

# **Analisis Sentimen Aplikasi Tokocrypto Berdasarkan Ulasan Pada Google Play Store Menggunakan Metode Naïve Bayes**

**Rizki Adi Saputra, Dion Parisda Ray, Faldy Irwiensyah\***

Fakultas Teknologi Industri dan Informatika, Teknik Informatika, Universitas Muhammadiyah Prof. Dr. Hamka, Jakarta, Indonesia

Email: <sup>1</sup>belajar7706@gmail.com, <sup>2</sup>dion7472@gmail.com, <sup>3,\*</sup>faldy@uhamka.ac.id

Email Penulis Korespondensi: faldy@uhamka.ac.id

**Abstrak**—Perkembangan teknologi yang semakin canggih telah membuat banyak perubahan dan kemudahan bagi manusia dalam segala hal termasuk di bidang keuangan, *cryptocurrency* muncul sebagai bentuk inovasi dalam dunia keuangan. Exchange *cryptocurrency* merupakan platform elektronik yang memungkinkan penjual dan pembeli melakukan transaksi perdagangan aset kripto melalui situs web atau aplikasi seluler. Saat ini banyak aplikasi exchange *cryptocurrency* yang buruk dalam pelayanan, keamanan yang tidak terjamin, penarikan dana yang cukup lama, biaya admin mahal dan masalah lainnya, sehingga banyak masyarakat Indonesia menggunakan review di Google Play Store untuk melihat ulasan pengguna sebelum memutuskan menggunakan aplikasi exchange *cryptocurrency* tersebut. Banyak dari masyarakat Indonesia mencari informasi aplikasi exchange *cryptocurrency* yang memberikan pelayanan terbaik untuk melakukan transaksi jual dan beli *cryptocurrency*, salah satu aplikasi exchange *cryptocurrency* yang menawarkan hal tersebut adalah aplikasi Tokocrypto berdasarkan ulasan di Google Play Store. Penelitian ini bermaksud untuk memahami sentimen terhadap ulasan pengguna pada aplikasi Tokocrypto dengan metode algoritma Naïve Bayes untuk klasifikasi data. Data yang diperoleh merupakan data ulasan Februari 2024 yang berjumlah sebanyak 2000 ulasan dari Google Play Store menggunakan Google Colaboratory. Tahapan penelitian ini terdiri dari scraping data menggunakan teknik web scraping, pelabelan data, preprocessing, pembobotan TF-IDF, implementasi algoritma Naïve Bayes dan evaluasi. Data bersih yang berhasil diperoleh sebanyak 1000 dengan jumlah sentimen positif sebanyak 396 dan sentimen negatif sebanyak 604. Hasil dari penelitian analisis sentimen dengan metode algoritma Naïve Bayes menunjukkan 74.22% untuk akurasi, 63.25% untuk presisi dan 81.40% untuk recall.

**Kata Kunci:** Tokocrypto; Naïve Bayes; Exchange Cryptocurrency; Ulasan; Data

**Abstract**—The advancement of increasingly sophisticated technology has brought numerous changes and conveniences for humans in all aspects, including the financial sector. *Cryptocurrency* has emerged as an innovation in the financial world. A *cryptocurrency* exchange is an electronic platform that enables sellers and buyers to conduct *cryptocurrency* trading transactions through a website or mobile application. Currently, many *cryptocurrency* exchange applications suffer from poor service, unreliable security, lengthy withdrawal processes, high administrative fees, and other issues. As a result, many people in Indonesia rely on reviews on the Google Play Store to check user feedback before deciding to use these *cryptocurrency* exchange applications. Many Indonesians seek information on *cryptocurrency* exchange applications that provide the best services for buying and selling *cryptocurrency*. One such application, according to reviews on the Google Play Store, is Tokocrypto. This study aims to understand the sentiment towards user reviews of the Tokocrypto application using the Naïve Bayes algorithm for data classification. The data obtained consists of 2,000 reviews from the Google Play Store in February 2024, collected using Google Colaboratory. The research stages include data scraping using web scraping techniques, data labeling, preprocessing, TF-IDF weighting, implementing the Naïve Bayes algorithm, and evaluation. The cleaned data resulted in 1,000 reviews, with 396 positive sentiments and 604 negative sentiments. The results of sentiment analysis research using the Naïve Bayes algorithm method show 74.22% for accuracy, 63.25% for precision, and 81.40% for recall.

**Keywords:** Tokocrypto; Naïve Bayes; Exchange Cryptocurrency; Reviews; Data

## **1. PENDAHULUAN**

Perkembangan inovasi dan teknologi di zaman ini merupakan yang paling pesat sehingga membuat banyak perubahan dan kemudahan bagi manusia dalam segala hal termasuk di bidang keuangan, *cryptocurrency* muncul sebagai bentuk inovasi dalam dunia keuangan [1]. *Cryptocurrency* merupakan teknologi berbasis blockchain yang umumnya digunakan sebagai bentuk mata uang digital [2]. *Exchange cryptocurrency* atau yang biasa disebut bursa perdagangan aset kripto merupakan platform elektronik yang memungkinkan penjual dan pembeli melakukan transaksi perdagangan aset kripto melalui situs web atau aplikasi seluler [3]. Salah satu aplikasi *exchange cryptocurrency* yang terkenal dari Indonesia untuk transaksi aset mata uang digital adalah aplikasi *exchange* Tokocrypto.

Saat ini banyak aplikasi *exchange cryptocurrency* yang buruk dalam pelayanan, keamanan yang tidak terjamin, penarikan dana yang cukup lama, biaya admin mahal dan masalah lainnya, sehingga banyak masyarakat Indonesia menggunakan review di Google Play Store untuk melihat pendapat publik terkait saran, isu, kritik dan opini-opini sebelum memutuskan menggunakan aplikasi *exchange cryptocurrency* tersebut. Banyak dari masyarakat Indonesia mencari informasi aplikasi *exchange cryptocurrency* yang memberikan pelayanan terbaik untuk melakukan transaksi jual dan beli *cryptocurrency*, salah satu aplikasi *exchange cryptocurrency* yang menawarkan hal tersebut adalah aplikasi Tokocrypto, menurut ulasan dan pandangan pengguna di Google Play Store. Google Play Store merupakan platform untuk konten digital, yang mengizinkan pengguna *smartphone android* pada mengunduh aplikasi dan produk digital lainnya, seperti game, film, e-book, dan sebagainya, dari yang berbayar hingga yang gratis [4]. Di Google Play Store, terdapat fitur penting yaitu rating dan ulasan. Fitur ini memiliki signifikansi besar karena dapat menjadi faktor utama yang memengaruhi calon pengguna dalam memutuskan untuk menggunakan suatu aplikasi.

