ANALISIS SENTIMEN X TERHADAP KEAMANAN DATA MENGGUNAKAN ALGORITMA NAÏVE BAYES CLASSIFIER: STUDI KASUS PUSAT DATA NASIONAL SEMENTARA (PDNS)

Disusun Oleh:

N a m a : Bagus Anugrah Prasetyo

NIM : 17523224

PROGRAM STUDI INFORMATIKA – PROGRAM SARJANA

FAKULTAS TEKNOLOGI INDUSTRI

UNIVERSITAS ISLAM INDONESIA

2024

HALAMAN PENGESAHAN DOSEN PEMBIMBING

ANALISIS SENTIMEN X TERHADAP KEAMANAN DATA MENGGUNAKAN ALGORITMA NAÏVE BAYES CLASSIFIER: STUDI KASUS PUSAT DATA NASIONAL SEMENTARA (PDNS)

TUGAS AKHIR

Disusun Oleh:

N a m a : Bagus Anugrah Prasetyo

NIM : 17523224

Yogyakarta, 17 Juli 2024

Pembimbing,

(Chandra Kusuma Dewa, S.Kom., M.Cs., Ph.D.)

HALAMAN PENGESAHAN DOSEN PENGUJI

ANALISIS SENTIMEN X TERHADAP KEAMANAN DATA MENGGUNAKAN ALGORITMA NAÏVE BAYES CLASSIFIER: STUDI KASUS PUSAT DATA NASIONAL SEMENTARA (PDNS)

TUGAS AKHIR

Telah dipertahankan di depan sidang penguji sebagai salah satu syarat untuk memperoleh gelar Sarjana Komputer dari Program Studi Informatika – Program Sarjana di Fakultas Teknologi Industri Universitas Islam Indonesia

Yogyakarta, 26 Juli 2024

Tim Penguji

Chandra Kusuma Dewa, S.Kom., M.Cs.,Ph.D. _____

Anggota 1

Hari Setiaji, S.Kom., M.Eng. _____

Anggota 2

Kholid Haryono, S.Kom., M.Kom. _____

Mengetahui,

Ketua Program Studi Informatika – Program Sarjana

Fakultas Teknologi Industri

Universitas Islam Indonesia

(Ir. Dhomas Hatta Fudholi, S.T., M.Eng., Ph.D.)

HALAMAN PERNYATAAN KEASLIAN TUGAS AKHIR

Yang bertanda tangan di bawah ini:

Nama : Bagus Anugrah Prasetyo

NIM : 17523224

Tugas akhir dengan judul:

ANALISIS SENTIMEN X TERHADAP KEAMANAN DATA MENGGUNAKAN ALGORITMA NAÏVE BAYES CLASSIFIER: STUDI KASUS PUSAT DATA NASIONAL SEMENTARA (PDNS)

Menyatakan bahwa seluruh komponen dan isi dalam tugas akhir ini adalah hasil karya saya sendiri. Apabila di kemudian hari terbukti ada beberapa bagian dari karya ini adalah bukan hasil karya sendiri, tugas akhir yang diajukan sebagai hasil karya sendiri ini siap ditarik kembali dan siap menanggung risiko dan konsekuensi apapun.

Demikian surat pernyataan ini dibuat, semoga dapat dipergunakan sebagaimana mestinya.

Yogyakarta, 17 Juli 2024



(Bagus Anugrah Prasetyo)

HALAMAN PERSEMBAHAN

Assalamualaikum Warahmatullahi Wabarakatuh.

Dengan menyebut nama Allah *Subhanahu wa ta'ala*, Yang Maha Pengasih lagi Maha Penyayang. Segala puji bagi Allah di mana dengan nikmat dan karunia yang telah diberikan oleh-Nya, tentu telah banyak hal berupa kebaikan, kemudahan dan kelancaran dalam menjalani kelangsungan kehidupan ini. Tak lupa shalawat beriring salam kita ucapakan bagi baginda Rasulullah Nabi Muhammad *Shalallaahu Alaihi Wassalam*. Izinkan saya menuliskan sebuah persembahan kali ini sebagai bentuk rasa syukur dan terimakasih dalam proses menyelesaikan karya ilmiah tugas akhir sebagai salah satu syarat kelulusan dalam menempuh pendidikan Program Sarjana Informatika di Universitas Islam Indonesia. Adapun Tugas akhir ini saya persembahkan untuk keluarga saya, terkhusus kedua orang tua dan adik tersayang. Keluarga yang selalu ada dan tak pernah luput selalu mendoakan, menasehati, dan mengajarkan saya.

Semoga Allah balas segala amal kebaikan kita semua, *Aamiin Ya Rabbal Alamin*

Wassalaikumsalam Warahmatullahi Wabarakatuh.

HALAMAN MOTO

"Libatkanlah Allah selalu dimanapun dan kapanpun, niatkan semua kegiatan baik atas dasar mencari ridho-Nya, atas izin Allah keberkahan akan selalu menyertaimu, bahkan disaat titik terendahpun, Allah tidak akan membiarkan hambanya dalam kesulitan, percaya dan selalu berprasangka baik ya sama Allah"- BAP

"Apakah kamu mengira bahwa kamu akan masuk surga, padahal belum datang kepadamu (cobaan) sebagaimana halnya orang-orang terdahulu sebelum kamu? Mereka ditimpa oleh malapetaka dan kesengsaraan, serta digoncangkan (dengan bermacam-macam cobaan) sehingga berkatalah Rasul dan orang-orang yang beriman bersamanya: "Bilakah datangnya pertolongan Allah?" Ingatlah, sesungguhnya pertolongan Allah itu amat dekat." (Q.S. Al Baqarah: 214)

"...Cukuplah Allah menjadi penolong bagi kami dan Dia sebaik-baik pelindung" (Q.S. Ali'Imran: 173)

"Hanya kepada Engkaulah kami menyembah dan hanya kepada Engkaulah kami mohon pertolongan" (Q.S. Al Fatihah: 5)

"Jadikanlah sabar dan shalat sebagai penolongmu" (Q.S. Al Baqarah: 45)

"...Barang siapa bertakwa kepada Allah, niscaya Dia akan mengadakan baginya jalan keluar. Dan memberinya rezeki dari arah yang tiada disangka-sangkanya. Dan barang siapa yang bertawakal kepada Allah, niscaya Allah akan mencukupkan (keperluan)nya.

Sesungguhnya Allah melaksanakan urusan yang (dikehendaki)-Nya, sesungguhnya Allah telah mengadakan ketentuan bagi tiap-tiap sesuatu" (Q.S. At Talaq: 2-3)

"Karena sesungguhnya sesudah kesulitan itu ada kemudahan, Sesungguhnya sesudah kesulitan itu ada kemudahan" (Q.S. Al Insyirah: 5-6)

KATA PENGANTAR

Assalaamu'alaykum warahmatullaahi wabarakatuh.

Alhamdulillaahirabbil 'alamiin. Segala puji bagi Allah *Subhanahu wa ta'ala* atas segala nikmat, karunia, hidayah, dan pertolongan-Nya, sehingga tugas akhir berjudul **"Analisis Sentimen X Terhadap Keamanan Data Menggunakan Agoritma Naïve Bayes Classifier: Studi Kasus Pusat Data Nasional Sementara (PDNS)"** ini dapat terselesaikan sebagaimana mestinya.

Tugas akhir ini dibuat sebagai salah satu syarat yang harus dipenuhi untuk menyelesaikan pendidikan jenjang sarjana dari Program Studi Informatika, Fakultas Teknologi Industri, Universitas Islam Indonesia.

Proses penyelesaian tugas akhir ini kurang dan lebihnya tentu ditemui berbagai rintangan dan hambatan oleh penulis. Namun, berbagai macam rintangan dan hambatan tersebut akhirnya dapat dilewati dikarenakan dukungan dari berbagai pihak. Maka dari itu penulis ingin mengucapkan terima kasih kepada:

1. Almh. nenek dan mbah saya, almh. Nenek Yulinur dan almh. Mbah Haji Qomariah, semasa hidupnya selalu menyemangati saya, mendoakan saya dan mendukung penuh saya untuk bisa berkuliah disini.
2. Kedua orang tua tercinta saya, Mama Yurisnel dan Bapak Bambang Resmiyadi, senantiasa memberikan doa dan dukungan yang sangat berarti selama ini.
3. Adik kesayangan saya, Keyza Indah Putri.
4. Bapak Fathul Wahid, S.T., M.Sc., Ph.D., selaku Rektor Universitas Islam Indonesia.
5. Bapak Prof. Dr. Ir. Hari Purnomo., M.T., IPU, ASEAN.Eng., selaku Dekan Fakultas Teknologi Industri, Universitas Islam Indonesia.
6. Bapak Dr. Raden Teduh Dirgahayu, S.T., M.Sc., selaku Ketua Jurusan Informatika, Universitas Islam Indonesia.
7. Bapak Ir. Dhomas Hatta Fudholi, S.T., M.Eng., Ph.D., selaku Ketua Program Studi Informatika – Program Sarjana, Universitas Islam Indonesia.
8. Bapak Chandra Kusuma Dewa, S.Kom., M.Cs., Ph.D., selaku dosen pembimbing saya yang telah berkenan membimbing, mengajari, mengarahkan dan memberikan ilmu, saran, serta masukan selama proses bimbingan tugas akhir ini.
9. Ibuk Fayruz Rahma, S.T., M.Eng., selaku dosen pembimbing akademik saya yang telah memberikan ilmu, saran, semangat, dan selalu perhatian selama ini.

10. Bapak Hari Setiaji, S.Kom., M.Eng., selaku dosen penguji satu saya yang telah bersedia menguji, mengarahkan, dan memberikan saran serta masukan.
11. Bapak Kholid Haryono, S.Kom., M.Kom., selaku dosen penguji dua saya yang telah bersedia menguji, mengarahkan, dan memberikan saran serta masukan.
12. Seluruh dosen dan staff Jurusan Informatika Universitas Islam Indonesia yang telah banyak memberikan ilmu bermanfaat serta membantu selama proses akademisi perkuliahan.
13. Seluruh keluarga besar saya, terutama makdang saya Bapak Jefrizon dan makwo saya Ibuk Pida.
14. Ibuk dan bapak kos saya, Ibuk Tin dan Bapak Agus yang selalu mendukung dan banyak membantu saya selama disini.
15. Seluruh teman yang selalu mendukung dan memberi semangat.
16. Pihak-pihak lain yang telah membantu kelancaran proses penelitian ini yang tidak dapat disebutkan satu persatu.
17. Terakhir saya ingin mengucapkan terima kasih kepada diri saya sendiri sebagai bentuk apresiasi karena telah berjuang selama ini dalam berbagai keadaan suka dan duka, atas izin Allah *Subhanahu wa ta'ala* saya bisa melalui semuanya dengan baik dan benar. *Alhamdulillaahirabbil 'alamiin.*

Dengan selesainya penulisan laporan tugas akhir ini, besar harapan penulis agar karya ini dapat memberikan manfaat dan dapat dikembangkan di masa yang akan datang.

Yogyakarta, 13 Agustus 2024



(Bagus Anugrah Prasetyo)

ABSTRACT

Data security has become a primary concern alongside rapid technological advancements. Managing and protecting data, especially on a large scale with thorough analysis, is crucial to prevent losses and safeguard the confidentiality of users' important documents. Modern technology and societal lifestyles support various studies enabling computers to understand, manipulate, and interpret human language. This research aims to analyze sentiments within application X regarding data security, particularly concerning incidents of breaches at the Temporary National Data Center (PDNS), using the Naïve Bayes Classifier algorithm. The data consists of Indonesian-language tweets collected within specific time frames using the keywords "pdns" and "pdns 2". This method is expected to provide deeper insights into public perceptions of data security context, including sentiment classifications such as positive, negative, and neutral visualized through graphs and word clouds. The findings of this study are anticipated to contribute to a better understanding of sentiment analysis on data security within application X and serve as a foundation for further research in related fields.

Keywords: Data, Security of Data, Sentiment Analysis

GLOSARIUM

Algoritma	langkah-langkah runtun menyelesaikan suatu permasalahan sederhana hingga kompleks dalam pemrograman komputer
Accuracy	salah satu metrik evaluasi yang umum digunakan untuk mengukur performa dari sebuah model atau sistem klasifikasi teks
Keyword	kata kunci dengan makna atau arti tertentu
Library	sekumpulan modul atau fungsionalitas yang digunakan untuk mempermudah proses dalam membangun sebuah sistem atau aplikasi
Machine Learning	cabang dari kecerdasan buatan *(artificial intelligence)*, berkaitan dengan pengembangan sistem yang dapat belajar dari data secara otomatis dan meningkatkan kinerjanya seiring waktu tanpa perlu secara eksplisit diprogram ulang
Precision	salah satu metrik evaluasi yang digunakan dalam konteks klasifikasi untuk mengukur seberapa baik model dapat mengidentifikasi teks
Pseudocode	tatanan cara penulisan algoritma bahasa pemrograman
Ransomware	jenis perangkat lunak berbahaya *(malware)* yang dirancang untuk mengenkripsi data atau mengunci akses ke sistem komputer atau *file* korban
Recall	salah satu metrik evaluasi yang digunakan dalam *information retrieval* (pemulihan informasi) dan *machine learning* untuk mengevaluasi kinerja model atau sistem

DAFTAR ISI

HALAMAN JUDUL i
HALAMAN PENGESAHAN DOSEN PEMBIMBING ii
HALAMAN PENGESAHAN DOSEN PENGUJI iii
HALAMAN PERNYATAAN KEASLIAN TUGAS AKHIR iv
HALAMAN PERSEMBAHAN v
HALAMAN MOTO vi
KATA PENGANTAR vii
ABSTRACT ix
GLOSARIUM x
DAFTAR ISI xi
DAFTAR TABEL xii
DAFTAR GAMBAR xiii
BAB I PENDAHULUAN 1
1.1 Latar Belakang 1
1.2 Rumusan Masalah 3
1.3 Tujuan Penelitian 3
1.4 Batasan Penelitian 3
1.5 Manfaat Penelitian 3
1.6 Sistematika Penulisan 4
BAB II LANDASAN TEORI 5
2.1 Penelitian Terdahulu 5
2.2 Analisis Sentimen 7
2.3 Naïve Bayes Classifier 8
2.4 Aplikasi X 9
2.5 Text Mining 9
2.6 Text Pre-Processing 9
BAB III METODOLOGI PENELITIAN 11
3.1 Langkah Penelitian 11
3.2 Uraian Penelitian 11
3.2.1 Early Stages 11
3.2.2 Data Analysis Process 16
3.2.3 Last Stage 19
BAB IV HASIL DAN PEMBAHASAN 20
4.1 Crawling Data 20
4.2 Processing Raw Data 25
4.3 Preprocessing 28
4.4 Translate 32
4.5 Labelling 33
4.6 Sentiment Classification 34
4.7 Visualization 39
4.8 Pembahasan 42
BAB V PENUTUP 45
5.1 Kesimpulan 45
5.2 Saran 45
DAFTAR PUSTAKA 46
LAMPIRAN 48

DAFTAR TABEL

Tabel 2.1 Penelitian Terdahulu 6
Tabel 4.2 Hasil Perolehan Analisis NBC 42
Tabel 4.3 Hasil Perolehan Analisis TextBlob 42

DAFTAR GAMBAR

Gambar 3.1 Flowchart Metode Penelitian 11
Gambar 3.2 Flowchart Early Stages 11
Gambar 3.3 Tweet Pengguna X 12
Gambar 3.4 Pos Berita 12
Gambar 3.5 Kumpulan Pos Berita 13
Gambar 3.6 Flowchart Data Analysis Process 16
Gambar 3.7 Flowchart Last Stage 19
Gambar 4.8 Tampilan Home Aplikasi X 21
Gambar 4.9 Tampilan Inspect X 21
Gambar 4.10 Tampilan Application Inspect X 22
Gambar 4.11 Tampilan Cookies Url X 22
Gambar 4.12 Pseudocode Program Install Pandas dan Node Js 23
Gambar 4.13 Proses Install Package Berhasil 23
Gambar 4.14 Pseudocode Program Crawling Data 24
Gambar 4.15 Proses Scrolling Pada Crawling Data 24
Gambar 4.16 Proses Crawling Data Selesai 24
Gambar 4.17 Pseudocode Program Menampilkan Hasil Crawling Data 25
Gambar 4.18 Tampilan Hasil Akhir Crawling Data 25
Gambar 4.19 Pseudocode Program Categorizing Data 26
Gambar 4.20 Hasil Categorizing Data 26
Gambar 4.21 Pseudocode Program dan Hasil Duplicates 27
Gambar 4.22 Pseudocode Program Data Cleaning 27
Gambar 4.23 Hasil Data Cleaning 27
Gambar 4.24 Pseudocode Program Case Folding 28
Gambar 4.25 Hasil Case Folding 28
Gambar 4.26 Pseudocode Program Normalized 29
Gambar 4.27 Hasil Normalized 29
Gambar 4.28 Proses Install Package Sastrawi 29
Gambar 4.29 Pseudocode Program Stopword 30
Gambar 4.30 Hasil Stopword 30
Gambar 4.31 Pseudocode Program Tokenized 30
Gambar 4.32 Hasil Tokenized 31

Gambar 4.33 Pseudocode Program Stemming 31
Gambar 4.34 Hasil Stemming.....32
Gambar 4.35 Proses Install Library Translate 32
Gambar 4.36 Data Hasil Preprocessing 33
Gambar 4.37 Pseudocode Program Translate.....33
Gambar 4.38 Proses Labelling.....34
Gambar 4.39 Pseudocode Program Install Library.....34
Gambar 4.40 Proses Install Library Berasil.....35
Gambar 4.41 Proses Import Library dan Download Package Punkt 35
Gambar 4.42 Pseudocode Program Klasifikasi Sentimen TextBlob 36
Gambar 4.43 Hasil Klasifikasi Sentimen TextBlob.....36
Gambar 4.44 Hasil Labelling Klasifikasi Sentimen TextBlob 37
Gambar 4.45 Pseudocode Program Mempersiapkan Dataset dan Data Train 37
Gambar 4.46 Pseudocode Program Proses Cek Akurasi dan Klasifikasi Sentimen NBC.....38
Gambar 4.47 Hasil Beserta Perbandingan Klasifikasi Sentimen NBC.....38
Gambar 4.48 Pseudocode Program Visualisasi Wordcloud 39
Gambar 4.49 Hasil Visualisasi Wordcloud.....39
Gambar 4.50 Pseudocode Program Visualisasi Bar Chart TextBlob 40
Gambar 4.51 Hasil Visualisasi Bar Chart TextBlob.....40
Gambar4.52 Pseudocode Program Visualisasi Bar Chart NBC 41
Gambar 4.53 Hasil Visualisasi Bar Chart NBC.....41
Gambar 4.54 Pseudocode Evaluasi Sentimen.....43
Gambar 4.55 Contoh Hasil Evaluasi Dari Tiap Sentimen 44

BAB I
PENDAHULUAN

1.1 Latar Belakang

Keamanan data menjadi sebuah tantangan yang timbul dari pesatnya perkembangan teknologi saat ini. Pengumpulan data dalam skala yang sangat besar beserta analisis yang mendalam menjadi sebuah pertimbangan dalam mengelola dan melindungi data penggunanya. Oleh sebab itu keamanan data sangat diperlukan sebagai bentuk pencegahan dari tindakan-tindakan yang dapat merugikan sekaligus untuk melindungi data berupa dokumen penting yang terkadang bersifat rahasia bagi penggunanya (Sabian, 2023).

