

Sentiment Analysis on Rupiah Depreciation Against USD Using XGBoost

Ni Komang Purnama Indrayuni ^{1*}, Ni Made Mila Rosa Desmayani ^{2*}, I Dewa Ayu Agung Tantri Pramawati ^{3*},
I Made Subrata Sandhiyasa ^{4*}, Komang Kurniawan Widiartha ^{5*}

* Program Studi Informatika, Fakultas Teknologi dan Informatika, Institut Bisnis dan Teknologi Indonesia Bali, Denpasar
purnamaindrayuni@gmail.com ¹, milarosadesmayani@instiki.ac.id ², dwayutantripramawati@gmail.com ³,
subrata.sandhiyasa@instiki.ac.id ⁴, komang.kurniawan@instiki.ac.id ⁵

Article Info

Article history:

Received 2025-08-11

Revised 2025-08-29

Accepted 2025-09-04

Keyword:

*Extreme Gradient Boosting,
Exchange Rates,
Sentiment Analysis,
Social Media.*

ABSTRACT

The depreciation of the rupiah against the United States dollar (USD) affects purchasing power and economic stability. Public responses are widely expressed through social media such as X and Instagram. This study aims to analyze public sentiment using the *Extreme Gradient Boosting* (XGBoost) algorithm. Data were collected through *crawling* and *scraping*, consisting of 13,443 X comments and 11,287 Instagram comments between January 2024 until April 2025. *Preprocessing* included emoji conversion, *cleaning*, *case folding*, *normalization*, *tokenization*, *stopwords removal*, and *Stemming*. Sentiment *labeling* was performed using the *InSet Lexicon*, TF-IDF weighting, and data *splitting* into 70:30, 80:20, and 90:10. The XGBoost model was trained with parameters: 100 *estimators*, *learning rate* 0.1, *max depth* 6, and *subsample* 0.8. Results showed accuracies of 74–76% on X data and stable 77% on Instagram. Model evaluation using *precision*, *recall*, and *F1-score* confirmed consistency: *precision* 0.76% – 0.84%, *recall* 0.86%–0.88%, and *F1-score* 0.82%–0.86%, reflecting a balance between accuracy and robustness in detecting sentiments. Sentiment distribution revealed that X is dominated by negative opinions (38%), while Instagram is more positive (41%). These findings confirm the effectiveness of XGBoost in sentiment classification and provide valuable insights for policymakers to design adaptive communication and monetary strategies based on digital public opinion.



This is an open access article under the [CC-BY-SA](https://creativecommons.org/licenses/by-sa/4.0/) license.

I. PENDAHULUAN

Mata uang merupakan indikator utama stabilitas ekonomi suatu negara karena mencerminkan kekuatan sekaligus kerentanannya. Rupiah sebagai mata uang nasional Indonesia, memiliki peran penting dalam kerangka ekonomi negara. Setiap negara memiliki mata uang dengan karakteristik dan nilai tukar yang berbeda terhadap mata uang asing. Nilai tukar rupiah terhadap dolar AS menjadi ukuran utama kekuatan ekonomi Indonesia karena dolar AS adalah mata uang utama dalam perdagangan dan keuangan internasional. Penyesuaian nilai tukar penting agar transaksi antarnegara berlangsung saling menguntungkan, mendorong hubungan ekonomi yang kuat. Sistem nilai tukar mencerminkan kondisi ekonomi negara dan memengaruhi pasar keuangan serta perdagangan global. Perbandingan nilai tukar antara mata uang suatu negara dengan negara lain

disebut dengan *Kurs*. Penelitian dari (Anindyntha dan Fuddin, 2023) menunjukkan bahwa volatilitas nilai tukar juga berdampak terhadap stabilitas sistem keuangan domestik serta arus investasi asing.

Kurs adalah jumlah mata uang domestik yang dibutuhkan untuk membeli satu unit mata uang asing, seperti dolar Amerika Serikat USD (*United States Dollar*). Ada dua jenis *Kurs*, yaitu *Kurs* jual dan *Kurs* beli. *Kurs* jual adalah jumlah yang dibebankan bank ketika menjual dolar kepada masyarakat, sedangkan kurs beli adalah jumlah yang diterima masyarakat saat menjual dolar kepada bank. *Kurs* dapat berubah karena berbagai faktor, termasuk suku bunga, inflasi, ketidakpastian ekonomi global, defisit perdagangan, dan pergerakan investasi asing. Ketika nilai tukar berfluktuasi banyak, hal itu dapat melemahkan mata uang lokal, seperti rupiah. Pelemahan ini dapat mengurangi daya beli masyarakat dan memengaruhi stabilitas ekonomi negara (Cahyadi dkk.,

2024), Hal ini sejalan dengan penelitian (Khusni dan Nurviliza2, 2025) menegaskan bahwa volatilitas nilai tukar rupiah secara signifikan memengaruhi indikator makroekonomi seperti pertumbuhan ekonomi dan cadangan devisa.

Berdasarkan berita dari Tempo.co (21 Mei 2025), nilai tukar rupiah terhadap dolar Amerika Serikat menunjukkan pelemahan yang cukup signifikan dalam kurun waktu April hingga Mei 2025, yakni mencapai Rp17.000 pada 7 April, menguat ke Rp16.353 pada 9 Mei, dan kembali melemah ke Rp 16.699 pada 21 Mei. Kondisi ini dipengaruhi oleh berbagai faktor, antara lain ketegangan perdagangan global, kebijakan tarif dari Amerika Serikat, serta penurunan daya beli masyarakat setelah momentum Hari Raya Idulfitri. Fenomena tersebut menimbulkan kekhawatiran terhadap laju inflasi, defisit neraca perdagangan, dan kestabilan ekonomi nasional secara keseluruhan (Wasifah Hanim, 2024). Studi ARDL terhadap data 2013–2023 menemukan inflasi dan suku bunga domestik sebagai kontributor utama fluktuasi USD/IDR, sementara harga minyak dunia memengaruhi secara jangka pendek (Hakim dan Aji, 2025).

Pemerintah Indonesia, melalui lembaga eksekutif, berperan penting dalam menjaga stabilitas nilai tukar rupiah terhadap dolar AS melalui intervensi pasar, penyesuaian suku bunga, serta kebijakan yang mendorong ekspor dan penggunaan produk dalam negeri. Kolaborasi antara kebijakan fiskal dan moneter ini memperkuat ketahanan ekonomi Indonesia, khususnya menghadapi pandemi COVID-19 dan ketidakpastian global. Pendekatan terkoordinasi ini menjadi kunci dalam mendorong stabilitas dan pertumbuhan ekonomi berkelanjutan (Samsul dkk., 2021)

Menyusun kebijakan yang efektif untuk menghadapi perubahan nilai tukar rupiah terhadap dolar AS tidak hanya melibatkan pemerintah dan lembaga keuangan, tetapi juga masyarakat, khususnya pelaku usaha. Ketika masyarakat memahami bagaimana perubahan nilai tukar dapat menyebabkan harga barang impor menjadi lebih tinggi dan daya beli menurun, hal tersebut dapat membantu menjaga stabilitas ekonomi. Penelitian (Lestari dan Anggraeni, 2021) menunjukkan bahwa sentimen publik selama pandemi COVID-19 berdampak signifikan terhadap nilai tukar rupiah. Kepanikan dan ketidakpastian selama masa tersebut menyebabkan banyak masyarakat mengalihkan asetnya ke dolar AS, yang meningkatkan tekanan terhadap rupiah untuk melemah. Di sisi lain, ketika masyarakat merasa optimis dan percaya terhadap tindakan pemerintah, hal tersebut membantu menstabilkan nilai tukar dengan mendorong aktivitas ekonomi dan investasi. Oleh karena itu, memahami sentimen publik sangat penting untuk menciptakan kebijakan yang efektif.

