



Analisis Sentimen Ulasan Aplikasi Indodax Pada Google Play Store Dengan Algoritma Random Forest

Muhammad Iqbal Maulana, Yusra*, Muhammad Fikry, Surya Agustian, Siti Ramadhani

Fakultas Sains dan Teknologi, Teknik Informatika, Universitas Islam Negeri Sultan Syarif Kasim, Pekanbaru, Indonesia
Email: ¹iqbalmaulana411@gmail.com, ^{2,*}yusra@uin-suska.ac.id, ³muhammad.fikry@uin-suska.ac.id, ⁴surya.agustian@uin-suska.ac.id, ⁵siti.ramadhani@uin-suska.ac.id
Email Penulis Korespondensi: yusra@uin-suska.ac.id

Abstrak—Aset kripto telah menjadi fenomena global dengan peningkatan jumlah *investor* yang signifikan di Indonesia. Indodax, sebagai *platform* jual beli aset kripto terbesar di Indonesia, telah berkontribusi dalam pertumbuhan ekosistem ini dan mendapatkan banyak ulasan pengguna melalui *Google Play Store*. Dengan lebih dari 5 juta unduhan dan 100 ribu ulasan, analisis sentimen menjadi alat penting untuk memahami persepsi pengguna terhadap layanan Indodax. Hasil pelabelan manual menunjukkan mayoritas ulasan bersentimen positif (3989 ulasan), sementara sentimen netral dan negatif masing-masing berjumlah 477 dan 534 ulasan. Dari penelitian dan pengujian yang telah dilakukan dengan metode *Random Forest* dan melakukan optimasi dengan *Hyperparameter Tuning GridSearchCV* pada 4 Skenario pengujian. Didapatkan hasil terbaik pada Skenario 4 (3 Tahap *Preprocessing* (*Cleaning*, *Case Folding*, dan *Tokenization*) + *Random Forest & Hyperparameter Tuning*) menghasilkan nilai yang terbaik, dengan *Precision* 81%, *Recall* 64%, *F1-Score* 70% dan *Accuracy* 89%. Dengan nilai parameter terbaik {'*criterion*': '*entropy*', '*max_depth*': *None*, '*max_features*': '*sqrt*', '*min_samples_leaf*': 1, '*min_samples_split*': 2, '*n_estimators*': 100}. Studi ini menunjukkan bahwa setiap model eksperimen yang dilakukan Optimasi Menghasilkan nilai lebih tinggi dari model eksperimen yang tidak dilakukan optimasi.

Kata Kunci: Cryptocurrency; Indodax; Analisis Sentimen; Google Play Store; Random Forest

Abstract—Crypto assets have become a global phenomenon with a significant increase in the number of investors in Indonesia. Indodax, as the largest crypto asset trading platform in Indonesia, has contributed to the growth of this ecosystem and received many user reviews through the *Google Play Store*. With more than 5 million downloads and 100 thousand reviews, sentiment analysis is an important tool to understand user perceptions of Indodax services. The results of manual labeling show that the majority of reviews are positive (3989 reviews), while neutral and negative sentiments are 477 and 534 reviews respectively. From the research and testing that has been carried out using the *Random Forest* method and optimizing with *Hyperparameter Tuning GridSearchCV* on 4 test scenarios. The best results were obtained in Scenario 4 (3 Preprocessing Stages (*Cleaning*, *Case Folding*, and *Tokenization*) + *Random Forest & Hyperparameter Tuning*) producing the best value, with *Precision* 81%, *Recall* 64%, *F1-Score* 70% and *Accuracy* 89%. With the best parameter values {'*criterion*': '*entropy*', '*max_depth*': *None*, '*max_features*': '*sqrt*', '*min_samples_leaf*': 1, '*min_samples_split*': 2, '*n_estimators*': 100}. This study shows that every experimental model that is optimized produces a higher value than experimental model that is not optimized.

Keywords: Cryptocurrency; Indodax; Analysis Sentiment; Google Play Store; Random Forest

1. PENDAHULUAN

Dalam beberapa tahun terakhir, aset kripto telah mengalami lonjakan popularitas yang signifikan di berbagai belahan dunia, termasuk Indonesia. Aset kripto, atau *cryptocurrency*, merupakan inovasi dalam sistem keuangan digital yang menggunakan teknologi kriptografi untuk menjamin keamanan transaksi, mengatur penciptaan unit baru, dan memverifikasi transfer aset melalui mekanisme desentralisasi. Sistem ini beroperasi melalui jaringan *blockchain*, yaitu buku besar digital yang tidak dapat diubah, terdistribusi, dan terbuka bagi publik [1]. Dengan sistem ini, setiap transaksi yang dilakukan akan terekam secara permanen dan diverifikasi oleh jaringan tanpa perlu lembaga perantara seperti bank sentral. Hal ini menciptakan sistem keuangan yang lebih terbuka, transparan, dan terdesentralisasi.

Aset kripto tidak hanya menjadi alat tukar digital, tetapi juga berkembang menjadi instrumen investasi alternatif yang digemari oleh berbagai kalangan. Salah satu alasan utama popularitas aset kripto adalah potensinya dalam memberikan imbal hasil yang tinggi dalam waktu relatif singkat, meskipun tetap disertai dengan risiko volatilitas yang besar. Perkembangan ini juga sejalan dengan meningkatnya literasi keuangan digital serta kemudahan akses terhadap teknologi internet dan aplikasi keuangan berbasis *smartphone*.

Fenomena ini sangat terlihat di Indonesia. Berdasarkan laporan Kajian Stabilitas Keuangan Nomor 37 yang diterbitkan oleh Bank Indonesia, tercatat bahwa pada bulan Juni 2021 jumlah investor aset kripto di Indonesia mencapai lebih dari ±6,5 juta orang. Angka ini menunjukkan pertumbuhan yang sangat pesat, bahkan telah melampaui jumlah investor di pasar saham domestik yang berjumlah sekitar ±2,4 juta pada periode yang sama [2]. Fakta ini mengindikasikan bahwa aset kripto tidak lagi dianggap sebagai fenomena eksklusif yang hanya dipahami oleh segelintir orang, melainkan telah menjadi bagian dari arus utama dalam strategi investasi masyarakat modern.

