

# ANALISIS SENTIMEN PADA ULASAN APLIKASI TOKOCRYPTO DENGAN MENGGUNAKAN METODE SUPPORT VECTOR MACHINE (SVM) PADA GOOGLE PLAY

<sup>1</sup>Dini Sumartini, <sup>2</sup>Lu'lu Mawaddah Wisudawati\*

<sup>1</sup>Fakultas Ilmu Komputer dan Teknologi Informasi Universitas Gunadarma, <sup>2</sup>Fakultas Teknologi Industri Universitas Gunadarma\*

Jl. Margonda Raya No. 100, Depok 16424, Jawa Barat

<sup>1</sup>dinismr@student.gunadarma.ac.id, <sup>2</sup>lulu\_mawadah@staff.gunadarma.ac.id\*

\*) Penulis Korespondensi

## Abstrak

Cryptocurrency adalah mata uang virtual yang digunakan sebagai mata uang alternatif, di mana mata uang tersebut dihasilkan dan diperdagangkan melalui proses kriptografi. Salah satu Aplikasi mata uang digital Cryptocurrency yaitu Tokocrypto, alat pembayaran yang sah hanya menggunakan rupiah. Metode klasifikasi sentimen pada penulisan ini menggunakan Support Vector Machine (SVM) dan parameter diuji secara ekstensif dengan menerapkan Teknik K-fold cross validation untuk menemukan konfigurasi yang optimal. Tahapan awal pada analisis sentimen ini yaitu tahap pengambilan data, pre-processing, yang terdiri dari case folding & cleaning, filtering (stopword removal), tokenizing, mengubah kata bernegasi, normalization dan stemming. Setelah itu, dilakukan pelabelan data dan pemberian skor menggunakan metode Lexicon Based. Dataset hasil dari pre-processing dan Lexicon Based digunakan untuk proses klasifikasi menggunakan SVM. Hasil Evaluasi terbaik didapatkan dengan AUC (Area Under The Curve) sebesar 90.50% dan Akurasi 85.80% dengan kernel Linear dan parameter C=1 dan menggunakan nilai cross validation K-fold 10. Visualisasi hasil analisis sentimen terhadap Aplikasi Tokocrypto ditampilkan dalam diagram batang dan wordcloud.

**Kata Kunci:** Analisis Sentimen, AUC, Lexicon Based, Support Vector Machine, Tokocrypto

## Abstract

Cryptocurrency is a virtual currency that is used as an alternative currency, where the currency is generated and traded through a cryptographic process. One of the Cryptocurrency digital currency applications is Tokocrypto, a legal means of payment using only rupiah. The sentiment classification method in this study uses Support Vector Machine (SVM) and the parameters are tested extensively by applying the K-fold cross validation technique to find the optimal configuration. The initial stages in this sentiment analysis are the data collection stage, pre-processing, which consists of case folding & cleaning, filtering (stopword removal), tokenizing, changing negated words, normalization and stemming. After that, data labeling and scoring are carried out using the Lexicon Based method. The dataset from pre-processing and Lexicon Based is used for the classification process using SVM. The best evaluation results were obtained with an AUC (Area Under the Curve) of 90.50% and an Accuracy of 85.80% with Linear kernel and parameter  $C = 1$  and using K-fold cross validation value of 10. Visualization of the results of the sentiment analysis of the Tokocrypto Application is displayed in a bar chart and word cloud.

**Keywords:** Sentimen Analysis, AUC, Lexicon Based, Support Vector Machine, Tokocrypto

## PENDAHULUAN

Dalam era digital saat ini, aplikasi trading berbasis teknologi *blockchain* seperti Tokocrypto telah menjadi sangat populer di kalangan investor dan pengguna *cryptocurrency*. Berdasarkan data Badan Pengawas Perdagangan Berjangka Komoditi (Bappebti) tahun 2023 mencatat adanya 18.51 juta investor aset kripto di Indonesia, meningkat 9.8% sejak awal tahun [1]. Banyak orang yang tertarik untuk berinvestasi dalam bentuk aset kripto. Selain itu, banyaknya bermunculan aplikasi kripto yang membuat para *trader* atau investor pemula harus memilih salah satu aplikasi kripto yang terbaik yang memberikan banyak fitur, kemudahan dalam penggunaannya, hingga potongan komisi yang ditawarkan dalam setiap transaksi di bursa kripto. Salah satu aplikasi trading kripto yaitu Tokocrypto.

Tokocrypto adalah sebuah *startup* yang bergerak di bidang *marketplace* yang menyediakan layanan masyarakat untuk melakukan jual beli *cryptocurrency*. Tokocrypto merupakan pedagang fisik aset kripto di Indonesia yang menyediakan aplikasi untuk melakukan transaksi jual beli aset kripto seperti Bitcoin, Ethereum, USDT (USD Tether) dan altcoin lainnya. Tokocrypto terdaftar resmi di Bappebti sejak tahun 2019 [2]. Dengan bertambahnya jumlah perusahaan rintisan *fintech*, Tokocrypto harus terus meningkatkan kualitas produk dan layanannya. Untuk unggul dalam persaingan,

perlu dilakukan analisis terhadap aplikasi Tokocrypto agar perusahaan memahami posisi dari proposisi nilai yang telah ditawarkan. Analisis ini memerlukan proposisi nilai yang unik dan inovatif sehingga aplikasi Tokocrypto dapat bersaing dengan perusahaan lain yang sejenis dan meningkatkan kesadaran merek (*brand awareness*) di kalangan pelanggan [3].

