frac{2 \cdot 1.00}{2.00} = 1.00$$

Class 2

$$F1\text{-Score}_2 = \frac{2 \cdot 1.00 \cdot 1.00}{1.00 + 1.00}$$

$$F1\text{-Score}_2 = \frac{2 \cdot 1.00}{2.00} = 1.00$$

Class 3

$$F1\text{-Score}_3 = \frac{2 \cdot 1.00 \cdot 1.00}{1.00 + 1.00}$$

$$F1\text{-Score}_3 = \frac{2 \cdot 1.00}{2.00} = 1.00$$

Class 4

$$F1\text{-Score}_4 = \frac{2 \cdot 1.00 \cdot 1.00}{1.00 + 1.00}$$

$$F1\text{-Score}_4 = \frac{2 \cdot 1.00}{2.00} = 1.00$$

Macro Average:

$$\text{Macro Precision} = \frac{\sum \text{Precision}_i}{\text{Jumlah Kelas}}$$

$$\text{Macro Precision} = \frac{1.00 + 1.00 + 1.00 + 1.00 + 1.00}{5} = 1.00$$

$$\text{Macro Recall} = \frac{\sum \text{Recall}_i}{\text{Jumlah Kelas}}$$

$$\text{Macro Recall} = \frac{1.00 + 1.00 + 1.00 + 1.00 + 1.00}{5} = 1.00$$

$$\text{Macro F1-Score} = \frac{\sum \text{F1-Score}_i}{\text{Jumlah Kelas}}$$

$$\text{Macro F1-Score} = \frac{1.00 + 1.00 + 1.00 + 1.00 + 1.00}{5} = 1.00$$

Weighted Average:

$$\text{Weighted Precision} = \frac{\sum (\text{Precision}_i \cdot \text{Support}_i)}{\text{Total Data}}$$

Weighted Precision

$$= \frac{(1.00 \cdot 1238) + (1.00 \cdot 1363) + (1.00 \cdot 851) + (1.00 \cdot 4832) + (1.00 \cdot 2076)}{10360}$$

$$\text{Weighted Precision} = \frac{1238 + 1363 + 851 + 4832 + 2076}{10360} = 1.00$$

$$\text{Weighted Recall} = \frac{\sum (\text{Recall}_i \cdot \text{Support}_i)}{\text{Total Data}}$$

Weighted Recall

$$= \frac{(1.00 \cdot 1238) + (1.00 \cdot 1363) + (1.00 \cdot 851) + (1.00 \cdot 4832) + (1.00 \cdot 2076)}{10360}$$

$$\text{Weighted Recall} = \frac{1238 + 1363 + 851 + 4832 + 2076}{10360} = 1.00$$

$$\text{Weighted F1-Score} = \frac{\sum(\text{F1-Score}_i \cdot \text{Support}_i)}{\text{Total Data}}$$

Weighted F1-Score

$$= \frac{(1.00 \cdot 1238) + (1.00 \cdot 1363) + (1.00 \cdot 851) + (1.00 \cdot 4832) + (1.00 \cdot 2076)}{10360}$$

$$\text{Weighted F1-Score} = \frac{1238 + 1363 + 851 + 4832 + 2076}{10360} = 1.00$$

BAB V

PENUTUP

5.1 Kesimpulan

Penelitian ini telah berhasil menganalisis sentimen kepercayaan publik terhadap Donald Trump menggunakan data dari aplikasi X dengan metode K-Means Clustering. Hasil analisis menunjukkan adanya variasi yang signifikan dalam sentimen publik, yang tercermin dari pembentukan beberapa cluster dengan karakteristik yang berbeda. Dominasi sentimen negatif dalam dataset memberikan gambaran bahwa banyak pengguna media sosial memiliki pandangan kritis terhadap Donald Trump. Selain itu, penelitian ini juga menyoroti efektivitas metode K-Means dalam mengelompokkan data sentimen, yang dapat dijadikan referensi untuk penelitian lebih lanjut di bidang analisis sentimen dan pengolahan data teks. Dengan demikian, penelitian ini tidak hanya memberikan wawasan yang lebih dalam tentang persepsi masyarakat, tetapi juga menegaskan pentingnya media sosial sebagai platform untuk memahami dinamika opini publik dalam konteks politik yang berkembang. Diharapkan hasil penelitian ini dapat menjadi acuan bagi peneliti selanjutnya dalam mengeksplorasi lebih lanjut hubungan antara media sosial dan sentimen publik terhadap isu-isu politik.