Untuk mendukung minat masyarakat yang tinggi terhadap aset digital ini, diperlukan *platform* yang dapat memfasilitasi kegiatan transaksi kripto secara cepat, aman, dan mudah diakses. Salah satu *platform* terkemuka di Indonesia yang menyediakan layanan tersebut adalah Indodax. Indodax merupakan perusahaan yang bergerak di bidang perdagangan aset kripto dan telah terdaftar secara resmi di bawah pengawasan Badan Pengawas Perdagangan Berjangka Komoditi (BAPPEBTI) sejak 29 Januari 2020 dengan nomor izin 002/BAPPEBTI/CP-AK/01/2020 [3]. Aplikasi Indodax telah diunduh lebih dari lima juta kali melalui *Google Play Store* dan telah menerima lebih dari seratus ribu ulasan dari

penggunanya. Angka tersebut menjadi indikator bahwa aplikasi ini telah banyak digunakan oleh masyarakat dan memiliki pengaruh signifikan dalam ekosistem investasi digital di Indonesia.

Banyaknya ulasan dari pengguna ini membuka peluang besar untuk mengeksplorasi persepsi dan pengalaman pengguna terhadap layanan yang diberikan. Melalui ulasan tersebut, pengguna menyampaikan berbagai ekspresi emosional berupa pujian, kritik, saran, atau bahkan keluhan. Namun, karena data tersebut berbentuk teks tidak terstruktur dan sangat beragam, maka diperlukan metode khusus untuk mengolah dan memahami makna dari teks-teks tersebut secara sistematis. Salah satu pendekatan yang efektif untuk tujuan ini adalah analisis sentimen.

Analisis sentimen adalah proses yang bertujuan untuk mengidentifikasi dan mengelompokkan opini atau ekspresi emosional dalam sebuah teks ke dalam kategori seperti positif, netral, dan negatif. Pendekatan ini memberikan nilai strategis dalam memahami bagaimana pengguna menilai suatu aplikasi atau layanan, sehingga informasi ini dapat dijadikan dasar dalam pengambilan keputusan pengembangan produk [4]. Selain itu, analisis sentimen juga memungkinkan pengembang untuk menangkap kebutuhan pengguna yang belum tersampaikan secara eksplisit.

Terdapat berbagai metode dalam menganalisis sentimen, salah satunya adalah *Random Forest Classifier*. *Random Forest* merupakan algoritma pembelajaran mesin berbasis *ensemble learning* yang membangun beberapa pohon keputusan (*decision tree*) dari sampel data acak, kemudian menggabungkan hasil prediksi dari setiap pohon untuk menghasilkan klasifikasi yang akurat [5]. Keunggulan utama dari algoritma ini terletak pada ketahanannya terhadap *overfitting*, kemampuannya dalam menangani data dalam skala besar, dan kestabilan performanya [6].

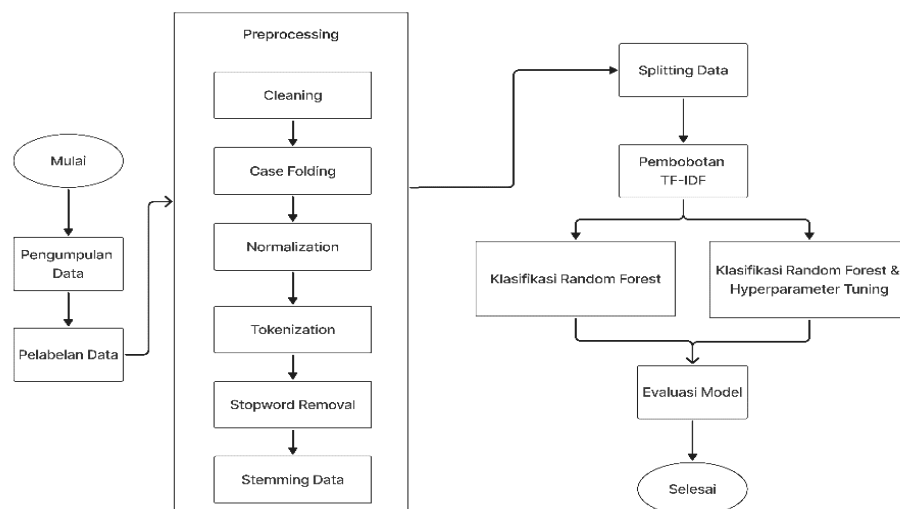
Dalam penelitian sebelumnya yang menganalisis ulasan aplikasi MyPertamina, kombinasi parameter jumlah pohon sebanyak 80 dan kedalaman pohon hingga 50 menunjukkan akurasi sebesar 90% [7]. Selain itu, ketika *Random Forest* dibandingkan dengan metode klasifikasi lain seperti *Naïve Bayes*, *Random Forest* menunjukkan performa yang lebih baik dengan akurasi mencapai 93,93%, nilai presisi 93,02%, nilai *recall* 89%, dan nilai *F1-score* sebesar 91,43% [8]. Temuan ini menunjukkan bahwa *Random Forest* sangat cocok untuk tugas klasifikasi data teks dalam konteks ulasan aplikasi.

Berdasarkan pemaparan tersebut, penelitian ini akan menerapkan pendekatan *Random Forest* untuk melakukan analisis sentimen terhadap ulasan pengguna aplikasi Indodax. Diharapkan hasil analisis ini dapat memberikan gambaran komprehensif mengenai kepuasan pengguna, fitur yang paling diapresiasi, serta aspek yang masih perlu diperbaiki dari layanan Indodax. Dengan demikian, penelitian ini tidak hanya berkontribusi terhadap pengembangan literatur analisis sentimen berbasis *machine learning*, tetapi juga memberikan manfaat praktis bagi perusahaan dalam meningkatkan kualitas dan daya saing layanannya di era digital yang kompetitif.

2. METODOLOGI PENELITIAN

2.1 Tahapan Penelitian

Studi ini bermaksud guna melaksanakan analisis sentiment pada tanggapan atau penilaian aplikasi Indodax di *Google Play Store* dengan menerapkan metode *Random Forest*. Prosedur studi ini disajikan secara visual di Gambar 1.