Pusat Data Nasional Sementara (PDNS) merupakan infrastruktur penting dalam penyimpanan dan pengolahan data nasional. Beberapa waktu yang lalu telah terjadi sebuah peristiwa peretasan Pusat Data Nasional Smentara (PDNS) oleh sekelompok *hacker* yang menamakan dirinya *Brain Chiper*. Peretasan terjadi ketika *data server* dari PDNS terkena *ransomeware lockbit 3.0. Ransomeware ini* langsung menyerang ke sistem pusat data yang mengakibatkan terganggunya beberapa aspek dan bidang keamanan di berbagai layanan publik dari kementrian beserta lembaga yang terintegrasi. Dilansir dari media berita *tempo.co*, beberapa dampak yang ditimbulkan dari serangan *ransomeware lockbit 3.0* diantaranya berdampak terhadap layanan imigrasi yang mengalami *down*, hal ini tentunya mengharuskan layanan untuk pindah ke *web* Amazon agar dapat tetap beroperasi, kemudian serangan ini juga berdampak terhadap hilangnya data pendaftar Kartu Indonesia Pintar Kuliah (KIP Kuliah), dan mengakibatkan kendala bagi ratusan instansi pemerintahan beserta puluhan layanan dari Kementrian Pendidikan dan Kebudayaan (Wuragil, 2024).

Kasus peretasan *Ransomware* pada PDNS 2 merupakan suatu kejadian yang mengkhawatirkan dan membutuhkan perhatian serius dari pihak terkait keamanan data negara khususnya Kementrian Komunikasi dan Informatika (KemKominfo) Indonesia. Peretasan ini merupakan jenis serangan dimana peretas mengenkripsi data pada sistem pusat server dan menuntut tebusan kepada pihak terkait agar data tersebut bisa dikembalikan. Hal ini tentunya memiliki dampak dan pengaruh besar terhadap pandangan masyarakat di Indonesia. Beragam tanggapan muncul di berbagai *platform social media* dan ramai menjadi perbincangan hangat terkait permasalahan ini. Salah satu diantaranya yaitu trending topik dengan tagar *#Blokirkominfo* di aplikasi X.

Perkembangan teknologi dan gaya hidup masyarakat dewasa ini sangat mendukung terhadap usaha dari berbagai bidang keilmuan yang dilakukan para peneliti di universitas dan industri (Heryadi & Wahyono, 2021). Salah satu dari berbagai bidang keilmuan itu ialah *Natural Language Processing (NLP)*. NLP merupakan bagian dari *machine learning* yang memungkinkan komputer mengolah bahasa manusia, sehingga komputer dapat memahami, memanipulasi, dan menginterpretasikan. Pengolahan bahasa manusia itu sendiri tidaklah luput dari data yang dapat berbentuk suara maupun teks. Salah satu pemanfaatan dari nlp yaitu menganalisis suatu data berupa sentimen atau biasanya disebut dengan nama analisis sentimen *(sentiment analysis)* (Lexalytics, 2021).

Analisis sentimen merupakan suatu deteksi komputasi dan pembelajaran berdasarkan pendapat *(opini)*, sentimen, emosi, dan subjektivitas di dalam teks (He et al., 2015). Analisis sentimen berhubungan dengan ekstraksi otomatis berupa opini positif atau negatif, sehingga memberikan bobot pada teks dari tiap entitas, topik, maupun tema tiap kalimat atau frasa (Lexalytics, 2021). Analisis sentimen juga melakukan proses memahami, mengekstrak dan mengolah data tekstual secara otomatis untuk mendapatkan informasi sentimen yang terkandung dalam suatu kalimat pendapat *(opini)* (Buntoro, 2017). Implementasi analisis sentimen tentunya membutuhkan suatu proses berupa klasifikasi teks. Beberapa algoritma yang sering digunakan dalam mengklasifikasikan teks yaitu *Naïve Bayes Classifier, Support Vector Machine, Decision Trees,* dan *Deep Learning* (Cahyaningtiyas, 2019).

Naïve Bayes Classifier merupakan salah satu bentuk algoritma klasifikasi teks yang cukup popular digunakan dan banyak diterapkan dalam bidang *text mining* dan *sentiment analysis.* Penelitian sebelumnya juga memaparkan beberapa parameter seperti nilai akurasi, presisi, dan recall menunjukkan kapabilitas yang tinggi (Nurfadillah, 2023), sehingga algoritma *Naïve Bayes Classifier* menjadi salah satu algoritma pilihan dalam mengolah analisis sentimen.

Berdasarkan pemaparan dari uraian permasalahan yang sudah dijelaskan, penulis merasa perlu untuk melakukan penelitian terhadap analisis sentimen keamanan data dengan studi kasus Pusat Data Nasional Sementara (PDNS). Adapun maksud dan tujuan dari penulis melakukan analisi sentimen ini untuk dapat mengetahui dan mengkategorisasikan sentimen publik yang beredar di kalangan masyarakat pada aplikasi X ke dalam beberapa kelompok sentimen positif, netral atau negatif.

1.2 Rumusan Masalah

Berdasarkan pemaparan latar belakang, pada penelitian ini dirumuskan permasalahan bagaimana analisis sentimen pada aplikasi X terhadap keamanan data berdasarkan permasalahan mengenai kejadian peretasan Pusat Data Nasional Sementara (PDNS) menggunakan algoritma *Naïve Bayes Classifier*.

1.3 Tujuan Penelitian

Berdasarkan rumusan masalah, adapun tujuan yang ingin dicapai dalam penelitian ini berupa perolehan proses dan hasil analisis sentimen pada aplikasi X terhadap keamanan data berdasarkan permasalahan mengenai kejadian peretasan Pusat Data Nasional Sementara (PDNS) menggunakan algoritma *Naïve Bayes Classifier*. Adapun sasaran dari tujuan penelitian yakni mengedukasi pembaca terhadap analisis sentimen keamanan data yang telah didapatkan dan sebagai bahan pembelajaran maupun referensi bagi penelitian selanjutnya.

1.4 Batasan Penelitian

Batasan masalah dalam penelitian ini adalah:

a) Data *tweet* berbahasa Indonesia.

b) Algoritma klasifikasi data menggunakan *Naïve Bayes Classifier*.

c) *Crawling data* sentimen postingan X dikumpulkan dalam rentang waktu setelah penyerangan peretasan pusat data nasional sementara terjadi yakni dimulai dari tanggal 01 Juli 2024 hingga 11 Juli 2024. Total penggunaan data dibatasi sebesar 150-an *tweet*, hal ini berkaitan dengan proses *translate* bawaan yang didukung oleh *Google Collab* sangat terbatas.

d) Pengambilan data menggunakan kata kunci *"pdns" dan "pdns 2"*.

e) Hasil akhir dari penelitian berupa klasifikasi positif, negatif, dan netral dengan persentase tiap sentimen, visualisasi yang akan ditampilkan berupa *grafik diagram* dan *wordcloud*.

1.5 Manfaat Penelitian

Manfaat yang diperoleh dalam penelitian ini adalah:

a) Memberikan informasi mengenai analisis sentimen terhadap keamanan data pada aplikasi X beserta perolehan klasifikasi yang akan didapatkan.

b) Memberikan informasi bagaimana proses dan gambaran perfoma dari metode

yang digunakan.

c) Sebagai bahan pembelajaran untuk lingkup penelitian selanjutnya dengan kemiripan kasus serupa yaitu analisis sentimen keamanan data.

1.6 Sistematika Penulisan

Sistematika penulisan laporan penelitian ini sebagai berikut:

BAB I Pendahuluan

Bab ini membahas bagaimana latar belakang penelitian, merumuskan masalah, batasan masalah penelitian, tujuan penelitian, manfaat penelitian, dan sistematika penulisan penelitian.

BAB II Landasan Teori

Bab ini membahas hal-hal dasar berisi pemaparan teori berkaitan dengan analisis sentimen beserta teori pendukung yang digunakan.

BAB III Metodologi Penelitian

Bab ini membahas bagaimana proses penelitian disusun beserta langkah-langkah yang dilakukan dalam penelitian analisis sentimen ini.

BAB IV Hasil dan Pembahasan

Bab ini menjelaskan dan memaparkan pembahasan hasil dari penelitian tentang analisis sentimen dengan algoritma *Naïve Bayes Classifier* berdasarkan metodologi penelitian yang telah dibuat.

BAB V Penutup

Bab ini memaparkan kesimpulan berupa rangkuman dari hasil penelitian dan beberapa saran dengan harapan dapat menjadi bahan pembelajaran penelitian di masa yang akan datang.

BAB II
LANDASAN TEORI

2.1 Penelitian Terdahulu

Penelitian terdahulu menjadi dasar acuan bagi penulis dalam menyusun laporan penelitian tugas akhir ini, sekaligus memunculkan ide dan menambah wawasan penulis terkait topik dan pokok permasalahan penelitian serupa. Adapun pemaparan dari penelitian dapat dilihat pada Tabel 2.1.

No	Penulis	Judul	Metode Penelitian	Akurasi
1	Joko Ade Nursiyono, Qorinul Huda	Analisis Sentimen Twitter Terhadap Perlindungan Data Pribadi Dengan Pendekatan Machine Learning	*Naïve Bayes Classifier, Random Forest*	Tingkat akurasi algoritma *Naïve Bayes Classifier* sebesar 99,84% dengan *interval* kepercayaan 95% antara 99,7% hingga 99,9%. Sedangkan tingkat akurasi algoritma *Random Forest* sebesar 99,82% dengan perolehan *interval* kepercayaan 95% antara 99,7% hingga 99,8%.
2	Ahmad Turmudi Zy, Wahyu Hadikristanto	Implementasi Algoritma Metode Naïve Bayes dan Support Vector Machine Tentang Pembobolan dan	*Naïve Bayes Classifier, Support Vector Machine*	Keakuratan dan kestabilan algoritma *Support Vector Machine* dalam mengklasifikasikan *tweet* memperoleh nilai *Precision 80%*

		Kebocoran Data di Twitter		dan nilai *Recall 93%*, sedangkan algoritma *Naïve Bayes Classifier* memperoleh nilai *Precision 97%* dan nilai *Recall 97%*.
3	Dinna Nurfadillah	Analisis Sentimen Mengenai Kesadaran Masyarakat Indonesia Terhadap Keamanan Siber Dalam Menghadapi Kebocoran Data Menggunakan Algoritma Naïve Bayes Classifier	*Naïve Bayes Classifier*	Tingkat akurasi yang diperoleh dari masing-masing parameter yakni *Accuracy 74%, Precision 77,15%,* dan *Recall 80%.*

Tabel 2.1 Penelitian Terdahulu

Berdasarkan beberapa penelitian terdahulu yang dimuat dalam Tabel 2.1 tertera penulis, judul, algoritma metode penelitian yang digunakan, dan akurasi dari tiap algoritma. Pada penelitian pertama (Nursiyono & Huda, 2023), berjudul *"Analisis Sentimen Twitter Terhadap Perlindungan Data Pribadi Dengan Pendekatan Machine Learning",* memaparkan bahwa dari total perolehan data sebesar 65.241 *tweet* memberikan gambaran dari kedua algoritma yaitu *Naïve Bayes Classifier* dan *Random Forest* memiliki kemampuan yang sangat baik dalam mengklasifikasikan *tweet* tiap sentimennya. Perolehan hasil persentase sentimen positif, negatif, dan netral dari algoritma *Naïve Bayes Classifier* berkisar dari 99,6% hingga 99,99%. Sedangkan algoritma Random Forest berkisar dari 99,41% hingga 100%. Pada penelitian kedua (Zy & Wahyu Hadikristanto, 2023), berjudul *"Implementasi Algoritma Metode Naïve Bayes dan Support Vector Machine Tentang Pembobolan dan Kebocoran Data di Twitter,* memaparkan dua metode yang digunakan yaitu *Naïve Bayes Classifier* dan *Support Vector Machine* dengan total pengolahan data 300 *tweet*, yang terdiri dari 115 *tweet* negatif, 128 *tweet*

positif, dan 17 *tweet* netral. Performa hasil yang diberikan dari kedua algoritma sangat mengesankan, hal ini dapat dilihat dari performa hasil algoritma *Naïve Bayes Classifier* dengan hasil sentimen negatif memiliki capaian nilai *Precision 97%* dan nilai *Recall 97%*, sentimen netral memiliki capaian nilai *Precision 81%* dan nilai *Recall 100%*, sentimen positif memiliki capaian nilai *Precision 100%* dan nilai *Recall 97%*. Sedangkan performa hasil algoritma *Support Vector Machine* memiliki capaian hasil sentimen negatif dengan nilai *Precision 80%* dan nilai *Recall 93%*, sentimen netral memiliki capaian nilai *Precision 81%* dan nilai *Recall 100%*, sentimen positif memiliki capaian nilai *Precision 80%* dan nilai *Recall 57%*. Pada penelitian ketiga (Nurfadillah, 2023), berjudul *"Analisis Sentimen Mengenai Kesadaran Masyarakat Indonesia Terhadap Keamanan Siber Dalam Menghadapi Kebocoran Data Menggunakan Algoritma Naïve Bayes Classifier"*, memaparkan bahwa penggunaan algorima *Naïve Bayes Classifier* pada 100 *tweet* setelah dilakukan pelabelan dan preprocessing memiliki kecendrungan hasil yang cukup baik dalam mengklasifikasikan tiap sentimen, menghasilkan perbandingan 5:9:3 tiap klasifikasi sentimen positif, netral, dan negatif. Pengujian model algoritma juga menggunakan metode *10-folds cross validation* dengan rentang persentase yang didapat sebesar 70% hingga 80%.

Berdasarkan pemaparan beberapa penelitian yang telah dilakukan sebelumnya, penulis mencoba untuk menggunakan algoritma *Naïve Bayes Classifier* sebagai metode utama yang akan digunakan dalam proses penelitian analisis sentimen keamanan data. Hal ini mengacu kepada performa hasil algoritma *Naïve Bayes Classifier* yang dinilai sangat mengesankan dalam melakukan klasifikasi. Sebagai pembanding dengan penelitian sebelumnya, penulis akan melakukan perbandingan hasil yang sederhana dari tiap klasifikasi termasuk algoritma *Naïve Bayes Classifier* menggunakan data yang sama namun dengan bahasa yang berbeda dengan *library* bawaan *phyton*.

2.2 Analisis Sentimen

Analisis sentimen (*Sentiment Analysis*) merupakan bidang studi keilmuan komputasi yang dapat mengolah dan memproses data untuk dapat dikenali sesuai ekspresi dari tiap pendapat, emosi, sikap, pandangan, dan subjektifitas dalam suatu teks (Latuny et al., 2021).

Pengolahan data analisis sentimen bertujuan dalam hal memahami tiap data yang berbentuk teks untuk dapat menemukan informasi berupa kalimat opini, sehingga memudahkan proses analisis dalam melihat kecendrungan dari tiap opini positif, negatif, dan netral terhadap suatu objek permasalahan (Zalyhaty, 2021).

2.3 Naïve Bayes Classifier

Naïve Bayes Classifier merupakan salah satu bentuk algoritma klasifikasi *machine learning* , penerapan pada umunya digunakan dalam bidang *text mining* dan *sentiment analysis.* Pemodelan dalam bidang matematika dan statistika yang ditemukan oleh salah seorang ilmuan Inggris bernama Thomas Bayes berhasil mengemukakan sebuah metode yang dapat memprediksikan probabilitas dimasa depan berdasarkan pengalaman di masa sebelumnya. Berdasarkan konsep *teori bayes* yang memaparkan peluang bersyarat *P(X/C)*, dimana X merupakan *posterior* dan C adalah *prior*. Prior itu sendiri merupakan sebuah pengetahuan dengan karakteristik berupa parameter atau pengalaman di masa lampau, sedangkan posterior merupakan sebuah pengetahuan dengan karakteristik berupa dugaan pada kejadian yang akan datang. Adapun persamaan yang digunakan dapat dilihat pada persamaan matematika berikut,

$$P(H/X) = \frac{P\,(H/X)\,P(H)}{P(X)}$$

Keterangan:

X : Data *class* yang belum diketahui

H : Hipotesis data *X* merupakan suatu *class* spesifik

P(H/X) : Probabilitas hipotesis *H* berdasarkan kondisi *X* (*posterior probabilitas*)

P(H) : Probabilitas hipotesis *H* (*prior probabilitas*)

P(H/X) : Probabilitas *X* berdasarkan kondisi

P(X) : Probabilitas dari *X*

Persamaan tersebut dapat dijabarkan sebagai data *training* dengan beberapa kategori sebagai berikut: *H = {H1, H2, Ha, …, Hk}* dimana *prior* probabilitas untuk semua kategori berupa *P(Hj)*, dengan nilai *j = 1, 2, 3,…, k*. Sedangkan untuk *X = {X1, …, Xj, … Xm}*, dengan *Xj* nilai *j = 1, 2, 3, …m.* Proses klasifikasi algoritma *Naïve Bayes Classifier* dibagi menjadi dua tahapan yaitu tahap pelatihan (*training*) dan tahap klasifikasi (*classification*). Nilai *prior* didapatkan saat proses pelatihan pertama kali dilakukan (Zy & Wahyu Hadikristanto, 2023).Pada penelitian analisis sentimen ini penulis akan menggunakan kurang lebih 150 *tweet* sentimen. Setelah perhitungan probabilitas, penulis akan melakukan perhitungan untuk data *testing* dengan kondisi probabilitas yang belum diketahui. Nantinya nilai dari data *testing* akan dikategorikan ke dalam dua kelas yaitu positif dan negatif.

2.4 Aplikasi X

X apps atau lebih akrab dikenal banyak orang sebagai *twitter* merupakan sebuah situs *microblog* yang dapat memfasilitasi penggunanya dalam mengirimkan pesan berupa *tweet* dengan ketentuan panjang maksimal 140 karakter (Achsanty, 2021). *Tweet* atau cuitan pada aplikasi X lebih sering digunakan kebanyakan orang sebagai tempat untuk berbagi cerita maupun sekedar informasi yang ingin mereka bagikan kepada para pengikutnya (*followers*). Pada umumnya *tweet* dapat mencakup data berupa teks, gambar, video, atau tautan. Tiap antar pengguna X dapat menggunakan *tweet* sebagai sarana untuk berinteraksi seperti membalas *tweet* (*reply*), mengirim ulang *tweet* orang lain ke pengikut mereka sendiri (me-*retweet*), atau hanya sekedar memberikan suka (*like*). Penelitian analisis sentimen keamanan data ini nantinya akan menggunakan *tweet* atau cuitan pada aplikasi X yang telah dikumpulkan melalui proses *crawling data*. Proses *crawling data tweet* aplikasi X , penulis menggunakan *tools* bernama *tweet harvest*.