5.2 Saran

Dari hasil penelitian ini, disarankan agar penelitian selanjutnya mengembangkan analisis sentimen dengan menggunakan dataset yang lebih luas dan beragam, termasuk data dari platform media sosial lainnya. Hal ini akan memberikan perspektif yang lebih komprehensif mengenai opini

publik terhadap tokoh politik. Selain itu, pendekatan multimodal yang menggabungkan analisis teks dengan elemen visual atau audio dapat meningkatkan pemahaman tentang dinamika opini masyarakat. Implementasi hasil penelitian ini dalam strategi komunikasi politik oleh politisi dan partai politik juga sangat penting, agar mereka dapat lebih responsif terhadap isu-isu yang relevan dan kebutuhan masyarakat. Terakhir, peningkatan kualitas data melalui teknik preprocessing yang lebih canggih, termasuk penggunaan metode Natural Language Processing (NLP) yang lebih maju, diharapkan dapat meningkatkan akurasi dalam analisis sentimen yang dilakukan.

DAFTAR PUSTAKA

- Alexandridis, G., Varlamis, I., Korovesis, K., Caridakis, G., & Tsantilas, P. (2021). A survey on sentiment analysis and opinion mining in greek social media. *Information (Switzerland)*, *12*(8), 1–17. <https://doi.org/10.3390/info12080331>
- Chaudhry, H. N., Javed, Y., Kulsoom, F., Mehmood, Z., Khan, Z. I., Shoaib, U., & Janjua, S. H. (2021). Sentiment analysis of before and after elections: Twitter data of U.S. election 2020. *Electronics (Switzerland)*, *10*(17), 1–26. <https://doi.org/10.3390/electronics10172082>
- Chemouil, P., Hui, P., Kellerer, W., Li, Y., Stadler, R., Wen, Y., & Zhang, Y. (2019). Special Issue on Artificial Intelligence and Machine Learning for Networking and Communications. *IEEE Journal on Selected Areas in Communications*, *37*(6), 1185–1191. <https://doi.org/10.1109/JSAC.2019.2909076>
- del Valle, E., & de la Fuente, L. (2023). Sentiment analysis methods for politics and hate speech contents in Spanish language: a systematic review. *IEEE Latin America Transactions*, *21*(3), 408–418. <https://latamt.ieeer9.org/index.php/transactions/article/view/7268>
- Goni, M. W., Gustian, D., & Sembiring, F. (2021). Implementasi K-Means Dalam Pengelompokan Penyebaran COVID-19 di Jawa Barat. *Progresif: Jurnal Ilmiah Komputer*, *17*(2), 107. <https://doi.org/10.35889/progresif.v17i2.648>
- Gustientiedina, G., Adiya, M. H., & Desnelita, Y. (2019). Penerapan Algoritma K-Means Untuk Clustering Data Obat-Obatan. *Jurnal Nasional Teknologi Dan Sistem Informasi*, *5*(1), 17–24. <https://doi.org/10.25077/teknosi.v5i1.2019.17-24>
- Haviluddin, H., Patandianan, S. J., Putra, G. M., Puspitasari, N., & Pakpahan, H. S. (2021). Implementasi Metode K-Means Untuk Pengelompokan Rekomendasi Tugas Akhir. *Informatika Mulawarman : Jurnal Ilmiah Ilmu Komputer*, *16*(1), 13.

<https://doi.org/10.30872/jim.v16i1.5182>

Irsyad, H., & Pribadi, M. R. (2020). Implementasi Text Mining Dalam Pengelompokan Data Tweet Pertanian Indonesia Dengan K-Means. *Kurawal - Jurnal Teknologi, Informasi Dan Industri*, 3(2), 164–172. <https://doi.org/10.33479/kurawal.v3i2.347>

Kurniawan, I., & Susanto, A. (2019). Implementasi Metode K-Means dan Naïve Bayes Classifier untuk Analisis Sentimen Pemilihan Presiden (Pilpres) 2019. *Eksplora Informatika*, 9(1), 1–10. <https://doi.org/10.30864/eksplora.v9i1.237>

Mahesh, B. (2020). Machine Learning Algorithms - A Review. *International Journal of Science and Research (IJSR)*, 9(1), 381–386. <https://doi.org/10.21275/art20203995>

