

PROJECT A

ANALISIS SENTIMEN, POS, DAN NER APLIKASI SATUSEHAT PADA ULASAN GOOGLE PLAY STORE DAN ARTIKEL BERITA



KELOMPOK 2

Kresna Dwipayana Ramadhani	5026211114
Anak Agung Istri Istadewanti	5026211143
Akhmad Nabiel Wicaksono	5026211196

PENGOLAHAN BAHASA ALAMI (A)

DEPARTEMEN SISTEM INFORMASI

FAKULTAS TEKNOLOGI ELEKTRO DAN INFORMATIKA CERDAS

INSTITUT TEKNOLOGI SEPULUH NOPEMBER

SEMESTER GASAL 2024/2025

DAFTAR ISI

ABSTRAK	3
1. LATAR BELAKANG	5
2. RUMUSAN MASALAH.....	5
3. DATA ACQUISITION	6
3.1 <i>Scraping</i> Ulasan Pengguna pada Google Play Store	6
3.2 <i>Scraping</i> Link Artikel Berita	8
3.3 <i>Scraping</i> Konten Artikel Berita	9
4. DATA PREPROCESSING	10
4.1 Data Preprocessing Ulasan Pengguna pada Google Play Store.....	10
4.2 Data Preprocessing Artikel Berita	12
5. DEFINISI DATASET.....	13
5.1 Dataset Ulasan Pengguna pada Google Play Store.....	13
5.2 Dataset Artikel Berita	14
5.2.1 Dataset Link Artikel Berita	14
5.2.2 Dataset Konten Artikel Berita	14
5.2.3 Dataset Hasil Preprocessing	15
6. ANALISIS SENTIMEN	15
6.1 Analisis Sentimen Ulasan Pengguna pada Google Play Store	15
6.2 Analisis Sentimen Konten Artikel Berita	19
7. POS DAN NER.....	24
7.1 POS dan NER Ulasan Pengguna pada Google Play Store.....	25
7.2 POS dan NER Konten Artikel Berita	27
8. KESIMPULAN	28
DAFTAR PUSTAKA.....	30

ABSTRAK

Aplikasi SatuSehat, yang sebelumnya dikenal dengan nama PeduliLindungi, merupakan sebuah inisiatif pemerintah Indonesia yang bertujuan untuk mendukung penanganan pandemi COVID-19 melalui digitalisasi layanan kesehatan. Dalam penelitian ini, ulasan pengguna dan pemberitaan terkait aplikasi SatuSehat akan dianalisis. Data diperoleh dari ulasan di Google Play Store dan artikel berita, yang kemudian diproses dengan menghapus *stopwords* dan normalisasi teks. Analisis sentimen menggunakan TextBlob menunjukkan dominasi sentimen negatif dalam ulasan pengguna, terutama setelah transformasi aplikasi, yang terkait dengan masalah akses sertifikat vaksin dan login. Di sisi lain, pemberitaan tentang aplikasi ini menunjukkan peningkatan perhatian seiring dengan transformasinya pada 2023, dengan fokus pada kata-kata seperti kesehatan, aplikasi, dan vaksin, mencerminkan peranannya dalam layanan kesehatan digital.

Kata kunci: SatuSehat, PeduliLindungi, analisis sentimen

ABSTRACT

The SatuSehat application, previously known as PeduliLindungi, is an initiative by the Indonesian government aimed at supporting the handling of the COVID-19 pandemic through the digitalization of healthcare services. This study analyzes user reviews and news articles related to the SatuSehat application. Data was obtained from reviews on the Google Play Store and news articles, which were then processed by removing stopwords and normalizing the text. Sentiment analysis using TextBlob shows a dominance of negative sentiment in user reviews, especially after the transformation of the app, which is linked to issues with vaccine certificate access and login problems. On the other hand, news coverage of the app shows increased attention following its transformation in 2023, with a focus on words like health, application, and vaccine, reflecting its role in digital healthcare services.

Keywords: *SatuSehat, PeduliLindungi, sentiment analysis*

1. LATAR BELAKANG

Perkembangan teknologi digital telah memberikan dampak signifikan dalam berbagai aspek kehidupan, termasuk dalam bidang kesehatan. Transformasi ini menciptakan kesempatan baru untuk meningkatkan akses dan kualitas layanan kesehatan melalui konsep yang dikenal sebagai *Digital Health* (Marpaung & Irwansyah, 2021). *Digital health* memanfaatkan teknologi dan data untuk memantau serta meningkatkan kesehatan pasien. Dengan ini, pemantauan kesehatan menjadi lebih lengkap dan tidak hanya mengandalkan informasi yang didapat saat kunjungan ke dokter (Lowery, 2020).

Salah satu aplikasi yang mencerminkan penerapan konsep *Digital Health* di Indonesia adalah PeduliLindungi. Aplikasi PeduliLindungi adalah alat yang digunakan untuk memantau kesehatan dan mengatasi penyebaran COVID-19. Aplikasi ini dibuat untuk membantu pemerintah dalam melacak penyebaran virus Corona (COVID-19) dan mencegahnya semakin meluas (Septiani et al., 2023). Aplikasi PeduliLindungi menawarkan berbagai fitur bermanfaat, seperti notifikasi untuk pengguna yang berada di kawasan berisiko, pengawasan pergerakan orang terinfeksi, serta akses sertifikat vaksin dan hasil tes COVID-19. Dengan aplikasi ini, masyarakat dapat dengan mudah mengakses layanan publik dan tetap mematuhi protokol kesehatan selama pandemi (Kemenkes, 2021). Setelah pandemi mereda, PeduliLindungi mengalami transformasi menjadi SatuSehat, yang tidak hanya berfokus pada penanganan pandemi tetapi juga merambah ke layanan kesehatan yang lebih luas. Perubahan ini dilakukan untuk menghadirkan aplikasi kesehatan yang lebih komprehensif, mencakup fitur-fitur yang memudahkan masyarakat dalam mengakses layanan kesehatan, konsultasi medis, hingga informasi kesehatan terkini (Kemenkes, 2023).

Untuk menilai sejauh mana efektivitas SatuSehat dalam memenuhi ekspektasi masyarakat, diperlukan analisis yang komprehensif terhadap pengalaman pengguna dan persepsi publik terhadap aplikasi ini. Dalam konteks ini, ulasan dari pengguna di Google Play Store menjadi sumber data yang sangat relevan karena mencerminkan pengalaman nyata dari pengguna aplikasi. Selain itu, artikel berita yang membahas perkembangan aplikasi ini memberikan gambaran lebih luas tentang respons pemerintah, kebijakan kesehatan yang terkait, dan pengaruh aplikasi terhadap masyarakat.

Pengumpulan data dilakukan melalui teknik *web scraping* dari ulasan aplikasi di Google Play Store serta berita terkait SatuSehat. Data yang terkumpul kemudian diproses untuk menghilangkan *noise*, melakukan tokenisasi, dan tahap *preprocessing* lainnya agar siap digunakan dalam analisis lebih lanjut. Analisis ini diharapkan dapat mengungkapkan pola dan tren sentimen publik terhadap SatuSehat, sehingga bisa menjadi dasar dalam mengevaluasi kinerja aplikasi ini dan memberikan rekomendasi perbaikan yang tepat, serta memahami peran aplikasi ini dalam ekosistem kesehatan digital di Indonesia.

2. RUMUSAN MASALAH

Berdasarkan latar belakang yang telah disampaikan, rumusan masalah dalam pembahasan ini adalah sebagai berikut:

1. Bagaimana cara melakukan *web scraping* untuk mengumpulkan tautan berita terkait SatuSehat dan mengekstrak konten artikel tersebut, serta mengambil ulasan pengguna pada aplikasi SatuSehat dari Google Play Store?
2. Bagaimana tahapan data *preprocessing* untuk mempersiapkan dataset pada analisis sentimen terkait SatuSehat?
3. Bagaimana membangun model untuk analisis sentimen terkait SatuSehat?

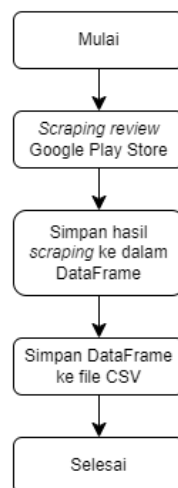
4. Bagaimana sentimen publik terhadap aplikasi SatuSehat?

3. DATA ACQUISITION

Data acquisition adalah proses pengumpulan dan pengambilan data dari berbagai sumber untuk tujuan analisis, penelitian, atau pemrosesan lebih lanjut. Pada proyek ini, pengumpulan data dilakukan untuk membentuk dataset yang diperlukan dalam analisis sentimen terkait pengalaman pengguna. Data yang diperlukan adalah ulasan pengguna dari Google Play Store dan artikel berita yang membahas mengenai aplikasi SatuSehat.

Proses pengumpulan data dilakukan menggunakan metode *web scraping* dengan memanfaatkan library Python, seperti Selenium, BeautifulSoup, dan Requests. Pengumpulan data dilakukan melalui tiga tahap:

3.1 Scraping Ulasan Pengguna pada Google Play Store



Gambar 3.1 Diagram Alir Proses *Scraping* Google Play Store

Pengumpulan data berupa review aplikasi SatuSehat pada Google Play Store dilakukan menggunakan library `google_play_scraper`. Proses *scraping* dimulai dengan memanggil fungsi `'scrap_satusehat_reviews'` yang secara berulang kali mengakses Google Play Store untuk mengumpulkan review aplikasi. Dengan parameter yang disesuaikan, *scraping* dilakukan dengan sejumlah *batch* hingga mencapai jumlah review yang diinginkan. Jumlah review yang didapatkan adalah 350.000 dari sekitar 1,1 juta review yang ada. Batasan jumlah review ini disebabkan oleh keterbatasan *library* dalam mengakses seluruh review.