Analisis sentimen adalah teknik *Natural Language Processing (NLP)* yang digunakan untuk menentukan sentimen yang mendasarinya positif, negatif, atau netral. Informasi subjektif dari teks dapat diekstraksi menggunakan analisis sentimen dengan mengenali konteks dan posisinya. Data dari berbagai sumber, seperti komentar jejaring sosial, berita, ulasan konsumen, dan lainnya dapat digunakan untuk analisis sentimen [4]. Sentimen analisis dapat diterapkan pada proses klasifikasi dan peringkasan teks, dan berbagai metode yang dapat digunakan untuk mengekstrak sentimen dari teks, termasuk algoritma pembelajaran mesin (*machine learning*) dan sistem berbasis aturan (*rule-based systems*). Analisis sentimen adalah alat penting bagi bisnis untuk memahami umpan balik pelanggan dan mengukur kepuasan dan loyalitas pelanggan [5].

Analisis sentimen dapat membantu Tokocrypto dalam meningkatkan kualitas layanan, meningkatkan kepuasan pengguna, dan meningkatkan reputasi. Selain itu, analisis sentimen pada ulasan aplikasi Tokocrypto juga dapat digunakan untuk memantau tren persepsi

pengguna dari waktu ke waktu. Terdapat beberapa algoritma yang dikembangkan dalam melakukan analisis sentimen seperti *Support Vector Machine* (SVM), *Naive Bayes*, *Maximum Entropy*, *Decision Tree*, *K-Nearest Neighbor* (K-NN) dan lainnya [6].

SVM adalah salah satu algoritma pembelajaran terawasi (*supervised learning*) paling populer yang diterapkan untuk analisis sentimen [7]. SVM telah menunjukkan akurasi yang luar biasa dalam banyak proses NLP termasuk analisis sentimen. Terdapat peneliti-peneliti sebelumnya yang menggunakan SVM dalam melakukan analisis. Peneliti Siti Aisah dkk. [8], melakukan penelitian mengenai analisis sentimen pendekatan *machine learning* menggunakan algoritma SVM pada ulasan pengguna Al-Quran, model yang dikembangkan dievaluasi dan diperoleh model terbaik dengan nilai akurasi sebesar 85,11%, tingkat presisi sebesar 88,61%, dan tingkat *recall* sebesar 83,44%. Peneliti selanjutnya Isnain dkk, melakukan pengembangan algoritma SVM untuk mengklasifikasi kebijakan *lockdown* pemerintah Jakarta dengan *tweet* berbahasa Indonesia menggunakan ekstraksi fitur *Term Frequency-Inverse Document Frequency* (*TF-IDF*) dan didapat bahwa penggunaan metode *TF-IDF* dalam ekstraksi fitur dapat mempengaruhi nilai akurasi yaitu sebesar 74%, *precision* sebesar 75%, *recall* sebesar 92% dan *F1-Score* sebesar 83% [9].

Peneliti Fenilinas Adi [10] melakukan pengujian sentimen analisis menggunakan algoritma SVM berbasis *Particle Swarm Optimization* (PSO) dengan data pada media sosial X dan melalui tahap *pre-processing*. Hasil klasifikasi didapatkan nilai akurasi sebesar 70% dan pengujian sentimen pada opini masyarakat pada media sosial X tentang anggota Kelompok Penyelenggara Pemungutan Suara (KPPS) didapatkan hasil bahwa 99% masyarakat memberikan opini yang positif, dan hanya 1% masyarakat yang memberikan opini negatif. Peneliti Darwis dkk [11] juga menggunakan SVM dan ekstraksi fitur menggunakan *TF-IDF*. Tahapan *pre-processing* terdiri dari *case folding*, *cleaning*, *tokenization*, *normalized*, *stemming* dan *stopwords removal*. Hasil *pre-processing* lalu dihitung nilai dari kemunculan kata untuk pelabelan yang menghasilkan sentimen positif, negatif dan netral. Berdasarkan hasil pengujian yang dihasilkan membuktikan bahwa tahapan *pre-processing* dapat meningkatkan hasil akurasi yaitu sebesar 82% dan menghasilkan sentimen dengan label negatif lebih besar dengan jumlah 77%, label positif 8% dan label netral 25%. Dalam pengembangan algoritma sentimen analisis peneliti Saputro dkk [12], mengimplementasikan *pre-processing* dan juga pelabelan leksikon menggunakan *TextBlob*, *Normalization*, *Word Vectorization* menggunakan *TF-IDF*, dan klasifikasi menggunakan *Naive Bayes* mendapatkan nilai akurasi sebesar 87.28%.

Penggunaan metode *Lexicon Based* juga digunakan oleh Diwandanu dalam proses pelabelan sentimen [13]. Proses transformasi sentimen analisis *TF-IDF* [14] menjadi salah satu metode yang banyak digunakan untuk pembobotan dan berpengaruh terhadap hasil evaluasi akhir.

Berdasarkan kelemahan dan kelebihan dari peneliti-peneliti sebelumnya, penelitian ini bertujuan untuk melakukan analisis sentimen terhadap ulasan pengguna aplikasi Tokocrypto pada google Play Store.

Tahapan *pre-processing* merupakan tahapan yang sangat penting dalam klasifikasi. Tahapan *pre-processing* terdiri dari *case folding*, *cleaning*, *filtering (stopword removal)*, *tokenizing*, mengubah kata bernegasi, *normalization* dan *stemming* digunakan sebagai proses awal dalam membersihkan teks sehingga teks dapat di proses lebih baik ke tahapan berikutnya. Selain itu, penelitian ini menggunakan gabungan metode *lexicon-based* dan *TF-IDF* dalam proses ekstraksi fitur yang diharapkan dapat meningkatkan akurasi.

