PENERAPAN NAÏVE BAYES DENGAN OPTIMASI *INFORMATION GAIN* DAN SMOTE UNTUK ANALISIS SENTIMEN PENGGUNA APLIKASI CHATGPT

Haikal Hidayatullah, Purwantoro, Yuyun Umaidah
Program Studi Informatika, Fakultas Ilmu Komputer, Universitas Singaperbangsa Karawang
Jl. HS. Ronggo Waluyo, Telukjambe Timur, Karawang, Jawa Barat, Indonesia - 41361
1910631170192@student.unsika.ac.id

ABSTRAK

ChatGPT yang dikembangkan oleh OpenAI yaitu aplikasi *chatbot* dengan rekor pertumbuhan tercepat ketika mencapai 100 juta pengguna aktif dua bulan setelah aplikasi ini diluncurkan. Keberhasilan aplikasi ini menyebabkan timbulnya berbagai macam komentar dari penggunanya di media sosial seperti Twitter. Komentar yang diberikan pengguna dapat dimanfaatkan oleh pengembang untuk meningkatkan kualitas aplikasi sesuai dengan kebutuhan pengguna karena komentar dapat berisi ulasan atau permasalahan ketika pengguna menggunakan aplikasi. Akan tetapi, mengolah informasi dari sejumlah besar komentar secara manual tidak memungkinkan dilakukan. Oleh karena itu, penelitian ini menerapkan analisis sentimen dengan pendekatan *machine learning* menggunakan algoritma Naïve Bayes dengan optimasi *Information Gain* dan SMOTE untuk mengolah informasi dari komentar-komentar pengguna. Data yang digunakan pada penelitian ini terdiri dari 721 komentar pengguna aplikasi ChatGPT yang dikumpulkan dari media sosial Twitter selama periode bulan Maret 2023 hingga April 2023. Hasil evaluasi menunjukkan bahwa model Opt2 menjadi model terbaik dengan nilai akurasi sebesar 87.20% dan nilai *F1-score* sebesar 84.74%. Selain itu, hasil evaluasi menunjukkan bahwa metode optimasi *Information Gain* dan SMOTE mampu meningkatkan nilai akurasi, *recall* dan *F1-score* dengan rata-rata peningkatan yang didapat sebesar 6.25% untuk nilai akurasi, 23.9% untuk nilai *recall*, dan 25.44% untuk nilai *F1-score*. Temuan penelitian ini menunjukkan bahwa meskipun aplikasi ChatGPT membantu pengguna dalam menyelesaikan tugas-tugas pengguna, pengembang perlu meningkatkan kualitas respons jawaban dari aplikasi dan menangani *error* yang terjadi saat pengguna menggunakan ChatGPT.

Kata kunci: Analisis Sentimen, Naïve Bayes, *Information Gain*, SMOTE, ChatGPT

1. PENDAHULUAN

Aplikasi *chatbot* yang dikenal sebagai ChatGPT (*Chat Generative Pre-Trained Transformer*) telah mencatat rekor pertumbuhan pengguna tercepat dengan 100 juta pengguna aktif pada bulan Januari 2023 tepat dua bulan setelah peluncurannya pada bulan November 2022 [1]. Menurut Hassani & Silva [2], ChatGPT memiliki kelebihan utama dalam hal fleksibilitas dan kemudahan penggunaannya. Namun, penelitian tersebut juga menyoroti kelemahan ChatGPT, yaitu ketergantungannya pada data pelatihan yang berjumlah besar. Walaupun model dapat diadaptasi untuk digunakan dengan kumpulan data yang lebih kecil, ada kemungkinan bahwa kinerjanya terbatas dibandingkan dengan model lain yang secara khusus dirancang untuk pengaturan sumber daya rendah, seperti ALBERT yang dikembangkan oleh Google [2]. Kelebihan dan kekurangan dari aplikasi ChatGPT menyebabkan banyaknya komentar pengguna berdasarkan pengalaman penggunaanya di berbagai media sosial seperti Twitter [3]. Komentar-komentar tersebut memiliki nilai informasi yang signifikan bagi pengembang aplikasi dalam usaha untuk meningkatkan kualitas perangkat lunak yang mereka miliki [4]. Hal ini dikarenakan komentar-komentar tersebut dapat berisi ulasan atau umpan balik dari pengguna mengenai masalah-masalah yang mereka alami saat menggunakan suatu aplikasi [3]. Namun, pengolahan informasi dari sejumlah besar komentar secara manual tidak memungkinkan [5]. Permasalahan ini dapat diatasi dengan menerapkan analisis sentimen menggunakan pendekatan *machine learning* untuk mengolah informasi dari komentar pengguna dengan lebih efisien [6].

Penelitian tentang analisis sentimen telah dilakukan sebelumnya oleh Hidayat, dkk. [7] melakukan penelitian mengenai analisis sentimen terkait pengembangan Pulau Rinca menggunakan algoritma *Support Vector Machine* (SVM) dan *Logistic Regression* (LR). Penelitian tersebut menghasilkan model dengan akurasi sebesar 87% dan F1-score sebesar 81%. Namun, penelitian ini memiliki kekurangan dalam penggunaan *dataset* yang tidak seimbang, sehingga dapat mempengaruhi hasil yang diperoleh. Selanjutnya, Fitri, dkk. [8] melakukan perbandingan model terhadap kampanye "anti-LGBT" di Indonesia dengan menggunakan algoritma Naïve Bayes (NB), *Decision Tree* (DT), dan *Random Forest* (RF). Hasilnya menunjukkan bahwa algoritma NB memiliki kinerja yang paling optimal, dengan akurasi sebesar 83.43%. Namun, penelitian ini belum dapat memproses kata-kata tidak baku menjadi kata baku, yang dapat mempengaruhi nilai akurasi yang diperoleh. Kemudian, Bani Baker, dkk. [9] melakukan deteksi penyebaran penyakit influenza melalui ulasan di Twitter menggunakan algoritma NB, SVM, DT, dan *K-Nearest Neighbor* (KNN) dalam pengembangan

