

Analyzing Public Sentiment on CORETAX Implementation: Insights from X Data

Kelompok 5

Timothy Hartanto

Yilmaz Trigumari Syah Putra

Nabila Avicenna

Muhammad Iqbal

Daftar Isi

- 01** Gambaran Umum
- 02** Pemahaman Data
- 03** Sentiment Polarity Analysis
- 04** Modelling
- 05** Kesimpulan & Rekomendasi

Gambaran Umum

Latar Belakang Instansi
Kondisi Eksisting
Pernyataan Masalah



Latar Belakang Instansi

Direktorat Jenderal Pajak (DJP)

DJP adalah lembaga pemerintah di bawah Kementerian Keuangan Indonesia yang bertanggung jawab atas pengelolaan administrasi perpajakan di Indonesia. DJP memiliki tugas untuk **memastikan pengumpulan pajak berjalan dengan efisien, akurat, dan sesuai dengan peraturan yang berlaku**. Dalam upaya meningkatkan layanan perpajakan, DJP meluncurkan sistem berbasis teknologi seperti **CORETAX** untuk memperbaiki proses administrasi, memodernisasi sistem perpajakan, dan memperkuat basis data pajak nasional guna mendukung kepatuhan wajib pajak serta meningkatkan pendapatan negara.



Kondisi Eksisting

Pada awal Januari, Direktorat Jenderal Pajak (DJP) Kementerian Keuangan Indonesia meluncurkan Sistem Administrasi Perpajakan Core (CORETAX), yang bertujuan untuk memperlancar proses administrasi dan meningkatkan basis data perpajakan.

Meskipun diharapkan meningkatkan efisiensi, akurasi, dan kepatuhan, CORETAX menghadapi berbagai masalah teknis, seperti kesalahan sistem dan akses yang lambat, yang memicu kritik publik. Media sosial, terutama Twitter, menjadi saluran bagi masyarakat untuk menyampaikan pendapat dan keluhan tentang CORETAX. **Memahami sentimen publik terhadap CORETAX penting untuk mengevaluasi penerimanya dan memberikan wawasan bagi perbaikan sistem ke depan.**



Pernyataan Masalah

Pernyataan Masalah

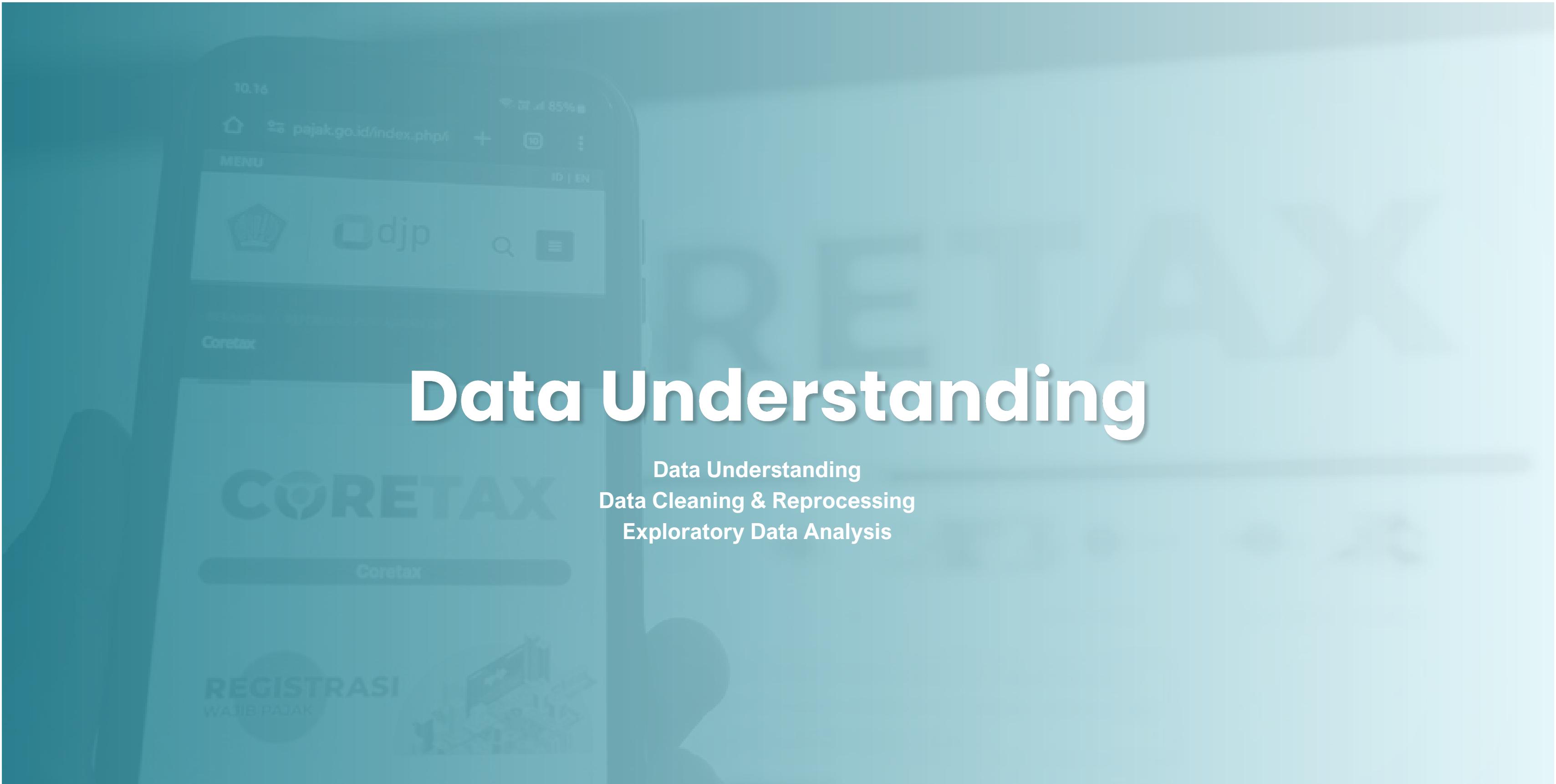
Bagaimana tanggapan publik terhadap implementasi CORETAX, dan Bagaimana kita dapat meningkatkan CORETAX berdasarkan respons publik dan kekhawatiran utama yang disampaikan di Twitter?

Pendekatan

- 1 Scrapping Data (Post Twitter)
- 2 Preprocessing Data
- 3 Memberi Label Sentimen
- 4 Interpretasi Hasil dan Analisis
- 5 Klasifikasi Sentimen

Data Understanding

Data Understanding
Data Cleaning & Reprocessing
Exploratory Data Analysis



Twitter Post Scraping

Menyiapkan Akun Twitter untuk Authentication

Kami membuat akun Twitter untuk mendapatkan token autentikasi guna mendapatkan akses ke API Twitter, yang memungkinkan kami mengambil tweet relevan tentang CORETAX.



Scraping Tweets Menggunakan TweetHarvest

Kami menggunakan `TweetHarvest`, alat pengikis berbasis Python, untuk mengekstrak tweet yang berisi kata kunci terkait dengan CORETAX.



Unduh Data dari Google Drive

Setelah tweet dikumpulkan, kami menyimpannya dalam file .csv dan memuatnya dari Google Drive untuk analisis lebih lanjut.