Proses klasifikasi menggunakan metode SVM dan metode evaluasi terhadap hasil pengujian menggunakan nilai akurasi, *precision*, dan juga *AUC (Area undet the Curve)* untuk melihat kualitas suatu aplikasi. Hasil penelitian ini diharapkan dapat membantu perusahaan Tokocrypto untuk meningkatkan kualitas layanan dan meningkatkan kepuasan pengguna.

## METODE PENELITIAN

Penelitian ini dilakukan analisis sentimen terhadap ulasan dari aplikasi Tokocrypto pada layanan *Google Play* untuk mengklasifikasi ulasan pengguna ke dalam 2 kategori kelas sentimen, yaitu positif dan negatif menggunakan metode SVM. Gambaran umum sistem dapat dilihat pada Gambar 1. Tahapan yang dilakukan pada analisis sentimen terdiri dari pengumpulan data, *pre-processing (case folding, cleaning, filtering (stopword removal), tokenizing, mengubah kata bernegasi, normalisasi dan stemming)*. Setelah itu, dilakukan pelabelan menggunakan *Lexicon Based* dan dilakukan pembobotan *TF-IDF*. Klasifikasi dilakukan menggunakan SVM dan evaluasi menggunakan *k-fold cross validation* dan *Confusion Matrix* dan hasil sentimen yang didapatkan dilakukan visualisasi data.

### Pengumpulan Data

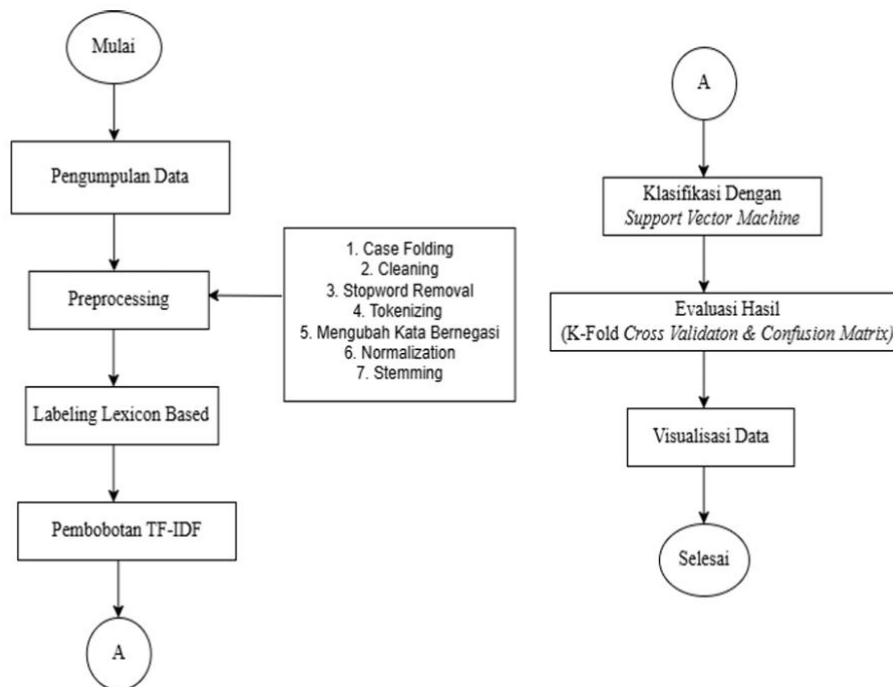
Pengambilan data dilakukan menggunakan *library python google\_play\_scraper*. Data yang digunakan dalam penelitian ini sebanyak 1500 data dengan pengambilan data mulai dari 5 Mei 2021 hingga 9 Juli 2024. Data tersebut di beri label positif dan negatif. Setelah data berhasil diperoleh, data akan disimpan di dalam dengan nama file "*scrapped\_Tokocrypto.csv*". Hasil pengumpulan data dapat terlihat pada Tabel 1 dimana terdapat *score* yang menunjukkan

penilaian aplikasi, tanggal dilakukan ulasan dan isi ulasan.

**Pre-processing**

Tahapan *pre-processing* merupakan tahap membersihkan data ulasan dari bagian yang tidak relevan agar data yang didapatkan lebih mudah diproses oleh program dan sesuai dengan kebutuhan. Tujuan dari *pre-processing*

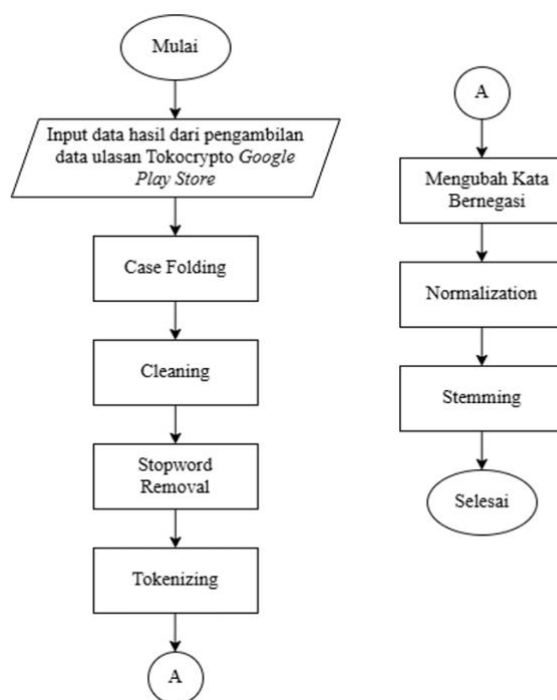
adalah untuk menghasilkan data yang relevan dengan memperoleh data yang isinya lebih mudah untuk diolah dalam sistem [15]. Tahapan ini (lihat Gambar 2) terdiri dari beberapa proses, yaitu penurunan huruf besar (*case folding*), pembersihan (*cleaning*), penghapusan kata umum (*stopword*), tokenisasi (*tokenizing*), mengubah kata bernegesi, normalisasi dan *stemming*.