2.5 Text Mining

Text Mining merupakan sebuah proses pencarian sebuah fakta maupun informasi pada sebuah teks, penerapan pemrosesan *text mining* diperuntukan untuk mencakup data dengan ukuran yang cukup besar dengan struktur kompleksitas, dan tingkat dimensi yang tinggi (Pravina et al., 2019). Hasil pencarian biasanya digunakan dalam menyelesaikan masalah data dengan lingkup permasalahan yang erat hubungannya dengan pemrosesan bahasa alami, sehingga pada tahapannya akan ditemukan proses seperti *text preprocessing*, pembobotan teks, maupun ekstraksi teks itu sendiri (Sabian, 2023).

2.6 Text Pre-Processing

Text Pre-Processing merupakan sebuah proses mempersiapkan data berupa teks sebelum dilakukan analisis maupun pengolahan lanjutan. Proses mempersiapkan teks sebelum dilakukannya analisis biasanya berfokus pada membersihkan dan merapikan data yang bersifat noise (informasi yang kurang lengkap atau hilang) (Pravina et al., 2019). Berikut adalah beberapa tahapan *text pre-processing* yang akan dilakukan penulis pada penelitian analisis sentimen keamanan data.

a) Case Folding

Proses *case folding* yang akan dilakukan berupa *lowercasing*, yaitu mengubah semua huruf pada suatu teks ataupun kalimat menjadi huruf kecil (Indraloka & Santosa, 2017),

misalnya mengubah “Ini merupakan SUATU Kalimat” menjadi “ini merupakan suatu kalimat”.

b) Data Cleaning

Data cleaning biasanya dilakukannya dengan cara membersihkan suatu data untuk kemudian dilakukannya penghapusan teradap karakter *non-teks* yang tidak relevan (karakter yang tidak termasuk ke dalam huruf *alfabet a – z*) (Pravina et al., 2019).

c) Translation

Translation diperuntukkan dalam hal menerjemahkan suatu kalimat dari suatu jenis bahasa ke dalam bentuk jenis bahasa lainnya. Pada umumnya proses ini banyak digunakan untuk menerjemahkan kalimat pada *tweet* dari Bahasa Inggris ke dalam bentuk Bahasa Indonesia, ataupun sebaliknya.

d) Tokenization

Tokenization merupakan suatu proses memisahkan setiap kata per kata dalam suatu kalimat atau teks yang dipisahkan dengan tanda baca koma (,) (Indraloka & Santosa, 2017).

e) Lemmatization

Lemmatization atau *stemming* merupakan suatu proses menormalisasikan teks dengan cara mengembalikan suatu kata pada suatu kalimat atau teks ke dalam bentuk kata dasar, penerapan *lemmatization* atau *stemming* pada teks Bahasa Indonesia umumnya dengan cara mengubah kata berimbuhan ke dalam bentuk kata dasarnya, misalnya : “menulis” menjadi “tulis”, “membaca” menjadi “baca”, “mengetik” menjadi “ketik”, dsb.

f) Stopword Removal

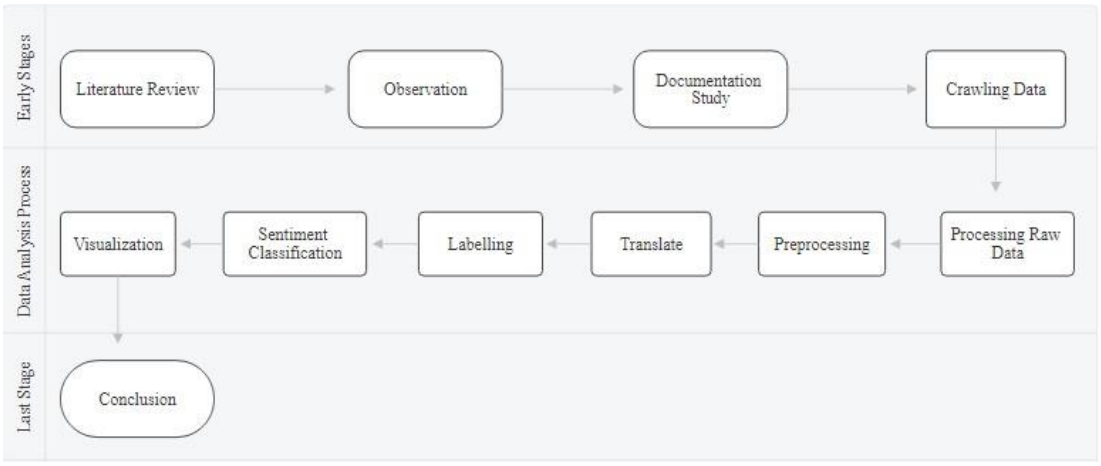
Stopword Removal merupakan proses menghapus kata-kata umum yang tidak memberikan nilai atau arti pada suatu teks, sehingga dapat mengurangi jumlah kata yang nantinya akan diproses dan disimpan oleh *corpus* atau sistem (Mukminin et al., 2021).

BAB III
METODOLOGI PENELITIAN

Pada bagian BAB III akan lebih berfokus terhadap pemaparan alur metodologi penelitian yang terbagi menjadi dua yakni bagian langkah penelitian menggambarkan secara garis besar proses yang akan dilakukan selama penelitian ini berlangsung dan bagian uraian penelitian menjelaskan secara detail dan rinci dari tiap proses yang telah dipaparkan pada bagian langkah penelitian.

3.1 Langkah Penelitian

Pada penelitian ini akan dilakukan langkah-langkah penelitian yang akan dibagi menjadi tiga tahapan yaitu e*arly stages, data analysis process,*dan *last stage*. Alur detail metodologi penelitian dapat dilihat pada Gambar 3.1.



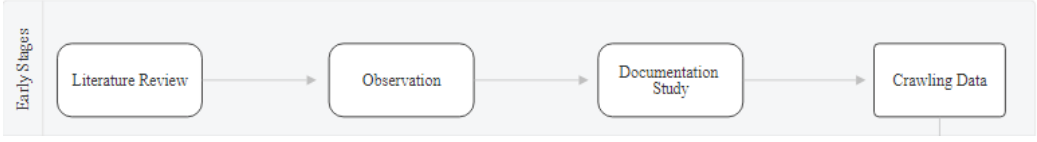
Gambar 3.1 Flowchart Metode Penelitian

3.2 Uraian Penelitian

Berikut ini akan diuraikan langkah dari alur penelitian ini sesuai dengan Gambar 3.1.

3.2.1 Early Stages

Early stages merupakan tahapan awal dari penelitian ini sebelum dilakukkannya proses analisis data (*data analysis process*). Adapun tahapan dari *early stages* dibagi menjadi empat tahap yaitu *Literature Review, Observation, Documentation Study,* dan *Crawling Data*.



Gambar 3.2 Flowchart Early Stages

3.2.1.1 Literature Review

Literature review atau kajian literatur merupakan langkah pertama yang penulis lakukan dalam melakukan penelitian ini. *Literature review* membantu penulis dalam melakukan kajian terhadap beragam literatur dari penelitian sebelumnya. Hal ini bertujuan sebagai acuan referensi dan bahan belajar bagi penulis dalam menyusun penelitian tugas akhir.

3.2.1.2 Observation

Observation merupakan langkah yang penulis lakukan sebagai bentuk melihat secara langsung, mengamati, dan memahami objek penelitian itu sendiri. Proses mengamati yang penulis lakukan dengan cara mengetikkan kata kunci pada kolom pencarian X berupa “pdns” dan “pdns 2”. Hasil pencarian memaparkan beragam cuitan (*tweet*) dari berbagai pengguna X dan pos berita, seperti pada Gambar 3.3 dan Gambar 3.4.



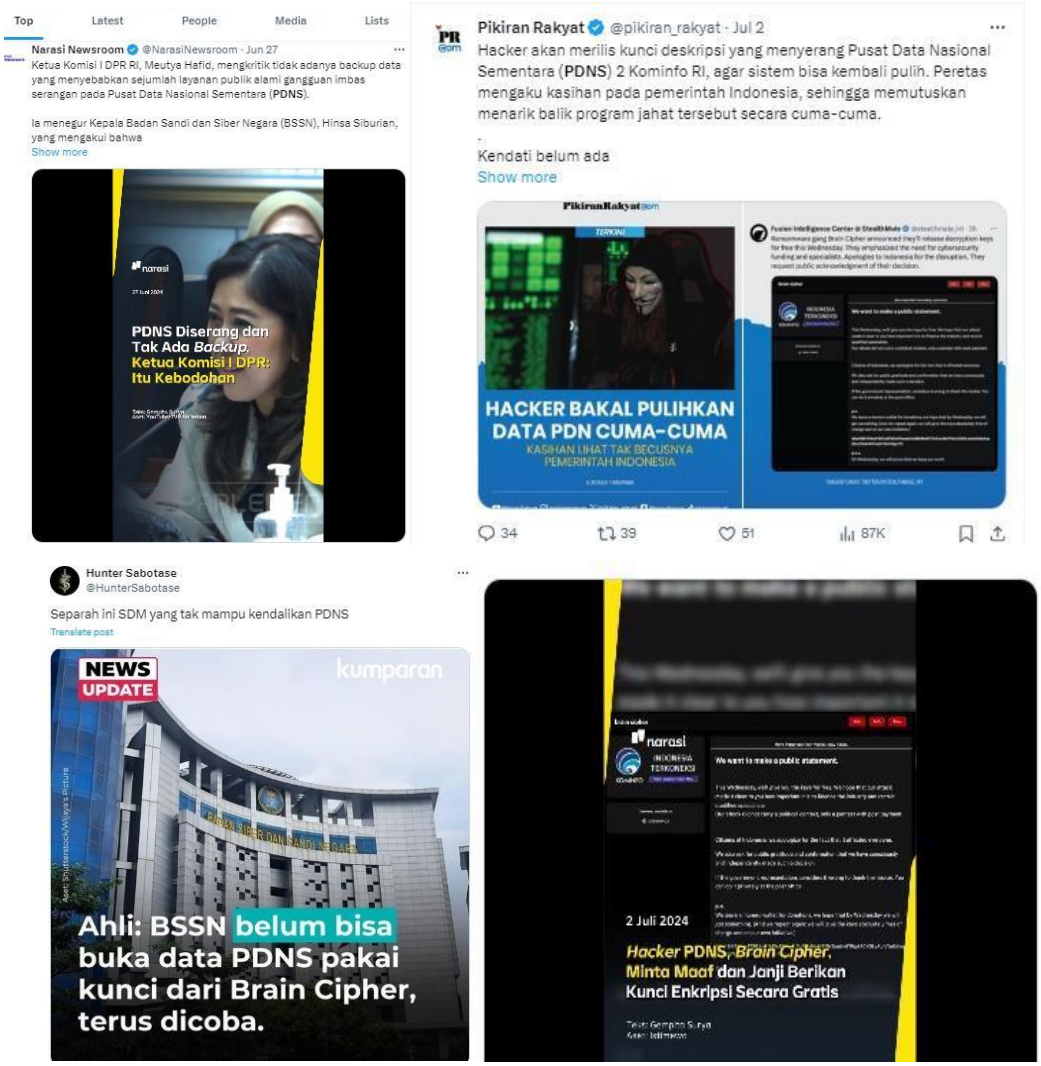
Gambar 3.3 Tweet Pengguna X



Gambar 3.4 Pos Berita

3.2.1.3 Documentation Study

Documentation Study merupakan lanjutan dari tahapan *observation* berupa hasil dari beragam cuitan (*tweet*) dan sumber seperti pada **Error! Reference source not found.** yang t elah dicari sebelumnya, penulis melakukan pencocokan dan pengerucutan suatu kesimpulan yang nantinya akan menjadi topik utama permasalahan pada penelitian ini. Temuan topik utama tersebut berupa permasalahan keamanan data pada Pusat Data Nasional Sementara (PDNS). Permasalahan keamanan data tersebut mengakibatkan terjadinya peretasan oleh *hacker* dengan nama *Brain Chiper*. Peretasan dilakukan dengan cara membobol sistem keamanan dan melakukan penyerangan sistem menggunakan *ransomeware LockBit 3.0.*



Gambar 3.5 Kumpulan Pos Berita

Hasil dari ketiga tahapan yakni *literature review, observation, dan documentation study* menghasilkan sebuah judul penelitian yaitu *"Analisis Sentimen X Terhadap Keamanan Data*

Menggunakan Algoritma Naive Bayes Classifier: Studi Kasus Pusat Data Nasional Sementara".

3.2.1.4 Crawling Data

Pada tahap crawling data, penulis melakukan pengambilan data menggunakan tools bernama *Tweet Harvest*. *Tweet Harvest* merupakan sebuah tools yang berbasis bahasa pemrograman *python* dan *javascript* menggunakan *node.js* sebagai *main runtime javascript* dalam menjalankan proses penambangan data *crawling* dari *X apps. Tools* ini membutuhkan *X authentication token* berupa token autentikasi yang berguna untuk mengotentikasi permintaan *API* yang diajukan ke X menggunakan *OAuth (Open Authorization)*. OAuth merupakan protokol otentikasi terbuka yang digunakan oleh X untuk mengamankan akses ke data pengguna oleh aplikasi pihak ketiga tanpa perlu mengungkapkan kata sandi pengguna. Data hasil *crawling* yang berhasil diperoleh sebanyak 1039 data *tweets.*

3.2.1.5 Analisis Kebutuhan Sistem

Analisis kebutuhan sistem mengambil peran dalam hal untuk dapat mengetahui kebutuhan apa saja yang diperlukan suatu penelitian. Sistem yang digunakan tersebut dapat berupa *perangkat lunak (software)* dan *perangkat keras (hardware)*.

Kebutuhan *perangkat lunak (software)* pada penelitian ini adalah sebagai berikut:

a. *Google Colaboratory* sebagai *tools* utama penulis dalam mengolah data pada penelitian ini.

b. Bahasa pemrograman yang penulis gunakan dalam penelitian ini yaitu *python 3* bawaan dari *Google Colaboratory.*

c. *Tools* untuk melakukan *crawling data* berupa *Tweet Harvest versi [v2.6.1]*, *tools* berbasis *node.js* ini banyak membantu penulis dalam menemukan temuan data *tweet* saat proses *crawling data*.

d. *Library pandas* bawaan *Google Colaboratory*, *library* ini digunakan penulis untuk menyimpan dan mengelola data dalam bentuk yang lebih terstruktur.

e. *Library seaborn* bawaan *Google Colaboratory, library* ini membantu penulis dalam pembuatan visualisasi data lebih menarik dan informatif.

f. *Library re (regular expression)* bawaan *Google Colaboratory, library* ini membantu penulis dalam hal mengurutkan karakter untuk dapat dilakukan pencocokan *string* yang lebih kompleks.

g. *Library matplotlib* bawaan *Google Colaboratory, library* ini berguna bagi penulis

dalam hal menyediakan fungsi-fungsi yang kuat dalam berbagai jenis plot dan grafik, mulai dari plot sederhana hingga plot yang sangat kustomisasi. *Library* ini cukup populer dalam hal visualisasi data di dalam ruang lingkup bahasa pemrograman *python.*

h. *Library Sastrawi* bawaan *Google Colaboratory, library* ini membantu penulis dalam hal mendukung proses *stemming* atau *lemmatization* dalam bahasa Indonesia.

i. *Library translate* bawaan *Google Colaboratory, library* ini berguna untuk menerjemahkan Bahasa Indonesia ke dalam bentuk Bahasa Inggris.

j. *Library tweet -preprocessor* bawaan *Google Colaboratory, library* ini memudahkan penulis dalam melakukan proses preprocessing pada *tweet* di X.

k. *Library textblob* bawaan *Google Colaboratory, library* ini membantu penulis memudahkan melakukan operasi-operasi pemrosesan teks dan bahasa alami seperi penelitian analisis sentimen ini.

l. *Library wordcloud* bawaan *Google Colaboratory, library* ini berguna dalam hal membuat visualisasi awan kata *(word cloud)*. Awan kata yang ditampilkan merupakan representasi visual dari data teks paling sering muncul dan memiliki bobot tiap katanyanya, dimana kata dengan ukuran lebih besar memiliki bobot lebih besar.

m. *Library nltk (natural language toolkit)* bawaan *Google Colaboratory, library* ini membantu penulis dalam hal menyediakan berbagai alat dan sumber daya membangun program dengan teks berbahasa Indonesia.

n. *Modul random* bawaan *Google Colaboratory*, modul ini menyediakan berbagai fungsi-fungsi untuk menghasilkan bilangan acak.