Data Understanding

	id_str	username	created_at	full_text	retweet_count	favorite_count
0	1885111788167172159	edwardyurianto	Thu Jan 30 23:44:08 +0000 2025	@inilahdotcom Mana ada aplikasi / web buatan n...	0	0
1	1885110547450126784	ValeValvuela	Thu Jan 30 23:39:12 +0000 2025	@Twilight_Loid Mau dibayar pake apa? Wong pen...	0	0
2	1885109876315349032	deni_hermawanto	Thu Jan 30 23:36:32 +0000 2025	@kring_pajak untuk pengisian eBUPOT pasal21 dg...	0	0
3	1885106389334581310	rzmazz	Thu Jan 30 23:22:41 +0000 2025	@CryptoAlifia @karifess Coretax itu apa kak k...	0	0
4	1885096002715263485	hajia_dirh30477	Thu Jan 30 22:41:25 +0000 2025	@DPR_RI harus panggil @kemenkeu @KemenkeuRI un...	0	3
...

Data merupakan data dari 1 Januari 2025 - 16 Februari 2026

Deskripsi

id_str	ID unik untuk setiap Tweet.	full_text	Isi tweet, digunakan untuk analisis sentimen.
username	Pegangan Twitter dari <i>user</i> yang memposting tweet.	retweet_count	Berapa kali tweet telah dibagikan.
created_at	Tanggal dan waktu ketika tweet di-posting.	favorite_count	Jumlah <i>likes</i> yang diterima tweet tersebut.

Data Cleaning & Reprocessing

Gambaran Umum

Pemahaman Data

Sentiment Polarity Analysis

Modelling

Kesimpulan dan Rekomendasi

Cleaning Text

Menghapus penyebutan, tagar, URL, angka, dan tanda baca.

Case Folding Text

Menjadikan semua teks menjadi huruf kecil

Tokenizing Text

Memisahkan teks menjadi kata-kata

Filtering Text

Menghapus kata-kata umum dalam bahasa Indonesia

Data Cleaning & Reprocessing

Core
Tax

Administration
System

Gambaran
Umum

Pemahaman
Data

Sentiment Polarity
Analysis

Modelling

Kesimpulan dan
Rekomendasi

Stemming Text

Mengubah kata menjadi bentuk
dasarnya

to Sentence

Mengubah kata-kata yang telah
ditokenisasi kembali menjadi suatu
kalimat.

Remove Slang Word

Mengubah kata-kata slang menjadi
kata baku sesuai KBBI

Data Cleaning

Result

Original Text

Misbakhun Soal Coretax Sering
Bermasalah: Hapus Sanksi Ditjen
Pajak Tanggung Jawab!
#bahlillahadalia #kuningkeren
#golkarindonesia #partaigolkar
#golkarpedia
<https://t.co/3aGNzBgujx>

Text Clean

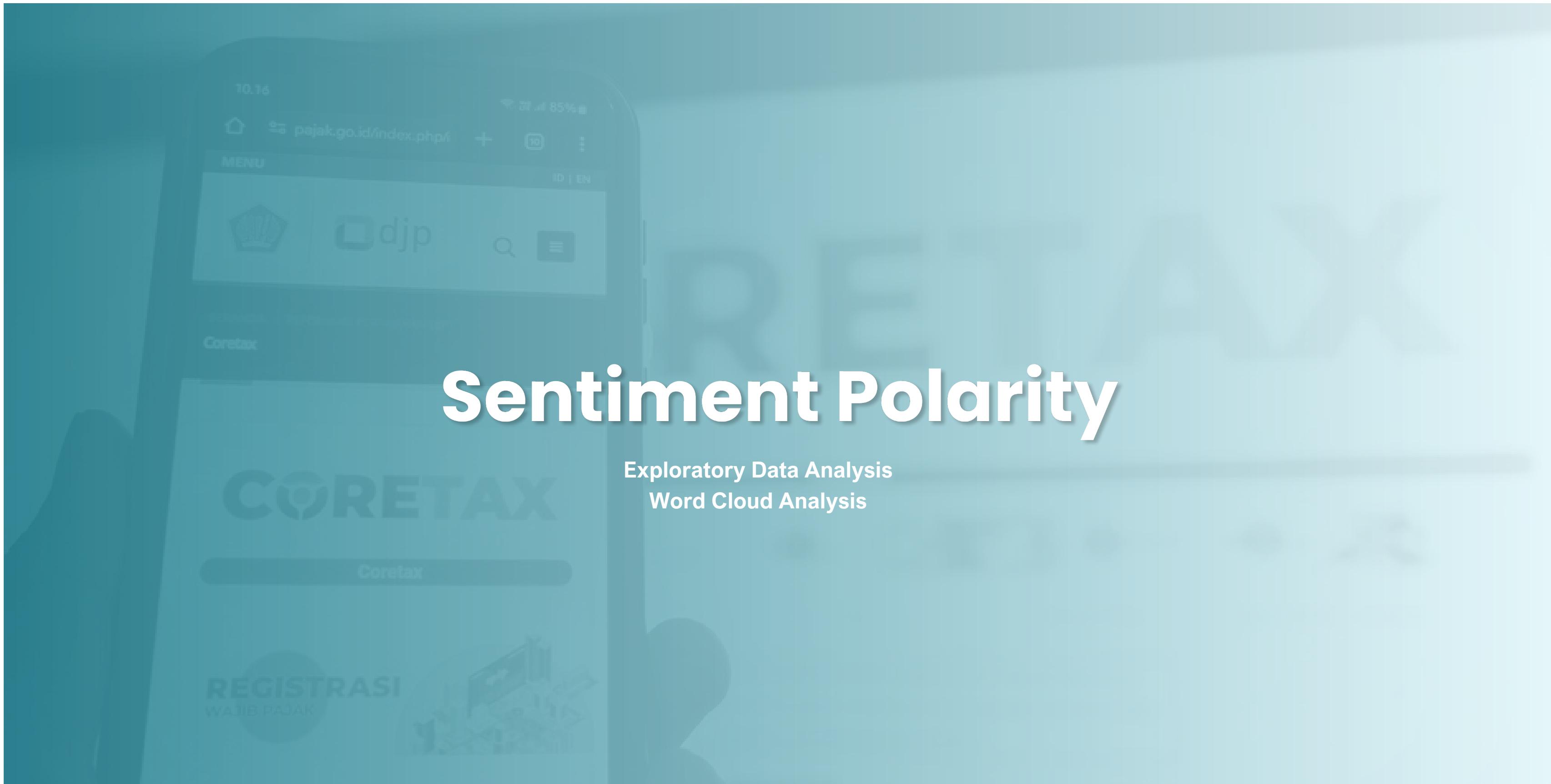
misbakhun soal coretax
sering bermasalah
hapus sanksi ditjen
pajak tanggung jawab