Sentimen publik terhadap fluktuasi nilai tukar rupiah dapat dianalisis melalui respons di media sosial X dan juga Instagram. Sebagai *Platform* yang banyak digunakan, media sosial X dan Instagram memungkinkan pengguna untuk menyampaikan pendapat dan berbagi informasi dengan cepat. Data yang dihasilkan dapat menjadi sumber yang berharga

untuk memahami persepsi publik terhadap isu-isu ekonomi seperti perubahan mata uang. Analisis tweet membantu mengukur persepsi publik, di mana sentimen positif mencerminkan dukungan dan sentimen negatif menunjukkan ketidakpuasan terhadap kebijakan. Oleh karena itu, analisis sentimen di media sosial penting untuk merumuskan kebijakan yang responsif dan tepat sasaran.

Analisis sentimen merupakan proses penting yang mengkaji dan mengelompokkan opini dalam teks ke dalam tiga kategori utama yaitu positif, negatif, dan netral. Dengan berkembangnya media sosial seperti X dan Instagram, berbagai metode analisis sentimen telah dievaluasi berdasarkan akurasi, *presisi*, dan kompleksitas algoritma. Salah satu metode populer adalah *Naïve Bayes*, yang dikenal sederhana dan cepat, namun akurasinya cenderung lebih rendah dalam menghadapi data besar dan kompleks. Studi oleh (Atmajaya dkk., 2023) menunjukkan bahwa *Naïve Bayes* memiliki akurasi 47%, sedangkan *Support Vector Machine* (SVM) lebih unggul dengan akurasi 59%, dengan skor *presisi* yang sama. Di sisi lain, *Extreme Gradient Boosting* (XGBoost) menjadi pesaing kuat karena kemampuannya mengurangi *overfitting* dan menghasilkan prediksi akurat. Penelitian oleh (Yulianti dkk., 2022) menjelaskan bahwa XGBoost mencapai akurasi 80,39% dan *presisi* 81,38%. Ini menegaskan efektivitas XGBoost dalam analisis sentimen, meskipun disertai dengan kompleksitas yang lebih tinggi dan kebutuhan penyetelan parameter yang cermat, menunjukkan adanya keseimbangan antara performa dan kerumitan model.

Random Forest dan *Decision Tree* adalah metode yang umum digunakan dalam analisis sentimen. *Decision Tree* mudah dipahami, namun rentan terhadap *overfitting* pada data besar. Sebaliknya, *Random Forest* mengatasi masalah ini dengan menggabungkan banyak pohon keputusan, sehingga meningkatkan akurasi dan *presisi*. Menurut (Muttaqin dan Kharisudin, 2021), SVM menunjukkan performa lebih baik daripada *K-Nearest Neighbors* (KNN), dengan akurasi 87,98% dibandingkan 82,14%. Sementara itu, *Regresi Logistik* dipilih karena kesederhanaannya, namun kurang efektif menangani pola non-*linier*, menghasilkan akurasi 78,57% (Assaidi dan Amin, 2022). KNN cocok untuk data kecil hingga sedang, tetapi kurang efisien pada data besar karena beban komputasi yang tinggi, selain itu penelitian (Oktafia dan Nugroho, 2024) menunjukkan bahwa XGBoost unggul dalam kecepatan dan performa klasifikasi, terutama pada data media sosial berskala besar.

Oleh karena itu, penelitian ini menggunakan metode *Extreme Gradient Boosting* (XGBoost), algoritma yang umum digunakan dalam klasifikasi sentimen di media sosial seperti X dan Instagram. Sehingga mampu mengelompokkan teks menjadi sentimen positif, negatif, dan netral.

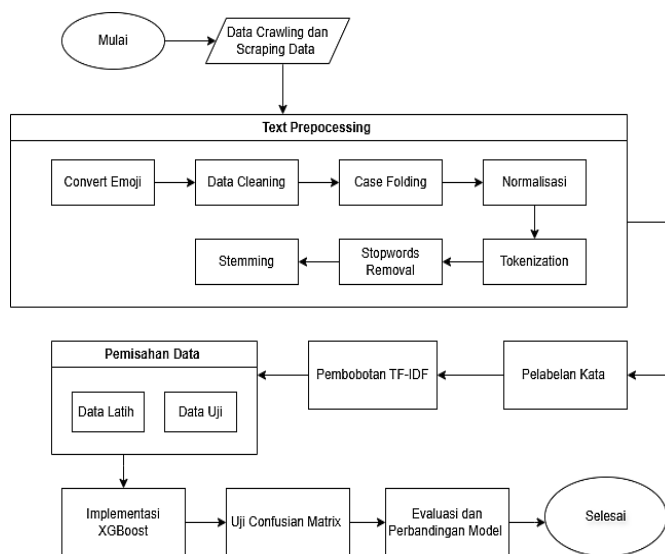
Melihat potensi tersebut, penelitian ini bertujuan untuk mengidentifikasi kecenderungan sentimen masyarakat terhadap melemahnya nilai tukar rupiah terhadap dolar AS dengan memanfaatkan data dari media sosial X dan Instagram. Selain itu, penelitian ini juga menganalisis

performa algoritma XGBoost berdasarkan metrik evaluasi seperti akurasi, *presisi*, *recall*, dan *F1-score*. Hasil dari penelitian ini diharapkan dapat memberikan kontribusi dalam pengembangan ilmu di bidang analisis sentimen dan ilmu data sosial, serta menjadi sumber referensi bagi pemangku kebijakan untuk memahami lebih dalam persepsi masyarakat. Dengan memanfaatkan teknologi dan pendekatan berbasis data, diharapkan tercipta kebijakan ekonomi yang lebih adaptif terhadap dinamika opini publik yang berkembang di ruang digital.

Dengan demikian, penelitian ini tidak hanya memberikan kontribusi akademis dalam analisis sentimen, tetapi juga menawarkan solusi strategis bagi pemerintah dalam merumuskan kebijakan stabilisasi rupiah berbasis opini publik digital.

II. METODE PENELITIAN

Proses pada penelitian ini akan dilakukan dengan alur penelitian seperti yang digambarkan pada Gambar 1.