Gambar 1. Tahapan Penelitian

a. Pengumpulan Data

Data empiris yang diterapkan pada studi ini dikumpulkan dengan Teknik *Web Scrapping*. *Web Scrapping* ialah teknik yang diterapkan dalam rangka mendapatkan data melalui platform *web* dengan menggunakan script untuk kemudian data tersebut dapat diolah dan disimpan dalam file CSV [9]. Data akan diambil sebanyak 5000 data.

b. Pelabelan Data



Proses pelabelan data dalam studi ini dijalankan dengan manual oleh peneliti. Sebanyak 5000 data akan dibagi kedalam 3 kelas sentimen, meliputi kategori positif, netral, serta negatif. Dimana kelas positif berisi dukungan dan kepuasan terhadap aplikasi. kelas netral berisi pertanyaan, harapan dan komentar yang tidak tertuju pada aplikasi. Kelas negatif berisi protes dan penilaian buruk terhadap aplikasi.

c. *Preprocessing*

Tahap *preprocessing* proses data mencakup tahapan pra-pemrosesan data set guna memastikan data dalam kondisi siap untuk dianalisis dengan pendekatan yang dipilih. Tahapan ini bertujuan guna memperbaiki keakuratan data dan memastikan data tersedia dalam format yang sistematis dan mudah dianalisis [10]. Sebelum informasi dianalisis secara mendalam, lebih dulu dilakukan tahapan transformasi ada supaya lebih sistematis dan mempermudah tahapan analisis [11]. Berikut tahapan dalam *text preprocessing*, antara lain:

1. *Cleaning*

Cleaning merupakan tahapan pembersihan pada atribut-atribut yang tidak ada hubungannya dengan teks seperti penghilangan tanda baca, simbol, dan karakter lain yang tidak dibutuhkan dalam proses analisis [12].

2. *Case Folding*

Case folding ialah teknik mengonversi huruf dalam bentuk teks ditransformasikan ke dalam huruf kecil sehingga dapat menjaga konsistensi dalam pengolahan kata menjadi teks [13].

3. *Normalization*

Normalization merupakan proses menyesuaikan kata yang tidak sesuai kaidah diubah ke bentuk baku berdasarkan ejaan KBBI [14].

4. *Tokenization*

Tokenisasi ialah proses memisahkan sebuah kalimat menjadi bagian-bagian yang lebih kecil, yaitu kata-kata individual. Tahapan ini bertujuan untuk mengidentifikasi setiap kata dalam kalimat sebagai satuan tersendiri yang dapat dianalisis lebih lanjut [15].

5. *Stopword Removal*

Stopword Removal ialah tahapan menghapus kata-kata umum yang tidak memiliki makna signifikan dalam analisis, seperti 'dan', 'atau', dan sebagainya. Proses ini dilakukan dengan menggunakan kamus *stopword* yang disusun berdasarkan hasil penelitian dan analisis terhadap kata-kata yang sering muncul dalam dataset yang digunakan.

6. *Stemming*

Proses *stemming* bertujuan untuk mengubah kata berimbuhan menjadi bentuk dasar dengan cara menghapus komponen seperti awalan, sisipan, dan akhiran agar diperoleh kata-kata penting dalam analisis [16]. Salah satu *library* yang dapat digunakan untuk proses *stemming* bahasa Indonesia, yaitu *Library Python Sastrawi*.

d. *Splitting Data*

Splitting data ialah tahapan memisahkan data ke dalam dua subset, yaitu data uji dan data latih secara acak [17]. Proses ini bertujuan untuk mengevaluasi kinerja model yang dibuat dengan data latih, sedangkan data uji digunakan untuk melakukan validasi. Data latih adalah data yang diterapkan melalui tahap pembelajaran, sedangkan data uji adalah data yang belum pernah diterapkan melalui tahap pembelajaran. Pada penelitian ini akan menggunakan *splitting data* dengan kombinasi 80 : 20.

e. Pembobotan TF-IDF

Usai dilaksanakannya *Splitting data*, berikutnya akan dilakukan pembobotan dengan TF-IDF. TF-IDF merupakan hasil dari perhitungan antara TF (*Term Frequency*) dan IDF (*Inverse Document Frequency*). TF (*Term Frequency*) merepresentasikan intensitas kemunculan suatu istilah dalam dokumen tertentu, dan IDF (*Inverse Document Frequency*) digunakan untuk mengidentifikasi bobot pentingnya istilah tersebut dalam seluruh kumpulan dokumen [18]. Frekuensi kata dalam dokumen mencerminkan tingkat kepentingan kata tersebut dalam suatu dokumen. Sementara, frekuensi dokumen yang mengandung kata tersebut mengindikasikan seberapa umum penggunaan kata tersebut [19].

f. Klasifikasi *Random Forest*

Penelitian ini akan menerapkan algoritma *Random Forest* untuk proses klasifikasi. *Random Forest* merupakan teknik klasifikasi berbasis ensemble yang mengombinasikan beberapa pohon keputusan, yang masing-masing dibentuk dari fitur acak independen dengan karakteristik berbeda [20]. *Random forest* adalah suatu algoritma yang dipergunakan untuk klasifikasi data dalam jumlah yang besar. *Random forest* memisahkan atribut dari kelas yang digunakan guna menemukan prediksi data yang belum muncul.

g. *Hyperparameter Tuning*

Penyetelan hiperparameter dilakukan dengan mencari kombinasi *hyperparameter* yang paling sesuai. Metodenya, hiperparameter diuji melalui proses coba-coba hingga diperoleh nilai yang paling optimal [21]. Studi ini memanfaatkan teknik *GridSearch* yang disediakan oleh *Scikit-learn* guna menentukan *hyperparameter* terbaik. *GridSearchCV* merupakan metode penyetelan *hyperparameter* yang memungkinkan dilakukannya pemindaian terhadap sejumlah opsi *hyperparameter* yang telah ditentukan.

h. Evaluasi

Pengukuran dalam studi ini akan memanfaatkan matriks kebingungan. Matriks kebingungan merupakan sebuah tabel yang memuat nilai tebakan tepat dan tebakan keliru dari suatu model pengelompokan. Dalam penilaian ketepatan hasil penelusuran, evaluasi dilakukan dengan memakai metrik *recall*, presisi, skor-f1, dan akurasi [6].