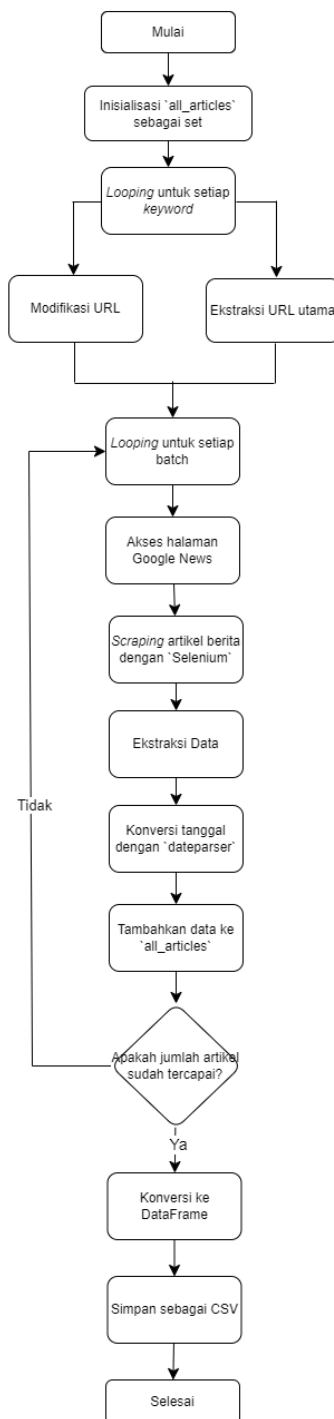
Setelah data terkumpul, proses analisis dilakukan dengan membagi data review ke dalam dua subset berdasarkan tanggal: sebelum 1 Maret 2023 dan setelahnya. Pembagian ini bertujuan untuk melihat perbedaan pola dan sentimen pada dua periode yang berbeda. Subset pertama, yaitu data review sebelum 1 Maret 2023, menyimpan review yang diambil pada periode ketika aplikasi masih dikenal sebagai PeduliLindungi. Subset kedua, yaitu data setelah 1 Maret 2023, berisi review yang berkaitan dengan penggunaan aplikasi setelah transformasi menjadi SatuSehat.

Review yang terkumpul kemudian disimpan ke dalam sebuah DataFrame dari library Pandas, yang mempermudah pengelolaan data untuk proses analisis lebih lanjut karena setiap review disusun dalam format tabular dengan kolom yang lengkap. Data yang telah disimpan

dalam DataFrame diekspor ke dalam dua file CSV, yaitu “satusehat_reviews_before_march_2023.csv”, “satusehat_reviews_after_march_2023.csv”. Proses ekspor ini menggunakan fungsi `to_csv` dari Pandas, sehingga setiap review dapat diakses dengan mudah dalam format CSV dan dapat digunakan kembali untuk proses analisis lainnya.

Informasi yang diambil pada tahap *scraping* review aplikasi SatuSehat meliputi “reviewId”, “userName”, “userImage”, “content”, “score”, “thumbsUpCount”, “reviewCreatedVersion”, “at”, “replyContent”, “repliedAt”, dan “appVersion”. Dengan membagi data review menjadi dua periode waktu, analisis lebih lanjut dapat dilakukan untuk mengevaluasi perubahan persepsi dan sentimen pengguna terkait aplikasi pada periode sebelum dan setelah transformasi menjadi SatuSehat.

3.2 *Scraping* Link Artikel Berita



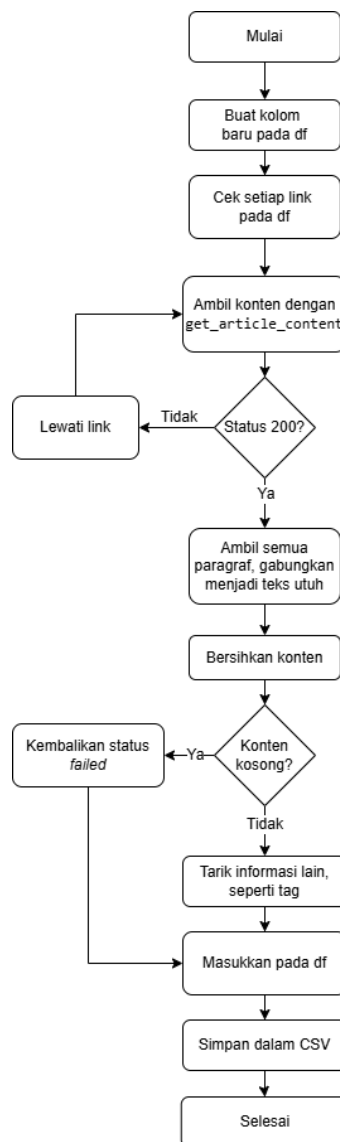
Gambar 3.2 Diagram Alir Proses *Scraping* Link Berita

Proses *scraping* link berita dimulai dengan menginisialisasi sebuah set bernama `'all_articles'`, yang berfungsi untuk menyimpan semua link berita yang ditemukan. Kemudian, program melakukan *loop* melalui daftar kata kunci yang relevan, seperti "Satu Sehat", "Peduli Lindungi", "Peduli Lindungi Covid 19", dan "layanan Peduli Lindungi". Setiap URL pencarian diubah menggunakan fungsi `'rearrange_news_link'`, yang memastikan format URL sesuai untuk pencarian di Google News. Setelah URL dimodifikasi, langkah

selanjutnya adalah mengekstrak URL utama dengan menggunakan fungsi ``get_google_main_url``.

Program kemudian melanjutkan dengan mengakses halaman Google News, di mana artikel-artikel relevan diambil menggunakan Selenium. Data yang diambil mencakup link, judul, tanggal, dan portal berita. Setiap artikel yang berhasil diambil ditambahkan ke dalam set ``all_articles``. Setelah jumlah artikel yang diinginkan tercapai, data yang terkumpul dikonversi menjadi sebuah DataFrame menggunakan library Pandas. Langkah terakhir adalah menyimpan DataFrame tersebut sebagai file CSV dengan nama "link_berita_satuselamat.csv", yang memudahkan untuk analisis lebih lanjut. Selama proses *scraping*, batasan atau target yang diberikan dalam menarik link artikel berita adalah 1.000 per kata kunci. Namun, pada kenyataannya, secara keseluruhan link yang terkumpul hanyalah 312. Hal ini dapat disebabkan oleh beberapa hal, seperti adanya keterbatasan library dalam melakukan *scraping*, pembatasan oleh Google News, atau jumlah berita yang sedikit.

3.3 Scraping Konten Artikel Berita



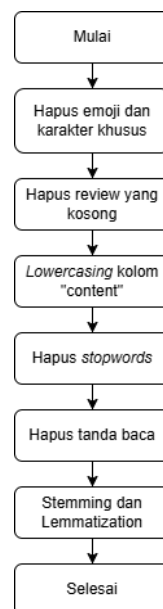
Gambar 3.3 Diagram Alir Proses *Scraping* Konten Artikel Berita

Proses pengambilan konten berita dimulai dengan inisialisasi kolom baru dalam DataFrame `saturehat_news_df` untuk menyimpan informasi terkait, yaitu konten, tags, dan status dari setiap artikel. Selanjutnya, program melakukan *loop* untuk setiap link yang terdapat dalam DataFrame. Pada setiap iterasi, program mencoba mengambil konten dari link yang diberikan dengan menggunakan fungsi `'get_article_content'`. Jika respons dari link menunjukkan status 200 (valid), program akan mengambil semua paragraf dari halaman dan menggabungkannya menjadi satu teks utuh. Konten yang diperoleh kemudian dibersihkan untuk menghapus elemen-elemen yang tidak diinginkan, seperti iklan atau teks promosi. Setelah konten dibersihkan, program memeriksa apakah konten tersebut kosong. Jika tidak, tags akan diambil dari meta tag yang relevan pada halaman. Program kemudian mengembalikan konten, tags, dan status. Apabila status adalah *success*, data yang diperoleh akan disimpan dalam kolom yang sesuai dalam DataFrame, dan jumlah artikel yang berhasil diambil serta yang gagal akan dihitung. Proses ini diulang untuk setiap link dalam DataFrame hingga semua link telah diproses. Di akhir proses, DataFrame yang berisi informasi berita disimpan dalam format CSV dan Excel untuk analisis lebih lanjut. Proses ini melibatkan penggunaan beberapa library Python, termasuk Requests untuk melakukan permintaan HTTP, BeautifulSoup untuk mem-*parsing* HTML, dan Pandas untuk mengelola data dan menyimpan hasil dalam format yang dapat dianalisis.

4. DATA PREPROCESSING

Data preprocessing merupakan tahapan di mana data mentah disiapkan dan dibersihkan sebelum digunakan dalam proses analisis atau pembuatan model machine learning.

4.1 Data Preprocessing Ulasan Pengguna pada Google Play Store



Gambar 4.1 Diagram Alir *Preprocessing* Review Google Play Store

Data yang telah dikumpulkan melalui `google_play_scraper` selanjutnya akan melalui tahap *preprocessing* untuk memastikan data bersih, konsisten, dan siap digunakan dalam analisis. Proses dimulai dengan menghapus emoji dan karakter khusus dari kolom “content” agar elemen-elemen yang tidak relevan tidak mengganggu hasil analisis. Setelah itu, dilakukan pengecekan apakah terdapat review kosong atau bernilai NaN, yang umumnya terjadi pada

review yang hanya berisi emoji atau karakter khusus saja. Review kosong ini kemudian dihapus untuk menjaga kualitas data.

Langkah berikutnya adalah mengonversi semua teks dalam kolom “content” menjadi huruf kecil (*lowercase*) untuk memastikan konsistensi dalam penulisan kata, sehingga variasi huruf kapital tidak memengaruhi analisis. Selain itu, dilakukan penghapusan *stopwords*, yaitu kata-kata umum seperti "dan", "di", atau "yang" yang tidak memberikan informasi signifikan dalam analisis sentimen. Proses ini membantu menyederhanakan data teks dan meningkatkan akurasi model. Selanjutnya, tanda baca juga dihapus untuk memastikan bahwa analisis hanya berfokus pada kata-kata yang relevan. Terdapat juga proses *stemming* atau menghapus imbuhan kata dan mereduksi kata ke bentuk dasar atau akarnya dan *lemmatization* yang mempertimbangkan konteks dan menghasilkan bentuk dasar kata yang lebih tepat. Setelah semua langkah *preprocessing* selesai, data disimpan kembali dalam format CSV untuk digunakan dalam tahap analisis lebih lanjut.