Kebutuhan perangkat keras *(hardware)* pada penelitian ini adalah sebagai berikut:

a. Laptop yang digunakan penulis pada penelitian ini memiliki *processor* yaitu *Intel(R) Core(TM) i3-5005U CPU @ 2.00GHz 2.00 GHz.*

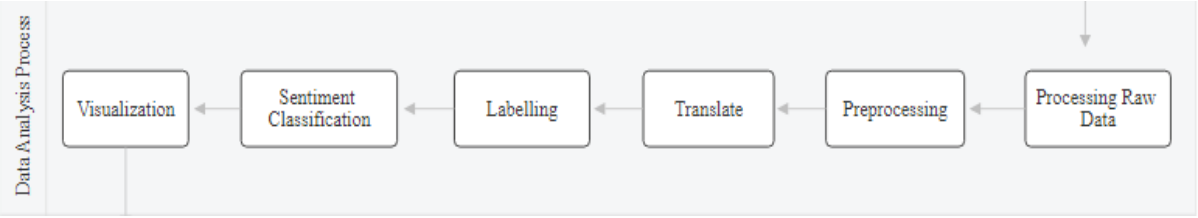
b. Laptop yang digunakan penulis menggunakan RAM (*Random Access Memory*) sebesar 4GB.

c. Laptop yang digunakan penulis memiliki monitor dengan ukuran *14inch* dengan resolusi *1366 x 768 pixel.*

d. Laptop yang digunakan penulis memiliki GPU (*Graphics Processing Unit*) *intel(R) HD Graphic 5500 NVDIA GeForce 920.*

3.2.2 Data Analysis Process

Tahap *Data Analysis Process* merupakan tahapan selanjutnya setelah dilakukannya tahap *Early Stages,* tahap ini berfokus pada proses analisis data. Adapun tahap *data analysis process* seperti pada Gambar 3.6 dibagi menjadi enam tahap yaitu *processing raw data, preprocessing, translate, labelling, sentiment classification,* dan *visualization.*



Gambar 3.6 Flowchart Data Analysis Process

3.2.2.1 Processing Raw Data

Processing Raw Data merupakan sebuah tahap awal dalam proses pengolahan data sebelum dilakukannya analisis data, proses ini akan mengolah data mentahan hasil *crawling data* sebelumnya, sehingga data siap untuk dilakukan tahap analisis selanjutnya. Tahap ini dilakukan pada tahapan awal memproses data mentah dan dibagi menjadi empat tahap yaitu *categorizing data, duplicates, data cleaning,* dan *case folding*. *Categorizing data* merupakan proses mengkategorisasikan data yang diperlukan ke dalam kolom kelompok yang diperlukan, hal ini akan memudahkan dalam memfokuskan pengolahan data yang diinginkan. Pada penelitian ini, proses *categorizing data* akan mengkategorikan data mentahan ke dalam tiga kelompok kolom yakni *full text, username, dan created at. Duplicates* merupakan proses menghilangkan data *tweet* kembar atau terduplikat, hal ini bertujuan untuk mengurangi data *tweet* berulang dalam proses analisis data nantinya. *Data cleaning* merupakan proses merapikan data dari beberapa simbol, karakter, tanda baca, dsb. Pada penelitian ini penulis melakukan *data cleaning* dari simbol *@username, tagar (#hashtag), retweet (RT), url pencarian (https:...), karakter alphanumeric, dan spasi berlebihan (space_). Case Folding* merupakan mengubah suatu kalimat berdasarkan ketentuan yang diinginkan sesuai kebutuhan, pada penelitian ini penulis mengubah kalimat *tweet* menjadi huruf kecil *(lowercase).*

3.2.2.2 Preprocessing

Preprocessing merupakan proses mempersiapkan data sebelum digunakan untuk dilakukan analisis ataupun pemodelan lebih lanjut, hal ini bertujuan untuk memastikan bahwa data yang akan digunakan nantinya telah bersih, relevan, dan siap untuk diproses. Pada penelitian ini akan dilakukan empat tahap *preprocessing* yaitu *normalized, stopword,*

tokenized, lemmatization atau *stemming. Normalized* merupakan proses menormalisasikan kata-kata yang nantinya akan diubah menjadi nilai-nilai bentuk variabel yang diinginkan dengan skala berbeda tiap kata, seperti mengubah kata *yg* menjadi kata *yang*, mengubah kata *blm* menjadi kata *belum*, *tdk* menjadi kata tidak. *Stopword* merupakan proses menghapus kata-kata umum yang sering muncul dalam suatu teks, baik itu kata yang ada bobot maupun sekedar tidak ada bobot. *Tokenized* merupakan proses mengubah suatu teks atau kalimat menjadi penggalan kata per kata yang lebih kecil. Seperti mengubah kalimat *hari ini saya akan makan nasi goreng*, kalimat tersebut akan berubah menjadi *(hari, ini, saya, akan, makan, nasi, goreng,). Lemmatization* atau *stemming* merupakan proses mengurangi atau mengembalikan kata dalam suatu teks menjadi bentuk kata dasarnya. Sebagai contoh seperti kata imbuhan berupa awalan dan akhiran. Sehingga kata yang tersisa hanya kata dasar, seperti mengubah kata *menyetir* menjadi kata *setir*, *memakan* menjadi kata *makan*, *memulai* menjadi kata *mulai*.

3.2.2.3 Translate

Translate merupakan suatu proses mengubah teks atau kalimat dari suatu bahasa ke bentuk bahasa lain, hal ini menjadi bagian tahapan dalam pemrosesan bahasa untuk diterjemahkan sehingga bahasa yang dihasilkan menjadi konsisten dan mudah dipahami oleh sistem saat analisis sentimen dilakukan. Proses *translate* dalam *preprocessing* bisa melibatkan beberapa tahapan seperti pengenalan bahasa, penerjemahan, evaluasi atau koreksi, dan pemrosesan lanjutan. Pada penelitian ini hanya menggunakan dua tahapan yaitu menerjemahkan dan pemrosesan lanjutan. Proses terjemahan data diperlukan karena analasis sentimen yang akan dilakukan pada penelitian ini menggunakan dua bahasa yaitu Bahasa Indonesia dan Bahasa Inggris.

3.2.2.4 Labelling

Labelling merupakan suatu proses memberikan pelabelan atau mengkategorikan suatu data yang tidak memiliki kategori label sebelumnya, hal ini menjadi bagian tahapan yang penting dalam banyak pengaplikasian *machine learning* dan analisis data yang memerlukan data terstruktur dan mudah dimengerti oleh algoritma atau model yang akan digunakan. Adapun proses dari labeling pada analisis sentimen terbagi menjadi empat tahap yakni pemilihan skala sentimen, anotasi manual, penggunaan algoritma, dan verifikasi. Tahap pemilihan skala sentimen merupakan langkah awal sebelum nantinya kegiatan analisis sentimen dilakukan, pada tahap ini skala pemilihan dari kategori sentimen akan digunakan meliputi sentimen bernilai positif, negatif, atau netral. Skala penentuan juga bergantung kepada kebutuan analisis dari analisis sentimen itu sendiri. Tahap anotasi manual memberikan

keterlibatan manusia sebagai seorang individu yang membaca suatu teks dan menentukan bobot sentimen yang terkandung di dalamnya berdasarkan aturan berupa panduan yang telah ditentukan. Sedangkan penggunaan algoritma pada labeling lebih menekankan terhadap penentuan algoritma atau model NLP seperti apa yang akan digunakan dalam melakukan prediksi analisis sentimen suatu teks. Hal ini perlu dilakukan untuk dapat menentukan validasi dan *crosscheck* (pengecekan ulang) oleh manusia itu sendiri dalam memastikan keakuratan. Tahap akhir akan dilakukan verifikasi, setelah semua tahapan labeling dilakukan pada analisis sentimen, sebuah koreksi dan verifikasi sangat diperlukan untuk dapat memastikan konsistensi labeling dari tiap teks berbeda. Sehingga hasil akhir berupa keakuratan dari labeling dapat dicapai.

3.2.2.5 Sentiment Classification

Sentiment Classification atau klasifikasi sentimen, merupakan suatu proses menganalisis teks atau kalimat dengan tujuan memudahkan dalam mengenali dan mengklasifikasikan jenis sentimen atau opini seperti apa yang terkandung dengan pertimbangan beberapa kategori-kateori yang sudah ditentukan. Klasifikasi sentimen ini dapat menentukan bobot suatu teks atau kalimat bersifat positif, negatif, atau netral berdasarkan ekspresi bahasa yang digunakan. Pada penelitian ini penulis menggunakan algoritma *Naive Bayes Classifier* sebagai metode utama dalam melakukan klasifikasi sentimen. Penulis juga menggunakan *library* bawaan seperti *TextBlob* untuk melakukan proses klasifikasi sederhana. Hasil dari klasifikasi sentimen nantinya akan membandingkan beberapa parameter seperti membandingkan hasil klasifikasi dari sentimen asal berbahasa Indonesia dengan sentimen hasil translate berbahasa Inggris. Setelah dilakukannya semua klasifikasi sentimen, penulis akan menampilkan visualisasi ke dalam bentuk *bar chart* dan awan kata (*word cloud*), sehingga data hasil analisis sentimen lebih mudah dipahami.

3.2.2.6 Visualization

Visualization atau visualisasi, merupakan proses visual yang merujuk kepada penggunaan teknik memvisualisasikan hasil dari analisis sentimen terhadap teks atau data yang telah diproses. Visualisasi ini berguna untuk dapat memudakan penulis dalam hal menyajikan informasi secara variatif, intuitif, dan mudah dimengerti, sehingga memberikan kemungkinan untuk dapat menarik kesimpulan dalam mengeksplorasikan data secara lebih efektif. Adapun tujuan utama visualisasi dalam analisis sentimen yakni memberikan penggambaran distribusi sentimen, mengidentifikasikan pola maupun *trend*, dan sebagai alat untuk perbandingan

analisis. Penggambaran distribusi sentimen diperlukan dalam visualisasi untuk dapat memudahkan dan menggambarkan bagaimana distribusi sentimen dari teks-teks yang dianalisis. Penyajian dalam bentuk *chart*, grafik, ataupun gambar berpola dapat digunakan sebagai wadah untuk menunjukkan proporsi dari tiap teks bersentimen positif, negatif, dan netral. Proses identifikasi pola maupun *trend* sendiri kebanyakan menggunakan grafik garis atau area, hal ini tentunya lebih memudahkan penelitian analisis sentimen dalam hal memvisualisasikan proses mengidentifikasikan pola dan tren dari sentimen itu sendiri berdasarkan rentang dari waktu ke waktu atau dari berbagai pristiwa yang sudah terjadi. Visualisasi membantu penulis untuk dapat membandingkan sentimen dari keberagaman kelompok tiap teks yang berbeda, sehingga memudahkan penulis untuk dapat melihat bagaimana tiap sentimen direpresntasikan dalam bentuk berbagai domain topik pembahasan.

3.2.3 Last Stage

Last stage merupakan tahapan akhir yang dilakukan setelah tahapan *early stages* dan *data analysis process* selesai. Pada tahap ini penulis memaparkan kesimpulan dan saran terkait penelitian ini dan berharap penelitian selanjutnya lebih baik lagi.



Gambar 3.7 Flowchart Last Stage

3.2.3.1 Conclusion

Conclusion atau Kesimpulan, tahap ini akan memaparkan sekaligus menjelaskan secara menyeluruh hasil analisis sentimen menggunakan algoritma *Naive Bayes Classifier* pada penelitian ini menggunakan data *tweet* terhadap keamanan data Pusat Data Nasional Sementara (PDNS), sehingga akan diketahui hasil dari klasifikasi data yang diperoleh bernilai sentimen positif,negatif, atau netral. Pada bagian ini juga menyertakan sekaligus saran, hal ini bertujuan sebagai masukan kepada para pembaca penelitian ini, penulis berharap penelitian ini dapat menjadi bahan refrensi ataupun acuan pada penelitian selanjutnya lebih baik lagi.

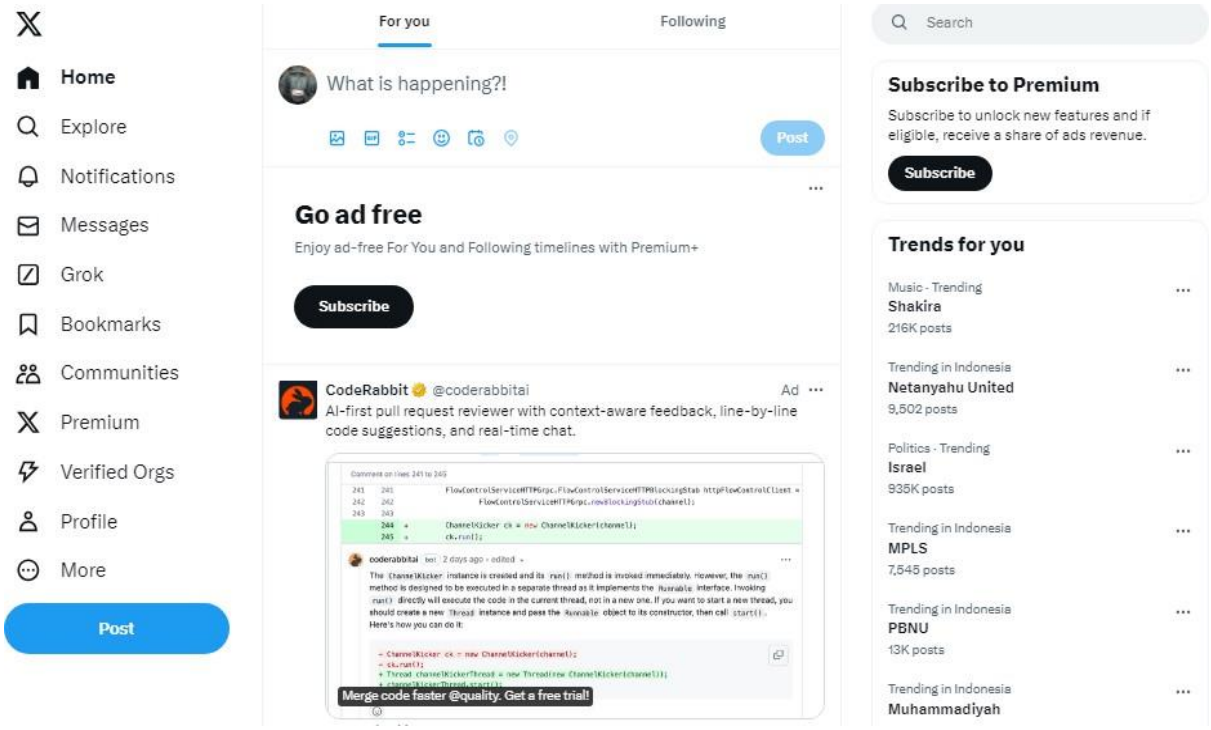
BAB IV
HASIL DAN PEMBAHASAN

Pada bagian BAB IV akan lebih berfokus terhadap berbagai rangkaian proses dan pembahasan hasil temuan yang telah dilakukan pada penelitian ini. Adapun urutan proses dimulai dengan pencarian data *(crawling data)*, memproses data mentah yang telah didapatkan *(processing raw data)*, mempersiapkan data *(preprocessing)*, menerjemahkan data *(translate)*, melakukan pelabelan pada data *(labelling)*, melakukan analisis klasifikasi sentimen *(sentiment classification)*, dan melakukan visualisasi data yang telah dianalisis *(visualization)*.

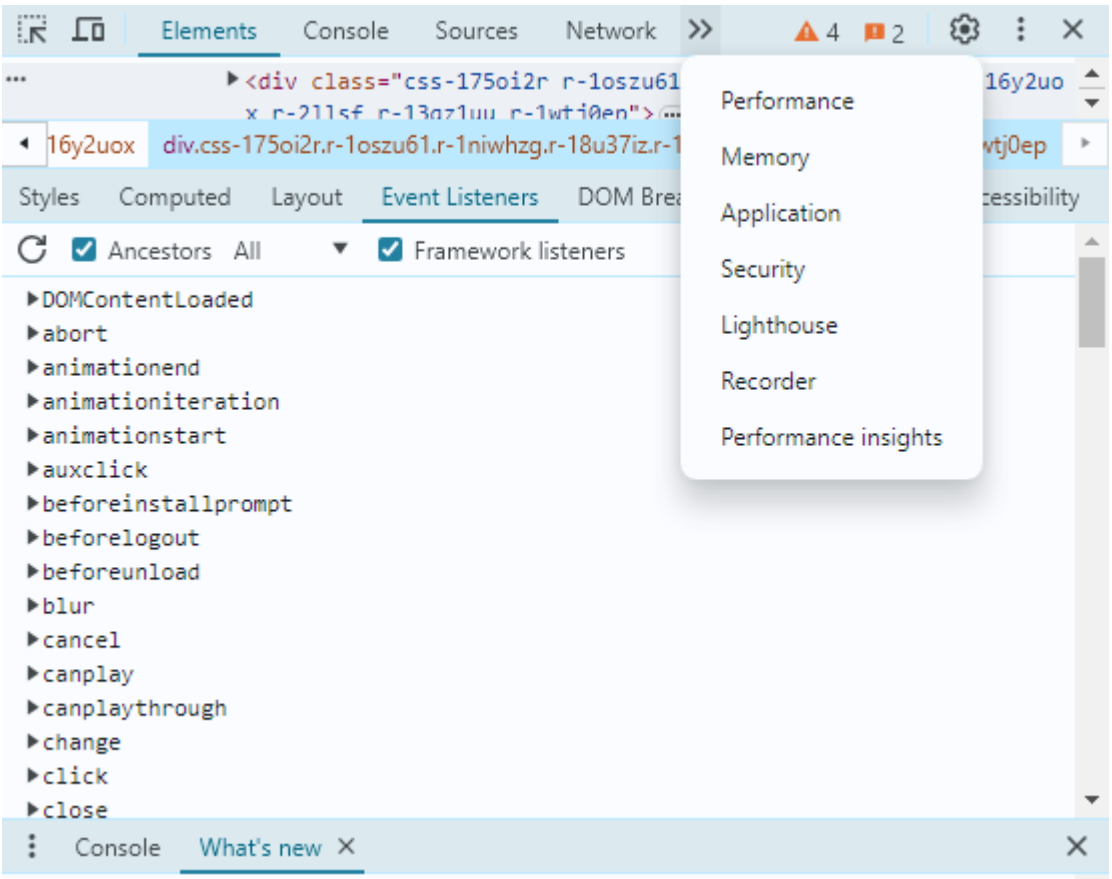
4.1 Crawling Data

Proses *crawling* akan melakukan pengambilan data menggunakan *tools* yang bernama *Tweet Harvest*- sebuah *tools* pencarian data berbasis *node js* dengan pengaplikasian pencarian secara otomatis pada aplikasi X melalui *automation scrolling* halaman per halaman hingga batas yang telah ditentukan. Sebelum memasuki tahap proses *crawling data*, langkah pertama yang harus dilakukan yakni pencarian token *authentication key* sebagai syarat utama untuk dapat mengakses aplikasi X. *Authentication key* terdiri dari token yang memiliki struktur berupa barisan deretan angka dan huruf acak tersusun secara enkripsi berdasarkan *password* pengguna (*user*) aplikasi X. Proses ini tentunya bertujuan untuk dapat melakukan izin terhadap proses *automation scrolling* yang nantinya akan dilakukan pada *crawling data* pada aplikasi X.