**Gambar 1. Gambaran Umum Sistem**

**Tabel 1. Hasil Pengumpulan Data**

<i>score</i>	<i>at</i>	Isi ulasan
5	5/26/2021	Aplikasi perdagangan crypto yang sangat rekomendasi sekali, karena berbagai fitur yg ada sangat mudah dipahami, tampilan nya pun jg menarik untuk tramsaksi deposit maupun withdraw jg cukup mudah & cepat, Terima kasih tokocrypto
5	9/5/2022	Aplikasi nya bagus sekali mudah di pahami dan mudah di guanakan untuk user baru seperti saya, Ui nya juga simpel. Aplikasinya wort it sekali dan keamanan datanya juga sudah terjmain.
5	7/17/2024	Simple, mudah, mantap
5	7/30/2024	Aplikasi terbaik untuk trading crypto, apalgi bagi pemula seperti saya, mudah dipahami dan gak ribet
5	7/30/2024	UInya simple gak ribet enak pakai nya



**Gambar 2. Tahapan *Pre-processing***

1. ***Case Folding*** merupakan proses yang dilakukan untuk mengubah semua bentuk huruf dalam teks menjadi huruf kecil. Karakter lain yang bukan huruf atau angka, seperti tanda baca dan spasi, dianggap sebagai pemisah (*delimiter*).
2. ***Cleaning*** adalah menghilangkan kata-kata tidak relevan untuk mengurangi *noise* pada klasifikasi. Tahapan ini bertujuan untuk menghilangkan elemen-elemen yang dapat mempengaruhi hasil analisis seperti simbol, tanda baca, angka, dan spasi berlebih. Kata yang mengandung *noise* dan dihilangkan dalam proses ini adalah kata yang memiliki *emoticon*, *hashtag*, nama pengguna, *link*, angka, dan beberapa karakter lainnya seperti simbol, tanda baca, spasi, karakter tunggal [16].
3. ***Stopword Removal*** merupakan tahapan untuk menghapus kata-kata yang tidak diperlukan atau tidak memiliki makna yang berpengaruh terhadap proses klasifikasi. Tujuan dari penggunaan *stopword* adalah untuk menghilangkan kata-kata dengan informasi rendah dalam sebuah teks, sehingga kita dapat lebih fokus pada kata-kata yang memiliki makna signifikan.
4. ***Tokenizing*** yaitu pemisahan teks menjadi bagian-bagian yang lebih kecil (token) untuk dianalisis lebih lanjut. Kata, angka, simbol, tanda baca, dan entitas penting lainnya dianggap sebagai token.

5. **Mengubah Kata Bernegasi** berarti menambahkan tanda "\_" di antara kata-kata yang memiliki makna negatif seperti "tidak" dan kata yang mengikutinya agar mendapatkan makna yang tepat.
6. **Normalization** yaitu mengubah kata-kata tidak baku menjadi kata-kata yang baku dan sesuai standar untuk diproses lebih lanjut. Kata-kata tidak baku yang dimaksud adalah yang tidak sesuai dengan Kamus Besar Bahasa Indonesia (KBBI).
7. **Stemming** merupakan proses mengubah kata menjadi bentuk dasarnya dengan menghapus semua infleksi, termasuk awalan, sisipan, akhiran, dan kombinasi imbuhan pada kata turunan.

#### **Labeling Lexicon-Based**

Tahapan ini dilakukan pelabelan menggunakan *Lexicon-Based VADER (Valence Aware Dictionary and sentiment Reasoner) Sentimen* [17][18], yang bertujuan untuk secara otomatis memberi label pada ulasan. Dalam berbagai kamus leksikon terdapat dua jenis kamus yang digunakan yaitu kamus kata tanpa nilai polaritas untuk setiap kata, dan kamus kata dengan nilai/skor untuk setiap kata sesuai dengan tingkat polaritasnya. Label yang diberikan terdiri dari dua kategori: positif dan negatif berdasarkan perhitungan skor bobot kata (*compound score*) dari *library*. Perhitungan bobot pada pelabelan *lexicon based* didapatkan dari analisis sentimen

menggunakan *library VADER* dengan menjumlahkan skor valensi setiap kata dalam leksikon sesuai dengan aturan dan dinormalisasi. Jika nilai *compound*  $\geq 0,05$  maka sentimen bernilai positif sedangkan jika nilai *compound*  $\leq 0,05$  maka sentimen bernilai negatif. Proses pelabelan dengan *Lexicon-Based VADER* dimulai perhitungan nilai bobot setiap kata dari data ulasan dari hasil *stemming*, kemudian program secara otomatis akan memberikan label berdasarkan hasil perhitungan *compound score* tersebut.

#### **Pembobotan TF-IDF**

Data yang dihasilkan akan diubah menjadi representasi numerik dengan menggunakan *TF-IDF* yang dapat digunakan oleh algoritma pembelajaran mesin. Representasi ini mempertimbangkan frekuensi kemunculan kata dalam dokumen (TF) serta keunikannya di seluruh koleksi dokumen (IDF). Bobot kata akan semakin besar jika sering muncul dalam suatu dokumen dan akan semakin kecil jika muncul dalam banyak dokumen. Vektor fitur ini kemudian digunakan sebagai input dalam klasifikasi SVM, yang akan mempelajari pola dalam data latih untuk mengidentifikasi sentimen yang berbeda.