Metcalf, J., Singh, R., Moss, E., Tafesse, E., & Watkins, E. A. (2023). Taking Algorithms to Courts: A Relational Approach to Algorithmic Accountability. *ACM International Conference Proceeding Series*, 1450–1462. <https://doi.org/10.1145/3593013.3594092>

Mihartinah, D., Anggarawati, S., & Nasution. (2023). Pengaruh Viral Marketing, Harga Dan Kepercayaan Konsumen Terhadap Keputusan Pembelian Online Melalui Media Sosial Instagram. *Student Journal of Business and Management*, 6(1), 21–42. <https://www.bps.go.id/>

Nugroho, G., Murdiansyah, D. T., & Lhaksana, K. M. (2021). Analisis Sentimen Pemilihan Presiden Amerika 2020 di Twitter Menggunakan Naïve Bayes dan Support Vector Machine. *E-Proceeding of Engineering*, 8(5), 10106–10115.

Şeker kaya, A. (2020). Contemporary Issues in Strategic Marketing. In *Contemporary Issues in Strategic Marketing*. <https://doi.org/10.26650/b/ss05.2020.002>

Sinaga, K. P., & Yang, M. S. (2020). Unsupervised K-means clustering algorithm.

IEEE Access, 8, 80716–80727. <https://doi.org/10.1109/ACCESS.2020.2988796>

Tan, J. J., Firdaus, A., & Aksar, I. A. (2024). Social Media for Political Information: A Systematic Literature Review. *Jurnal Komunikasi: Malaysian Journal of Communication*, 40(1), 77–98. <https://doi.org/10.17576/JKMJC-2024-4001-05>

Wirayasa, I. K. A., & Santoso, H. (2022). Analisis Employee Satisfaction Menggunakan Teknik Clustering Dan Classification Machine Learning. *Progresif: Jurnal Ilmiah Komputer*, 18(1), 1. <https://doi.org/10.35889/progresif.v18i1.766>

Lampiran

1. analisa.py

- Streamlit

```
import pandas as pd
import numpy as np
import streamlit as st
import matplotlib.pyplot as plt
from sklearn.feature_extraction.text import TfidfVectorizer
from sklearn.cluster import KMeans

# Load dataset
@st.cache
def load_data():
    return pd.read_csv('tweets_01-08-2021.csv')

# Main function
def main():
    st.title("K-Means Clustering of Tweets on Economy and Security")

    # Load data
    data = load_data()
    st.write("Dataset Loaded:")
    st.dataframe(data.head())

    # Define keywords for filtering
    economy_keywords = ['economy', 'jobs', 'growth', 'tax']
    security_keywords = ['security', 'safety', 'crime', 'law']

    # Filter relevant tweets
    data['economy_relevant'] = data['text'].apply(lambda x: any(word in x.lower() for word in economy_keywords))
    data['security_relevant'] = data['text'].apply(lambda x: any(word in x.lower() for word in security_keywords))

    economy_tweets = data[data['economy_relevant']]
    security_tweets = data[data['security_relevant']]
```

```

# Display pie charts for economy and security tweets
st.subheader("Distribution of Relevant Tweets")
labels = ['Ec'] (variable) economy_tweets: DataFrame
sizes = [len(economy_tweets), len(security_tweets)]
colors = ['#ff9999', '#66b3ff']

plt.figure(figsize=(8, 6))
plt.pie(sizes, labels=labels, colors=colors, autopct='%1.1f%%', startangle=140)
plt.axis('equal') # Equal aspect ratio ensures that pie is drawn as a circle.
st.pyplot(plt)

if economy_tweets.empty and security_tweets.empty:
    st.write("No relevant tweets found.")
    return

# Preprocess data for economy tweets
if not economy_tweets.empty:
    vectorizer_economy = TfidfVectorizer(stop_words='english')
    X_economy = vectorizer_economy.fit_transform(economy_tweets['text'])

    # K-Means parameters for economy
    k_economy = st.slider("Select number of clusters for Economy (K):", 1, 10, 3)
    kmeans_economy = KMeans(n_clusters=k_economy, random_state=0)
    kmeans_economy.fit(X_economy)

    # Assign clusters
    economy_tweets['cluster'] = kmeans_economy.labels_

    # Show cluster distribution for economy
    st.subheader("Economy Cluster Distribution")
    st.write(economy_tweets['cluster'].value_counts())

# Visualize economy clusters
plt.figure(figsize=(10, 6)) (variable) economy_tweets: DataFrame
plt.scatter(economy_tweets['cluster'], economy_tweets['favorites'], alpha=0.5)
plt.title("Economy Cluster Visualization based on Favorites")
plt.xlabel("Cluster")
plt.ylabel("Number of Favorites")
st.pyplot(plt)

# Display tweets in each economy cluster
for i in range(k_economy):
    st.subheader(f"Economy Tweets in Cluster {i}")
    cluster_tweets_economy = economy_tweets[economy_tweets['cluster'] == i]
    st.write(cluster_tweets_economy[['text', 'favorites', 'date']])

Preprocess data for security tweets
not security_tweets.empty:
    vectorizer_security = TfidfVectorizer(stop_words='english')
    X_security = vectorizer_security.fit_transform(security_tweets['text'])

    # K-Means parameters for security
    k_security = st.slider("Select number of clusters for Security (K):", 1, 10, 3)
    kmeans_security = KMeans(n_clusters=k_security, random_state=0)
    kmeans_security.fit(X_security)

    # Assign clusters
    security_tweets['cluster'] = kmeans_security.labels_

    # Show cluster distribution for security