Analisis sentimen merupakan proses penggalian informasi dari teks yang bersifat subjektif, yang dapat memberikan bantuan kepada pihak yang memerlukan [5]. Tugas terpenting dalam analisis sentimen adalah

mengklasifikasikan polaritas teks yang terdapat dalam sebuah tingkat aspek, dokumen atau kalimat. Ini mencakup penentuan apakah opini yang diungkapkan dalam sebuah kalimat, dokumen atau tingkat aspek tersebut bersifat opini positif atau negatif [6]. Metode yang dipakai pada penelitian ini adalah metode algoritma Naïve Bayes Classifier. Metode algoritma Naïve Bayes Classifier ini adalah sebuah teknik klasifikasi algoritma yang memperkirakan statistik dan probabilitas untuk prediksi. Pemanfaatan metode Naïve Bayes Classifier ini memberikan keuntungan karena pada prosesnya untuk klasifikasi data, hanya diperlukan sedikit data untuk perhitungannya [7].

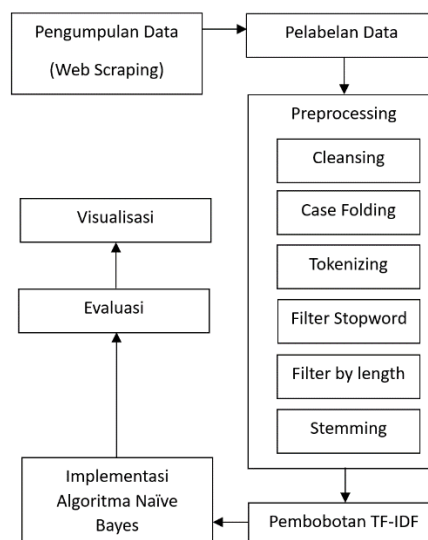
Beberapa penelitian sebelumnya mengenai analisis sentimen adalah penelitian Fachri Zaini yang melakukan penelitian analisis sentimen masyarakat terhadap kegagalan indonesia menjadi Tuan Rumah U-20 menggunakan *multinomial Naïve Bayes Classifier*. Hasil penelitian ini adalah dari 214 data tweets yang ada mengenai keyword kata dunia, piala, indonesia, pildun, gagal, ganjar, israel dan kecewa. dalam bahasa Indonesia, ditemukan sebanyak 121 kata negatif dan 93 kata positif dengan nilai 97.26% untuk akurasi, 93.33% untuk presisi dan 100% untuk *Recall* [8]. Penelitian lainnya tentang analisis sentimen adalah penelitian Fauzan Setya Ananto yaitu Analisis Sentimen Terhadap Ulasan Pengguna di Google Play Store pada Aplikasi MyPertamina dengan Algoritma Naïve Bayes. Hasil penelitian ini menunjukkan bahwa dari 1500 dataset yang diperoleh mengenai tanggapan pengguna di *Google Play Store* tentang Aplikasi MyPertamina, Setelah proses klasifikasi dan evaluasi, ditemukan bahwa tingkat akurasi mencapai 77.42%, presisi sebesar 49.98%, dan *recall* sebesar 76.87% [9].

Perlu dilakukan perbandingan antara metode algoritma *Naïve Bayes* dengan pendekatan lainnya untuk menilai seberapa efektifnya metode tersebut dalam analisis sentimen. Penelitian sebelumnya yang menggunakan metode berbeda adalah penelitian Caesar Rio Anggina Toruan yang menggunakan metode algoritma *Long Short-Term Memory (LSTM)* pada Tokocrypto, sehingga hal ini memungkinkan untuk melakukan perbandingan antara kedua metode tersebut. Tingkat akurasi yang diperoleh metode algoritma *Long Short-Term Memory (LSTM)* sebesar 95%, sedangkan tingkat akurasi yang diperoleh metode algoritma *Naïve Bayes* hanya sebesar 74.22%. Hal ini menunjukkan bahwa metode *Long Short-Term Memory (LSTM)* lebih efektif daripada metode *Naïve Bayes* dalam menganalisis sentimen. Berdasarkan penelitian sebelumnya, dengan metode algoritma *Long Short-Term Memory (LSTM)* untuk menganalisis sentimen Tokocrypto, Setelah dilakukan pengujian dan evaluasi model, didapatkan hasil nilai skor-fi 0.95, presisi 0.98, dan *recall* 0.92 [10].

Berdasarkan uraian masalah yang telah dipaparkan sebelumnya, maka penelitian ini akan menggunakan metode algoritma *Naïve Bayes* dalam menganalisis sentimen. Metode ini dipilih berdasarkan hasil penelitian terdahulu yang telah terbukti akurat dan memiliki banyak kelebihan daripada metode lainnya. Penelitian ini juga bertujuan untuk mengetahui pendapat dan tingkat kepuasan pengguna dalam melakukan transaksi jual dan beli *cryptocurrency* di aplikasi Tokocrypto. Diharapkan juga hasil dari penelitian ini, mengenai tanggapan pengguna dijadikan evaluasi kepada pengembang aplikasi Tokocrypto, serta memberikan kontribusi dan manfaat bagi masyarakat umum.

## 2. METODOLOGI PENELITIAN

Pada tahap awal, penelitian ini melibatkan serangkaian langkah yang dilakukan secara teratur dan berurutan. Dimulai dari pengumpulan data dengan teknik *Web Scraping*, pelabelan data, preprocessing, pembobotan TF-IDF, implementasi algoritma Naïve Bayes, hingga evaluasi. Proses *Web Scraping* akan menggunakan *Google Colaboratory* untuk memperoleh data ulasan pengguna aplikasi Tokocrypto di *Google Play Store*. Untuk proses klasifikasi data, perangkat lunak RapidMiner akan digunakan. Setelah proses klasifikasi data dengan metode algoritma *Naïve Bayes*, hasilnya akan dievaluasi dan divisualisasikan. Berikut ilustrasi tahapan alur penelitian pada Gambar 1.

