ANALISIS SENTIMEN OPINI PUBLIK KEMENTERIAN KEUANGAN MENGGUNAKAN *NAIVE BAYES* DAN *SUPPORT VECTOR MACHINE*

LAPORAN SKRIPSI

Muhammad Alviansyah Gustama

00000051526

PROGRAM STUDI SISTEM INFORMASI

FAKULTAS TEKNIK DAN INFORMATIKA

UNIVERSITAS MULTIMEDIA NUSANTARA

TANGERANG

2024

ANALISIS SENTIMEN OPINI PUBLIK KEMENTERIAN KEUANGAN MENGGUNAKAN NAIVE BAYES DAN SUPPORT VECTOR MACHINE

LAPORAN SKRIPSI

Diajukan sebagai Salah Satu Syarat untuk Memperoleh

Gelar Sarjana Komputer (S.Kom)

Muhammad Alviansyah Gustama

00000051526

PROGRAM STUDI SISTEM INFORMASI

FAKULTAS TEKNIK DAN INFORMATIKA

UNIVERSITAS MULTIMEDIA NUSANTARA

TANGERANG

2024

HALAMAN PERNYATAAN TIDAK PLAGIAT

Dengan ini saya,

Nama	: Muhammad Alviansyah Gustama
Nomor Induk Mahasiswa	: 00000051526
Program studi	: Sistem Informasi

Laporan Skripsi dengan judul: "ANALISIS SENTIMEN OPINI PUBLIK KEMENTERIAN KEUANGAN MENGGUNAKAN NAIVE BAYES DAN SUPPORT VECTOR MACHINE"

merupakan hasil karya saya sendiri bukan plagiat dari karya ilmiah yang ditulis oleh orang lain, dan semua sumber, baik yang dikutip maupun dirujuk, telah saya nyatakan dengan benar serta dicantumkan di Daftar Pustaka.

Jika di kemudian hari terbukti ditemukan kecurangan/penyimpangan, baik dalam pelaksanaan skripsi maupun dalam penulisan laporan skripsi, saya bersedia menerima konsekuensi dinyatakan TIDAK LULUS untuk Tugas Akhir yang telah saya tempuh.

Tangerang, 23 Oktober 2024

Muhammad Alviansyah Gustama

HALAMAN PERSETUJUAN PUBLIKASI KARYA ILMIAH

Yang bertanda tangan dibawah ini:

Nama : Muhammad Alviansyah Gustama

NIM : 00000051526

Program Studi : Sistem Informasi

Jenjang : ~~D3~~/S1/~~S2~~

Judul Karya Ilmiah :

ANALISIS SENTIMEN OPINI PUBLIK KEMENTERIAN KEUANGAN MENGUNAKAN NAIVE BAYES DAN SUPPORT VECTOR MACHINE

Menyatakan dengan sesungguhnya bahwa saya bersedia* **(pilih salah satu):**

☑ Saya bersedia memberikan izin sepenuhnya kepada Universitas Multimedia Nusantara untuk mempublikasikan hasil karya ilmiah saya ke dalam repositori Knowledge Center sehingga dapat diakses oleh Sivitas Akademika UMN/Publik. Saya menyatakan bahwa karya ilmiah yang saya buat tidak mengandung data yang bersifat konfidensial.

☐ Saya tidak bersedia mempublikasikan hasil karya ilmiah ini ke dalam repositori Knowledge Center, dikarenakan: dalam proses pengajuan publikasi ke jurnal/konferensi nasional/internasional (dibuktikan dengan *letter of acceptance*) **.

Tangerang, 23 Oktober 2024

(Muhammad Alviansyah Gustama)

HALAMAN PENGESAHAN

Skripsi dengan judul

ANALISIS SENTIMEN OPINI PUBLIK KEMENTERIAN KEUANGAN MENGGUNAKAN NAIVE BAYES DAN SUPPORT VECTOR MACHINE

Oleh

Nama	: Muhammad Alviansyah Gustama
NIM	: 00000051526
Program Studi	: Sistem Informasi
Fakultas	: Teknik dan Informatika

Telah diujikan pada hari Rabu, 23 Oktober 2024
Pukul 10.00 s.d 12.00 dan dinyatakan
LULUS
Dengan susunan penguji sebagai berikut.

Ketua Sidang

Dr. Friska Natalia, S.Kom., M.T.
0601128307

Penguji

Wella, S.Kom., M.M.S.I.
305119101

Pembimbing 1

Dr. Santo Fernandi Wijaya S.Kom., M.M
081435

Pembimbing 2

Iwan Prasetiawan, S.Kom., M.M
00552

Ketua Program Studi Sistem Informasi

Ririn Ikana Desanti, S.Kom., M.Kom

HALAMAN PERSETUJUAN PUBLIKASI KARYA ILMIAH UNTUK KEPENTINGAN AKADEMIS

Sebagai civitas academica Universitas Multimedia Nusantara, saya yang bertanda tangan di bawah ini:

Nama : Muhammad Alviansyah Gustama

NIM : 00000051526

Program Studi : Sistem Informasi

Fakultas : Fakultas Teknik Dan Informatika

JenisKarya : *~~Tesis~~/Skripsi/~~Tugas Akhir~~ (*coret salah satu)

Demi pengembangan ilmu pengetahuan, menyetujui untuk memberikan kepada Universitas Multimedia Nusantara Hak Bebas Royalti Nonekslusif (*Non-exclusive Royalty-Free Right*) atas karya ilmiah saya yang berjudul.

ANALISIS SENTIMEN OPINI PUBLIK KEMENTERIAN KEUANGAN MENGGUNAKAN NAIVE BAYES DAN SUPPORT VECTOR MACHINE

Beserta perangkat yang ada (jika diperlukan). Dengan Hak Bebas Royalti Noneksklusif ini, Universitas Multimedia Nusantara berhak menyimpan, mengalihmediakan/mengalihformatkan, mengelola dalam bentuk pangkalan data (*database*), merawat, dan memublikasikan tugas akhir saya selama tetap mencantumkan nama saya sebagai penulis/pencipta dan sebagai pemilik Hak Cipta. Demikian pernyataan ini saya buat dengan sebenarnya.

Tangerang, 23 Oktober 2024

Yang menyatakan,

Muhammad Alviansyah Gustama

KATA PENGANTAR

Puji Syukur atas selesainya penulisan Laporan Skripsi ini dengan judul: "ANALISIS SENTIMEN OPINI PUBLIK KEMENTERIAN KEUANGAN MENGGUNAKAN NAIVE BAYES DAN SUPPORT VECTOR MACHINE" dilakukan untuk memenuhi salah satu syarat untuk mencapai gelar sarjana Jurusan Sistem Informasi pada Fakultas Teknik dan Informatika Universitas Multimedia Nusantara. Saya menyadari bahwa, tanpa bantuan dan bimbingan dari berbagai pihak, dari masa perkuliahan sampai pada penyusunan tugas akhir ini, sangatlah sulit bagi saya untuk menyelesaikan tugas akhir ini. Oleh karena itu, saya mengucapkan terima kasih kepada

Mengucapkan terima kasih

1. Dr. Ninok Leksono, selaku Rektor Universitas Multimedia Nusantara.
2. Dr. Eng. Niki Prastomo, S.T., M.Sc., selaku Dekan Fakultas Teknik dan Informatika Universitas Multimedia Nusantara.
3. Ririn Ikana, S.Kom., M.Kom., selaku Ketua Program Studi Sistem Informasi Universitas Multimedia Nusantara.
4. Dr. Santo Fernandi Wijaya sebagai Pembimbing 1 yang telah banyak meluangkan waktu untuk memberikan bimbingan, arahan dan motivasi atas terselesainya skripsi ini.
5. Iwan Prasetiawan, S.Kom., M.M. sebagai Pembimbing 2 yang telah banyak meluangkan waktu untuk memberikan bimbingan, arahan dan motivasi atas terselesainya skripsi ini.
6. Orang Tua yang telah memberikan bantuan dukungan material dan moral, sehingga penulis dapat menyelesaikan skripsi ini.
7. Nurul Aini Lativah yang telah memberikan bantuan dukungan material dan moral, sehingga penulis dapat menyelesaikan skripsi ini.

Semoga karya ilmiah ini bermanfaat, baik sebagai sumber informasi maupun sumber inspirasi, bagi para pembaca.

Tangerang, 23 Oktober 2024

Muhammad Alviansyah Gustama

ANALISIS SENTIMEN OPINI PUBLIK KEMENTERIAN KEUANGAN MENGGUNAKAN NAIVE BAYES DAN SUPPORT VECTOR MACHINE

Muhammad Alviansyah Gustama

ABSTRAK

Kementerian Keuangan seringkali menjadi pusat perhatian dan diskusi di kalangan masyarakat. Opini publik tentang kinerja dan kebijakan Kementerian sangat berpengaruh dalam merumuskan langkah-langkah yang lebih tepat dan responsif. Pada penelitian untuk memahami lebih dalam pandangan masyarakat terhadap Kementerian Keuangan.

Penelitian ini bertujuan untuk membandingkan analisis sentimen opini publik terkait pandangan masyarakat terhadap Kementerian Keuangan menggunakan dua model algoritma, yaitu *Naive Bayes* dan *Support Vector Machine (SVM).* Analisis sentimen digunakan untuk memahami opini publik, sehingga dapat membantu Kementerian Keuangan dalam mengidentifikasi persepsi masyarakat. Metode yang digunakan adalah framework CRISP-DM, yang membantu mengatur proses analisis data secara sistematis dari pemahaman bisnis hingga penerapan hasil.

Hasil penelitian menunjukkan bahwa model algoritma *Support Vector Machine (SVM)* memberikan kinerja yang lebih baik dengan nilai akurasi sebesar 90%, dibandingkan dengan *Naive Bayes* yang hanya mencapai akurasi sebesar 85.32%. Hal ini menunjukkan bahwa *Support Vector Machine (SVM)* lebih efektif dalam mengenali dan mengklasifikasikan sentimen opini publik terhadap Kementerian Keuangan. Penelitian ini membantu Kementerian Keuangan dengan mengukur dan memahami sentimen publik, sehingga lembaga bisa merumuskan kebijakan dan strategi komunikasi yang lebih efektif dan sesuai dengan kebutuhan serta pandangan masyarakat. Pemahaman yang lebih mendalam tentang opini publik, lembaga dapat menemukan area yang memerlukan perbaikan dalam hal keterbukaan dan transparansi.

Kata kunci: Analisis Sentimen, Kementerian Keuangan, *Naive Bayes, Support Vector Machine.*

Comparison of Public Opinion Sentiment Analysis for the Ministry of Finance Using Naive Bayes And Support Vector Machine

Muhammad Alviansyah Gustama

ABSTRACT (English)

The Ministry of Finance is often the center of attention and discussion among the public. Public opinion on the performance and policies of the Ministry greatly influences the formulation of more appropriate and responsive steps. In this study, we aim to better understand the public's views on the Ministry of Finance.

This study aims to compare the analysis of public opinion sentiment related to the public's views on the Ministry of Finance using two algorithm models, namely Naive Bayes and Support Vector Machine (SVM). Sentiment analysis is used to understand public opinion, so that it can help the Ministry of Finance in identifying public perceptions. The method used is the CRISP-DM framework, which helps organize the data analysis process systematically from business understanding to implementing results.

The results of the study showed that the Support Vector Machine (SVM) algorithm model provided better performance with an accuracy value of 90%, compared to Naive Bayes which only achieved an accuracy of 85.32%. This shows that Support Vector Machine (SVM) is more effective in recognizing and classifying public opinion sentiment towards the Ministry of Finance. This study helps the Ministry of Finance by measuring and understanding public sentiment, so that institutions can formulate more effective communication policies and strategies that are in accordance with the needs and views of the public. With a deeper understanding of public opinion, institutions can identify areas that need improvement in terms of openness and transparency.

Keywords*: Ministry of Finance, Naive Bayes, Support Vector Machine, Sentiment Analysis.*

DAFTAR ISI

DAFTAR TABEL

DAFTAR GAMBAR

DAFTAR RUMUS

DAFTAR LAMPIRAN

BAB I

PENDAHULUAN

1.1 Latar Belakang

Dalam era globalisasi dan kemajuan teknologi informasi dan keterbukaan informasi, dinamika hubungan antara pemerintah dan masyarakat mengalami perubahan signifikan. Opini publik merupakan pendapat, gagasan, atau pun kritik yang disampaikan masyarakat melalui kanal media tertentu [1]. Opini publik, sebagai ekspresi pandangan dan perasaan kolektif masyarakat, telah menjadi kekuatan yang mampu membentuk dan mempengaruhi kebijakan publik. Penyampaian opini tidak lagi terbatas pada forum tradisional, tetapi juga merambah ke ranah digital melalui berbagai *platform online*, seperti media sosial, *blog*, dan forum diskusi daring. Saat ini, menyampaikan pendapat tidak lagi menjadi hal yang sulit karena perkembangan teknologi telah menjadi kekuatan signifikan dalam membentuk opini publik. Media komunikasi, khususnya media sosial sebagai alat penyebaran informasi, memainkan peran penting dalam menyebarluaskan isu-isu yang berkembang di masyarakat dan opini publik dalam konteks komunikasi politik [2].

Penggunaan media sosial semakin meningkat, pemerintah turut memanfaatkan untuk berinteraksi dengan masyarakat yang dilayani. Media sosial memungkinkan pemerintah untuk memperkuat partisipasi dalam demokrasi, publik didorong agar lebih aktif terlibat dalam proses pembuatan kebijakan, meningkatkan layanan, mengumpulkan ide, dan memperkuat transparansi. Hal ini membuat lembaga pemerintah memanfaatkan media sosial sebagai platform untuk berinteraksi dengan masyarakat, memberikan informasi, mempromosikan layanan publik, serta mendorong partisipasi dalam perancangan ide layanan masa depan [3]. Salah satu lembaga pemerintah yang memegang peran sentral dalam menjaga stabilitas ekonomi suatu negara adalah Kementerian Keuangan. Kementerian Keuangan adalah salah satu lembaga

pemerintah yang memiliki peran sentral dalam mengelola keuangan Negara [4]. Fungsi utama sebagai pengelola kebijakan fiskal, Kementerian Keuangan memiliki tanggung jawab dalam perencanaan, pengelolaan, dan pengawasan keuangan publik. Seiring dengan peran strategis, keberhasilan implementasi kebijakan oleh Kementerian Keuangan tidak hanya bergantung pada aspek teknis dan regulasi, tetapi juga pada kepercayaan yang dimiliki oleh masyarakat terhadap lembaga tersebut.

Kepercayaan masyarakat terhadap Kementerian Keuangan bukan sekadar indikator keberhasilan, melainkan juga menjadi landasan utama untuk mendukung efektivitas kebijakan yang diimplementasikan. Dalam kondisi di mana informasi dengan mudah tersebar dan diakses oleh masyarakat melalui berbagai saluran digital, pemahaman terhadap sentimen dan pandangan masyarakat menjadi semakin kompleks dan dinamis. Oleh sebab itu, diperlukan analisis mendalam untuk memahami dan menggali esensi dari opini publik terkait kepercayaan masyarakat terhadap Kementerian Keuangan. Menurut data dari *We Are Social*, jumlah pengguna X global mencapai 564,1 juta pada bulan Juli 2023 [5]. Indonesia menunjukkan peningkatan peringkat dari posisi keenam menjadi peringkat keempat di antara negara-negara dengan pengguna X terbesardi dunia pada bulan yang sama. Hal ini berdasarkan laporan terbaru, yang menunjukkan pertumbuhan signifikan dalam jumlah pengguna X di Indonesia. X dianggap sebagai *platform* yang erat dengan isu-isu viral yang sedang populer pada saat tertentu dan dapat digunakan sebagai sumber informasi untuk topik-topik tertentu, termasuk evaluasi kinerja pemerintahan. Informasi yang tersedia dalam bentuk *tweet* di X dapat dianalisis menggunakan metode analisis sentimen.

Kementerian Keuangan dapat menggunakan *platform* X untuk memantau opini masyarakat, karena *platform* ini menyediakan berbagai manfaat. X menawarkan wawasan langsung dari masyarakat yang mungkin tidak mencerminkan seluruh keragaman opini yang ada di masyarakat. Hal ini dapat

menyebabkan pemahaman yang kurang lengkap atau bahkan salah tentang apa yang dibutuhkan dan diharapkan oleh masyarakat secara keseluruhan. X memberikan respons langsung dari masyarakat terkait kebijakan dan keputusan, memungkinkan tanggapan cepat terhadap perubahan sentimen atau kekhawatiran yang timbul. Selain itu, X juga berfungsi sebagai alat promosi untuk kebijakan dan program Kementerian Keuangan, meningkatkan pemahaman publik dan transparansi informasi yang disampaikan. Dengan melakukan analisis sentimen Kementerian Keuangan dapat memahami pandangan, perasaan, dan respon masyarakat terhadap kebijakan tertentu, memberikan kontribusi yang berharga dalam merancang kebijakan dan sesuai dengan kebutuhan serta harapan publik. Analisis sentimen melakukan analisis mendalam terhadap pendapat, sentimen, evaluasi, sikap, dan emosi seseorang dalam menyampaikan perasaannya terkait suatu topik yang mampu digunakan untuk membuat keputusan langsung terkait dengan komentar, produk, kebijakan, dan hal lainnya. *VADER* (*Valence Aware Dictionary and Sentiment Reasoner*) digunakan untuk mendapatkan sentimen dari opini yang diungkapkan oleh masyarakat melalui *tweet* di X [6]. *VADER* adalah suatu teknik analisis sentimen berbasis leksikon dan berdasarkan aturan yang menghasilkan informasi mengenai sentimen yang bersifat positif, negatif, atau netral. algoritma *SVM (Support Vector Machine)* serta algoritma *NBC (Naïve Bayes Classifier).* .

Penelitian terdahulu yang berkaitan dengan analisis sentimen menggunakan algoritma *Naïve Bayes* dan *Support Vector Machine (SVM)* menunjukkan berbagai hasil akurasi. Analisis sentimen netizen Twitter terhadap isu Kementerian Keuangan menghasilkan akurasi sebesar 71,7% untuk *Naïve Bayes* dan 74% untuk *SVM* [7]. Penelitian yang dilakukan oleh Yulius Bambang Seran dan Supatman menggunakan algoritma Support Vector Machine (*SVM*) untuk analisis data [8]. Hasil penelitian menunjukkan bahwa model *SVM* mencapai tingkat akurasi sebesar 66%. Analisis sentimen terhadap kenaikan cukai rokok di Twitter, *Naïve Bayes* menghasilkan akurasi sebesar 74% [9].

Penerapan algoritma *SVM* untuk analisis sentimen di Twitter mengenai Komisi Pemberantasan Korupsi (KPK) Republik Indonesia menghasilkan akurasi sebesar 82% [10]. Penelitian menggunakan algoritme *Naïve Bayes* dengan akurasi terbaik sebesar 74% dengan menggunkan SMOTE untuk menangani ketidakseimbangan data, sehingga model dapat lebih baik dalam mengenali kelas minoritas dan memberikan hasil klasifikasi yang lebih akurat [9]. Oleh karena itu, SMOTE kembali diterapkan dalam penelitian ini untuk menangani tantangan ketidakseimbangan data yang serupa.

Kontribusi dalam penelitian ini, 1) Perbandingan model algoritma *Naive Bayes* dan *SVM* digunakan untuk menganalisis opini publik terkait Kementerian Keuangan. 2) *Data source* yang digunakan berasal dari X yang didapatkan dari periode Agustus 2023 – 2024. 3) Berdasarkan evaluasi performa dua algoritma klasifikasi yang diterapkan dalam penelitian ini, yaitu *Naive Bayes* dan *Support Vector Machines (SVM)*, hasil menunjukkan bahwa *SVM* memiliki kinerja terbaik dengan akurasi mencapai 90%. 4) Dilakukan pembuatan *dashboard* visualisasi dalam bentuk *website* untuk proses *deploy* dan memperlihatkan hasil analisis.

1.2 Rumusan Masalah

Berdasarkan latar belakang, rumusan masalah yang didapatkan sebagai berikut:

1. Bagaimana hasil perbandingan performa antara algoritma Naive Bayes dan Support Vector Machines (*SVM*) dalam mengukur dan mengevaluasi akurasi model analisis sentimen terkait opini publik pada Kementerian Keuangan?
2. Bagaimana hasil opini yang dihasilkan dari analisis sentimen publik terhadap kemenkeu diimplementasikan?

1.3 Batasan Masalah

Berdasarkan rumusan masalah, batasan masalah yang ditetapkan sebagai berikut:

1. Data yang diambil berasal dari media *social* X yang terkait Kementerian Keuangan.
2. Data yang dianalisis diambil dari rentang waktu Agustus 2023 hingga Agustus 2024 dengan total jumlah data sebanyak 2354.
3. Penelitian ini membahas hasil perbandingan performa pada opini publik terhadap kemenkeu menggunakan algoritma *Naive Bayes,* dan *Support Vector Machines (SVM)* [11].
4. Hasil klasifikasi yang ditampilkan mencakup pengelompokan opini menjadi tiga kategori, yaitu positif, negatif, dan netral [12].
5. Penelitian ini memanfaatkan google *Colab* dan menggunakan bahasa pemrograman *Python* sebagai alat untuk menjalankan analisis dan untuk merancang model Analisis Jaringan Sosial [13].
6. Penelitian ini akan menerapkan kerangka kerja *CRISP-DM* untuk meneliti opini publik tentang Kementerian Keuangan [7], [14].

1.4 Tujuan dan Manfaat Penelitian

1.4.1 Tujuan Penelitian

1. Penelitian ini bertujuan untuk evaluasi performa algotima ML *Naive Bayes*, dan *Support Vector Machines (SVM)* melalui pengukuran *akurasi, precision, recall, dan F1-score* yang dihasilkan dari analisis sentimen publik pada Kementerian Keuangan.
2. Penelitian ini bertujuan untuk mengimplementasikan hasil analisis sentimen yang digunakan untuk menganalisis opini publik pada Kementerian Keuangan.