Text Preprocessed

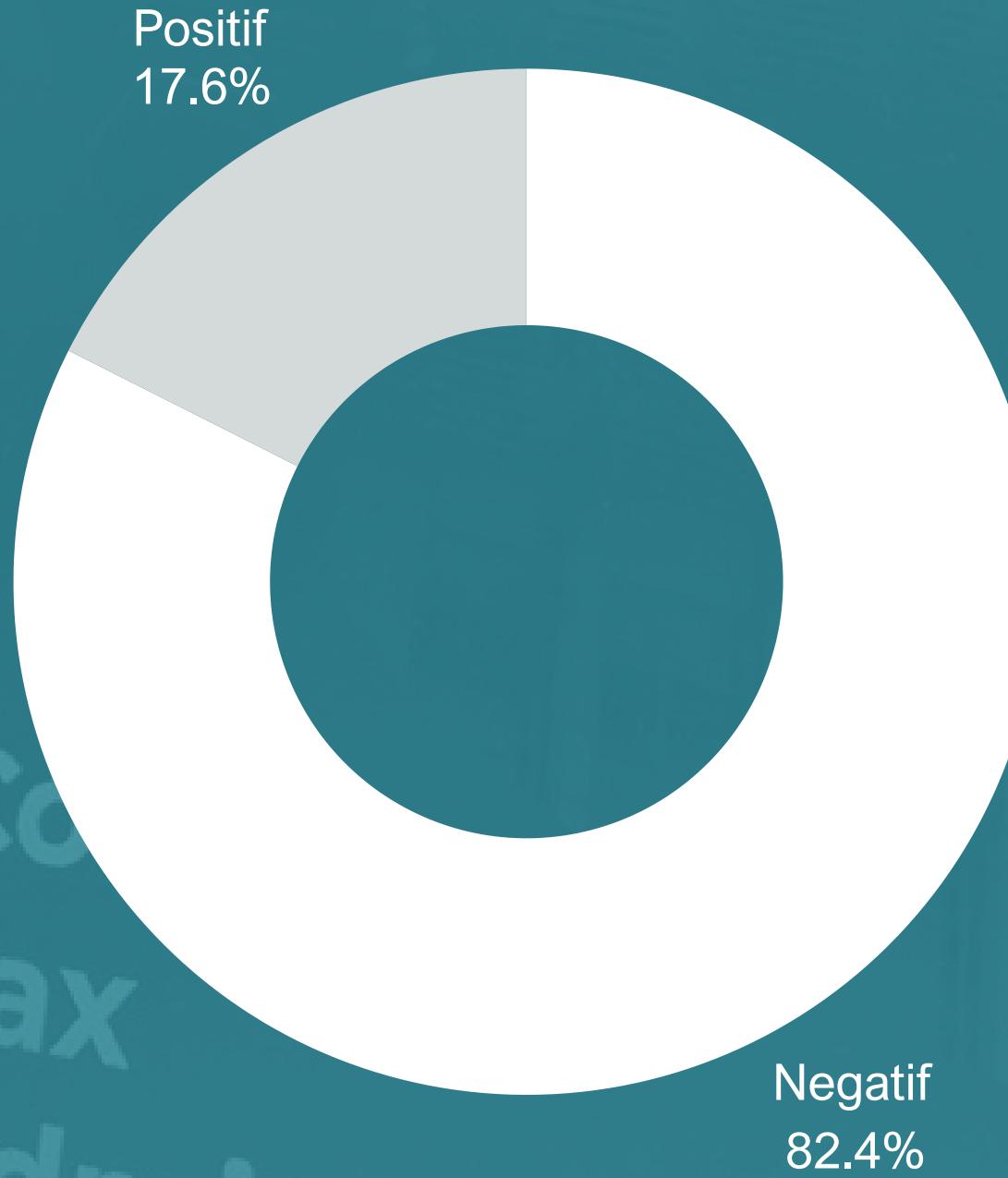
misbakhun coretax
masalah hapus sanksi
ditjen pajak tanggung

Sentiment Polarity

Exploratory Data Analysis
Word Cloud Analysis



Sentiment Polarity on Tweets data



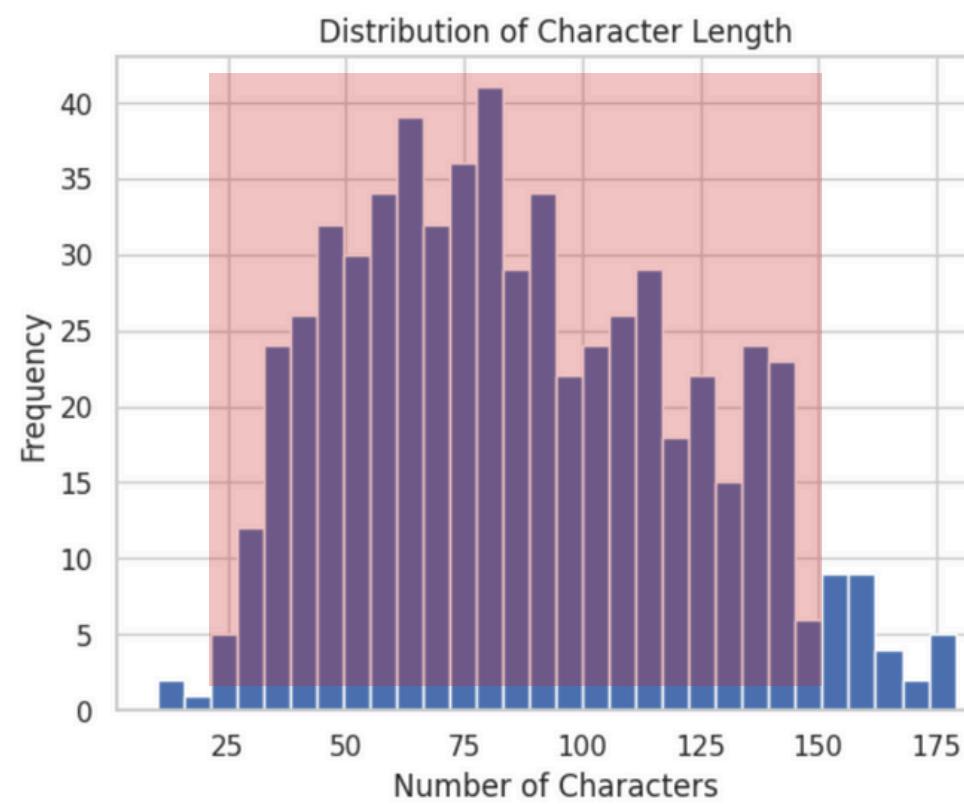
Sentiment Polarity

Dengan mencocokkan kata dalam tweet dengan leksikon positif dan negatif, diperoleh hasil bahwa **82.4% tweet bernada negatif**, menunjukkan ketidakpuasan tinggi terhadap CORETAX, sementara hanya **17.6% yang positif**.

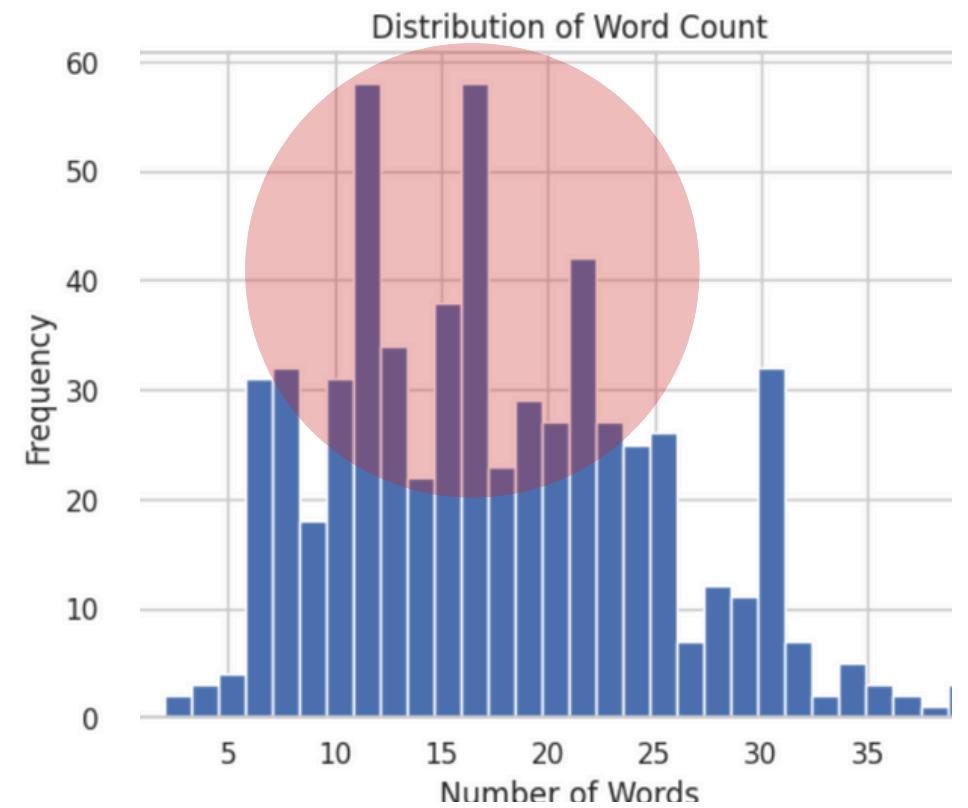
Temuan ini menyoroti kekhawatiran publik dan perlunya investigasi lebih lanjut, didukung dengan contoh tweet positif dan negatif.