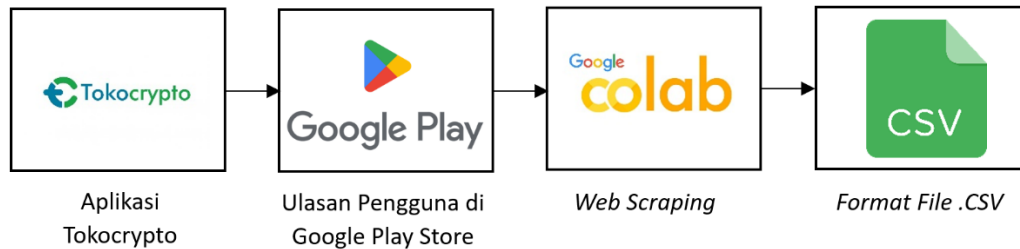


Gambar 1. Tahapan Alur Penelitian

Pada gambar 1 merupakan alur tahapan penelitian ini, mulai dari *Web Scraping* hingga *Evaluasi*.

## 2.1 Pengumpulan Data

*Web scraping* merupakan teknik pengumpulan data web terstruktur secara otomatis yang dilakukan oleh aplikasi khusus atau kode pemrograman [11]. Dengan menggunakan teknik *Web Scraping* untuk mengumpulkan dataset dari *Google Play Store*, dataset yang diperoleh menjadi terstruktur secara otomatis. Dataset yang berhasil diperoleh dengan *web scraping* menggunakan website *Google Colaboratory* kemudian disimpan dalam format file .CSV sesuai dengan jumlah data ulasan pengguna yang dibutuhkan.



**Gambar 2.** Proses Web Scraping

Pada gambar 2 merupakan proses *Scraping Data* ulasan pengguna aplikasi Tokocrypto dari *Google Play Store* menggunakan website *Google Colaboratory*, dataset yang berhasil diperoleh akan disimpan dalam format file .CSV.

## 2.2 Pelabelan Dataset

Dataset yang telah diperoleh melalui teknik *web scraping*, kemudian data masuk ke tahap proses pelabelan dataset. Ulasan dalam dataset dianalisis dengan memberikan label pada setiap ulasan, yang bertujuan untuk mengklasifikasikan data ke dalam dua jenis kategori sentimen yaitu kategori positif dan kategori negatif [12]. Proses pelabelan data penting untuk menetapkan kelas dari suatu ulasan dalam dokumen, apakah ulasan tersebut termasuk dalam kelas yang berlabel positif atau negatif. Kelas positif mencakup sanjungan, kepuasan, kesenangan, dan kebahagiaan dari pengguna aplikasi, sedangkan kelas negatif mencakup ketidakpuasan, ketidaksetujuan, dan ketidaksenangan [13].

## 2.3 Preprocessing

*Preprocessing* adalah tahap awal dalam persiapan data sebelum diproses lebih lanjut[14]. Dataset akan diproses melalui serangkaian tahapan preprocessing teks termasuk *cleaning* (Membersihkan), *case folding* (mengganti huruf kapital), *tokenizing* (tokenisasi), *Stopword*, dan *stemming*. Melalui tahapan ini, dataset akan tersusun dengan bersih, yang akan mempermudah proses sistem [15]. Tahap *preprocessing* dapat dibagi menjadi beberapa proses yaitu:

### a. Cleaning

Merupakan proses menghapus semua karakter dalam teks yang bukan huruf alfabet, dengan tujuan mengurangi karakter atau simbol yang tidak diinginkan atau tidak bermakna dalam analisis sentimen [16].

### b. Case Folding

Merupakan tahapan di mana huruf kapital dalam teks diubah menjadi huruf kecil secara menyeluruh, sehingga terjadi penyamarataan huruf [17].

### c. Tokenizing

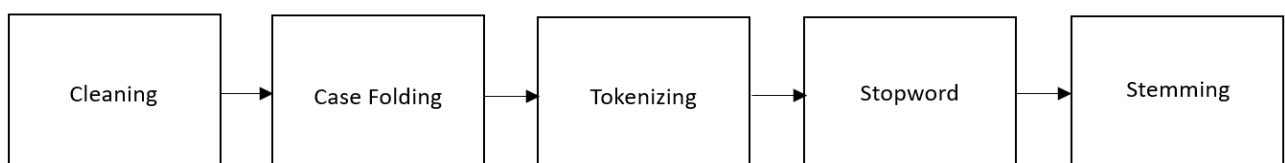
Adalah tahap di mana kalimat dalam data dipecah menjadi beberapa bagian kata yang lebih kecil, agar mudah untuk diproses [9].

### d. Stopword

Merupakan proses untuk mengidentifikasi daftar kata yang tidak mewakili isi dalam sebuah dokumen teks. Proses *Stopword* berfungsi untuk menyaring kata-kata atau istilah yang tidak bermakna. [15].

### e. Stemming

Merupakan proses untuk mengidentifikasi sebuah kata dasar dengan menghapus imbuhan yang ada pada kata tersebut [18].



**Gambar 3.** Proses *Preprocessing* Dataset

Pada gambar 3 merupakan tahapan proses *preprocessing* dataset, mulai dari *cleaning* hingga *stemming*.

## 2.4 Implementasi Algoritma Naïve Bayes

Naïve Bayes adalah metode klasifikasi probabilistik yang sederhana tentang perhitungan semua peluang dengan cara mengumpulkan berbagai kombinasi dan terkait nilai frekuensi dari sebuah basis data yang diperoleh [19]. Teorema ini diajukan oleh ilmuwan Inggris, dengan nama Thomas Bayes, yang menjelaskan tentang prediksi dan penjelasan tentang peluang yang akan muncul di masa depan yang didasarkan oleh pengalaman di masa lalu [20]. Algoritma Naïve Bayes diterapkan untuk memberikan bobot pada setiap ulasan yang telah melewati tahap *preprocessing* data. Ketika frekuensi kemunculan kata-kata yang serupa yang dipilih untuk pembobotan meningkat, nilai bobotnya juga akan mengalami peningkatan seiringnya.