Selain itu, pada *preprocessing* ulasan pengguna terkait SatuSehat pada Google Play Store juga melakukan percobaan lain dimana data yang ditarik sekitar 389 ribu. Adapun data yang diambil meliputi kolom-kolom penting seperti `reviewId`, `userName`, `content`, `score`, `reviewCreatedVersion`, `at`, dan `appVersion`. Tipe data pada kolom tanggal (`at`) kemudian dikonversi menjadi `datetime`. Hal ini bertujuan untuk mempermudah analisis berbasis waktu, seperti pemisahan data berdasarkan periode tertentu. Dengan langkah ini, pengolahan data berbasis waktu dapat dilakukan dengan lebih efisien. Data lalu disimpan dalam format `DataFrame` Pandas, yang akan diekspor ke file CSV untuk mempermudah proses analisis selanjutnya.

Untuk memfasilitasi analisis yang lebih mendalam, data ulasan dipisahkan menjadi dua subset berdasarkan waktu: sebelum dan sesudah 1 Maret 2023. Pembagian ini bertujuan untuk membedakan analisis sentimen pada dua periode yang berbeda, yakni ketika aplikasi masih bernama PeduliLindungi dan setelah berganti menjadi SatuSehat.

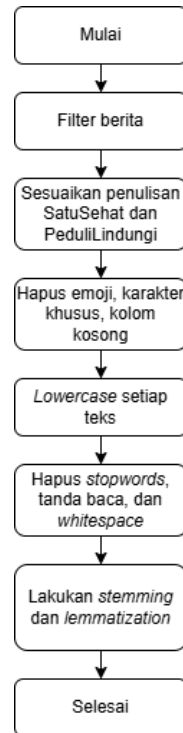
Analisis sentimen dilakukan dengan menggunakan `TextBlob`, yang menghitung nilai *polarity* dan *subjectivity* untuk setiap ulasan. *Polarity* menunjukkan sentimen ulasan (negatif, netral, atau positif) dengan rentang nilai antara -1 hingga 1, sementara *subjectivity* menunjukkan sejauh mana ulasan bersifat subjektif, dengan rentang nilai antara 0 (objektif) hingga 1 (subjektif). Hasil analisis ini digunakan untuk membuat kolom tambahan yang mengelompokkan ulasan ke dalam kategori sentimen positif, netral, dan negatif berdasarkan *score*.

Hasil visualisasi data meliputi histogram distribusi skor ulasan dan scatter plot *polarity vs subjectivity*. Histogram distribusi skor ulasan membantu menampilkan jumlah ulasan berdasarkan skor untuk dua periode yang berbeda, yaitu sebelum dan sesudah 1 Maret 2023. Scatter plot digunakan untuk memvisualisasikan hubungan antara *polarity* dan *subjectivity*, yang dapat membantu mengidentifikasi pola distribusi sentimen ulasan pada kedua periode tersebut.

Penghapusan *stopwords* juga dilakukan sebagai bagian dari *preprocessing*. Daftar *stopwords* Bahasa Indonesia diunduh dan diterapkan pada kolom `content` untuk menghapus kata-kata yang kurang bermakna. *Stopwords* tambahan seperti "agak", "bakal", dan "kemungkinan" juga ditambahkan ke dalam daftar ini. Hasil dari penghapusan *stopwords* adalah teks ulasan yang lebih bersih dan siap digunakan untuk analisis lebih lanjut, seperti klasifikasi dan identifikasi kata kunci.

Setelah melalui proses *preprocessing*, hasilnya disimpan dalam file CSV terpisah untuk setiap subset data (sebelum dan sesudah 1 Maret 2023). Penyimpanan ini bertujuan untuk mempermudah akses dan penggunaan kembali data dalam proses analisis lanjutan, seperti clustering atau pembuatan model prediksi. Langkah-langkah *preprocessing* ini memberikan landasan penting untuk analisis teks berikutnya dan evaluasi performa aplikasi berdasarkan ulasan pengguna.

4.2 Data Preprocessing Artikel Berita



Gambar 4.2 Diagram Alir *Preprocessing* Artikel Berita

Proses pembersihan data dimulai dengan proses pengecekan dan pemfilteran judul serta konten berita secara manual. Setiap artikel yang tidak menggunakan Bahasa Indonesia dihapus dari data agar hanya artikel yang relevan yang diproses lebih lanjut. Jumlah akhir setelah melakukan pemfilteran adalah 161 link. Langkah ini membantu memastikan konsistensi dan kualitas data yang akan dianalisis.

Setelah pemfilteran, data yang telah dipilih diunggah ke dalam Google Colab dan disimpan dalam bentuk DataFrame. Di tahap ini, dilakukan penyesuaian pada nama variabel di kolom "judul" dan "content" untuk menjaga konsistensi terminologi, seperti mengganti kata "Peduli Lindungi" menjadi "PeduliLindungi" dan "Satu Sehat" menjadi "SatuSehat." Penyesuaian ini bertujuan untuk memastikan keseragaman kata kunci penting dalam analisis lebih lanjut.

Tahap berikutnya adalah pembersihan karakter khusus dalam teks. Emoji dan simbol yang tidak relevan dihapus menggunakan teknik *regular expressions* (Regex) untuk menghindari noise dalam data. Pada tahap ini juga dilakukan penghapusan entri di kolom "content" yang kosong. Selanjutnya, teks diubah menjadi huruf kecil (*lowercasing*) agar tidak ada perbedaan antara huruf besar dan kecil, yang penting dalam proses analisis teks.

Proses pembersihan dilanjutkan dengan penghapusan kata-kata yang tidak memiliki makna signifikan, atau yang biasa disebut *stopwords*. Meski demikian, masih ada kemungkinan beberapa kata atau simbol yang tidak relevan tetap ada pada kolom "content". Oleh karena itu, dilakukan visualisasi untuk menampilkan 100 kata teratas dalam bentuk diagram, sehingga kata

atau simbol yang tidak bermakna dapat diidentifikasi dan dihapus. Selain itu, terdapat langkah penghapusan tanda baca dan *whitespace* berlebih untuk memastikan teks lebih rapi dan seragam.

Pada tahap berikutnya, dilakukan proses *stemming* dan *lemmatization*. Proses *stemming*, yang menggunakan library Sastrawi untuk Bahasa Indonesia, menghapus imbuhan pada kata sehingga menghasilkan bentuk dasar yang lebih sederhana. Sementara itu, *lemmatization*, yang dilakukan dengan library NLTK, mengubah kata menjadi bentuk leksikal dasarnya. Kedua proses ini tidak dilakukan bersamaan karena masing-masing memiliki tujuan berbeda. *Stemming* berfokus pada pemangkasan imbuhan, sedangkan *lemmatization* mempertahankan makna dasar kata dalam konteks bahasa.

5. DEFINISI DATASET

Bagian ini menjelaskan struktur dataset yang diperoleh setelah melalui tahapan *data acquisition* dan *data preprocessing*.

5.1 Dataset Ulasan Pengguna pada Google Play Store

Dataset ini berisi ulasan aplikasi SatuSehat yang diperoleh melalui scraping di Google Play Store menggunakan library `google_play_scraper`. Ulasan dikumpulkan dalam format tabular dengan berbagai informasi terkait review pengguna. Berikut adalah penjelasan tentang kolom-kolom dalam dataset ulasan:

No.	Kolom	Definisi
1.	<i>reviewId</i>	ID unik untuk setiap ulasan pengguna
2.	<i>userName</i>	Nama pengguna yang memberikan ulasan
3.	<i>content</i>	Isi ulasan yang diberikan pengguna
4.	<i>score</i>	Skor ulasan yang diberikan pengguna (1-5)
5.	<i>thumbsUpCount</i>	Jumlah 'like' yang diberikan untuk ulasan tersebut
6.	<i>reviewCreatedVersion</i>	Versi aplikasi SatuSehat saat ulasan diberikan
7.	<i>at</i>	Tanggal dan waktu ketika ulasan dibuat
8.	<i>replyContent</i>	Isi balasan dari pengembang aplikasi, jika ada
9.	<i>repliedAt</i>	Tanggal dan waktu ketika balasan diberikan oleh pengembang
10.	<i>appVersion</i>	Versi aplikasi yang digunakan pengguna saat memberikan ulasan
11.	<i>sentiment_polarity</i>	Skor polaritas sentimen ulasan (dihitung menggunakan TextBlob)
12.	<i>sentiment_subjective</i>	Skor subjektivitas ulasan (dihitung menggunakan TextBlob)
13.	<i>sentiment_rating</i>	Kategori sentimen (Positif, Netral, Negatif) berdasarkan skor ulasan
14.	<i>wordCount</i>	Jumlah kata dalam konten ulasan
15.	<i>uppercaseCharCount</i>	Jumlah karakter huruf kapital dalam konten ulasan
16.	<i>specialCharCount</i>	Jumlah karakter spesial dalam konten ulasan
17.	<i>content_without_stopwords</i>	Konten ulasan setelah penghapusan stopwords untuk analisis teks lebih lanjut

Dataset ini mencakup ulasan yang diambil dalam dua periode berbeda: sebelum 1 Maret 2023 dan setelahnya. Hal ini dilakukan untuk melihat perbedaan sentimen dan karakteristik ulasan antara periode sebelum perubahan nama aplikasi dari PeduliLindungi menjadi SatuSehat, serta setelah perubahan tersebut. Dataset dipecah menjadi dua file CSV:

- `satussehat_reviews_before_march_2023_stopwords_removed.csv` untuk ulasan sebelum 1 Maret 2023.
- `satussehat_reviews_after_march_2023_stopwords_removed.csv` untuk ulasan setelah 1 Maret 2023.

Dataset ini memuat ulasan lengkap pengguna dan dilengkapi dengan analisis sentimen yang mendalam, informasi statistik, serta pembersihan teks untuk keperluan analisis lebih lanjut, seperti model klasifikasi atau pengelompokan ulasan berdasarkan karakteristik tertentu.