modelnya. Penelitian ini menunjukkan bahwa model NB memiliki performa terbaik di antara algoritma lainnya, dengan akurasi mencapai 83.20%. Namun, prediksi kelas positif dan negatif dari model tersebut tidak optimal, sehingga diperlukan penambahan metode optimasi guna meningkatkan nilai akurasi dan *F1-score* yang diperoleh. Walaupun demikian, NB didasarkan pada Teorema Bayes yang mengasumsikan bahwa setiap fitur saling independen (naif). Asumsi naif ini memungkinkan model Naive Bayes untuk melakukan perhitungan dengan cepat dan efisien [10]. Pada penelitian Özkurt, dkk. [11] NB menunjukkan kinerja komputasi yang lebih tinggi dengan waktu eksekusi sekitar 2 detik dibandingkan dengan model lain seperti SVM, RF, dan KNN dalam melakukan analisis sentimen terkait suatu produk.

Berdasarkan pembahasan tersebut, penulis mengusulkan pengembangan model klasifikasi teks menggunakan algoritma Naïve Bayes untuk mengidentifikasi layanan-layanan di ChatGPT yang perlu ditingkatkan atau dipertahankan. Pada penelitian ini, akan dilakukan *normalize* atau pengubahan kata-kata tidak baku menjadi baku sebagai proses yang dilakukan [12]. Selanjutnya, metode optimasi *Information Gain* diterapkan dengan tujuan untuk meningkatkan nilai akurasi dan *F1-score*, sejalan dengan penelitian yang dilakukan oleh Negara, dkk. [13] yang menggunakan metode optimasi seleksi fitur *Information Gain* dalam analisis sentimen terhadap maskapai penerbangan menggunakan algoritma NB. Hasil penelitian tersebut menunjukkan bahwa penggunaan *Information Gain* dapat meningkatkan rata-rata nilai akurasi dan *F1-score* sebesar 6.6% untuk akurasi dan 12.9% untuk *F1-score*. Kemudian penelitian ini menerapkan SMOTE untuk mengatasi ketidakseimbangan *dataset* yang digunakan, sebagaimana disebutkan dalam penelitian yang dilakukan oleh Safitri & Muslim [14] yang melakukan klasifikasi terhadap data kredit Germany dengan menggunakan optimasi SMOTE untuk menangani ketidakseimbangan data. Hasil penelitian ini menunjukkan bahwa metode SMOTE dapat meningkatkan nilai akurasi sebesar 1.918%. Manfaat penelitian ini diharapkan dapat memaksimalkan pengembangan aplikasi ChatGPT berdasarkan komentar pengguna di Twitter.

2. TINJAUAN PUSTAKA

2.1. *Chat Generative Pre-Trained Transformer* (ChatGPT)

Chat Generative Pre-Trained Transformer (ChatGPT) merupakan sebuah model bahasa alami yang termasuk dalam kategori *Large Language Model* (LLM) yang dikembangkan oleh OpenAI dan diluncurkan pada bulan November tahun 2022 [1], [15]. LLM merupakan jenis algoritma kecerdasan buatan baru yang dilatih untuk memprediksi kemungkinan urutan kata tertentu berdasarkan konteks kata sebelumnya [15]. ChatGPT dirancang dalam format dialog yang memungkinkannya untuk merespons pertanyaan-pertanyaan sebelumnya dari pengguna, menolak permintaan yang tidak pantas, mengakui kesalahan, dan menggugah premis yang keliru [16].

2.2. Analisis Sentimen

Analisis sentimen adalah proses di mana informasi diekstraksi dari sebuah entitas dan secara otomatis mengidentifikasi subjektivitas yang terkandung di dalamnya. Tujuan utamanya adalah untuk menentukan apakah suatu teks memiliki polaritas positif, negatif, atau netral. Proses ini melibatkan penerapan metode pemrosesan teks khusus yang memungkinkan pengolahan teks secara otomatis [17].

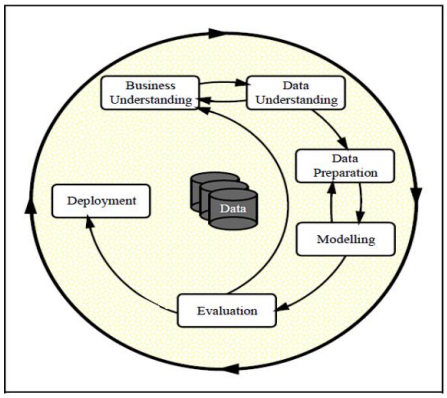
2.3. *Text Mining*

Text mining merupakan proses interaktif yang menggunakan berbagai alat analisis untuk memahami kumpulan dokumen dalam periode tertentu [18]. Tujuannya adalah untuk mengekstrak informasi dari data dokumen dengan mengidentifikasi dan mengeksplorasi pola yang ada. Namun, sebelum dapat dilakukan analisis, data dalam text mining perlu melalui tahap preprocessing agar menjadi lebih terstruktur karena data tersebut awalnya tidak terstruktur [19]. Tahapan-tahapan *preprocessing* data, sebagaimana dijelaskan oleh A. E. Sari, dkk. [12] dan Fikri, dkk. [20], meliputi:

a. *Data cleansing*, merupakan proses untuk membersihkan elemen-elemen yang tidak relevan pada data.
b. *Case folding*, yaitu proses untuk menyamakan format pada data menjadi *lower case*.
c. *Tokenization*, yaitu proses memisahkan setiap kata dalam kalimat menjadi satuan kata atau yang biasa disebut dengan *token*.
d. *Normalize*, yaitu proses mengubah kata tidak baku menjadi kata baku.
e. *Stopword removal*, merupakan proses untuk menghilangkan kata-kata yang tidak memiliki pengaruh signifikan dalam kalimat.
f. *Stemming*, merupakan proses untuk menghilangkan imbuhan-imbuhan pada kata.