```

```

# Assign clusters
security_tweets['cluster'] = kmeans_security.labels_

# Show cluster distribution for security
st.subheader("Security Cluster Distribution")
st.write(security_tweets['cluster'].value_counts())

# Visualize security clusters
plt.figure(figsize=(10, 6))
plt.scatter(security_tweets['cluster'], security_tweets['favorites'], alpha=0.5)
plt.title("Security Cluster Visualization based on Favorites")
plt.xlabel("Cluster")
plt.ylabel("Number of Favorites")
st.pyplot(plt)

# Display tweets in each security cluster
for i in range(k_security):
    st.subheader(f"Security Tweets in Cluster {i}")
    cluster_tweets_security = security_tweets[security_tweets['cluster'] == i]
    st.write(cluster_tweets_security[['text', 'favorites', 'date']])

if __name__ == "__main__":
    main()

```

- Analisis.py

```

import pandas as pd
import numpy as np
import streamlit as st
import matplotlib.pyplot as plt
from sklearn.feature_extraction.text import TfidfVectorizer
from sklearn.cluster import KMeans
from sklearn.metrics import f1_score
import nltk
from wordcloud import WordCloud
from textblob import TextBlob

# Unduh stopwords NLTK (jalankan sekali)
nltk.download('stopwords')

# Judul Aplikasi
st.title("Analisis Tweet Donald Trump")

# 1. Memuat data
data = pd.read_csv('Dataset_new_real_donaldtrump_insult_tweets_2014_to_2021.csv')
tweets = data['tweet'].astype(str)

# 2. Preprocessing teks
def preprocess_text(text):
    text = text.lower()
    text = ''.join(char for char in text if char.isalpha() or char.isspace())
    return text

# Terapkan preprocessing
tweets = tweets.apply(preprocess_text)

# 3. Ekstraksi fitur menggunakan TF-IDF
from nltk.corpus import stopwords

```

```

from nltk.corpus import stopwords
vectorizer = TfidfVectorizer(stop_words=stopwords.words('indonesian'))
X = vectorizer.fit_transform(tweets)

# 4. K-Means Clustering
num_clusters = 5
kmeans = KMeans(n_clusters=num_clusters, random_state=42)
kmeans.fit(X)

# 5. Menambahkan hasil clustering (variable) labels_ ndarray
data['cluster'] = kmeans.labels_

# 6. Menambahkan kolom tahun
data['date'] = pd.to_datetime(data['date'])
data['year'] = data['date'].dt.year

# 7. Analisis Sentimen
def get_sentiment(text):
    analysis = TextBlob(text)
    if analysis.sentiment.polarity > 0:
        return 'Positif'
    elif analysis.sentiment.polarity < 0:
        return 'Negatif'
    else:
        return 'Netral'

data['sentiment'] = data['tweet'].apply(get_sentiment)

# 8. Visualisasi

# a. Jumlah Tweet per Cluster
st.subheader('Jumlah Tweet per Cluster')