5.2 Dataset Artikel Berita

5.2.1 Dataset Link Artikel Berita

Dataset ini berisi link artikel berita yang diperoleh melalui proses *scraping* menggunakan library Selenium dan Newspaper pada Google News. Link-link ini dikumpulkan melalui *looping* berdasarkan pencarian dengan kata kunci seperti "Satu Sehat," "Peduli Lindungi," "Peduli Lindungi Covid-19," dan "layanan Peduli Lindungi." Dari hasil pencarian tersebut, diperoleh *link* sebanyak 312. Berikut merupakan penjelasan lebih lanjut terkait dataset.

No.	Kolom	Definisi
1.	<i>link</i>	URL artikel berita yang diambil dari hasil <i>scraping</i>
2.	<i>judul</i>	Judul artikel yang mencerminkan isi dari berita tersebut
3.	<i>tanggal</i>	Tanggal publikasi artikel berita
4.	<i>portal</i>	Nama sumber atau portal berita tempat artikel diterbitkan

5.2.2 Dataset Konten Artikel Berita

Dataset ini berisi konten artikel berita yang berhubungan dengan aplikasi SatuSehat yang didapatkan melalui proses *scraping* menggunakan library BeautifulSoup dan requests. Selain itu, dataset ini juga dilengkapi dengan informasi pendukung lainnya, beberapa diantaranya merupakan bagian dari dataset link artikel berita. Penjelasan lebih lanjut terkait dataset dapat dilihat pada tabel di bawah ini.

No.	Kolom	Definisi
1.	<i>link</i>	URL artikel berita yang diambil dari hasil <i>scraping</i>
2.	<i>judul</i>	Judul artikel yang mencerminkan isi dari berita tersebut
3.	<i>tanggal</i>	Tanggal publikasi artikel berita
4.	<i>portal</i>	Nama sumber atau portal berita tempat artikel diterbitkan
5.	<i>content</i>	Konten lengkap dari artikel berita yang telah diambil
6.	<i>tags</i>	Kata kunci atau <i>tag</i> yang relevan dengan isi artikel
7.	<i>status</i>	Status pengambilan artikel, yang menunjukkan apakah proses pengambilan berhasil atau tidak.

Link yang berhasil diambil isi kontennya adalah 174, sedangkan yang gagal adalah 101. Kemudian dengan pemfilteran *link* secara manual, jumlah *link* saat ini adalah 161.

5.2.3 Dataset Hasil *Preprocessing*

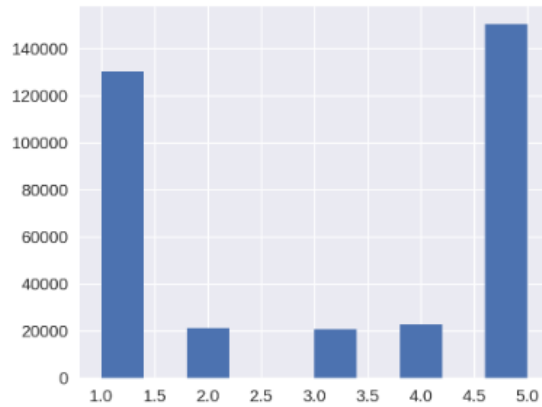
Dataset ini berisi konten artikel berita yang telah dibersihkan serta mengalami proses *stemming* dan *lemmatization* dari tahap *preprocessing*. Dataset ini terdiri dari sejumlah kolom, di mana definisi masing-masing kolom dapat ditemukan pada tabel di bawah ini.

No.	Kolom	Definisi
1.	<i>link</i>	URL artikel berita yang diambil dari hasil <i>scraping</i>
2.	<i>judul</i>	Judul artikel yang mencerminkan isi dari berita tersebut
3.	<i>tanggal</i>	Tanggal publikasi artikel berita
4.	<i>portal</i>	Nama sumber atau portal berita tempat artikel diterbitkan
5.	<i>content</i>	Konten lengkap dari artikel berita yang telah diambil
6.	<i>tags</i>	Kata kunci atau <i>tag</i> yang relevan dengan isi artikel
7.	<i>status</i>	Status pengambilan artikel, yang menunjukkan apakah proses pengambilan berhasil atau tidak.
8.	<i>tahun</i>	Tahun publikasi artikel, diambil dari kolom tanggal
9.	<i>wordCount</i>	Jumlah total kata dalam konten artikel
10.	<i>content_without_stopwords</i>	Konten artikel setelah menghapus kata-kata yang tidak bermakna (<i>stopwords</i>)
11.	<i>wordcount_after_stopwords</i>	Jumlah total kata dalam konten setelah penghapusan <i>stopwords</i>
12.	<i>content_stem_without_stopwords</i>	Konten artikel setelah dilakukan <i>stemming</i> dan penghapusan <i>stopwords</i>
13.	<i>wordcount_stem_without_stopwords</i>	Jumlah total kata dalam konten setelah dilakukan <i>stemming</i> dan penghapusan <i>stopwords</i>
14.	<i>content_lemmatized</i>	Konten artikel setelah dilakukan <i>lemmatization</i> untuk mengubah kata-kata ke bentuk dasarnya
15.	<i>wordcount_lemmatized</i>	Jumlah total kata dalam konten setelah dilakukan <i>lemmatization</i>

6. ANALISIS SENTIMEN

6.1 Analisis Sentimen Ulasan Pengguna pada Google Play Store

Analisis sentimen terhadap ulasan aplikasi SatuSehat di Google Play Store menunjukkan gambaran menarik tentang persepsi pengguna. Pada awal percobaan, terdapat sekitar 350 ribu data lebih ulasan yang dianalisis, persebaran rating didominasi oleh rating 5 dan 1 seperti pada Gambar 6.1, yang mencerminkan opini pengguna yang sangat positif atau sangat negatif. Dalam analisis ini, dibuat *sentiment_rating* dimana sentimen positif (rating 4 dan 5), netral (rating 3), dan negatif (rating 1 dan 2). Setelah itu, model TextBlob digunakan untuk mengevaluasi polaritas (positif/negatif) dan subjektivitas (opini personal atau fakta) dari setiap ulasan.



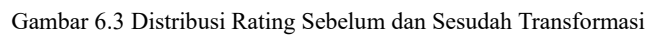
Gambar 6.1 Persebaran Rating Pengguna pada Google Play Store 350 ribu Data

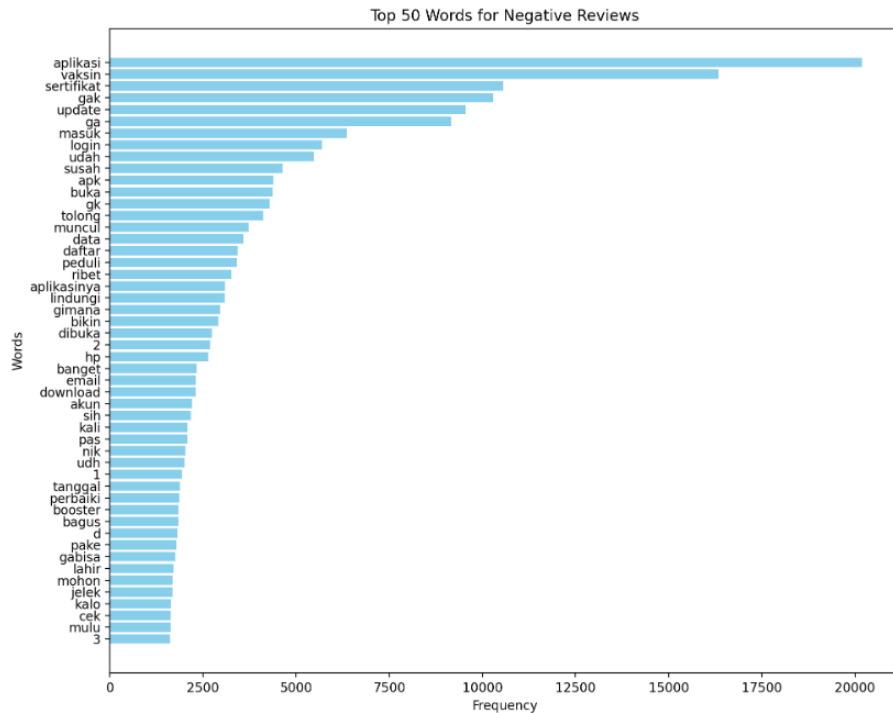
Berdasarkan tabel dan visualisasi pada *Gambar 6.2*, TextBlob menghasilkan analisis yang kurang akurat, terutama pada sentimen netral yang tidak berada persis di tengah dalam visualisasi scatterplot. Meskipun begitu, persebaran sentimen positif dan negatif cukup terwakili dengan baik. Hal ini dikarenakan TextBlob dirancang untuk analisis teks dalam bahasa Inggris, sehingga tidak sepenuhnya efektif untuk ulasan dalam bahasa Indonesia.