1.4.2 Manfaat Penelitian

1. Manfaat Praktis

   Penelitian ini diharapkan akan bermanfaat untuk membantu pemahaman yang lebih mendalam terkait dengan opini dan sentimen masyarakat terhadap Kementerian Keuangan. Ini dapat membantu lembaga tersebut memahami persepsi masyarakat dan meresponsnya secara lebih efektif. Pemahaman yang lebih baik tentang opini publik, lembaga dapat mengidentifikasi area yang perlu ditingkatkan dalam keterbukaan dan transparansi.

2. Manfaat Teoritis

   Bagi peneliti, pembaca, dan masyarakat secara umum, diharapkan bahwa penelitian ini akan memberikan kontribusi tambahan dalam memperluas pemahaman dan referensi terkait prediksi opini masyarakat terhadap lembaga pemerintah.

1.5 Sistematika Penulisan

Sistematika penulisan pada penelitian ini terdiri dari:

1. **BAB I PENDAHULUAN**

   Pada bab ini, terdiri dari latar belakang masalah, rumusan masalah, batasan masalah, tujuan dan manfaat penelitian, dan sistematika penulisan.

2. **BAB II LANDASAN TEORI**

   Pada bab ini, berisi penjelasan mengenai teori yang mendukung penelitian, serta terdapat beberapa penelitian terdahulu yang sesuai dengan topik penelitian.

3. **BAB III METODOLOGI PENELITIAN**

   Pada bab ini, berisi penjelasan mengenai metode dan tahapan yang akan dilaksanakan pada penelitian.

4. **BAB IV ANALISIS DAN HASIL PENELITIAN**

   Pada bab ini, menjelaskan analisis yang telah dilakukan, dengan memaparkan kode program dan menjelaskan temuan penelitian.

5. **BAB V SIMPULAN DAN SARAN**

   Pada bab ini, berisi kesimpulan dan saran dari hasil penelitian yang telah dilakukan.

BAB II

LANDASAN TEORI

2.1 Penilitian Terdahalu

Penelitian yang telah dilakukan sebelumnya penting sebagai landasan dalam melaksanakan penelitian yang akan dilakukan. Tabel 2.1 merupakan beberapa penelitian terdahulu.

Tabel 2. 1 Penelitian Terdahulu

1.	Nama Jurnal	Buletin Sistem Informasi dan Teknologi Islam, Vol 4, No 4, (2023) [7]
	Author	A. Anugrah Aqsaa, Irawatia, Lukman Syafieb,
	Metode	Metode penelitian yang digunakan adalah Metode *Naïve Bayes* dan *Support Vector Machine*
	Permasalahan	Permasalahan pada penilitian adanya dugaan transaksi mencurigakan di Kementerian Keuangan (Kemenkeu) yang menjadi sorotan di media sosial, terutama di Twitter
	Hasil dan Kesimpulan	Hasil dan kesimpulan pada penilitian Naïve Bayes mendapatkan nilai akurasi sebesar 71,7%, presisi sebesar 55,2%, recall sebesar 45,3%, dan f1-score sebesar 44,8%, sedangkan pada *SVM* mendapatkan nilai akurasi sebesar 74%, presisi sebesar 87,8%, recall sebesar 49,1%, dan f1-score sebesar 49,8%.
2.	Nama Jurnal	Jurnal Mahasiswa Teknik Informatika, Vol. 8, No. 4 (2024) [8]
	Author	Yulius Bambang Seran, Supatman
	Metode	Penilitian ini menggunakan metode *Support Vector Machine*
	Permasalahan	Permasalahan pada penilitian adanya pro dan kontra terhadap kinerja kerja Presiden Joko Widodo
	Hasil dan Kesimpulan	Hasil dan kesimpulan pada penilitian menunjukkan bahwa model *Support Vector Machine* mencapai

		tingkat akurasi sebesar 66%. Model ini menunjukkan performa bagusdalam identifikasikelaspositif, dengan nilai recall98% dan f1-score77%
3.	Nama Jurnal	Jurnal Pengembangan Teknologi Informasi dan Ilmu Komputer, *Vol. 7, No. 1, (2023)* [9]
	Author	Denny Manuel Yeremia Sinurat, Dian Eka Ratnawati, Dwija Wisnu Brata
	Metode	Penilitian ini menggunakan metode *Naïve Bayes Classifier* dan *SMOTE*
	Permasalahan	Permasalahan pada penilitian ini reaksi masyarakat terhadap kebijakan kenaikan cukai rokok sebesar 10% yang diberlakukan mulai tahun 2023
	Hasil dan Kesimpulan	Hasil penelitian ini menunjukkan akurasi tertinggi sebesar 74%, yang dicapai menggunakan algoritma *Naïve Bayes* dengan data seimbang hasil dari metode *SMOTE.*
4.	Nama Jurnal	Techno.COM, Vol. 21, No. 4, (2022) [15]
	Author	Fefbiansyah Hasibuan, Wowon Priatna, Tyastuti Sri Lestari
	Metode	Pada penelitian metode algoritma klasifikasi *Naive Bayes*
	Permasalahan	Permasalahan pada penilitan ini kelangkaan minyak goreng yang terjadi di Indonesia, yang menyebabkan berbagai opini dari masyarakat di media twitter terkait pera Kementerian Perdagangan Republik
	Hasil dan Kesimpulan	Hasil dan kesimpulan pada penelitian menggunakan algoritma Naïve Bayes menunjukkan nilai akurasi sebesar 84,24%.
5.	Nama Jurnal	Jurnal Pengembangan Teknologi Informasi dan Ilmu Komputer, Vol. 6, No. 5, (2022) [16]
	Author	Jasico Da Comoro Aruan, Bayu Rahayudi, Achmad Ridok
	Metode	Pada penelitian ini menggunakan metode *Support Vector Machine*
	Permasalahan	Permasalahan pada penilitian pentingnya menilai sentimen masyarakat terhadap layanan RSUD untuk meningkatkan mutu pelayanan kesehatan dan mendukung proses akreditasi RSUD.

	Hasil dan Kesimpulan	Hasil dan kesimpulan pada penilitian ini metode *Support Vector Machine (SVM)* menghasilkan nilai akurasi sebesar 88%. Selain itu, nilai recall sebesar 87,5%, precision sebesar 90%, dan f1-score sebesar 87,5%
6.	Nama Jurnal	Jurnal Ilmiah Edutic, Vol.7, No.1, (2020) [10]
	Author	Dedi Darwis, Eka Shintya Pratiwi, A. Ferico Octaviansyah Pasaribu
	Metode	Pada penelitian ini menggunakan metode *Support Vector Machine*.
	Permasalahan	Permasalahan pada penilitan ini memahami opini masyarakat terhadap kinerja Komisi Pemberantasan Korupsi (KPK RI), terutama dalam konteks pemberantasan tindak pidana korupsi.
	Hasil dan Kesimpulan	Hasil dan kesimpulan penelitian ini menunjukkan nilai akurasi sebesar 82% dengan menggunakan algoritma *Support Vector Machine*, di mana sentimen dengan label negatif mendominasi sebesar 77%, diikuti oleh label netral sebesar 25%, dan label positif sebesar 8%.
7.	Nama Jurnal	Jurnal Teknologi Informasi dan Komunikasi, 6(4) (2022) [17]
	Author	Ali Ahmad, Windu Gata
	Metode	Pada penelitian ini menggunakan metode *Support Vector Machine*
	Permasalahan	Permasalahan pada penilitan ini memahami sentimen masyarakat Indonesia terhadap teknologi metaverse, terutama mengingat dampak signifikan yang ditimbulkan oleh perkembangan teknologi
	Hasil dan Kesimpulan	Hasil penelitian ini menunjukkan teknologi metaverse yang menunjukkan 66% bersikap netral, 17% negatif dan 16% positif, sedangkan dari hasil pengujian dengan algorithma *SVM* didapatkan hasil performansi *SVM* sebesar 87% .
8.	Nama Jurnal	Jurnal Informatika dan Teknik Elektro Terapan, Vol. 10 No. 1, (2022) [18]
	Author	Dianati Duei Putri, Gigih Forda Nama, Wahyu Eko Sulistiono
	Metode	Pada penelitian ini menggunakan metode *Naïve Bayes*

	Permasalahan	Permasalahan pada penelitian ini bagaimana masyarakat menyampaikan opini atau sentimen mereka terhadap kinerja Dewan Perwakilan Rakyat (DPR) melalui media social twitter
	Hasil dan Kesimpulan	Hasil penelitian ini menunjukkan bahwa algoritma *naive bayes* mendapatkan accuracy score sebesar 80%
9.	Nama Jurnal	Information Technology and Engineering (2023) [19]
	Author	Raymond Oetama, Yanfi Yanfi, Masagus M. Ikhsan Assiddiq
	Metode	Pada penelitian ini menggunakan algoritma *Support Vector Machine*
	Permasalahan	Permasalahan pada penelitian ini kasus penipuan yang melibatkan platform trading online seperti Binomo, yang sempat populer di Indonesia berkat promosi dari beberapa influencer.
	Hasil dan Kesimpulan	Hasil penelitian ini menunjukkan bahwa algoritma *SVM* memiliki akurasi sebesar 86% untuk data training dan 80% untuk data testing
10.	Nama Jurnal	*JOURNAL OF MULTIDISCIPLINARY ISSUES, 2(2) 1 - 21 (2022)* [20]
	Author	Vinson Phoan, Johan Setiawan
	Metode	Pada penelitian ini menggunakan algoritma *Support Vector Machine*
	Permasalahan	Permasalahan pada penelitian adanya fenomena pelecehan seksual, yang melibatkan tindakan verbal dan nonverbal dengan unsur pemaksaan terhadap korban.
	Hasil dan Kesimpulan	Hasil penelitian menggunkan *Support Vector Machine* menunjukkan bahwa akurasi 55,14% dan data pelecehan seksual yang dikumpulkan pada 16 Maret 2022, dengan 287 data yang diperoleh dari situs Twitter

Pada tabel 2.1 berisi referensi dari studi-studi sebelumnya yang menjadi landasan atau dasar bagi penelitian yang sedang dilakukan. Dalam Analisis sentimen, terdapat beberapa algoritma machine learning yang populer, yaitu *Naive Bayes* dan *Support Vector Machine*. Kedua algoritma tersebut menjadi populer

karena menghasilkan nilai akurasi yang tinggi. Terdapat pada penelitian [15] menghasilkan nilai akurasi *Support Vector Machine* sebesar 88%. Pada penelitian [14] menghasilkan nilai akurasi *Naive Bayes* sebesar 84.24%. Penelitian [11] membahas tentang objek Kementerian keuangan yang menghasilkan nilai akurasi *Support Vector Machine* sebesar 74% dan nilai akurasi *Naive Bayes* sebesar 71.7% dengan pelabelan secara manual. Model *Naïve Bayes* yang menggunkan teknik *SMOTE* [13] menghasilkan nilai akurasi sebesar 74% dengan pelabelan data secara manual.

Fokus penelitian ini adalah pada analisis sentimen mengenai opini publik terhadap pemerintahan khususnya Kementerian Keuangan. Penelitian ini membandingkan dua algoritma yaitu *Support Vector Machine* dan *Naive Bayes*. Dataset yang digunakan mengalami ketidakseimbangan, teknik SMOTE diterapkan untuk menyimbangkan data dan meningkatkan performa model prediksi. Hasil penelitian akan divisualisasikan dalam bentuk *dashboard* pada sebuah *website* yang menampilkan sentimen analisis dari hasil opini dan persepsi masyarakat terhadap Kementerian Keuangan di Indonesia.

2.1 Tinjauan Teori

2.1.1 X

X merupakan *platform* media sosial ini memungkinkan pengguna untuk membagikan pesan singkat yang disebut *"tweet.* Setiap *tweet* dibatasi hingga 280 karakter dan dapat mencakup teks, gambar, video, atau tautan. X digunakan oleh jutaan orang di seluruh dunia untuk menyampaikan pemikiran, berita, opini, dan informasi lainnya secara *real-time*. Analisis sentimen *X* melibatkan mengumpulkan, memproses, dan menganalisis *tweet* untuk menilai apakah opini atau perasaan yang disampaikan dalam *tweet* tersebut termasuk dalam kategori positif, negatif, atau netral terhadap topik tertentu [21].

Opini publik mengacu pada pandangan, pendapat, atau sikap yang dimiliki oleh sekelompok orang dalam masyarakat terhadap suatu topik atau entitas. Dalam penggunaan *X* untuk memprediksi opini publik terhadap Kementerian Keuangan analisis sentimen dilakukan untuk menilai dan memahami respons

atau tanggapan publik yang diungkapkan melalui *tweet* terkait Kementerian Keuangan. Menganalisis *tweet* dapat memperoleh wawasan yang berguna untuk mengukur respons publik dan memprediksi tren opini yang berkaitan dengan kinerja atau kebijakan Kementerian Keuangan.

2.1.2 Analisis Sentimen

Analisis sentimen adalah menganalisis teks digital untuk menentukan apakah pesan tersebut memiliki emosional yang positif, negatif, atau netral Analisis sentimen bermanfaat untuk memahami pendapat yang terkandung dalam ulasan atau komentar yang digunakan oleh pengguna internet. Berbagai penelitian dalam bidang analisis sentimen telah dilakukan oleh sejumlah peneliti sebelumnya dengan tujuan untuk mendapatkan informasi dari suatu kumpulan data mengenai penilaian subjek yang diteliti. [22]. Proses analisis sentimen dengan pengumpulan data teks yang relevan dengan topik yang ingin dianalisis. Data tersebut kemudian diproses untuk menghilangkan noise, seperti tanda baca, kata-kata yang tidak relevan, dan emotikon. Setelah itu, dilakukan analisis sentimen untuk menentukan apakah sentimen dalam teks tersebut bersifat positif, negatif, atau netral.

Teknik yang digunakan bisa berupa analisis kata kunci, analisis statistik, atau menggunakan model *Machine Learning*. Hasil dari analisis sentimen kemudian diinterpretasikan untuk memahami kesimpulan atau tren umum yang terkait dengan topik atau entitas yang dianalisis. Analisis sentimen juga membantu dalam mendeteksi isu-isu yang penting bagi masyarakat, mengukur tingkat kepercayaan publik, serta mengevaluasi efektivitas komunikasi dan strategi pemerintah [23]. Pemahaman yang lebih baik tentang sentimen masyarakat, lembaga atau kementerian dapat merespons secara lebih efektif dan memperbaiki kinerja untuk kepentingan publik.

2.1.3 Text Preprocessing

Preprocessing teks adalah langkah pertama dalam menghilangkan gangguan atau *noise* pada data teks agar dapat diolah lebih lanjut dengan lebih efisien [24]. Proses *preprocessing* teks meliputi serangkaian rutinitas dan langkah untuk menyiapkan data agar dapat digunakan dalam fungsi pengambilan data dari sistem penambangan teks. Tujuan dari langkah-langkah ini adalah untuk melakukan pembersihan dan persiapan data teks agar dapat dijalankan dengan lebih efisien dalam proses analisis atau penambangan informasi, di antaranya [25]:

1. *Case Folding*

   Proses mengubah teks dalam kalimat menjadi huruf kecil dilakukan meskipun terdapat nama kota dan entitas lainnya. Hal ini disebabkan karena ulasan yang diperoleh dari pengumpulan data tidak memiliki format yang seragam seperti ulasan lainnya.

2. *Cleaning*

   Proses menghilangkan teks yang mengandung awalan tertentu, tag *HTML*, tanda baca, serta angka menggunakan ekspresi reguler (regex).

3. *Stop Word Removal*

   Proses penghapusan kata yang tidak relevan dalam kalimat dilakukan berdasarkan daftar *Stop Word* dalam bahasa Inggris atau bahasa Indonesia.

4. *Tokenization*

   Proses membagi teks menjadi unit-unit kecil yang disebut token, seperti kata-kata, frasa, atau kalimat, untuk memudahkan analisis dan pemrosesan data.

5. *Stemming*

   Proses mengubah kata-kata ke dalam bentuk dasarnya untuk mengenali kata-kata yang memiliki arti yang sama.

2.1.4 TF-IDF

TF-IDF (Term Frequency-Inverse Document Frequency) merupakan salah satu teknik yang digunakan dalam pemrosesan bahasa alami dan analisis teks

untuk mengevaluasi pentingnya sebuah kata dalam suatu dokumen atau dalam keseluruhan korpus teks [26]. Metode ini menghitung skor untuk kata-kata berdasarkan dua faktor utama, yaitu seberapa sering kata tersebut muncul dalam dokumen (TF) dan seberapa umum kata tersebut muncul dalam seluruh korpus teks (IDF). Dengan menggabungkan nilai TF dan IDF, skor TF-IDF memberikan informasi yang lebih tepat mengenai pentingnya suatu kata dalam konteks teks tertentu.

2.1.5 Wordcloud

Wordcloud merupakan cara visual untuk menampilkan kata-kata yang sering muncul dalam sebuah teks. Semakin sering kata tersebut muncul, semakin besar ukurannya dalam wordcloud. Visualisasi ini memudahkan melihat kata-kata kunci atau tema utama dari teks secara cepat. Wordcloud sering digunakan dalam analisis teks untuk memberikan gambaran cepat mengenai kata-kata yang dominan atau penting dalam sebuah dokumen, artikel, atau kumpulan data seperti tweet. Alat ini membantu mengidentifikasi tema atau topik utama dari teks secara visual dan mudah dipahami [27].

2.1.6 Scraping

Scraping merupakan proses mengumpulkan data dari sebuah website secara otomatis menggunakan program atau *script*. Tujuannya adalah Untuk mengambil informasi yang ditampilkan di halaman web, seperti teks, gambar, tanpa harus menyalin secara manual. Data ini kemudian bisa digunakan untuk berbagai keperluan, seperti analisis, penelitian, atau pembuatan aplikasi. Scraping biasanya dilakukan dengan bantuan alat atau bahasa pemrograman, seperti Python menggunakan library seperti BeautifulSoup atau Scrapy [28].

2.1.7 SMOTE

SMOTE singkatan dari (*Synthetic Minority Over-sampling Technique*) teknik oversampling dalam *Machine Learning* untuk menangani ketidakseimbangan data. Teknik ini membuat data sintetis baru dari kelas

minoritas dengan memilih titik data minoritas dan menambah titik sintetis di antara tetangga terdekatnya. Hal ini berguna untuk menambah jumlah data pada kelas minoritas serta mencegah saat melatih model pada dataset yang tidak seimbang [29]. *SMOTE* dapat meningkatkan jumlah sampel pada kelas minoritas tanpa melakukan duplikasi data.

2.1.8 Confusion Matrix

Confusion matrix merupakan tabel yang digunakan untuk melihat seberapa baik model klasifikasi dalam memprediksi sesuatu. Tabel ini menunjukkan empat kemungkinan hasil. Pertama, *True Positive (TP)* adalah ketika model benar memprediksi sesuatu sebagai positif. Kedua, *True Negative (TN),* ketika model benar memprediksi sesuatu sebagai negatif. Ketiga, False Positive (FP), saat model salah memprediksi positif padahal sebenarnya negatif. Terakhir, False Negative (FN), ketika model salah memprediksi negatif padahal sebenarnya positif. Confusion matrix bisa memahami kesalahan yang dibuat oleh model dan menghitung seberapa akurat model [30] .

2.1.9 Accuracy

Akurasi merupakan ukuran kinerja model klasifikasi yang menunjukkan seberapa sering model memberikan prediksi yang benar. Akurasi dihitung dengan cara membagi jumlah prediksi yang benar, baik *True Positive* maupun *True Negative,* dengan total prediksi yang dibuat. [31].

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN}$$

Rumus 2. 1 Rumus Akurasi

2.1.10 Precision

Presisi adalah metrik yang digunakan dalam evaluasi model klasifikasi untuk menunjukkan seberapa tepat model saat memprediksi kelas positif. Presisi mengukur berapa banyak prediksi positif yang benar dibandingkan dengan semua prediksi positif yang dibuat oleh model [32].

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

Rumus 2. 2 Rumus Presisi

2.1.11 Recall

Recall merupakan metrik evaluasi untuk model klasifikasi yang menunjukkan seberapa efektif model dalam mengidentifikasi semua contoh positif yang sebenarnya. Recall mengukur berapa banyak kasus positif yang sebenarnya (*True Positive*) berhasil diprediksi dengan benar oleh model dibandingkan dengan seluruh jumlah kasus positif yang ada [33], [34]

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

Rumus 2. 3 Rumus Recall

2.1.12 F-measure

F-measure dikenal sebagai F1-score, adalah metrik yang menggabungkan presisi dan recall menjadi satu angka untuk memberikan gambaran yang seimbang tentang kinerja model klasifikasi. F1-score digunakan Ketika ingin menyeimbangkan antara presisi dan *recall* terutama jika ada ketidakseimbangan antara jumlah kelas positif dan negative [34].

$$F1 = 2 \text{ x } \frac{Precision\ x\ Recall}{Precision + Recall}$$

Rumus 2. 4 Rumus F-Measure

2.2 Algoritma dan *Framework*

2.2.1 *Machine Learning*

Machine Learning merupakan bidang yang terus berkembang, berfokus pada pengenalan pola dan pembelajaran berbasis komputer dalam kecerdasan buatan. *Machine Learning* digunakan berbagai algoritma pembelajaran, baik yang bersifat diawasi (*supervised*) maupun tidak diawasi (*unsupervised*) untuk memprediksi dan mendukung pengambilan

keputusan otomatis berdasarkan sekumpulan data [35]. Tujuan utama *ML* adalah memberikan kemampuan kepada sistem komputer untuk belajar dari pengalaman, memahami pola data, dan meningkatkan kinerjanya seiring waktu tanpa pemrograman eksplisit. *Machine Learning* tidak hanya diinstruksikan untuk mengeksekusi tugas tertentu, tetapi juga diberikan kemampuan untuk belajar dari pengalaman. Proses implementasi *Machine Learning* dapat disusun ke dalam tiga tahap, yaitu persiapan, pemodelan data, dan evaluasi.

1. Persiapan

   Tahap persiapan dimulai dengan pemilahan atribut atau parameter yang akan digunakan dalam pemodelan *Machine Learning*. Proses ini melibatkan pemilihan variabel atau fitur yang signifikan untuk memastikan kontribusi maksimal dalam pembentukan model yang akurat.
2. Pemodelan data

   Tahap pemodelan data adalah representasi matematis dari hubungan antara fitur dan label. Selama tahap pelatihan, model mempelajari parameter dan karakteristik dari data latihan.
3. Evaluasi.