2.4. *Cross Industry Standard Process for Data Mining* (CRISP-DM)

Menurut Wirth & Hipp [21], CRISP-DM adalah sebuah metodologi pemodelan yang memberikan kerangka kerja untuk *data mining* yang dapat diterapkan dalam sektor industri dan teknologi. Singgalen [22] menyebutkan bahwa pendekatan CRISP-DM dapat digunakan untuk mengidentifikasi, memproses, dan mengevaluasi ulasan yang menggambarkan perasaan konsumen terhadap layanan atau komoditas tertentu, dengan menggunakan model yang sesuai dengan karakteristik bisnis yang relevan. Tahapan-tahapan yang dilakukan dalam pendekatan CRISP-DM, seperti yang dijelaskan oleh Wirth & Hipp [21] dapat dilihat pada Gambar 1.



Gambar 1. Tahapan CRISP-DM [23]

a. *Business understanding* merupakan tahapan yang dilakukan untuk memahami permasalahan dan menentukan tujuan penelitian.
b. *Data understanding* adalah tahapan yang dilakukan untuk memahami data yang dibutuhkan dalam penelitian. Tahapan ini melibatkan proses pengumpulan dan evaluasi kelayakan data.
c. *Data preparation* merupakan tahap pemrosesan data pengumpulan awal atau data mentah menjadi *dataset* akhir yang sudah siap untuk diolah.
d. *Modeling* adalah tahapan pembangunan alur kerja pengolahan data untuk memilih teknik pemodelan dan parameter yang akan digunakan.
e. *Evaluation* merupakan tahapan di mana model yang telah dibangun dievaluasi untuk memeriksa apakah model tersebut telah mencapai tujuan penelitian yang telah ditetapkan.
f. *Deployment* merupakan tahapan dokumentasi keseluruhan tahapan penelitian yang sudah dilakukan, disajikan dalam bentuk laporan atau jurnal publikasi.

2.5. Naïve Bayes

Menurut Alsaeedi & Khan [24] Naïve Bayes adalah metode klasifikasi yang mengandalkan Teorema Bayes dengan asumsi bahwa setiap pasangan fitur dan kelas adalah saling independen. Algoritma Naïve Bayes merupakan metode klasifikasi yang efisien dan sederhana yang digunakan dalam analisis sentimen [13]. Rumus atau persamaan dari Teorema Bayes dapat dihitung menggunakan Persamaan 1 [20].

$$P(M|N) = \frac{P(N|M)P(M)}{P(N)} \quad (1)$$

Keterangan:

N : Data Sampel
M :Hipotesis bahwa data *N* merupakan data dari suatu kelas
$P(M \mid N)$: Peluang terjadinya hipotesis *M*, setelah diberikan kondisi *N* (*posterior probability*)
$P(M)$: Peluang terjadinya hipotesis *M*, setelah diberikan kondisi *N* (*posterior probability*)
$P(N \mid M)$: Peluang terjadinya *N*, setelah diberikan kondisi hipotesis *M*
$P(N)$: Peluang terjadinya *N* (*predictor prior probabilit*)

2.6. *Information Gain*

Information Gain adalah sebuah metode seleksi fitur yang mengukur sejauh mana kehadiran atau ketidakhadiran sebuah kata berpengaruh dalam membuat keputusan klasifikasi yang akurat pada kelas data tertentu. Tujuannya adalah untuk mengurangi dimensi data dan meningkatkan akurasi klasifikasi [13]. Rumus-rumus yang digunakan untuk menghitung nilai *Information Gain* dapat ditemukan pada Persamaan 2, 3, dan 4 [12].

$$Info(D) = -\sum_{i=1}^{c} pi\ log_2(pi) \quad (2)$$

$$Info_A(D) = -\sum_{j=1}^{v} \frac{|D_j|}{|D|} * Info(D_j) \quad (3)$$

$$InfoGain(A) = -Info(D) - |Info(A)| \quad (4)$$

Keterangan:

c : Jumlah partisi dari data *D*
pi : Jumlah sampel yang termasuk dalam kelas i dibandingkan dengan jumlah total sampel
$Info(D)$: Nilai *entropy* dari kelompok data *D*
A : Atribut
$|D|$: Keseluruhan data sampel
$|D_j|$: Jumlah sampel pada bagian *j*
v : Total seluruh bagian atribut *A*
$Info(D_j)$: Nilai e*ntropy* untuk bagian *j*
$Info_A(D)$: Pengukuran seberapa banyak informasi yang bisa didapatkan dari atribut A untuk membantu memprediksi kelompok data *D*
$Info(A)$: Nilai *entropy* dari *A*
$InfoGain(A)$: Nilai *Information Gain* dari *A*

2.7. *Synthetic Minority Over-Sampling Technique* (SMOTE)

Menurut Siringoringo [25], SMOTE adalah sebuah metode yang digunakan untuk mengatasi masalah ketika kita memiliki data yang tidak seimbang, di mana jumlah data pada satu kelompok jauh lebih banyak daripada kelompok lainnya. Masalah ini dapat menyebabkan model yang dikembangkan menjadi terlalu fokus pada kelompok data mayoritas, sehingga tidak dapat memprediksi dengan baik pada kelompok minoritas. SMOTE mampu membuat data tambahan pada kelompok minoritas sehingga kedua kelompok memiliki jumlah data yang seimbang. Hal ini membantu model yang dikembangkan untuk belajar secara adil dan menghindari *overfitting* [26], [27]. SMOTE bekerja dengan cara mencari tetangga terdekat dari setiap data yang termasuk dalam kelompok yang jumlahnya lebih sedikit. Kemudian, SMOTE membuat data baru di

antara data kelompok yang jumlahnya lebih sedikit dan tetangga terdekat yang dipilih secara acak [25], [27].

