

## COMPARISON OF NAIVE BAYES AND RANDOM FOREST METHODS IN SENTIMENT ANALYSIS ON THE GETCONTACT APPLICATION

Juan Pala Arisula<sup>\*1</sup>, Parjito<sup>2</sup>

<sup>1</sup>Information System, Universitas Teknokrat Indonesia, Indonesia

<sup>2</sup>Informatics, Universitas Teknokrat Indonesia, Indonesia

Email: [juan\\_pala\\_arisula@teknokrat.ac.id](mailto:juan_pala_arisula@teknokrat.ac.id), [djito@teknokrat.ac.id](mailto:djito@teknokrat.ac.id)

(Article received: April 24, 2024; Revision: May 16, 2024; published: October 20, 2024)

### Abstract

The rapid growth in the use of social media and instant messaging platform apps has significantly changed the way people communicate. One of the most popular apps is GetContact, a platform focused on identifying the phone numbers of irresponsible people and reducing the impact of spam calls. In cases like this, sentiment analysis is important to understand user responses to the service. In performing sentiment analysis, there are two classification methods that will be used, namely the Naive Bayes and Random Forest methods. This research utilizes the SMOTE technique to handle data imbalance, and the results show that the application of SMOTE successfully improves classification accuracy. The Random Forest model performed better than Naive Bayes, with 80% accuracy, 84% precision, 77% recall, and 80% F1 score for positive sentiments, while Naive Bayes achieved 77% accuracy, 79% precision, 79% recall, and 79% F1 score. Although Random Forest is superior in precision, recall, and F1 score for positive sentiments, it performs almost on par with Naive Bayes in classifying negative sentiments, with 76% precision, 84% recall, and 80% F1 score for Random Forest, and 76% precision, 76% recall, and 76% F1 score for Naive Bayes. This shows that both models provide similar results in identifying negative sentiment overall.

**Keywords:** GetContact, Naive Bayes, Random Forest, Sentiment Analysis.

## PERBANDINGAN METODE NAIVE BAYES DAN RANDOM FOREST DALAM ANALISIS SENTIMEN PADA APLIKASI GETCONTACT

### Abstrak

Perkembangan yang pesat dalam penggunaan aplikasi platform media sosial dan pesan instan telah mengubah cara komunikasi masyarakat secara signifikan. Salah satu aplikasi yang populer adalah GetContact, sebuah platform yang difokuskan untuk mengidentifikasi nomor telepon orang-orang yang tidak bertanggung jawab dan mengurangi dampak panggilan spam. Pada kasus seperti ini analisis sentimen menjadi penting digunakan untuk memahami respons pengguna terhadap layanan ini. Dalam melakukan analisis sentimen terdapat dua metode klasifikasi yang akan digunakan yaitu metode Naive Bayes dan Random Forest. Penelitian ini memanfaatkan teknik SMOTE untuk menangani ketidakseimbangan data, dan didapatkan hasil yang menunjukkan bahwa penerapan SMOTE berhasil meningkatkan akurasi klasifikasi. Model Random Forest menunjukkan performa lebih baik daripada Naive Bayes, dengan akurasi 80%, precision 84%, recall 77%, dan F1 score 80% untuk sentimen positif, sedangkan Naive Bayes mencapai akurasi 77%, precision 79%, recall 79%, dan F1 score 79%. Meskipun Random Forest lebih unggul dalam presisi, recall, dan F1 score untuk sentimen positif, performanya hampir setara dengan Naive Bayes dalam mengklasifikasikan sentimen negatif, dengan precision 76%, recall 84%, dan F1 score 80% untuk Random Forest, serta precision 76%, recall 76%, dan F1 score 76% untuk Naive Bayes. Ini menunjukkan bahwa kedua model memberikan hasil yang serupa dalam mengidentifikasi sentimen negatif secara keseluruhan.

**Kata kunci:** Analisis Sentimen, GetContact, Naive Bayes, Random Forest.

### 1. PENDAHULUAN

Perkembangan teknologi di ranah internet telah berlangsung dengan cepat, mempengaruhi berbagai aspek kehidupan manusia. Ini mencakup perubahan

dalam perdagangan melalui *e-commerce*, transformasi dalam metode pendidikan dengan adopsi pembelajaran online, serta dampak signifikan pada cara manusia berinteraksi dan berkomunikasi

secara sosial melalui beragam platform dan media sosial [1]. Menurut survei dari Asosiasi Penyelenggara Jasa Internet Indonesia (APJII), penggunaan internet di Indonesia pada periode 2022-2023 telah mencapai 215,63 juta pengguna. Angka ini terus meningkat setiap tahunnya seiring dengan perkembangan teknologi yang semakin canggih. Laporan dari We Are Social mencatat bahwa pada Januari 2023, jumlah pengguna internet di Indonesia mencapai 213 juta, meningkat sebesar 5,44% dibandingkan tahun sebelumnya. Pertumbuhan ini didorong oleh internet yang kini menjadi kebutuhan esensial bagi masyarakat [2].