#### **Klasifikasi SVM**

Klasifikasi SVM dilakukan dengan mencari hyperplane atau garis pembatas (decision boundary) yang memisahkan antara satu kelas dengan kelas yang lain. Hyperplane terbaik

dapat ditentukan dengan mengukur margin hyperplane dan mencari titik maksimum. Margin adalah jarak antara hyperplane dengan pola terdekat dari masing masing kelas. Komponen SVM terdiri atas optimal hyperplane, hyperplane kelas positif, hyperplane kelas negatif dan margin. Langkah awal klasifikasi adalah pemilihan model SVM yang tepat, termasuk menentukan kernel dan parameter yang paling sesuai.

Dalam penelitian ini, kernel yang diuji coba adalah linear dan RBF (Radial Basis Function) dengan menguji 3 parameter C yaitu antara 1, 10, dan 100. Hal tersebut dilakukan untuk mengevaluasi pengaruh kekuatan regulasi pada model dan menilai kinerjanya dalam memprediksi label sentimen dari data ulasan Tokocrypto. Setelah melalui tahap pelatihan, model SVM dihadapkan pada data uji yang baru untuk mengukur performa dalam memprediksi sentimen pada ulasan Tokocrypto.

### **Cross Validation**

Cross validation yaitu memisahkan data menjadi dua bagian: satu bagian untuk proses pelatihan (training) dan bagian lainnya untuk pengujian (testing) dengan 3 parameter nilai K-fold yang akan digunakan, yaitu K= 2, 5, dan 10.

Metode K-fold bekerja dengan cara melakukan perulangan sambil mengacak atribut input sehingga sistem diuji dengan berbagai atribut input yang dipilih secara acak

dapat memilih model yang memiliki kinerja terbaik dan paling stabil [19].

### **Evaluasi**

Evaluasi dilakukan untuk mengukur kinerja klasifikasi yang dilakukan oleh algoritma SVM dengan tujuan untuk mengevaluasi seberapa akurat model algoritma yang dibuat.

Evaluasi kinerja dilakukan dengan mengutamakan akurasi sebagai parameter evaluasi untuk memperoleh nilai AUC, akurasi, precision dan recall menggunakan confusion matrix. Confusion matrix berfungsi sebagai alat analisis untuk mengevaluasi kemampuan model klasifikasi dalam mengenali tupel data yang berbeda. AUC digunakan untuk mengukur seberapa baik model membedakan antara kelas positif dan negatif tanpa memperhatikan ambang batas spesifik. Nilai AUC berkisar antara 0 hingga 1, di mana nilai 1 menunjukkan bahwa model sempurna dalam membedakan kelas, sedangkan nilai 0.5 menunjukkan performa yang sama dengan pengacakan.

Nilai akurasi didapatkan dari jumlah data bernilai positif yang diprediksi positif dan data bernilai negatif yang diprediksi negatif dibagi dengan jumlah seluruh data di dalam dataset. Precision adalah peluang kasus yang diprediksi positif yang pada kenyataannya termasuk kasus kategori positif. Recall merupakan salah satu perhitungan akurasi prediksi yang digunakan sebagai ukuran



tingkat keberhasilan sistem dalam memulihkan informasi.

*stopwords removal* pada ulasan dapat dilihat pada Tabel 3.

## HASIL DAN PEMBAHASAN

### Hasil *Pre-processing*

Hasil tahapan *pre-processing* terdiri dari *case folding dan cleaning, filtering (stopword removal), tokenizing*, mengubah kata bernegasi, *normalization* dan *stemming*. Hasil *case folding dan cleaning* pada ulasan dapat dilihat pada Tabel 2. Dapat dilihat pada ulasan “Sangat mudah digunakan untuk pemula seperti saya gak ribet lagi” , simbol yang terdapat pada ulasan tersebut hilang. Pada tahap *stopwords removal* adalah tahap untuk menghapus kata-kata yang tidak diperlukan atau tidak memiliki makna yang berpengaruh terhadap proses klasifikasi. Hasil tahap

Pada ulasan “**sangat mudah digunakan untuk pemula seperti saya** gak ribet lagi”, kemudian dihilangkan kata-kata yang tidak diperlukan, diubah menjadi “mudah pemula gak ribet”.

Pada tahap *tokenizing* yaitu proses membagi teks menjadi bagian - bagian lebih kecil sehingga menjadi sebuah token. Hasil tahap *tokenizing* pada ulasan dapat dilihat pada Tabel 4. Pada tahap mengubah kata bernegasi, mengubah kata yang memiliki negasi berarti menambahkan tanda "\_" di antara kata-kata yang memiliki makna negatif, seperti "tidak" dan kata yang mengikutinya agar mendapatkan makna yang tepat. Hasil tahap mengubah kata bernegasi pada ulasan dapat dilihat pada Tabel

**Tabel 2. Hasil *Case Folding dan Cleaning***

<b>Input</b>	<b><i>Case Folding dan Cleaning</i></b>
Sangat mudah digunakan untuk pemula seperti saya gak ribet lagi	sangat mudah digunakan untuk pemula seperti saya gak ribet lagi
Awalnya ragu tapi coba coba bagus dan mulai ketagihan good pokonya terbaik best	awalnya ragu tapi coba coba bagus dan mulai ketagihan good pokonya terbaik best

**Tabel 3. Hasil *Filtering (Stopword Removal)***

<b>Input</b>	<b><i>Stopword Removal</i></b>
sangat mudah digunakan untuk pemula seperti saya gak ribet lagi	mudah pemula gak ribet
awalnya ragu tapi coba coba bagus dan mulai ketagihan good pokonya terbaik best.	ragu tapi coba coba bagus ketagihan good pokonya terbaik best