## 2.5 Evaluasi

Tahap berikutnya yaitu tahapan Evaluasi dengan melibatkan penggunaan metode *Confusion Matrix* untuk menghitungnya. *Confusion Matrix* merupakan sebuah matriks yang dipakai untuk menggambarkan performa dari sistem klasifikasi berbasis data [21]. Pada tahap ini merupakan tahap perhitungan untuk menentukan nilai dari akurasi, presisi dan terakhir adalah *recall*. Berikut perumusan untuk menilai tingkat akurasinya:

$$Accuracy = \frac{(TP+TN)}{(TP+TN+FP+FN)} \quad (1)$$

$$Precision = \frac{TP}{(TP+FP)} \quad (2)$$

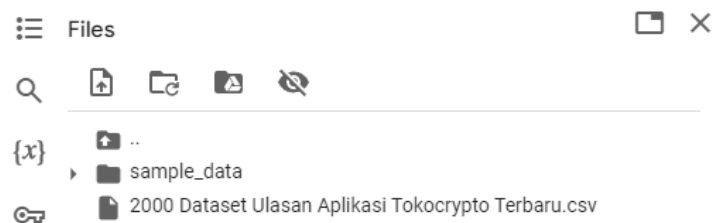
$$Recall = \frac{TP}{(TP+FN)} \quad (3)$$

Ada empat konsep yang mewakili hasil klasifikasi. Ini meliputi *True Positive* (TP), *True Negative* (TN), *False Positive* (FP), dan *False Negative* (FN) [22].

# 3. HASIL DAN PEMBAHASAN

## 3.1 Scraping Data

Pengumpulan dataset ulasan pengguna terhadap aplikasi Tokocrypto di *Google Play Store* dilakukan dengan metode *Web Scraping* melalui *Google Colaboratory* memakai bahasa pemrograman *Python*, dataset yang diperoleh merupakan data ulasan febuari 2024. Jumlah total dataset yang dikumpulkan adalah 2000 ulasan dalam bahasa Indonesia, yang mewakili berbagai sentimen yang berbeda. Berikut adalah hasil proses *Scraping Data* ulasan pengguna aplikasi Tokocrypto menggunakan *Google Colaboratory*:



Gambar 4. Hasil Proses *Scraping Data* menggunakan *Google Colaboratory*

Pada gambar 4 merupakan hasil proses *Scraping Data* menggunakan *Google Colaboratory* dengan bahasa pemrograman *Python* dan lokasi tujuan dataset yang di *scraping* adalah 'com.binance.cloud.tokocrypto'. Hasil proses tersebut berisi dataset yang berjumlah 2000 ulasan dalam format file .CSV, bahasa yang digunakan adalah Indonesia, dan penyortiran dataset berdasarkan data yang terbaru.

## 3.2 Pelabelan Data

Dataset yang berhasil diperoleh dari *scraping data* sebelumnya, selanjutnya dataset tersebut masuk ke tahap proses *labeling* dataset. Proses *labeling* dataset akan dikerjakan secara manual dengan cara memberikan label pada setiap ulasan. Proses *labeling* dataset bertujuan untuk mengelompokkan data ke dalam beberapa kategori sentimen. Pelabelan dataset ini akan dibagi menjadi dua jenis kategori sentimen, yaitu kategori sentimen dengan label positif dan label negatif, label yang ditandai positif akan berisi ulasan baik mengenai aplikasi Tokocrypto berupa pujian, kepuasan dan persetujuan, lalu untuk label yang ditandai negatif akan berisi ulasan buruk mengenai aplikasi Tokocrypto berupa kritik, saran, ketidakpuasan dan ketidaksetujuan.

Tabel 1. Hasil Pelabelan Dataset

No	Ulasan	Label
----	--------	-------

1	Aplikasi tolol	Negatif
2	Mudah digunakan	Positif
3	Penarikan mahal	Negatif
4	Mantap	Positif
5	Sangat puas	Positif

Pada tabel 1 memperlihatkan hasil dari proses pelabelan data. Ulasan berupa pujian, kepuasan akan diberi label positif sedangkan ulasan berupa ketidakpuasan, kritik dan saran akan diberi label negatif.

### 3.3 Preprocessing

Pada tahap proses *preprocessing* dataset, perangkat lunak RapidMiner digunakan sebagai alat bantu untuk mengolah dataset agar lebih terstruktur dan bersih. Berikut beberapa tahapan *preprocessing*:

- Tahap awal *preprocessing* adalah dengan melakukan proses *cleaning* pada dataset ulasan, proses ini bertujuan untuk menghilangkan emoji, simbol, angka dan tautan link.



Gambar 5. Proses *Cleaning Data*

Pada gambar 5 merupakan proses awal dari *preprocessing* yaitu *cleaning data* menggunakan dataset yang sebelumnya sudah beri label positif dan negatif setiap ulasannya. Proses *preprocessing* ini menggunakan tools RapidMiner. Pada proses *cleaning data* diperlukan operator *Replace* dan *Trim* yang akan diintegrasikan ke dalam operator *Subprocess*.

Tabel 2. Hasil Proses *Cleaning Data*

Data sebelum proses Cleaning	Data sesudah proses Cleaning
Habis upgrade tambah jelek, susah masuk /login. SEGERA HAPUS APLIKASI INI. SANGAT MERUGIKAN. POTONGAN KEMAHALAN LOGIN SUSAH APLIKASI SAMPAH	Habis upgrade tambah jelek susah masuk login SEGERA HAPUS APLIKASI INI SANGAT MERUGIKAN POTONGAN KEMAHALAN LOGIN SUSAH APLIKASI SAMPAH

Pada tabel 2 memperlihatkan hasil dari proses *cleaning data* menggunakan operator *Replace*. Pada teks ulasan di atas sebelum data di *cleaning* terdapat simbol atau karakter seperti ./!/ lalu data tersebut dibersihkan sehingga simbol atau karakter sebelumnya menghilang dan hanya menyisakan kata.