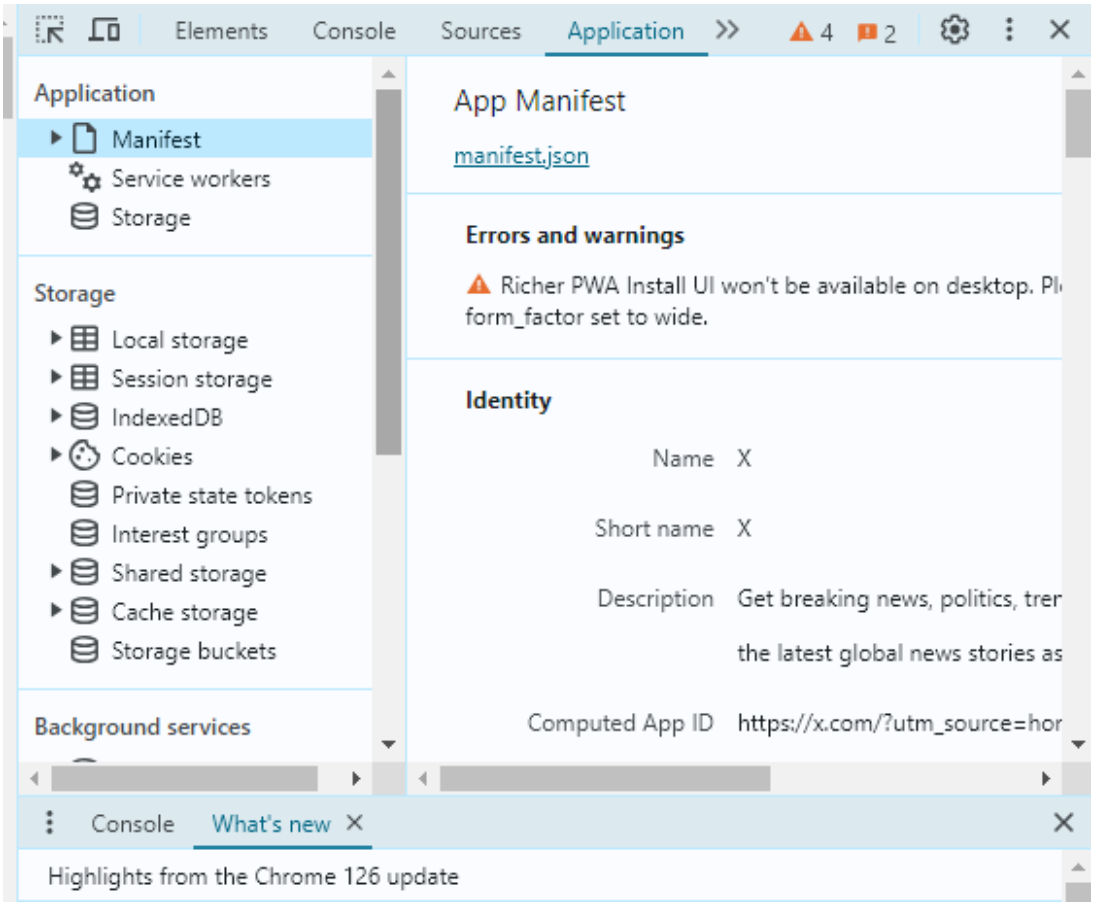
Proses pencarian *auth_token* dibagi menjadi beberapa tahapan langkah. Pada tahap awal akan dilakukan proses mendapatkan *Authentication token* dengan cara melakukan *log in* terlebih dahulu pada aplikasi X. Setelah masuk ke dalam halaman utama dari aplikasi X seperti pada Gambar 4.8, langkah selanjutnya melakukan *inspect element* hingga muncul tampilan seperti pada Gambar 4.9. Lakukan pencarian terhadap bagian *application* seperti pada Gambar 4.10 hingga memunculkan bagian *cookies* dari lamaan utama pengguna aplikasi X tersebut seperti pada Gambar 4.11. Bagian *url-cookies* nantinya akan memuat data yang merupakan keterangan dari lamaan tampilan *home* utama aplikasi X yang telah dibuka. Pada bagian *auth_token* akan berisi deretan angka dan huruf acak sebagai representasi dari *password* pengguna aplikasi X. Salinan dari *auth_token* akan berguna sebagai *main key* dalam menjalankan *tools Tweets Harvest*.



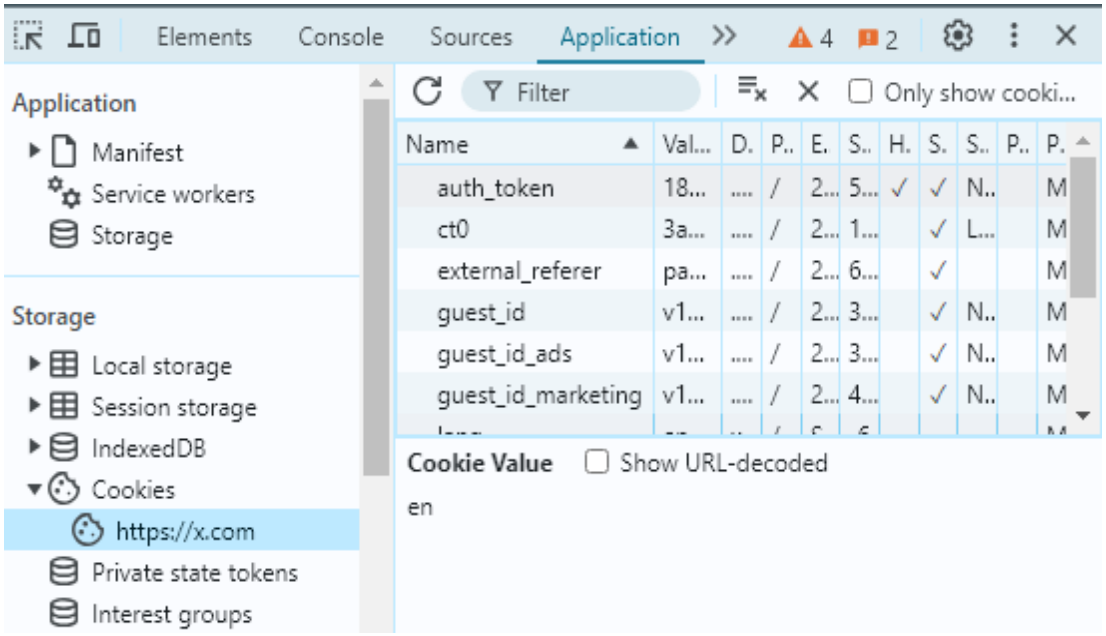
Gambar 4.8 Tampilan Home Aplikasi X



Gambar 4.9 Tampilan Inspect X



Gambar 4.10 Tampilan Application Inspect X



Gambar 4.11 Tampilan Cookies Url X

Setelah mendapatkan *authentication key* barulah proses *crawling data* dapat dilakukan. Proses *crawling* data dimulai dengan melakukan *install package pandas* dan *install node.js*, seperti pada Gambar 4.12.

```
#import python package 'pandas'
!pip install pandas

#install node.js (tweet-harvest)
!sudo apt-get update
!sudo apt-get install -y ca-certificates curl gnupg
!sudo mkdir -p /etc/apt/keyrings
!curl -fsSL https://deb.nodesource.com/gpgkey/nodesource-repo.gpg.key | sudo gpg --dearmor -o /etc/apt/keyrings/nodesource.gpg
!NODE_MAJOR=20 && echo "deb [signed-by=/etc/apt/keyrings/nodesource.gpg] https://deb.nodesource.com/node_$NODE_MAJOR.x nodistro main" |

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

!node -v
```

Gambar 4.12 Pseudocode Program Install Pandas dan Node Js

```
Requirement already satisfied: pandas in /usr/local/lib/python3.10/dist-packages (2.0.3)
Requirement already satisfied: python-dateutil>=2.8.2 in /usr/local/lib/python3.10/dist-packages (from pandas) (2.8.2)
Requirement already satisfied: pytz>=2020.1 in /usr/local/lib/python3.10/dist-packages (from pandas) (2023.4)
Requirement already satisfied: tzdata>=2022.1 in /usr/local/lib/python3.10/dist-packages (from pandas) (2024.1)
Requirement already satisfied: numpy>=1.21.0 in /usr/local/lib/python3.10/dist-packages (from pandas) (1.25.2)
Requirement already satisfied: six>=1.5 in /usr/local/lib/python3.10/dist-packages (from python-dateutil>=2.8.2->pandas) (1.16.0)
Hit:1 http://security.ubuntu.com/ubuntu jammy-security InRelease
Hit:2 https://cloud.r-project.org/bin/linux/ubuntu jammy-cran40/ InRelease
Hit:3 https://developer.download.nvidia.com/compute/cuda/repos/ubuntu2204/x86_64  InRelease
Hit:4 https://deb.nodesource.com/node_20.x nodistro InRelease
Hit:5 http://archive.ubuntu.com/ubuntu jammy InRelease
Get:6 http://archive.ubuntu.com/ubuntu jammy-updates InRelease [128 kB]
Hit:7 https://ppa.launchpadcontent.net/c2d4u.team/c2d4u4.0+/ubuntu jammy InRelease
Hit:8 https://ppa.launchpadcontent.net/deadsnakes/ppa/ubuntu jammy InRelease
Hit:9 https://ppa.launchpadcontent.net/graphics-drivers/ppa/ubuntu jammy InRelease
Hit:10 http://archive.ubuntu.com/ubuntu jammy-backports InRelease
Hit:11 https://ppa.launchpadcontent.net/ubuntugis/ppa/ubuntu jammy InRelease
Fetched 128 kB in 1s (96.4 kB/s)
Reading package lists... Done
Reading package lists... Done
Building dependency tree... Done
Reading state information... Done
ca-certificates is already the newest version (20230311ubuntu0.22.04.1).
curl is already the newest version (7.81.0-1ubuntu1.16).
gnupg is already the newest version (2.2.27-3ubuntu2.1).
0 upgraded, 0 newly installed, 0 to remove and 52 not upgraded.
gpg: cannot open '/dev/tty': No such device or address
curl: (23) Failed writing body
deb [signed-by=/etc/apt/keyrings/nodesource.gpg] https://deb.nodesource.com/node_20.x nodistro main
Hit:1 https://cloud.r-project.org/bin/linux/ubuntu jammy-cran40/ InRelease
Hit:2 https://developer.download.nvidia.com/compute/cuda/repos/ubuntu2204/x86_64  InRelease
```

Gambar 4.13 Proses Install Package Berhasil

Setelah berhasil melakukan *install package* seperti pada Gambar 4.13, proses akan dilakukan dengan menginisialisasi nama *file* yang nantinya akan disimpan, pada penelitian ini *file* data akan disimpan dengan nama *'pdns.csv'*, perolehan *file output* yang akan didapatkan nantinya setelah proses *crawling data* selesai dilakukan berupa csv *(comma separated value)*. Kata kunci dari pencarian c*rawling data* pada penelitian ini yaitu 'pdns 2'. Pemilihan rentang waktu pencarian data dimulai saat peretasan Pusat Data Nasional Sementara telah selesai yakni tanggal 01 Juli 2024 s/d 11 Juli 2024. Bahasa utama pencarian data *tweet* yang dinginkan berupa Bahasa Indonesia. *Pseudocode* program dari proses *crawling data* dapat dilihat pada Gambar 4.14.

```
[ ]  #proses crawl_data
     filename = 'PdnsV2.csv'
     search_keyword = 'pdns 2 since:2024-07-1 until:2024-07-11 lang:id'
     limit = 150

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

Gambar 4.14 Pseudocode Program Crawling Data

```
This script uses Chromium Browser to crawl data from Twitter with your Twitter auth token.
Please enter your Twitter auth token when prompted.

Note: Keep your access token secret! Don't share it with anyone else.
Note: This script only runs on your local device.


Opening twitter search page...


Found existing file ./tweets-data/PdnsV2.csv, renaming to ./tweets-data/PdnsV2.old.csv

-- Scrolling... (1)
Filling in keywords: pdns 2 since:2024-07-1 until:2024-07-11 lang:id

 (2) (3) (4)

Your tweets saved to: /content/tweets-data/PdnsV2.csv
Total tweets saved: 15


Your tweets saved to: /content/tweets-data/PdnsV2.csv
Total tweets saved: 33
```

Gambar 4.15 Proses Scrolling Pada Crawling Data

```
Your tweets saved to: /content/tweets-data/PdnsV2.csv
Total tweets saved: 120

-- Scrolling... (1)

Your tweets saved to: /content/tweets-data/PdnsV2.csv
Total tweets saved: 135

-- Scrolling... (1) (2) (3) (4)

Your tweets saved to: /content/tweets-data/PdnsV2.csv
Total tweets saved: 153
Got 153 tweets, done scrolling...
```

Gambar 4.16 Proses Crawling Data Selesai

Selama proses *crawling data* program akan otomatis melakukan pencarian dengan melakukan *scrolling* secara berulang melalui aplikasi X hingga mencapai limit yang telah

ditentukan, sebagai contoh pada penelitian ini pencarian hanya dibatasi sebanyak 153 data *tweet.*

Setelah proses pencarian *Crawling Data* selesai seperti pada Gambar 4.16, maka akan ditampilkan data hasil *crawling* ke dalam sebuah tabel yang berisikan banyak kolom seperti *full_text, created_at, image_url, id_str* dsb. *Pseudocode* program untuk menampilkan hasil akhir *crawling data* dapat dilihat pada Gambar 4.17 dan Gambar 4.18.

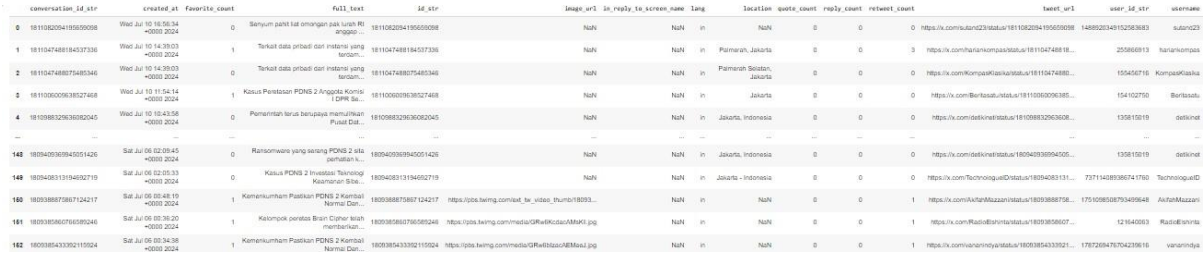
```
[ ]  import pandas as pd

     #Spesifikasi File CSV
     file_path = f"tweets-data/{filename}"

     #Membaca File CSV kedalam pandas DataFrame
     df = pd.read_csv(file_path, delimiter=",")

     #Menampilkan DataFrame
     display(df)
```

Gambar 4.17 Pseudocode Program Menampilkan Hasil Crawling Data



	conversation_id_str	created_at	favorite_count	full_text	id_str	image_url	in_reply_to_screen_name	lang	location	quote_count	reply_count	retweet_count	tweet_url	user_id_str	username
0	1811082094195659098	Wed Jul 10 16:56:34 +0000 2024	0	Senyum pahit tak omongan pak Ketua RI anggap ...	1811082094195659098	NaN	NaN	in	NaN	0	0	0	https://x.com/sutanl23/status/1811082094195659098	1488928349152583683	sutanl23
1	1811047488184537336	Wed Jul 10 14:39:03 +0000 2024	1	Terkait data pribadi dari instansi yang terdam...	1811047488184537336	NaN	NaN	in	Pamarah, Jakarta	0	0	3	https://x.com/hariankompas/status/1811047488184...	255866913	hariankompas
2	1811047488075485346	Wed Jul 10 14:39:03 +0000 2024	0	Terkait data pribadi dari instansi yang terdam...	1811047488075485346	NaN	NaN	in	Pamarah Selatan, Jakarta	0	0	0	https://x.com/KompasKlasika/status/1811047488075...	155456716	KompasKlasika
3	1811006009638527468	Wed Jul 10 11:54:14 +0000 2024	1	Kasus Peretasan PDNS 2 Anggota Komisi I DPR Sa...	1811006009638527468	NaN	NaN	in	Jakarta	0	0	0	https://x.com/Beritasatu/status/1811006009638527...	154102750	Beritasatu
4	1810988329636082045	Wed Jul 10 10:43:58 +0000 2024	0	Pemerintah terus berupaya memulihkan Pusat Dat...	1810988329636082045	NaN	NaN	in	Jakarta, Indonesia	0	0	0	https://x.com/detikinet/status/1810988329636082...	135815019	detikinet
...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...
148	1809409369945051426	Sat Jul 06 02:09:45 +0000 2024	0	Ransomware yang serang PDNS 2 sita perhatian k...	1809409369945051426	NaN	NaN	in	Jakarta, Indonesia	0	0	0	https://x.com/detikinet/status/1809409369945051...	135815019	detikinet
149	1809408313194692719	Sat Jul 06 02:05:33 +0000 2024	0	Kasus PDNS 2 Investasi Teknologi Keamanan Sibe...	1809408313194692719	NaN	NaN	in	Jakarta - Indonesia	0	0	0	https://x.com/TechnologueID/status/1809408313194...	737114089366741760	TechnologueID
150	1809388875867124217	Sat Jul 06 00:48:19 +0000 2024	1	Kemenkumham Pastikan PDNS 2 Kembali Normal Dan...	1809388875867124217	https://pbs.twimg.com/ext_tw_video_thumb/18093...	NaN	in	NaN	0	0	1	https://x.com/AkfahMazzani/status/1809388875867...	1751098508793499648	AkfahMazzani
151	1809385860766585246	Sat Jul 06 00:36:20 +0000 2024	1	Kelompok peretas Brain Cipher telah memberikan ...	1809385860766585246	https://pbs.twimg.com/media/GRw6KcdaCAMsKII.jpg	NaN	in	NaN	0	0	1	https://x.com/RadioBishinta/status/1809385860767...	121640063	RadioBishinta
152	1809385433392115924	Sat Jul 06 00:34:38 +0000 2024	1	Kemenkumham Pastikan PDNS 2 Kembali Normal Dan...	1809385433392115924	https://pbs.twimg.com/media/GRw6blzacAEMeeJ.jpg	NaN	in	NaN	0	0	1	https://x.com/vanarindya/status/1809385433392111...	1787269476704239616	vanarindya

Gambar 4.18 Tampilan Hasil Akhir Crawling Data

4.2 Processing Raw Data

Proses mengolah data mentah hasil dari *crawling data* akan dibagi menjadi empat tahapan yaitu *categorizing data, duplicates, data cleaning, case folding.* Pada proses mengkategorisasikan data (*categorizing data)* akan dilakukan terlebih dahulu pembagian data menjadi tiga bagian kolom yang akan digunakan sebagai bahan analisis pada tahap selanjutnya yaitu *full_text, username,* dan *created_at*. Kolom *Full text* berisikan baris isi dari berbagai macam *tweet* X, *username* berisikan baris identitas pengguna X, dan *created_at* berisikan baris waktu dari pembuatan *tweet* tersebut. *Pseudocode* program untuk melakukan kategorisasi data dapat dilihat pada Gambar 4.19 dan akan menghasilkan baris beserta kolom seperti pada Gambar 4.20.