```

```

# a. Jumlah Tweet per Cluster
st.subheader('Jumlah Tweet per Cluster')
fig1, ax1 = plt.subplots()
data['cluster'].value_counts().plot(kind='bar', color='skyblue', ax=ax1)
ax1.set_title('Jumlah Tweet per Cluster')
ax1.set_xlabel('Cluster')
ax1.set_ylabel('Jumlah Tweet')
st.pyplot(fig1)

# b. Kata Paling Umum per Cluster
for i in range(num_clusters):
    cluster_tweets = data[data['cluster'] == i]['tweet']
    text = ' '.join(cluster_tweets)
    wordcloud = WordCloud(width=800, height=400, background_color='white').generate(text)

    st.subheader(f'Word Cloud untuk Cluster {i}')
    plt.figure(figsize=(10, 6))
    plt.imshow(wordcloud, interpolation='bilinear')
    plt.axis('off')
    st.pyplot()

# c. Grafik Distribusi Panjang Tweet per Cluster
data['tweet_length'] = data['tweet'].apply(len)

st.subheader('Rata-rata Panjang Tweet per Cluster')
fig2, ax2 = plt.subplots()
data.groupby('cluster')['tweet_length'].mean().plot(kind='bar', color='lightcoral', ax=ax2)
ax2.set_title('Rata-rata Panjang Tweet per Cluster')
ax2.set_xlabel('Cluster')
ax2.set_ylabel('Rata-rata Panjang Tweet')
st.pyplot(fig2)

# d. Diagram Lingkaran Jumlah Tweet per Cluster
st.subheader('Distribusi Tweet per Cluster')
fig3, ax3 = plt.subplots()
data['cluster'].value_counts().plot(kind='pie', autopct='%1.1f%%', startangle=90, colors=plt.cm.Paired.colors, ax=ax3)
ax3.set_title('Distribusi Tweet per Cluster')
ax3.set_ylabel('')
st.pyplot(fig3)

# e. Jumlah Tweet per Tahun
st.subheader('Jumlah Tweet per Tahun')
fig4, ax4 = plt.subplots()
data['year'].value_counts().sort_index().plot(kind='bar', color='lightgreen', ax=ax4)
ax4.set_title('Jumlah Tweet per Tahun')
ax4.set_xlabel('Tahun')
ax4.set_ylabel('Jumlah Tweet')
st.pyplot(fig4)

# f. Distribusi Sentimen
st.subheader('Distribusi Sentimen Tweet')
sentiment_counts = data['sentiment'].value_counts()

fig5, ax5 = plt.subplots()
sentiment_counts.plot(kind='bar', color=['lightblue', 'lightcoral', 'lightgreen'], ax=ax5)
ax5.set_title('Distribusi Sentimen Tweet')
ax5.set_xlabel('Sentimen')
ax5.set_ylabel('Jumlah Tweet')
ax5.set_xticklabels(sentiment_counts.index, rotation=0)
st.pyplot(fig5)

# g. Memisahkan Sentimen
st.subheader('Tweet Positif, Negatif, dan Netral')

```

```

# Menghitung jumlah tweet berdasarkan sentimen
positive_tweets = data[data['sentiment'] == 'Positif']
negative_tweets = data[data['sentiment'] == 'Negatif']
neutral_tweets = data[data['sentiment'] == 'Netral']

st.write(f"Jumlah Tweet Positif: {len(positive_tweets)}")
st.write(f"Jumlah Tweet Negatif: {len(negative_tweets)}")
st.write(f"Jumlah Tweet Netral: {len(neutral_tweets)}")

# Menampilkan beberapa contoh tweet
st.write("Contoh Tweet Positif:")
st.write(positive_tweets['tweet'].sample(5).values)

st.write("Contoh Tweet Negatif:")
st.write(negative_tweets['tweet'].sample(5).values)

st.write("Contoh Tweet Netral:")
st.write(neutral_tweets['tweet'].sample(5).values)

# h. Menghitung F1 Score
if 'true_sentiment' in data.columns: # Pastikan kolom ini ada
    # Mencetak F1 Score
    f1 = f1_score(data['true_sentiment'], data['sentiment'], average='weighted')
    st.write(f"F1 Score: {f1:.2f}")
else:
    st.write("Kolom 'true_sentiment' tidak ditemukan dalam dataset. F1 Score tidak dapat dihitung.")