2.2 Metode Analisis Data

Penelitian ini ditujukan guna menganalisis data ulasan aplikasi Indodax Pada *Google Play Store* dengan menerapkan algoritma *random forest*. Algoritma *random forest* merupakan klasifikasi yang membuahkan lebih dari satu pohon keputusan (*decision tree*) menerapkan subset sampel serta *variable* data latih yang dipilih secara acak yang dapat meningkatkan akurasi[22]. Algoritma tersebut dinilai efektif karena mampu menghasilkan akurasi yang tinggi, menangani missing data dengan baik, memiliki toleransi terhadap data pencilan, serta hemat dalam manajemen penyimpanan [13]. Prosedur atau mekanisme kerja algoritma *Random Forest* dapat dijelaskan sebagai berikut:

- Langkah awal dalam penerapan teknik *bootstrap* yakni melaksanakan pemilihan contoh secara acak berjumlah n dari informasi yang ada
- Menerapkan informasi hasil *bootstrap*, dibuatlah pohon keputusan dengan himpunan sebagian atribut yang dipilih secara acak untuk melaksanakan pelatihan.
- Melaksanakan ulang tahap 1 serta 2 sejumlah k kali, di mana k merupakan total pohon keputusan yang akan dibentuk
- Output perkiraan dari semua pohon keputusan akan dihimpun dan dilakukan pemungutan suara terbanyak untuk memperoleh keluaran akhir

2.3 Metode Optimasi

Penelitian ini akan menggunakan metode *hyperparameter tuning* untuk melakukan optimasi terhadap klasifikasi *random forest*. *Hyperparameter tuning* merupakan proses penyesuaian nilai-nilai parameter dalam model *random forest*. Salah satu metode untuk parameter tuning yang optimal adalah menggunakan algoritma *grid search*. Seperti namanya, *Grid Search* akan mencari parameter "*Grid*" dan *Cross Validation (CV)* untuk mengevaluasi kinerja. *Grid Search Cross Validation* adalah metode yang secara otomatis menguji dan memvalidasi kombinasi parameter dan *hyperparameter* untuk menemukan model dengan performa terbaik untuk prediksi.[23].

2.3 Evaluasi

Penelitian ini akan menerapkan *Confusion Matrix* guna memperoleh informasi terkait nilai-nilai yang diprediksikan serta hasil yang aktual, biasa digunakan untuk menghitung *accuracy*, *recall*, *precision* dan *F1-score*[24]. *Accuracy*, *precision*, *recall*, dan *F1-score* adalah metrik evaluasi yang digunakan untuk menilai kinerja model klasifikasi. *Accuracy* mengukur proporsi prediksi yang benar dari seluruh prediksi yang dilakukan, namun bisa menyesatkan jika data tidak seimbang. *Precision* menunjukkan seberapa banyak prediksi positif yang benar dibandingkan dengan semua prediksi positif, berguna saat kita ingin meminimalkan kesalahan positif palsu. *Recall* mengukur seberapa banyak data positif yang berhasil dikenali dengan benar oleh model, penting ketika kita ingin menghindari kesalahan negatif palsu. Sementara itu, *F1-score* adalah rata-rata dari *precision* dan *recall*, memberikan keseimbangan antara keduanya, terutama saat distribusi kelas tidak merata. Cara kerja dari *confusion matrix* yaitu meninjau perbedaan antara hasil pengklasifikasian yang aktual dengan hasil klasifikasi dari model yang digunakan[25]. Rumus dari *Confusion Matrix* bisa diamati pada penelitian[26].

3. HASIL DAN PEMBAHASAN

3.1 Pengumpulan Data

Pengumpulan data ulasan diterapkan melalui Teknik *web scrapping*. Data diambil sebanyak 5000 data ulasan. Keluaran dari proses ini berupa file .csv yang memuat atribut nama pengguna, tanggal ulasan, penilaian, dan isi komentar. Data ulasan bisa dilihat pada Gambar 2.

	Review ID	Username	Date	Rating	Review Text
0	c4162212-8127-44b0-be0e-1a064b953e3b	Pengguna Google	5/20/2025 4:49	5	mantap
1	071d6477-aabf-41da-abb8-ef8de683d614	Pengguna Google	5/20/2025 4:42	5	lumayan
2	6f66feea-ed2a-46e0-8f0f-da65a46023c9	Pengguna Google	5/20/2025 4:40	5	sangat mudah dan gampang untuk pemula belajar ...
3	b4b73b1a-9a5a-4ab9-a420-14340b366da9	Pengguna Google	5/20/2025 4:40	2	gajelas, marketnya di setting biar kita rugi
4	f1831e07-24ae-48a2-9920-75ffe2325706	Pengguna Google	5/19/2025 13:50	1	woi kenapa coin bitcoin gold gua ga bisa di wi...
5	86d7317b-51bf-4236-89dd-ca322c8909ee	Pengguna Google	5/20/2025 4:24	5	Aplikasinya sangat bagus
6	6a5d0aa3-a9da-4516-b17e-beb778e3dd24	Pengguna Google	5/20/2025 4:18	5	apk bagus Gampang di pahami dan invest di apk ...

Gambar 2. Data Ulasan



3.2 Pelabelan Data

Selanjutnya manual labeling dilakukan terhadap setiap entri ulasan untuk memastikan ketepatan kategori sentimen. Hasilnya ulasan sentimen positif sebanyak 3989 ulasan, 477 ulasan dengan sentiment netral, dan 534 ulasan negatif. Data hasil pelabelan dapat diamati pada Tabel 1.