   Tahap evaluasi menggunakan data yang tidak terlihat sebelumnya untuk memastikan bahwa model dapat memberikan prediksi yang akurat dan dapat diandalkan pada situasi dunia nyata.

2.2.2 Naive Bayes

Naive Bayes merupakan metode klasifikasi sederhana yang dapat mengestimasi probabilitas dengan menggabungkan variasi dan frekuensi nilai dari dataset yang tersedia. Algoritma ini menggunakan teorema Bayes untuk memperkirakan probabilitas atribut yang independen satu sama lain, diberikan nilai pada variabel kelas. *Naive Bayes* berdasarkan asumsi sederhana bahwa nilai atribut secara bersyarat saling bebas jika nilai *output* telah diketahui [36]. Nilai *output* sudah diketahui, probabilitas mengamati

bersama-sama adalah hasil kali dari probabilitas individu. Keunggulan *Naive Bayes* terletak pada kebutuhan jumlah data pelatihan yang relatif kecil untuk menentukan estimasi parameter yang diperlukan dalam proses klasifikasi [37]. Metode ini sering memberikan kinerja yang baik dalam kebanyakan situasi dunia nyata yang kompleks dibandingkan dengan ekspektasinya. Prediksi *Bayes* didasarkan pada teorema Bayes sebagai berikut :

$$\mathrm{P}(X) = \frac{P(H)\ X\ P(H)}{P(x)}$$

Rumus 2. 5 Rumus Naive Bayes

Ket:

- X : Data dengan class yang belum diketahui
- H : Hipotesis data merupakan suatu class spesifik
- P(H|X) : Probabilitas hipotesis H berdasar kondisi X (posteriori probabilitas)
- P(H) : Probabilitas hipotesis H (prior probabilitas)
- P(X|H) : Probabilitas X berdasarkan kondisi pada hipotesis H
- P(X) : Probabilitas X

2.2.3 Support Vector Machine (SVM)

SVM (*Support Vector Machine*) adalah suatu algoritma yang dikenal baik karena mampu menghasilkan solusi yang optimal dalam melakukan klasifikasi [38]. Algoritma ini diperkenalkan oleh Vapnik sebagai model *Machine Learning* berbasis kernel yang dapat digunakan untuk klasifikasi dan regresi. *SVM* bekerja dengan membangun *hyperplane* optimal atau batas keputusan dalam ruang fitur, yang memisahkan berbagai kelas data. Pendekatan ini memungkinkan *SVM* untuk memberikan solusi klasifikasi yang baik, terutama dalam konteks data yang tidak linier dan memiliki dimensi tinggi.

Algoritma *Support Vector Machine (SVM)* digunakan untuk menemukan *hyperplane* terbaik dalam ruang N-dimensi yang dengan jelas

memisahkan titik data. *Hyperplane* merupakan suatu fungsi yang berperan sebagai pemisah antara kelas. Fungsi utama dari *Support Vector Machine (SVM)* adalah untuk memisahkan data menjadi dua kelas yang berbeda dengan menggunakan *hyperplane*. *Hyperplane* ini bisa berupa garis atau bidang yang memiliki margin maksimum, yang memisahkan kedua kelas. tersebut secara optimal. Rumus *SVM* untuk klasifikasi sebagai berikut:

$$Largef(x) = sign\left(sum_{i=1}^{n} yi\ alpha_{\ i}\ K(x_{i,}x\right) + b)$$

Rumus 2. 6 Rumus Support Vector Machine

Ket:

- (x): fungsi prediksi
- x: vektor fitur input
- y: label kelas (+1 atau -1)
- α: vektor bobot
- K(xi,x): fungsi kernel yang menghitung jarak antara dua vektor fitur
- b: bias

2.2.4 *CRISP-DM*

CRISP-DM atau *Cross Industry Standard Process* for *Data mining*, adalah suatu standar dalam pemrosesan *data mining* yang telah dikembangkan. Dalam standar ini, data melewati serangkaian fase yang terstruktur dan jelas, mengikuti metodologi yang efisien . Metodologi ini digagas oleh *CRISP-DM* Consortium, sebuah kelompok yang menghasilkan standar industri yang diterima secara luas dalam bidang *data mining* dan analisis data. Metodologi ini terdiri dari enam tahapan yaitu *Business Understanding*, *Data Understanding*, *Data Preparation*, *Modelling*, *Evaluation*, dan *Deployment*. Metodologi ini melibatkan enam tahapan yang dapat diuraikan sebagai berikut [39]:

1. *Business Understanding*

Pada tahap ini, beberapa kegiatan melibatkan pemahaman kebutuhan dan tujuan dari perspektif bisnis. Selanjutnya, pengetahuan diartikan ke dalam bentuk pendefinisian masalah dalam *data mining*, dan kemudian merumuskan rencana serta strategi untuk mencapai tujuan *data mining*.

2. *Data Understanding*

   Tahap ini dimulai dengan pengumpulan data, dilanjutkan dengan deskripsi data, dan mengevaluasi kualitas data. Data dijelajahi untuk memahami karakteristiknya, pola yang mungkin ada, serta masalah atau kekurangan yang perlu diatasi.

3. *Data Preparation*

   Pada tahap ini, langkah-langkah melibatkan pembentukan dataset akhir dari data mentah. Beberapa kegiatan yang dilakukan mencakup pembersihan data, pemilihan data (*Data Selection*) untuk rekaman dan atribut, serta transformasi data (*Data Transformation*). Semua langkah ini bertujuan untuk menyediakan input yang bersih dan relevan untuk proses pemodelan berikutnya.

4. *Modelling*

   Tahap ini melibatkan pemilihan dan penerapan model *data mining* yang sesuai dengan tujuan proyek. Model ini dapat mencakup teknik seperti regresi, klasifikasi, clustering, atau yang lainnya.

5. *Evaluation*

   Tahap ini melibatkan evaluasi kinerja pola yang dihasilkan oleh algoritma. Parameter untuk evaluasi komparatif algoritma mencakup *Confusion Matrix*, yang mengacu pada nilai akurasi, presisi, dan *recall*.

6. *Deployment*

   Setelah model berhasil dievaluasi, langkah terakhir melibatkan implementasi model ke dalam lingkungan produksi. Pada tahap ini, dilakukan pembuatan laporan dan artikel jurnal dengan menggunakan model yang telah dihasilkan.

2.2.5 *SEMMA*

SEMMA merupakan singkatan dari *Sample, Emplore, Modify, Model*, dan *Assess*, sebuah metodologi yang dikembangkan oleh SAS Institute. Metode ini dirancang untuk membantu pengguna dalam melakukan proses *data mining* dengan lebih efisien. Dengan lima tahapan yang jelas yaitu *Sample, Explore, Modify, Model*, dan *Assess*. *SEMMA* memberikan kerangka kerja yang mudah dipahami dan digunakan untuk memprediksi variabel-variabel yang relevan dalam proyek *data mining*. Tahapan-tahapan ini masing-masing memiliki peran uniknya sendiri dalam keseluruhan proses dan memberikan manfaat yang signifikan dalam mengelola proyek *data mining* dengan efektif [40].

2.2.6 *KDD*

Knowledge Discovery in Database Process (KDD) merupakan metode yang digunakan dalam *data mining* untuk menemukan informasi berharga dan pola yang tersembunyi dalam data. Definisi KDD oleh Fayyed et al. (1996) menggambarkan *KDD* sebagai proses menggunakan teknik *data mining* untuk mengidentifikasi pola yang signifikan, melibatkan algoritma untuk mengekstrak pola dari data. Dunham (2003) merangkum proses *KDD* dalam beberapa tahapan, yaitu seleksi data, pra-proses data, transformasi data, *data mining*, dan interpretasi dan evaluasi. Seleksi data melibatkan pemilihan data yang relevan, pra-proses data adalah pembersihan dan integrasi data, transformasi data mengubah format data, *data mining* adalah inti proses *KDD* dengan penggunaan algoritma, dan tahap terakhir adalah interpretasi dan evaluasi hasil untuk memahami temuan dan mengevaluasi [41].

2.3 Software dan *Tools* yang digunakan

2.3.1 *Google Colab*

Colaboratory atau *Colab* merupakan produk yang dikembangkan oleh *Google Research*. *Colab* memungkinkan pengguna untuk menulis dan menjalankan kode Python secara bebas melalui browser, dan sangat cocok untuk keperluan *Machine Learning*, analisis data, serta pembelajaran. Secara teknis, *Colab* adalah layanan notebook *Jupyter* yang dihosting dan dapat digunakan tanpa memerlukan konfigurasi tambahan. Selain itu, *Colab* juga memberikan akses gratis ke sumber daya komputasi, termasuk GPU. [42].

Google *Colab* menyediakan fungsi kolaborasi yang memungkinkan beberapa pengguna untuk bekerja bersama dalam satu notebook secara real-time, memudahkan kerja tim pada proyek analisis sentimen. Integrasi dengan Google Drive juga menyederhanakan manajemen proyek, sementara tersedianya berbagai library *Python* seperti *Pandas, NumPy, Matplotlib,* dan *Tweepy* mempermudah pengguna dalam analisis dan visualisasi data. Dokumentasi yang komprehensif dan tutorial online juga membantu pengguna memahami dan memanfaatkan fitur-fitur Google *Colab* dengan lebih baik, menjadikannya opsi populer untuk analisis sentimen dan pembelajaran mesin lainnya.

2.3.2 *Python*

Python adalah bahasa pemrograman tingkat tinggi yang banyak digunakan di berbagai bidang, seperti pengembangan *web*, analisis data, kecerdasan buatan, dan pengembangan perangkat lunak. Bahasa ini dikenal karena sintaksisnya yang mudah dimengerti dan dipelajari. *Python* menjadi pilihan yang populer bagi para pengembang karena kemudahannya dalam mengelola kode [43]. Dalam konteks analisis sentimen, *Python* sangat berguna karena kemampuannya dalam pengolahan data, analisis statistik, dan pemodelan. Beberapa library yang sering digunakan seperti Pandas untuk manipulasi data, *NLTK* untuk

pemrosesan bahasa alami, dan Scikit-learn untuk pemodelan *Machine Learning*, semuanya dapat diintegrasikan dengan *Python* untuk melakukan analisis sentimen dengan efektif. Dengan menggunakan kombinasi library dan tools ini, para analis dapat membersihkan data, menganalisis sentimen teks, membangun model klasifikasi, dan menghasilkan visualisasi untuk mendukung interpretasi hasil analisis. *Python* memberikan fleksibilitas dan kekuatan yang dibutuhkan untuk mengelola dan menganalisis data teks dengan efisien dalam konteks analisis sentiment.

BAB III

METODOLOGI PENELITIAN

3.1 Gambaran Umum Objek Penelitian

Objek penelitian yang dilakukan dalam penelitian ini adalah analisis sentimen terhadap opini publik tentang Kementerian Keuangan, dengan fokus pada pemahaman dan klasifikasi sentimen masyarakat terhadap lembaga tersebut, yang tercermin dalam platform media sosial X. Kementerian Keuangan, sebagai salah satu lembaga di bawah pengawasan dan bertanggung jawab kepada Presiden Republik Indonesia, memiliki fungsi melaksanakan tugas pemerintahan terkait keuangan negara dan kekayaan negara. Tugas utama Kementerian Keuangan mencakup pelaksanaan urusan pemerintahan di sektor keuangan negara dan pengelolaan kekayaan negara[44]. Dalam konteks ruang lingkup pemerintahan, media sosial seperti X dapat dijadikan sebagai alat penting untuk menyampaikan berbagai kebijakan dan informasi yang dikeluarkan oleh pemerintah. X dipilih karena popularitasnya yang global memungkinkan akses ke beragam opini dari berbagai latar belakang dan wilayah. Platform ini juga memungkinkan informasi disampaikan secara langsung dan real-time, memastikan data yang digunakan selalu terkini [45].

Objek penelitian ini mencakup opini-opini yang dinyatakan oleh masyarakat terkait pandangan, evaluasi terhadap kinerja, dan kebijakan yang diterapkan oleh Kementerian Keuangan. Kementerian Keuangan membutuhkan opini publik karena kebijakan dan tindakan yang diambil oleh Kementerian Keuangan dapat memiliki dampak yang signifikan terhadap masyarakat secara keseluruhan. Opini publik memberikan umpan balik yang penting bagi Kementerian Keuangan untuk memahami pandangan, kebutuhan, dan harapan masyarakat terkait dengan kebijakan fiskal, pengelolaan keuangan publik, dan kegiatan ekonomi lainnya yang berkaitan dengan keuangan negara [46]. Analisis sentimen dilakukan untuk mengeksplorasi sentimen yang muncul, seperti opini yang bersifat positif, negatif, atau netral, terhadap Kementerian Keuangan di *platform* media sosial X. Analisis

sentimen dilakukan untuk mengeksplorasi sentimen yang muncul, seperti opini yang bersifat positif, negatif, atau netral, terhadap Kementerian Keuangan di platform media sosial X. Data opini publik Kemenkeu diperoleh menggunakan keyword kemenkeu dengan rentang waktu Agustus 2023 hingga Agustus 2024. Penggunaan algoritma *Naive Bayes* dan *Support Vector Machines (SVM)* dipilih untuk mengklasifikasikan opini secara otomatis dalam analisis sentimen, memanfaatkan data yang tersedia di X terkait dengan Kementerian Keuangan.

3.2 Metode Penelitian

Metode penelitian mengacu pada pendekatan yang terstruktur dan sistematis yang digunakan oleh peneliti untuk merancang, melaksanakan, dan mengevaluasi suatu studi atau penelitian. Dalam penelitian analisis sentimen opini publik terkait Kementerian Keuangan, pendekatan kualitatif digunakan dengan memanfaatkan data dari media sosial X. Proses pengumpulan data dilakukan melalui *scraping*, untuk mengakses dan mengumpulkan informasi opini masyarakat secara langsung dari *platform* tersebut. Pemilihan metode kualitatif ini bertujuan untuk memberikan pemahaman yang lebih terukur dan statistik terkait sentimen pengguna X terhadap Kementerian Keuangan.

Metode penelitian terdapat beberapa jenis, metode yang paling sering digunakan adalah metode kualitatif dan kualitatif. Metode kualitatif adalah metode yang menggunakan pengamatan, wawancara. Metode kuantitatif adalah metode yang menggunakan analisis statistik terhadap data angka. Penelitian ini menggunakan metode kuantitatif yang dikarenakan menggunakan dataset di ambil dari *social media* X. Dataset diambil menggunakan teknik *scraping* dengan bantuan bahasa pemograman *Python*. Pada penelitian dengan metode kuantitatif, akan digunakan perbandingan antara algoritma *Naive Bayes*, *Support Vector Machines (SVM*) untuk analisis sentiment.

Pada proses *data mining*, terdapat beberapa metode yang sering digunakan, antara lain *Cross Industry Standard Process (CRISP-DM), Knowledge Discovery in Database (KDD) [41], dan SEMMA (Sample, Explore, Modify, Model, and*

Assess)[47]. Setiap metode ini memiliki tahapan-tahapan yang berbeda. Berikut adalah rangkuman tentang tahapan dari masing-masing metode.

Tabel 3. 1 Perbandingan *CRISP-DM, SEMMA, KDD*

Metode	Kelebihan	Kekurangan	Tahapan
CRISP-DM	*CRISP-DM* memiliki struktur yang terorganisir dengan tahapan-tahapan yang jelas namun juga fleksibel, serta panduan yang detail untuk setiap tahapan proses *data mining* memudahkan alur kerja	Memerlukan waktu yang lebih lama untuk menyelesaikan seluruh tahapan proses	*Business Understanding, Data Understanding, Data Preparation, Modeling, Evaluation, Deployment*
SEMMA	Metodologi ini fokus pada langkah-langkah kunci dalam proses *data mining* dan mudah untuk diimplementasikan serta cocok digunakan untuk analisis data yang langsung, tidak terlalu komplek	SEMMA tidak cocok untuk proyek-proyek *data mining* memerlukan analisis yang lebih mendalam atau lebih banyak tahapan yang terstruktur dan tidak terlalu rumit	*Sample, Explore, Modify, Model, Assesment*
KDD	Metodologi ini memungkinkan untuk pemahaman yang lebih mendalam terhadap data dan masalah yang dihadapi karena pemahaman data dan seleksi data yang cermat	KDD seringkali memakan waktu dan memerlukan sumber daya yang cukup besar terutama pada tahap pemrosesan data serta model yang dihasilkan kompleks untuk diinterpretasikan	*Pre KDD, Selection, Pro processing, Transformation, Data mining, Interpretation/Evaluation, Post KDD*

Pada tabel 3.1 dari perbandingan *CRISP-DM, SEMMA, KDD,* menggunakan *CRISP-DM*. Metodologi *CRISP-DM* memberikan panduan langkah-demi-langkah yang mudah dipahami dan terstruktur dalam melakukan proses *data mining*. Struktur yang terorganisir dapat mengikuti alur kerja yang sistematis dari pemahaman masalah hingga evaluasi hasil. Hal ini membantu mengelola proyek-proyek *data mining* dengan lebih efisien dan memastikan tidak ada tahapan yang terlewatkan. Teknik *CRISP-DM* (*Cross Industry Standard Process for Data mining*) sebagai metodologi. *CRISP-DM* ini terdiri dari enam tahap utama yang disusun secara sistematis untuk membimbing penelitian atau proyek *data mining*. Tahap-tahap tersebut meliputi Pemahaman Bisnis (*Business Understanding*), Pemahaman Data (*Data Understanding*), Persiapan Data (*Data Preparation*), Pembuatan Model (*Modeling*), Evaluasi (*Evaluation*), dan Penyebaran (*Deployment*). *CRISP-DM* memberikan kerangka kerja yang terstruktur untuk memahami masalah bisnis, mengumpulkan dan mempersiapkan data, mengembangkan model, mengevaluasi hasil, dan mengimplementasikan solusi.

3.3 Variabel Penelitian

Penelitian ini memanfaatkan dataset yang diperoleh dari *platform* media sosial X. Dataset ini dihasilkan melalui proses *scraping* data. Dalam kerangka variabel penelitian, terdapat dua jenis variabel, yaitu variabel independen dan variabel dependen.

3.3.1 Variabel Independen

Variabel independen adalah variabel yang tidak terikat atau bersifat bebas, mampu memberikan pengaruh pada variabel lainnya. Dalam konteks penelitian ini, variabel independen terutama terfokus pada variabel "full_text". Variabel ini menjadi fokus dalam menganalisis dampaknya terhadap variabel lainnya.

3.3.2 Variabel Dependen

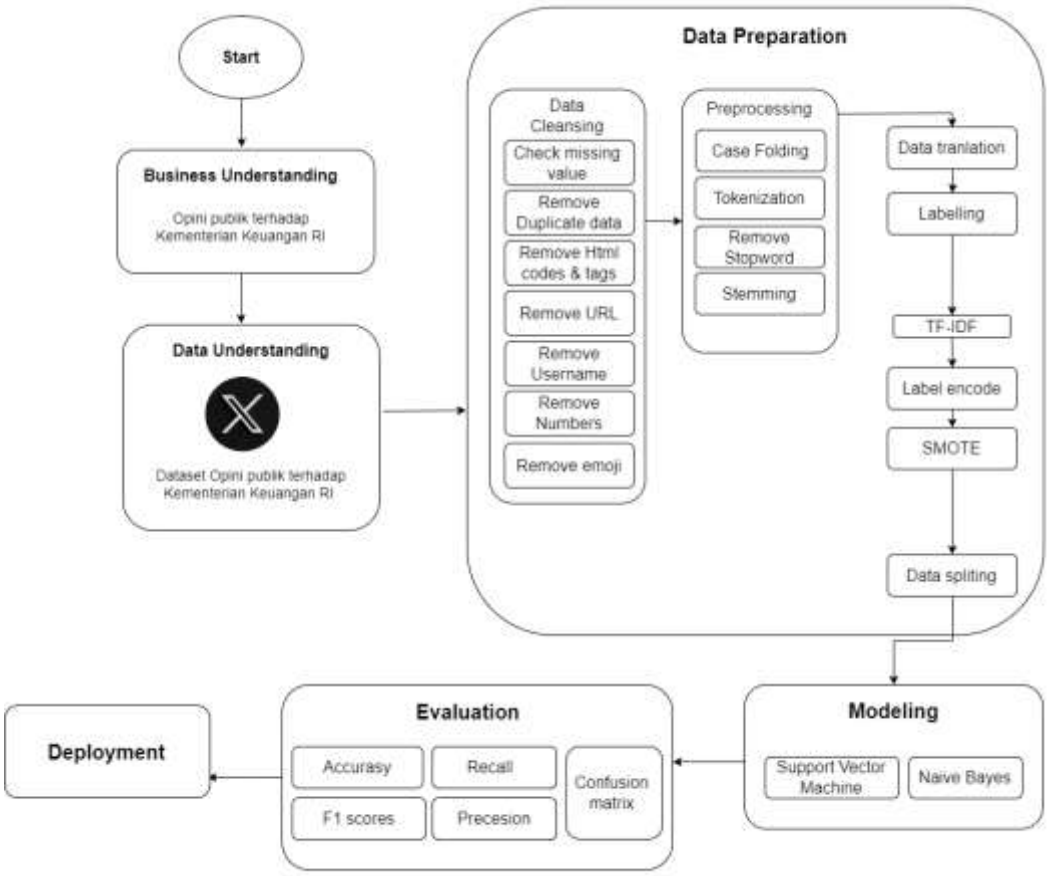
Variabel dependen merupakan variabel yang bersifat terikat atau dipengaruhi oleh variabel independen. Dalam penelitian ini, variabel dependen terfokus pada variabel "label" yang memiliki kategori positif, negatif, dan netral. Variabel ini menjadi objek penelitian untuk mengidentifikasi bagaimana variabel independen, yaitu " *full_text* ", mempengaruhi atau berhubungan dengan label-label tersebut.