```
[ ]  #Membagi Data CSV ke dalam Tiga Bagian Kolom (full_text,username,created_at)

     df = df[['full_text','username','created_at']]

     df.head(100)
```

Gambar 4.19 Pseudocode Program Categorizing Data

	full_text	username	created_at
0	Senyum pahit liat omongan pak lurah RI anggap ...	sutand23	Wed Jul 10 16:56:34 +0000 2024
1	Terkait data pribadi dari instansi yang terdam...	hariankompas	Wed Jul 10 14:39:03 +0000 2024
2	Terkait data pribadi dari instansi yang terdam...	KompasKlasika	Wed Jul 10 14:39:03 +0000 2024
3	Kasus Peretasan PDNS 2 Anggota Komisi I DPR Se...	Beritasatu	Wed Jul 10 11:54:14 +0000 2024
4	Pemerintah terus berupaya memulihkan Pusat Dat...	detikinet	Wed Jul 10 10:43:58 +0000 2024
...	...	...	...
95	ransomware ke PDNS 2 tergolong ke dalam aksi t...	mishaameliaa	Tue Jul 09 03:00:00 +0000 2024
96	Serangan PDNS 2 ransomware sudah ternasuk tero...	JudikaKova	Tue Jul 09 03:00:00 +0000 2024
97	Bak Drama Korea Peretas PDNS 2 Brain Chiper Bi...	SINDOnews	Tue Jul 09 02:45:06 +0000 2024
98	Pakar Pertanyakan Standar Keamanan PDNS 2 http...	Beritasatu	Tue Jul 09 01:31:34 +0000 2024
99	Peningkatan keamanan itu dilakukan agar PDN me...	Azzubairi93	Mon Jul 08 21:46:52 +0000 2024

100 rows × 3 columns

Gambar 4.20 Hasil Categorizing Data

Setelah proses mengkategorisasi data selesai dilakukan. Data akan dibersihkan dari berbagai macam data yang terindikasi duplikat *(duplicates)*, hal ini berguna untuk memudahkan proses menghilangkan data kembar. Proses *duplicates* dimulai dengan mengecek terlebih dahulu jumlah baris dan kolom, setelah itu data barulah akan dilakukan pembersihan duplikat dengan syarat data telah berhasil memalui tahap pengecekan ulang terhadap duplikat data *tweet*. Secara terperinci *pseudocode* dari keseluruhan proses *duplicates* dapat dilihat pada Gambar 4.21.

Setelah proses menghilangkan data duplikat selesai dilakukan, tahap selanjutnya melakukan pembersihan data *(data cleaning)*. Proses membersihkan data ini sendiri bertujuan untuk menghilangan atau menghapus karakter-karakter yang tidak diperlukan seperti *@username, #hashtag, RT(retweet), url* pencarian, karakter *non-alphanumeric*, dan spasi yang berlebihan. *Pseudocode* program dan hasil akhir dari *data cleaning* dapat dilihat pada Gambar 4.22 dan Gambar 4.23.

```
[ ] #Mengecek Jumlah Baris dan Kolom
    df.shape

    (153, 3)

[ ] #Menghilangkan Data Tweet yang Kembar (Terduplikasi)
    df = df.drop_duplicates(subset=['full_text'])

[ ] #Mengecek Apakah Masih Ada Data Tweet yang Kembar (Terduplikasi)
    df.duplicated().sum()

    0

[ ] #Menghapus Nilai Variabel yang Kosong
    df = df.dropna()

[ ] #Menampilkan Hasil Nilai Variabel yang Kosong
    df.isnull().sum()

    full_text     0
    username      0
    created_at    0
    dtype: int64

[ ] #Mengecek Kembali Hasil dari Proses Penyaringan Data
    df.shape

    (153, 3)
```

Gambar 4.21 Pseudocode Program dan Hasil Duplicates

```
[ ] #Proses Membersihkan Data

    def clean_twitter_text(text):

      #Menghapus nama pengguna x
      text = re.sub(r'@[A-Za-z0-9_]+', '', text)

      #Menghapus tagar (hashtag)
      text = re.sub(r'#\w+', '', text)

      #Menghapus retweet (RT)
      text = re.sub(r'RT[\s]+', '', text)

      #Menghapus url pencarian
      text = re.sub(r'https?://\S+', '', text)

      #Menghapus karakter non-alphanumeric
      text = re.sub(r'[^A-Za-z0-9\s]', '', text)

      #Menghapus spasi berlebihan
      text = re.sub(r'\s+', ' ', text).strip()

      return text

    df ['full_text'] = df['full_text'].apply(clean_twitter_text)
```

Gambar 4.22 Pseudocode Program Data Cleaning

```
[ ] #Menampilkan Data yang Telah Dibersihkan

    df.head()
```

	full_text	username	created_at
0	Senyum pahit liat omongan pak lurah RI anggap ...	sutand23	Wed Jul 10 16:56:34 +0000 2024
1	Terkait data pribadi dari instansi yang terdam...	hariankompas	Wed Jul 10 14:39:03 +0000 2024
2	Terkait data pribadi dari instansi yang terdam...	KompasKlasika	Wed Jul 10 14:39:03 +0000 2024
3	Kasus Peretasan PDNS 2 Anggota Komisi I DPR Se...	Beritasatu	Wed Jul 10 11:54:14 +0000 2024
4	Pemerintah terus berupaya memulihkan Pusat Dat...	detikinet	Wed Jul 10 10:43:58 +0000 2024

Gambar 4.23 Hasil Data Cleaning

Proses selanjutnya setelah dilakukan pembersihan data yakni melakukan *case folding*, yaitu proses menyamaratakan kata pada tiap kalimatnya dengan tujuan mempermudah proses analisis selanjutnya. Pada penelitian ini *case folding* akan dilakukan dengan melakukan *lower casing* terhadap data dengan maksud mengubah semua huruf menjadi huruf kecil. *Pseudocode* program dan hasil akhir proses *case folding* dapat dilihat pada Gambar 4.24 dan Gambar 4.25.

```
[ ]    #Mengubah Tweet Menjadi Huruf Kecil

       df ['full_text'] = df['full_text'].str.lower()
```

```
[ ]    #Menampilkan Data Lower Case

       df.head(100)
```

Gambar 4.24 Pseudocode Program Case Folding

	full_text	username	created_at
0	senyum pahit liat omongan pak lurah ri anggap ...	sutand23	Wed Jul 10 16:56:34 +0000 2024
1	terkait data pribadi dari instansi yang terdam...	hariankompas	Wed Jul 10 14:39:03 +0000 2024
2	terkait data pribadi dari instansi yang terdam...	KompasKlasika	Wed Jul 10 14:39:03 +0000 2024
3	kasus peretasan pdns 2 anggota komisi i dpr se...	Beritasatu	Wed Jul 10 11:54:14 +0000 2024
4	pemerintah terus berupaya memulihkan pusat dat...	detikinet	Wed Jul 10 10:43:58 +0000 2024
...	...	...	...
95	ransomware ke pdns 2 tergolong ke dalam aksi t...	mishaameliaa	Tue Jul 09 03:00:00 +0000 2024
96	serangan pdns 2 ransomware sudah ternasuk tero...	JudikaKova	Tue Jul 09 03:00:00 +0000 2024
97	bak drama korea peretas pdns 2 brain chiper bi...	SINDOnews	Tue Jul 09 02:45:06 +0000 2024
98	pakar pertanyakan standar keamanan pdns 2	Beritasatu	Tue Jul 09 01:31:34 +0000 2024
99	peningkatan keamanan itu dilakukan agar pdn me...	Azzubairi93	Mon Jul 08 21:46:52 +0000 2024

100 rows × 3 columns

Gambar 4.25 Hasil Case Folding

4.3 Preprocessing

Setelah proses mengolah data mentah *(processing raw data)* selesai dilakukan, proses akan dilanjutkan dengan melakukan *preprocessing* data yakni sebuah proses mempersiapkan data yang nantinya akan dilakukan analisis. Tahapan dari *preprocessing* data pada penelitian ini dibagi menjadi empat yaitu *normalized*, *stopword*, *tokenized*, dan *stemming*. Tahap *normalized* dilakukan dengan cara menormalisasikan data berupa kata-kata yang nantinya akan

diubah menjadi nilai-nilai bentuk variabel dengan skala berbeda tiap katanya. Pada penelitian ini proses menormalisasikan dilakukan terhadap bentuk kata seperti *'liat'* menjadi kata *'lihat', 'khabar'* menjadi kata *'kabar', 'se7'* menjadi kata *'setuju', 'pj'* menjadi kata *'penanggung jawab', 'wako'* menjadi kata *'wali kota'*, dan *'fyi'* menjadi kata *'sekedar tahu'. Pseudocode* program dan hasil dari proses *normalized* dapat dilihat pada Gambar 4.26 dan Gambar 4.27.

```
[ ]  #Normalized

     norm = {' liat ':' lihat ',' khabar ':' kabar ',' se7 ':' setuju ',' pj ':' penanggung jawab ',' wako ': ' wali kota ',' fyi ':' sekedar tahu '}

     def normalisasi(str_text):
       for i in norm:
         str_text = str_text.replace(i, norm[i])
         return str_text

     df['full_text'] = df['full_text'].apply(lambda x: normalisasi (x))
     df.head()
```

Gambar 4.26 Pseudocode Program Normalized

	full_text	username	created_at
0	senyum pahit lihat omongan pak lurah ri anggap...	sutand23	Wed Jul 10 16:56:34 +0000 2024
1	terkait data pribadi dari instansi yang terdam...	hariankompas	Wed Jul 10 14:39:03 +0000 2024
2	terkait data pribadi dari instansi yang terdam...	KompasKlasika	Wed Jul 10 14:39:03 +0000 2024
3	kasus peretasan pdns 2 anggota komisi i dpr se...	Beritasatu	Wed Jul 10 11:54:14 +0000 2024
4	pemerintah terus berupaya memulihkan pusat dat...	detikinet	Wed Jul 10 10:43:58 +0000 2024

Gambar 4.27 Hasil Normalized

Setelah proses normalized berhasil dilakukan, proses selanjutnya melakukan *install package Sastrawi* seperti pada Gambar 4.28. *Install package Sastrawi* berguna untuk dapat melakukan *import* fungsi yang berguna pada proses analisis selanjutnya.

```
#Install Package Sastrawi

!pip install Sastrawi
```

```
Collecting Sastrawi
  Downloading Sastrawi-1.0.1-py2.py3-none-any.whl.metadata (909 bytes)
Downloading Sastrawi-1.0.1-py2.py3-none-any.whl (209 kB)
   ━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━ 209.7/209.7 kB 4.5 MB/s eta 0:00:00
Installing collected packages: Sastrawi
Successfully installed Sastrawi-1.0.1
```

Gambar 4.28 Proses Install Package Sastrawi

Setelah proses *install package Sastrawi* barulah proses *stopword* dapat dilakukan. Proses *stopword* merupakan proses menghapus kata-kata umum yang sering muncul dalam suatu teks, baik itu kata yang ada bobot maupun sekedar tidak ada bobot. *Pseudocode* program dan hasil dari proses *stopword* dapat dilihat pada Gambar 4.29 dan Gambar 4.30.

```
[ ] #Stopword

    import Sastrawi
    from Sastrawi.StopWordRemover.StopWordRemoverFactory import StopWordRemoverFactory, StopWordRemover, ArrayDictionary
    more_stop_words = []

    stop_words = StopWordRemoverFactory().get_stop_words()
    stop_words.extend(more_stop_words)

    new_array = ArrayDictionary(stop_words)
    stop_words_remover_new = StopWordRemover(new_array)

    def stop_words(str_text):
      str_text = stop_words_remover_new.remove(str_text)
      return str_text

    df['full_text'] = df['full_text'].apply(lambda x: stop_words(x))

    df.head()
```

Gambar 4.29 Pseudocode Program Stopword

	full_text	username	created_at
0	senyum pahit lihat omongan pak lurah ri anggap...	sutand23	Wed Jul 10 16:56:34 +0000 2024
1	terkait data pribadi instansi terdampak ganggu...	hariankompas	Wed Jul 10 14:39:03 +0000 2024
2	terkait data pribadi instansi terdampak ganggu...	KompasKlasika	Wed Jul 10 14:39:03 +0000 2024
3	kasus peretasan pdns 2 anggota komisi i dpr se...	Beritasatu	Wed Jul 10 11:54:14 +0000 2024
4	pemerintah terus berupaya memulihkan pusat dat...	detikinet	Wed Jul 10 10:43:58 +0000 2024

Gambar 4.30 Hasil Stopword

Setelah proses *stopword* berhasil dilakukan, proses selanjutnya melakukan *tokenized*, proses *tokenized* akan dilakukan pengubahan suatu teks atau kalimat menjadi penggalan kata per kata yang lebih kecil. Seperti mengubah kalimat *(senyum pahit liat omongan pak lurah)*, kalimat tersebut akan berubah menjadi *(senyum, pahit, liat, omongan, pak, lurah)*. *Pseudocode* program dan hasil dari proses *tokenized* dapat dilihat pada Gambar 4.31 dan Gambar 4.32.

```
[ ] #Tokenized

    tokenized = df ['full_text'].apply(lambda x:x.split())
    tokenized
```

Gambar 4.31 Pseudocode Program Tokenized

```
[ ]  #Tokenized

     tokenized = df ['full_text'].apply(lambda x:x.split())
     tokenized
```

```
0      [senyum, pahit, lihat, omongan, pak, lurah, ri...
1      [terkait, data, pribadi, instansi, terdampak, ...
2      [terkait, data, pribadi, instansi, terdampak, ...
3      [kasus, peretasan, pdns, 2, anggota, komisi, i...
4      [pemerintah, terus, berupaya, memulihkan, pusa...
                             ...
148    [ransomware, serang, pdns, 2, sita, perhatian,...
149    [kasus, pdns, 2, investasi, teknologi, keamana...
150    [kemenkumham, pastikan, pdns, 2, normal, dibac...
151    [kelompok, peretas, brain, cipher, memberikan,...
152    [kemenkumham, pastikan, pdns, 2, normal, dibac...
Name: full_text, Length: 153, dtype: object
```

Gambar 4.32 Hasil Tokenized

Setelah proses *tokenized* berhasil dilakukan, langkah berikutnya yaitu melakukan proses *stemming* atau *lemmatization*, proses ini akan mengurangi atau mengembalikan kata dalam suatu teks menjadi bentuk kata dasarnya seperti kata imbuhan berupa awalan dan akhiran, sehingga kata yang tersisa hanya kata dasar. *Pseudocode* program dan hasil dari proses stemming dapat dilihat pada Gambar 4.33 dan Gambar 4.34.

```
[ ]  #Lemmatization or Stemming

     from Sastrawi.Stemmer.StemmerFactory import StemmerFactory

     def stemming(text_cleaning):
       factory = StemmerFactory()
       stemmer = factory.create_stemmer()
       do = []

       for w in text_cleaning:
         dt = stemmer.stem(w)
         do.append(dt)
       d_clean = []
       d_clean = " ".join(do)

       print(d_clean)
       return d_clean

     tokenized = tokenized.apply(stemming)

     #Menyimpan Data Hasil Stemming

     tokenized.to_csv("/content/drive/My Drive/Skripsi/tokenized2_data.csv", index=False)
     data = pd.read_csv("/content/drive/My Drive/Skripsi/tokenized2_data.csv", encoding='latin1')
```

Gambar 4.33 Pseudocode Program Stemming

```
senyum pahit lihat omong pak lurah ri anggap kecil hacker pdns nunggu serang ransomware babu jilid 2
kait data pribadi instansi dampak ganggu pdns 2 brain chiper duga sandera data sistem
kait data pribadi instansi dampak ganggu pdns 2 brain chiper duga sandera data sistem
kasus retas pdns 2 anggota komisi i dpr sebut retas kelola data pidana
perintah terus upaya pulih pusat data nasional pdns 2 kena serang ransomware brain cipher sejak 20 juni lalu
buntut pdns 2 retas erick thohir dorong audit investigasi telkomsigma
kirain update pdns pdns apa khabar min
akamsi desak jokowi tanggung jawab atas bobol pdns 2
akamsi geruduk kominfo buntut pdns 2 bobol desak budi arie mundur
pdns 2 bobol hacker erick thohir kalau kompeten copot
hacker ransomware brain cipher klaim hapus semua data pdns 2
belakang media sosial ramai bahas kena kasus retas pdns tarik perhati banyak orang
perintah indonesia kini hadap tantang besar pulih pusat data nasional pdns 2 ganggu akibat serang ransomwar
satu firma intelijen siber klaim brain cipher hapus data mereka curi server pusat data nasional pdns 2 suraba
hai kak saat sedang masalah pdns2 sistem kip kuliah kembali operasi sepenuh paling lambat 29 juli 2024 proses
manajemen pt indosat tbk buka suara kait berita kena libat persero kasus atas pusat data nasional pdns 2
ingin tahu apa benar jadi putar ransomware pdns sila hadir mas yohanes ungkap benar 2 ransomware bukan cuma 1
kemenkominfo jelas tiga zona pulih data imbas insiden pdns 2
kemenkominfo jelas tiga zona pulih data imbas insiden pdns 2
pakar it ridho rahmadi respons bobol pusat data nasional pdns 2 surabaya pun dorong perintah perlu timbang un
ahli bicara alas balik retas pdns 2 dadak baik hati
ransomware serang pdns rakyat jd korban 1 kalian terlalu bodoh jaga aman 2 kalian saking bodoh ndak punya bac
pasca serang hacker pdns 2 pj wako pekanbaru tinjau layan mpp
pasca serang hacker pdns 2 pj wako pekanbaru tinjau layan mpp
pusat data nasional pdns 2 dalam proses pulih layan usai kena ransomware apjii minta perintah bentuk satgas a
3 zona pulih data imbas serang siber pdns 2
brain cipher klaim hapus data curi pdns 2
kunci dekripsi serang ransomware pdns 2 hasil buka simak lapor lengkap tahu kembang baru
perintah tetap strategi pulih jangka pendek tengah panjang pdns 2
perintah tetap strategi pulih jangka pendek tengah panjang pdns 2
imbas retas ransomware pdns 2 surabaya dirjen aptika kemenkominfo semuel abrijani pangerapan mundur jabat eff
usai serang siber pdns 2 perintah terap tiga zona pulih data
perintah tegas proyek bangun pusat data nasional pdn tetap lanjut meski serang ransomware pdns 2
```

Gambar 4.34 Hasil Stemming

4.4 Translate

Setelah proses *preprocessing* dilakukan, langkah selanjutnya data hasil *preprocessing* melalui proses *translate*. Proses *translate* bertujuan untuk mengubah teks atau kalimat dari suatu bahasa ke bentuk bahasa lain yang diinginkan. Pada penelitian ini data akan dierjemahkan ke dalam Bahasa Inggris. Hal ini tentu untuk mendapatkan dua jenis bahasa yang diinginkan yaitu Bahasa Indonesia dan Bahasa Inggris. Sebelum melakukan *translate*, proses dimulai dengan *install library translate* seperti pada Gambar 4.35. Setelah itu akan dilakukan pemanggilan data hasil *preprocessing* seperti pada Gambar 4.36, pada proses ini hanya menampilkan kolom *full_text* yang berisi cuitan pengguna X (t*weets*), hal ini dikarenakan pada penelitian analisis sentimen ini hanya berfokus pada opini atau pendapat saja. Setelah itu barulah melakukan proses *translate* sekaligus hasil data *translate* akan disimpan dalam format *csv* pada folder *data2_english.csv* seperti pada Gambar 4.37.