**Tabel 4. Hasil *Tokenizing***

<b>Input</b>	<b><i>Tokenizing</i></b>
mudah pemula gak ribet	['mudah', 'pemula', 'gak', 'ribet']
ragu tapi coba coba bagus ketagihan good pokonya terbaik best	['ragu', 'tapi', 'coba', 'coba', 'bagus', 'ketagihan', 'good', 'pokonya', 'terbaik', 'best']

**Tabel 5. Hasil Mengubah Kata Bernegasi**

Sebelum Kata Bernegasi	Sesudah Kata Bernegasi
mudah digunakan untuk pemula soalnya aplikasinya user friendly banget simpel dan aman juga pokoknya tentunya tidak ribet enak banget sih aplikasi dipakai pemula yang aman tidak ribet tradingg lancar cuann bos	mudah digunakan untuk pemula soalnya aplikasinya user friendly banget simpel dan aman juga pokoknya tentunya <b>tidak_ribet</b> enak banget sih apk dipakai pemula yang aman <b>tidak_ribet</b> tradingg lancar cuann bos

**Tabel 6. Hasil Normalization dan Stemming**

Input	Normalization dan Stemming
['lancar', 'ga', 'ribet', 'pemula', 'cocok', 'banget', 'withdrawnya', 'gampang', 'the', 'best', 'trading', 'lengkap']	['lancar', 'tidak', 'susah', 'pemula', 'cocok', 'sangat', 'withdraw', 'mudah', 'the', 'best', 'trading', 'lengkap']
['mudah', 'pemula', 'gak', 'ribet']	['mudah', 'pemula', 'tidak', 'susah']

**Tabel 7. Hasil Labeling Lexicon-Based**

Stemming	Compound Score	Sentimen
['tepat', 'sekali', 'pakai', 'tokocrypto', 'tuh', 'penggunaan', 'nya', 'mudah', 'pahami', 'menu', 'aplikasi', 'nya', 'tidak', 'susah', 'fee', 'nya', 'yang', 'rendah', 'deposit', 'biaya', 'potongan', 'buat', 'suka', 'tokocrypto', 'karena', 'cocok', 'sangat', 'pemula', 'fitur', 'terlaksana', 'overall', 'sudah', 'bagus', 'mudah', 'pintar', 'investasi', 'mari', 'pakai', 'tokocrypto']	0.8074	Positif
['verifikasi', 'kyc', 'aplikasi', 'jatuh', 'restart', 'mohon', 'tetap', 'bintang']	-0.4019	Negatif

*Normalization* yaitu mengubah setiap kata yang tidak baku menjadi kata yang baku dan siap untuk diproses. Kata tidak baku yang dimaksud adalah kata-kata yang tidak sesuai dengan KBBI dan proses *stemming* mengubah kata menjadi bentuk dasarnya dengan menghapus semua infleksi, termasuk awalan, sisipan, akhiran, dan kombinasi imbuhan pada kata turunan. Hasil tahap *normalization* dan *stemming* pada ulasan dapat dilihat pada Tabel 6.

#### Hasil Labeling Lexicon-Based

Pada tahap ini, dilakukan pelabelan menggunakan *Lexicon-Based VADER*

sentimen, dengan tujuan memberikan label otomatis pada ulasan.

Label yang diberikan terdiri dari dua kategori yaitu positif dan negatif. Hasil *Labeling Lexicon-based* dapat dilihat pada Tabel 7.

#### Hasil Pembobotan TF-IDF

Dalam proses ini, data ulasan yang sudah diproses masih berupa kata-kata, namun ketika menganalisis klasifikasi data harus berbentuk numerik atau angka. Oleh karena itu, data harus diubah ke format numerik dengan memberikan bobot pada kata-kata

menggunakan metode *TF-IDF*. Proses ini akan menghasilkan *term* dengan bobot paling terbesar (lihat Tabel 8).

### Hasil Klasifikasi SVM

Penelitian ini menggunakan metode SVM dengan dua jenis kernel SVM, yaitu *Linear* dan *RBF/Non-Linear* serta diuji dengan tiga nilai parameter *C* yang berbeda (1, 10, 100) menggunakan nilai *cross validation K-fold* 2, 5 dan 10. Proses ini dilakukan untuk mengukur performa model dalam memprediksi label sentimen ulasan aplikasi Tokocrypto dengan evaluasi seperti *AUC*,

akurasi, *precision*, dan *Recall*. Metrik-metrik tersebut digunakan untuk menilai seberapa efektif model dalam memprediksi sentimen dari data ulasan. Hasil uji coba dapat dilihat pada Tabel 9 dengan parameter *C*=1, Tabel 10 dengan parameter *C*=10, dan Tabel 11 dengan parameter *C*=100. Hasil penelitian ini menunjukkan bahwa kernel *linear* dengan parameter *C* = 1 dan rasio pembagian pada data *K-fold* 10 memberikan performa terbaik dengan nilai *AUC* sebesar 90.5% dan akurasi sebesar 85.80%, *precision* 88,81% dan *recall* 78,15%. Tabel 12. merupakan hasil *confusion matrix* dari nilai *AUC* tertinggi.