3.3 Teknik Pengumpulan Data

Teknik pengumpulan data mengacu pada cara atau metode yang digunakan untuk mengumpulkan informasi atau data dalam sebuah penelitian atau studi. Data primer merujuk pada informasi yang dikumpulkan secara langsung dari sumber aslinya untuk keperluan penelitian tertentu. Data sekunder adalah informasi yang telah dikumpulkan oleh pihak lain untuk tujuan yang berbeda dari penelitian yang sedang dilakukan. Data tersier adalah data yang telah diolah atau diinterpretasikan oleh pihak ketiga, yang kemudian disajikan dalam bentuk yang siap digunakan oleh peneliti. Pada penelitian ini, teknik pengumpulan data dilakukan secara primer yang diperoleh melalui *scraping* data dari media sosial X dengan data teks yang dikumpulkan dari bulan Oktober 2023 hingga Februari 2024. Dataset ini berisi opini publik mengenai pandangan masyarakat terhadap Kementerian Keuangan.

3.4 Teknik Analisis Data

Dalam penelitian ini, akan menerapkan pendekatan model *CRISP-DM* (*Cross Industry Standard Process for Data mining*). Berdasarkan gambar 3.1 menampilkan enam tahap dalam *CRISP-DM* akan diterapkan dalam penelitian ini, dengan penjelasan sebagai berikut [46].



Gambar 3. 1 Teknik Analisis Data

3.4.2 *Business Understanding*

Tahapan *Business Understanding* adalah langkah awal dalam metodologi *CRISP-DM* (*Cross-Industry Standard Process for Data mining*). Pada penelitian ini tujuan utama adalah untuk analisis sentimen terhadap opini publik tentang pandangan masyarakat terhadap Kementerian Keuangan dengan pemilihan algoritma *Naive Bayes* dan *Support Vector Machine*. Pemahaman terhadap perspektif masyarakat mengenai pemerintahan, kebijakan keuangan, dan peran Kementerian Keuangan. Analisis sentimen secara spesifik, seperti mengidentifikasi pola sentimen positif, negatif, atau netral, terkait pandangan masyarakat terhadap Kemenkeu.

3.4.3 *Data Understanding*

Pada langkah ini, penelitian ini melakukan pengambilan data dengan menggunakan kata kunci terkait Kementerian Keuangan dan Kemenkeu. Dataset yang dihasilkan mencakup sekitar 2354 data, menjadi sumber informasi yang signifikan untuk dilibatkan dalam analisis sentimen terhadap opini publik terkait Kementerian Keuangan. Data yang digunakan berasal dari *platform* media sosial X dan diperoleh melalui proses *scraping* dengan Google collab.

3.4.4 *Data Preparation*

Pada tahapan ini, dilaksanakan proses pembersihan data yang dimulai dengan cleansing data dan tahap *text preprocessing*. Pada tahap *text preprocessing*, beberapa proses dilakukan, termasuk *case folding, tokenizing, stemming,* dan *stopword*. Proses ini bertujuan untuk memastikan data yang digunakan dalam analisis lebih terstruktur dan bersih, memungkinkan hasil yang lebih akurat dalam langkah-langkah analisis selanjutnya. Tahapan *data preparation* terdiri dari *data cleansing, preprocessing, data translation, labeling, TF-IDF, label encode, SMOTE, data splitting.*

3.4.4.1 *Data Cleansing*

Data cleansing merupakan proses membersihkan data dari kesalahan, duplikasi, atau informasi yang tidak lengkap, untuk memastikan data yang digunakan akurat, konsisten, dan siap untuk dianalisis atau diproses lebih lanjut. Tujuannya adalah meningkatkan kualitas data sehingga analisis atau keputusan yang dibuat berdasarkan data tersebut menjadi lebih valid. Berikut merupakan tahapan data cleansing yang digunakan pada penelitian ini *check missing value, remove duplicate data, remove html codes & tags, remove URL, remove username, remove numbers, remove emoji.*

3.4.4.2 *Preprocessing*

Tahapan preprocessing merupakan proses persiapan data mentah, khususnya data teks, sebelum dilakukan analisis atau digunakan dalam model machine learning. Langkah ini bertujuan untuk membersihkan dan menyederhanakan data agar lebih mudah dipahami oleh algoritma. Preprocessing penting untuk meningkatkan kualitas data dan hasil analisis.

1. Case Folding: Mengubah semua huruf dalam teks menjadi huruf kecil (lowercase) agar konsisten, sehingga perbedaan huruf besar dan kecil tidak mempengaruhi analisis. Misalnya, "Kemenkeu" dan "kemenkeu" dianggap sama.
2. Tokenization: Memecah teks menjadi unit-unit kecil seperti kata atau frasa. Misalnya, kalimat "Kemenkeu mengelola anggaran" akan dipecah menjadi ["Kemenkeu", "mengelola", "anggaran"].
3. Remove Stopwords: Menghilangkan kata-kata umum yang tidak memiliki makna penting dalam analisis, seperti "dan", "yang", "atau". Tujuannya untuk fokus pada kata-kata kunci yang lebih bermakna.
4. Stemming: Mengubah kata-kata menjadi bentuk dasarnya. Misalnya, kata "mengelola" akan diubah menjadi "kelola", sehingga berbagai bentuk kata dianggap sebagai satu entitas.

3.4.4.3 *Data Translation*

Data translation adalah proses mengubah data yang telah dikumpulkan dan dibersihkan menjadi bahasa inggris dengan bantuan *library deep translator*. Data yang di *translate* adalah data yang telah dibersihkan dan diproses sebelumnya. Tahap ini penting untuk mempersiapkan data sebelum pelabelan menggunakan *library VADER*.

3.4.4.4 *Labeling*

Labeling merupakan proses memberi label atau kategori pada data berdasarkan kriteria tertentu. Analisis sentimen, labeling biasanya dilakukan untuk mengklasifikasikan teks, seperti tweet atau ulasan, ke

dalam kategori seperti positif, negatif, atau netral. Proses ini penting untuk memudahkan analisis lebih lanjut dan pengembangan model pembelajaran mesin. Pada proses pelabelan data, menggunakan *library VADER (Valence Aware Dictionary Sentiment Reasoner)*. *VADER* menilai setiap teks dan memberikan skor positif, negatif, atau netral. Skor-skor ini kemudian dijumlahkan untuk menghasilkan nilai skor komposit. *Compound score* ukuran yang mempertimbangkan semua skor yang dinormalisasi dalam rentang -1 hingga +1. Nilai komposit di atas 0,05 dianggap sebagai sentimen positif, sedangkan nilai di bawah -0,05 dianggap negatif. Jika nilai berada di antara -0,05 dan 0,05, maka dikategorikan sebagai netral. Pada tabel 3.2 perbandingan yang bisa menjelaskan perbedaan dan kegunaan utama VADER dan SMOTE dalam konteks analisis sentimen:

Tabel 3. 2 Perbandingan Vader danSMOTE

Aspek	**VADER**	**SMOTE**
Jenis Analisis	Sentimen berbasis **lexicon** (daftar kata)	Teknik **resampling** untuk penyeimbangan data yang tidak seimbang dalam klasifikasi
Penelitian	Digunakan untuk **menganalisis sentimen tweet** terkait Kemenkeu (positif, negatif, netral)	Membantu dalam **penyeimbangan data** untuk melatih model SVM dan Naive Bayes pada data tweet
Kata Kunci Utama	"Positive," "Negative," "Neutral," "Compound Score"	"Oversampling," "Synthetic Minority," "Data Imbalance"
Pendekatan	**Lexicon-based**: Menilai sentimen teks pendek seperti tweet	**Data Resampling**: Menyeimbangkan dataset agar model tidak bias ke kelas mayoritas

3.4.4.5 *TF-IDF*

TF-IDF (Term Frequency-Inverse Document Frequency) adalah metode yang digunakan dalam pengolahan teks untuk menilai seberapa penting suatu kata dalam sebuah dokumen dalam konteks koleksi dokumen lainnya. Metode ini memberikan bobot pada setiap kata dan menghitung nilai invers berdasarkan kemunculannya dalam kalimat. *TF-IDF* bertujuan untuk mengubah teks menjadi vektor numerik dengan mempertimbangkan frekuensi kata dalam dokumen serta frekuensi kemunculannya di seluruh kumpulan data.

3.4.4.6 *Label Encode*

Label Encoding adalah metode pengkodean yang mengubah label kategori menjadi format numerik. Dalam hal ini, label sentimen seperti positif, negatif, dan netral diubah menjadi angka, yaitu ‘negatif: 0’, ‘netral: 1’, dan ‘positif: 2’. Proses ini dilakukan setelah tahap TF-IDF untuk mempermudah model dalam memproses data numerik.

3.4.4.7 *SMOTE*

SMOTE (Synthetic Minority Over-sampling Technique) merupakan teknik yang digunakan untuk mengatasi masalah ketidakseimbangan kelas dalam dataset. *SMOTE* membantu meningkatkan performa model dalam mengklasifikasikan data, sehingga model menjadi lebih baik dalam mengenali pola dari kelas yang kurang. Dataset yang telah dilabeli menggunakan *VADER*, langkah selanjutnya adalah menyesuaikan parameter untuk memastikan jumlah sampel sintetis yang dihasilkan sebanding dengan kelas mayoritas. Berikut merupakan kelebihan dan kekurangan menggunakan Teknik *SMOTE, ROS* dan *Data augmentation* [48], [49]:

Tabel 3. 3 Kelebihan dan kekurangan SMOTE, ROS, Data augmentation

Teknik	Kelebihan	Kekurangan
SMOTE	- Menghasilkan sampel sintetis yang menambah variasi dalam kelas minoritas. - Cocok untuk data numerik dan bermanfaat dalam klasifikasi.	- Berisiko overfitting jika sampel sintetis terlalu mirip dengan sampel asli.
ROS	- Dapat digunakan untuk semua jenis data tanpa modifikasi tambahan. - Mudah diterapkan dan tidak membutuhkan komputasi berat.	- Risiko overfitting tinggi karena hanya menyalin data minoritas secara acak. - Tidak menambah variasi dalam kelas minoritas.
Data augmentation	- Meningkatkan ukuran data secara signifikan untuk memperkaya variasi - Cocok untuk data visual, teks, dan suara.	- Berisiko menambahkan noise atau distorsi yang tidak diinginkan pada data. - Memerlukan metode khusus untuk setiap tipe data (misalnya, rotasi gambar, penggeseran teks).

SMOTE digunakan karana menghasilkan sampel sintetis melalui interpolasi antara titik data yang ada, menambah variasi dan kedalaman pada kelas minoritas. Sementara ROS hanya menyalin data yang sudah ada, berisiko menyebabkan model terjebak pada pola yang sama dan meningkatkan kemungkinan *overfitting*. Selain itu, meskipun data augmentation efektif untuk data visual dan teks, teknik ini memerlukan pemahaman mendalam tentang transformasi yang relevan, yang mungkin tidak selalu mudah diterapkan pada semua jenis data. Dengan demikian, *SMOTE* terbukti lebih efektif dalam meningkatkan kinerja model pada dataset yang tidak seimbang, membantu model mengenali pola dalam kelas minoritas, dan meningkatkan metrik evaluasi seperti akurasi dan *recall*.

3.4.4.8 *Data Splitting*

Data splitting merupakan proses membagi dataset menjadi dua atau lebih subset untuk tujuan pelatihan dan pengujian model. Tujuan utama dari pemisahan ini adalah untuk mengevaluasi seberapa baik model yang dibangun dapat memprediksi data yang tidak terlihat sebelumnya. Rasio pembagian data, seperti 90:10, 80:20, atau 70:30, angka pertama menunjukkan persentase data yang digunakan untuk pelatihan (training set), sementara angka kedua menunjukkan persentase data yang digunakan untuk pengujian (test set). Hasil penelitian tersebut menunjukkan bahwa rasio 80:20 menghasilkan tingkat akurasi yang tinggi

3.4.5 Modeling

Pada langkah ini, akan dilakukan analisis terhadap pemodelan menggunakan algoritma klasifikasi yang telah dipilih yaitu *Naive Bayes* dan *Support Vector Machines (SVM)*. Pemilihan algoritma-algoritma ini berdasarkan hasil penelitian terdahulu yang telah menggunakan *Naive Bayes* dan *SVM* dalam analisis sentimen. Proses pemodelan akan menggunakan bahasa pemrograman *Python* untuk mengimplementasikan algoritma-algoritma tersebut secara efektif.

3.4.6 *Evaluation*

Pada langkah ini, akan dilakukan evaluasi kinerja dari model yang telah dibuat. Evaluasi ini mencakup penyajian *akurasi, precision, recall, dan F1-score* dan *confusion matrix* yang telah dihasilkan oleh algoritma *Naive Bayes,* dan *Support Vector Machines (SVM)* serta hasil dari pengujian model.

3.4.7 *Deployment*

Langkah terakhir dalam proses ini adalah menerapkan hasil yang telah didapatkan. Dalam penelitian ini, tahap implementasi akan mencakup pembuatan sebuah *dashboard* yang akan menampilkan visualisasi data terkait analisis sentimen terhadap pendapat publik tentang Kementerian Keuangan. *Dashboard* ini akan memberikan gambaran yang jelas dan mudah dipahami mengenai pola opini yang

beredar di masyarakat terkait lembaga tersebut. Demikian, informasi yang disajikan dapat digunakan sebagai dasar untuk mengambil keputusan yang lebih baik di masa depan.

BAB IV

ANALISIS DAN HASIL PENELITIAN

Penelitian ini menerapkan kerangka kerja *CRISP-DM (Cross Industry Standard Process for Data mining)* sebagai kerangka kerja yang digunakan untuk memandu proses analisis data, mulai dari pemahaman bisnis hingga evaluasi hasil. Dalam opini publik terhadap Kementerian Keuangan, kerangka kerja ini akan membantu dalam mengelola data yang diperoleh dari media sosial seperti X. Dengan penggunaan *CRISP-DM* akan memungkinkan untuk sistematis menganalisis sentimen dan dinamika opini publik terhadap kebijakan dan keputusan Kementerian Keuangan.

4.1 Business Understanding

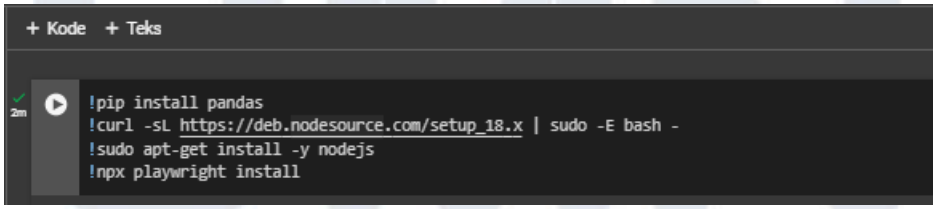
Tahap pertama dalam *CRISP-DM*, yakni "*Business Understanding*" atau Pemahaman Bisnis, sangat penting. Pada tahap awal ini, diperlukan pemahaman yang lebih mendalam terhadap sentimen masyarakat terhadap kebijakan tertentu atau evaluasi terhadap persepsi publik mengenai kinerja Kementerian Keuangan. Kementerian Keuangan perlu memahami opini publik karena ini memberikan wawasan berharga untuk merancang kebijakan yang lebih responsif. Selain itu, penting bagi Kementerian Keuangan agar setiap kebijakan publik dapat dipahami sepenuhnya oleh masyarakat.

Memahami secara mendalam bagaimana masyarakat merespons kebijakan atau program tertentu, Kementerian Keuangan dapat mengidentifikasi pola-pola dalam respons publik yang dapat membantu menyesuaikan strategi komunikasi. Kementerian dapat menyesuaikan komunikasi untuk lebih menekankan manfaat yang dirasakan oleh masyarakat atau menjelaskan dengan lebih baik tujuan dan alasan di balik kebijakan tersebut. Hal ini memungkinkan Kementerian untuk membangun komunikasi yang lebih efektif dan membuat kebijakan yang lebih diterima oleh publik. Oleh karena itu, dalam penilitian ini dilakukan analisis

sentiment terkait opini publik terhadap Kementerian Keuangan dengan melakukan analisis sentimen menggunakan algoritma *Naive Bayes*, dan *Support Vector Machines (SVM).*

4.2 Data Understanding

Pada tahap *Data Understanding*, dilakukan pemahaman terhadap data yang akan digunakan dalam penelitian. Penelitian ini menggunakan data yang diperoleh dari media sosial X dengan rentang waktu dari bulan Agustus 2023 hingga Agustus 2024. Rentang waktu ini dipilih berdasarkan periode yang dianggap signifikan dalam mengamati opini publik terkait Kementerian Keuangan pada saat itu. Data yang digunakan dihasilkan melalui proses *scraping* data dari media sosial X dengan menggunakan bahasa pemrograman *Python* dan Google *Colab*. Proses *scraping* data ini dilakukan untuk mengumpulkan sejumlah besar data *tweet* yang berkaitan dengan topik opini publik tentang Kementerian Keuangan selama periode yang ditentukan. Data yang diperoleh melalui *scraping* mencakup teks *tweet*, metadata seperti tanggal posting, jumlah re*tweet* dan like, serta informasi pengguna yang dapat memberikan konteks lebih lanjut terkait dengan opini yang disampaikan. Dengan memanfaatkan kombinasi bahasa pemrograman *Python* dan Google *Colab* proses pengumpulan dan pemrosesan data.



Gambar 4. 1 Menginstal Library Pandas dan Node.js

Pada gambar 4.1 merupakan *code* perintah yang digunakan untuk menginstal beberapa paket dan alat yang diperlukan dalam proses pengolahan data dan *scraping* menggunakan *Python* dan alat pengujian otomatis bernama Playwright.

1. !pip install pandas

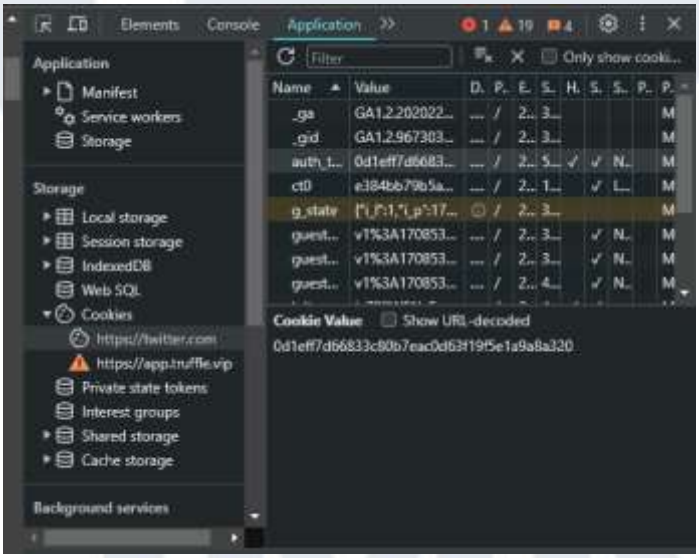
Perintah ini menginstal library Pandas untuk pengolahan data yang akan digunakan dalam pengolahan data nanti. Pandas adalah salah satu library *Python* yang sangat populer digunakan untuk analisis data.

2. !c*URL* -sL https://deb.nodesource.com/setup_18.x | sudo -E bash -:
   Perintah ini digunakan untuk mengunduh dan menginstal Node.js versi 18.x. Node.js digunakan untuk menjalankan JavaScript di luar browser, dan sering digunakan dalam pengembangan *web* dan aplikasi.
3. !sudo apt-get install -y nodejs:
   Perintah ini menginstal Node.js setelah proses unduhnya selesai. "-y" digunakan untuk memberikan persetujuan secara otomatis pada saat instalasi.
4. !npx playwright install:
   Perintah untuk menginstal Playwright, sebuah alat pengujian otomatis yang memungkinkan melakukan interaksi otomatis dengan browser. Playwright sangat berguna dalam skenario pengujian *web* dan pengambilan data dari *web* (*scraping*).

Gambar 4. 2 Crawl Data

Pada gambar 4.2 merupakan pengambilan data dari X menggunakan tweet-harvest dilakukan melalui perintah dalam terminal atau lingkungan *Python* yang mendukung eksekusi perintah shell. Alat tweet-harvest memungkinkan pengguna untuk melakukan *scraping* atau pengambilan data secara spesifik berdasarkan kata kunci yang diinginkan dari *platform* X. Data yang diperoleh dari X menggunakan

keyword "kemenkeu,". Pemilihan *keyword* 'Kemenkeu' bertujuan menjaga fokus analisis pada sentimen publik yang spesifik terhadap Kementerian Keuangan dan kebijakannya, sehingga data lebih relevan dengan tujuan penelitian. Dengan menggunakan *keyword* tersebut, data yang diambil melalui tweet-harvest akan lebih terfokus dan memberikan informasi yang lebih bermakna terkait dengan pekerjaan dan kebijakan yang ada di Kementerian Keuangan. Setelah menjalankan perintah untuk menggunakan paket tweet-harvest. Tahapan berikutnya adalah memasukkan token otentikasi untuk login. Token ini harus dimasukkan dari akun yang sudah masuk ke X dengan cara login ke akun tersebut melalui browser, seperti yang terlihat pada gambar.