- Tahap selanjutnya setelah proses *cleaning* adalah proses *case folding*. Pada proses ini, tujuannya adalah untuk mengonversi atau mengubah semua huruf kapital atau besar dalam dokumen teks menjadi huruf kecil secara keseluruhan.

Tabel 3. Hasil Proses *Case Folding Data*

Data sebelum proses Case Folding	Data sesudah proses Case Folding
Habis upgrade tambah jelek susah masuk login SEGERA HAPUS APLIKASI INI SANGAT MERUGIKAN POTONGAN KEMAHALAN LOGIN SUSAH APLIKASI SAMPAH	habis upgrade tambah jelek susah masuk login segera hapus aplikasi ini sangat merugikan potongan kemahalan login susah aplikasi sampah

Pada tabel 3 memperlihatkan hasil dari proses *Case Folding Data*. Pada proses ini menggunakan operator *Transform Case*, proses ini mengubah teks ulasan pada kolom kiri yang sebelumnya terdapat huruf kapital dan huruf kecil menjadi huruf kecil secara menyeluruh, hasil prosesnya dapat terlihat pada kolom kanan.

- Tahap berikutnya adalah proses *tokenizing*, proses ini bertujuan untuk memecah kalimat dalam data menjadi potongan kata yang lebih kecil.

Tabel 4. Hasil Proses *Tokenizing Data*

Data sebelum proses Tokenizing	Data sesudah proses Tokenizing
habis upgrade tambah jelek susah masuk login segera hapus aplikasi ini sangat merugikan potongan kemahalan login susah aplikasi sampah	habis, upgrade, tambah, jelek, susah, masuk, login, segera, hapus, aplikasi, ini, sangat, merugikan, potongan, kemahalan, login, susah, aplikasi, sampah

Pada tabel 4 memperlihatkan hasil dari proses *Tokenizing Data*. Pada proses ini menggunakan operator *Tokenize*, pada kolom kiri teks ulasan masih dalam bentuk sebuah kalimat, lalu kalimat tersebut di proses dengan *Tokenize* sehingga

ulasan yang awalnya dalam bentuk kalimat sekarang di pecah menjadi kata-kata yang dipisahkan menggunakan tanda baca, hasil prosesnya dapat terlihat pada kolom kanan.

- d. Tahap berikutnya adalah proses *stopword*, Proses ini bertujuan untuk menghapus daftar kata yang tidak mencerminkan isi dalam dokumen teks. *Stopword* digunakan untuk menghapus daftar kata atau frasa yang tidak memiliki makna, berdasarkan kamus *stopword* bahasa Indonesia yang tersedia di *website* [www.kaggle.com](http://www.kaggle.com).

**Tabel 5.** Hasil Proses *Stopword* Data

Data sebelum proses <i>Stopword</i>	Data sesudah proses <i>Stopword</i>
habis, upgrade, tambah, jelek, susah, masuk, login, segera, hapus, aplikasi, ini, sangat, merugikan, potongan, kemahalan, login, susah, aplikasi, sampah	habis, upgrade, jelek, susah, masuk, login, hapus, aplikasi, merugikan, potongan, kemahalan, login, susah, aplikasi, sampah

Pada tabel 5 memperlihatkan hasil dari proses *Stopword Data*. Pada proses ini menggunakan operator *Filter Stopwords (Dictionary)* dengan menggunakan kamus *stopwords* bahasa Indonesia dari *website* <https://www.kaggle.com/oswinrh/indonesian-stoplist> sebagai *list stopwords* nya. Proses ini dilakukan untuk menghapus teks atau kata yang tidak bermakna, terlihat perbedaan dari teks sebelum di proses *stopword* pada kolom kiri dan hasil dari teks yang sudah melalui proses *stopword* pada kolom kanan.

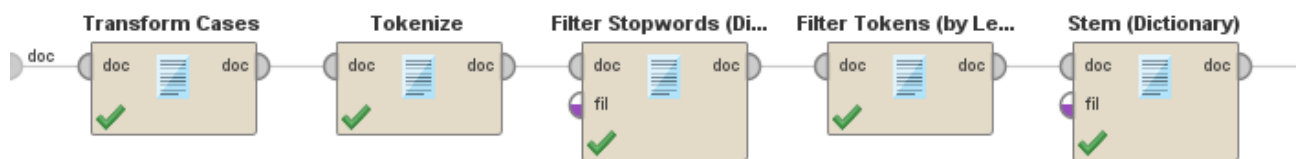
- e. Tahap selanjutnya adalah proses *Filter tokens by length* dan *stemming*, Proses *filter tokens by length* diperlukan untuk membatasi jumlah kata dalam data ulasan sesuai dengan batas minimal dan maksimal. Batasan yang digunakan adalah minimal 4 karakter dan maksimal 25 karakter. Selain itu, dilakukan juga *stemming* untuk mengubah kata yang didalamnya terdapat imbuhan menjadi kata dasar, dengan mengacu pada kamus *stemming* bahasa Indonesia.

**Tabel 6.** Hasil Proses *Filter by length* dan *Stemming*

Data sebelum proses <i>Filter by length</i> dan <i>Stemming</i>	Data sesudah proses <i>Filter by length</i> dan <i>Stemming</i>
habis, upgrade, jelek, susah, masuk, login, hapus, aplikasi, merugikan, potongan, kemahalan, login, susah, aplikasi, sampah	habis, upgrade, jelek, susah, masuk, login, hapus, aplikasi, rugi, potong, mahal, sampah

Pada tabel 6 menunjukkan hasil dari proses *Filter by length* dan *Stemming*. Pada proses ini, digunakan operator *Filter tokens (by length)* untuk menghapus kata-kata yang memiliki jumlah karakter sekurang-kurangnya adalah 4 dan yang melebihi dari 25 karakter. Selain itu, untuk proses *Stemming*, digunakan operator *Stem (Dictionary)* untuk mengubah kata yang didalamnya terdapat imbuhan menjadi kata dasar, dengan menggunakan kamus *stemming* bahasa Indonesia. Terlihat perbedaan antara teks sebelum dan setelah proses *Filter by length* dan *Stemming*, di kolom kiri dan kanan tabel.