2.8. *Confusion Matrix*

Confusion Matrix adalah suatu metode yang digunakan untuk mengukur kinerja suatu sistem dengan membandingkan hasil prediksi sistem dengan data aktual yang ada. [28]. *Confusion matrix* dapat dilihat pada Tabel 1.

Tabel 1. *Confusion Matrix*

Predicted Value	***Actual Value***	
	Positive	**Negative**
Positive	*True Positive* (TP)	*False Positive* (FP)
Negative	*False Negative* (FN)	*True Negative* (TN)

Keterangan:
TP : Data positif yang diprediksi positif
TF : Data negatif yang diprediksi negatif
FP : Data negatif yang diprediksi positif
FN : Data positif yang diprediksi negatif

Confusion Matrix menghasilkan *output* berupa *F1-score* yaitu nilai rata-rata harmonik dari *precision* dan *recall*., akurasi (*accuracy*) untuk mengukur sejauh mana metode tersebut mampu memprediksi output dengan benar, *recall* yaitu rasio antara jumlah sampel positif yang terdeteksi dengan benar terhadap jumlah total sampel yang sebenarnya positif, dan presisi (*precision*) yaitu proporsi antara jumlah data positif yang berhasil diklasifikasikan secara akurat oleh sistem klasifikasi dengan jumlah keseluruhan data yang diklasifikasikan sebagai data positif oleh sistem klasifikasi [6], [29]. Setiap *output* ini dinyatakan dalam bentuk persentase, dengan rentang nilai antara 1 hingga 100%. Berikut merupakan rumus yang digunakan untuk menghitung nilai dari masing-masing *output confusion matrix* menurut Wardani & Purbolaksono [30]:

$$F1 - score = 2 * \frac{Recall * Precision}{Recall + Precision} \quad (5)$$

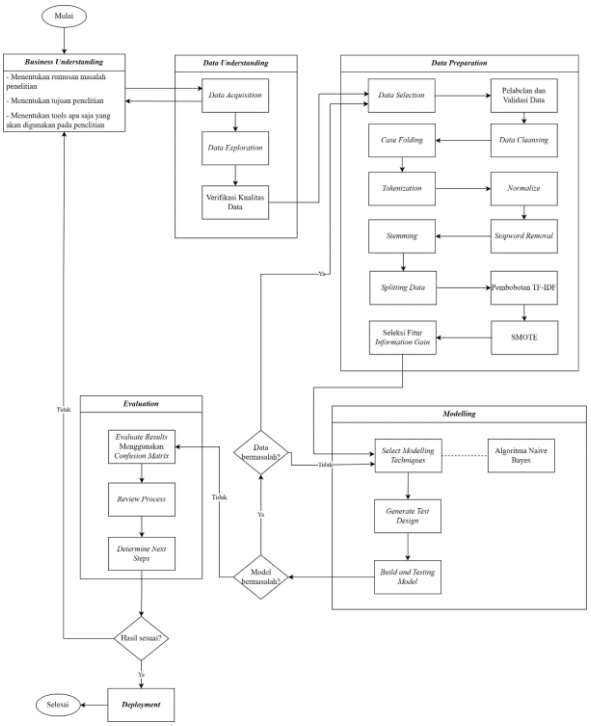
$$Accuracy = \frac{TP + TN}{Total\ Data} \quad (6)$$

$$Recall = \frac{TP}{TP + FN} \quad (7)$$

$$Precision = \frac{TP}{TP + FP} \quad (8)$$

3. METODE PENELITIAN

Metode penelitian ini berdasarkan pada CRISP-DM yang terdiri dari 6 tahapan utama yakni *business understanding*, *data understanding*, *data preparation*, *modeling*, *evaluation*, dan *deployment*. Adapun *flowchart* atau diagram alir yang digunakan pada penelitian ini dapat dilihat pada Gambar 2.



Gambar 2. *Flowchart* penelitian

4. HASIL DAN PEMBAHASAN

4.1. *Business Understanding*

Pemahaman masalah berdasarkan latar belakang yang sudah dibahas sebelumnya adalah kepopuleran, kelebihan, dan kelemahan aplikasi ChatGPT menyebabkan timbulnya banyak komentar pengguna di media sosial Twitter. Banyaknya komentar yang timbul menyebabkan proses pengolahan informasi tidak memungkinkan untuk dilakukan secara manual. Oleh karena itu, tujuan penelitian ini adalah untuk mengatasi permasalahan pengolahan informasi melalui penerapan analisis sentimen dengan pendekatan *machine learning*, dan mengetahui kinerja dari pendekatan yang dilakukan. Penelitian ini menggunakan algoritma Naïve Bayes dengan optimasi *Information Gain* dan SMOTE untuk meningkatkan nilai akurasi dan *F1-score* sekaligus mengatasi *dataset* yang tidak seimbang. Adapun *tools* yang digunakan adalah Google Colaboratory dengan bahasa pemrograman Python dan penggunaan *library* Sastrawi, NLTK, *Scikit-learn*, Pandas, serta Snscrape.