**Tabel 8. Hasil Pembobotan TF-IDF**

<i>Term</i>	<i>Weight</i>
<b>mudah</b>	0.032283
<b>trading</b>	0.027433
<b>koin</b>	0.021315
<b>lengkap</b>	0.020119
<b>crypto</b>	0.013661
<b>paham</b>	0.013021

**Tabel 9. Tabel Parameter C=1**

<b>kernel</b>	<b>Linear</b>				<b>RBF</b>			
	<i>K-fold</i>	<i>AUC</i>	<i>Accuracy</i>	<i>Precision</i>	<i>Recall</i>	<i>AUC</i>	<i>Accuracy</i>	<i>Precision</i>
<b>2</b>	87.64	81.93	84.45	72.60	85.89	76.53	83.50	72.75
<b>5</b>	90.37	85.27	88.36	77.30	89.14	80.33	87.27	68.60
<b>10</b>	<b>90.50</b>	<b>85.80</b>	<b>88.81</b>	<b>78.15</b>	89.51	80.80	87.09	81.42

**Tabel 10. Tabel Parameter C=10**

<b>kernel</b>	<b>Linear</b>				<b>RBF</b>			
	<i>K-fold</i>	<i>AUC</i>	<i>Accuracy</i>	<i>Precision</i>	<i>Recall</i>	<i>AUC</i>	<i>Accuracy</i>	<i>Precision</i>
<b>2</b>	86.50	82.47	80.17	77.83	86.63	79.87	82.87	79.20
<b>5</b>	88.42	85.00	83.58	80.46	89.52	83.47	86.61	81.61
<b>10</b>	88.20	85.27	83.46	81.61	89.76	84.20	86.97	75.82

**Tabel 11. Tabel Parameter C=100**

<b>kernel</b>	<b>Linear</b>				<b>RBF</b>			
	<i>K-fold</i>	<i>AUC</i>	<i>Accuracy</i>	<i>Precision</i>	<i>Recall</i>	<i>AUC</i>	<i>Accuracy</i>	<i>Precision</i>
<b>2</b>	85.81	81.80	79.50	76.49	86.63	79.87	82.87	75.30
<b>5</b>	87.60	83.53	81.78	78.74	89.52	83.47	86.68	81.61
<b>10</b>	87.04	84.20	82.50	79.78	89.76	84.20	86.97	75.82

Hasil penelitian terdahulu terkait dengan analisis sentimen menggunakan metode klasifikasi SVM [8] [9] mendapatkan nilai akurasi sebesar 85.11% dan 74%. Selain itu, metode ekstraksi fitur menggunakan *TF-IDF* dan *lexicon-based* juga digunakan pada peneliti [12] dan mendatkan nilai akurasi sebesar 87.28%. Tahapan *pre-processing* dalam analisis sentimen dilakukan oleh peneliti [10][11] dan mendapatkan nilai akurasi masing-masing sebesar 70% dan 82%. Penelitian ini menggunakan gabungan dari kelebihan peneliti-peneliti sebelumnya, yaitu menggunakan metode *pre-processing*, metode ekstraksi fitur *TF-IDF* dan *lexicon based* dan metode klasifikasi SVM. Hasil *AUC* dan akurasi yang didapatkan sebesar 90.50% dan 85.80%. Hasil evaluasi menunjukkan nilai *AUC* yang tinggi dan adanya peningkatan nilai akurasi dari peneliti-peneliti sebelumnya. Hal tersebut menunjukkan metode yang digunakan mampu memberikan hasil yang optimal.

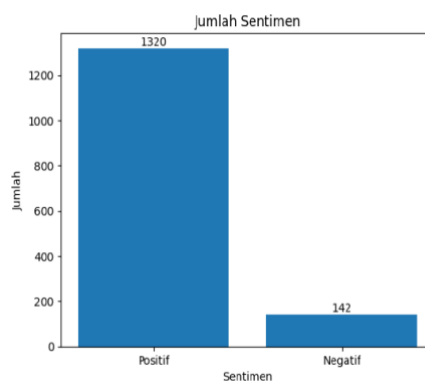
## Hasil Visualisasi Data

Visualisasi *chart* dan *wordcloud* dari hasil klasifikasi menunjukkan bahwa hasil prediksi (diagram batang) aplikasi Tokocrypto memiliki data positif 1320 dan data negatif 142. Hasil tersebut menunjukkan bahwa hasil ulasan lebih banyak mengandung kata positif dibandingkan dengan kata negatif (lihat Gambar 3).

*Wordcloud* hasil prediksi pada kelas positif pada Gambar 4 menunjukkan bahwa adanya kepuasan pengguna dan menganggap Tokocrypto merupakan aplikasi yang bagus dan berguna seperti kata yang terlihat atau muncul seperti ‘terbaik’, ‘lengkap’ dan lain-lain. *Wordcloud* Hasil prediksi pada kelas negatif pada Gambar 5 menunjukkan bahwa adanya kepuasan pengguna dan menganggap Tokocrypto merupakan aplikasi yang kurang bagus seperti kata yang terlihat yaitu ‘susah’, ‘gagal’ dan lain-lain.

**Tabel 12. Hasil Confusion Matrix**

<i>Confusion Matrix</i>	<i>Linear, C=1, K-fold 10</i>	
	<i>Predicted Negative</i>	<i>Predicted Positive</i>
<i>Actual Negative</i>	70	72
<i>Actual Positive</i>	17	1303