Gambar 6.2 Sentiment Analysis TextBlob 350 ribu Data

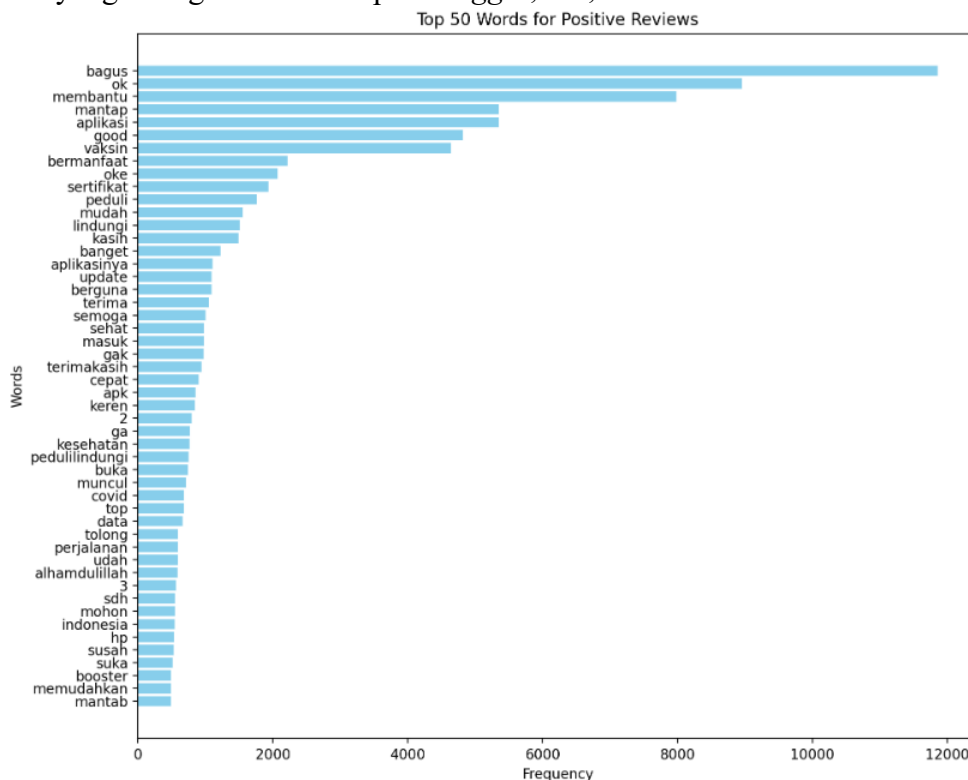
Selanjutnya, terdapat percobaan kedua, yaitu menarik sekitar 389 ribu ulasan. Pada percobaan ini, data dibagi menjadi 2, yaitu berdasarkan periode sebelum dan setelah aplikasi bertransformasi dari PeduliLindungi menjadi SatuSehat pada 1 Maret 2023. Dapat dilihat pada *Gambar 6.3* bahwa terdapat perubahan yang signifikan dimana ulasan dengan rating 1



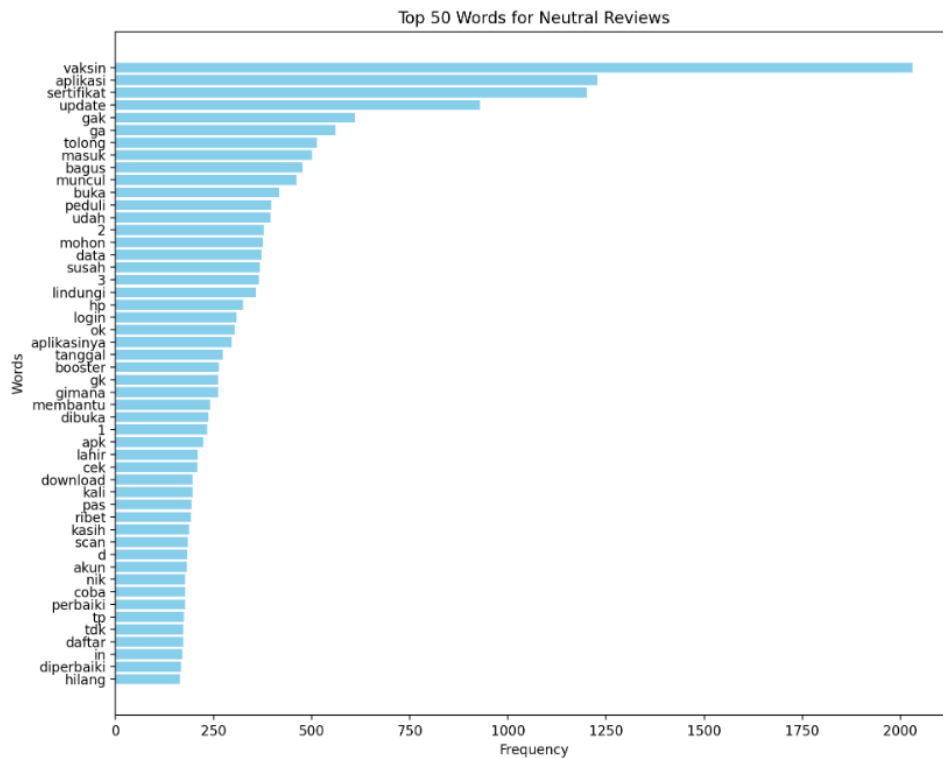


Gambar 6.5 Tren Kata 'Negative' terkait Aplikasi SatuSehat pada Google Play Store

Di sisi lain, ulasan positif tetap ada, meskipun jumlahnya lebih sedikit dibandingkan ulasan negatif. Kata-kata seperti *bagus*, *membantu*, dan *berguna* yang dapat dilihat pada Gambar 6.6 sering ditemukan dalam ulasan yang memberikan apresiasi terhadap aplikasi. Untuk ulasan netral seperti pada Gambar 6.7, analisis kurang akurat karena banyak ulasan dengan polaritas negatif yang diberi rating 3, sehingga polaritas sebenarnya tidak tercermin dengan baik. Kata-kata netral yang sering muncul meliputi *tanggal*, *cek*, dan *scan*.



Gambar 6.6 Tren Kata 'Positive' terkait Aplikasi SatuSehat pada Google Play Store

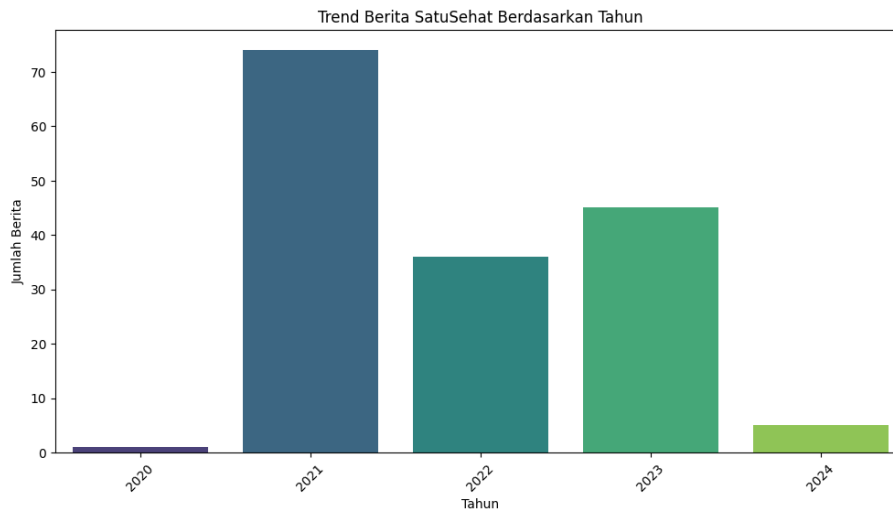


Gambar 6.7 Tren Kata 'Neutral' terkait Aplikasi SatuSehat pada Google Play Store

Kesimpulannya, meskipun aplikasi SatuSehat diharapkan menjadi pengembangan dari PeduliLindungi, banyak pengguna yang menganggapnya sebagai *downgrade*. Penurunan rating positif dan peningkatan ulasan negatif menunjukkan perlunya perbaikan pada fitur, aksesibilitas, dan pengalaman pengguna untuk memulihkan kepercayaan dan kepuasan pengguna.

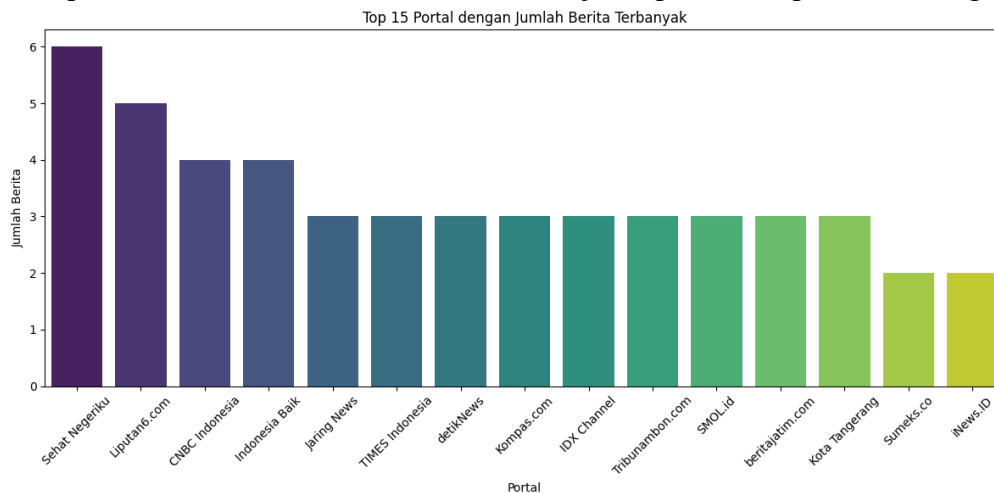
6.2 Analisis Sentimen Konten Artikel Berita

Aplikasi SatuSehat, yang sebelumnya dikenal sebagai PeduliLindungi, dirancang sebagai respons terhadap pandemi COVID-19. Berdasarkan analisis tren berita dari rentang tahun 2020 hingga 2024 yang dapat dilihat pada *Gambar 6.8*, puncak pembahasan aplikasi ini terjadi pada tahun 2021. Tahun ini merupakan masa kritis pandemi, dengan kebijakan pemerintah seperti kewajiban penggunaan PeduliLindungi untuk mengakses ruang publik. Pemberitaan terkait aplikasi ini menurun pada tahun-tahun berikutnya, dengan lonjakan signifikan kembali pada tahun 2023, yang ditandai dengan transformasi PeduliLindungi menjadi SatuSehat. Sebaliknya, pemberitaan pada tahun 2020, meski menjadi awal pandemi, sangat minim, dengan hanya satu artikel tercatat. Hal ini kemungkinan disebabkan oleh keterbatasan *library scraping* dalam menarik data berita lama.



Gambar 6.8 Tren Artikel Berita SatuSehat berdasarkan Tahun

Analisis portal berita yang dapat dilihat pada Gambar 6.9 menunjukkan bahwa pemberitaan SatuSehat melibatkan 120 portal, dengan dominasi oleh *Sehat Negeriku* yang membahas aplikasi ini sebanyak 6 kali, diikuti oleh *Liputan6* (5 kali), serta *CNBC Indonesia* dan *Indonesia Baik* masing-masing 4 kali. Sebagian besar portal berita lainnya hanya membahas aplikasi ini satu kali. Visualisasi lebih lanjut dapat dilihat pada file Google Colab.