Process Documents from Data



**Gambar 6.** Proses *Preprocessing Data*

Pada gambar 6 menggambarkan proses *Preprocessing Data*. Pada tahap *Preprocessing Data*, dataset yang sebelumnya telah melewati proses *cleaning* akan diproses melalui beberapa operator. Ini termasuk *Transform Cases*, *Tokenize*, *Filter Stopwords*, *Filter by length*, dan *Stem*, yang akan diintegrasikan ke dalam operator *Process Documents from Data*.

**3.4 Pembobotan TF-IDF**

Dalam tahap awal implementasi, dilakukan pembobotan menggunakan metode TF - IDF. TF - IDF adalah sebuah proses dengan menggunakan algoritma untuk menilai bobot pada setiap kata yang sering digunakan dalam sebuah kalimat [23]. Proses ini bertujuan untuk menentukan bobot nilai dari sebuah kata dengan cara menggabungkan dua konsep (TF) dan (IDF) menggunakan operator *Process Documents from Data* pada RapidMiner.

**Tabel 7.** Hasil Proses Pembobotan TF - IDF

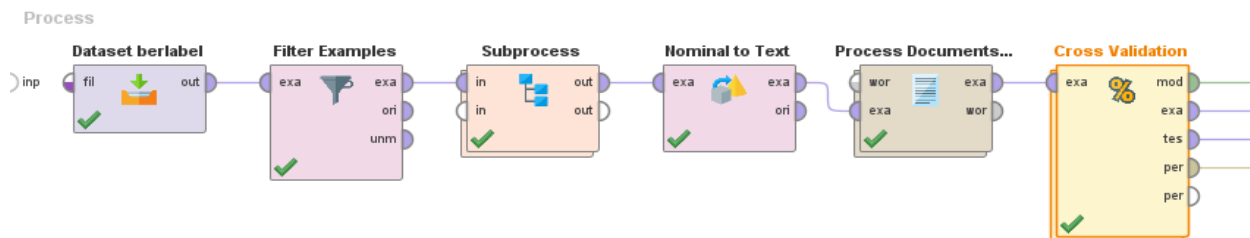
No	Text	abis	account	ada	adain
1	loginnya susah beberapa kali	0	0	0	0
2	salah nyerah abis beli uninstall	0.349	0	0	0
3	penjualan instan jajan gagal	0	0	0	0
4	aplikasi bagus mudah pakai	0	0	0	0
5	hallo daftar udah lumayan pake	0	0	0	0

6	bang adain dana dong bank	0	0	0	0.378
7	diperbarui susah verifikasi	0	0	0	0

Pada tabel 7 memperlihatkan hasil proses dari pembobotan TF - IDF, hasil dari proses ini adalah setiap kata memiliki bobot nilai berdasarkan frekuensi kemunculannya.

### 3.5 Implementasi Algoritma Naïve Bayes

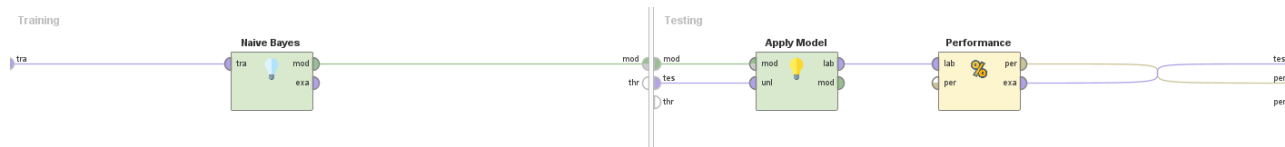
Dalam tahap implementasi algoritma Naïve Bayes, data yang telah diproses sebelumnya melalui *preprocessing* dan pembobotan *Term Frequency-Inverse Document Frequency* akan diklasifikasikan menggunakan metode Naïve Bayes. Proses klasifikasi data, teknik yang digunakan adalah *cross validation* dengan bobot nilai *k-fold* sebanyak 10. Tujuan dari penggunaan teknik ini adalah untuk membagi data agar menjadi sesuai.



Gambar 7. Proses Klasifikasi Data

Pada gambar 7 merupakan tampilan dari seluruh proses klasifikasi data menggunakan metode Naïve Bayes. Semua operator dihubungkan mulai dari Dataset berlabel hingga *Cross Validation*, hal ini bertujuan untuk melihat Tingkat kinerja dari algoritma tersebut.

Dalam *cross validation*, data diolah menggunakan metode algoritma Naïve Bayes dengan melalui operator Naïve Bayes, kemudian dihubungkan dengan operator *apply model*. Setelah itu, digunakan operator *performance* untuk menampilkan tingkat kinerja dari algoritma tersebut dalam proses klasifikasi data.



Gambar 8. Proses klasifikasi data dengan algoritma Naïve Bayes

Pada gambar 8 merupakan proses klasifikasi data dengan menggunakan operator *Naïve Bayes* di dalam *cross validation* operator.

### 3.6 Evaluasi

Evaluasi pemodelan algoritma Naïve Bayes dibantu perangkat lunak RapidMiner menghasilkan tingkat akurasi 74.22%. *Confusion matrix* menunjukkan nilai 315 untuk *True Positive* (TP), 183 untuk *False Positive* (FP), 72 untuk *False Negative* (FN), dan 419 untuk *True Negative* (TN).

accuracy: 74.21% +/- 3.63% (micro average: 74.22%)

	true negatif	true positif	class precision
pred. negatif	419	72	85.34%
pred. positif	183	315	63.25%
class recall	69.60%	81.40%	

Gambar 9. Hasil Evaluasi pemodelan algoritma Naïve Bayes

Pada gambar 9 merupakan hasil evaluasi pemodelan algoritma Naïve Bayes pada proses klasifikasi data menggunakan perangkat lunak RapidMiner. Berikut menggunakan perumusan untuk mengukur tingkat akurasi:

$$Accuracy = \frac{(TP+TN)}{(TP+TN+FP+FN)} = \frac{(315+419)}{(315+419+183+72)} = \frac{734}{989} = 0,742 \quad (4)$$

$$Precision = \frac{TP}{(TP+FP)} = \frac{315}{(315+183)} = \frac{315}{498} = 0,632 \quad (5)$$

$$Recall = \frac{TP}{(TP+FN)} = \frac{315}{(315+72)} = \frac{315}{387} = 0,814 \quad (6)$$

Berdasarkan perhitungan diatas dengan memakai rumus *confusion matrix*, nilainya sesuai dengan hasil klasifikasi data yang diperoleh menggunakan metode algoritma Naïve Bayes melalui perangkat lunak RapidMiner.