Kemajuan dalam bidang teknologi membuat banyak informasi yang dapat diterima oleh masyarakat [3]. Dalam kasus ini, manusia mendapatkan dampak positif dan negatif dalam kemudahan mengakses suatu aplikasi atau sistem [4]. Namun tidak dapat dipungkiri, bahwa internet dapat berkaitan dengan kejahatan yang dapat merugikan penggunaannya. Maka dari itu pengguna harus berhati-hati dengan data pribadi mereka agar orang yang tidak bertanggung jawab tidak dapat menyalahgunakannya [5]. Aplikasi media sosial dan pesan instan telah menjadi sarana utama bagi masyarakat dalam berinteraksi dan berkomunikasi [6]. Salah satu aplikasi yang populer adalah *GetContact*, yang membantu pengguna dalam mengidentifikasi nomor telepon penipu dan mencegah panggilan *spam* [7]. Fitur-fitur yang dimiliki aplikasi ini memungkinkan pengguna untuk melindungi diri dari panggilan yang tidak diinginkan serta memberikan pengalaman yang baik dalam mengelola kontak telepon [8]. *GetContact* menciptakan lingkungan yang memungkinkan pengguna untuk saling berinteraksi dan berbagi informasi kontak, meningkatkan kesamaan dan kepercayaan dalam komunikasi [9].

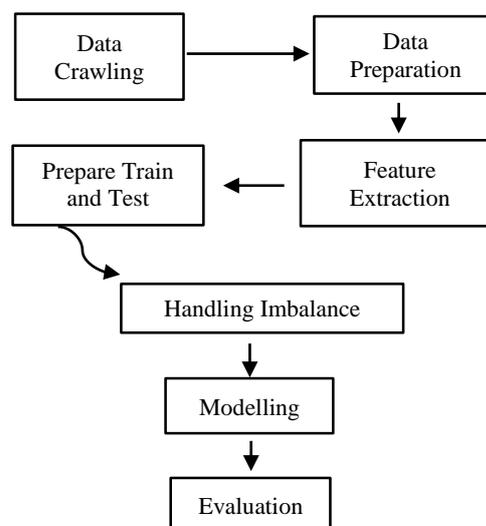
Untuk memahami persepsi dan pandangan pengguna terhadap layanan yang disediakan oleh *GetContact* maka perlu dilakukan analisis sentimen terhadap aplikasi tersebut [10]. Analisis sentimen mencakup pengumpulan, pengelompokan, dan penafsiran opini atau perasaan yang terkandung dalam teks, yang memungkinkan pemahaman mendalam tentang preferensi dan kebutuhan pengguna [11].

Penelitian sebelumnya yang dilakukan oleh Ronny Julianto, Evi Dianti Bintari, dan Indrianti dari STMIK PPKIA Tarakanita Rahmawati menggunakan metode *Naive Bayes Classification* untuk menganalisis sentimen pada *data training*, mencapai akurasi 79%. Dengan mengacu pada *The Text Mining Handbook, Sentiment Analysis* pada Review Barang Berbahasa Indonesia, dan *An Analysis of Factors Used in Search Engine Ranking*, penelitian ini berhasil mengklasifikasikan sentimen secara akurat melalui *confusion matrix*, membuktikan keefektifan metode tersebut dalam analisis sentimen [12].

Penelitian ini menggunakan algoritma *Random Forest Classification* dan teknik *Randomized Search CV* untuk memfilter email *spam* secara efektif, mencapai akurasi tinggi dan nilai ROC AUC dalam kumpulan data pelatihan dan pengujian. Model *random forest classifier* masih menunjukkan kinerja yang sangat baik dengan tingkat akurasi sekitar 97% dan nilai ROC AUC sekitar 99%, memprediksi email *spam* secara akurat dalam data baru. Hasilnya menyoroti keandalan model dalam mengklasifikasikan email *spam* dan menyarankan area potensial untuk penelitian lebih lanjut guna meningkatkan generalisasi dan mengurangi masalah *overfitting* [13]. Dalam melakukan analisis sentimen, pemilihan metode klasifikasi yang tepat menjadi kunci. Dua metode klasifikasi yang akan digunakan adalah *Naive Bayes* dan *Random Forest* [14].