4.2. *Data Understanding*

4.2.1. *Data Acquisition*

Proses *data acquisition* merupakan proses mengumpulkan data. Data yang dikumpulkan pada penelitian ini adalah data komentar pengguna aplikasi ChatGPT, berbahasa Indonesia di media sosial Twitter dalam rentang waktu bulan Maret 2023 sampai dengan bulan April 2023. Pengumpulan data dilakukan dengan bantuan *library* Python Snscrape. Data awal yang dikumpulkan pada penelitian ini adalah sebanyak 15.537 data komentar pengguna.

4.2.2. *Data Exploration*

Proses ini bertujuan untuk mengetahui informasi dari data yang sudah dikumpulkan. Adapun informasi pada data penelitian yang didapatkan adalah sebagai berikut:

a. Semua data yang diperoleh melalui proses pengumpulan data adalah berupa komentar yang berkaitan dengan ChatGPT.
b. Terdapat total 15.537 komentar yang telah dikumpulkan.
c. Semua komentar yang terkumpul dalam data ini menggunakan bahasa Indonesia.

Data akan dikategorikan menjadi dua kategori, yaitu kategori positif dan kategori negatif. Proses pengkategorian ini akan dilakukan secara manual oleh penulis. Selanjutnya, kategori yang telah ditetapkan akan divalidasi oleh seorang ahli tata bahasa untuk memastikan ketepatan klasifikasi yang dilakukan.

4.2.3. Verifikasi Kualitas Data

Tahap verifikasi dilakukan dengan cara memeriksa setiap komentar pengguna secara manual untuk memastikan kesesuaiannya dengan kebutuhan. Hasil dari tahap ini menunjukkan bahwa data yang telah dikumpulkan sesuai dengan kebutuhan, yaitu berupa komentar pengguna tentang ChatGPT. Dengan demikian, tahap selanjutnya dapat dilanjutkan untuk mempersiapkan data sebelum proses klasifikasi dilakukan.

4.3. *Data Preparation*

Tahap *data preparation* dilakukan dengan tujuan mengubah data yang awalnya tidak terstruktur menjadi data yang terstruktur. Hal ini bertujuan untuk mempermudah proses klasifikasi yang akan dilakukan. Pada tahap ini, data akan diatur dan disesuaikan dengan format yang dibutuhkan, sehingga dapat dimanfaatkan secara lebih efisien dalam analisis dan pengklasifikasian. Proses yang dilakukan pada tahapan ini terdapat 12 proses seperti terlihat pada Gambar 2, mulai dari seleksi data (*data selection*) sampai dengan seleksi fitur *Information Gain*.

Proses pertama yaitu seleksi data, proses ini melibatkan pemilihan data komentar pengguna yang mengalami duplikasi dan penghapusan komentar dengan sentimen netral. Data awal yang terdiri dari 15.537 komentar pengguna, terdapat 11.709 komentar yang merupakan duplikasi dan 3.107 komentar dengan sentimen netral. Setelah proses seleksi ini, jumlah total data komentar pengguna ChatGPT pada penelitian ini menjadi 721 komentar.

Proses kedua yaitu pelabelan dan validasi data, proses ini dilakukan secara manual oleh penulis dengan memberikan label positif atau negatif dari 721 data komentar yang digunakan. Setelah itu, hasil pelabelan dilakukan validasi oleh seorang ahli tata bahasa yaitu Ibu Cut Nuraini, S.Pd., M.Pd. Proses ini menghasilkan data dengan label positif sebanyak 512 dan data dengan label negatif sebanyak 209 data.

Proses ketiga *data cleansing* sampai dengan proses kedelapan *stemming* disajikan dalam bentuk tabel yang dapat dilihat pada Tabel 2.

Tabel 2. Hasil *data cleansing* sampai *stemming*

Tahap	Hasil
Komentar Awal	ternyata ChatGPT blm bisa dijadiin konsultan pajak ges. bikin stres doang konsul sama ChatGPT soal PPh 21 :(
Data Cleansing	ternyata ChatGPT blm bisa dijadiin konsultan pajak ges bikin stres doang konsul sama ChatGPT soal PPh
Case Folding	ternyata chatgpt blm bisa dijadiin konsultan pajak ges bikin stres doang konsul sama chatgpt soal pph
Tokenization	['ternyata', 'chatgpt', 'blm', 'bisa', 'dijadiin', 'konsultan', 'pajak', 'ges', 'bikin', 'stres', 'doang', 'konsul', 'sama', 'chatgpt', 'soal', 'pph']
Normalize	['ternyata', 'chatgpt', 'belum', 'bisa', 'dijadiin', 'konsultan', 'pajak', 'ges', 'bikin', 'stres', 'doang', 'konsul', 'sama', 'chatgpt', 'soal', 'pph']
Stopword Removal	['ternyata', 'chatgpt', 'bisa', 'dijadiin', 'konsultan', 'pajak', 'ges', 'bikin', 'stres', 'doang', 'konsul', 'sama', 'chatgpt', 'soal', 'pph']
Stemming	['nyata', 'chatgpt', 'bisa', 'dijadiin', 'konsultan', 'pajak', 'ges', 'bikin', 'stres', 'doang', 'konsul', 'sama', 'chatgpt', 'soal', 'pph']

Proses kesembilan yaitu *splitting data*, proses ini dilakukan dengan cara memisahkan data menjadi *data training* (data latih) untuk pelatihan model dan *data testing* (data uji) untuk menguji model yang dikembangkan. Pembagian data yang digunakan pada penelitian ini dibagi menjadi 5 skenario pembagian data. Adapun hasil dari pembagian data berdasarkan 5 skenario dapat dilihat pada Tabel 3.