Tabel 1. Data Hasil Pelabelan

No.	Review Text	Sentimen
1.	Aplikasi yang sangat bagus dan mudah dipahami untuk pemula yang ingin trading atau investasi crypto.	POSITIF
2.	lumayan INDODAX platform terbaik dan terpercaya yang saya jumpai	POSITIF
3.	Bismillah dapat hadiah 3.000.000 wkwwk ngarep banget njir gw	NETRAL
4.	Virtual account top up cuman ada permata tambah bank lain dong.. transfer bank juga cuman BCA, pas top up pake mandiri sama BSI gagal terus udah error selama sebulan	NEGATIF
5.	Kenapa order jualnya lama sekali. Bikin kecewa dan emosi karena tidak real time.	NEGATIF

3.3 Preprocessing

Setelah dilakukan pelabelan, selanjutnya data akan dilakukan proses *preprocessing* agar didapatkan data yang bersih dan terstruktur, proses *preprocessing* adalah:

- a. *Cleaning*: tahapan penghilangan frasa yang tidak konsektual misalnya *URL*, *hashtag*, dan simbol. Hasil *Cleaning* bisa diamati pada Tabel 2.

Tabel 2. *Cleaning*

data	Cleaning
1.	Aplikasi yang sangat bagus dan mudah dipahami untuk pemula yang ingin trading atau investasi crypto
2.	INDODAX platform terbaik dan terpercaya yang saya jumpai
3.	Bismillah dapat hadiah wkwwk ngarep banget njir gw
4.	Virtual account top up cuman ada permata tambah bank lain dong transfer bank juga cuman BCA pas top up pake mandiri sama BSI gagal terus udah error selama sebulan
5.	Kenapa order jualnya lama sekali Bikin kecewa dan emosi karena tidak real time

- b. *Case Folding*: proses menyamakan format dari setiap huruf pada setiap kata. Hasil *Case Folding* bisa diamati pada Tabel 3.

Tabel 3. *Case Folding*

data	Case Folding
1.	aplikasi yang sangat bagus dan mudah dipahami untuk pemula yang ingin trading atau investasi crypto
2.	indodax platform terbaik dan terpercaya yang saya jumpai
3.	bismillah dapat hadiah wkwwk ngarep banget njir gw
4.	virtual account top up cuman ada permata tambah bank lain dong transfer bank juga cuman bca pas top up pake mandiri sama bsi gagal terus udah error selama sebulan
5.	kenapa order jualnya lama sekali bikin kecewa dan emosi karena tidak real time

- c. *Normalization*: Proses menyesuaikan frasa informal dikonversi ke format standar. Hasil *Normalization* bisa diamati Pada Tabel 4.

Tabel 4. *Normalization*

data	Normalization
1.	aplikasi yang sangat bagus dan mudah dipahami untuk pemula yang ingin trading atau investasi crypto
2.	indodax platform terbaik dan terpercaya yang saya jumpai
3.	bismillah dapat hadiah wkwwk mengharap banget anjir gue
4.	virtual account top up cuman ada permata tambah bank lain dong transfer bank juga cuman baca pas top up pakai mandiri sama bsi gagal terus sudah error selama sebulan
5.	kenapa order jualnya lama sekali bikin kecewa dan emosi karena tidak real time

- d. *Tokenization*: Proses memisahkan kalimat menjadi kumpulan kata tersendiri. Hasil *Tokenization* bisa diamati Pada Tabel 5.

Tabel 5. *Tokenization*

data	Tokenization
1.	['aplikasi', 'yang', 'sangat', 'bagus', 'dan', 'mudah', 'dipahami', 'untuk', 'pemula', 'yang', 'ingin', 'trading', 'atau', 'investasi', 'crypto']



data	Tokenization
2.	['indodax', 'platform', 'terbaik', 'dan', 'terpercaya', 'yang', 'saya', 'jumpai']
3.	['bismillah', 'dapat', 'hadiah', 'wkwwk', 'mengharap', 'banget', 'anjir', 'gue']
4.	['virtual', 'account', 'top', 'up', 'cuman', 'ada', 'permata', 'tambah', 'bank', 'lain', 'dong', 'transfer', 'bank', 'juga', 'cuman', 'baca', 'pas', 'top', 'up', 'pakai', 'mandiri', 'sama', 'bsi', 'gagal', 'terus', 'sudah', 'error', 'selama', 'sebulan']
5.	['kenapa', 'order', 'jualnya', 'lama', 'sekali', 'bikin', 'kecewa', 'dan', 'emosi', 'karena', 'tidak', 'real', 'time']

- e. *Stopword Removal*: Tahapan menghilangkan istilah konjungsi pada frasa, seperti “atau”, “dan”, “sebuah”. Hasil *Stopword Removal* bisa diamati Pada Tabel 6.

Tabel 6. *Stopword Removal*

data	Stopword Removal
1.	['aplikasi', 'bagus', 'mudah', 'dipahami', 'pemula', 'trading', 'investasi', 'crypto']
2.	['indodax', 'platform', 'terbaik', 'terpercaya', 'jumpai']
3.	['bismillah', 'hadiah', 'wkwwk', 'mengharap', 'banget', 'anjir', 'gue']
4.	['virtual', 'account', 'top', 'up', 'cuman', 'permata', 'bank', 'transfer', 'bank', 'cuman', 'baca', 'pas', 'top', 'up', 'pakai', 'mandiri', 'bsi', 'gagal', 'error', 'sebulan']
5.	['order', 'jualnya', 'bikin', 'kecewa', 'emosi', 'real', 'time']

- f. *Stemming*: ialah tahapan bertujuan guna mengkonversi kata berimbuhan ke bentuk dasar. Tabel 7 menampilkan hasil dari proses tersebut.

Tabel 7. *Stemming*

data	Stemming
1.	aplikasi bagus mudah paham mula trading investasi crypto
2.	indodax platform baik percaya jumpa
3.	bismillah hadiah wkwwk harap banget anjir gue
4.	virtual account top up cuman permata bank transfer bank cuman baca pas top up pakai mandiri bsi gagal error bulan
5.	order jual bikin kecewa emosi real time

3.4 Splitting Data

Setelah didapatkan data hasil *preprocessing*, selanjutnya dilakukan proses *splitting data*, yaitu memisahkan data menjadi dua bagian data, yaitu data uji dan data latih. Penelitian ini menggunakan kombinasi 80:20. Informasi mengenai proporsi data latih dan data uji bisa ditemukan pada Tabel 8.