### 3.7 Visualisasi WordCloud

Pada tahap visualisasi wordcloud, data klasifikasi diubah menjadi representasi visual berdasarkan frekuensi kata yang biasanya muncul dalam ulasan pengguna aplikasi Tokocrypto. Visualisasi kata akan ditampilkan sebagai *wordcloud*, dengan tujuan untuk memberikan representasi visual tentang kata-kata yang biasanya muncul dan sering digunakan.



Gambar 10. Visualisasi Wordcloud

Pada gambar 10 merupakan bentuk visualisasi *wordcloud* dari kata yang sering muncul menggunakan perangkat lunak RapidMiner.

## 4. KESIMPULAN

Berdasarkan hasil dari penelitian ini, didapatkan sebuah kesimpulan, proses klasifikasi data menggunakan metode *Naive bayes* pada aplikasi Tokocrypto diawali dengan melakukan scraping data menggunakan teknik web scraping di *google colab*. Data yang diperoleh merupakan data ulasan febuari 2024 yang berjumlah sebanyak 2000 ulasan dan setelah dilakukan proses *preprocessing* data bersih yang diperoleh sebanyak 1000 dengan jumlah 396 setimen positif dan 604 sentimen negatif. Lalu setelah data bersih dilakukan pembobotan kata agar mengetahui nilai dari setiap kata, melalui pembobotan TF-IDF. Pada tahap klasifikasi, data diimplementasi algoritma *Naive bayes*, diperoleh nilai 74.22% untuk akurasi, 63.25% untuk presisi dan 81.40% untuk *recall* dari hasil klasifikasi data ulasan aplikasi Tokocrypto di *google play store*. Saran untuk penelitian selanjutnya digunakan algoritma yang berbeda, sehingga hasil dari klasifikasi data bisa dijadikan bahan perbandingan untuk menentukan tingkat akurasi terbaik dari masing-masing algoritma.

## REFERENCES

- [1] S. Sajidin, "LEGALITAS PENGGUNAAN CRYPTOCURRENCY SEBAGAI ALAT PEMBAYARAN DI INDONESIA," *Arena Hukum*, vol. 14, no. 2, pp. 245–267, Aug. 2021, doi: 10.21776/ub.arenahukum.2021.01402.3.
- [2] Rizaldi Azhar, Adi Surahman, and Christina Juliane, "Analisis Sentimen Terhadap Cryptocurrency Berbasis Python TextBlob Menggunakan Algoritma Naïve Bayes," *Jurnal Sains Komputer & Informatika (J-SAKTI)*, vol. 6, no. 1, pp. 267–281, Mar. 2022, Accessed: Feb. 24, 2024. [Online]. Available: <https://tunasbangsa.ac.id/ejurnal/index.php/jsakti>
- [3] A. Sentimen et al., "Sentiment Analysis of Cryptocurrency Exchange Application on Twitter Using Naïve Bayes Classifier Method," *Jurnal Informatika dan Teknologi Informasi*, vol. 20, no. 1, pp. 15–30, 2023, doi: 10.31515/telematika.v20i1.9044.
- [4] P. Aditiya, U. Enri, and I. Maulana, "Analisis Sentimen Ulasan Pengguna Aplikasi Myim3 Pada Situs Google Play Menggunakan Support Vector Machine," *JURIKOM (Jurnal Riset Komputer)*, vol. 9, no. 4, p. 1020, Aug. 2022, doi: 10.30865/jurikom.v9i4.4673.
- [5] Dwi Normawati and Surya Allit Prayogi, "Implementasi Naïve Bayes Classifier Dan Confusion Matrix Pada Analisis Sentimen Berbasis Teks Pada Twitter," *Jurnal Sains Komputer & Informatika (J-SAKTI)*, vol. 5, no. 2, pp. 697–711, Sep. 2021, Accessed: Feb. 24, 2024. [Online]. Available: <https://tunasbangsa.ac.id/ejurnal/index.php/jsakti>
- [6] L. Ardiani, H. Sujaini, and T. Tursina, "Implementasi Sentiment Analysis Tanggapan Masyarakat Terhadap Pembangunan di Kota Pontianak," *Jurnal Sistem dan Teknologi Informasi (Justin)*, vol. 8, no. 2, p. 183, Apr. 2020, doi: 10.26418/justin.v8i2.36776.
- [7] A. Nurian, M. S. Ma'arif, I. N. Amalia, and C. Rozikin, "ANALISIS SENTIMEN PENGGUNA APLIKASI SHOPEE PADA SITUS GOOGLE PLAY MENGGUNAKAN NAÏVE BAYES CLASSIFIER," *Jurnal Informatika dan Teknik Elektro Terapan*, vol. 12, no. 1, Jan. 2024, doi: 10.23960/jitet.v12i1.3631.
- [8] F. Zaini, J. W. Sari, and F. N. Hasan, "ANALYSIS OF PUBLIC SENTIMENT RELATED TO THE FAILURE OF INDONESIA TO HOST U-20 USING MULTINOMIAL NAÏVE BAYES CLASSIFIER," *Jurnal Teknik Informatika (Jutif)*, vol. 4, no. 6, pp. 1409–1418, Dec. 2023, doi: 10.52436/1.jutif.2023.4.6.1209.
- [9] F. Setya Ananto and F. N. Hasan, "Implementasi Algoritma Naïve Bayes Terhadap Analisis Sentimen Ulasan Aplikasi MyPertamina pada Google Play Store," *Jurnal ICT : Information Communication & Technology*, vol. 23, no. 1, pp. 75–80, 2023, [Online]. Available: <https://ejournal.ikmi.ac.id/index.php/jict-ikmi>
- [10] Caesar Rio Anggina Toruan, Novanto Yudistra, and Rizal Setya Perdana, "Analisis Sentimen Tokocrypto pada Twitter menggunakan Metode Long Short-Term Memory," *Jurnal Pengembangan Teknologi Informasi dan Ilmu Komputer*, vol. 7, no. 2, pp. 719–726, Feb. 2023, Accessed: Feb. 23, 2024. [Online]. Available: <http://j-ptiik.ub.ac.id/>
- [11] D. Rudini, D. Gita Purnama, and A. Achmad Khan, "PENGGUNAAN TEKNIK WEB SCRAPING DALAM APLIKASI PENGAMBILAN DATA DARI GOOGLE MAPS UNTUK MENUNJANG DIGITAL MARKETING," *Lentera: Multidisciplinary Studies*, vol. 2, no. 1, pp. 10–19, 2023, [Online]. Available: <https://lentera.publikasiku.id/index.php>