Sentiment Polarity EDA

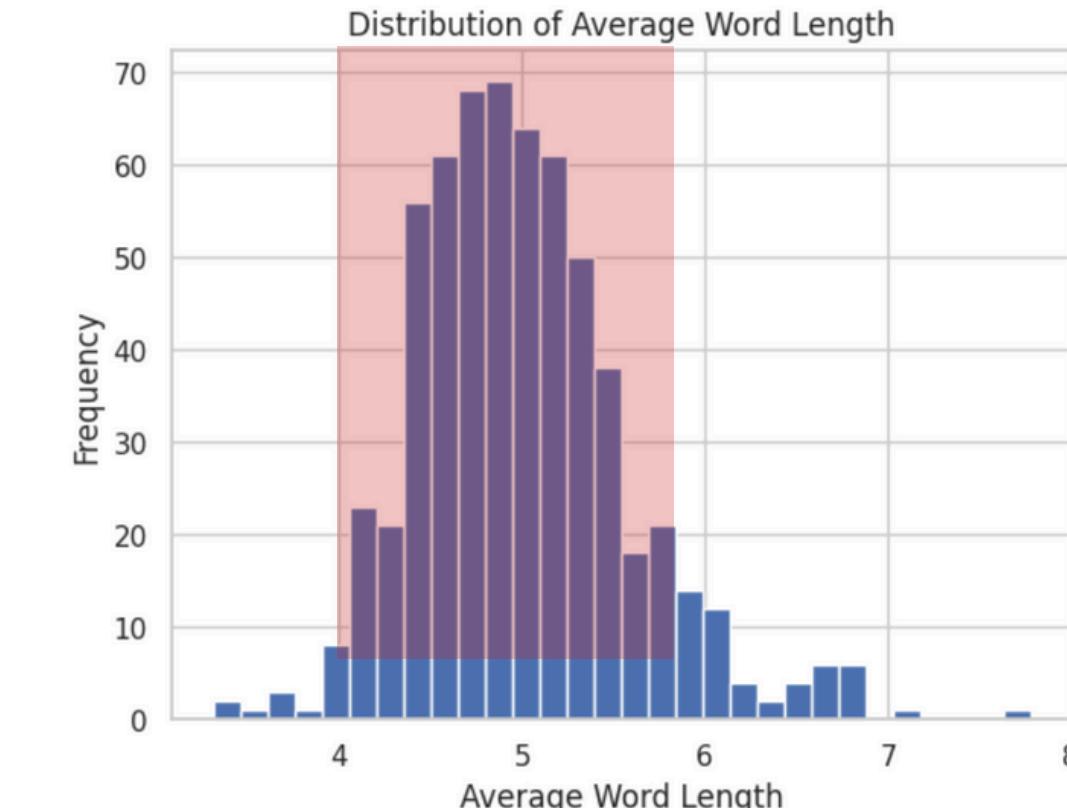
Distribusi Hasil Sentiment Polarity



Sebagian besar teks memiliki panjang sekitar 75-100 karakter.



Mayoritas teks terdiri dari 10 hingga 20 kata



Rata-rata panjang kata adalah sekitar 5 karakter per kata.

Sentiment Polarity

Contoh Post Positif

sehari bisa input faktur pajak invoice pas jam kerja aja udah beruntung biasa juga paling banyak setiap hari harus lembur karena coretax lebih lancar diakses setelah jam kerja apalagi di atas jam malam

pajak iyaa kak jadi kalau kita buat faktur di efaktur tidak perlu pakai hitungan dpp nilai lain jadi langsung dppx di coretax ppn kan ada perhitung dpp nilai lainnya tapi sebetulnya secara nilai kan ppn nya tetap sama



hai open jasa lapor spt tahunan pegawai pt cv badan perhitungan pajak pembuatan npwp tanggapan spdk konsultasi perpajakan pembuatan laporan keuangan pembuatan lapkeu spt lapor spt masa baca bio sebelum order ya kaa

ternyata ini sungguh visioner aplikasinya lancar klo dini hari sengaja biar orang latian bangun sahur dah gitu ngajarin memperbanyak istighfar astaghfirullaherror meneh iki ya allah makasih lo

Sentiment Polarity

Contoh Post Negatif

hadapi rezim dalam sistem ppn di indonesia ikuti webinar ini perubahan skema ppn dan dampaknya penghitungan ppn terutang administrasi pajak faktur pajak nota pembatalan nota retur coretax djp pmk simak selengkapnya

konyol bgt si █████ owo kalau janji target pajak tapi masalah coretax █████ yani ga bisa beresin dalam satu semester ini mah pecat aja menjabat cuma tahun kalau satu semester aja pendapatan pajak jeblok efeknya bisa berlarut larut apalagi sambil nombok utang j██owi