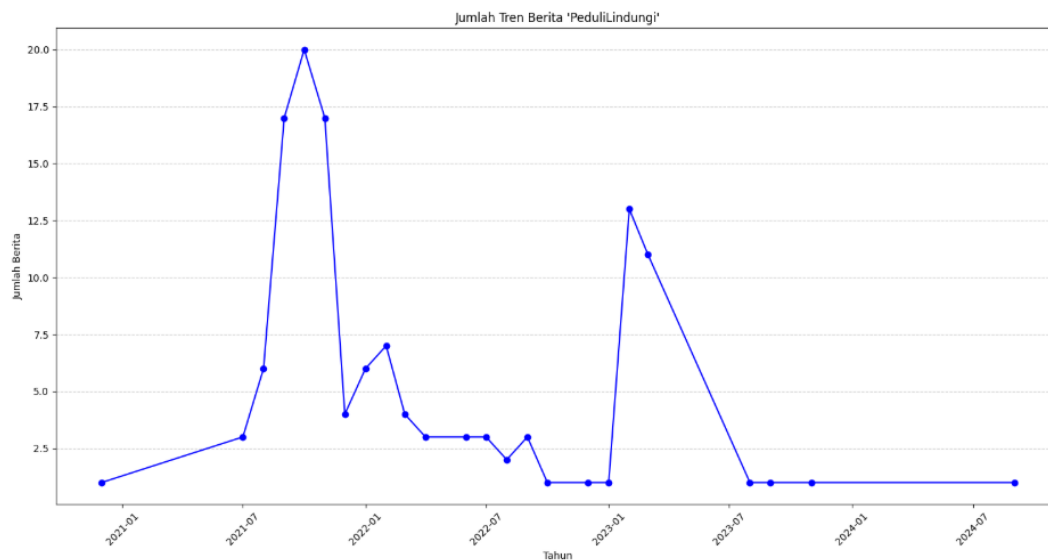


Gambar 6.9 Tren Portal Berita terkait SatuSehat

Dari segi terminologi, istilah "PeduliLindungi" ditemukan dalam berita sejak tahun 2020 hingga 2024 seperti pada, dengan puncaknya pada Juli hingga November 2021 yang dapat dilihat pada Gambar 6.11, yang merupakan periode intensif vaksinasi dan penggunaan aplikasi ini secara luas. Topik dominan meliputi kata kunci seperti *PeduliLindungi*, *aplikasi*, *masyarakat*, *COVID-19*, *vaksin*, *QR*, *vaksinasi*, *kesehatan*, dan *pemerintah* seperti pada Gambar 6.12.



Gambar 6.10 Tren Kata 'PeduliLindungi' dalam Artikel Berita



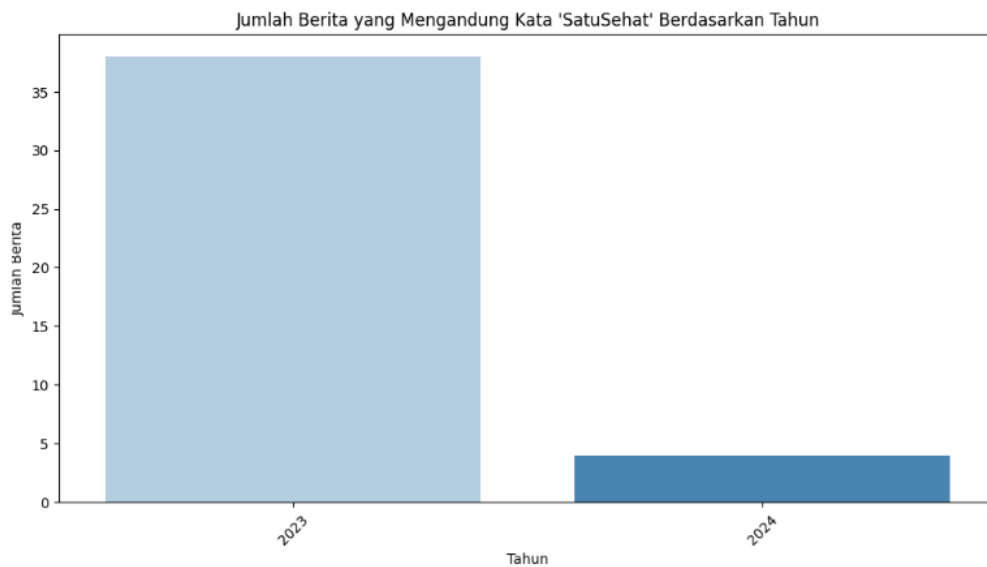
Gambar 6.11 Visualisasi Lanjutan Tren Kata 'PeduliLindungi' dalam Artikel Berita

	word	frequency
0	PeduliLindungi	420
1	aplikasi	393
2	masyarakat	225
3	covid19	164
4	vaksin	153
5	qr	139
6	vaksinasi	137
7	code	133
8	kesehatan	129
9	pemerintah	128

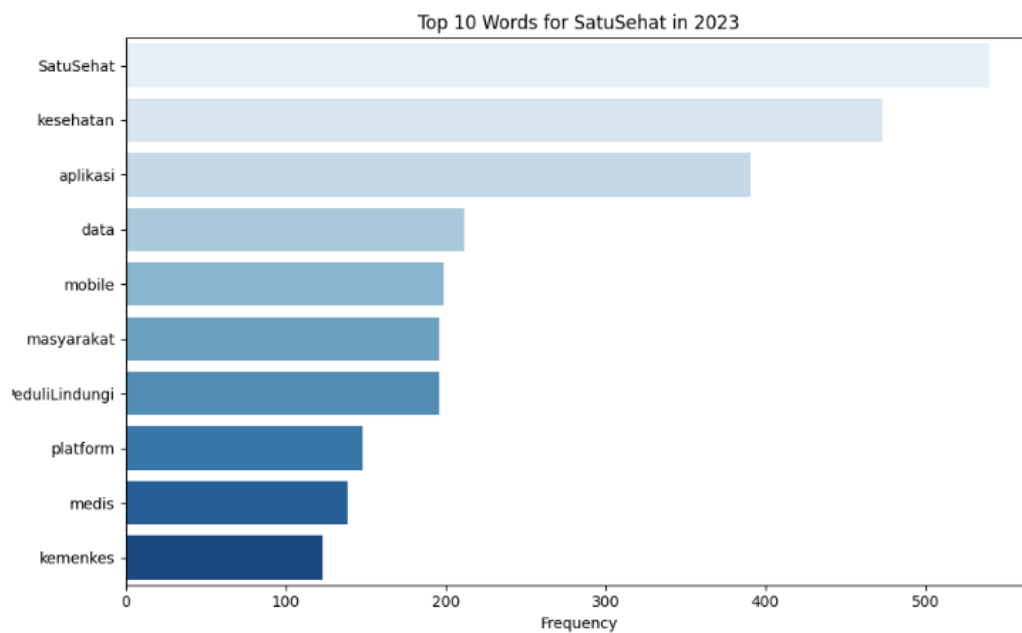
Gambar 6.12 Kata yang Paling Banyak Muncul terkait 'PeduliLindungi'

Sebaliknya, pemberitaan tentang SatuSehat baru mulai muncul pada tahun 2023 seperti pada Gambar 6.13, bersamaan dengan transformasi aplikasi dari PeduliLindungi. Sepuluh kata paling dominan pada tahun tersebut yang mengandung 'SatuSehat' adalah SatuSehat (540 kali), kesehatan (473 kali), aplikasi (391 kali), data (212 kali), mobile (199 kali), masyarakat (196 kali), PeduliLindungi (196 kali), platform (148 kali), medis (139 kali), dan Kemenkes (123 kali) yang dapat dilihat pada Gambar 6.14. Pada tahun 2024, fokus pemberitaan bergeser dengan

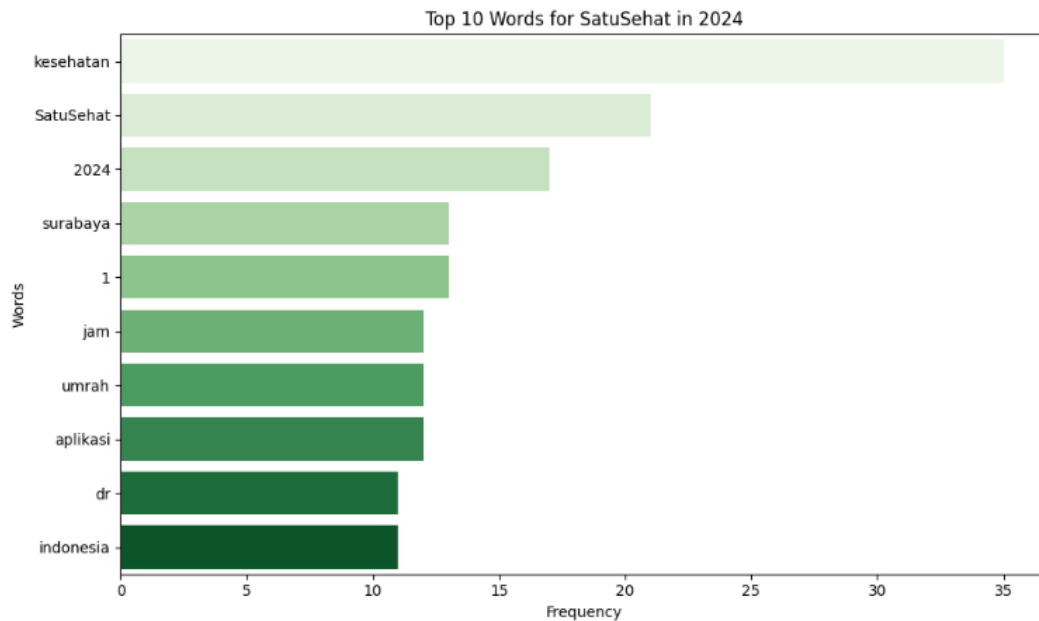
dominasi kata seperti kesehatan (35 kali), SatuSehat (21 kali), dan sejumlah istilah baru seperti umrah (12 kali) serta Surabaya (13 kali) yang dapat dilihat pada Gambar 6.15, mencerminkan penekanan pada integrasi aplikasi dengan layanan kesehatan yang lebih luas.