```

- Model

Evaluasi Model

PENGUJIAN MODEL

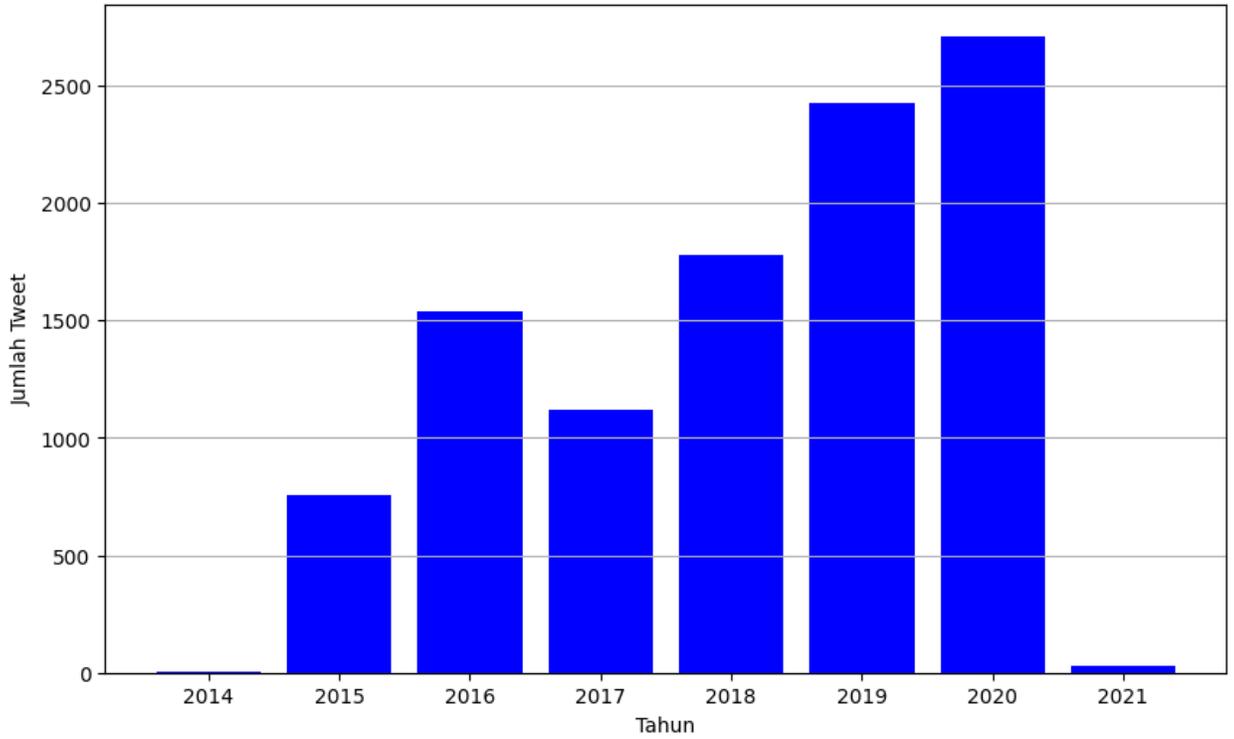
```
# Impor yang diperlukan
from sklearn.metrics import classification_report

# Hitung evaluasi model
report = classification_report(data['cluster'], kmeans.labels_)

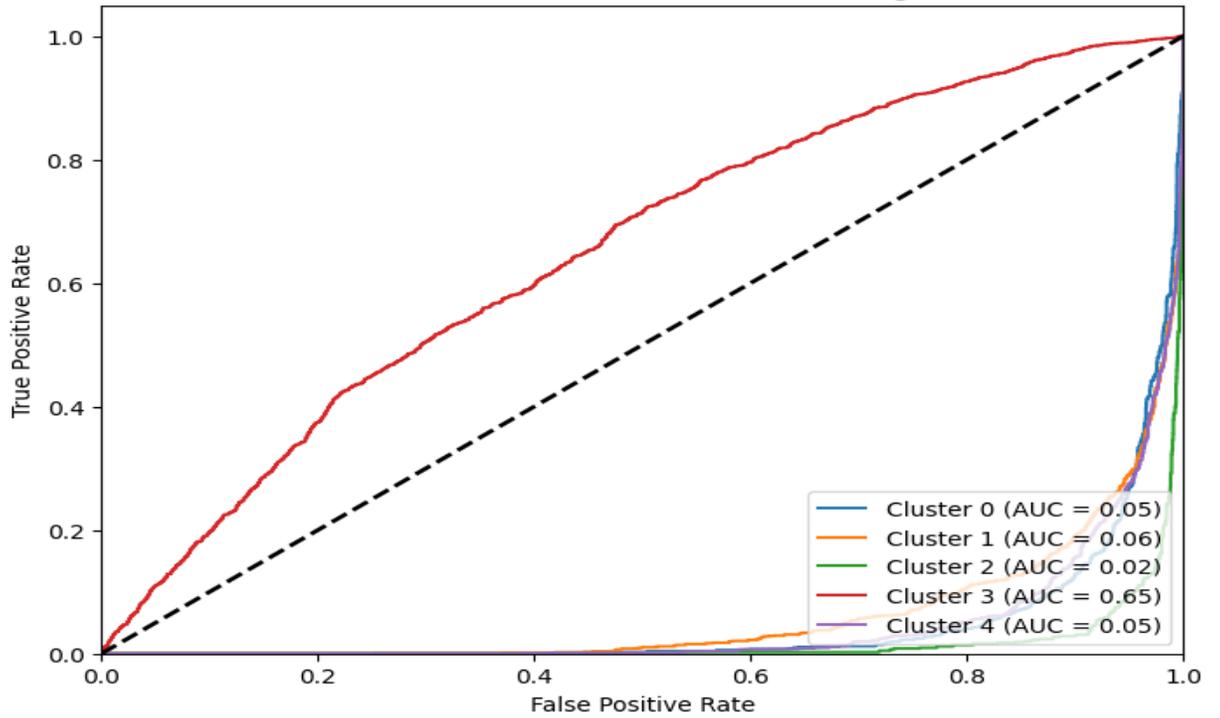
# Tampilkan hasil
print(report)
```