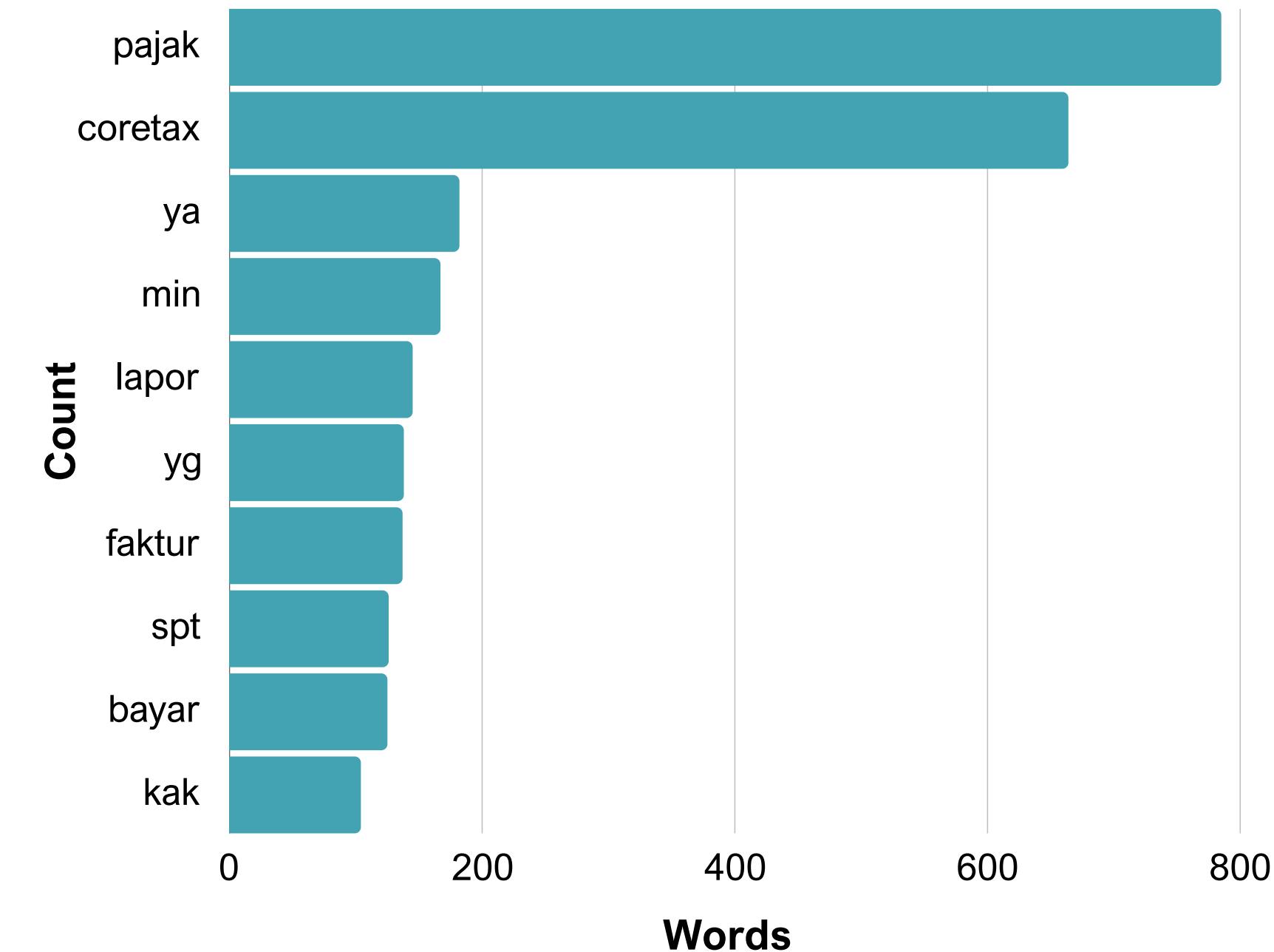
minimal sebagai drafter coretax bisa download bupot ppn unifikasi masa kudu minta atasan perkara download doank hadeh mana lemot parah pajak

fzn rakjel biasa yg mau taat pajak aja dipersulit gini coretax error terus web djp lama ga ada opsi taun giliran telat bayar pajak diuber dikasi denda macem kayak ngegarong duit negara