Gambar 6.13 Tren Kata 'SatuSehat' dalam Artikel Berita

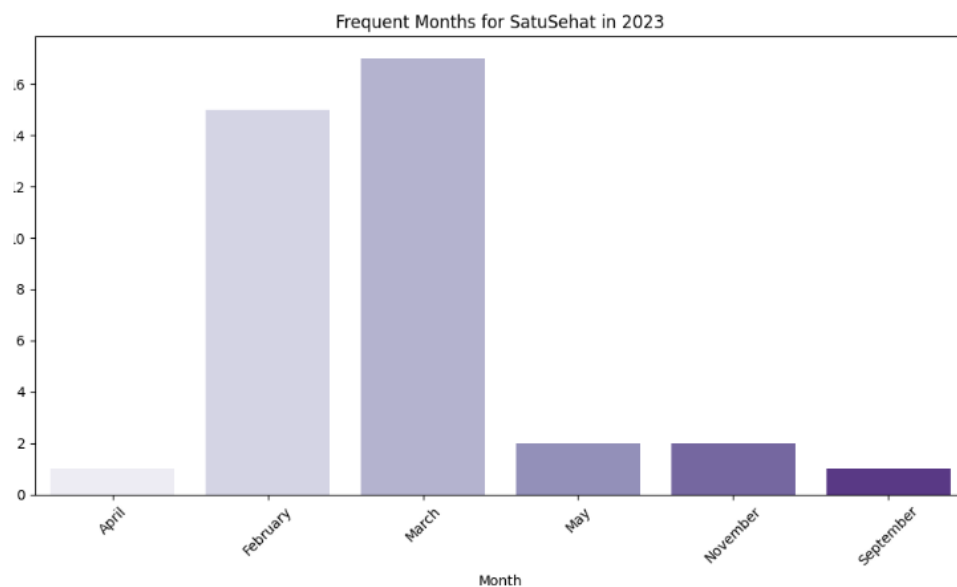


Gambar 6.14 Kata yang Paling Banyak Muncul terkait 'SatuSehat' pada Tahun 2023



Gambar 6.15 Kata yang Paling Banyak Muncul terkait 'SatuSehat' pada Tahun 2024

Frekuensi pemberitaan terkait SatuSehat pada tahun 2023 menunjukkan lonjakan signifikan pada bulan Februari (15 kali) dan Maret (17 kali) seperti pada Gambar 6.16. Periode ini diwarnai oleh kegiatan edukasi dan transformasi layanan menuju SatuSehat, sementara bulan-bulan lain lebih berfokus pada pengenalan aplikasi untuk layanan kesehatan yang lebih komprehensif. Sayangnya, transformasi ini juga diiringi oleh laporan kasus penipuan terkait SatuSehat, yang menjadi perhatian publik dan media massa. Salah satu contohnya seperti berita yang berjudul “hati-hati, penipuan berkedok link update aplikasi SatuSehat”.



Gambar 6.16 Bulan yang Paling Banyak Membahas 'SatuSehat' pada Tahun 2023

Berdasarkan analisis *word cloud* pada Gambar 6.17 yang dihasilkan dari artikel berita terkait SatuSehat setelah teks dibersihkan dari *stopwords*, terlihat bahwa pembahasan dominan berpusat pada kata-kata seperti aplikasi, PeduliLindungi, kesehatan, masyarakat, vaksin, data, COVID-19, dan sertifikat. Kata-kata ini mencerminkan fokus utama pemberitaan, yakni fungsi aplikasi sebagai sarana digital untuk mendukung penanganan pandemi, pelacakan kontak, serta

[illegible]

Kesimpulannya, meski aplikasi SatuSehat berhasil melanjutkan peran PeduliLindungi dalam mendukung layanan kesehatan digital di Indonesia, tren pemberitaan menunjukkan dinamika yang dipengaruhi oleh momentum pandemi, transformasi aplikasi, dan munculnya isu-isu baru. Dengan dominasi kata seperti kesehatan dan aplikasi, dapat disimpulkan bahwa media memusatkan perhatian pada dampak aplikasi ini terhadap masyarakat, baik dalam konteks pandemi maupun pengembangan layanan kesehatan digital yang lebih luas.

Part-of-Speech (POS) Tagging adalah proses menganalisis teks untuk mengidentifikasi kategori gramatikal atau peran kata dalam sebuah kalimat, seperti kata benda (*noun*), kata kerja (*verb*), kata sifat (*adjective*), dan lainnya. *POS tagging* membantu memahami struktur sintaksis kalimat dan maknanya secara lebih mendalam. Dalam konteks aplikasi SatuSehat, *POS tagging* bisa digunakan untuk menganalisis ulasan atau artikel yang membahas fitur-fitur aplikasi atau pengalaman pengguna, sehingga memberikan wawasan yang lebih dalam tentang bagaimana aplikasi tersebut dipersepsikan.

- "Aplikasi" sebagai *noun* (kata benda)
- "membantu" sebagai *verb* (kata kerja)
- "dengan cepat" sebagai *adverb phrase* (frasa kata keterangan)
- "rekam medis" sebagai *noun phrase* (frasa kata benda).

6.1 POS dan NER Ulasan Pengguna pada Google Play Store

a) POS Ulasan Pengguna terkait SatuSehat pada Google Play Store

Analisis ulasan pengguna terkait aplikasi SatuSehat mengungkap beberapa pola linguistik menarik berdasarkan analisis *Part-of-Speech* (POS) dan frekuensi kata. Dari keseluruhan teks dihasilkan Gambar 7.1, kata benda (NN) mendominasi dengan jumlah mencapai 599.437, yang mencerminkan fokus ulasan pada aspek konkret seperti "aplikasi" dan "sertifikat." Kata sifat (JJ) sebanyak 86.266 menunjukkan bahwa pengguna sering menggambarkan pengalaman mereka, baik positif seperti "bagus" maupun negatif seperti "gak." Selain itu, kata kerja (VBP, VBD, VBZ) sebanyak 26.697 menyoroti berbagai tindakan yang dilaporkan oleh pengguna, seperti "masuk" dan "update".

Tag	Count	Unique Tokens
NN	599437	43901
JJ	86266	8283
DT	2844	37
CD	17608	1759
WP	3715	74
VBZ	6323	1543
IN	6360	633
FW	7583	1115
VBD	9463	1921
NNS	17842	2228
VBP	10911	2644
RB	4221	740
VBG	2653	426
RP	571	25
VB	2130	708
PRP	432	19
VCN	561	126
MD	356	77
UH	37	2
CC	1240	250
RBS	50	23
JJS	344	11
WRB	118	28
NNP	1310	467
PDT	10	5
TO	211	2
PRP\$	109	7
JJR	94	44
WDT	11	5
RBR	27	10
EX	2	1

Gambar 7.1 POS Tagging Ulasan Pengguna terkait SatuSehat pada Google Play Store

Frekuensi kata pada Gambar 7.2 juga mengungkapkan bahwa pengguna sering menyebutkan istilah seperti "aplikasi," "vaksin," dan "sertifikat," menandakan fitur inti yang menjadi perhatian utama. Kata "bagus" dan "membantu" mencerminkan pengalaman positif, sedangkan kata seperti "gak," "ga," dan "masuk" menunjukkan keluhan atau tantangan, terutama terkait aksesibilitas dan performa.

Top 10 most common words: [('aplikasi', 26768), ('vaksin', 22642), ('bagus', 14168), ('sertifikat', 13683),
Top 10 most common POS tags: [('NN', 599437), ('JJ', 86266), ('NNS', 17842), ('CD', 17608), ('VBP', 10911),

Gambar 7.2 Tren Kata dan POS tags terkait SatuSehat pada Google Play Store

Dengan pola ini, dapat disimpulkan bahwa ulasan mencerminkan evaluasi langsung terhadap fungsi dan kualitas aplikasi, di mana sebagian besar diskusi berpusat pada masalah teknis, sertifikat vaksin, serta kebutuhan untuk meningkatkan pengalaman pengguna secara keseluruhan.

b) NER

Named Entity Recognition (NER) pada ulasan pengguna aplikasi SatuSehat di Google Play Store seperti pada Gambar 7.4 menunjukkan bahwa entitas yang paling sering terdeteksi adalah PERSON, diikuti oleh ORG (organisasi) dan GPE (geopolitical entity). Entitas PERSON merujuk pada nama individu atau penyebutan yang berhubungan dengan orang, meskipun beberapa entitas seperti "gausah jelek" atau "akun ga" yang seharusnya tidak terdeteksi sebagai PERSON seperti pada Gambar 7.3. Selain itu, entitas ORG lebih sering muncul, seperti pada "pedulilindungi daftar" yang merujuk pada aplikasi sebelumnya. Namun, banyak entitas yang salah deteksi atau tidak relevan, yang menunjukkan adanya keterbatasan dalam kemampuan NER untuk mengenali konteks yang tepat, terutama dengan bahasa sehari-hari atau penggunaan istilah yang tidak baku. Entitas GPE yang mengacu pada lokasi juga ditemukan, meskipun tidak sebanyak PERSON dan ORG. Analisis ini menunjukkan pentingnya perbaikan dalam model NER untuk meningkatkan akurasi dan relevansi deteksi entitas dalam ulasan pengguna aplikasi SatuSehat.

```

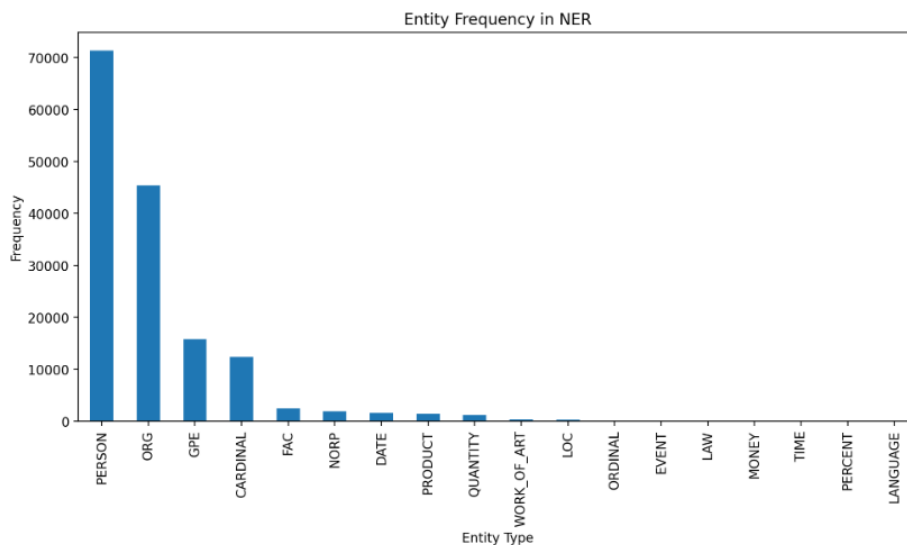
review_without_stopwords \
0      kalo gk niat gausah jelek
1      gak no daftar ane
2  pedulilindungi daftar akun login web masuk amb...
3      membantu
4      akun ga log in

ner_entities
0      [(gausah jelek, PERSON)]
1      [(gak no daftar, ORG)]
2      [(pedulilindungi daftar, PERSON), (1, CARDINAL...
3      []
4      [(akun ga, ORG)]

kalo gk niat  gausah jelek  PERSON