Tabel 3. *Splitting data*

Skenario	Data Latih	Data Uji
50:50	360	361
60:40	432	289
70:30	504	217
80:20	576	145
90:10	648	73

Proses kesepuluh yaitu pembobotan TF-IDF, proses ini dilakukan dengan cara memberikan bobot pada setiap kata pada data yang sudah disiapkan sebelumnya. Selanjutnya adalah proses kesebelas penerapan SMOTE, yaitu menyeimbangkan *dataset* penelitian. Kemudian, proses kedua belas yaitu penerapan seleksi fitur *Information Gain* yang bertujuan untuk menyeleksi jumlah fitur hasil pembobotan TF-IDF. Adapun hasil dari ketiga proses yang dilakukan dapat dilihat pada Tabel 4 untuk hasil TF-IDF dan *Information Gain*, serta pada Tabel 5 untuk hasil penerapan SMOTE.

Tabel 4. Hasil TF-IDF dan *Information Gain*

Skenario	TF-IDF	*Information Gain*
50:50	1.587	887
60:40	1.838	1.066
70:30	2.024	1.181
80:20	2.210	1.287
90:10	2.368	1.388

Tabel 5. Hasil penerapan SMOTE

Skenario	Sebelum Penerapan SMOTE		Setelah Penerapan SMOTE	
	Positif	Negatif	Positif	Negatif
50:50	256	104	256	256
60:40	300	132	300	300
70:30	346	158	346	346
80:20	399	177	399	399
90:10	452	196	452	452

4.4. *Modeling*

Proses pertama yang dilakukan pada adalah memilih teknik pemodelan yang akan digunakan, pada penelitian ini menggunakan algoritma Naïve Bayes. Model yang akan dikembangkan pada penelitian terdiri dari 2 model utama yang dibagi lagi menjadi 5 model untuk masing-masing model utama berdasarkan skenario pembagian data, penerapan *Information Gain*, dan SMOTE. Adapun keterangan dari setiap model yang akan dikembangkan dapat dilihat pada Tabel 6.

Tabel 6. Model penelitian

Model Ori	Optimasi	Model Opt	Optimasi
Ori1 (50:50)	×	Opt1 (50:50)	✓
Ori2 (60:40)	×	Opt2 (60:40)	✓
Ori3 (70:30)	×	Opt3 (70:30)	✓
Ori4 (80:20)	×	Opt4 (80:20)	✓
Ori5 (90:10)	×	Opt5 (90:10)	✓

Proses selanjutnya *generate test design*, yaitu memilih *test design* yang digunakan pada penelitian. Penelitian ini menggunakan *confusion matrix* untuk mengukur kinerja model yang dikembangkan. Kemudian, dilanjutkan dengan *training and testing model*, yaitu proses pembangunan sekaligus pengujian setiap model yang dikembangkan. Hasil dari proses ini dievaluasi pada tahapan berikutnya yaitu tahap *evaluation*.

4.5. *Evaluation*

Tahapan ini bertujuan untuk mengevaluasi hasil dari tahap *modeling* yang dilakukan. Proses pertama pada tahapan ini adalah *evaluate result*, yaitu mengevaluasi setiap hasil dari model yang dikembangkan. Adapun hasil dari proses ini dapat dilihat pada Tabel 7.

Tabel 7. Hasil setiap model penelitian

Model Ori (Sebelum Optimasi)			Model Opt (Setelah Optimasi)			Dampak
Model	Pengukuran	Hasil	Model	Pengukuran	Hasil	
Ori1	Akurasi	72.30%	**Opt1**	Akurasi	82.83%	+10.53%
	Recall	52.38%		*Recall*	79.75%	+27.37%
	Precision	85.96%		*Precision*	79.11%	-6.85%
	F1-Score	46.38%		*F1-Score*	79.41%	+33.03%
Ori2	Akurasi	76.12%	**Opt2**	Akurasi	87.20%	+11.08%
	Recall	55.19%		*Recall*	87.55%	+32.36%
	Precision	87.72%		*Precision*	83.09%	-4.63%
	F1-Score	52.41%		*F1-Score*	84.74%	+32.33%
Ori3	Akurasi	79.26%	**Opt3**	Akurasi	83.41%	+4.15%
	Recall	55.88%		*Recall*	81.01%	+25.13%
	Precision	89.34%		*Precision*	77.06%	-10.03%
	F1-Score	54.56%		*F1-Score*	78.59%	+24.03%
Ori4	Akurasi	80.00%	**Opt4**	Akurasi	85.52%	+5.52%
	Recall	54.69%		*Recall*	72.27%	+17.58%
	Precision	89.79%		*Precision*	79.31%	-10.48%
	F1-Score	52.89%		*F1-Score*	78.20%	+25.31%
Ori5	Akurasi	84.93%	**Opt5**	Akurasi	84.93%	0
	Recall	57.69%		*Recall*	69.74%	+12.05%
	Precision	92.25%		*Precision*	74.44%	-17.81%
	F1-Score	59.13%		*F1-Score*	71.62%	+12.49%

Setelah itu, dilakukan *review process* untuk meninjau langkah-langkah yang sudah dilakukan apakah sudah sesuai dengan tujuan penelitian. Kemudian, langkah selanjutnya ditentukan pada proses *determine next steps*. Penelitian ini sudah melakukan keseluruhan tahapan dengan didapatkan model yang dapat melakukan klasifikasi komentar pengguna, sehingga dapat dilanjutkan ke tahapan *deployment*.

4.6. *Deployment*

Tahap ini dilakukan dokumentasi dengan menggambarkan setiap tahapan yang telah dilakukan dalam bentuk laporan penelitian yaitu publikasi jurnal.

Hasil analisis yang diperoleh divisualisasikan dalam bentuk *word cloud* yang menampilkan visualisasi kata-kata yang paling umum muncul dalam komentar dengan label positif (Gambar 3) dan komentar dengan label negatif (Gambar 4).