Gambar 1 Alur Penelitian

Alur penelitian yang dijabarkan pada Gambar 1 menjelaskan bagaimana penelitian ini dilakukan :

1. Data Crawling dan Scraping Data

Proses *Crawling* data dan *Scraping* data atau pengambilan data dari media sosial X dan Instagram dengan memanfaatkan *OAuth Token* yang dimiliki setiap pengguna. Data set yang diperoleh masih berbentuk *file raw* yang terdiri dari kolom-kolom yang tidak relevan dan tidak diperlukan untuk penelitian, seperti *conversation_id_str*, *created_at*, *favorite_count*, *full_text* (tweet), *id_str*, *image_url*, *in_reply_to_screen_name*, *lang*, *location*, *quote_count*, *reply_count*, *retweet_count*, *tweet_url*, dan *username* sehingga sebelum memasuki tahap *Preprocessing*, peneliti melakukan *editing* secara

manual terhadap *file raw* tersebut sehingga yang tersisa hanya kolom *username* dan *full_text* (tweet /komentar) total data yang berhasil dikumpulkan terdiri dari 13.443 data dari X dan 11.287 data dari Instagram. Berikut merupakan contoh dari data mentah X yang ditunjukkan pada Tabel 1 dan Instagram pada Tabel 2.

TABEL 1
DATA MENTAH X

created_at	full_text	username
Thu Apr 17 03:52:40	Setiap kali nilai tukar Rupiah turun, yang jadi korban selalu rakyat kecil. Harga barang naik, gaji segitu-gitu aja. Pemerintah cuma diam!	gwathrenin_
Sun Apr 13 01:42:29	@cloudythvnder Rupiah melemah sih, tapi ini kesempatan bagus buat pelaku ekspor makin berkembang. Semangat UMKM lokal!"	Berita_satu
Wed Apr 09 15:10:34	@Kimberley_PS08 Harus diakui pemerintahan sekarang memang pintar buktinya rupiah semakin melemah.	shalatlah

TABEL 2
DATA MENTAH INSTAGRAM

userName	text	date
Afriansyah pongoh92	Kalau yg paham global saat ini seperti apa pasti mengerti tidak salahkan pemerintah 🤔👍	12/02/2025
ahmadtarzana	Rupiah semakin hampir tidak ada harganya. Inilah bumi dengan semboyan gemah ripah loh jinawi 🔥🔥🔥	28/02/2025
maw1633	Gaji stuck, rupiah terjun gimana ga lesuh per ekonomian 🤔	12/03/2025

2. Text Preprocessing

Preprocessing adalah proses membersihkan teks sehingga menjadi berkualitas dan memenuhi persyaratan untuk dieksekusi pada sebuah algoritma atau siap dianalisis oleh model *machine learning* (Larasati dkk., 2022). Tahap ini penting untuk dilakukan karena data yang diperoleh melalui tahap *crawling* seringkali dalam bentuk raw, tidak lengkap, atau memiliki format yang tidak konsisten. Berikut tahapan *Preprocessing* adalah sebagai berikut:

- *Convert Emoji*: *Convert* emoji adalah proses mengubah emoticon, yaitu gabungan simbol atau karakter yang merepresentasikan ekspresi wajah atau emosi, menjadi bentuk teks yang memiliki makna jelas sehingga dapat digunakan untuk keperluan analisis atau pemrosesan data lanjutan. Berikut merupakan contoh kamus yang digunakan dalam *Convert* emoji yang ditunjukkan pada Tabel 3.

TABEL 3
KAMUS CONVERT EMOJI

Emoji	Arti
thinking face	Wajah Berpikir
rolling_on_the_floor_lau ghing	Tertawa Terbahak- bahak
smiling_face_with_sungl asses	pakai kacamata hitam

- *Data Cleaning*: Proses membersihkan data yang tidak relevan, seperti URL, Hashtag, mention, dan karakter atau simbol yang tidak standar seperti: ('~&#[0-9]+;').
- *Case folding*: Proses mengubah semua huruf dalam data menjadi huruf kecil untuk *konsistensi* dan memudahkan analisis. *Case folding* akan menyederhanakan data dan mengurangi variasi kata karena perbedaan kapitalisasi huruf. Penyederhanaan ini perlu dilakukan karena komputer memiliki perbedaan perlakuan terhadap kapitalisasi besar dan kecil meskipun itu kata yang sama.
- *Normalisasi*: Proses yang bertujuan untuk merubah format penulisan kata agar lebih konsisten dan mudah dianalisis, misalnya 'tdk' menjadi 'tidak', 'gpp' menjadi 'tidak apa apa', 'kl' menjadi 'kalau'. Dengan memperbaiki format penulisan, algoritma atau model yang dibangun dapat memahami teks dengan lebih baik dan menghasilkan analisis yang lebih bisa diandalkan. Berikut merupakan contoh kamus kata baku yang digunakan dalam proses normalisasi yang ditunjukkan pada Tabel 4.

TABEL 4
KAMUS KATA BAKU

tidak baku	kata baku
skrg	sekarang
udah	sudah
dri	dari

- *Tokenization*: Proses memecah atau membagi teks dalam dokumen menjadi kata-kata individual sehingga menjadi lebih kecil dan lebih bermakna, yang disebut token. *Tokenization* memecah teks menjadi kalimat dengan menggunakan tanda baca

seperti titik, koma, dan tanda tanya, kemudian kalimat yang diperoleh akan dipecah lagi dengan memanfaatkan spasi putih yang memisahkan setiap kata.

- *Stopwords removal*: Proses menghapus kata-kata yang umum dan tidak memiliki makna yang signifikan, seperti: yang, dan, di, ke, dari, pada, menurut, dengan, ini, serta itu. Kata-kata ini perlu dihilangkan karena tidak memberikan informasi yang berguna untuk proses analisis.
- *Stemming*: Mengubah kata menjadi bentuk dasarnya dengan menghilangkan imbuhan. Tahap ini bertujuan untuk menyederhanakan kata dan meningkatkan konsistensi (misalnya "berlari" → "lari").

3. Pelabelan Kata

Proses pelabelan kata atau sentimen *labeling* dalam analisis sentimen mengidentifikasi opini dan sentimen dalam teks dengan mengklasifikasikan setiap kata ke dalam kelas positif, negatif, atau netral berdasarkan maknanya. *InSet Lexicon* adalah kamus sentimen bahasa Indonesia yang dikembangkan oleh tim peneliti Universitas Melbourne. Kamus *InSet Lexicon* cukup populer digunakan dalam proses *labeling* karena memuat 3.609 kata positif dan 6.609 kata negatif dengan bobot nilai atau *polarity Score* yang diklasifikasi Total skor dari setiap teks digunakan untuk menentukan label sentimen: positif jika skor > 0, netral jika skor = 0, dan negatif jika skor < 0. Contoh klasifikasi kata positif dan negatif pada *InSet Lexicon* dan *polarity Score* ditunjukkan pada Tabel 5.