```
#Install library

!pip install translate

Collecting translate
  Downloading translate-3.6.1-py2.py3-none-any.whl.metadata (7.7 kB)
Requirement already satisfied: click in /usr/local/lib/python3.10/dist-packages (from translate) (8.1.7)
Requirement already satisfied: lxml in /usr/local/lib/python3.10/dist-packages (from translate) (4.9.4)
Requirement already satisfied: requests in /usr/local/lib/python3.10/dist-packages (from translate) (2.31.0)
Collecting libretranslatepy==2.1.1 (from translate)
  Downloading libretranslatepy-2.1.1-py3-none-any.whl.metadata (233 bytes)
Requirement already satisfied: charset-normalizer<4,>=2 in /usr/local/lib/python3.10/dist-packages (from requests->translate) (3.3.2)
Requirement already satisfied: idna<4,>=2.5 in /usr/local/lib/python3.10/dist-packages (from requests->translate) (3.7)
Requirement already satisfied: urllib3<3,>=1.21.1 in /usr/local/lib/python3.10/dist-packages (from requests->translate) (2.0.7)
Requirement already satisfied: certifi>=2017.4.17 in /usr/local/lib/python3.10/dist-packages (from requests->translate) (2024.7.4)
Downloading translate-3.6.1-py2.py3-none-any.whl (12 kB)
Downloading libretranslatepy-2.1.1-py3-none-any.whl (3.2 kB)
Installing collected packages: libretranslatepy, translate
Successfully installed libretranslatepy-2.1.1 translate-3.6.1
```

Gambar 4.35 Proses Install Library Translate

```
[ ]  #Menampilkan Data Hasil Preprocessing

     data = pd.read_csv("/content/drive/My Drive/Skripsi/tokenized2_data.csv", encoding='latin1')

     data.head()
```

	full_text
0	senyum pahit lihat omong pak lurah ri anggap k...
1	kait data pribadi instansi dampak ganggu pdns ...
2	kait data pribadi instansi dampak ganggu pdns ...
3	kasus retas pdns 2 anggota komisi i dpr sebut ...
4	perintah terus upaya pulih pusat data nasional...

Gambar 4.36 Data Hasil Preprocessing

```
[ ]  #Melakukan Proses Translate

     from translate import Translator

     def convert_eng(tweet):
         translator = Translator(to_lang="en", from_lang="id")
         translation = translator.translate(tweet)
         return translation

     data['tweet_english'] = data['full_text'].apply(convert_eng)
     data.to_csv("/content/drive/My Drive/Skripsi/data2_english.csv")
```

Gambar 4.37 Pseudocode Program Translate

4.5 Labelling

Setelah proses *translate* berhasil dilakukan, langkah selanjutnya melakukan *labelling* terhadap data yang akan dianalisis. *Labelling* merupakan proses pemberian label atau mengkategorikan suatu data yang tidak memiliki kategori pelabelan sebelumnya, hal ini menjadi bagian tahapan yang penting dalam banyak pengaplikasian *machine learning* dan analisis data yang memerlukan data terstruktur, sehingga memberikan kemudahan sistem model analisis dalam memahami serta penerapan pengaplikasian analisis yang akan digunakan. Sebelum melakukan *labelling*, proses dimulai dengan pemanggilan terhadap data hasil *translate* terlebih dahulu, *data2_english* akan menampilkan dua data yang sudah terlabel yakni data dengan label berbahasa Indonesia pada kolom *full_text* dan berbahasa Inggris pada kolom *tweet_english.* Adapun hasil dari *labelling* dapat dilihat pada Gambar 4.38.

```
[ ]  #Menampilkan Labelling dan Data Hasil Translate

     data = pd.read_csv("/content/drive/My Drive/Skripsi/data2_english.csv", index_col=0)

     data.head(100)
```

	full_text	tweet_english
0	senyum pahit lihat omong pak lurah ri anggap k...	bitter smile see talk sir lurah ri think small...
1	kait data pribadi instansi dampak ganggu pdns ...	hook personal data agency impact disturb pdns ...
2	kait data pribadi instansi dampak ganggu pdns ...	hook personal data agency impact disturb pdns ...
3	kasus retas pdns 2 anggota komisi i dpr sebut ...	hack case pdns 2 members of commission i dpr c...
4	perintah terus upaya pulih pusat data nasional...	order to continue efforts to recover the natio...
...	...	...
95	ransomware pdns 2 golong dalam aksi terorisme ...	ransomware pdns 2 golong in cyber terrorism ac...
96	serang pdns 2 ransomware ternasuk terorisme si...	attack pdns 2 ransomware including cyber terro...
97	bak drama korea retas pdns 2 brain chiper bila...	korean drama like hack pdns 2 brain chiper sai...
98	pakar tanya standar aman pdns 2	experts ask pdns 2 safe standards
99	tingkat aman laku pdn jadi pusat data kuat tid...	the level of safe selling of pdn so a strong d...

100 rows × 2 columns

Gambar 4.38 Proses Labelling

4.6 Sentiment Classification

Setelah proses pelabelan ke dalam Bahasa Inggris, langkah berikutnya adalah melakukan proses *sentiment classification* (klasifikasi sentimen). Klasifikasi sentimen dapat diartikan sebagai suatu proses menganalisis teks atau kalimat dengan tujuan memudahkan dalam mengenali dan mengklasifikasikan jenis sentimen atau opini seperti apa yang terdapat di dalam suatu teks. Sebelum melakukan klasifikasi sentimen, terlebih dahulu melakukan proses *install* beberapa *library* yakni *tweet-preprocessor*, *textblob*, dan *nltk*. Proses *install library* dapat dilihat pada Gambar 4.39 dan Gambar 4.40. Setelah dilakukannya *install library*, langkah selanjutnya melakukan *import library* sekaligus melakukan *download punkt* seperti pada Gambar 4.41.

```
[ ]  #Install Library

     !pip install tweet-preprocessor
     !pip install textblob
     !pip install wordcloud
     !pip install nltk
```

Gambar 4.39 Pseudocode Program Install Library

```
[ ] #Install Library

    !pip install tweet-preprocessor
    !pip install textblob
    !pip install wordcloud
    !pip install nltk
```

```
Collecting tweet-preprocessor
  Downloading tweet_preprocessor-0.6.0-py3-none-any.whl.metadata (5.9 kB)
Downloading tweet_preprocessor-0.6.0-py3-none-any.whl (27 kB)
Installing collected packages: tweet-preprocessor
Successfully installed tweet-preprocessor-0.6.0
Requirement already satisfied: textblob in /usr/local/lib/python3.10/dist-packages (0.17.1)
Requirement already satisfied: nltk>=3.1 in /usr/local/lib/python3.10/dist-packages (from textblob) (3.8.1)
Requirement already satisfied: click in /usr/local/lib/python3.10/dist-packages (from nltk>=3.1->textblob) (8.1.7)
Requirement already satisfied: joblib in /usr/local/lib/python3.10/dist-packages (from nltk>=3.1->textblob) (1.4.2)
Requirement already satisfied: regex>=2021.8.3 in /usr/local/lib/python3.10/dist-packages (from nltk>=3.1->textblob) (2024.5.15)
Requirement already satisfied: tqdm in /usr/local/lib/python3.10/dist-packages (from nltk>=3.1->textblob) (4.66.4)
Requirement already satisfied: wordcloud in /usr/local/lib/python3.10/dist-packages (1.9.3)
Requirement already satisfied: numpy>=1.6.1 in /usr/local/lib/python3.10/dist-packages (from wordcloud) (1.25.2)
Requirement already satisfied: pillow in /usr/local/lib/python3.10/dist-packages (from wordcloud) (9.4.0)
Requirement already satisfied: matplotlib in /usr/local/lib/python3.10/dist-packages (from wordcloud) (3.7.1)
Requirement already satisfied: contourpy>=1.0.1 in /usr/local/lib/python3.10/dist-packages (from matplotlib->wordcloud) (1.2.1)
Requirement already satisfied: cycler>=0.10 in /usr/local/lib/python3.10/dist-packages (from matplotlib->wordcloud) (0.12.1)
Requirement already satisfied: fonttools>=4.22.0 in /usr/local/lib/python3.10/dist-packages (from matplotlib->wordcloud) (4.53.1)
Requirement already satisfied: kiwisolver>=1.0.1 in /usr/local/lib/python3.10/dist-packages (from matplotlib->wordcloud) (1.4.5)
Requirement already satisfied: packaging>=20.0 in /usr/local/lib/python3.10/dist-packages (from matplotlib->wordcloud) (24.1)
Requirement already satisfied: pyparsing>=2.3.1 in /usr/local/lib/python3.10/dist-packages (from matplotlib->wordcloud) (3.1.2)
Requirement already satisfied: python-dateutil>=2.7 in /usr/local/lib/python3.10/dist-packages (from matplotlib->wordcloud) (2.8.2)
Requirement already satisfied: six>=1.5 in /usr/local/lib/python3.10/dist-packages (from python-dateutil>=2.7->matplotlib->wordcloud) (1.16.0)
Requirement already satisfied: nltk in /usr/local/lib/python3.10/dist-packages (3.8.1)
Requirement already satisfied: click in /usr/local/lib/python3.10/dist-packages (from nltk) (8.1.7)
```

Gambar 4.40 Proses Install Library Berasil

```
[ ] #Import Library

    import preprocessor as p
    import nltk
    from textblob import TextBlob
    from nltk.stem import PorterStemmer
    from nltk.tokenize import word_tokenize

    nltk.download('punkt')
```

```
[nltk_data] Downloading package punkt to /root/nltk_data...
[nltk_data]   Unzipping tokenizers/punkt.zip.
True
```

Gambar 4.41 Proses Import Library dan Download Package Punkt

Setelah proses *install* dan *import library*, dilanjutkan dengan *download package punkt* agar memudahkan dalam proses memecah teks lebih akurat menjadi unit-unit yang lebih kecil kata-per kata. Setelah proses *download* selesai barulah proses klasifikasi sentimen sederhana dilakukan menggunakan *TextBlob*, proses ini melibatkan kondisi perulangan sekaligus kondisional sehingga tiap kalimat akan dipecah menjadi beberapa bagian, setelah itu barulah akan dilakukan analisis sesuai dengan kondisi yang telah ditentukan. *Pseudocode* program dan hasil proses klasifikasi sentimen dapat dilihat pada Gambar 4.42.

```
[ ]  #Menentukan Sentimen Masyarakat | TextBlob

     data_tweet = list(data['tweet_english'])
     polaritas = 0

     status = []
     total_positif = total_negatif = total_netral = total = 0

     for i, tweet in enumerate(data_tweet):
       analysis = TextBlob(tweet)
       polaritas += analysis.polarity

       if analysis.sentiment.polarity > 0.0:
         total_positif += 1
         status.append('Positif')

       elif analysis.sentiment.polarity == 0.0:
         total_netral += 1
         status.append('Netral')

       else:
         total_negatif += 1
         status.append('Negatif')

       total += 1

     print(f'Hasil Analisis Data:\nPositif = {total_positif}\nNetral = {total_netral}\nNegatif = {total_negatif}')
     print(f'\nTotal Data : {total}')
```

Gambar 4.42 Pseudocode Program Klasifikasi Sentimen TextBlob

```
Hasil Analisis Data
Positif = 78
Netral = 43
Negatif = 32

Total Data : 153
```

```
Hasil Analisis Data
Positif = 78
Netral = 43
Negatif = 32

Total Data : 153
```

Gambar 4.43 Hasil Klasifikasi Sentimen TextBlob

```
Hasil Analisis Data
Positif = 78
Netral = 43
Negatif = 32

Total Data : 153
```

Proses klasifikasi sentimen pada Gambar 4.42 dilakukan sebanyak tiga kali dan menghasilkan tiga kategori sentimen beserta kategori dari setiap sentimen seperti pada Gambar 4.43. Hasil analisis sentimen menunjukkan konsistensi dari ketiga klasifikasi yang dilakukan sentimen bernilai positif berjumlah 78 *tweets*, sentimen bernilai negatif berjumlah 32 *tweets*, sentimen bernilai netral berjumlah 43 *tweets*. Setelah proses analisis tersebut dilakukan, dilanjutkan dengan melabeli terlebih dahulu data yang sudah didapatkan seperti pada Gambar 4.44. Setelah itu akan dilakukan pembuatan *dataset* sekaligus *data training* seperti pada *pseudocode* program Gambar 4.45.

```
[ ] #Menampilkan Data Sentimen Masyarakat | TextBlob

    data ['klasifikasi'] = status

    data.head(100)
```

	full_text	tweet_english	klasifikasi
0	senyum pahit lihat omong pak lurah ri anggap k...	bitter smile see talk sir lurah ri think small...	Negatif
1	kait data pribadi instansi dampak ganggu pdns ...	hook personal data agency impact disturb pdns ...	Netral
2	kait data pribadi instansi dampak ganggu pdns ...	hook personal data agency impact disturb pdns ...	Netral
3	kasus retas pdns 2 anggota komisi i dpr sebut ...	hack case pdns 2 members of commission i dpr c...	Negatif
4	perintah terus upaya pulih pusat data nasional...	order to continue efforts to recover the natio...	Negatif
...	...	...	...
95	ransomware pdns 2 golong dalam aksi terorisme ...	ransomware pdns 2 golong in cyber terrorism ac...	Positif
96	serang pdns 2 ransomware ternasuk terorisme si...	attack pdns 2 ransomware including cyber terro...	Positif
97	bak drama korea retas pdns 2 brain chiper bila...	korean drama like hack pdns 2 brain chiper sai...	Netral
98	pakar tanya standar aman pdns 2	experts ask pdns 2 safe standards	Positif
99	tingkat aman laku pdn jadi pusat data kuat tid...	the level of safe selling of pdn so a strong d...	Positif

100 rows × 3 columns

Gambar 4.44 Hasil Labelling Klasifikasi Sentimen TextBlob

```
[ ] import random

    set_positif = []
    set_negatif = []
    set_netral = []

    for n in dataset:
      if(n[1] == 'Positif'):
        set_positif.append(n)
      elif(n[1] == 'Negatif'):
        set_negatif.append(n)
      else:
        set_netral.append(n)

    #Membagi Data Train
    set_positif = random.sample(set_positif, k=int(len(set_positif)/2))
    set_negatif = random.sample(set_negatif, k=int(len(set_negatif)/2))
    set_netral = random.sample(set_netral, k=int(len(set_netral)/2))

    #Akumulasi Data Train
    train = set_positif + set_negatif + set_netral
    train_set = []

    for n in train:
      train_set.append(n)
```

Gambar 4.45 Pseudocode Program Mempersiapkan Dataset dan Data Train

Proses setelah menyiapkan *dataset* dan *data train.* Tahap selanjutnya mengecek tingkat akurasi yang didapatkan dan melakukan klasifikasi menggunakan algoritma *Naïve Bayes Classifier* seperti pada Gambar 4.46. Hasil perbandingan ketiga percobaan klasifikasi tiap sentimen dimuat dalam tabel pada Gambar 4.47.

```
[ ] #Cek akurasi
    from textblob.classifiers import NaiveBayesClassifier
    cl = NaiveBayesClassifier(train_set)
    print("Akurasi Test: ", cl.accuracy(dataset))
```

```
Akurasi Test:  0.7777777777777778
```

```
#NaiveBayesClassifier
data_tweet = list(data['tweet_english'])
polaritas = 0

status = []
total_positif = total_negatif = total_netral = total = 0

for i, tweet in enumerate(data_tweet):
    analysis = TextBlob(tweet, classifier=cl)

    if analysis.classify() == 'Positif':
        total_positif += 1
    elif analysis.classify() == 'Netral':
        total_netral += 1
    else:
        total_negatif += 1

        status.append(analysis.classify())
        total += 1

print(f'\nHasil Analisis Data NBC:\nPositif = {total_positif}\nNetral = {total_netral}\nNegatif = {total_negatif}')
print(f'\nTotal Data : {total}')
```

Gambar 4.46 Pseudocode Program Proses Cek Akurasi dan Klasifikasi Sentimen NBC

```
Hasil Analisis Data NBC
Positif = 84
Netral = 47
Negatif = 22

Total Data : 153
```

```
Hasil Analisis Data NBC
Positif = 89
Netral = 42
Negatif = 22

Total Data : 153
```

```
Hasil Analisis Data NBC
Positif = 93
Netral = 44
Negatif = 16

Total Data : 153
```

```
[ ] data.head()
```

	full_text	tweet_english	klasifikasi	klasifikasi_bayes
0	senyum pahit lihat omong pak lurah ri anggap k...	bitter smile see talk sir lurah ri think small...	Negatif	Negatif
1	kait data pribadi instansi dampak ganggu pdns ...	hook personal data agency impact disturb pdns ...	Netral	Negatif
2	kait data pribadi instansi dampak ganggu pdns ...	hook personal data agency impact disturb pdns ...	Netral	Negatif
3	kasus retas pdns 2 anggota komisi i dpr sebut ...	hack case pdns 2 members of commission i dpr c...	Negatif	Negatif
4	perintah terus upaya pulih pusat data nasional...	order to continue efforts to recover the natio...	Negatif	Negatif

Gambar 4.47 Hasil Beserta Perbandingan Klasifikasi Sentimen NBC

4.7 Visualization

Tahapan akhir dari proses analisis sentimen yaitu Visualisasi (*Visualization*). Proses visualisasi merupakan penggunaan teknik memvisualisasikan hasil dari analisis sentimen terhadap teks atau data yang telah diproses. Proses ini memudakan dalam hal menyajikan informasi data secara variatif, intuitif, dan mudah dimengerti, hal ini tentu bertujuan agar dapat memberikan kemungkinan menarik kesimpulan dalam mengeksplorasikan data secara lebih efektif. Pada penelitian ini akan dilakukan visualisasi menggunakan *chart diagram* dan *wordcloud.* Adapun *pseudocode* program dan hasil visualisasi analisis sentimen menggunakan *wordcloud* dapat dilihat pada Gambar 4.48 dan Gambar 4.49.

```
[ ]  #Melakukan Visualisasi WordCloud NBC

     from wordcloud import WordCloud, STOPWORDS
     import matplotlib.pyplot as plt

     def plot_cloud(wordcloud):
       plt.figure(figsize=(10, 8))
       plt.imshow(wordcloud,interpolation='bilinear')
       plt.axis('off')
       plt.show()

     all_words = ' '.join([tweets for tweets in data['tweet_english']])

     wordcloud = WordCloud(
           width=3000,
           height=2000,
           random_state=3,
           background_color='white',
           colormap='Blues_r',
           collocations=False,
           stopwords=STOPWORDS
     ).generate(all_words)

     plot_cloud(wordcloud)
```

Gambar 4.48 Pseudocode Program Visualisasi Wordcloud



Gambar 4.49 Hasil Visualisasi Wordcloud

Pemanfaatan visualisasi pada penelitian ini juga dilakukan menggunakan *bar chart*. Berdasarkan hasil analisis yang sudah didapatkan pada Gambar 4.43 dan Gambar 4.46, diperoleh chart diagram berdasarkan *pseudocode* program dari visualisasi analisis sentimen yakni *TextBlob* dan *NBC* seperti pada Gambar 4.50, Gambar 4.51, Gambar 4.52 dan Gambar 4.53.