Gambar 4. 3 Token Otentikasi

Setelah berhasil mengambil data, data akan disimpan dalam format CSV di folder komputer seperti yang ditunjukkan dalam Gambar 4.4 Data yang diperoleh dari proses scraping mencakup jumlah sebanyak 2.354 data dengan rentang waktu dari Agustus 2023 hingga Agustus 2024.

full_text	id_str	image_url	ply_to_screen_	lang	location	quote_count	reply_count	retweet_count	tweet_url	user_id_str	username
Tetap jaga pak a	174296395	https://pbs.twimg.com/ext_tw_vi		in		153	362	914	https://twitter.cor	1614451668170	AbunawasRetun
Telnyata ini Orat	174296264	https://pbs.twimg.com/ext_tw_vi		in	Jabodetab	7	46	61	https://twitter.cor	1667233216091	Andria75777
Jangan Terkesa	174298787	https://pbs.twimg.com/ext_tw_vi		in	Jabodetab	3	3	119	https://twitter.cor	1667233216091	Andria75777
Semoga ALLAH	174295018	https://pbs.twimg.com/ext_tw_vi		in	NANGROE	0	17	84	https://twitter.cor	1344480749954	CutSanna5
Ada pesan dari	174297714	https://pbs.twimg.com/ext_tw_vi		in		4	12	233	https://twitter.cor	1603593688005	Malika6027
Abah Anies Sen	174305486	https://pbs.twimg.com/ext_tw_vi		in		2	11	147	https://twitter.cor	1605780365866	Putri96960977

Gambar 4. 4 Data Scraping X

4.3 Data Preparation

4.3.1 Check Missing Value

Pemeriksaan nilai yang hilang (missing value) dalam data yang baru diambil untuk analisis sangat penting karena berbagai alasan yang berdampak langsung pada kualitas analisis dan hasil yang dihasilkan. Salah satu alasan adalah untuk memastikan konsistensi dan integritas data yang digunakan dalam analisis. Ketika ada nilai yang hilang, ini dapat mengganggu integritas data dan menyebabkan kesimpulan yang bias atau tidak akurat dalam analisis. Selain itu, kualitas model analisis juga tergantung pada data yang lengkap dan representatif.

```
[ ]  # Check Missing Value

     dataset.isnull().sum()

     conversation_id_str          0
     created_at                   0
     favorite_count               0
     full_text                    0
     id_str                       0
     image_url                 1911
     in_reply_to_screen_name   1936
     lang                         0
     location                  1029
     quote_count                  0
     reply_count                  0
     retweet_count                0
     tweet_url                    0
     user_id_str                  0
     username                     0
     dtype: int64
```

Gambar 4. 5 Missing value

Pada dataset yang ditampilkan dalam gambar 4.5 dapat diperhatikan bahwa terdapat kekosongan data pada kolom seperti image_*URL*, in_reply_to_screen_name, dan location. Data kosong tidak memberikan kontribusi yang signifikan dalam analisis data yang ingin di buat. Oleh

karena itu, data yang kosong tersebut dihapus dari dataset. Penghapusan data kosong ini bertujuan untuk membersihkan dataset dari elemen yang tidak diperlukan untuk meningkatkan kualitas dataset secara keseluruhan. Menghilangkan data yang tidak digunakan seperti pada gambar 4.6 agar hasil analisis yang dihasilkan akan lebih akurat dan dapat diandalkan. Pada gambar. Pada gambar 4.7 adalah hasil output dari kolom yang telah dihapus.

```
# Menghapus kolom dataset
dataset.drop(['image_url', 'in_reply_to_screen_name','id_str','location' ], axis=1, inplace=True)
```

Gambar 4. 6 Code hapus kolom dataset

```
#  informasi tentang dataset
dataset.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 1129 entries, 0 to 1128
Data columns (total 11 columns):
 #   Column              Non-Null Count  Dtype  
---  ------              --------------  -----  
 0   conversation_id_str  1129 non-null   float64
 1   created_at          1129 non-null   object 
 2   favorite_count      1129 non-null   float64
 3   full_text           1129 non-null   object 
 4   lang                1129 non-null   object 
 5   quote_count         1129 non-null   float64
 6   reply_count         1129 non-null   float64
 7   retweet_count       1129 non-null   float64
 8   tweet_url           1129 non-null   object 
 9   user_id_str         1129 non-null   float64
 10  username            1129 non-null   object 
dtypes: float64(6), object(5)
memory usage: 97.1+ KB
```

Gambar 4. 7 kolom yang telah dihapus

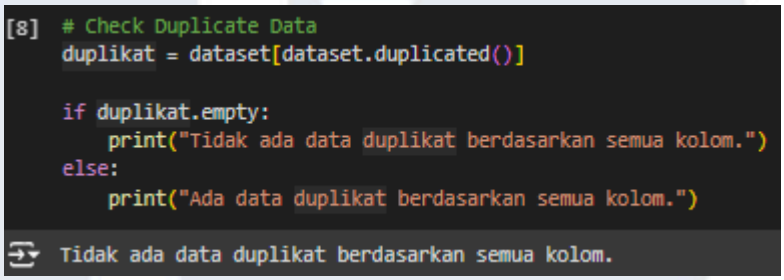
4.3.2 *Removing Duplicate Data*

```
# Check Duplicate Data
duplikat = dataset[dataset.duplicated()]

if duplikat.empty:
    print("Tidak ada data duplikat berdasarkan semua kolom.")
else:
    print("Ada data duplikat berdasarkan semua kolom.")
```

```
Ada data duplikat berdasarkan semua kolom.
```

Gambar 4. 8 Code Duplicate Data

Pada gambar 4.8 merupakan perintah untuk menjalankan check duplicate data. Pada tahap ini bagian penting untuk membersihkan dataset dari entri yang identik atau duplikat. Proses penghapusan data duplikat ini membantu memastikan integritas dan konsistensi dataset sebelum melakukan analisis lebih lanjut. Pada gambar di atas terlihat terdapat data yang duplikat.

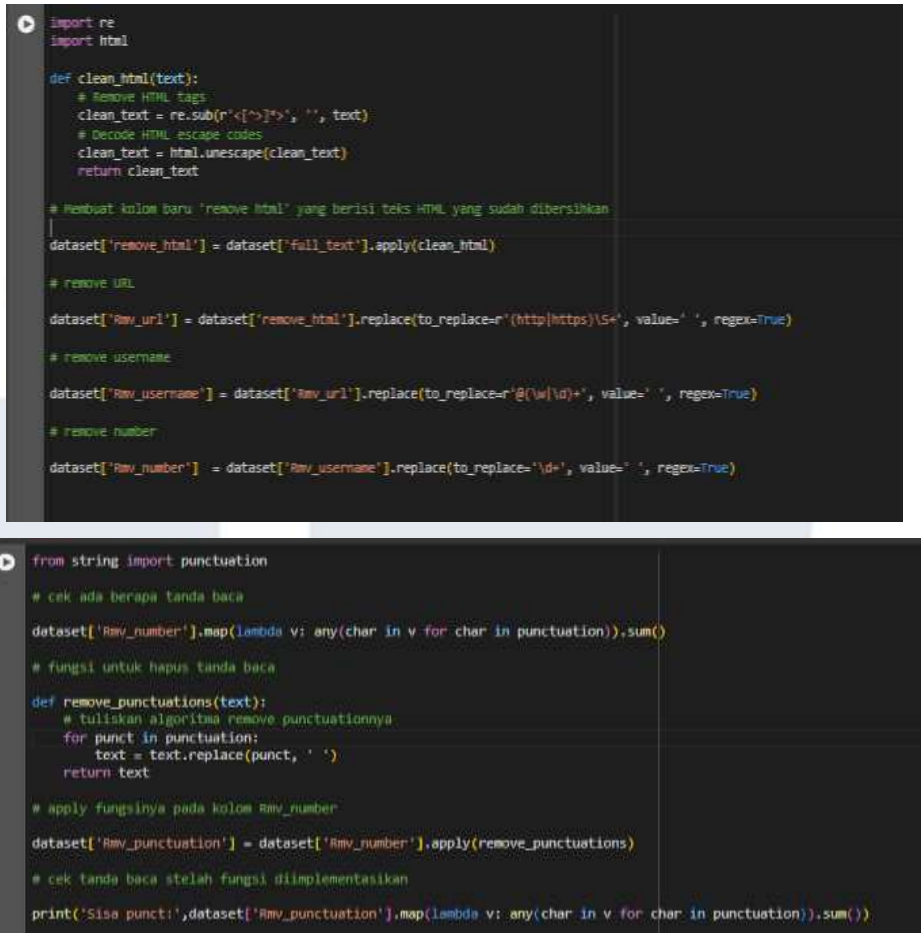


Gambar 4. 9 Ouput data yang tidak ada duplikat dari semua kolom

Gambar 4.9 menunjukkan output data yang sudah diproses sehingga “Tidak ada duplikat berdasarkan semua kolom”. Setiap baris dalam data tersebut tidak ada nilai yang sama di antara baris-baris yang ada. Proses ini untuk memastikan bahwa analisis selanjutnya tidak terpengaruh oleh data yang berulang, yang dapat menyebabkan hasil yang tidak akurat.

4.3.3 Data Cleansing

Data cleansing adalah tahapan yang sangat krusial dalam proses analisis data. Tujuannya adalah untuk membersihkan data dari segala jenis kesalahan, ketidakakuratan, atau format yang tidak sesuai. Dalam penelitian ini, data cleansing menghapus *HTML code* & tag, *URL*, username, angka, tanda baca, emoji, seperti pada gambar 4.10 langkah-langkah ini bertujuan untuk memastikan bahwa data yang digunakan dalam analisis benar-benar bersih dan dapat dipercaya.



```
import re
import html

def clean_html(text):
    # Remove HTML tags
    clean_text = re.sub(r'<[^>]*>', '', text)
    # Decode HTML escape codes
    clean_text = html.unescape(clean_text)
    return clean_text

# Membuat kolom baru 'remove html' yang berisi teks HTML yang sudah dibersihkan

dataset['remove_html'] = dataset['full_text'].apply(clean_html)

# remove URL

dataset['Rmv_url'] = dataset['remove_html'].replace(to_replace=r'(http|https)\S+', value=' ', regex=True)

# remove username

dataset['Rmv_username'] = dataset['Rmv_url'].replace(to_replace=r'@[\w|\d]+', value=' ', regex=True)

# remove number

dataset['Rmv_number']  = dataset['Rmv_username'].replace(to_replace='\d+', value=' ', regex=True)
```

```
from string import punctuation

# cek ada berapa tanda baca

dataset['Rmv_number'].map(lambda v: any(char in v for char in punctuation)).sum()

# fungsi untuk hapus tanda baca

def remove_punctuations(text):
    # tuliskan algoritma remove punctuationnya
    for punct in punctuation:
        text = text.replace(punct, ' ')
    return text

# apply fungsinya pada kolom Rmv_number

dataset['Rmv_punctuation'] = dataset['Rmv_number'].apply(remove_punctuations)

# cek tanda baca stelah fungsi diimplementasikan

print('Sisa punct:',dataset['Rmv_punctuation'].map(lambda v: any(char in v for char in punctuation)).sum())
```

Gambar 4. 10 Code data cleansing

Pada tabel 4.2 menunjukkan perbandingan teks sebelum dan sesudah data cleansing yaitu teks yang awalnya mengandung berbagai elemen seperti *HTML* tag, *URL*, angka, emoji, dan tanda baca. Angka dan emoji dihilangkan untuk mengurangi gangguan pada saat analisis. Misalnya, emoji senyum menunjukkan sentimen positif. Namun, angka dan emoji juga bisa membingungkan dalam beberapa konteks kalimat.

Tabel 4. 1 Sebelum dan sesudah data cleansing

Data sebelum di cleansing	Data sesudah di cleansing
Soal Transaksi Rp 349 T di Kemenkeu Siaga 98: Mahfud MD Harus Jujur https://t.co/B9ny9HdtK0M	Soal Transaksi Rp T di Kemenkeu Siaga Mahfud MD Harus Jujur
Transaksi Janggal Rp349 T di Kemenkeu Satgas TPPU: Ada Potensi Tersangka Baru https://t.co/zwScMwSWm6	Transaksi Janggal Rp T di Kemenkeu Satgas TPPU Ada Potensi Tersangka Baru
Kemenkeu Lelang Online 60 Motor Royal Enfield Harga Mulai Rp23 Juta #Kemenkeu #Lelang #LelangMotor #RoyalEnfield https://t.co/lIYXGmCR4y https://t.co/aY759TDGFK	Kemenkeu Lelang Online Motor Royal Enfield Harga Mulai Rp Juta Kemenkeu Lelang LelangMotor RoyalEnfield

4.3.4 Preprocessing

Preprocessing dilakukan pada data sebelum data tersebut digunakan dalam analisis atau pemodelan. Tujuan utama dari preprocessing adalah untuk mempersiapkan data agar lebih mudah dipahami dan diolah oleh algoritma pemrosesan data atau pembelajaran mesin. Beberapa langkah yang biasanya dilakukan dalam *preprocessing* yaitu *Case Folding, tokenization, removing stopwords* dan *stemming.*

1. *Case Folding*

   Case folding merupakan proses mengubah semua huruf dalam sebuah teks menjadi huruf kecil (lowercase) bertujuan untuk memudahkan analisis. Pada gambar 4.11 merupakan kode dari *case folding.*

```
def Rmv_lowercase(text):
    # Mengonversi teks menjadi lowercase
    text_lower = text.lower()
    # Menghapus URL dari teks menggunakan regex
    return re.sub(r'(http|https)\S+', ' ', text_lower)

# Mengaplikasikan fungsi remove_url_and_lower pada kolom 'Rmv_emoji' dan menambahkan hasilnya ke kolom baru 'lowercase'
dataset['lowercase'] = dataset['Rmv_emoji'].apply(Rmv_lowercase)
```

Gambar 4. 11 Code Case Folding

Pada tabel 4.3 memperlihatkan perbedaan antara sebelum dan sesudah menerapkan *case folding.*

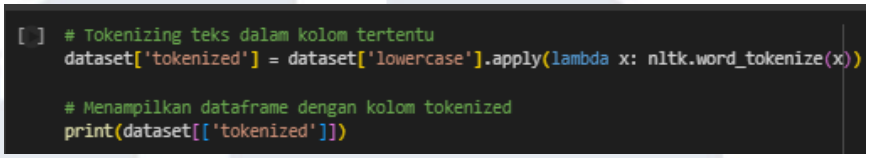
Tabel 4. 2 Sebelum dan sesudah case folding

Sebelum case folding	Sesudah case folding
Soal Transaksi Rp T di Kemenkeu Siaga Mahfud MD Harus Jujur	soal transaksi rp t di kemenkeu siaga mahfud md harus jujur
Transaksi Janggal Rp T di Kemenkeu Satgas TPPU Ada Potensi Tersangka Baru	transaksi janggal rp t di kemenkeu satgas tppu ada potensi tersangka baru
Kemenkeu Lelang Online Motor Royal Enfield Harga Mulai Rp Juta Kemenkeu Lelang LelangMotor RoyalEnfield	kemenkeu lelang online motor royal enfield harga mulai rp juta kemenkeu lelang lelangmotor royalenfield

2. *Tokenization*

   Tokenization merupakan proses memecah teks atau dokumen menjadi unit-unit yang lebih kecil, seperti kata-kata, frasa, atau simbol-simbol tertentu yang disebut dengan token. Tujuan utama dari

tokenization adalah untuk memudahkan pengolahan dan analisis teks. Pada tampilan kode di Gambar 4.12 merupakan proses dari *tokenization*



Gambar 4. 12 Code Tokenization

Pada tabel 4.4 memperlihatkan perbedaan antara sebelum dan sesudah menerapkan *tokenization.*

Tabel 4. 3 Sebelum dan sesudah tokenization.

Sebelum tokenization	Sesudah tokenization
soal transaksi rp 1 di kemenkeu siaga mahfud md harus jujur	['soal', 'transaksi', 'rp', '1', 'di', 'kemenkeu', 'siaga', 'mahfud', 'md', 'harus', 'jujur']
transaksi janggal rp 1 di kemenkeu satgas tppu ada potensi tersangka baru	['transaksi', 'janggal', 'rp', '1', 'di', 'kemenkeu', 'satgas', 'tppu', 'ada', 'potensi', 'tersangka', 'baru']
kemenkeu lelang online motor royal enfield harga mulai rp juta kemenkeu lelang lelangmotor royalenfield	['kemenkeu', 'lelang', 'online', 'motor', 'royal', 'enfield', 'harga', 'mulai', 'rp', 'juta', 'kemenkeu', 'lelang', 'lelangmotor', 'royalenfield']

3. *Removing stopwords*

Removing stopwords merupakan proses menghilangkan kata-kata yang umum dan tidak memberikan nilai tambah pada analisis teks. Kata-kata ini sering muncul dalam bahasa namun tidak membawa makna atau informasi penting dalam analisis. Pada tabel 4.5 adalah contoh kata-kata yang termasuk dalam stopwords.

Tabel 4. 4 Kata-kata masuk dalam stopwords

Kata-kata *stopwords*
['rp', 'yg', 'sbg', 'sebagai', 'jg', 'juga', 'wkwkwk', 'u', 'tp', 'tapi', 'itu', 'dan', 'akan', 'nggak','eh', 'ke', 'dari', 'lu', 'dr', 'ada', 'by', 'telah', 'mau', 't', 'tidak', 'ini', 'dia', 'blm', 'belum', 'belom', 'sama', 'loh', 'iya', 'ya', 'aja', 'gak', 'ga', 'ngga', 'engga', 'trus', 'nya', 'di', 'pak', 'yang', 'kita', 'jadi', 'oleh', 'dalam', 'karena', 'utk', 'harus', 'lain',

'bangun', 'masih', 'adalah', 'lebih', 'terus', 'kalau', 'kalo', 'atau', 'semua', 'sudah', 'banyak','dengan', 'pada', 'untuk', 'bisa', 'laku', 'hari', 'cara', 'lalu', 'hanya', 'apa', 'the', 'dgn', 'nyata', 'lagi', 'pasti', 'tuju', 'bukan', 'sebut"]

Pada tampilan kode di gambar 4.13 merupakan proses dari *removing stopwords.*

```
[ ]  #bikin fungsi untuk stopwords'
     import re

     def remove_stopword(text):
     #  gunakan kolom Slang Removed dan simpan di kolom baru Stopwords Removed
         for i in range(len(stop)):
             if stop[i] in text:
                 text = re.sub(r'\b(?:'+stop[i]+r')\b', '', text)
                 text = text.replace('  ',' ')
         return text

     #apply funsi tersebut

     dataset['Rmv_stopwors'] = dataset['lowercase'].apply(remove_stopword)
```

Gambar 4. 13 Code Removing Stopwords.

Pada tabel 4.6 memperlihatkan perbedaan antara sebelum dan sesudah menerapkan *removing stopwords*.

Tabel 4. 5 Sebelum dan sesudah removing stopwords

Sebelum removing stopwords	Sesudah removing stopwords	Removed
soal transaksi rp t di kemenkeu siaga mahfud md harus jujur	soal transaksi kemenkeu siaga mahfud md jujur	['rp', 't', 'di', 'harus']
transaksi janggal rp t di kemenkeu satgas tppu ada potensi tersangka baru	transaksi janggal kemenkeu satgas tppu potensi tersangka baru	['rp', 't', 'di', 'ada',]
kemenkeu lelang online motor royal enfield harga mulai rp juta kemenkeu lelang lelangmotor royalenfield	kemenkeu lelang online motor royal enfield harga mulai juta kemenkeu lelang lelangmotor royalenfield	['rp',]

4. *Stemming*

Stemming merupakan proses dalam pemrosesan teks yang bertujuan untuk mengubah kata-kata menjadi bentuk dasarnya dengan menghapus akhiran kata yang tidak diperlukan seperti imbuhan dan kata yang serupa. Membantu untuk mengurangi variasi kata yang sama agar kata-kata tersebut

dapat dianggap sebagai satu entitas saat analisis teks dilakukan. Pada tampilan kode di gambar 4.14 merupakan proses dari *Stemming.*

```
[ ]  #bikin fungsi untuk melakukan stemming

     def stem_sentences(sentence):
     #tulis step-step pengaplikasian stemmer di atas
         tokens = sentence.split()
         stemmed_tokens = [stemmer.stem(token) for token in tokens]
         return " ".join(stemmed_tokens)

     # apply fungsi diatas

     dataset['Stem_tweet'] = dataset['Rmv_stopwors'].apply(stem_sentences)
```

Gambar 4. 14 Code Stemming

Pada tabel 4.7 memperlihatkan perbedaan antara sebelum dan sesudah menerapkan *stemming.*

Tabel 4. 6 Sebelum dan sesudah stemming.

Sebelum Stemming	Sesudah Stemming
soal transaksi kemenkeu siaga mahfud md jujur	soal transaksi kemenkeu siaga mahfud md jujur
transaksi janggal kemenkeu satgas tppu potensi tersangka baru	transaksi janggal kemenkeu satgas tppu potensi sangka baru
kemenkeu lelang online motor royal enfield harga mulai juta kemenkeu lelang lelangmotor royalenfield	kemenkeu lelang online motor royal enfield harga mulai juta kemenkeu lelang lelangmotor royalenfield

4.3.5 Data Translation

Data translation merupakan proses mengubah teks dari satu bahasa ke bahasa lain menggunakan algoritma komputer. Proses *translation* dilakukan dengan menggunakan library deep_translator. Kode yang digunakan untuk melakukan *translation* dapat dilihat dalam gambar 4.15

```
!pip install deep_translator

from deep_translator import GoogleTranslator
import pandas as pd

# Inisialisasi translator
translator = GoogleTranslator(source='auto', target='en')

# Fungsi untuk menerjemahkan teks
def translate_text(text):
    try:
        translation = translator.translate(text)
        return translation
    except Exception as e:
        return "Error: " + str(e)


# Menerjemahkan kolom 'full_text'
dataset['tweet'] = dataset['Stemming'].apply(lambda x: translate_text(x))

print(dataset)
```

Gambar 4. 15 Code Data Translation

Pada tabel 4.8 adalah sebuah contoh perubahan sebelum dan setelah proses *translation* yang bisa dilihat di bawah ini.