```

Gambar 7.3 Contoh NER Ulasan Pengguna terkait SatuSehat pada Google Play Store



Gambar 7.4 Jumlah Entity pada NER Ulasan Pengguna terkait SatuSehat pada Google Play Store

6.2 POS dan NER Konten Artikel Berita

a) POS Artikel Berita SatuSehat

	Tag	Count	Unique Tokens
0	NN	30544	5091
1	NNP	2719	121
2	RB	158	78
3	JJ	6279	1670
4	FW	382	125
5	VBZ	835	251
6	CD	2318	466
7	VBD	1092	422
8	IN	330	97
9	NNS	1924	280
10	VBP	1587	583
11	WRB	7	5
12	VB	229	78
13	PRP\$	31	4
14	MD	12	5
15	VRN	66	22
16	CC	40	11
17	DT	100	9
18	PRP	16	4
19	TO	28	1
20	VBG	240	64
21	WP	82	6
22	JJS	9	2
23	PDT	2	1
24	RBS	4	4
25	JJR	9	5
26	RP	7	3
27	RBR	7	1

Gambar 7.5 POS Tagging Artikel Berita SatuSehat

Berdasarkan hasil analisis *part-of-speech* (POS) pada dataset berita SatuSehat, terdapat berbagai jenis tag POS yang mengindikasikan struktur gramatikal dan makna kata dalam teks seperti pada Gambar 7.5. Sebagai contoh, tag *NN* (kata benda tunggal) muncul paling banyak dengan 30.544 kemunculan dan 5.091 token unik, yang menunjukkan dominasi kata benda dalam berita ini. Kemudian, tag *NNP* (kata benda terhitung yang merupakan nama diri) muncul sebanyak 2.719 kali dengan 121 token unik, mengindikasikan adanya banyak referensi kepada entitas tertentu seperti nama orang atau organisasi. Tag *RB* (kata keterangan) muncul 158 kali, dengan 78 token unik, mengindikasikan penggunaan kata keterangan dalam kalimat untuk memperjelas makna atau konteks aksi yang dijelaskan. Tag *JJ* (kata sifat) muncul sebanyak 6.279 kali, dengan 1.670 token unik, yang menunjukkan adanya banyak deskripsi atau kualifikasi terhadap kata benda dalam berita. Serta tag lainnya.

Beberapa tag dengan frekuensi lebih rendah seperti *WRB* (kata tanya), *PRP\$* (kata ganti milik), dan *MD* (kata kerja modal) muncul dalam jumlah terbatas, yang mungkin mencerminkan elemen sintaksis yang lebih jarang digunakan namun tetap penting untuk memahami konteks atau makna dari kalimat yang lebih kompleks. Selain itu, pada tahap ini setiap kata yang terdapat dalam “content” juga dihitung frekuensi kemunculannya seperti pada Gambar 7.6.

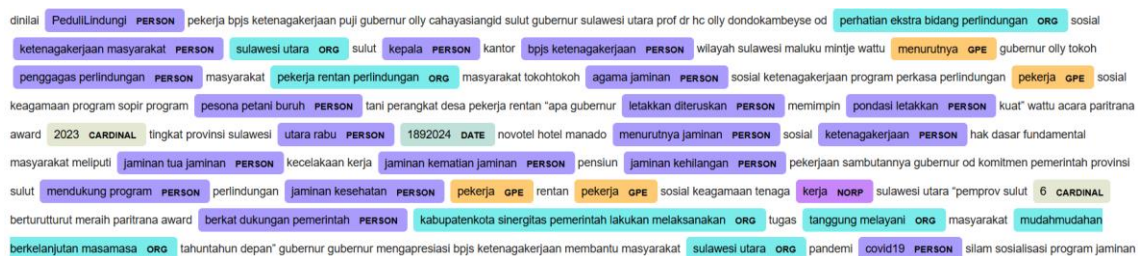
Word	POS Tag	Frequency
aplikasi	NN	1120
PeduliLindungi	None	967
kesehatan	NN	863
masyarakat	NN	604
SatuSehat	None	566
vaksin	NN	478
data	NNS	462
covid19	NN	389
sertifikat	NN	341
vaksinasi	NN	295

Gambar 7.6 Frekuensi Kata beserta POS Tag pada Artikel Berita SatuSehat

Secara keseluruhan, analisis POS ini memberikan gambaran tentang struktur gramatikal yang dominan dalam berita SatuSehat, yang banyak menggunakan kata benda, kata kerja, serta kata sifat dan keterangan untuk menyampaikan informasi secara deskriptif dan kuantitatif.

b) NER Artikel Berita SatuSehat

Hasil *Named Entity Recognition* (NER) pada artikel berita terkait SatuSehat seperti Gambar 7.7 menunjukkan bahwa model NER yang digunakan tidak sepenuhnya akurat dalam mengenali entitas dalam teks. Sebagai contoh, beberapa entitas seperti *PeduliLindungi* dan *SatuSehat* yang semestinya dikategorikan sebagai ORGANIZATION (ORG) malah diabaikan atau tidak dikenali dengan benar. Selain itu, entitas *gubernur Olly Dondokambey* dan *kepala kantor BPJS Ketenagakerjaan Wilayah Sulawesi Maluku Mintje Wattu* yang seharusnya dikategorikan sebagai PERSON terdeteksi secara parsial atau keliru. Dalam beberapa kasus, nama-nama seperti *sulawesi utara* dan *sulut* yang merupakan nama wilayah (LOC) malah terdeteksi sebagai ORGANIZATION (ORG).



Gambar 7.7 Contoh NER Artikel Berita SatuSehat

Ketidakakuratan ini disebabkan oleh beberapa faktor. Model NER yang digunakan mungkin tidak dirancang khusus untuk bahasa Indonesia, sehingga pemrosesan bahasa alami untuk entitas non-Inggris menjadi kurang optimal. Selain itu, konteks dalam berita sering kali kompleks, mencampurkan nama, jabatan, institusi, dan wilayah, sehingga model kesulitan memisahkan atau mengkategorikan entitas dengan benar. Misalnya, *prof dr hc olly dondokambeyse od* terdeteksi sebagai PERSON, tetapi jabatan atau gelarnya tidak dikenali sebagai bagian terpisah.

8. KESIMPULAN

Pengambilan data dilakukan dengan mengumpulkan ulasan pengguna dari Google Play Store terkait aplikasi SatuSehat serta artikel berita yang membahas aplikasi tersebut. Proses *preprocessing* melibatkan penghapusan tanda baca, *stopwords*, dan normalisasi teks untuk meningkatkan kualitas analisis. Pada analisis sentimen ulasan pengguna, data diolah menggunakan TextBlob untuk menilai polaritas dan subjektivitas, meskipun keterbatasan TextBlob dalam memahami bahasa Indonesia mengakibatkan hasil yang kurang akurat, terutama untuk ulasan netral. Secara umum, ulasan didominasi oleh sentimen negatif, terutama setelah transformasi dari PeduliLindungi menjadi SatuSehat, dengan banyak keluhan terkait akses sertifikat vaksin, fitur yang dihilangkan, dan kesulitan login. Sedangkan analisis sentimen artikel berita menunjukkan bahwa SatuSehat, yang sebelumnya dikenal sebagai PeduliLindungi, menjadi pusat perhatian media selama pandemi COVID-19 dan transformasinya pada 2023. Kata-kata dominan seperti kesehatan, aplikasi, vaksin, dan data mencerminkan peran aplikasi dalam mendukung vaksinasi, verifikasi sertifikat kesehatan, dan pengelolaan data medis. Transformasi ini juga diwarnai oleh edukasi publik dan isu penipuan

yang menarik perhatian media. Secara keseluruhan, SatuSehat berperan penting dalam pengembangan layanan kesehatan digital di Indonesia.

DAFTAR PUSTAKA

- Kemenkes. (2021). Masyarakat Perlu Tahu Manfaat Aplikasi PeduliLindungi. In *Ayosehat.Kemkes.Go.Id*. <https://ayosehat.kemkes.go.id/masyarakat-perlu-tahu-manfaat-aplikasi-pedulilindungi>
- Kemenkes. (2023). PeduliLindungi Resmi Berubah Menjadi SatuSehat. In 2 Maret 2023. <https://promkes.kemkes.go.id/pedulilindungi-resmi-berubah-menjadi-satusehat>
- Lowery, C. (2020). What Is Digital Health and What Do I Need to Know About It? In *Obstetrics and Gynecology Clinics of North America* (Vol. 47, Issue 2, pp. 215–225). <https://doi.org/10.1016/j.ogc.2020.02.011>
- Marpaung, Y. N. M., & Irwansyah, I. (2021). Aplikasi Kesehatan Digital Sebagai Konstruksi Sosial Teknologi Media Baru. *Jurnal Komunikasi Dan Kajian*, 5, 243–258.
- Septiani, D., Ruhama, S., & Astuti, I. (2023). Implementasi Metode Pieces Untuk Menganalisis Tingkat Kepuasan Pengguna Aplikasi Peduli Lindungi. *JIKI (Jurnal Ilmu Komputer Dan Informatika)*, 4(1), 53–64.