- [12] Ni Putu Gita Naraswati, Delvira Cindy Rosmilda, Dinda Desinta, Fadhilatul Khairi, Riska Damaiyanti, and Rani Nooraeni, "Analisis Sentimen Publik dari Twitter Tentang Kebijakan Penanganan Covid-19 di Indonesia dengan Naive Bayes Classification," *SISTEMASI: Jurnal Sistem Informasi*, vol. 10, no. 1, pp. 222–238, Jan. 2021, Accessed: Feb. 24, 2024. [Online]. Available: <http://sistemasi.ftik.unisi.ac.id/>
- [13] H. Parasian Doloksaribu and Y. T. Samuel, "KOMPARASI ALGORITMA DATA MINING UNTUK ANALISIS SENTIMEN APLIKASI PEDULILINDUNGI," *JURNAL TEKNOLOGI INFORMASI(JTI) Jurnal Keilmuan dan Aplikasi Bidang Teknik Informatika*, vol. 16, no. 1, 2022, doi: 10.47111/JTI.
- [14] A. Rhamadanti, A. Rifa'i, F. Dikananda, and K. Anam, "ANALISIS SENTIMEN PADA ULASAN ACCESS BY KERETA API INDONESIA DENGAN K-NEAREST NEIGHBOR," *Jurnal Informatika dan Teknik Elektro Terapan*, vol. 12, no. 1, pp. 2830–7062, 2024, doi: 10.23960/jitet.v12i1.3691.
- [15] E. Yuniar, D. S. Utsalinah, and D. Wahyuningsih, "Implementasi Scrapping Data Untuk Sentiment Analysis Pengguna Dompok Digital dengan Menggunakan Algoritma Machine Learning," *Jurnal Janitra Informatika dan Sistem Informasi*, vol. 2, no. 1, pp. 35–42, Apr. 2022, doi: 10.25008/janitra.v2i1.145.
- [16] A. Witanti, B. Yogyakarta Ji Raya Wates-Jogjakarta, K. Sedayu, K. Bantul, and D. Istimewa Yogyakarta, "ANALISIS SENTIMEN MASYARAKAT TERHADAP VAKSINASI COVID-19 PADA MEDIA SOSIAL TWITTER MENGGUNAKAN ALGORITMA SUPPORT VECTOR MACHINE (SVM)," *Jurnal Sistem Informasi dan Informatika (Simika) P-ISSN*, vol. 5, pp. 2622–6901, 2022.
- [17] Yuyun, Nurul Hidayah, and Supriadi Sahibu, "Algoritma Multinomial Naïve Bayes Untuk Klasifikasi Sentimen Pemerintah Terhadap Penanganan Covid-19 Menggunakan Data Twitter," *Jurnal RESTI (Rekayasa Sistem dan Teknologi Informasi)*, vol. 5, no. 4, pp. 820–826, Aug. 2021, doi: 10.29207/resti.v5i4.3146.
- [18] Dicki Nugraha and Dudih Gustian, "Analisis Sentimen Penggunaan Aplikasi Transportasi Online Pada Ulasan Google Play Store dengan Metode Naive Bayes Classifier," *KESATRIA: Jurnal Penerapan Sistem Informasi (Komputer & Manajemen)*, vol. 5, no. 1, pp. 326–335, Jan. 2024, Accessed: Feb. 24, 2024. [Online]. Available: <https://tunasbangsa.ac.id/pkm/index.php/kesatria/index>
- [19] P. Ayu, W. Purnama, and T. A. Putra, "Klasifikasi Penjualan Produk Menggunakan Algoritma Naive Bayes pada Konter HP Bayu Cell," *Remik: Riset dan E-Jurnal Manajemen Informatika Komputer*, vol. 8, no. 1, pp. 286–292, 2024, doi: 10.33395/remik.v8i1.13207.
- [20] M. Afriansyah, J. Saputra, V. Yoga Pudya Ardhana, Y. Sa, and U. Qamarul Huda Badaruddin, "ALGORITMA NAIVE BAYES YANG EFISIEN UNTUK KLASIFIKASI BUAH PISANG RAJA BERDASARKAN FITUR WARNA," *Hal. 236 Journal of Information Systems Management and Digital Business (JISMDB)*, vol. 1, no. 2, pp. 236–248, 2024, Accessed: Feb. 24, 2024. [Online]. Available: <https://journal.ppmi.web.id/index.php/jismdb>
- [21] D. A. Prawinata, U. Pembangunan, N. Veteran, and J. Timur, "Analisis Sentimen Kendaraan Listrik Pada Twitter Menggunakan Metode Long Short Term Memory Ani Dijah Rahajoe I Gede Susrama Mas Diyasa," vol. 2, no. 1, pp. 300–313, 2024, doi: 10.59841/saber.v2i1.857.
- [22] M. N. Hidayat and R. Pramudita, "Analisis Sentimen Terhadap Pembelajaran Secara Daring Pasca Pandemi Covid-19 Menggunakan Metode IndoBERT," *Information Management for Educators and Professionals*, vol. 8, no. 2, pp. 161–170, 2023, Accessed: Feb. 24, 2024. [Online]. Available: <https://ejournal-binainsani.ac.id/index.php/IMBI>
- [23] S. Riyadi, "Analisis Sentimen Opini Masyarakat Terhadap Stadion Jakarta Internasional Stadium (JIS) Pada Twitter Dengan Perbandingan Metode Naive Bayes Dan Support Vector Machine," *Jurnal Sains dan Teknologi*, vol. 5, no. 3, pp. 801–809, 2024, doi: 10.55338/saintek.v5i1.2790.