Tabel 4. 7 Sebelum dan setelah proses translation

Sebelum translation	Sesudah translation
soal transaksi kemenkeu siaga mahfud md jujur	regarding the Ministry of Finance transaction, Mahfud MD is honest
transaksi janggal kemenkeu satgas tppu potensi sangka baru	strange transactions ministry of finance tppu task force potential new suspects
kemenkeu lelang online motor royal enfield harga mulai juta kemenkeu lelang lelangmotor royalenfield	Ministry of Finance online auction of Royal Enfield motorbikes, prices starting at millions, Ministry of Finance auction, Royal Enfield motorbike auction

4.3.*5 Labelling*

Labelling merupakan langkah penting dalam analisis data di mana memberikan label atau tag pada data untuk mengidentifikasi atau mengkategorikan informasi yang terdapat di dalamnya. Pada gambar 4.16 menggunakan library *Vader* untuk melakukan labelling pada data yang berhasil di translate pada variable tweet. Label yang digunakan adalah *positive*, *negative*, dan *neutral* untuk mengklasifikasikan sentimen dari teks yang ada dalam dataset tersebut [50]. Label positif berisikan opini yang menunjukkan perasaan atau pandangan yang baik. Label negatif berisikan

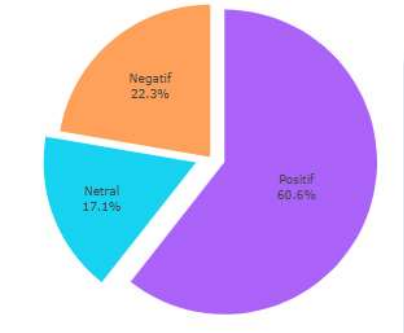
opini yang menunjukkan perasaan atau pandangan yang buruk. Label netral berisikan opini yang tidak menunjukkan perasaan yang kuat baik positif maupun negatif. Proses labelling ini membantu dalam pemahaman dan analisis lebih lanjut terhadap sentimen yang terkandung dalam data tersebut.

```
[ ]  nltk.download('vader_lexicon')
     from nltk.sentiment.vader import SentimentIntensityAnalyzer

     # Membuat objek SentimentIntensityAnalyzer
     sid = SentimentIntensityAnalyzer()

     # Fungsi untuk mendapatkan skor sentimen
     def get_sentiment_score(sentence):
         skor_sentimen = sid.polarity_scores(sentence)
         return skor_sentimen['compound']

     # Menambahkan kolom skor sentimen ke DataFrame
     dataset['skor_sentimen'] = dataset['tweet'].apply(get_sentiment_score)

     # Menampilkan DataFrame dengan skor sentimen
     print(dataset)
```

```
[ ]  # Fungsi untuk mendapatkan label sentimen
     def get_sentiment_label(score):
         if score >= 0.05:
             return "Positif"
         elif score <= -0.05:
             return "Negatif"
         else:
             return "Netral"

     # Menambahkan kolom label sentimen ke DataFrame
     dataset['label_sentimen'] = dataset['skor_sentimen'].apply(get_sentiment_label)

     # Menampilkan DataFrame dengan label sentimen
     print(dataset)
```

Gambar 4. 16 Code Labelling

Pada gambar 4.17 merupakan hasil dari label sentiemen data opini public Kementerian keuangan. Pada gambar 4.17 merupakan sebuah visualisasi *pie chart* dari sentiment data opini publik terhadap Kementerian keuangan. Pada gambar di bawah memperlihatkan hasil label "*positive*" menjadi label yang paling dominan dengan frekuensi kemunculan sebanyak 60.6%. Hasil pelabelan sentimen ini menunjukkan bahwa mayoritas opini publik terhadap Kementerian Keuangan cenderung positif.



Gambar 4. 17 Visualisasi Sentimen

.

4.3.6 *TF-IDF*

```
from sklearn.feature_extraction.text import TfidfVectorizer

# Vectorization menggunakan TfidfVectorizer
vectorizer = TfidfVectorizer()
x = vectorizer.fit_transform(dataset['Stemming'])
```

Gambar 4. 18 Code TF-IDF

Pada gambar 4.18 merupakan kode untuk menjalankan proses *tf-idf* merupakan proses mengubah data mentah menjadi representasi fitur yang dapat digunakan dalam model pembelajaran mesin. Salah satu metode untuk ekstraksi fitur dari teks adalah dengan menggunakan *TF-IDF (Term Frequency-Inverse Document Frequency)*. Pada kode di atas *TfidfVectorizer* berguna untuk mengonversi kumpulan dokumen teks menjadi representasi vektor TF-IDF dan mempersiapkan data teks untuk digunakan dalam model pembelajaran mesin.

4.3.7 *Label Encode*

Label Encoding proses yang mengubah nilai kategorikal menjadi nilai numerik dalam analisis data. Hal ini membantu algoritma pembelajaran mesin memahami hubungan antara nilai-nilai kategorikal dalam data. Penggunaan *Label Encoding* seringkali diperlukan sebelum melatih model untuk memproses data yang memiliki variabel kategorikal.

```
from sklearn.preprocessing import LabelEncoder

# Encode label kategori menjadi nilai numerik
label_encoder = LabelEncoder()
y = label_encoder.fit_transform(dataset['label_sentimen'])
```

Gambar 4. 19 Code Label Encoding

Pada gambar 4.19 merupakan kode untuk menjalankan proses *label encoding*. *Label encoding* adalah proses mengonversi nilai kategori menjadi nilai numerik. Pada kode di atas untuk mengimpor kelas *LabelEncoder* mengubah nilai-nilai kategori menjadi bilangan bulat sesuai dengan urutan kemunculan nilai kategori dalam data.

4.3.8 *SMOTE*

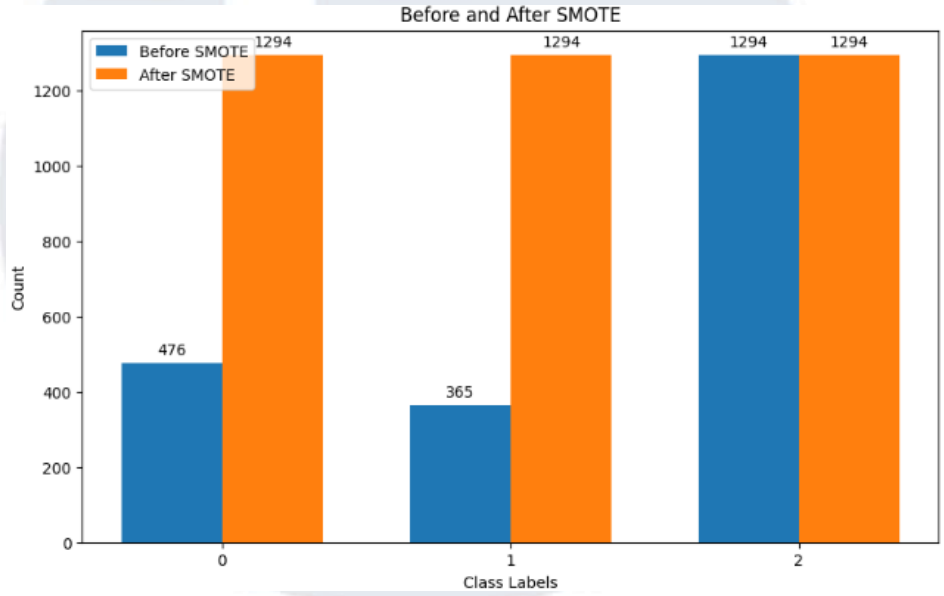
SMOTE langkah yang digunakan untuk menangani ketidakseimbangan kelas dalam dataset. Tahap ini dilakukan dengan menambahkan data tambahan pada kelompok yang kurang representatif, memastikan bahwa komputer dapat belajar dengan lebih efisien dan tepat. *SMOTE* membantu meningkatkan keberagaman data pada kelompok minoritas memungkinkan model *Machine Learning* untuk belajar dari ketidakseimbangan dalam dataset.

```
from imblearn.over_sampling import SMOTE
# Lakukan SMOTE terlebih dahulu
smote = SMOTE(random_state=42)
X_resampled, y_resampled = smote.fit_resample(x, y)
```

Gambar 4. 20 Code SMOTE

Pada gambar 4.20 merupakan kode yang digunakan untuk menjalankan *SMOTE* dari pustaka *imbalanced-learn* guna melakukan *oversampling*. Pada hasil labeling terdapat data yang tidak seimbang dapat dapat mengurangi akurasi model secara keseluruhan. *SMOTE* membantu mengatasi masalah ketidakseimbangan kelas dalam pembelajaran mesin dengan membuat data menjadi lebih seimbang. Teknik *SMOTE* digunakan untuk mengidentifikasi kelas mayoritas dan minoritas dalam dataset yang

telah dilabel menggunakan VADER. Selanjutnya, parameter disesuaikan agar jumlah sampel sintetis yang dihasilkan sama dengan kelas mayoritas. Sampel sintetis ini dibuat secara artifisial untuk menyeimbangkan jumlah data antara kelas mayoritas dan minoritas dalam dataset [29].



Gambar 4. 21 Visualisasi Before and After SMOTE

Pada gambar 4.21 terlihat perbandingan antara data sebelum dan setelah penggunaan *SMOTE*. Sebelum menggunakan *SMOTE* terdapat ketidakseimbangan data di mana jumlah sampel pada beberapa kelas berbeda. Pada sentiment ditandai 'Negatif': 0, 'Netral': 1, 'Positif': 2. Pada kelas positif terdapat 1294 data, kelas negatif memiliki 476 data, dan kelas netral hanya memiliki 365 data. Setelah menerapkan metode *SMOTE* dataset menjadi lebih seimbang secara keseluruhan dan berhasil menambahkan data pada kelas-kelas yang kurang.

4.3.9 *Data Splitting*

Tahap *data splitting* merupakan langkah penting dalam pengembangan model *Machine Learning* yang membagi dataset menjadi subset untuk pelatihan, validasi, dan pengujian. Pemisahan dataset menjadi bagian-bagian ini membantu mengidentifikasi performa model dan

mencegah *overfitting* atau *underfitting*. Proses *data splitting* dilakukan setelah SMOTE untuk memastikan bahwa data yang digunakan untuk pelatihan dan pengujian model tetap mewakili keseluruhan data. Data spilitting menggunakan keseluruhan data yang telah dilakukan SMOTE, data SMOTE dibagi menjadi 80:20.

```
from sklearn.model_selection import train_test_split


# Melakukan data splitting pada data yang sudah di-SMOTE
X_train, X_test, y_train, y_test = train_test_split(X_resampled, y_resampled, test_size=0.2, random_state=42)

# Menampilkan ukuran data training dan data testing setelah SMOTE
print("Jumlah data training setelah SMOTE:", X_train.shape[0])
print("Jumlah data testing setelah SMOTE:", X_test.shape[0])
```

Gambar 4. 22 Code Data Splitting

Pada gambar 4.22 adalah kode untuk membagi dataset yang telah di *SMOTE* menggunakan fungsi *train_test_split* dari *Scikit-Learn*. Kode tersebut memisahkan data menjadi 80% untuk *data training* dan 20% untuk *data testing* dengan menggunakan nilai *test_size*=0.2. Pada kode *X_resampled* dan *y_resampled* merupakan dataset yang sudah melalui proses *SMOTE*.

4.4 Modeling

4.4.1 Word Cloud

Gambar 4. 23 *Visualisai Word Cloud*

Pada gambar 4.23 merupakan visualisasi *word cloud* opini public pada Kementerian Keuangan. Visualisasi *word cloud* adalah grafis dari teks di mana kata-kata yang paling sering muncul dalam teks ditampilkan dalam ukuran yang lebih besar daripada kata-kata yang jarang muncul. *Word cloud* digunakan untuk memberikan gambaran visual tentang frekuensi kata-kata dalam teks dan mengidentifikasi tema atau topik utama yang muncul. Pada gambar di atas, ukuran kata yang lebih besar adalah 'Kemenkeu, uang, dan pajak' memiliki frekuensi kemunculan yang lebih tinggi menunjukkan bahwa kata tersebut salah satu kata-kata yang paling sering muncul dalam dataset tersebut.

4.4.2 Naïve Bayes

```
from sklearn.model_selection import GridSearchCV
from sklearn.metrics import accuracy_score, classification_report
from sklearn.naive_bayes import MultinomialNB

# Inisialisasi model Naive Bayes
nb_model = MultinomialNB()

# Tentukan parameter grid untuk pencarian
param_grid = {'alpha': [0.1, 0.5, 1.0, 1.5, 2.0]}

# Inisialisasi GridSearchCV
nb_grid_search = GridSearchCV(nb_model, param_grid, cv=5)

# Melakukan pencarian parameter terbaik
nb_grid_search.fit(X_train, y_train)

# Ambil parameter terbaik dari hasil pencarian
best_nb_params = nb_grid_search.best_params_

# Inisialisasi model Naive Bayes dengan parameter terbaik
best_nb_model = MultinomialNB(alpha=best_nb_params['alpha'])

# Melatih model Naive Bayes dengan parameter terbaik
best_nb_model.fit(X_train, y_train)

# #Output naive bayes memberikan hasil dibawah ini:
print("Parameter terbaik untuk Naive Bayes:", best_nb_params)
```

```
Parameter terbaik untuk Naive Bayes: {'alpha': 0.1}
```

Gambar 4. 24 Code Parameter Naive Bayes

Pada gambar 4.24 merupakan kode untuk melakukan optimisasi parameter untuk model klasifikasi *Naive Bayes* menggunakan metode *Grid Search Cross-Validation*. Pada kode di atas, *Naive Bayes* digunakan untuk melakukan klasifikasi pada data yang dimasukkan. Alpha dalam kode

tersebut adalah parameter smoothing yang digunakan dalam model Naive Bayes. Nilai alpha yang optimal dapat ditemukan dengan menggunakan Grid Search untuk meningkatkan kinerja model. Parameter terbaik untuk *Naive Bayes* yang ditemukan adalah {'alpha': 0.1}. Cv=5 pada *GridSearchCV* dipilih karena memberikan keseimbangan optimal antara akurasi evaluasi model dan efisiensi waktu komputasi. Dengan cv=5, data dibagi menjadi 5 fold, sehingga model dilatih pada 80% data dan diuji pada 20%. Pembagian ini menghasilkan evaluasi yang cukup akurat tanpa terlalu membebani komputasi, berbeda dengan nilai yang lebih tinggi seperti cv=10 yang membutuhkan waktu lebih lama [16].

4.4.3 Support Vector Machine

```
from sklearn.model_selection import train_test_split
from sklearn.svm import SVC
from sklearn.model_selection import GridSearchCV

# Parameter grid untuk SVC (Support Vector Classifier)
svc_param_grid = {
    'C': [0.1, 1, 10],
    'kernel': ['linear', 'rbf'],
    'gamma': ['scale', 'auto']
}

# Inisialisasi GridSearchCV untuk SVC
svc_grid_search = GridSearchCV(SVC(random_state=42), svc_param_grid, cv=5, n_jobs=-1)

# Latih model SVC dengan parameter terbaik yang ditemukan oleh GridSearchCV
svc_grid_search.fit(X_train, y_train)

# Ambil model SVC terbaik dari hasil pencarian
best_svc_model = svc_grid_search.best_estimator_

#Output SVM memberikan hasil dibawah ini:
print("Parameter terbaik untuk SVC:", best_svc_model)
```

```
Parameter terbaik untuk SVC: SVC(C=10, random_state=42)
```

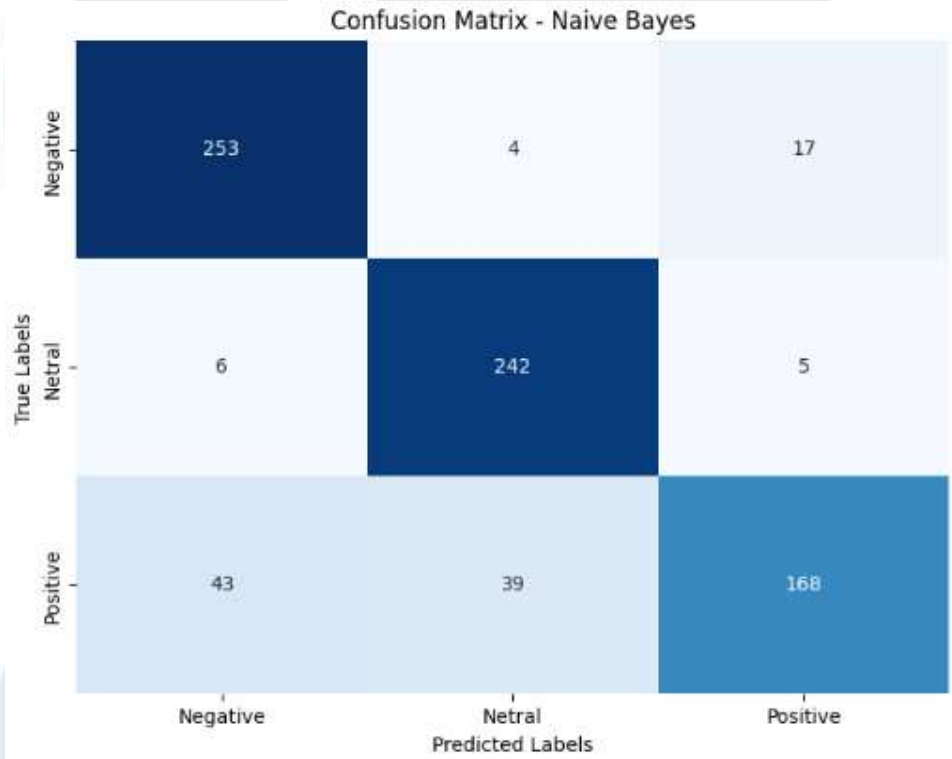
Gambar 4. 25 Code Parameter Support Vector Classsifer

Pada gambar 4.25 menggunakan *GridSearchCV* untuk menemukan parameter terbaik untuk model *Support Vector Classifier*. *SVM* digunakan untuk klasifikasi atau regresi serta mencari *hyperplane* terbaik dari kedua kelas. Parameter *Grid* untuk menentukan nilai yang akan diuji untuk parameter C, kernel dan gamma. Parameter terbaik untuk model *SVM* yang ditemukan dalam hasil *GridSearchCV* adalah C=10. *GridSearchCV* menemukan bahwa C=10 memberikan hasil yang paling baik pada dataset

tersebut. *Random_state*=42 memastikan hasil yang ditemukan yakni mempermudah dalam memvalidasi bahwa hasil yang diperoleh konsisten dan dapat dibuat.

4.5 Evaluation

4.5.2 Naïve Bayes



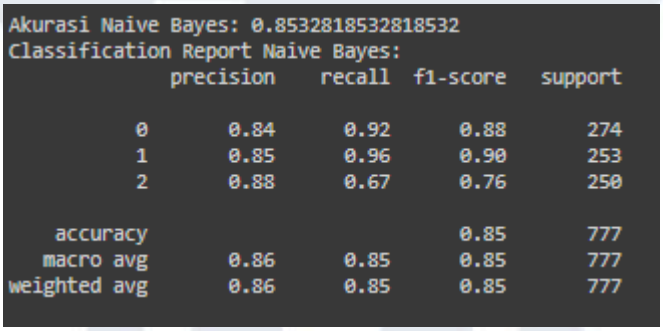
Gambar 4. 26 Confusion Matrix Naïve Bayes

Pada Gambar 4.26, ditampilkan hasil dari confusion matrix yang menunjukkan performa model dalam memprediksi tiga kelas sentimen: negatif, netral, dan positif. Sentimen negatif, dari total 274 data yang sebenarnya memiliki sentimen negatif, model berhasil mengklasifikasikan 253 di antaranya dengan benar sebagai negatif (*True Negative*). Namun, terdapat 17 data yang salah diklasifikasikan sebagai positif (*False Positive*) dan 4 data lainnya salah diklasifikasikan sebagai netral (False Neutral).

Selanjutnya, pada sentimen netral, terdapat total 242 data yang benar-benar memiliki sentimen netral, dan model mampu memprediksi

seluruhnya dengan benar sebagai netral (*True Neutral*). Hal ini menunjukkan bahwa model berhasil mengenali data netral dengan sangat baik, tanpa adanya kesalahan klasifikasi ke dalam kategori negatif atau positif, yang berarti akurasi untuk kelas netral cukup tinggi.

Sentimen positif, dari 250 data yang sebenarnya memiliki sentimen positif, sebanyak 168 di antaranya berhasil diklasifikasikan dengan benar sebagai positif (*True Positive*). Namun, terdapat 39 data positif yang salah diklasifikasikan sebagai netral (False Neutral) dan 43 data positif lainnya yang salah diklasifikasikan sebagai negatif (*False Negative).* Hasil confusion matrix ini secara keseluruhan memberikan gambaran mendalam mengenai ketepatan dan kelemahan model dalam mengklasifikasikan setiap kelas sentimen.

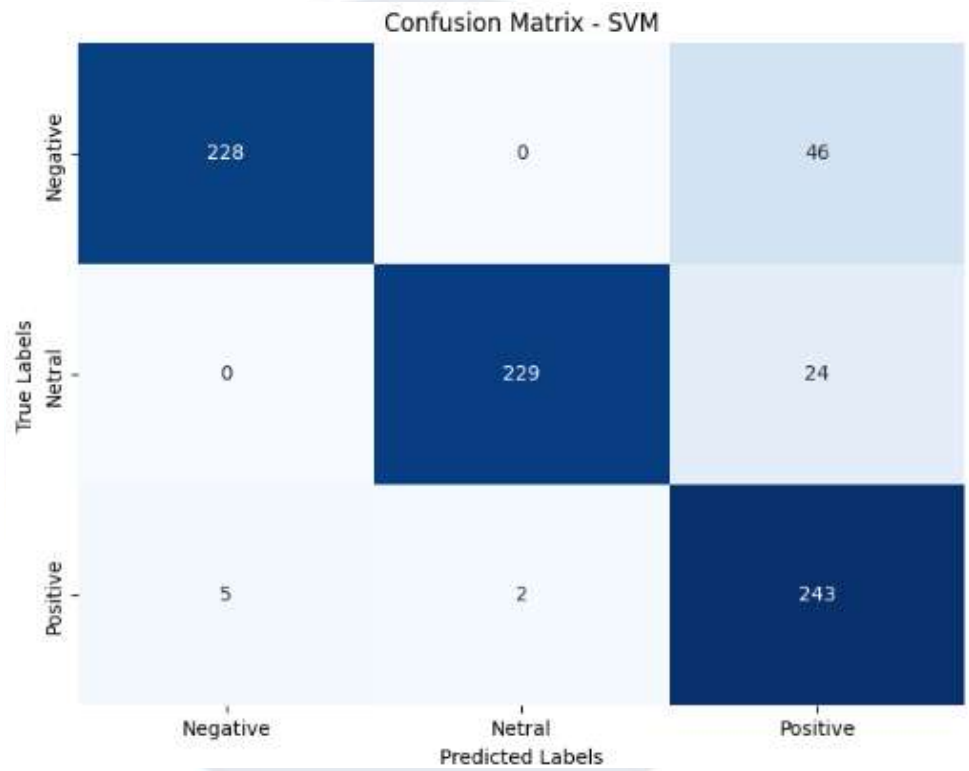


```
Akurasi Naive Bayes: 0.8532818532818532
Classification Report Naive Bayes:
              precision    recall  f1-score   support

           0       0.84      0.92      0.88       274
           1       0.85      0.96      0.90       253
           2       0.88      0.67      0.76       250

    accuracy                           0.85       777
   macro avg       0.86      0.85      0.85       777
weighted avg       0.86      0.85      0.85       777
```

Gambar 4. 27 Performa Naïve Bayes

Pada gambar 4.27 menampilkan hasil dari penerapan metode Naïve Bayes mencakup nilai *accuracy, precision, recall, dan F1-score*. Hasil analisis menunjukkan bahwa penggunaan metode *Naïve Bayes* menghasilkan nilai akurasi sebesar 85.32%. Selain itu, presisi memiliki nilai sebesar 86%, *recall* dan *f1-score* masing-masing sebesar 85%. Model *Naïve Bayes* memberikan performa yang baik dalam mengklasifikasikan data secara akurat.