Gambar 3. *Word cloud* komentar positif

Gambar 4 *Word cloud* komentar negatif

Visualisasi komentar positif menunjukkan bahwa pengguna pengguna menganggap aplikasi ChatGPT sangat membantu dalam pekerjaan mereka, ditunjukkan oleh kata-kata seperti "bantu" dan "kerja" dalam *word cloud*. Namun, komentar dengan sentimen negatif menunjukkan bahwa aplikasi ChatGPT memberikan referensi yang salah dan mengalami *error* ketika digunakan oleh pengguna. Hal ini terlihat dari keberadaan kata-kata seperti "referensi", "salah"", "jurnal", dan "*error*" dalam *word cloud*.

5. KESIMPULAN DAN SARAN

Penelitian ini telah menerapkan analisis sentimen dengan pendekatan *machine learning* menggunakan algoritma Naïve Bayes dengan optimasi *Information Gain* dan SMOTE untuk komentar pengguna aplikasi ChatGPT. Hasil evaluasi dari 10 model yang dikembangkan diperoleh model Opt2 sebagai model terbaik dengan nilai akurasi yang didapatkan sebesar 87.20% dan nilai *F1-score* sebesar 84.74%. Rata-rata peningkatan nilai akurasi yang didapat adalah sebesar 6.25% dan sebesar 25.44% untuk nilai *F1-score*. Berdasarkan komentar pengguna dari visualisasi *word cloud*, pengembang dapat mempertahankan fitur-fitur yang membantu pengguna dalam pekerjaan mereka sembari meningkatkan kualitas respon jawaban aplikasi dan mengatasi masalah-masalah *error* yang sering terjadi saat pengguna menggunakan ChatGPT. Pada penelitian selanjutnya, disarankan untuk memperluas penggunaan metode klasifikasi dan metode optimasi yang berbeda untuk membandingkan kinerja dengan metode yang telah diterapkan sebelumnya. Selain itu, penelitian selanjutnya dapat memperluas rentang waktu pengambilan data agar model yang dikembangkan lebih akurat.

DAFTAR PUSTAKA

[1] K. Hu, "ChatGPT Sets Record for Fastest-Growing User Base - Analyst Note," Feb. 2023. https://www.reuters.com/technology/chatgpt-sets-record-fastest-growing-user-base-analyst-note-2023-02-01/ (accessed Mar. 16, 2023).

[2] H. Hassani and E. S. Silva, "The Role of ChatGPT in Data Science: How AI-Assisted Conversational Interfaces Are Revolutionizing the Field," *Big Data and Cognitive Computing*, vol. 7, no. 2, p. 62, Mar. 2023, doi: 10.3390/bdcc7020062.

[3] M. U. Haque, I. Dharmadasa, Z. T. Sworna, R. N. Rajapakse, and H. Ahmad, "'I Think This Is the Most Disruptive Technology': Exploring Sentiments of ChatGPT Early Adopters Using Twitter Data," Dec. 2022, [Online]. Available: http://arxiv.org/abs/2212.05856

[4] Y. Jiang, T. Hu, and H. Zhao, "Users' Comment Mining for App Software's Quality-in-Use," *Communications in Computer and Information Science*, pp. 510–525, 2019, doi: https://doi.org/10.1007/978-981-15-1377-0_40.

[5] M. Lu and P. Liang, "Automatic Classification of Non-Functional Requirements from Augmented App User Reviews," in *ACM International Conference Proceeding Series*, Association for Computing Machinery, Jun. 2017, pp. 344–353. doi: 10.1145/3084226.3084241.

[6] A. R. Alaei, S. Becken, and B. Stantic, "Sentiment Analysis in Tourism: Capitalizing on Big Data," *Journal of Travel Research*, vol. 58, no. 2. SAGE Publications Ltd, pp. 175–191, Feb. 01, 2019. doi: 10.1177/0047287517747753.

[7] T. H. J. Hidayat, Y. Ruldeviyani, A. R. Aditama, G. R. Madya, A. W. Nugraha, and M. W. Adisaputra, "Sentiment Analysis of Twitter Data Related to Rinca Island Development Using Doc2Vec and SVM and Logistic Regression as Classifier.," in *Procedia Computer Science*, Elsevier B.V., 2021, pp. 660–667. doi: 10.1016/j.procs.2021.12.187.

[8] V. A. Fitri, R. Andreswari, and M. A. Hasibuan, "Sentiment Analysis of Social Media Twitter with Case of Anti-LGBT Campaign in Indonesia using Naïve Bayes, Decision Tree, and Random Forest Algorithm," in *Procedia Computer Science*, Elsevier B.V., 2019, pp. 765–772. doi: 10.1016/j.procs.2019.11.181.

[9] Q. Bani Baker, F. Shatnawi, S. Rawashdeh, M. Al-Smadi, and Y. Jararweh, "Detecting Epidemic Diseases Using Sentiment Analysis of Arabic Tweets," *Journal of Universal Computer Science*, vol. 26, no. 1, pp. 50–70, 2020.

[10] H. Chen, S. Hu, R. Hua, and X. Zhao, "Improved Naive Bayes Classification Algorithm for Traffic

Risk Management," *EURASIP J Adv Signal Process*, vol. 2021, no. 1, Dec. 2021, doi: 10.1186/s13634-021-00742-6.

[11] N. Özkurt, T. Yıldırım, Yaşar Üniversitesi, Institute of Electrical and Electronics Engineers. Turkey Section., and Institute of Electrical and Electronics Engineers, "Sentiment Analysis in Turkish Text with Machine Learning Algorithms," in *2019 Innovations in Intelligent Systems and Applications Conference (ASYU)*, 2019.