Berdasarkan latar belakang tersebut penelitian ini bertujuan untuk membandingkan efektivitas metode *Naive Bayes* dan *Random Forest* dalam melakukan analisis sentimen terhadap data yang dihasilkan dari ulasan pengguna *GetContact*. Dengan membandingkan keduanya, diharapkan dapat diidentifikasi metode yang lebih efektif dalam mengklasifikasikan sentimen pengguna terhadap aplikasi tersebut [15]. Selain itu, penelitian ini juga bertujuan untuk meningkatkan pemahaman tentang penerapan teknik klasifikasi ini dalam konteks aplikasi media sosial dan pesan instan. Oleh karena itu, penelitian ini akan memberikan manfaat tidak hanya bagi pengguna *GetContact*, tetapi juga bagi pengembang aplikasi dan peneliti yang tertarik dalam analisis sentimen serta penggunaan teknologi klasifikasi dalam konteks media sosial dan pesan instan.

## 2. METODE PENELITIAN



Gambar 1. Tahapan Penelitian

Metode penelitian ini meliputi tahapan yang terstruktur, dimulai dari pengumpulan data, persiapan data, ekstraksi fitur, penyusunan *dataset* latih dan uji,

pemodelan, hingga evaluasi. Dengan pendekatan ini, memastikan bahwa setiap langkah dalam proses analisis sentimen dikembangkan secara sistematis dan terkoordinasi untuk mencapai hasil yang akurat dan relevan dalam penelitian ini. Tahapan penelitian dapat dilihat pada Gambar 1.

### 2.1. Data Crawling

Proses pengumpulan data dilakukan melalui ekstraksi ulasan pengguna dari *platform Google Play Store*, khususnya untuk aplikasi bernama *GetContact* [16]. Data ini diperoleh menggunakan metode *scraping* yang diimplementasikan menggunakan lingkungan pengembangan *Google Colab*. Dengan demikian, data yang diperoleh mencakup berbagai ulasan, peringkat, dan komentar pengguna terhadap aplikasi tersebut, yang akan digunakan untuk analisis lebih lanjut terkait performa dan *respons* pengguna terhadap *GetContact*.

### 2.2. Data Preparation

*Data preparation* merupakan tahapan dalam proses analisis data mentah dipersiapkan dan diproses untuk dilakukan analisis lebih lanjut. Proses ini mencakup pembersihan data, penanganan *missing values*, transformasi data, dan penghapusan data yang tidak relevan. Tujuan dari *data preparation* untuk memastikan bahwa data yang digunakan dalam analisis berkualitas dan siap untuk diproses oleh model atau algoritma tertentu [17]. Proses tersebut melalui beberapa tahap yaitu *Data cleaning*, *Case folding*, *Stopword removal*, *Tokenizing* dan *Labelling*.

#### a. Data Cleaning

*Data cleaning* merupakan proses penghapusan emoji, URL, dan *hashtag* dari teks atau data untuk membersihkan konten dari elemen-elemen yang tidak diinginkan atau tidak relevan.

#### b. Case Folding

*Case folding* merujuk pada proses mengubah semua karakter dalam teks menjadi huruf kecil untuk memastikan konsistensi dalam analisis teks, terutama dalam konteks pemrosesan bahasa alami.

#### c. Tokenizing

Tokenisasi adalah proses memisahkan teks menjadi unit-unit yang lebih kecil seperti kata-kata atau frasa, yang disebut sebagai token, untuk mempermudah analisis dan pemrosesan lebih lanjut.

#### d. Stopword

*Stopword removal & filtering* adalah langkah dalam pemrosesan teks yang melibatkan eliminasi kata-kata umum yang tidak memberikan nilai tambah pada pemahaman konten, seperti "dan", "atau", dan "dari".

#### e. Stemming

*Stemming* adalah teknik dalam pemrosesan teks yang menghapus imbuhan dari kata untuk menghasilkan bentuk dasar kata yang disebut stem,

sehingga kata-kata yang berasal dari akar yang sama dapat diidentifikasi dengan benar [18].

#### f. Labeling Data

*Labeling data* adalah proses penambahan label atau *tag* pada setiap data dalam sebuah *dataset* berdasarkan karakteristik atau atribut tertentu yang dimiliki oleh data tersebut. Dalam hal analisis sentimen, hasil evaluasi sentimen menggunakan *library python* membantu untuk mengklasifikasikan ulasan dari *GetContact*. Ulasan dengan skor lebih dari 3 dianggap sebagai ulasan positif, sementara skor evaluasi kurang dari atau sama dengan 3, maka ulasan tersebut dikategorikan sebagai negatif. Berdasarkan klasifikasi ini, jumlah ulasan *GetContact* yang memiliki sentimen positif adalah 1854 data, sedangkan ulasan dengan sentimen negatif berjumlah 1381 data.