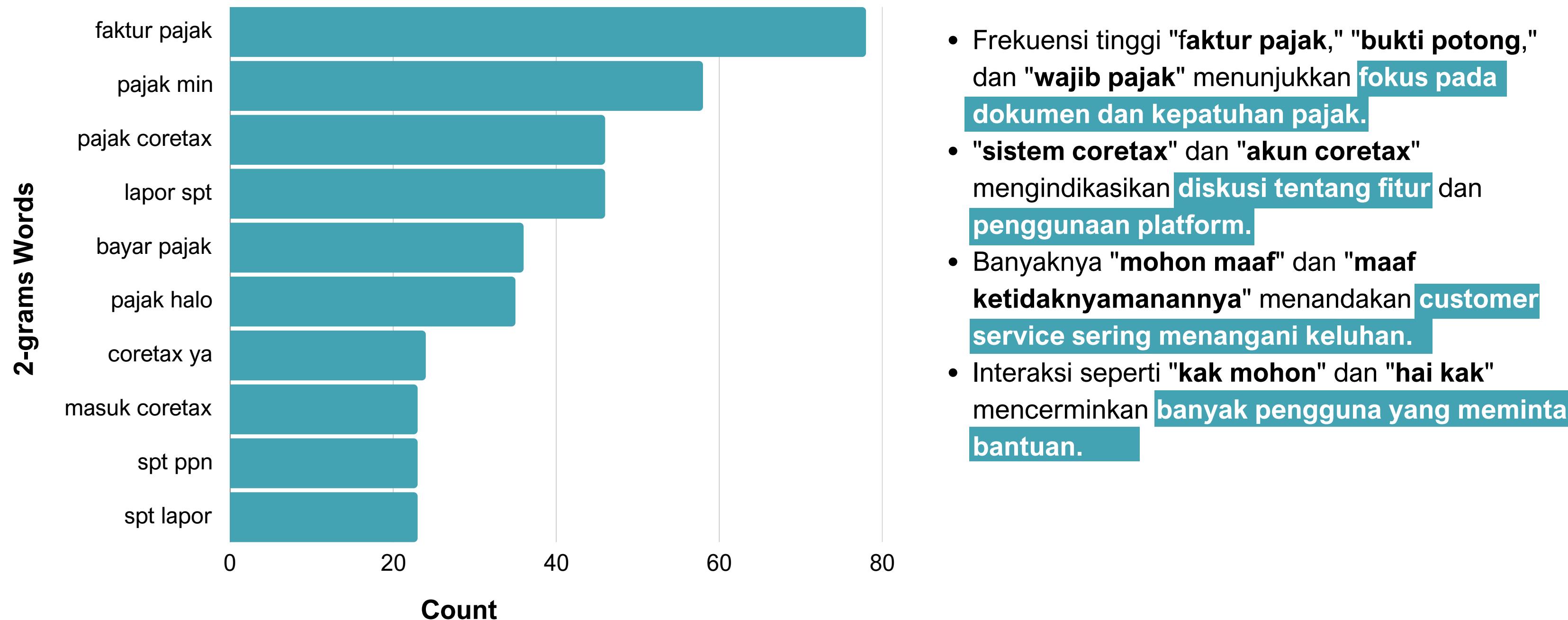
Sentiment Polarity EDA

Top 10 Most Common Words



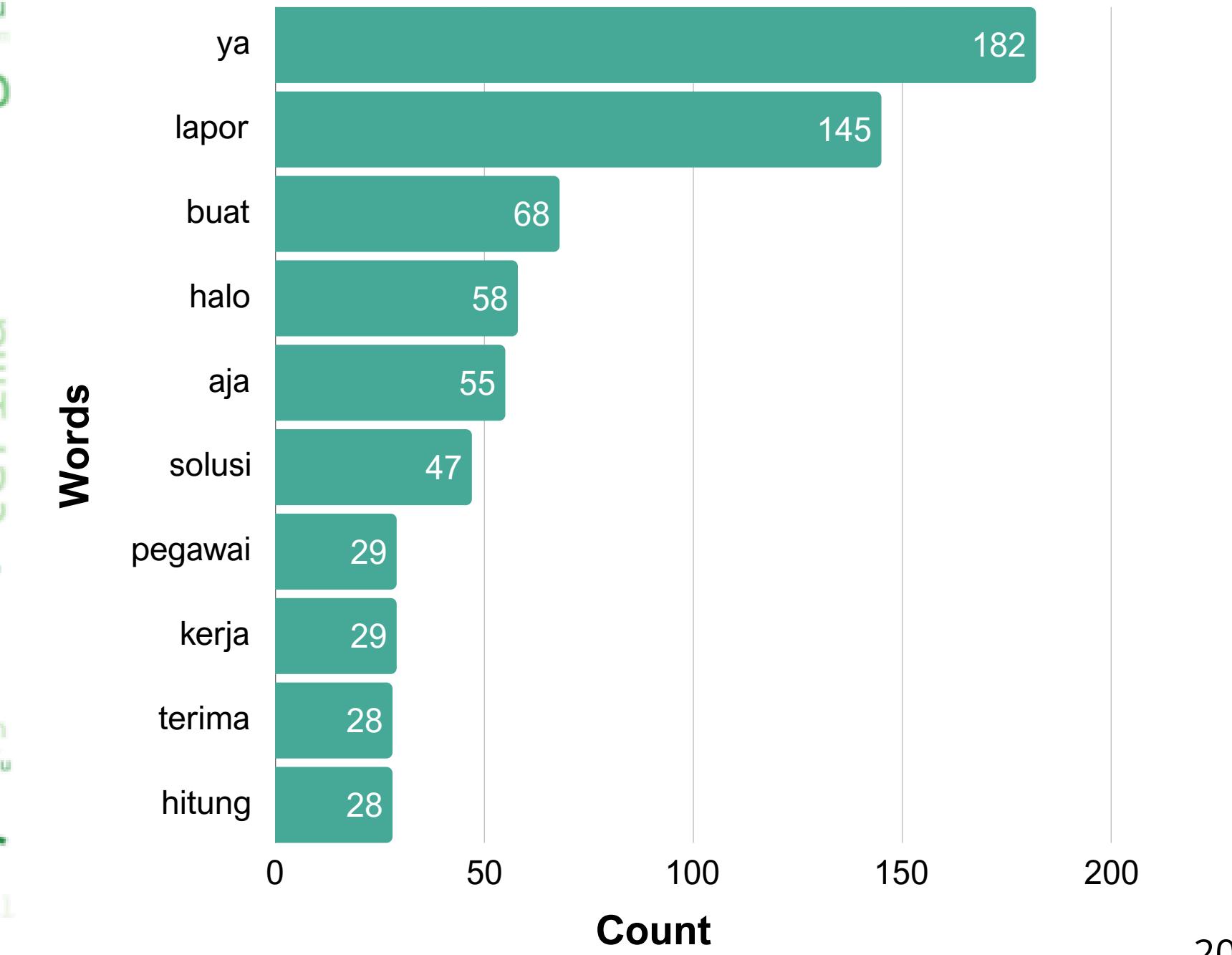
Sentiment Polarity EDA

Top 10 Most Common 2-grams



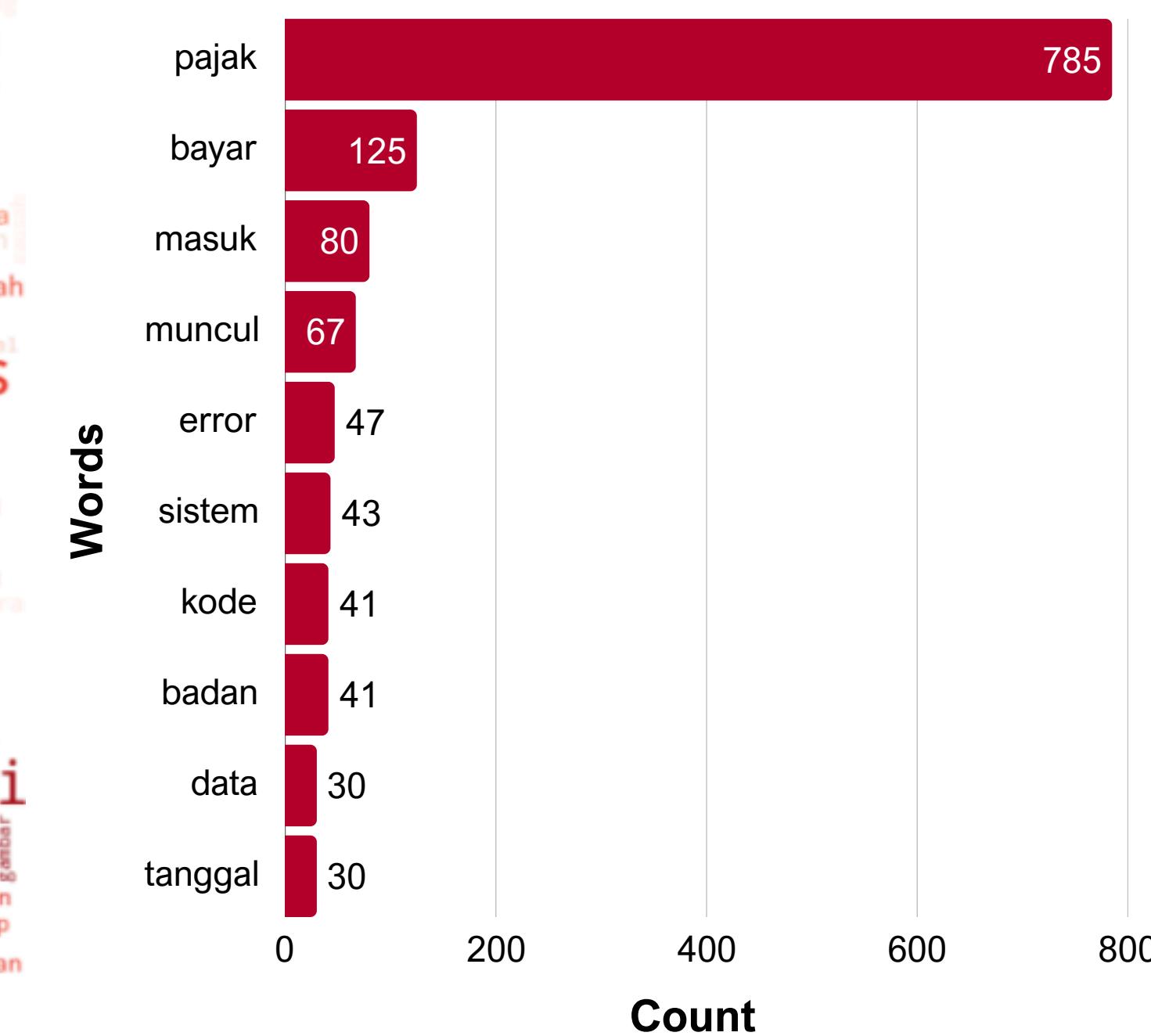
WordClouds Data Tweets

Positive Words

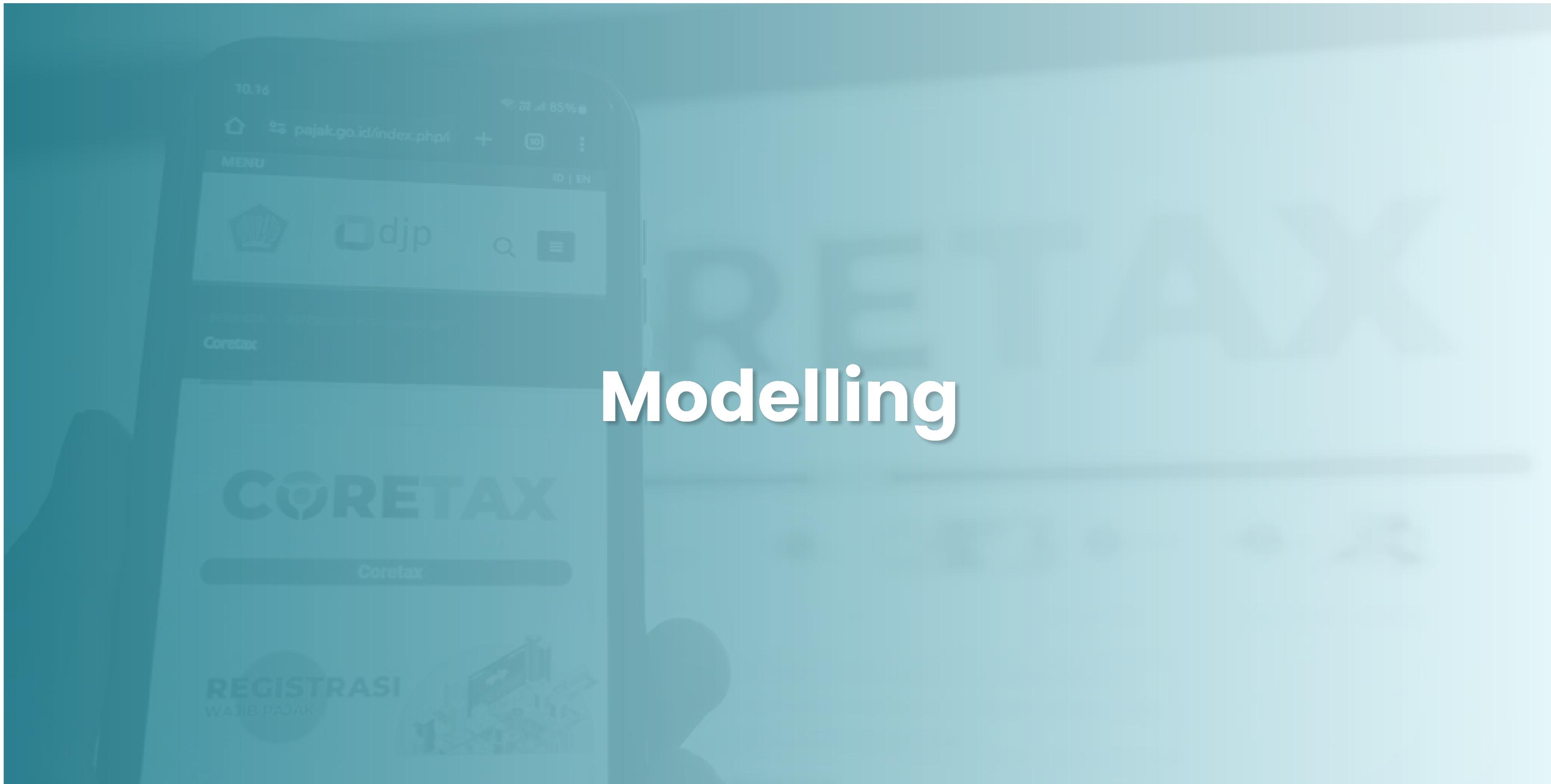


WordClouds Data Tweets

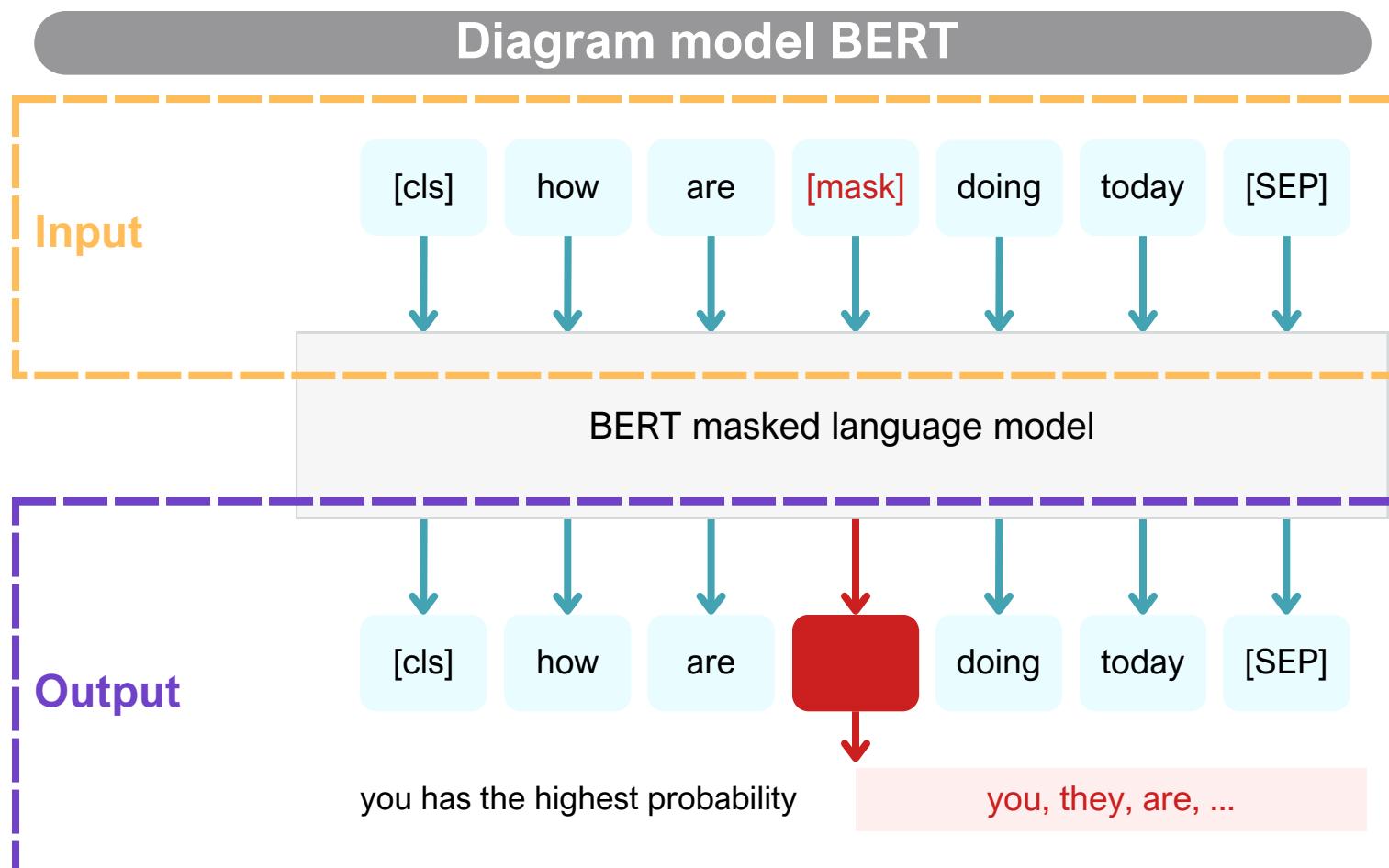
Negative Words



Modelling



Natural Language Processing (NLP) with IndoBERT



Dioptimalkan untuk Teks Bahasa Indonesia:

Versi terbaru IndoBERT dengan pemahaman kontekstual yang lebih baik.

Dibangun dengan Arsitektur BERT:

Dilatih menggunakan korpus bahasa Indonesia yang luas untuk meningkatkan performa NLP.

Analisis Sentimen pada Tweet Terkait Coretax:

Mampu mengklasifikasikan sentimen secara akurat dengan menangkap nuansa bahasa dan slang.

Modelling

Pre-Processing

Lakukan pengkodean numerik pada label sentimen dan bagi dataset menjadi data pelatihan/pengujian untuk mempersiapkan dataframe sebelum tokenisasi.

Tokenization

Ubah teks menjadi format token menggunakan tokenizer IndoBERT-base-2, memastikan kompatibilitas dengan kosakata pra-terlatih model.

Model Initialization

Muat model IndoBERT-base-2 dengan bobot pra-terlatih dan konfigurasikan untuk tugas klasifikasi sentimen.