Tabel 8. Jumlah Data Latih dan Data Uji

Data latih	Data uji
4000	1000

3.5 Pembobotan TF-IDF

Selanjutnya menerapkan pembobotan kata melalui TF-IDF. Dari proses TF-IDF yang telah dilakukan. Pada saat meng-*generate vector* TF-IDF untuk input *tweet*, parameter yang dapat dipilih melalui *library TfidfVectorizer*, antara lain adalah jumlah minimal dokumen yang mengandung token (*min_df*), jumlah maksimal dokumen yang mengandung token (*max_df*), jumlah *n-gram* yang dibentuk (*ngram_range*), jumlah fitur maksimal (*max_feature*). Dalam penelitian ini, digunakan parameter *default*, yaitu *min_df* = 1, *max_df* = 1, *ngram_range* = (1,1) dan *max_feature* = 'none' (berapa pun jumlah fitur yang terbentuk dari proses tokenisasi terhadap data train).

3.6 Eksperimen Set Up

Dalam pengujian, dilakukan 4 skenario pengujian yaitu:

- Skenario 1, data dengan 6 tahap *Preprocessing* (*Cleaning*, *Case Folding*, *Normalization*, *Tokenization*, *Stopword Removal*, serta *Stemming*) lalu diterapkan Klasifikasi *Random Forest*.
- Skenario 2, data dengan 3 tahap *Preprocessing* (*Cleaning*, *Case Folding*, dan *Tokenization*) lalu dilakukan Klasifikasi *Random Forest*.
- Skenario 3, data dengan 6 tahap *Preprocessing* (*Cleaning*, *Case Folding*, *Normalization*, *Tokenization*, *Stopword Removal*, serta *Stemming*) lalu dilakukan Klasifikasi *Random Forest* dan *Hyperparameter Tuning*.
- Skenario 4, data dengan 3 tahap *Preprocessing* (*Cleaning*, *Case Folding*, dan *Tokenization*) lalu dilakukan Klasifikasi *Random Forest* dan *Hyperparameter Tuning*.



3.7 Klasifikasi *Random Forest*

Setelah melakukan pembobotan TF-IDF, proses selanjutnya adalah klasifikasi dengan *random forest*. Untuk hasil klasifikasi *Random Forest* pada Skenario 1, data yang telah lewat 6 tahap *Preprocessing* (*Cleaning*, *Case Folding*, *Normalization*, *Tokenization*, *Stopword Removal*, dan *Stemming*) dan pada Skenario 2, data dengan 3 tahap *preprocessing* (*Cleaning*, *Case Folding*, dan *Tokenization*) dapat dilihat pada Tabel 9.

Tabel 9. Hasil Klasifikasi *Random Forest*

Data	Precision	Recall	F1-Score	Accuracy
Skenario 1 (6 Tahap <i>Preprocessing</i>)	0.73	0.60	0.65	0.87
Skenario 2 (3 Tahap <i>Preprocessing</i>)	0.79	0.64	0.68	0.88

3.8 *Hyperparameter Tuning*

Studi ini menentukan nilai *hyperparameter tuning* yang akan digunakan. Nilai *hyperparameter* dapat dilihat pada Tabel 10.

Tabel 10. *Hyperparameter Set*

Hyperparameter	Nilai
<i>Criterion</i>	<i>Gini</i> , <i>Entropy</i> , <i>Log_Loss</i>
<i>Max_Depth</i>	<i>None</i> , 150, 300
<i>N_Estimator</i>	100, 150, 300
<i>Max_Feature</i>	<i>Sqrt</i> , <i>log2</i>
<i>Min_Samples_Split</i>	2, 4, 6
<i>Min_Samples_Leaf</i>	1, 2

Selanjutnya dilakukan *Hyperparameter Tuning* untuk Skenario 3, data yang telah lewat 6 tahap *Preprocessing* (*Cleaning*, *Case Folding*, *Normalization*, *Tokenization*, *Stopword Removal*, dan *Stemming*) + *Random Forest* dan Skenario 4, data dengan 3 tahap *preprocessing* (*Cleaning*, *Case Folding*, dan *Tokenization*) + *Random Forest*. Didapatkan hasil parameter terbaik untuk Skenario 3 yaitu {'*criterion*': '*entropy*', '*max_depth*': *None*, '*max_features*': '*sqrt*', '*min_samples_leaf*': 1, '*min_samples_split*': 2, '*n_estimators*': 200}. Dan hasil parameter terbaik untuk Skenario 4 yaitu {'*criterion*': '*entropy*', '*max_depth*': *None*, '*max_features*': '*sqrt*', '*min_samples_leaf*': 1, '*min_samples_split*': 2, '*n_estimators*': 100}. Hasil optimasi dapat dilihat pada Tabel 11.

Tabel 11. Hasil Optimasi dengan *Hyperparameter Tuning*

Data	Precision	Recall	F1-Score	Accuracy
Skenario 3 (6 Tahap <i>Preprocessing</i> + <i>Random Forest</i>)	0.76	0.63	0.68	0.88
Skenario 4 (3 Tahap <i>Preprocessing</i> + <i>Random Forest</i>)	0.81	0.64	0.70	0.89

3.9 Evaluasi

Selanjutnya melakukan perbandingan dari seluruh Skenario yang telah dilakukan Pengujian. Untuk mengetahui skenario dengan tingkat *accuracy* dan *F1-Score* yang tertinggi. Perbandingan hasil dapat dilihat pada Tabel 12.