TABEL 5
KAMUS INSET LEXICON

No	Kata Positif	<i>Polarity Score</i>	Kata Negatif	<i>Polarity Score</i>
1	kesempatan	3	salah	-4
2	rakyat	4	lesu	-2
3	hemat	4	lemah	-4
4	Tertawa	5	rug	-5
5	harga	3	turun	-3

4. Pembobotan TF – IDF

TF-IDF (*Term Frequency-Inverse Document Frequency*) adalah metode pembobotan kata dalam teks untuk keperluan klasifikasi. TF menghitung seberapa sering kata muncul dalam dokumen, sedangkan IDF memberi bobot lebih besar pada kata yang jarang muncul di seluruh dokumen. Nilai TF-IDF diperoleh dari hasil perkalian TF dan IDF, yang mencerminkan pentingnya kata dalam dokumen (Wati dkk., 2023)

- *Term Frequency* (TF)
Rasio kemunculan suatu kata terhadap total kata dalam dokumen untuk menentukan tingkat kepentingannya.

$$tf(t, d) = \frac{\text{Jumlah kemunculan kata } t \text{ dalam dokumen } d}{\text{Jumlah total kata dalam dokumen } d}$$

- *Inverse Document Frequency (IDF)*
pembobotan kata dalam TF-IDF yang memberikan bobot yang lebih tinggi pada kata-kata yang jarang muncul.
$$idf(t, d) = \log \frac{\text{Total jumlah data dalam dokumen } d}{\text{Jumlah data dimana kata } t \text{ muncul}}$$
- *Term Frequency-Inverse Document Frequency (TF-IDF)*
Statistik yang menunjukkan tingkat kepentingan suatu kata dalam dokumen berdasarkan frekuensi kemunculannya dan seberapa jarang kata tersebut muncul di seluruh dokumen.

$$TF - IDF(t, d) = tf \times id$$

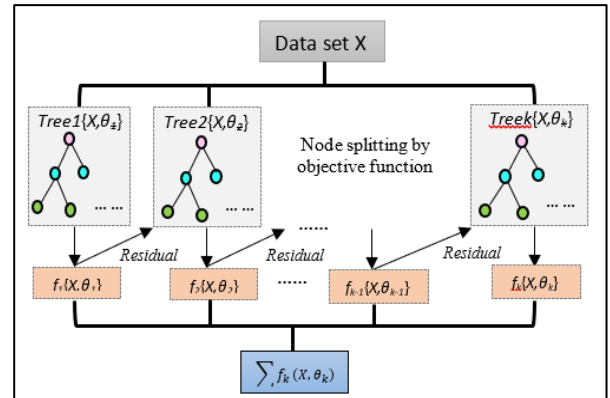
5. Data Splitting

Setelah melakukan pembobotan kata, selanjutnya dilakukan data *splitting* (pembagian data) menjadi data latih dan data uji. Data latih digunakan untuk membuat dan mengembangkan model. Model kemudian akan melakukan pengujian terhadap data uji untuk menguji performa model yang telah dibangun. Dalam penelitian ini, untuk mendapat performa terbaik maka pembagian data latih dan data uji dibagi dengan perbandingan 70:30, 80:20, 90:10. Pemilihan perbandingan yang tepat berpengaruh pada hasil akhir model. Perbandingan yang terlalu kecil untuk data uji dapat memberikan evaluasi yang kurang akurat, sedangkan perbandingan yang terlalu kecil untuk data latih dapat menyebabkan model *underfitting* yang menyebabkan model kurang terlatih dalam menganalisa data.

6. Implementasi Extreme Gradient Boosting

Extreme Gradient Boosting (XGBoost) merupakan algoritma *machine learning* berbasis pohon keputusan yang dirancang untuk menyelesaikan masalah klasifikasi maupun regresi. XGBoost menggunakan pendekatan *boosting*, yaitu teknik *ensemble* yang bekerja secara bertahap untuk membentuk pohon-pohon keputusan secara berurutan, di mana setiap pohon baru bertugas untuk memperbaiki kesalahan prediksi dari pohon sebelumnya. Proses ini memungkinkan XGBoost untuk meminimalkan fungsi kerugian (*loss function*) secara

lebih efisien, menghasilkan model yang lebih akurat, dan lebih tahan terhadap *overfitting*.



Gambar 2 Diagram Skema XGboost

Gambar 2 memperlihatkan alur proses komputasi algoritma XGBoost. Algoritma ini membangun *regression tree* secara teratur untuk menghasilkan model klasifikasi yang optimal dan mengurangi risiko *overfitting*. Dalam pelatihannya, XGBoost mengoptimalkan fungsi objektif yang terdiri dari dua komponen, yaitu *training loss* untuk mengukur akurasi prediksi, dan regularisasi untuk mengontrol kompleksitas model (Yulistiani, 2024). Pemilihan XGBoost dalam penelitian ini didasarkan pada keunggulannya dibandingkan algoritma lain, seperti *Naïve Bayes* atau *Logistic Regression* yang sederhana namun kurang mampu menangani data dalam jumlah besar. XGBoost mendukung proses komputasi paralel, dilengkapi dengan regularisasi (*L1* dan *L2*) untuk mengurangi *overfitting*.

7. Confusion Matrix

Tabel evaluasi yang membandingkan hasil prediksi model dengan data aktual untuk menilai kinerja klasifikasi. Tabel ini menunjukkan jumlah prediksi yang benar dan salah untuk masing-masing kelas, serta memungkinkan perhitungan metrik seperti *accuracy*, *precision*, *recall*, dan *F1-score* (Rininda dkk., 2024).

TABEL 6
CONFUSION MATRIX

		Prediksi		
		Positif	Netral	Negatif
Actual	Positif	True Positif (TP)	False Netral (FLp)	False Negatif (FNp)
	Netral	False Positif (FPI)	True Netral (TL)	False Negatif (FNn)
	Negatif	False Positif (Fpn)	False Netral (FLn)	True Negatif (TN)

Berdasarkan tabel *Confusion Matrix* pada Tabel 6 diperoleh:

- a) True Positif (TP): Data dengan aktual positif dan diprediksi positif.
- b) False Positif (FP): Data dengan aktual bukan positif yang diprediksi positif.
- c) True Negatif (TN): Data dengan aktual negatif dan diprediksi negatif.
- d) False Negatif (FN): Data dengan aktual bukan negatif yang diprediksi negatif.
- e) True Netral (TL): Data dengan aktual netral dan diprediksi netral.
- f) False Netral (FL): Data dengan aktual bukan netral yang diprediksi netral

Rumus *confusion matrix* untuk menghitung *accuracy*, *precision*, dan *recall* seperti berikut:

$$Accuracy = \frac{TP + TL + TN}{Total}$$

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

$$Recall = \frac{TP}{TP + FLp + FNp}$$

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

III. HASIL DAN PEMBAHASAN

1. Data Crawling

Pada tahap awal dilakukan proses pengumpulan dan pelabelan data. Data diperoleh melalui teknik *crawling* untuk media sosial X dan web *scraping* untuk Instagram. Hasil pengambilan data menghasilkan 13.443 komentar dari X dan 11.287 komentar dari Instagram. Setelah melalui tahap *preprocessing*, jumlah data yang dapat digunakan menjadi 11.638 dari X dan 8.188 dari Instagram. Seluruh data dikumpulkan dalam rentang waktu Januari 2024 hingga April 2025.

2. Text Preprocessing

Berikut merupakan tahapan dari *preprocessing* dataset yang ditampilkan pada Tabel 7.