4.5.2 Support Vector Machine



Gambar 4. 28 Confusion Matrix Support Vector Machine

Pada Gambar 4.28, ditampilkan hasil dari confusion matrix yang menunjukkan performa model SVM dalam memprediksi tiga kelas sentimen: negatif, netral, dan positif. Sentimen negatif, dari total 274 data yang sebenarnya memiliki sentimen negatif, model SVM berhasil mengklasifikasikan 228 di antaranya dengan benar sebagai negatif (*True Negative*). Namun, terdapat 46 data yang salah diklasifikasikan sebagai positif (*False Positive*), menunjukkan beberapa kesalahan klasifikasi yang perlu diperhatikan.

Pada sentimen netral, terdapat total 229 data yang benar-benar memiliki sentimen netral, dan model SVM berhasil memprediksi seluruhnya dengan benar sebagai netral (True Neutral). Tidak ada kesalahan klasifikasi ke dalam kategori negatif atau positif pada kelas ini, yang menunjukkan performa yang sangat baik dalam mengenali sentimen netral secara akurat.

Sentimen positif, dari 250 data yang sebenarnya memiliki sentimen positif, sebanyak 243 di antaranya diklasifikasikan dengan benar sebagai positif (*True Positive*). Namun, terdapat 5 data positif yang salah diklasifikasikan sebagai negatif (*False Negative*) dan 2 data positif yang salah diklasifikasikan sebagai netral (*False Neutral*). Secara keseluruhan, hasil dari confusion matrix ini memberikan wawasan yang berharga mengenai kekuatan dan kelemahan model SVM dalam mengklasifikasikan setiap kelas sentimen dengan lebih mendalam.

```
Akurasi SVC: 0.9009009009009009
Classification Report SVC:
              precision    recall  f1-score   support

           0       0.98      0.83      0.90       274
           1       0.99      0.91      0.95       253
           2       0.78      0.97      0.86       250

    accuracy                           0.90       777
   macro avg       0.92      0.90      0.90       777
weighted avg       0.92      0.90      0.90       777
```

Gambar 4. 29 Perfoma Support Vector Machine

Performa yang terlihat dalam gambar 4.29 menampilkan hasil dari penerapan metode *Support Vector Machine* mencakup nilai *accuracy*, *precision*, *recall*, dan *F1-score*. Hasil analisis menunjukkan bahwa penggunaan metode *Support Vector Machine* menghasilkan nilai akurasi sebesar 90%. Selain itu, presisi memiliki nilai sebesar 92%, *recall* dan *f1-score* masing-masing sebesar 91%. Akurasi ini menunjukkan bahwa model *SVM* dapat mengklasifikasikan dan memprediksi dengan tepat.

4.6 Deployment

Tahap deployment dalam CRISP-DM merupakan langkah penting dalam pembuatan *website*, termasuk dalam pengembangan aplikasi SATUKU. *Deployment* dengan visualisasi bertujuan untuk memberikan kemudahan dan interaksi kepada pengguna, sehingga mereka dapat dengan mudah menganalisis data opini publik seputar Kementerian Keuangan. Visualisasi yang disajikan,

seperti pie chart, word cloud, dan bar chart, membantu memperjelas informasi yang terkandung dalam data dan mempermudah pemahaman terhadap pola-pola sentimen masyarakat.

SATUKU merupakan nama website yang dibuat dalam penelitian ini. SATUKU adalah aplikasi yang dirancang khusus untuk menganalisis sentimen publik terhadap Kementerian Keuangan (Kemenkeu), berdasarkan data yang dikumpulkan dari media sosial platform X. Aplikasi ini menggunakan teknik Machine Learning untuk mengidentifikasi serta membandingkan opini masyarakat, baik yang bersifat positif, negatif, maupun netral. Dengan adanya visualisasi yang interaktif, SATUKU memungkinkan pengguna untuk memahami dan menginterpretasi sentimen publik secara lebih efektif dan efisien.

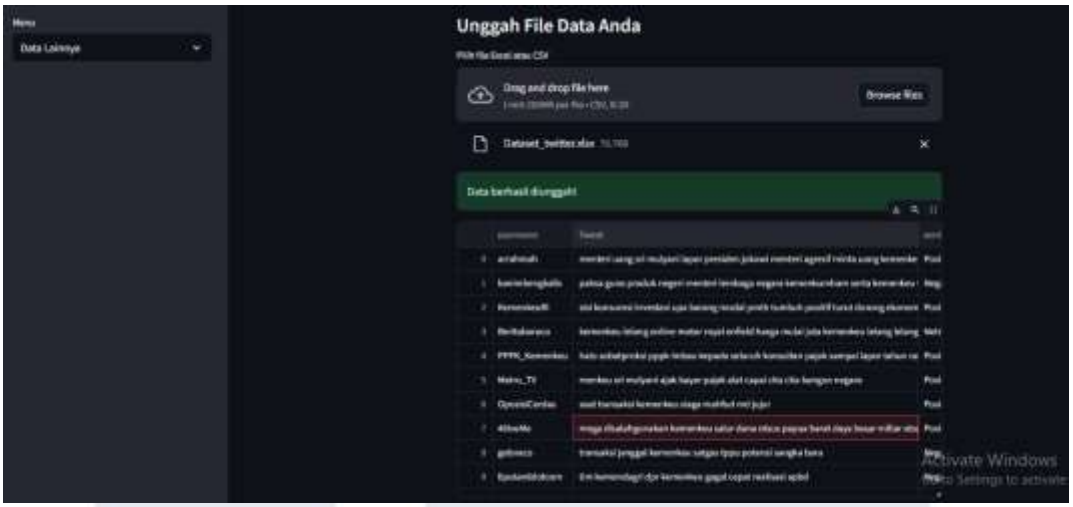
Gambar 4. 30 Tampilan website opini publik pada Kementerian Keuangan.

Pada Gambar 4.30 ditampilkan tampilan awal dari website SATUKU, sebuah aplikasi yang dirancang untuk menganalisis sentimen publik terhadap Kementerian Keuangan. Melalui dropdown menu, pengguna dapat memilih beberapa opsi utama seperti data periode Agustus 2023-2024, visualisasi, dan data lainnya yang berhubungan dengan tweet tentang Kementerian Keuangan. Di menu "Data Periode Agustus 2023-2024," disajikan informasi yang diperoleh melalui proses scraping dari platform X, yang memberikan gambaran jelas tentang persepsi

publik terhadap Kementerian Keuangan selama periode tersebut. Fitur ini memudahkan pengguna untuk memahami sentimen publik, baik itu positif, negatif, maupun netral, terhadap kebijakan Kementerian Keuangan.

Gambar 4. 31 Tampilan menu visualisasi

Pada Gambar 4.31, pengguna dapat melihat berbagai opsi visualisasi yang menggambarkan opini publik mengenai Kementerian Keuangan. Opsi-opsi visualisasi ini mencakup pie chart yang efektif untuk menunjukkan distribusi sentimen masyarakat—apakah positif, negatif, atau netral—terhadap kebijakan dan tindakan Kementerian. Selain itu, terdapat word cloud yang menampilkan kata-kata yang paling sering muncul dalam opini publik, memberikan gambaran mengenai isu-isu utama yang menjadi perhatian masyarakat. Visualisasi bar chart juga ditampilkan untuk menunjukkan username teratas yang aktif dalam diskusi mengenai Kementerian Keuangan, memberikan informasi tentang pengaruh individu atau akun tertentu dalam percakapan ini. Visualisasi ini, pengguna dapat memperoleh wawasan yang lebih luas dan mendalam terkait opini publik, sehingga dapat memahami lebih baik bagaimana masyarakat merespons kebijakan Kementerian Keuangan.



Gambar 4. 32 Tampilan menu data lainnya

Pada Gambar 4.32, ditampilkan menu "Data Lainnya" yang memberikan kemudahan bagi pengguna untuk mengunggah file data mereka sendiri dalam format CSV atau XLSX. Setelah file berhasil diunggah, pengguna dapat langsung melihat data tersebut ditampilkan dalam aplikasi. Salah satu visualisasi yang disajikan adalah *wordcloud* yang memperlihatkan kata-kata kunci yang paling sering muncul dalam dataset, sehingga pengguna dapat dengan cepat mengidentifikasi tema atau isu yang mendominasi. Selain itu, *pie chart* akan memberikan gambaran mengenai distribusi sentimen, menunjukkan proporsi opini publik yang positif, negatif, dan netral terhadap Kementerian Keuangan. *Bar chart* akan menampilkan username teratas yang berkontribusi pada diskusi ini, memberikan wawasan tentang individu atau akun yang paling berpengaruh. Fitur ini, pengguna tidak hanya dapat menganalisis data mereka sendiri tetapi juga mendapatkan perspektif yang lebih jelas mengenai opini publik terhadap Kementerian Keuangan.

Gambar 4. 33 Visualisasi dari data yang diunggah

Pada gambar 4.33 memperlihatkan visualisasi dari contoh data yang diunggah, yang mencakup beberapa elemen analisis, seperti word *cloud* yang menggambarkan kata-kata yang paling sering muncul dalam data, distribusi sentimen yang menunjukkan proporsi sentimen positif, negatif, dan netral, serta daftar *top usernames* yang membantu mengidentifikasi pengguna yang paling berpengaruh dalam percakapan. Visualisasi ini memberikan gambaran umum mengenai pola yang muncul dari data yang telah diolah dan memberikan insight awal terkait tren dan persepsi publik yang terekam dalam dataset.

4.7 Hasil dan Pembahasan

Evaluasi dari algoritma *Naïve Bayes* dan *Support Vector Machine* akan dibandingkan dengan penelitian sebelumnya. Pada tabel 4.9 merupakan tabel peneliti terdahulu, Penelitian yang dilakukan oleh A. Anugrah Aqsaa, Irawatia, dan Lukman Syafieb membandingkan kinerja dua algoritma, yaitu Naïve Bayes dan Support Vector Machine (*SVM*) [11]. Hasil penelitian tersebut, *SVM* terbukti lebih unggul dengan akurasi sebesar 74%, dibandingkan Naïve Bayes yang memiliki akurasi lebih rendah. Sementara itu, Penelitian yang dilakukan oleh Yulius Bambang Seran dan Supatman menggunakan algoritma Support Vector Machine (*SVM*) untuk analisis data [12]. Hasil penelitian menunjukkan bahwa model *SVM* mencapai tingkat akurasi sebesar 66%.

Tabel 4. 8 Peneliti Terdahulu

Penelitian Terdahulu		
Peneliti	Algoritma	Akurasi
A. Anugrah Aqsaa, Irawatia, Lukman Syafieb, [11]	*Support Vector Machine*	74%
	Naïve Bayes	71,7%
Yulius Bambang Seran, Supatman [12]	*Support Vector Machine*	66%
Penelitian Ini		
Algoritma		Akurasi
Support Vector Machine		90%
Naïve Bayes		85.32%.

Model *Support Vector Machine* mampu memberikan prediksi yang seimbang dan akurat terhadap pandangan positif dan negatif terhadap Kementerian Keuangan dengan tingkat akurasi mencapai 90%, algoritma *Support Vector Machine* dapat mengklasifikasikan dan memprediksi dengan tingkat ketepatan yang sangat baik. Algoritma *Naïve Bayes* menghasilkan tingkat akurasi sebesar 85.32%. Hasil penelitian ini dapat disimpulkan bahwa penggunaan metode yang melibatkan algoritma *Naïve Bayes* dan *Support Vector Machine* menghasilkan kinerja yang lebih unggul. *SVM* berhasil mengklasifikasikan 228 dari 274 data negatif sebagai negatif (*True Negative*). Sementara *Naïve Bayes* memang mengklasifikasikan 253 data dengan benar, *SVM* cenderung memiliki kesalahan klasifikasi yang lebih sedikit. Pada sentimen netral, *SVM* memprediksi semua 229 data netral dengan benar (*True Neutral*) tanpa satu pun kesalahan. Pada sentimen positif, dari 250 data positif, *SVM* berhasil mengklasifikasikan 243 dengan benar (*True Positive*), sementara *Naïve Bayes* menunjukkan lebih banyak kesalahan. *SVM* lebih efektif dalam mengklasifikasikan sentimen karena lebih mampu menangani kompleksitas

data dan memiliki tingkat kesalahan klasifikasi yang lebih rendah dibandingkan *Naïve Bayes*. Hasil evaluasi terkait akurasi, presisi, *recall*, dan *F1-score* menunjukkan peningkatan yang signifikan dibandingkan dengan penelitian sebelumnya. Metode-metode tersebut terbukti efektif dalam melakukan evaluasi sentimen terhadap opini masyarakat tentang kementerian keuangan. Hasil penelitian ini memberikan wawasan yang penting dalam memahami dinamika opini publik terkait Kementerian Keuangan Indonesia di media sosial X. Analisis jaringan sosial ini dapat menjadi landasan untuk lebih memahami persepsi dan respons masyarakat terhadap isu-isu terkait Kementerian Keuangan.

.

BAB V
SIMPULAN DAN SARAN

5.1 Simpulan

Berdasarkan penelitian yang dilakukan terkait opini publik terhadap Kementerian Keuangan berikut merupakan kesimpulan dan saran yang diambil

1. Algoritma Support Vector Machine (SVM) memiliki performa yang lebih baik dibandingkan dengan model Naïve Bayes dalam mengklasifikasikan opini publik terhadap Kementerian Keuangan, dengan nilai akurasi *SVM* mencapai 90% dibandingkan 85.32% untuk *Naïve Bayes*. Hal ini menunjukkan bahwa SVM lebih efektif dalam menangani data yang kompleks dan tidak normal, yang sering ada pada opini publik. Hasil ini dapat membantu Kementerian Keuangan dalam memahami dan merespons opini masyarakat dengan lebih baik.
2. Penelitian ini, dilakukan perancangan dan implementasi *User Interface (UI)* menggunakan *Streamlit* untuk menganalisis opini publik terhadap Kementerian Keuangan. *Website* yang dibuat memuat hasil visualisasi dari data yang diperoleh melalui proses *Scraping* di X. Visualisasi yang disajikan mencakup distribusi sentimen dalam bentuk *Pie Chart, Word Cloud* untuk menampilkan kata-kata yang paling sering muncul.

5.2 Saran

Berdasarkan analisis kesimpulan yang telah dibuat terdapat sejumlah rekomendasi yang dapat bermanfaat bagi peneliti pada masa yang akan datang.

1. Disarankan untuk memperluas cakupan data dengan memasukkan *platform* media sosial lainnya selain X, seperti facebook, tiktok, dan youtube untuk mendapatkan gambaran yang lebih luas tentang opini publik.
2. Disarankan untuk analisis sentimen yang lebih luas dengan mendeteksi ekspresi emosional seperti marah, sedih, senang, dan takut. Analisis sentimen untuk mencakup ekspresi emosional dapat memberikan wawasan yang lebih mendalam tentang bagaimana opini publik terhadap Kementerian Keuangan.

DAFTAR PUSTAKA

[1] A. Aridho, T. A. Situmeang, D. R. Tinambunan, K. N. Ramadhani, M. W. Lase, and J. Ivanna, "Peran Media Massa Dalam Membentuk Opini Publik: Demokratisasi Pasca-Reformasi," *IJEDR: Indonesian Journal of Education and Development Research*, vol. 2, no. 1, pp. 206–210, Jan. 2024, doi: 10.57235/IJEDR.V2I1.1693.

[2] R. Korniawan, "Opini publik media massa terhadap masalah penghindaran pajak: perbandingan Indonesia dan Irlandia," *PRofesi Humas Jurnal Ilmiah Ilmu Hubungan Masyarakat*, vol. 4, no. 2, p. 237, Feb. 2020, doi: 10.24198/PRH.V4I2.20108.

[3] Andhini Hastrida, "Proses Pengelolaan Media Sosial Pemerintah : Manfaat Dan Risiko - Yahoo Search Results," *Jurnal Penelitian Komunikasi dan Opini Publik*, 2021.

[4] "Profil Kementerian Keuangan." Accessed: Oct. 16, 2024. [Online]. Available: https://djpb.kemenkeu.go.id/kppn/tuban/id/profil/sejarah/114-profil/2814

[5] Cindy Mutia Annur, "Ada 27 Juta Pengguna Twitter di Indonesia, Terbanyak ke-4 Global." Accessed: Oct. 16, 2024. [Online]. Available: https://databoks.katadata.co.id/teknologi-telekomunikasi/statistik/75dd4b36866dc54/ada-27-juta-pengguna-twitter-di-indonesia-terbanyak-ke-4-global

[6] W. A. Anggraeni, F. Fahru Roji, and M. Alkautsar, "Analisis Sentimen Publik terhadap Kebijakan Insentif Perpajakan Dengan Pendekatan VADER (Valence Aware Dictionary And Sentiment Reasoner)," *Jurnal Proaksi*, vol. 10, no. 4, pp. 465–477, Oct. 2023, doi: 10.32534/JPK.V10I4.4732.

[7] "Perbandingan Metode Naïve Bayes dan *SVM* dalam Analisis Sentimen Netizen Twitter Terhadap Isu Kemenkeu | Aqsa | Buletin Sistem Informasi dan Teknologi Islam." Accessed: Oct. 16, 2024. [Online]. Available: https://jurnal.fikom.umi.ac.id/index.php/BUSITI/article/view/1824

[8] Y. Bambang Seran and S. Supatman, "ANALISIS SENTIMEN MASYARAKAT TERHADAP KINERJA KERJA PRESIDEN JOKO WIDODO ENGGUNAKAN ALGORITMA SUPPORT VECTOR MACHINE," *JATI (Jurnal Mahasiswa Teknik Informatika)*, vol. 8, no. 4, pp. 7190–7195, Jun. 2024, doi: 10.36040/JATI.V8I4.10171.

[9] D. Manuel, Y. Sinurat, D. E. Ratnawati, and D. W. Brata, “Analisis Sentimen Terhadap Kenaikan Cukai Rokok pada Media Sosial Twitter menggunakan Algoritma Naive Bayes Classifier,” *Jurnal Pengembangan Teknologi Informasi dan Ilmu Komputer*, vol. 7, no. 1, pp. 17–25, Feb. 2023, Accessed: Oct. 16, 2024. [Online]. Available: https://j-ptiik.ub.ac.id/index.php/j-ptiik/article/view/12089

[10] D. Darwis, E. Shintya Pratiwi, A. Ferico, and O. Pasaribu, “PENERAPAN ALGORITMA *SVM* UNTUK ANALISIS SENTIMEN PADA DATA TWITTER KOMISI PEMBERANTASAN KORUPSI REPUBLIK INDONESIA,” *Jurnal Ilmiah Edutic : Pendidikan dan Informatika*, vol. 7, no. 1, pp. 1–11, Nov. 2020, doi: 10.21107/EDUTIC.V7I1.8779.

[11] C. Zai and A. Rahman Isnain, “Komparasi Algoritma Naïve Bayes dan Support Vector Machine (*SVM*) pada Analisis Sentimen Capcut,” vol. 9, no. 1, p. 2024.

[12] Y. A. Singgalen, “Analisis Sentimen Pengunjung Pulau Komodo dan Pulau Rinca di Website Tripadvisor Berbasis CRISP-DM,” *Journal of Information System Research (JOSH)*, vol. 4, no. 2, pp. 614–625, Jan. 2023, doi: 10.47065/JOSH.V4I2.2999.

[13] R. G. Guntara, “Deteksi Atap Bangunan Berbasis Citra Udara Menggunakan Google Colab dan Algoritma Deep Learning YOLOv7,” *Jurnal Manajemen Sistem Informasi (JMASIF)*, vol. 2, no. 1, pp. 9–18, May 2023, doi: 10.59431/JMASIF.V2I1.156.

[14] Y. A. Singgalen, “Analisis Sentimen Pengunjung Pulau Komodo dan Pulau Rinca di Website Tripadvisor Berbasis CRISP-DM,” *Journal of Information System Research (JOSH)*, vol. 4, no. 2, pp. 614–625, Jan. 2023, doi: 10.47065/JOSH.V4I2.2999.

[15] F. Hasibuan, F. Hasibuan, W. Priatna, and T. S. Lestari, “Analisis Sentimen Terhadap Kementrian Perdagangan Pada Sosial Media Twitter Menggunakan Metode Naïve Bayes,” *Techno.Com*, vol. 21, no. 4, pp. 741–752, Nov. 2022, doi: 10.33633/tc.v21i4.6565.

[16] J. Da, C. Aruan, B. Rahayudi, and A. Ridok, “Analisis Sentimen Opini Masyarakat terhadap Pelayanan Rumah Sakit Umum Daerah menggunakan Metode Support Vector Machine dan Term Frequency - Inverse Document Frequency,” *Jurnal Pengembangan Teknologi Informasi dan Ilmu Komputer*, vol. 6, no. 5, pp. 2072–2078, Mar. 2022, Accessed: Oct. 16, 2024. [Online]. Available: https://j-ptiik.ub.ac.id/index.php/j-ptiik/article/view/10976

[17] A. Ahmad and W. Gata, “Sentimen Analisis Masyarakat Indonesia di Twitter Terkait Metaverse dengan Algoritma Support Vector Machine,” *Jurnal JTIK (Jurnal Teknologi Informasi dan Komunikasi)*, vol. 6, no. 4, pp. 548–555, Mar. 2022, doi: 10.35870/JTIK.V6I4.569.