Tabel 12. Perbandingan Seluruh Skenario

Data	Precision	Recall	F1-Score	Accuracy
Skenario 1 (6 Tahap <i>Preprocessing</i> + <i>Random Forest</i>)	0.73	0.60	0.65	0.87
Skenario 2 (3 Tahap <i>Preprocessing</i> + <i>Random Forest</i>)	0.79	0.64	0.68	0.88
Skenario 3 (6 Tahap <i>Preprocessing</i> + <i>Random Forest</i> & <i>Hyperparameter Tuning</i>)	0.76	0.63	0.68	0.88
Skenario 4 (3 Tahap <i>Preprocessing</i> + <i>Random Forest</i> & <i>Hyperparameter Tuning</i>)	0.81	0.64	0.70	0.89

Dari Data pada Tabel 11 menunjukkan bahwa data pada Skenario 4 (3 Tahap *Preprocessing* (*Cleaning*, *Case Folding*, dan *Tokenization*) + *Random Forest* & *Hyperparameter Tuning*) menghasilkan nilai yang terbaik, dengan *Precision* 81%, *Recall* 64%, *F1-Score* 70% dan *Accuracy* 89%. Dari tabel juga dapat dilihat bahwa setiap eksperimen model yang dilakukan Optimasi Menghasilkan nilai lebih tinggi dari eksperimen model yang tidak dilakukan optimasi.



Peningkatan nilai dari eksperimen model sebelum dan sesudah optimasi yaitu *f1-Score* mengalami peningkatan sebanyak 0.02 dan *accuracy* sebanyak 0.01.

4. KESIMPULAN

Studi ini menerapkan algoritma *Random Forest* guna menerapkan analisis sentimen pada penilaian aplikasi Indodax pada *Google Play Store*. Hasil pelabelan dilakukan secara manual pada 5000 data ulasan, didapatkan sentimen positif sebanyak 3989 ulasan, sentiment netral sebanyak 477 ulasan, dan sentiment negatif sebanyak 534 ulasan. Dari penelitian dan pengujian yang telah dilakukan dengan metode *Random Forest* dan melakukan optimasi dengan *Hyperparameter Tuning GridSearchCV* pada 4 Skenario pengujian. Didapatkan hasil terbaik pada Skenario 4 (3 Tahap *Preprocessing* (*Cleaning*, *Case Folding*, dan *Tokenization*) + *Random Forest & Hyperparameter Tuning*) menghasilkan nilai yang terbaik, dengan *Precision* 81%, *Recall* 64%, *F1-Score* 70% dan *Accuracy* 89%. Dengan nilai parameter terbaik {'*criterion*': '*entropy*', '*max_depth*': *None*, '*max_features*': '*sqrt*', '*min_samples_leaf*': 1, '*min_samples_split*': 2, '*n_estimators*': 100}. Dapat dilihat bahwa setiap eksperimen model yang dilakukan Optimasi Menghasilkan nilai lebih tinggi dari eksperimen model yang tidak dilakukan optimasi. Peningkatan nilai dari eksperimen model sebelum dan sesudah optimasi yaitu *f1-Score* mengalami peningkatan sebanyak 0.02 dan *accuracy* sebanyak 0.01. Penelitian ini hanya membandingkan metode *Random Forest* dan *Random Forest* yang dilakukan optimasi dengan *Hyperparameter tuning*. Untuk penelitian selanjutnya diharapkan dapat melakukan perbandingan dengan metode klasifikasi lainnya.

REFERENCES

- [1] M. Arbina and M. I. F. Putuhena, "Tata Kelola Pembentukan Regulasi Terkait Perdagangan Mata Uang Kripto (Cryptocurrency) sebagai Aset Kripto (Crypto Asset)," *Mahadi Indones. J. Law*, vol. 1, no. 1, pp. 33–57, 2022, doi: 10.32734/mah.v1i1.8314.
- [2] M. D. Firgayanda and H. M. Jumhur, "Analisis Pengaruh Financial Literacy Terhadap Keputusan Investasi Di Pasar Cryptocurrency (Studi Pada Pengguna Platform Indodax)," in *e-Proceeding of Management*, 2022, pp. 1768–1771.
- [3] A. Kusuma, E. Ermatita, and H. N. Irmada, "Analisis Sentimen Pada Ulasan Aplikasi Indodax di Google Play Store Menggunakan Metode Support Vector Machine," in *Prosiding Seminar Nasional Mahasiswa Bidang Ilmu Komputer dan Aplikasinya*, 2022, pp. 773–782.
- [4] N. Z. Rania and R. D. Syah, "Analisis Sentimen Terhadap Aplikasi Gojek Pada Play Store Menggunakan Metode Random Forest Classifier," *J. Ilm. Inform. Komput.*, vol. 29, no. 2, pp. 144–153, 2024, doi: 10.35760/ik.2024.v29i2.11877.
- [5] M. R. U. Pulungan, D. E. Ratnawati, and B. Rahayudi, "Analisis Sentimen Ulasan Aplikasi PeduliLindungi dengan Metode Random Forest," *J. Pengemb. Teknol. Inf. dan Ilmu Komun.*, vol. 6, no. 9, pp. 4378–4385, 2022, [Online]. Available: <http://j-ptiik.ub.ac.id/index.php/j-ptiik/article/download/11582/5142>
- [6] V. Fransisco and D. B. Rarasati, "Analisis Sentimen Aplikasi Polri Super App Menggunakan Algoritma Random Forest," *J. Ilm. Sains dan Teknol.*, vol. 8, no. 2, pp. 183–195, 2024, doi: 10.47080/saintek.v8i2.3383.
- [7] C. G. Indrayanto, D. E. Ratnawati, and B. Rahayudi, "Analisis Sentimen Data Ulasan Pengguna Aplikasi MyPertamina di Indonesia pada Google Play Store menggunakan Metode Random Forest," *J. Pengemb. Teknol. Inf. dan Ilmu Komput.*, vol. 7, no. 3, pp. 1131–1139, 2023, [Online]. Available: <http://j-ptiik.ub.ac.id>
- [8] A. Miftahusalam, H. Pratiwi, and I. Slamet, "Perbandingan Metode Random Forest dan Naive Bayes pada Analisis Sentimen Review Aplikasi BCA Mobile," in *SIPTEK : Seminar Nasional Inovasi dan Pengembangan Teknologi Pendidikan*, 2023, pp. 1–8.
- [9] D. Reinaldo and Y. F. Riti, "Analisis Sentimen Ulasan Aplikasi Genshin Impact di Play Store Menggunakan Random Forest," in *Seminar Teknik Elektro, Sistem Informasi, dan Teknik Informatika*, 2024, pp. 578–586. doi: 10.31284/p.snestik.2024.5883.
- [10] T. A. U. Azmi, L. Hakim, D. C. R. Novitasari, and W. D. U. D. Utami, "Application Random Forest Method for Sentiment Analysis in Jamsostek Mobile Review," *Telemat. J. Inform. dan Teknol. Inf.*, vol. 20, no. 1, p. 117, 2023, doi: 10.31315/telematika.v20i1.8868.
- [11] S. Nur Adhan, G. N. A. Wibawa, D. C. Arisona, I. Yahya, and R. Ruslan, "Analisis Sentimen Ulasan Aplikasi Wattpad Di Google Play Store Dengan Metode Random Forest," *AnoatiK J. Teknol. Inf. dan Komput.*, vol. 2, no. 1, pp. 6–15, 2024, doi: 10.33772/anoatik.v2i1.32.
- [12] D. N. I. Huda, C. Prianto, and R. M. Awangga, "Analisis Sentimen Perbandingan Layanan Jasa Pengiriman Kurir Pada Ulasan Play Store Menggunakan Metode Random Forest dan Descision Tree," *J. Ilm. Inform.*, vol. 11, no. 2, pp. 150–158, 2023, doi: 10.33884/jif.v11i02.7952.
- [13] M. F. Y. Herjanto and C. Carudin, "Analisis Sentimen Ulasan Pengguna Aplikasi Sirekap Pada Play Store Menggunakan Algoritma Random Forest Classifier," *J. Inform. dan Tek. Elektro Terap.*, vol. 12, no. 2, pp. 1204–1210, 2024, doi: 10.23960/jitet.v12i2.4192.
- [14] O. Manullang, C. Prianto, and N. H. Harani, "Analisis Sentimen Untuk Memprediksi Hasil Calon Pemilu Presiden Menggunakan Lexicon Based Dan Random Forest," *J. Ilm. Inform.*, vol. 11, no. 2, pp. 159–169, 2023, doi: 10.33884/jif.v11i02.7987.
- [15] Tangguh Prana Welas Sukma and M. R. Pribadi, "Analisis Sentimen Review Pengguna Viu Pada Play Store Dengan Algoritma Random Forest," *J. Softw. Eng. Comput. Intell.*, vol. 2, no. 01, pp. 9–16, 2024, doi: 10.36982/jseci.v2i01.4016.
- [16] N. C. Ramadani, "Analisis Sentimen Untuk Mengukur Ulasan Pengguna Aplikasi Mobile Legend Menggunakan Algoritma Naive Bayes, SVM, Random Fores, Decision Tree, dan Logistic Regression," *J. Sist. Inf.*, vol. 16, no. 1, pp. 123–138, 2024, doi: 10.18495/jsi.v16i1.152.
- [17] M. A. A. Jihad, Adiwijaya, and W. Astuti, "Analisis Sentimen Terhadap Ulasan Film Menggunakan Algoritma Random Forest," in *eProceedings of Engineering*, 2021, pp. 10153–10166.
- [18] F. A. Larasati, D. E. Ratnawati, and B. T. Hanggara, "Analisis Sentimen Ulasan Aplikasi Dana dengan Metode Random Forest," *J. Pengemb. Teknol. Inf. dan Ilmu Komput.*, vol. 6, no. 9, pp. 4305–4313, 2022.