TABEL 7
HASIL PREPROCESSING

Tahapan Preprocessing	Data Preprocessing	Hasil Preprocessing
Convert Emotikon	Kalau yg paham global saat ini seperti apa pasti mengerti tidak	Kalau yg paham global saat ini seperti apa pasti mengerti tidak

	salahkan pemerintah 🤔👍	salahkan pemerintah pakai kacamata hitam jempol
	Setiap kali nilai tukar Rupiah turun, yang jadi korban selalu rakyat kecil. Harga barang naik, gaji segitu-gitu aja. Pemerintah cuma diam!	Setiap kali nilai tukar Rupiah turun, yang jadi korban selalu rakyat kecil. Harga barang naik, gaji segitu-gitu aja. Pemerintah cuma diam!
Data Cleaning	Kalau yg paham global saat ini seperti apa pasti mengerti tidak salahkan pemerintah pakai kacamata hitam jempol	Kalau yg paham global saat ini seperti apa pasti mengerti tidak salahkan pemerintah pakai kacamata hitamjempol
	Setiap kali nilai tukar Rupiah turun, yang jadi korban selalu rakyat kecil. Harga barang naik, gaji segitu-gitu aja. Pemerintah cuma diam!	Setiap kali nilai tukar Rupiah turun yang jadi korban selalu rakyat kecil Harga barang naik gaji segitugitu aja Pemerintah cuma diam
Case folding	Kalau yg paham global saat ini seperti apa pasti mengerti tidak salahkan pemerintah pakai kacamata hitamjempol	kalau yg paham global saat ini seperti apa pasti mengerti tidak salahkan pemerintah pakai kacamata hitam jempol
	Setiap kali nilai tukar Rupiah turun yang jadi korban selalu rakyat kecil Harga barang naik gaji segitugitu aja Pemerintah cuma diam	setiap kali nilai tukar rupiah turun yang jadi korban selalu rakyat kecil harga barang naik gaji segitugitu aja pemerintah cuma diam
Normalisasi	kalau yg paham global saat ini seperti apa pasti mengerti tidak salahkan pemerintah pakai kacamata hitam jempol	kalau yang paham global saat ini seperti apa pasti mengerti tidak salahkan pemerintah pakai kacamata hitamjempol
	setiap kali nilai tukar rupiah turun yang jadi korban selalu rakyat kecil harga barang naik gaji segitugitu aja	setiap kali nilai tukar rupiah turun yang jadi korban selalu rakyat kecil harga barang naik gaji segitugitu

	pemerintah cuma diam	saja pemerintah cuma diam
Tokenization	kalau yang paham global saat ini seperti apa pasti mengerti tidak salahkan pemerintah pakai kacamata hitamjempol	['kalau', 'yang', 'paham', 'global', 'saat', 'ini', 'seperti', 'apa', 'pasti', 'mengerti', 'tidak', 'salahkan', 'pemerintah', 'pakai', 'kacamata', 'hitam', 'jempol']
	setiap kali nilai tukar rupiah turun yang jadi korban selalu rakyat kecil harga barang naik gaji segitugitu saja pemerintah cuma diam	['setiap', 'kali', 'nilai', 'tukar', 'rupiah', 'turun', 'yang', 'jadi', 'korban', 'selalu', 'rakyat', 'kecil', 'harga', 'barang', 'naik', 'gaji', 'segitugu', 'gitu', 'saja', 'pemerintah', 'cuma', 'diam']
Stopwords removal	['kalau', 'yang', 'paham', 'global', 'saat', 'ini', 'seperti', 'apa', 'pasti', 'mengerti', 'tidak', 'salahkan', 'pemerintah', 'pakai', 'kacamata', 'hitam', 'jempol']	['paham', 'global', 'mengerti', 'salahkan', 'pemerintah', 'pakai', 'kacamata', 'hitam', 'jempol'],
	['setiap', 'kali', 'nilai', 'tukar', 'rupiah', 'turun', 'yang', 'jadi', 'korban', 'selalu', 'rakyat', 'kecil', 'harga', 'barang', 'naik', 'gaji', 'segitugu', 'gitu', 'saja', 'pemerintah', 'cuma', 'diam']	['kali', 'nilai', 'tukar', 'rupiah', 'turun', 'korban', 'rakyat', 'harga', 'barang', 'gaji', 'segitugitu', 'pemerintah', 'diam'],
Stemming	['paham', 'global', 'mengerti', 'salahkan', 'pemerintah', 'pakai', 'kacamata', 'hitam', 'jempol'],	paham global mengerti salah perintah pakai kacamata hitam jempol
	['kali', 'nilai', 'tukar', 'rupiah', 'turun', 'korban', 'rakyat', 'harga', 'barang', 'gaji', 'segitugitu', 'pemerintah', 'diam'],	kali nilai tukar rupiah turun korban rakyat harga barang gaji segitugitu perintah diam

3. Pelabelan Kata

Hasil pelabelan kata ini dilakukan dengan menggunakan kamus *Inset Lexicon*, di mana setiap kata diberi bobot tertentu. Dari bobot-bobot tersebut dihitung *polarity Score* yang akan menunjukkan apakah kalimat tersebut memiliki sentimen positif, negatif, atau netral, seperti yang ditampilkan pada Tabel 8.

TABEL 8
HASIL PELABELAN KATA

No	Hasil Preprocessing	Bobot Kata	Polarity Score	Sentimen
1	paham global mengerti salah perintah pakai kacamata hitam jempol	[4,0,3, -5,2,-4, -5,0,3,]	-2	Negative
2	kali nilai tukar rupiah turun korban rakyat harga barang gaji segitugitu perintah diam	[0,5,1,0, -4,-3,2, -2,0,1, 0,-5,2]	-3	Negative

4. Data Splitting

Berikut merupakan hasil dari data *splitting* yang ditunjukkan pada Tabel 9.

Tabel 9
Hasil Data *Splitting*

Sumber	Skenario	Data Latih	Data Uji	Total
X	70:30	8.146	3.492	11.638
	80:20	9.310	2.328	11.638
	90:10	10.474	1.164	11.638
Instagram	70:30	5.731	2.457	8.188
	80:20	6.550	1.638	8.188
	90:10	7.369	819	8.188

5. Pembobotan TF – IDF

Berikut merupakan Hasil dari pembobotan TF-IDF yang sudah dilakukan yang dapat dilihat pada Tabel 10.