[12] A. E. Sari, S. Widowati, and K. M. Lhaksmana, "Klasifikasi Ulasan Pengguna Aplikasi Mandiri Online di Google Play Store dengan Menggunakan Metode Information Gain dan Naive Bayes Classifier," *e-Proceeding of Engineering*, vol. 6, no. 2, pp. 1–15, Aug. 2019.

[13] A. B. P. Negara, H. Muhardi, and I. M. Putri, "Analisis Sentimen Maskapai Penerbangan Menggunakan Metode Naive Bayes dan Seleksi Fitur Information Gain," vol. 7, no. 3, pp. 599–606, 2020, doi: 10.25126/jtiik.202071947.

[14] A. R. Safitri and A. Muslim, "Improved Accuracy of Naive Bayes Classifier for Determination of Customer Churn Uses SMOTE and Genetic Algorithms," *Journal of Soft Computing Exploration*, vol. 1, no. 1, pp. 70–75, 2020, doi: https://doi.org/10.52465/joscex.v1i1.5.

[15] T. H. Kung *et al.*, "Performance of ChatGPT on USMLE: Potential for AI-Assisted Medical Education Using Large Language Models," *PLOS Digital Health*, vol. 2, no. 2, Feb. 2023, doi: 10.1371/journal.pdig.0000198.

[16] OpenAI, "Introducing ChatGPT," Nov. 2022. https://openai.com/blog/chatgpt#OpenAI (accessed Mar. 16, 2023).

[17] N. C. Dang, M. N. Moreno-García, and F. De la Prieta, "Sentiment Analysis Based on Deep Learning: A Comparative Study," *Electronics (Switzerland)*, vol. 9, no. 3, Mar. 2020, doi: 10.3390/electronics9030483.

[18] A. Rahman Isnain, A. Indra Sakti, D. Alita, and N. Satya Marga, "Sentimen Analisis Publik terhadap Kebijakan Lockdown Pemerintah Jakarta Menggunakan Algoritma SVM," *JDMSI*, vol. 2, no. 1, pp. 31–37, 2021, [Online]. Available: https://t.co/NfhnfMjtXw

[19] A. K. Fauziyyah and D. H. Gautama, "Analisis Sentimen Pandemi COVID-19 pada Streaming Twitter dengan Text Mining Python," *Jurnal Ilmiah SINUS*, vol. 18, no. 2, p. 31, Jul. 2020, doi: 10.30646/sinus.v18i2.491.

[20] M. I. Fikri, T. S. Sabrila, and Y. Azhar, "Perbandingan Metode Naïve Bayes dan Support Vector Machine pada Analisis Sentimen Twitter," *SMATIKA*, vol. 10, no. 2, pp. 71–76, 2020.

[21] R. Wirth and J. Hipp, "CRISP-DM: Towards a Standard Process Model for Data Mining," *Proceedings of the 4th International Conference on the Practical Applications of Knowledge Discovery and Data Mining*, vol. 01, no. 01, 2000.

[22] Y. A. Singgalen, "Analisis Sentimen Wisatawan terhadap Kualitas Layanan Hotel dan Resort di Lombok Menggunakan SERVQUAL dan CRISP-DM," *Technology and Science (BITS)*, vol. 4, no. 4, pp. 1870–1882, 2023, doi: 10.47065/bits.v4i4.3199.

[23] Y. Mahmud, S. Shaeeali, and S. Mutalib, "Comparison of Machine Learning Algorithms for Sentiment Classification on Fake News Detection," *IJACSA (International Journal of Advanced Computer Science and Applications)*, vol. 12, no. 10, pp. 658–665, 2021, [Online]. Available: www.ijacsa.thesai.org

[24] A. Alsaeedi and M. Z. Khan, "A Study on Sentiment Analysis Techniques of Twitter Data," *International Journal of Advanced Computer Science and Applications*, vol. 10, no. 2, pp. 361–374, 2019, doi: 10.14569/ijacsa.2019.0100248.

[25] R. Siringoringo, "Klasifikasi Data Tidak Seimbang Menggunakan Algoritma SMOTE dan k-Nearest Neighbor," *Jurnal ISD*, vol. 3, no. 1, pp. 44–49, 2018.

[26] N. V Chawla, K. W. Bowyer, L. O. Hall, and W. P. Kegelmeyer, "SMOTE: Synthetic Minority Over-Sampling Technique," *Journal of Artificial Intelligence Research*, vol. 16, pp. 321–357, 2002.

[27] V. Rupapara, F. Rustam, H. F. Shahzad, A. Mehmood, I. Ashraf, and G. S. Choi, "Impact of SMOTE on Imbalanced Text Features for Toxic Comments Classification Using RVVC Model," *IEEE Access*, vol. 9, pp. 78621–78634, 2021, doi: 10.1109/ACCESS.2021.3083638.

[28] Y. Deta Kirana and S. Al Faraby, "Sentiment Analysis of Beauty Product Reviews Using the K-Nearest Neighbor (KNN) and TF-IDF Methods with Chi-Square Feature Selection," *OPEN ACCESS J DATA SCI APPL*, vol. 4, no. 1, pp. 31–042, 2021, doi: 10.34818/JDSA.2021.4.71.

[29] M. Hussein and F. Özyurt, "A New Technique for Sentiment Analysis System Based on Deep Learning Using Chi-Square Feature Selection Methods," *Balkan Journal of Electrical and Computer Engineering9*, vol. 9, no. 4, Oct. 2021, doi: 10.17694/bajece.887339.

[30] A. P. P. Wardani and M. D. Purbolaksono, "Sentiment Analysis on Beauty Product Review Using Modified Balanced Random Forest Method and Chi-Square," *Journal of Information System Research*, vol. 4, no. 1, pp. 1–7, Oct. 2022, doi: 10.47065/josh.v4i1.2047