	precision	recall	f1-score	support
0	1.00	1.00	1.00	1238
1	1.00	1.00	1.00	1363
2	1.00	1.00	1.00	851
3	1.00	1.00	1.00	4832
4	1.00	1.00	1.00	2076
accuracy			1.00	10360
macro avg	1.00	1.00	1.00	10360
weighted avg	1.00	1.00	1.00	10360

Jumlah Tweet Donald Trump per Tahun

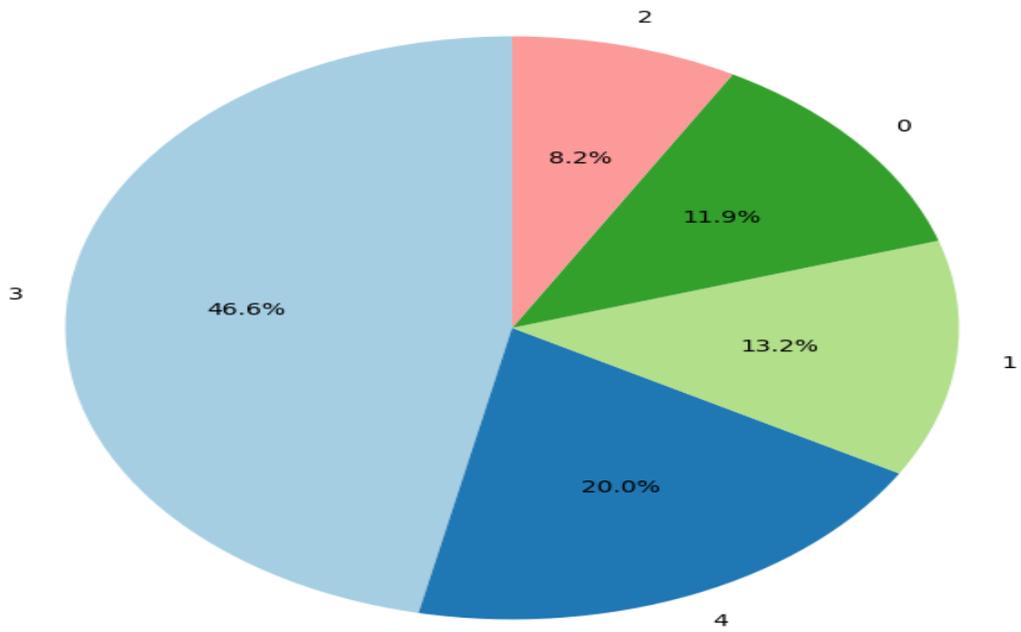


Kurva ROC Model K-Means Clustering



- Visualisasi Data

Distribusi Tweet per Cluster



2. Turnitin Draf Skripsi

 Page 2 of 63 - Integrity Overview Submission ID trnoid::361882972406

18% Overall Similarity

The combined total of all matches, including overlapping sources, for each database.