TABEL 10
HASIL PEMBOBOTAN TF - IDF

Data	Term Frequency (TF)			df	D/df	IDF (log D/df)	TF-IDF (W = TF × IDF)		
	D1	D2	D3				D1	D2	D3
rupiah	1	1	1	3	1	0.000	0.000	0.000	0.000
melemah	1	0	1	2	1.5	0.176	0.176	0.000	0.176
tukar	1	0	0	1	3	0.477	0.477	0.000	0.000
rakyat	1	0	0	1	3	0.477	0.477	0.000	0.000
bagus	0	1	0	1	3	0.477	0.000	0.477	0.000
pemerintah	1	0	1	2	1.5	0.176	0.176	0.000	0.176
ekspor	0	1	0	1	3	0.477	0.000	0.477	0.000
pintar	0	0	1	1	3	0.477	0.000	0.000	0.477
berkembang	0	1	0	1	3	0.477	0.000	0.477	0.000

6. Implementasi *Extream Gradient Boosting*

Dari hasil model klasifikasi yang telah dibangun, penerapan algoritma XGBoost pada kedua *platform* menunjukkan performa yang cukup baik. *Pseudo code* untuk menghitung dan menampilkan akurasi model dapat dilihat pada Gambar 3.

```
y_pred_encoded = xgb_model.predict(x_test_tfidf)
y_pred = label_encoder.inverse_transform(y_pred_encoded)

#evaluasi performa model menggunakan y_test dan y_pred
accuracy = accuracy_score(y_test, y_pred)
print(f'Akurasi: {accuracy:.2f}')
```

Gambar 3 Pseudo code Performa Algoritma Xgboost

Hasil Akurasi yang didapatkan ialah sebagai berikut :

TABEL 11
HASIL AKURASI SKENARIO

Sumber	Skenario	Tingkat Akurasi
X	70:30	76%
	80:20	76%
	90:10	74%
Instagram	70:30	77%
	80:20	77%
	90:10	77%

Berdasarkan hasil klasifikasi yang ditampilkan pada Tabel 11, algoritma XGBoost menunjukkan performa yang cukup baik dan konsisten dalam mengklasifikasikan sentimen pada data media sosial X dan Instagram. *Akurasi* tertinggi pada data X dicapai pada skenario pembagian data 70:30 dan 80:20 sebesar 76%, sementara pada skenario 90:10 sedikit menurun menjadi 74%. Di sisi lain, *akurasi* pada data Instagram stabil di angka 77% pada ketiga skenario, yang menunjukkan bahwa model XGBoost mampu beradaptasi dengan baik terhadap struktur data Instagram. Perbedaan hasil yang tidak terlalu mencolok antara kedua *platform* menunjukkan kemampuan generalisasi XGBoost yang cukup tinggi. Hal ini menunjukkan bahwa XGBoost adalah algoritma yang cukup baik dan bisa digunakan untuk mengklasifikasikan sentimen secara otomatis, meskipun digunakan pada data dari *platform* yang berbeda.

7. Evaluasi Model

Meskipun akurasi model XGBoost telah diperoleh, evaluasi lanjutan tetap diperlukan karena akurasi saja tidak cukup untuk menilai kinerja model secara menyeluruh. Oleh karena itu, pengujian menggunakan *Confusion Matrix* dilakukan pada tiga skenario pembagian data, yaitu 70:30,

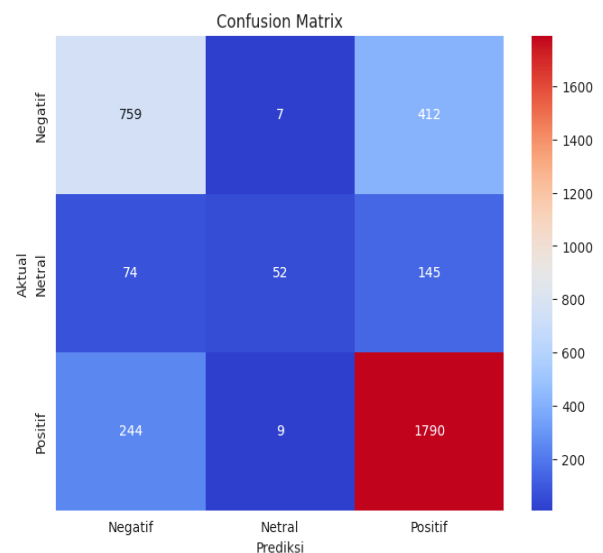
80:20, dan 90:10, untuk mengevaluasi performa model dalam kondisi yang berbeda. Dengan metode ini, dapat dianalisis sejauh mana model mampu mengklasifikasikan data secara benar maupun salah, termasuk distribusi *True Positive*, *True Negative*, *False Positive*, dan *False Negative*. Model XGBoost memiliki beberapa keterbatasan. Sensitivitas terhadap pengaturan *hyperparameter* dan kemungkinan terpengaruh oleh noise atau distribusi data yang tidak seimbang dapat menyebabkan akurasi model tidak selalu optimal. Oleh karena itu, evaluasi dengan *Confusion Matrix* menjadi penting untuk memberikan gambaran kinerja model secara lebih menyeluruh dan representatif. *Pseudo code* untuk pengujian *Confusion Matrix* ditunjukkan pada Gambar 4.

```
from sklearn.metrics import confusion_matrix
import seaborn as sns
import matplotlib.pyplot as plt

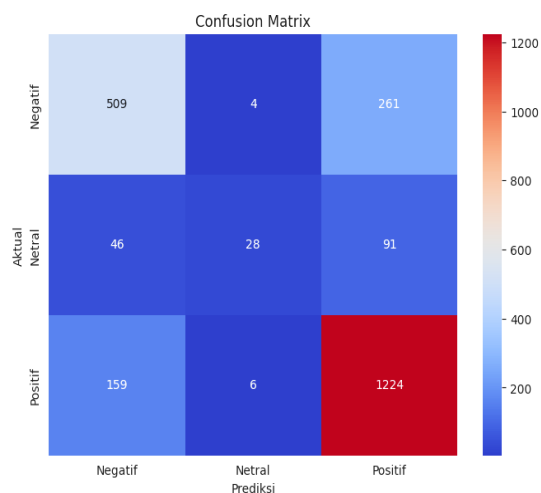
#menghitung confusion matrix
conf_matrix= confusion_matrix(y_test, y_pred)

plt.figure(figsize=(8, 6))
sns.heatmap(conf_matrix, annot=True, fmt='d', cmap='coolwarm',
xticklabels=['Negatif', 'Netral', 'Positif'], yticklabels=
['Negatif', 'Netral', 'Positif'], cbar='coolwarm')
plt.xlabel('Prediksi')
plt.ylabel('Aktual')
plt.title('Confusion Matrix')
plt.show
```

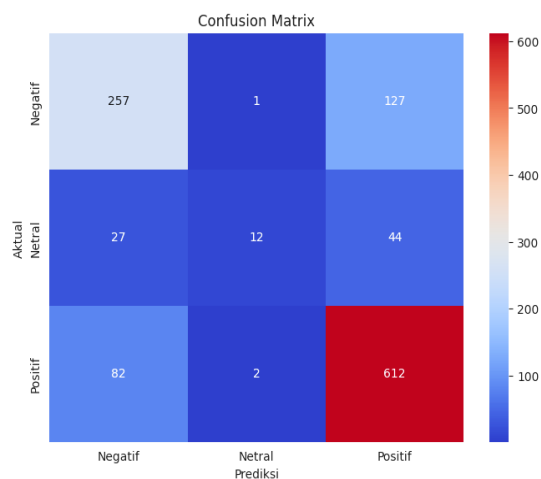
Gambar 4 Pseudo code Uji Confusion Matrix