[18] D. D. Putri, G. F. Nama, and W. E. Sulistiono, “Analisis Sentimen Kinerja Dewan Perwakilan Rakyat (DPR) Pada Twitter Menggunakan Metode Naive Bayes Classifier,” *Jurnal Informatika dan Teknik Elektro Terapan*, vol. 10, no. 1, p. 10, Jan. 2022, doi: 10.23960/jitet.v10i1.2262.

[19] R. S. Oetama, Y. Yanfi, M. M. Ikhsan Assiddiq Up, and S. Muhamad Isa, “Sentiment Analysis in Indonesian Trading using Lexicon-based and Support Vector Machine,” *ICCoSITE 2023 - International Conference on Computer Science, Information Technology and Engineering: Digital Transformation Strategy in Facing the VUCA and TUNA Era*, pp. 744–749, 2023, doi: 10.1109/ICCOSITE57641.2023.10127736.

[20] V. Phoan and J. Setiawan, “SENTIMENT ANALYSIS OF COMMENTS ON SEXUAL HARASSMENT IN COLLEGES ON FOUR POPULAR SOCIAL MEDIA,” *Journal of Multidisciplinary Issues*, vol. 2, no. 2, pp. 1–21, Aug. 2022, doi: 10.53748/JMIS.V2I2.33.

[21] A. Alsaeedi and M. Z. Khan, “A Study on Sentiment Analysis Techniques of Twitter Data,” *International Journal of Advanced Computer Science and Applications*, vol. 10, no. 2, pp. 361–374, 2023, doi: 10.14569/IJACSA.2019.0100248.

[22] R. Situmorang, U. M. Husni Tamyis, and L. S. Andar Muni, “ANALISIS SENTIMEN DESTINASI WISATA DI JAWABARAT PADA TWITTER MENGGUNAKAN ALGORITMA NAIVE BAYES CLASSIFIER,” *Simtek : jurnal sistem informasi dan teknik komputer*, vol. 8, no. 2, pp. 339–342, Oct. 2023, doi: 10.51876/SIMTEK.V8I2.287.

[23] D. R. Manalu, M. C. L. Tobing, and M. Yohanna, “ANALISIS SENTIMEN TWITTER TERHADAP WACANA PENUNDAAN PEMILU DENGAN METODE SUPPORT VECTOR MACHINE,” *METHOMIKA Jurnal Manajemen Informatika dan Komputerisasi Akuntansi*, vol. 6, no. 6, pp. 149–156, Oct. 2022, doi: 10.46880/JMIKA.VOL6NO2.PP149-156.

[24] N. Bin Aras, R. Risawandi, and L. Rosnita, “ANALISIS SENTIMEN KEPUASAN CUSTOMER TERHADAP EKSPEDISI TIKI, SICEPAT EXPRESS DAN NINJA EXPRESS MENGGUNAKAN ALGORITMA NAIVE BAYES,” *JOURNAL OF INFORMATICS AND COMPUTER SCIENCE*, vol. 9, no. 1, p. undefined-undefined, Apr. 2023, doi: 10.33143/JICS.V9I1.2943.

[25] S. Khairunnisa, A. Adiwijaya, and S. Al Faraby, “Pengaruh Text Preprocessing terhadap Analisis Sentimen Komentar Masyarakat pada Media Sosial Twitter (Studi Kasus Pandemi COVID-19),” *JURNAL MEDIA INFORMATIKA BUDIDARMA*, vol. 5, no. 2, p. 406, Apr. 2021, doi: 10.30865/MIB.V5I2.2835.

[26] D. ’ Rohannisa, F. Daud, B. Irawan, and A. Bahtiar, “PENERAPAN METODE NAIVE BAYES PADA ANALISIS SENTIMEN APLIKASI MCDONALDS DI GOOGLE PLAY STORE,” *JATI (Jurnal Mahasiswa Teknik Informatika)*, vol. 8, no. 1, pp. 759–766, Mar. 2024, doi: 10.36040/JATI.V8I1.8784.

[27] Irvandi, B. Irawan, and O. Nurdiawan, “NAIVE BAYES DAN WORDCLOUD UNTUK ANALISIS SENTIMEN WISATA HALAL PULAU LOMBOK,” *INFOTECH journal*, vol. 9, no. 1, pp. 236–242, May 2023, doi: 10.31949/INFOTECH.V9I1.5322.

[28] V. V. Ramaswamy *et al.*, “Beyond web-scraping: Crowd-sourcing a geographically diverse image dataset,” *arXiv.org*, 2023, doi: 10.48550/ARXIV.2301.02560.

[29] D. Elreedy, A. F. Atiya, and F. Kamalov, “A theoretical distribution analysis of synthetic minority oversampling technique (SMOTE) for imbalanced learning,” *Mach Learn*, vol. 113, no. 7, pp. 4903–4923, Jul. 2024, doi: 10.1007/S10994-022-06296-4.

[30] M. Heydarian, T. E. Doyle, and R. Samavi, “MLCM: Multi-Label Confusion Matrix,” *IEEE Access*, vol. 10, pp. 19083–19095, 2022, doi: 10.1109/ACCESS.2022.3151048.

[31] D. Krstinic, L. Seric, and I. Slapnicar, “Comments on ‘MLCM: Multi-Label Confusion Matrix,’” *IEEE Access*, vol. 11, pp. 40692–40697, 2023, doi: 10.1109/ACCESS.2023.3267672.

[32] I. Markoulidakis, G. Kopsiaftis, I. Rallis, and I. Georgoulas, “Multi-Class Confusion Matrix Reduction method and its application on Net Promoter Score classification problem,” *Petra*, pp. 412–419, Jun. 2021, doi: 10.1145/3453892.3461323.

[33] D. Normawati and S. A. Prayogi, “Implementasi Naïve Bayes Classifier Dan Confusion Matrix Pada Analisis Sentimen Berbasis Teks Pada Twitter,” *J-SAKTI (Jurnal Sains Komputer dan Informatika)*, vol. 5, no. 2, pp. 697–711, Sep. 2021, doi: 10.30645/J-SAKTI.V5I2.369.

[34] L. Lavazza and S. Morasca, “Comparing ϕ and the F-measure as performance metrics for software-related classifications,” *Empir Softw Eng*,

vol. 27, no. 7, pp. 1–38, Dec. 2022, doi: 10.1007/S10664-022-10199-2/TABLES/5.

[35] R. G. Wardhana, G. Wang, and F. Sibuea, "PENERAPAN MACHINE LEARNING DALAM PREDIKSI TINGKAT KASUS PENYAKIT DI INDONESIA," *Journal of Information System Management (JOISM)*, vol. 5, no. 1, pp. 40–45, Jul. 2023, doi: 10.24076/JOISM.2023V5I1.1136.

[36] R. Rachman, R. N. Handayani, and I. Artikel, "Klasifikasi Algoritma Naive Bayes Dalam Memprediksi Tingkat Kelancaran Pembayaran Sewa Teras UMKM," *Jurnal Informatika*, vol. 8, no. 2, pp. 111–122, Aug. 2021, doi: 10.31294/JI.V8I2.10494.

[37] A. Triawan and D. Lintang Melinda, "Penerapan Metode Naïve Bayes Untuk Rekomendasi Topik Tugas Akhir Berdasarkan Daftar Hasil Studi Mahasiswa di Perguruan Tinggi," *Teknois : Jurnal Ilmiah Teknologi Informasi dan Sains*, vol. 10, no. 2, pp. 58–70, Nov. 2020, doi: 10.36350/JBS.V10I2.91.

[38] R. Nanda, E. Haerani, S. K. Gusti, and S. Ramadhani, "Klasifikasi Berita Menggunakan Metode Support Vector Machine," *Jurnal Nasional Komputasi dan Teknologi Informasi (JNKTI)*, vol. 5, no. 2, pp. 269–278, Apr. 2022, doi: 10.32672/JNKTI.V5I2.4193.

[39] M. A. Hasanah, S. Soim, and A. S. Handayani, "Implementasi CRISP-DM Model Menggunakan Metode Decision Tree dengan Algoritma CART untuk Prediksi Curah Hujan Berpotensi Banjir," *Journal of Applied Informatics and Computing*, vol. 5, no. 2, pp. 103–108, Oct. 2021, doi: 10.30871/JAIC.V5I2.3200.

[40] "Proses Data Mining SEMMA – School of Information Systems." Accessed: Oct. 16, 2024. [Online]. Available: https://sis.binus.ac.id/2021/09/30/proses-data-mining-semma/

[41] "Proses Data Mining KDD – School of Information Systems." Accessed: Oct. 16, 2024. [Online]. Available: https://sis.binus.ac.id/2021/09/30/proses-data-mining-kdd/

[42] G. I. E. Soen, M. Marlina, and R. Renny, "Implementasi Cloud Computing dengan Google Colaboratory pada Aplikasi Pengolah Data Zoom Participants," *JITU : Journal Informatic Technology And Communication*, vol. 6, no. 1, pp. 24–30, Jun. 2022, doi: 10.36596/JITU.V6I1.781.

[43] A. Lavanya *et al.*, "Assessing the Performance of Python Data Visualization Libraries: A Review," *International Journal of Computer*

Engineering in Research Trends, vol. 10, no. 1, pp. 28–39, Jan. 2023, doi: 10.22362/IJCERT/2023/V10/I01/V10I0104.

[44] "Menkeu Tekankan Pentingnya Pelayanan Publik yang Lebih Baik - Nasional Tempo.co." Accessed: Oct. 16, 2024. [Online]. Available: https://nasional.tempo.co/read/1792358/menkeu-tekankan-pentingnya-pelayanan-publik-yang-lebih-baik

[45] P. Pasek, O. Mahawardana, G. M. A. Sasmita, P. Agus, and E. Pratama, "Analisis Sentimen Berdasarkan Opini dari Media Sosial Twitter terhadap 'Figure Pemimpin' Menggunakan Python," *JITTER : Jurnal Ilmiah Teknologi dan Komputer*, vol. 3, no. 1, p. 810, Jan. 2022, doi: 10.24843/JTRTI.2022.V03.I01.P17.

[46] Y. A. Singgalen, "Analisis Sentimen Pengunjung Pulau Komodo dan Pulau Rinca di Website Tripadvisor Berbasis CRISP-DM," *Journal of Information System Research (JOSH)*, vol. 4, no. 2, pp. 614–625, Jan. 2023, doi: 10.47065/JOSH.V4I2.2999.

[47] O. Firas, "A combination of SEMMA & CRISP-DM models for effectively handling big data using formal concept analysis based knowledge discovery: A data mining approach," *World Journal of Advanced Engineering Technology and Sciences*, vol. 8, no. 1, pp. 009–014, Jan. 2023, doi: 10.30574/WJAETS.2023.8.1.0147.

[48] A. Nur Azizah, M. Falach Asy'ari, I. Wisma Dwi Prastya, and D. Purwitasari, "Easy Data Augmentation untuk Data yang Imbalance pada Konsultasi Kesehatan Daring," *Jurnal Teknologi Informasi dan Ilmu Komputer*, vol. 10, no. 5, pp. 1095–1104, Oct. 2023, doi: 10.25126/JTIIK.20231057082.

[49] D. Ajeng Kristiyanti, S. Ady Sanjaya, V. Christio Tjokro, and J. Suhali, "Dealing imbalance dataset problem in sentiment analysis of recession in Dealing imbalance dataset problem in sentiment analysis of recession in Indonesia," *IAES International Journal of Artificial Intelligence (IJ-AI*, vol. 13, no. 2, pp. 2060–2072, 2024, doi: 10.11591/ijai.v13.i2.pp2060-2072.

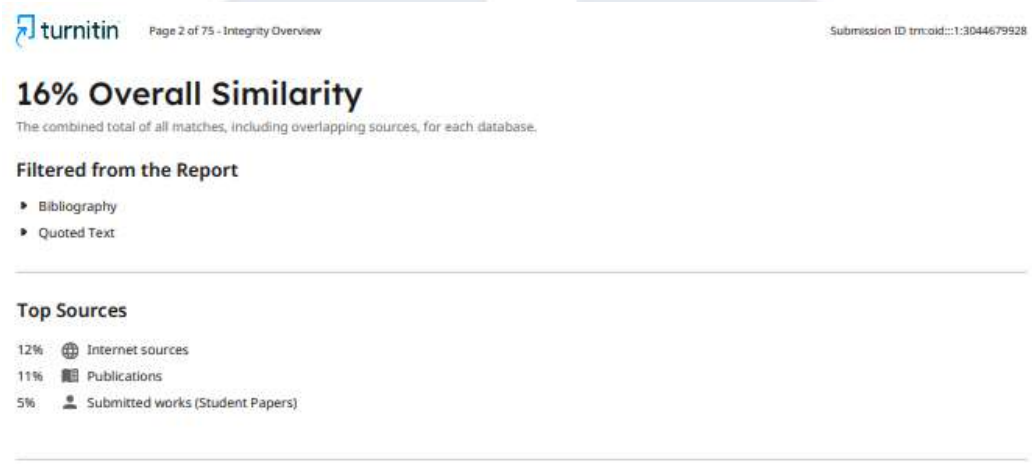
[50] Y. Asri, W. N. Suliyanti, D. Kuswardani, and M. Fajri, "Pelabelan Otomatis Lexicon Vader dan Klasifikasi Naive Bayes dalam menganalisis sentimen data ulasan PLN Mobile," *PETIR*, vol. 15, no. 2, pp. 264–275, Nov. 2022, doi: 10.33322/PETIR.V15I2.1733.

[51] K. Hulliyah, N. S. A. A. Bakar, A. R. Ismail, and M. O. Pratama, "A Benchmark of Modeling for Sentiment Analysis of the Indonesian Presidential Election in 2019," *2019 7th International Conference on Cyber*

and IT Service Management, CITSM 2019, Nov. 2019, doi: 10.1109/CITSM47753.2019.8965387.

LAMPIRAN

Lampiran A Turnitin



turnitin Page 2 of 75 - Integrity Overview Submission ID trn:oid:::1:3044679928

16% Overall Similarity

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

Filtered from the Report

- Bibliography
- Quoted Text

Top Sources

12% Internet sources
11% Publications
5% Submitted works (Student Papers)

Lampiran B Dataset Awal

created_at	full_text	lang	tweet_url	username	move_htr	Rmv_url	hv_usernar	mv_numb	v_punctua	Rmv_emoj	lowercase	tokenized
Sun Aug 0	Menteri K	in	https://x.(	arrahmah	Menteri K	Menteri K	Menteri K	Menteri K	Menteri K	Menteri K	menteri k	['menteri'
Fri Aug 04	Pemaksaa	in	https://x.(	kanimben	Pemaksaa	Pemaksaa	Pemaksaa	Pemaksaa	Pemaksaa	Pemaksaa	pemaksaa	['pemaksa
Tue Aug 0	Tidak han	in	https://x.(	Kemenke	Tidak han	Tidak han	Tidak han	Tidak han	Tidak han	Tidak han	tidak hany	['tidak', 'h
Fri Aug 04	Kemenke	in	https://x.(	Beritabar	Kemenke	Kemenke	Kemenke	Kemenke	Kemenke	Kemenke	kemenke	['kemenke
Wed Aug (	Halo #Sob	in	https://x.(	PPPK_Ken	Halo #Sob	Halo #Sob	Halo #Sob	Halo #Sob	Halo Soba	Halo Soba	halo soba	['halo', 'so
Tue Aug 0	Menkeu S	in	https://x.(	Metro_TV	Menkeu S	Menkeu S	Menkeu S	Menkeu S	Menkeu S	Menkeu S	menkeu s	['menkeu'
Tue Aug 0	Soal Trans	in	https://x.(	OposisiCe	Soal Trans	Soal Trans	Soal Trans	Soal Trans	Soal Trans	Soal Trans	soal trans	['soal', 'tra
Mon Aug (	Semoga ti	in	https://x.(	46iveMe	Semoga ti	Semoga ti	Semoga ti	Semoga ti	Semoga ti	Semoga ti	semoga ti	['semoga',
Wed Aug (	Transaksi	in	https://x.(	geloraco	Transaksi	Transaksi	Transaksi	Transaksi	Transaksi	Transaksi	transaksi j	['transaksi
Sat Aug 05	Tim Keme	in	https://x.(	liputan6d	Tim Keme	Tim Keme	Tim Keme	Tim Keme	Tim Keme	Tim Keme	tim keme	['tim', 'ker
Tue Aug 0	Perekono	in	https://x.(	Kemenke	Perekono	Perekono	Perekono	Perekono	Perekono	Perekono	perekono	['perekon
Wed Aug (	Direktorat	in	https://x.(	Metro_TV	Direktorat	Direktorat	Direktorat	Direktorat	Direktorat	Direktorat	direktorat	['direktora

Lampiran C Dataset Akhir

created_at	tweet_url	username	tweet	tweet_english	skor_sentimen	label_sentimen	Keyword
Wed Oct 18 0	https://twitter	geloraco	dpr tagih janji ma	DPR claims Mahfud	-0.1531	Negatif	Kemenkeu
Wed Oct 18 0	https://twitter	geloraco	apa kabar satgas (	What's up with the I	0	Netral	Kemenkeu
Thu Oct 26 02	https://twitter	liputan6dot	prabowo gibran ja	Prabowo Gibran pro	0.3818	Positif	Kemenkeu
Wed Oct 18 1	https://twitter	Kanseulir	apa kabar satgas (	what's up with the M	-0.4588	Negatif	Kemenkeu
Wed Oct 18 2	https://twitter	hipohan	mahfud pilih tepa	Mahfud chose the ri	-0.5696	Negatif	Kemenkeu
Sat Oct 14 09:	https://twitter	ajies4ra	bagi rb rice cooke	for RB free rice cook	0.9229	Positif	Kemenkeu
Wed Oct 18 0	https://twitter	HermanBud	potret buram era	A blurry portrait of I	0.836	Positif	Kemenkeu
Thu Oct 26 14	https://twitter	madHink	tau apa sebab dui	Do you know why m	0	Netral	Kemenkeu
Mon Oct 23 0	https://twitter	MRifkiano	kemenkeu ajar se	Ministry of Finance	0	Netral	Kemenkeu
Sun Oct 29 09	https://twitter	ApriyantoLu	kemenkeu sindir	Ministry of Finance	0.8271	Positif	Kemenkeu
Wed Oct 25 0	https://twitter	0jenry1	gosa jauh cek inst	gosa far, check the t	0.7501	Positif	Kemenkeu
Tue Oct 17 03	https://twitter	WajahPribu	perintah lalu mer	last order from the I	0.8225	Positif	Kemenkeu
Tue Oct 24 16	https://twitter	ecosocright	yah salah analisa	Yes, it's wrong to an	-0.1027	Negatif	Kemenkeu
Mon Oct 23 1(	https://twitter	NcangBeni	siape kemaren ng	Who said yesterday	0.2168	Positif	Kemenkeu
Tue Oct 24 10	https://twitter	pejabrut	menteri mana ko	Which minister com	-0.7783	Negatif	Kemenkeu
Fri Oct 27 05:	https://twitter	BlueNeo81	apaaa kemenkeu	Is it true that the Mi	-0.0116	Netral	Kemenkeu
Mon Oct 23 1	https://twitter	UpsOdd	apa event tahun r	What non-Ministry (	0	Netral	Kemenkeu
Wed Oct 25 1(	https://twitter	Metro_TV	presiden joko wi	President Joko Wid	0.6705	Positif	Kemenkeu
Wed Oct 18 2	https://twitter	NikoAkmal	kasus kemenkeu	The Ministry of Fina	0	Netral	Kemenkeu

Lampiran D Form Bimbingan

FORMULIR KONSULTASI SKRIPSI – FAKULTAS TEKNIK & INFORMATIKA

Dosen Pembimbing : **Dr. Santo Fernandi Wijaya**
Jurusan : **Sistem Informasi**
Semester : **9**
Nama : **Muhammad Alviansyah Gustama**
NIM : **00000051526**

Tanggal Konsultasi	Agenda/Pokok Bahasan	Saran Perbaikan
August 30, 2024	**Menjelaskan terkait skirpsi**	**Mencari topik dan jurnal**
September 06, 2024	**Update Progress skripsi**	**Mencari jurnal dan Membuat BAB I – BAB III**
September 13, 2024	**Revisi laporan BAB I - III dan SLR**	**Perbaiki laporan BAB IV – BAB V**
September 18, 2024	**Revisi laporan BAB I - III dan SLR**	**Perbaiki BAB I - III dan SLR**
September 27, 2024	**Menyampaikan hasil revisi SLR dan BAB IV – BAB V**	**Lanjutkan penulisan SLR dan BAB IV – BAB V**
October 04, 2024	**Menyampaikan progress laporan BAB IV – BAB V**	**Perbaiki laporan BAB IV – BAB V**
October 05, 2024	**Menyampaikan progress laporan BAB IV-BAB V**	**Perbaiki laporan BAB IV – BAB V**
October 07, 2024	**Menyampaikan progress laporan BAB IV-BAB V**	**Perbaiki laporan BAB IV – BAB V**

Catatan : Form ini wajib dibawa pada saat konsultasi & dilampirkan didalam skripsi (**Minimal 8 kali Konsultasi**)

Tangerang, Oktober 2024

Dr. Santo Fernandi Wijaya, S.Kom., M.M

FORMULIR KONSULTASI SKRIPSI – FAKULTAS TEKNIK & INFORMATIKA

Dosen Pembimbing : **Iwan Prasetiawan, S.Kom., M.M**

Jurusan : **Sistem Informasi**

Semester : **9**

Nama : **Muhammad Alviansyah Gustama**

NIM : **00000051526**

Tanggal Konsultasi	Agenda/Pokok Bahasan	Saran Perbaikan
September 07, 2024	**Progress laporan skripsi**	**Memperbaiki laporan skrispsi**
September 21, 2024	**Bimbingan laporan skripsi dan update olah data**	**Memperbaiki laporan dan olah data**
September 23, 2024	**Update olah data**	**Memperbaiki olah data**
October 05, 2024	**Update final laporan skripsi**	**Memperbaiki laporan skrispsi**

Catatan : Form ini wajib dibawa pada saat konsultasi & dilampirkan didalam skripsi (**Minimal 4 kali Konsultasi**)

Tangerang, Oktober 2024

Iwan Prasetiawan, S.Kom., M.M