Modelling

Training Setup

Tentukan fungsi loss, optimizer, dan jadwal learning rate untuk mengoptimalkan kinerja model selama pelatihan.

Hyperparameter Search

Uji berbagai hyperparameter (misalnya, batch size, learning rate, dropout) untuk menemukan konfigurasi terbaik dalam melatih IndoBERT-base-2.

Train the Model

Latih IndoBERT-base-2 menggunakan data sentimen berlabel dengan memanfaatkan akselerasi GPU untuk meningkatkan akurasi.

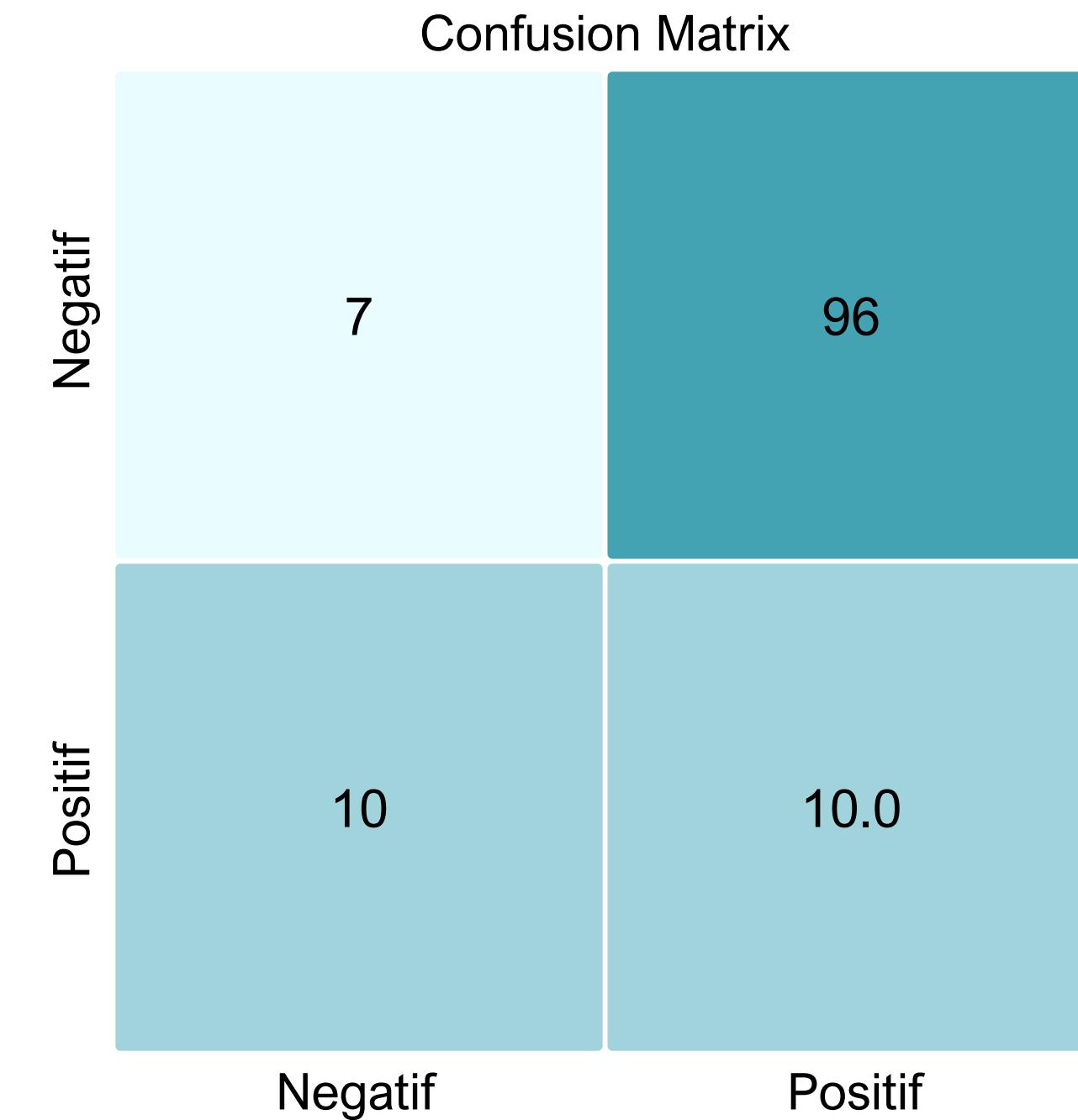
Hasil

Metrics	Training	Testing
Accuracy	0.88	0.86
Recall	0.89	0.86
Precision	0.88	0.85
F1 Score	0.85	0.73

Hyperparameter terbaik yang diperoleh:

- Learning Rate: 4.45e-05
- Epochs: 5
- Weight Decay: 0.04
- Warmup Ratio: 0.145
- Batch Size: 32

Dengan kombinasi ini, model mencapai akurasi 86% pada dataset uji, namun masih perlu peningkatan pada recall untuk kelas positif.



Kesimpulan dan Rekomendasi



Kesimpulan

1. Sentimen Majoritas Negatif

- 1 Analisis Lexicon-Based menunjukkan banyak keluhan terkait error sistem, kendala pelaporan, dan kebingungan pengguna.
- 2 Model IndoBERT memiliki akurasi 88%, namun recall untuk sentimen positif masih rendah (0.25), menunjukkan bias terhadap sentimen negatif.

2. Topik Utama dalam Data

- 1 Kata dominan dalam EDA & WordCloud: "pajak," "coretax," "lapor," "spt," "bayar".
- 2 WordCloud negatif menampilkan kata-kata seperti "error," "gagal," "tidak bisa", memperkuat temuan bahwa kendala teknis adalah isu utama.
- 3 WordCloud positif jauh lebih kecil, sering kali berupa apresiasi terhadap layanan pajak yang berjalan lancar.

Rekomendasi

Coretax sebagai Jasa Layanan Pajak Online

Optimasi Sistem & Layanan Coretax

- Perbaiki *bug* & *error* sistem untuk meningkatkan stabilitas layanan.
- Evaluasi kembali alur pelaporan pajak agar lebih intuitif dan efisien.
- Tingkatkan dukungan pelanggan dengan respons lebih cepat terhadap keluhan.

Peningkatan Transparansi & Komunikasi

- Gunakan analisis sentimen otomatis untuk memantau opini pengguna secara real-time.
- Publikasikan pembaruan sistem & solusi kendala teknis untuk mengurangi kebingungan pengguna.

Rekomendasi Perspektif Teknologi (IndoBERT)

Optimalisasi Model Sentimen

- Lakukan rebalancing dataset & data augmentation untuk mengurangi bias sentimen negatif.
- Fine-tune IndoBERT dengan dataset pajak yang lebih luas untuk meningkatkan akurasi.

Penerapan NLP dalam Monitoring Coretax

- Gunakan IndoBERT untuk deteksi dini isu sistem melalui analisis pola keluhan pengguna.
- Implementasikan dashboard sentimen real-time untuk pemantauan otomatis layanan Coretax.

Rekomendasi

Rekomendasi Umum untuk Peningkatan Layanan Digital

Edukasi & Literasi Pajak Digital

- Buat FAQ interaktif dan chatbot AI untuk membantu pengguna mengatasi kendala teknis.
- Adakan webinar & sosialisasi untuk meningkatkan pemahaman wajib pajak.

Analisis Sentimen Berkelanjutan

- Lakukan monitoring sentimen berkala untuk mengukur efektivitas perbaikan layanan.
- Integrasikan hasil sentimen dengan pengambilan kebijakan di Coretax.