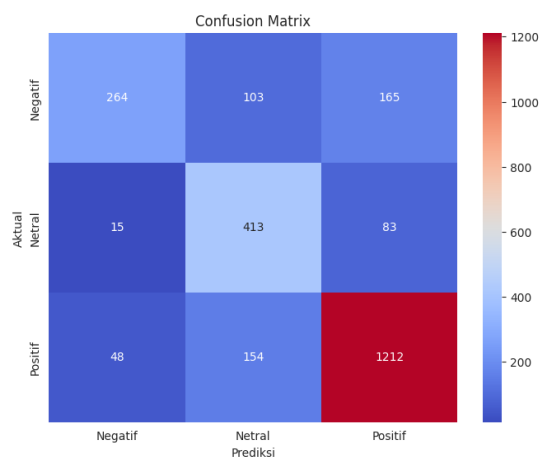
Gambar 5 Confusion Matrix pada Dataset X 70:30



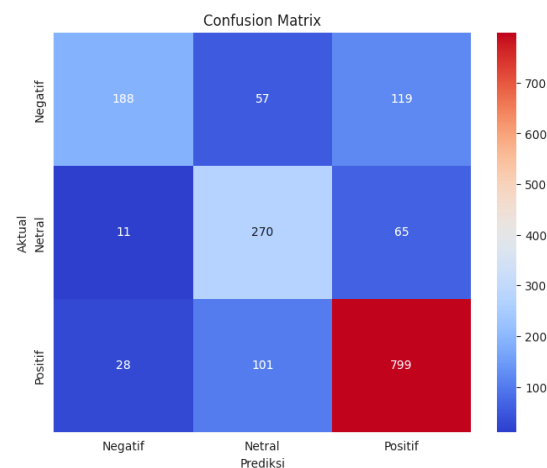
Gambar 6 Confusion Matrix pada Dataset X 80:20



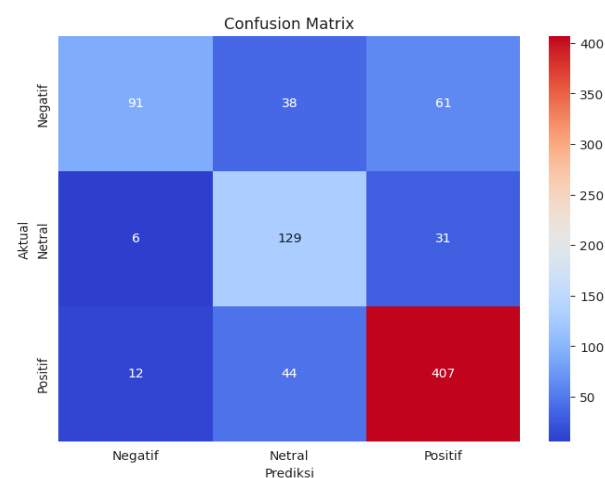
Gambar 7 Confusion Matrix pada Dataset X 90:10



Gambar 8 Confusion Matrix pada Dataset Ig 70:30



Gambar 9 Confusion Matrix pada Dataset Ig 80:20



Gambar 10 Confusion Matrix pada Dataset Ig 90:10

Dari hasil *Confusion Matrix* yang ditunjukkan oleh Gambar 5 hingga Gambar 10, metrik evaluasi untuk setiap skenario pembagian data dapat dihitung secara detail, sebagaimana ditampilkan pada Tabel 12. Parameter yang digunakan pada model XGBoost meliputi jumlah *estimators* sebanyak 100, *learning rate* 0,1, *max depth* 6, serta *subsample* 0,8. Perlu dicatat bahwa dalam penelitian ini tidak dilakukan penanganan data *imbalance* menggunakan teknik seperti SMOTE, model hanya bekerja dengan data asli tanpa penyesuaian *class weights*.

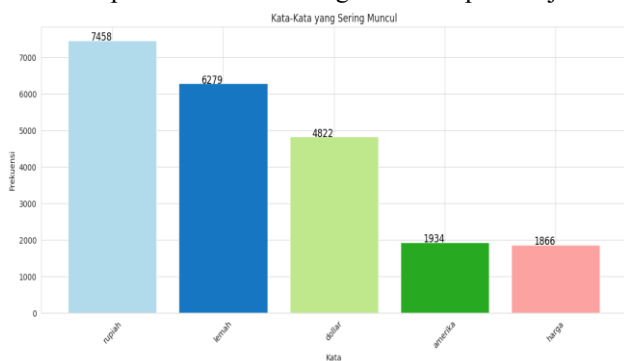
TABEL 12
HASIL IMPLEMENTASI MODEL

Skenario	Data	Presisi	Recall	F1-Score
70 : 30	X	0,76%	0,88%	0,82%
	Ig	0,83%	0,86%	0,84%
80 : 20	X	0,78%	0,88%	0,83%
	Ig	0,81%	0,86%	0,83%
90 : 20	X	0,84%	0,88%	0,86%
	Ig	0,81%	0,88%	0,85%

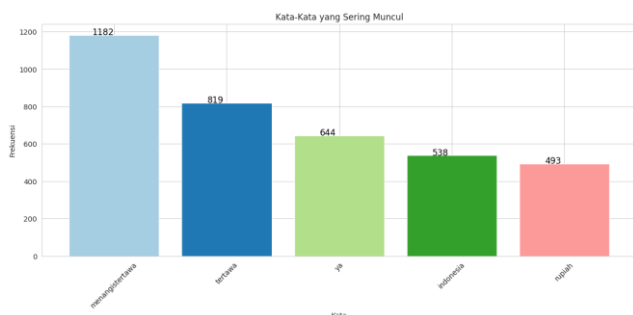
cloud Ig menunjukkan bahwa opini kurang intens terhadap permasalahan namun tetap menarik.

10. Visualisasi Data Diagram Batang

Visualisasi diagram batang merupakan metode yang digunakan untuk menggambarkan frekuensi kemunculan kata dalam dataset. Diagram ini memudahkan pembaca untuk melihat seberapa sering kata-kata tertentu muncul pada masing-masing media sosial, sehingga pola atau topik dominan dapat diidentifikasi dengan lebih cepat dan jelas.



Gambar 14 Visualisasi Diagram Batang Data X



Gambar 15 Visualisasi Diagram Batang Data Ig

Gambar 13 menunjukkan bahwa percakapan di X didominasi kata-kata informatif terkait isu pelemahan rupiah. Sebaliknya, Gambar 14 memperlihatkan Instagram lebih banyak berisi kata-kata bernuansa emosional. Hal ini menunjukkan X cenderung digunakan untuk diskusi ekonomi, sementara Instagram menonjolkan ekspresi reaksi publik.

11. Keterbatasan Model Penelitian

Meskipun algoritma XGBoost mampu mengklasifikasikan sentimen publik dengan baik, penelitian ini memiliki beberapa keterbatasan. Akurasi model berada pada kisaran 74–77%, yang dipengaruhi oleh komentar yang ambigu, sarkastik, atau ironis, serta komentar sangat singkat atau menggunakan bahasa campuran (Indonesia dan Inggris), meskipun data telah melalui *preprocessing*. Distribusi kelas sentimen yang tidak seimbang juga berpotensi menimbulkan bias, karena teknik *balancing*

seperti SMOTE atau *class weighting* tidak diterapkan. Penelitian ini hanya menggunakan XGBoost tanpa perbandingan metode lain, dan data yang dianalisis terbatas pada periode Januari 2024 hingga April 2025 serta hanya mencakup *platform* X dan Instagram, sehingga hasilnya belum tentu dapat digeneralisasikan ke konteks waktu atau media sosial lain.