Filtered from the Report

- Bibliography
- Quoted Text
- Methods and Materials
- Submitted works
- Crossref posted content database

Top Sources

15%  Internet sources
10%  Publications
0%  Submitted works (Student Papers)

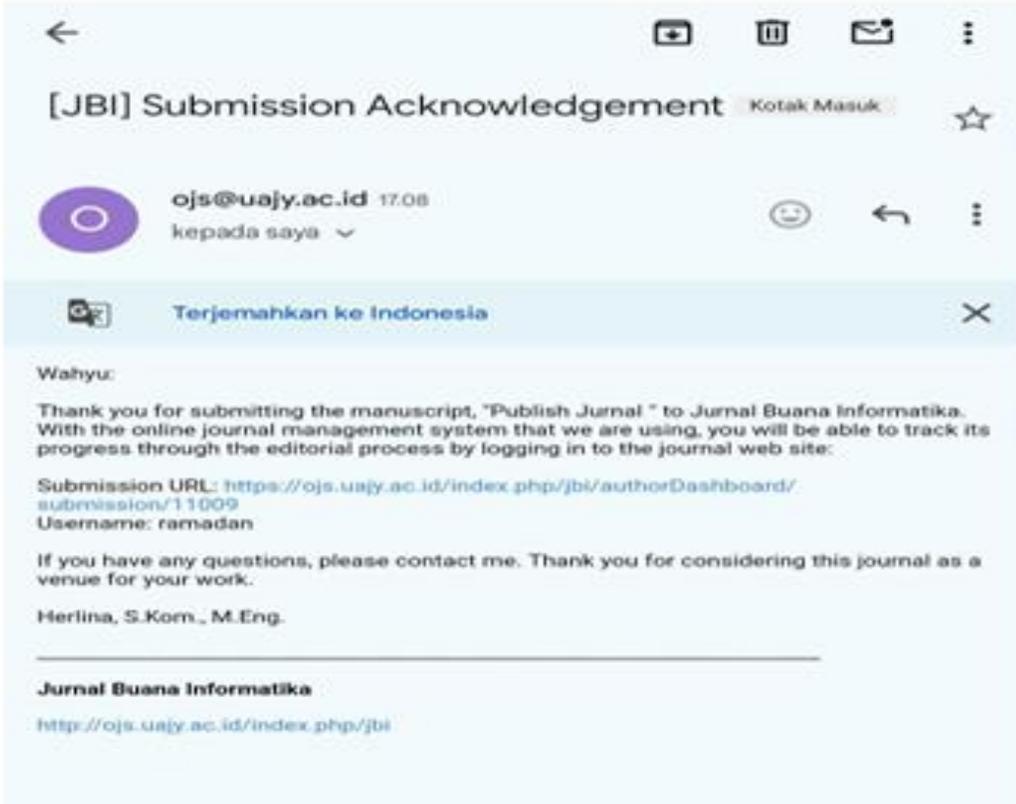
Integrity Flags

0 Integrity Flags for Review
No suspicious text manipulations found.

Our system's algorithms look deeply at a document for any inconsistencies that would set it apart from a normal submission. If we notice something strange, we flag it for you to review.

A Flag is not necessarily an indicator of a problem. However, we'd recommend you focus your attention there for further review.

3. Publish Jurnal



[JBI] Submission Acknowledgement Kotak Masuk

 **ojs@uajy.ac.id** 17.08
kepada saya

 **Terjemahkan ke Indonesia**

Wahyu:

Thank you for submitting the manuscript, "Publish Jurnal " to Jurnal Buana Informatika. With the online journal management system that we are using, you will be able to track its progress through the editorial process by logging in to the journal web site:

Submission URL: <https://ojs.uajy.ac.id/index.php/jbi/authorDashboard/submission/11009>
Username: ramadan

If you have any questions, please contact me. Thank you for considering this journal as a venue for your work.

Herlina, S.Kom., M.Eng.

Jurnal Buana Informatika
<http://ojs.uajy.ac.id/index.php/jbi>