**Gambar 3. Grafik Hasil Prediksi (Diagram Batang)**



- Ekonomi Digital”, April, 2024. [Online]. Available: <https://www.kemendag.go.id/berita/pojok-media/bappebti-perdagangan-kripto-jadi-strategi-percepatan-ekonomi-digital>. [Accessed Agustus. 12, 2024].
- [2] Tokocrypto, “Kenalan Dulu dengan Tokocrypto”. [Online]. Available: <https://support.tokocrypto.com/hc/id/articles/17502415349389-Kenalan-Dulu-dengan-Tokocrypto>. [Accessed Agustus. 30, 2024].
- [3] P. Ladita, “Analisis Penerapan Aplikasi Android Tokocrypto Menggunakan Pendekatan Design Thinking Dibantu Dengan Platform Design Toolkit V.2,” Anal. Penerapan Apl. Android Tokocrypto Menggunakan Pendekatan Des. Think. Dibantu Dengan Platf. Des. Toolkit V.2, pp. 1–132, 2020.
- [4] Aftab, F., Bazai, S.U., Marjan, S., Baloch, L., Aslam, S., Amphawan, A., Neo, T.-K., “A Comprehensive Survey on Sentimen Analysis Techniques”. *International Journal of Technology*. Volume 14(6), pp. 1288-1298. 2023.
- [5] S. Sindhu, S. Kumar and A. Noliya, "A Review on Sentimen Analysis using Machine Learning," *2023 International Conference on Innovative Data Communication Technologies and Application (ICIDCA)*, Uttarakhand, India, pp. 138-142, doi: 10.1109/ICIDCA56705.2023.10099665 . 2023.
- [6] Monika, P., C. Kulkarni, N. Harish Kumar, S. Shruthi, and V. Vani. “Machine Learning Approaches for Sentimen Analysis: A Survey”. *International Journal of Health Sciences*, vol. 6, no. S4, Apr., pp. 1286-00, 2022. doi:10.53730/ijhs.v6nS4.6119.
- [7] Moraes, Rodrigo & Valiati, Joao & Gavião Neto, Wilson. Document-level sentimen classification: An empirical comparison between SVM and ANN. *Expert Systems with Applications*. 40. 621–633. 2013. 10.1016/j.eswa.2012.07.059.
- [8] I. Siti Aisah, B. Irawan, and T. Suprapti, “Algoritma Support Vector Machine (Svm) Untuk Analisis Sentimen Ulasan Aplikasi Al Qur’an Digital”, *JATI (Jurnal Mhs. Tek. Inform)*, vol. 7, no. 6, pp. 3759–3765, 2024. doi: 10.36040/jati.v7i6.8263.
- [9] Isnain, Auliya Rahman et al. “Sentimen Analisis Publik Terhadap Kebijakan Lockdown Pemerintah Jakarta Menggunakan Algoritma Svm”. *Jurnal Data Mining dan Sistem Informasi*, 2021.
- [10] Artanto, Fenilinas Adi, “Support Vector Machine Berbasis Particle Swarm Optimization Pada Analisis Sentimen Anggota KPPS.” *JURNAL FASILKOM* 14 (1). LPPM Universitas

- Muhammadiyah Riau: 75–79. 2024. doi:10.37859/jf.v14i1.6795.
- [12] Darwis, Dedi & Pratiwi, Eka & Pasaribu, A. ‘Penerapan Algoritma Svm Untuk Analisis Sentimen Pada Data Twitter Komisi Pemberantasan Korupsi Republik Indonesia’. *Edutic - Scientific Journal of Informatics Education*. Vol. 7.No.1.2020.10.21107/edutic.v7i1.8779
- [13] Saputro, M., & Alamsyah, A. “Comparison of Naive Bayes Classifier and K-Nearest Neighbor Algorithms with Information Gain and Adaptive Boosting for Sentimen Analysis of Spotify App Reviews”. *Recursive Journal of Informatics*, 2(1), 37-44. 2024. <https://doi.org/10.15294/rji.v2i1.68551>
- [14] Diwandanu, M.T., & Wisudawati, L.M. “Analisis Sentimen Terhadap Twit Maxim Pada Twitter Menggunakan R Programming Dan K Nearest Neighbors”. *Jurnal Ilmiah Informatika Komputer*.2023.DOI: <http://dx.doi.org/10.35760/ik.2023.v28i1.7909>
- [15] Pamungkas, Adji Surya & Cahyono, N. “Analisis Sentimen Review ChatGPT di Play Store menggunakan Support Vector Machine dan K-Nearest Neighbor”. *Edumatic: Jurnal Pendidikan Informatika*. Vol. 8 No. 1, Juni, Hal. 1-10, 2024. DOI: 10.29408/edumatic.v8i1.24114
- [16] P. Aditiya, U. Enri, and I. Maulana, “Analisis Sentimen Ulasan Pengguna Aplikasi Myim3 Pada Situs Google Play Menggunakan Support Vector Machine,” *JURIKOM (Jurnal Ris. Komputer)*, vol. 9, no. 4, p. 1020, Aug. 2022. doi: 10.30865/jurikom.v9i4.4673.
- [17] G. A. Buntoro, “Analisis sentimen calon gubernur DKI Jakarta 2017 di Twitter,” *Integer J.*, vol. 2, no. 1, pp. 32–41, 2017, [Online]. Available: <https://t.co/jrvaMsgBdH>.
- [18] Hutto, C., & Gilbert, E. VADER: A Parsimonious Rule-Based Model for Sentimen Analysis of Social Media Text. *Proceedings of the International AAAI Conference on Web and Social Media*, 8(1), 216-225. 2014. <https://doi.org/10.1609/icwsm.v8i1.14550>
- [19] Asri, Yessy & Suliyanti, Widya & Kuswardani, Dwina & Fajri, Muhamad. “Pelabelan Otomatis Lexicon Vader dan Klasifikasi Naive Bayes dalam menganalisis sentimen data ulasan PLN Mobile: Analisis Sentimen”. *PETIR*. 15. 264-275. 2022. 10.33322/petir.v15i2.1733.
- [20] Berrar, Daniel. Cross-Validation. 2018. 10.1016/B978-0-12-809633-8.20349-X.