- [19] F. M. Rayhan *et al.*, “Analisis Sentimen Root Cause Analisis Kepuasan,” *J. Pengemb. Teknol. Inf. dan Ilmu Komput.*, vol. 8, no. 8, pp. 1–10, 2024.
- [20] R. Aryanti, T. Misriati, and A. Sagiyanto, “Analisis Sentimen Aplikasi Primaku Menggunakan Algoritma Random Forest dan SMOTE untuk Mengatasi Ketidakseimbangan Data,” *J. Comput. Syst. Informatics*, vol. 5, no. 1, pp. 218–227, 2023, doi: 10.47065/josyc.v5i1.4562.
- [21] I. Muhamad Malik Matin, “Hyperparameter Tuning Menggunakan GridsearchCV pada Random Forest untuk Deteksi Malware,” *J. Multinetics*, vol. 9, no. 1, pp. 43–50, 2023, doi: 10.32722/multinetics.v9i1.5578.
- [22] H. Chyntia Morama, D. E. Ratnawati, and I. Arwani, “Analisis Sentimen berbasis Aspek terhadap Ulasan Hotel Tentrem Yogyakarta menggunakan Algoritma Random Forest Classifier,” *J. Pengemb. Teknol. Inf. dan Ilmu Komput.*, vol. 6, no. 4, pp. 1702–1708, 2022.
- [23] A. A. Putri, S. Agustian, R. Abdillah, and Pizaini, “Penerapan Metode Logistic Regression Untuk Klasifikasi Sentimen Pada Dataset Twitter Terbatas,” *J. Sist. Inf.*, vol. 7, no. 1, pp. 95–107, 2025, doi: 10.31849/zn.v7i1.24804.
- [24] S. Nanda, D. Mualfah, and D. A. Fitri, “Analisis Sentimen Kepuasan Pengguna Terhadap Layanan Streaming Mola Menggunakan Algoritma Random Forest,” *J. Apl. Teknol. Inf. dan Manaj.*, vol. 3, no. 2, pp. 210–219, 2022, doi: 10.31102/jatim.v3i2.1592.
- [25] Y. R. Nasution, A. H. Lubis, and T. F. Eliza, “Analisis Sentimen Platform X Mengenai Pro Kontra Rekrutmen Guru Melalui Marketplace Menggunakan Metode Naïve Bayes,” *Bull. Comput. Sci. Res.*, vol. 4, no. 6, pp. 406–713, 2024, doi: 10.47065/bulletincsr.v4i6.368.
- [26] P. Yohana, S. Agustian, and S. K. Gusti, “Klasifikasi Sentimen Masyarakat terhadap Kebijakan Vaksin Covid-19 pada Twitter dengan Imbalance Classes Menggunakan Naive Bayes,” in *Seminar Nasional Teknologi Informasi, Komunikasi dan Industri*, 2022, pp. 69–80. [Online]. Available: <http://ejournal.uin-suska.ac.id/index.php/SNTIKI/article/view/19012>