### 2.3. Feature Extraction

Feature Extraction TF-IDF (*Term Frequency-Inverse Document Frequency*) adalah metode yang digunakan dalam pemrosesan teks untuk mengekstraksi dan menentukan kepentingan relatif dari sebuah kata dalam sebuah dokumen atau kumpulan dokumen [19]. Dalam TF-IDF, setiap kata di dalam dokumen diberi skor yang mencerminkan seberapa pentingnya kata tersebut dalam konteks dokumen tersebut, dibandingkan dengan kumpulan dokumen secara keseluruhan. Skor ini dihitung dengan menggabungkan dua faktor: TF (*Term Frequency*), yang mencerminkan seberapa sering sebuah kata muncul dalam sebuah dokumen, dan IDF (*Inverse Document Frequency*), yang mencerminkan seberapa unik sebuah kata dalam kumpulan dokumen. Rumus TF dapat dilihat pada Persamaan 1.

$$TF(t_i, d_j) = f(t_i, d_j) \quad (1)$$

$f(t_i, d_j)$  merupakan jumlah kemunculan *term*  $i$  pada dokumen  $j$ . Dengan menggabungkan TF dan IDF, TF-IDF memberikan skor yang tinggi untuk kata-kata yang sering muncul dalam sebuah dokumen tetapi jarang muncul dalam kumpulan dokumen secara keseluruhan, sehingga menyoroti kata-kata yang unik atau spesifik untuk dokumen tersebut. Skor TF-IDF ini digunakan sebagai fitur atau atribut dalam analisis teks, seperti dalam klasifikasi teks atau pencarian informasi.

### 2.4. Prepare Train and Test Dataset

*Prepare Train and Test Dataset* adalah tahapan dalam proses pengolahan data yang bertujuan untuk membagi *dataset* menjadi dua *subset* terpisah: *subset* pelatihan (*train*) dan *subset* pengujian (*test*) [20]. *Subset* pelatihan digunakan untuk melatih model atau algoritma pembelajaran mesin, sedangkan *subset* pengujian digunakan untuk menguji kinerja model yang telah dilatih. Pembagian *dataset* ini penting untuk menghindari *overfitting* dan memastikan

bahwa model dapat melakukan generalisasi dengan baik pada data yang belum pernah dilihat sebelumnya.

### 2.5. Handling Imbalance

Dalam menangani ketidakseimbangan kelas pada *dataset*, salah satu pendekatan yang umum digunakan adalah teknik *oversampling* seperti *Synthetic Minority Over-sampling Technique* (SMOTE). SMOTE bekerja dengan cara membuat sampel sintesis baru dalam kelas minoritas. Selain itu, SMOTE juga membantu dalam meningkatkan generalisasi model dengan menyediakan variasi tambahan dalam data, yang dapat membantu model untuk belajar pola yang lebih baik dari kelas minoritas. Oleh karena itu, pengoptimalan SMOTE merupakan langkah yang penting dalam menghadapi masalah ketidakseimbangan kelas, juga dapat meningkatkan kinerja dan akurasi model klasifikasi.

### 2.6. Modeling

*Modelling* melibatkan penggunaan model statistik atau algoritma pembelajaran mesin tertentu untuk menganalisis pola-pola yang terdapat dalam data dan membuat prediksi atau klasifikasi. Dua model algoritma yang digunakan dalam penelitian ini yaitu *Naive Bayes* dan *Random Forest*.

#### a. Naive Bayes

Algoritma klasifikasi yang berbasis pada *Teorema Bayes* dengan asumsi sederhana bahwa fitur-fitur dalam data adalah independen satu sama lain [21]. Meskipun asumsi ini terkadang tidak realistis, *Naive Bayes* sering memberikan kinerja yang baik, terutama pada *dataset* dengan dimensi tinggi. Keunggulan *Naive Bayes* termasuk kecepatan pelatihan dan prediksi yang tinggi serta kinerja yang baik pada *dataset* yang cukup besar [22]. Rumus *Naive Bayes* dapat dilihat pada Persamaan 2.

$$P(X|Y) = \frac{P(Y|X) \times P(X)}{P(Y)} \quad (1)$$

$P(X|Y)$  adalah probabilitas bahwa X terjadi jika Y telah terjadi, sementara  $P(Y|X)$  adalah probabilitas bahwa Y terjadi jika X telah terjadi, dengan  $P(X)$  dan  $P(Y)$  sebagai probabilitas sebelumnya dari X dan Y secara berturut-turut.