```
]  #Menampilkan Diagram Batang

   import seaborn as sns

   sns.set_theme()

   labels = ['Positif', 'Negatif', 'Netral']
   counts = [total_positif, total_negatif, total_netral]

   def show_bar_chart(labels, counts, title):
     fig, ax = plt.subplots(figsize= (8, 6))
     bars = ax.bar(labels, counts, color=['#fffafa','#f72323','#fac343'])

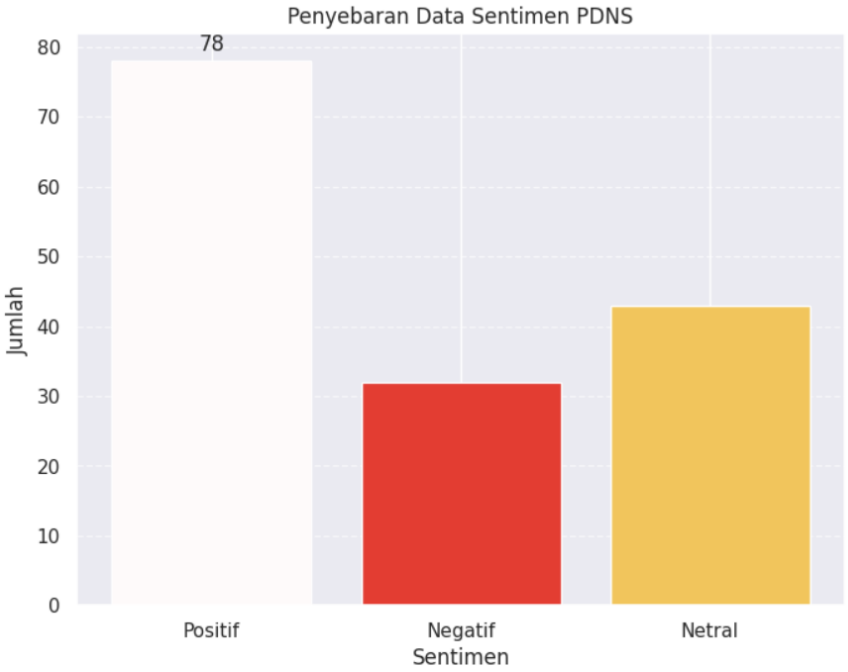
     for bar, count in zip(bars, counts):
         height = bar.get_height()
         ax.annotate(f'{count}', xy = (bar.get_x() + bar.get_width() / 2, height),
                     xytext = (0, 3),
                      textcoords = 'offset points',
                      ha = 'center', va = 'bottom')

         ax.grid(axis = 'y', linestyle='--', alpha=0.7)
         ax.set_xlabel('Sentimen')
         ax.set_ylabel('Jumlah')
         ax.set_title(title)

         plt.show()

   show_bar_chart(labels, counts, "Penyebaran Data Sentimen PDNS")
```

Gambar 4.50 Pseudocode Program Visualisasi Bar Chart TextBlob



Gambar 4.51 Hasil Visualisasi Bar Chart TextBlob

```
[ ]  #Menampilkan Diagram Batang Hasil NBC

     import seaborn as sns

     sns.set_theme()

     labels = ['Positif', 'Negatif', 'Netral']
     counts = [total_positif, total_negatif, total_netral]

     def show_bar_chart(labels, counts, title):
       fig, ax = plt.subplots(figsize=(10,8))
       bars = ax.bar(labels, counts, color=['#FFFAFA','#F72323','#FAC343'])

       for bar, count in zip(bars, counts):
         height = bar.get_height()
         ax.annotate(f'{count}', xy = (bar.get_x() + bar.get_width() / 2, height),
                     xytext = (0,3),
                     textcoords = 'offset points',
                     ha = 'center', va = 'bottom')

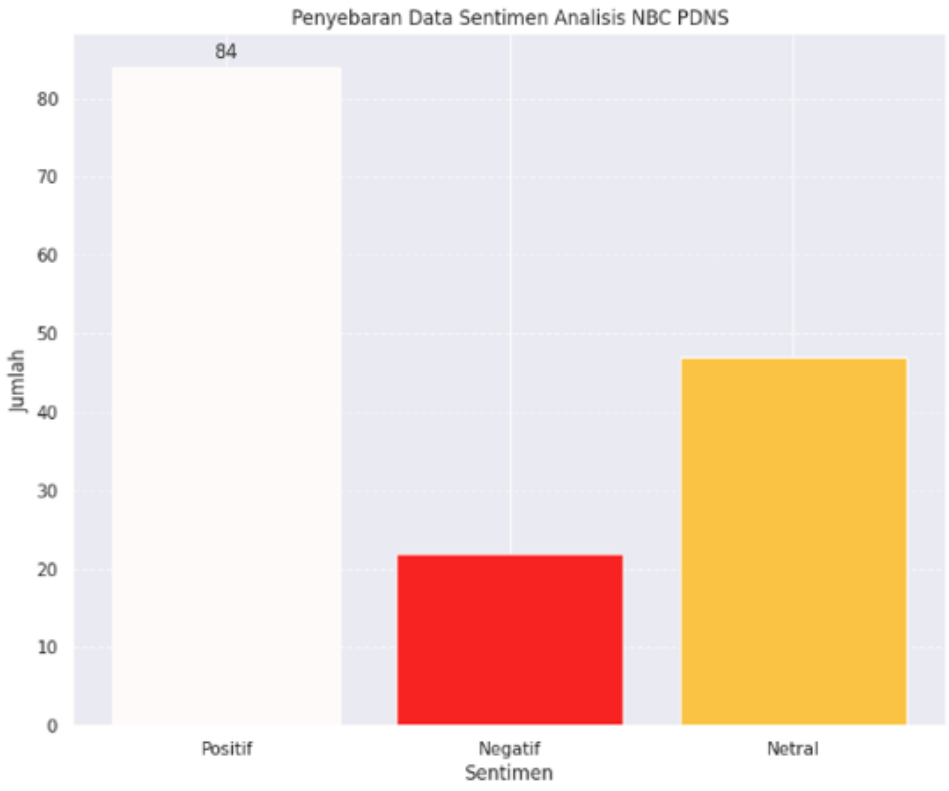
         ax.grid(axis = 'y', linestyle ='--', alpha = 0.7)

         ax.set_xlabel('Sentimen')
         ax.set_ylabel('Jumlah')
         ax.set_title(title)

         plt.show()

     show_bar_chart(labels, counts, "Penyebaran Data Sentimen Analisis NBC PDNS")
```

Gambar4.52 Pseudocode Program Visualisasi Bar Chart NBC



Gambar 4.53 Hasil Visualisasi Bar Chart NBC

4.8 Pembahasan

Percobaan Analisis	Sentimen Positif	Sentimen Netral	Sentimen Negatif	Akurasi
Pertama	84	47	22	0.77
Kedua	89	42	22	0.82
Ketiga	93	44	16	0.78
Rata-rata	88.7	44.3	20	0.79

Tabel 4.2 Hasil Perolehan Analisis NBC

Percobaan Analisis	Sentimen Positif	Sentimen Netral	Sentimen Negatif	Akurasi
Pertama	78	43	32	0.78
Kedua	78	43	32	0.78
Ketiga	78	43	32	0.78
Rata-rata	78	43	32	0.78

Tabel 4.3 Hasil Perolehan Analisis TextBlob

Berdasarkan proses penelitian yang telah dilakukan, diperoleh berbagai temuan hasil terkait analisis sentimen keamanan data sebagai berikut:

a) Hasil *crawling data* yang telah dikumpulkan pada penelitian ini secara menyeluruh sebesar 1039 *data tweets*. Namun sangat disayangkan atas keterbatasan *resource* penulis untuk pengolahan data yang bisa dilakukan hanya sebesar 153 *data tweets* dengan pertimbangan mengambil secara acak dari *total tweets* yang telah didapatkan.

b) Penelitian ini menggunakan dua metode analisis yakni analisis menggunakan *TextBlob* dan *Naïve Bayes Classifier*. Perolehan hasil secara menyeluruh dapat dilihat pada Tabel 4.2 dan Tabel 4.3.

Berdasarkan hasil tersebut diperoleh temuan bahwa perolehan hasil menggunakan analisis sentimen *TextBlob* lebih stabil ketimbang perolehan hasil menggunakan analisis sentimen *Naïve Bayes Classifier*, dengan perolehan rerata pada NBC yakni sentimen positif 88.7, sentimen netral 44.3, sentimen negatif 20, dan akurasi 0.79. Sedangkan perolehan rerata pada *TextBlob* yakni sentimen positif 78, sentimen netral 43, sentimen negatif 32, dan akurasi 0.78.

c) Berdasarkan *pseudocode* pada Gambar 4.42 dan Gambar 4.46, proses analisis sentimen dimulai dengan pengambilan data berupa *DataFrame* yang berisi *tweet* berbahasa Indonesia dan Inggris. Data yang telah diperoleh akan dilakukan penjumlahan total nilai polaritas dari semua *tweet* dengan list status yang diinisia dengan "[]", dan akan dilakukan penghitungan dari tiap variable untuk mendapatkan total jumlah dari masing-masing *tweet* bernilai positif, negatif, dan netral. Implementasi perulangan pada *looping enumerate* untuk iterasi melalui setiap *tweets* dalam data yang didapatkan, dengan "nilai i" sebagai nilai indeks. "analysis.polarity" akan memberikan *nilai polaritas dari -1 (negatif) hingga 1 (positif)*. Nilai ini nantinya akan ditambahkan ke variable polaritas.

Setelah itu barulah akan dilakukan pemeriksaan dari nilai polaritas untuk menentukan apakah sentimen tersebut dikategorikan kedalam positif, negatif, atau netral, dengan beberapa pertimbangan sebagai berikut*: jika polaritas lebih dari 0 maka dianggap sentimen positif, jika polaritas sama dengan 0 maka dianggap sentimen netral, dan jika polaritas kurang dari 0 maka dianggap sentimen negatif*. Setelah itu barulah akan dilakukan peningkatan dari total jumlah kategori yang sesuai dengan menambahkan dari tiap status sentimen ke dalam list "status" sebelumnya dan akan dilakukan *print* dari masing-masing kategori sentimen yakni positif, negatif, dan netral.

d) Hasil visualisasi yang telah dilakukan yakni menggunakan *Bar Chart* dan *WordCloud* yang dapat dilihat pada Gambar 4.51 dan Gambar 4.53. Visualisasi tersebut tentunya menjelaskan bahwa sentimen positif memiliki nilai tertinggi disusul dengan sentimen negatif dan netral. Pada visualisasi *wordcloud* beberapa kata yang sering muncul diantaranya *ransomware, cyber, attack, national, terrorism, pdn,* dan *center*.

```
[ ]  #Melakukan Evaluasi

     data_eval = [tuple(x) for x in data.to_records(index=False)]

     for n in data_eval:
       if len(n) >= 4:
         if n[2] != n[3]:
           print(f'Text: {n[0]}\nClassifier: {n[2]}\nClassifier Bayes: {n[3]}\n')
         else:
           print("Tuple tidak memiliki cukup elemen.")
```

Gambar 4.54 Pseudocode Evaluasi Sentimen

```
Text: kait data pribadi instansi dampak ganggu pdns 2 brain chiper duga sandera data sistem
Classifier: Netral
Classifier Bayes: Negatif

Text: kait data pribadi instansi dampak ganggu pdns 2 brain chiper duga sandera data sistem
Classifier: Netral
Classifier Bayes: Negatif
```

Gambar 4.55 Contoh Hasil Evaluasi Dari Tiap Sentimen

e) Berdasarkan Gambar 4.54 dan Gambar 4.55, penulis mencoba melakukan evaluasi analisis sentimen dari tiap kalimat untuk dapat ditentukan nilai dari tiap sentimen berupa positif, netral, atau negatif. Implementasi evaluasi sentimen menggunakan pengubahan pada *dataframe* menjadi *list of tuples*. Konversi data akan diubah menjadi struktur array *'numpy'* di mana setiap baris dalam dataframe akan diubah menjadi *nilai tuple* dengan *nilai indeks* berupa *'false'*. *List* tersebut akan mengubah struktur array yang telah ditentukan sebelumnya yang akan berisi beberapa *list tuple* dari dataframe yakni data *tweets*.

Selanjutnya akan dilakukan iterasi setiap *tuple 'n'* dalam *list 'data_eval'*. Kemudian barulah akan dilakukan pengecekan melalui kondisional dari *fungsi 'if'* yang telah ditentukan sebagai berikut: jika dari tiap ketiga elemen memenuhi kriteria informasi nilai evaluasi, maka akan dimasukkan kedalam kondisional *n[0]* hingga *n[3]* sebagai elemen pemisah antar analisis tiap sentimen. Namun apabila elemen ketiga dan keempat sama, maka akan mencetak pesan *"Tuple tidak memiliki cukup elemen"*. Setelah itu barulah akan dilakukan *print* terhadap perbandingan antar evaluasi sentimen.

BAB V
PENUTUP

5.1 Kesimpulan

Berdasarkan hasil penelitian tentang Analisis Sentimen X Terhadap Keamanan Data Menggunakan Algoritma Naïve Bayes Classifier: Studi Kasus Pusat Data Sementara (PDNS) dapat disimpulkan bahwa:

- Penelitian yang telah dilakukan menghasilkan dua model analisis sentimen yakni analisis sentimen sederhana menggunakan *TextBlob* dan analisis sentimen menggunakan *Naïve Bayes Classifier (NBC).*
- Penelitian berhasil melakukan perubahan data dari teks Bahasa Indonesia ke dalam bentuk teks Bahasa Inggris, sekaligus melakukan pelabelan dan analisis sentimen sederhana.
- Penelitian menghasilkan data hasil analisis dari dua model *(TextBlob dan Naïve Bayes Classifier)*, Hasil analisis dari total 153 data *tweets*, model TextBlob memaparkan total rerata sentimen positif berjumlah 78 *tweets*, sentimen netral 43 *tweets*, sentimen negatif 32 *tweets*, dan total akurasi 0.78. Sedangkan model *Naïve Bayes Classifier (NBC)* memaparkan total rerata sentimen positif berjumlah 88.7 *tweets*, sentimen netral 44.3 *tweets*, sentimen negatif 20 *tweets*, dan total akurasi 0.79.

5.2 Saran

Berdasarkan penelitian yang telah dilakukan, penulis berharap penelitian ini dapat menjadi wadah untuk pembelajaran dan dapat dikembangkan lagi oleh peneliti lainnya. Penulis memiliki beberapa saran untuk penelitian berikutnya:

- Penelitian selanjutnya diharapkan dapat menggunakan lebih banyak data untuk dilakukan analisis sentimen, mohon maaf atas keterbatasan penulis hanya dapat menganalisis sedikit data sentimen saja.
- Penelitian selanjutnya diharapkan dapat menggunakan berbagai macam algoritma analisis lainnya dan dapat melakukan perbandingan antar algoritma tersebut.
- Penelitian selanjutnya diharapkan dapat menggunakan berbagai macam teknik visualisasi, teknik visual tidak hanya pada *wordcloud* dan *chart diagram* saja.

DAFTAR PUSTAKA

Achsanty, R. A. (2021). *Voting dan Fandom K-Pop (Analisis Komunikasi Antar Penggemar TREASURE dalam Ajakan Voting di Twitter)*. 1–24. http://eprints.ums.ac.id/id/eprint/93813%0Ahttp://eprints.ums.ac.id/93813/2/Naskah Publikasi.pdf

Buntoro, G. A. (2017). Analisis Sentimen Calon Gubernur DKI Jakarta 2017 Di Twitter. *Journal Of Information Technology*.

Cahyaningtiyas, S. (2019). Pemodelan Topik Terhadap Analisis Sentimen Calon Presiden Republik Indonesia Pada Media Sosial Twitter. *Informatika UII*, 80.

He, W., Wu, H., Yan, G., Akula, V., & Shen, J. (2015). A novel social media competitive analytics framework with sentiment benchmarks. *Information and Management*, *52*(7), 801–812. https://doi.org/10.1016/j.im.2015.04.006

Heryadi, Y., & Wahyono, T. (2021). *Dasar-Dasar Deep Learning dan Implementasinya* (I–2021). Penerbit Gava Media.

Indraloka, D. S., & Santosa, B. (2017). Penerapan Text Mining untuk Melakukan Clustering Data Tweet Shopee Indonesia. *Jurnal Sains Dan Seni ITS*, *6*(2), 6–11. https://doi.org/10.12962/j23373520.v6i2.24419

Latuny, W., Lawalata, V. O., Pailin, D. B., & Ohoirenan, R. (2021). Sentiment Analysis of Consumers for Determining the Packaging Features of Eucalyptus Oil Products. *Jurnal Ilmiah Teknik Industri*, *20*(1), 71–80. https://doi.org/10.23917/jiti.v20i1.13461

Lexalytics. (2021). *Sentiment Analysis Explained*. https://www.lexalytics.com/technology/sentiment-analysis

Mukminin, A., Sunarto, M. . D., & Nurcayawati, V. (2021). Analisis Sentimen Publik Terhadap Pelayanan Tes Swab -PCR Covid-19 Di Indonesia Menggunakan Algoritma Support Vector Machine. *Jsika*, *10*(4), 1–13.

Nurfadillah, D. (2023). Analisis Sentimen Mengenai Kesadaran Masyarakat Indonesia Terhadap Keamanan Siber Dalam Menghadapi Kebocoran Data Menggunakan Algoritma Naive Bayes Classifier. *Jurnal Elektro Luceat*, *9*(1), 1–9.

Nursiyono, A. J., & Huda, Q. (2023). Sentiment Analysis Of Personal Data Protection With Machine Learning Approach. *Jurnal Pertahanan & Bela Negara*, *13*(1), 1–16.

Pravina, A. M., Cholissodin, I., & Adikara, P. P. (2019). Analisis Sentimen Tentang opini Maskapai Penerbangan pada Dokumen Twitter Menggunakan Algoritme Support Vector

Machine (SVM). *Jurnal Pengembangan Teknologi Informasi Dan Ilmu Komputer*, *3*(3), 2789–2797. https://doi.org/10.1061/9780784415313.ch06

Sabian, R. (2023). Analisis Sentimen Publik Terhadap BJORKA Dalam Insiden Kebocoran Data KOMINFO Menggunakan Algoritma Support Vector Machine. *Universitas Dinamika*, *55*, ii–55.

Wuragil, Z. (2024). *6 Dampak Serangan Ransomware ke Server PDNS*. Tempo.Co. https://tekno.tempo.co/read/1886038/6-dampak-serangan-ransomware-ke-server-pdns#:~:text=Salah satu dampak besar dari,pendaftar KIP-K ikut raib.

Zalyhaty, L. Q. (2021). Analisis Sentimen Tanggapan Masyarakat Terhadap Vaksin Covid-19 Menggunakan Algoritma Support Vector Machine (SVM). *Universitas Dinamika*, *21*, 1–49. https://doi.org/10.1016/j.solener.2019.02.027%0Ahttps://www.golder.com/insights/block-caving-a-viable-alternative/%0A???

Zy, A., & Wahyu Hadikristanto. (2023). Implementasi Algoritma Metode Naive Bayes dan Support Vector Machine Tentang Pembobolan dan Kebocoran Data di Twitter. *Bulletin of Information Technology (BIT)*, *4*(1), 49–56. https://doi.org/10.47065/bit.v4i1.493

LAMPIRAN