IV. KESIMPULAN

Hasil penelitian menunjukkan bahwa bahwa algoritma *Extreme Gradient Boosting* (XGBoost) mampu mengklasifikasikan sentimen publik terkait pelemahan rupiah terhadap dolar AS dengan kinerja konsisten pada dua *platform* media sosial, X dan Instagram. Hasil pengujian menunjukkan akurasi mencapai 74%–76% pada data X dan stabil 77% pada data Instagram, dengan *precision* 0,76%–0,84%, *recall* 0,86%–0,88%, dan *F1-score* 0,82%–0,86%. Model dibangun menggunakan parameter 100 *estimators*, *learning rate* 0,1, *max depth* 6, dan *subsample* 0,8 yang terbukti efektif menjaga stabilitas klasifikasi. Distribusi sentimen menunjukkan dominasi opini negatif pada X (38%) dan dominasi opini positif pada Instagram (41%). Visualisasi *word cloud* dan diagram batang mengungkap bahwa percakapan di X didominasi istilah yang informatif dan relevan langsung dengan isu ekonomi seperti rupiah, lemah, dan dollar, mencerminkan diskusi serius dan analitis. Sebaliknya, pada Instagram kata-kata yang muncul lebih bersifat emosional dan ekspresif seperti ya, menangis, dan tertawa, yang menandakan respons publik lebih reaktif dibanding analitis. Temuan ini menegaskan efektivitas XGBoost dalam klasifikasi sentimen sekaligus memperlihatkan perbedaan pola opini publik antar *platform*. Namun demikian, penelitian ini masih memiliki keterbatasan berupa distribusi data yang tidak seimbang, adanya komentar ambigu atau sarkastik, serta cakupan data terbatas pada dua *platform* dalam periode tertentu.

DAFTAR PUSTAKA

- [1] Anindynta, F. A., dan Fuddin, M. K. 2023. "How do macroeconomic variables and financial inclusion affect financial stability in Indonesia?". *Jurnal Perspektif Pembiayaan dan Pembangunan Daerah*, 11(5), 359–370. <https://doi.org/10.22437/ppd.v11i5.25818>.
- [2] Assaidi, S. A., dan Amin, F. 2022. "Analisis Sentimen Evaluasi Pembelajaran Tatap Muka 100 Persen pada Pengguna Twitter menggunakan Metode Logistic Regression". *Jurnal Pendidikan Tambusai*, 6(2), 13217–13227.
- [3] Atmajaya, D., Febrianti, A., dkk. 2023. "Metode SVM dan Naive Bayes untuk Analisis Sentimen ChatGPT di Twitter". *The Indonesian Journal of Computer Science*, 12(4), 2173–2181. <https://doi.org/10.33022/ijcs.v12i4.3341>.
- [4] Cahyadi, R., Rama Aditya Yunanda, R. S. P., dkk. 2024. "Analisis Faktor-Faktor Melemahnya Kurs Rupiah pada Era Digital", 5(2), 412–427.
- [5] Hakim, Y. R., dan Aji, T. S. 2025. "Factors Affecting the Exchange Rate of the Indonesian Rupiah Against the United States Dollar: January 2013 - December 2023". *Formosa Journal of*

- Multidisciplinary Research, 4(5), 2043–2062. <https://doi.org/10.55927/fjmr.v4i5.175>.
- [6] Khusni, K., dan Nurviliza2, O. 2025. "Navigating Economic Turbulence: Exchange Rate Volatility And Macroeconomic Resilience In Open". Jurnal Ilmiah Ilmu Pendidikan dan sosial, 14(1), 917–932.
- [7] Larasati, F. R., Ratnawati, D. E., dkk. 2022. "Analisis Sentimen Ulasan Aplikasi Dana dengan Metode Random Forest". Jurnal Pengembangan Teknologi Informasi dan Ilmu Komputer, 6(9), 4305–4313.
- [8] Lestari, M. I., dan Anggraeni, D. 2021. "Analisis dampak sentimen masyarakat selama pandemi covid-19 terhadap kurs rupiah (Studi kasus pandemi covid-19 di Indonesia)". Jurnal EMBA, 9(1), 1–14.
- [9] Muttaqin, M. N., dan Kharisudin, I. 2021. "Analisis Sentimen Pada Ulasan Aplikasi Gojek Menggunakan Metode Support Vector Machine dan K Nearest Neighbor". UNNES Journal of Mathematics, 10(2), 22–27. diambil dari <http://journal.unnes.ac.id/sju/index.php/ujm>.
- [10] Oktafia, O., dan Nugroho, R. S. A. 2024. "Comparison of Support Vector Machine(Svm), Xgboost and Random Forest for Sentiment Analysis of Bumble App User Comments". Proxies: Jurnal Informatika, 6(1), 32–46. <https://doi.org/10.24167/proxies.v6i1.12453>.
- [11] Rininda, G., Hartami Santi, I., dkk. 2024. "Penerapan Svm Dalam Analisis Sentimen Pada Edlink Menggunakan Pengujian *Confusion Matrix*". JATI (Jurnal Mahasiswa Teknik Informatika), 7(5), 3335–3342. <https://doi.org/10.36040/jati.v7i5.7420>.
- [12] Samsul, E. M., Indriani, N., dkk. 2021. "Kebijakan Fiskal dan Moneter Indonesia serta Pengaruhnya terhadap Pasar Domestik di masa Pandemi COVID-19". Jurnal Accounting Information System (AIMS), 4(2), 46–57. <https://doi.org/10.32627/aims.v4i2.228>.
- [13] Wasifah Hanim 2024. "Analisis Fluktuasi Nilai Tukar Rupiah Terhadap Dollar AS Ditinjau Dari Faktor Internal dan Eksternal", 9(1), 7–19.
- [14] Wati, R., Ernawati, S., dkk. 2023. "Pembobotan TF-IDF Menggunakan Naïve Bayes pada Sentimen Masyarakat Mengenai Isu Kenaikan BIPIH". Jurnal Manajemen Informatika (JAMIKA), 13(1), 84–93. <https://doi.org/10.34010/jamika.v13i1.9424>.
- [15] Yulianti, S. E. H., Oni Soesanto, dkk. 2022. "Penerapan Metode Extreme Gradient Boosting (XGBOOST) pada Klasifikasi Nasabah Kartu Kredit". Journal of Mathematics: Theory and Applications, 4(1), 21–26. <https://doi.org/10.31605/jomta.v4i1.1792>.
- [16] Yulistiani, S. 2024. "Analisis Sentimen Terhadap Calon Presiden Indonesia 2024 dengan Metode Extreme Gradient Boosting (XGBOOST)". Jurnal Informatika: Jurnal pengembangan IT, 9(3), 322–328. <https://doi.org/10.30591/jpit.v9i3.6127>.