#### b. Random Forest

*Random Forest* merupakan algoritma pembelajaran mesin *ensemble* yang menggunakan sejumlah besar pohon keputusan secara independen untuk membuat prediksi [18]. Setiap pohon keputusan dibangun dengan menggunakan *dataset* latih yang diambil secara acak dengan penggantian (*bootstrap sampling*) dan pemilihan fitur secara acak pada setiap simpul pemisahan. Prediksi akhir dari *Random Forest* diperoleh melalui agregasi hasil prediksi dari semua pohon keputusan, seperti *voting* mayoritas untuk klasifikasi atau rata-rata untuk regresi, sehingga menghasilkan model yang kuat,

tangguh, dan mampu menangani berbagai jenis data [23]. Beberapa kelebihan dari algoritma *Random Forest* yaitu kinerja yang tinggi dalam klasifikasi dan regresi, ketangguhan terhadap *overfitting* karena penggunaan *ensemble learning*, kemampuan untuk mengatasi data yang tidak seimbang, serta kemudahan penggunaan dan disesuaikan dengan *hyperparameter* yang intuitif.

Dengan menggunakan *Naive Bayes* dan *Random Forest* sebagai model dalam proses *modelling*, kita dapat memanfaatkan keunggulan masing-masing model untuk mendapatkan prediksi atau klasifikasi yang akurat dan berguna dalam analisis data. Selain itu, perbandingan kinerja kedua model ini juga dapat memberikan wawasan tambahan dalam memilih model yang paling sesuai untuk *dataset* dan tujuan analisis yang spesifik.

### 2.7. Evaluation

Dalam analisis kinerja sistem, menggunakan *Classification report*, Terdapat metrik evaluasi penting yang digunakan untuk mengevaluasi keakuratan dan keefektifan suatu model yaitu akurasi, presisi, dan *recall*. Dalam penelitian ini, evaluasi dilakukan dengan menggunakan nilai akurasi sebagai metrik evaluasi utama [24]. Akurasi menjadi parameter penting yang digunakan untuk mengukur sejauh mana model klasifikasi atau prediksi mampu memberikan hasil yang akurat atau sesuai dengan data yang tersedia [25]. Akurasi mencerminkan presentase keberhasilan model dalam memprediksi kelas atau label secara tepat, serta mengukur sejauh mana model mampu mengklasifikasikan data dengan benar secara keseluruhan [26]. Sedangkan presisi menilai seberapa banyak dari prediksi positif yang benar. Sementara itu, *recall* mengukur kemampuan sistem untuk menemukan kembali semua *instance* dari kelas yang relevan [27].

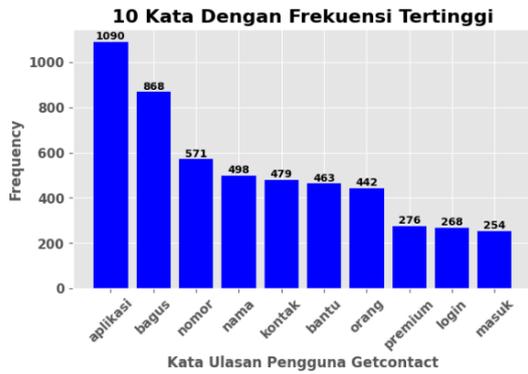
## 3. HASIL DAN PEMBAHASAN

### a. Dataset

*Dataset* diperoleh dari ulasan pengguna pada *Google Play Store* melalui proses *crawling*, menghasilkan sebanyak 3235 data. Selanjutnya, dilakukan tahap *preprocessing* teks pada *dataset*, diikuti dengan pelabelan atau pembuatan kelas. Dalam data tersebut, terdapat 10 kata dengan frekuensi tertinggi, antara lain "aplikasi", "bagus", "nomor", "nama", "kontak", "bantu", "orang", "premium", "login", "masuk".

Gambar 2 menampilkan 10 kata dengan frekuensi tertinggi yang muncul dalam ulasan pengguna. Dengan melihat grafik ini, kita dapat mengidentifikasi kata-kata yang paling sering digunakan oleh pengguna dalam ulasan mereka. Informasi ini memberikan wawasan tentang aspek-aspek produk yang paling sering ditekankan atau dibahas oleh pengguna. Dengan demikian, analisis ini dapat membantu dalam memahami preferensi,

kebutuhan, atau masalah yang paling umum dihadapi pengguna terkait produk tersebut.



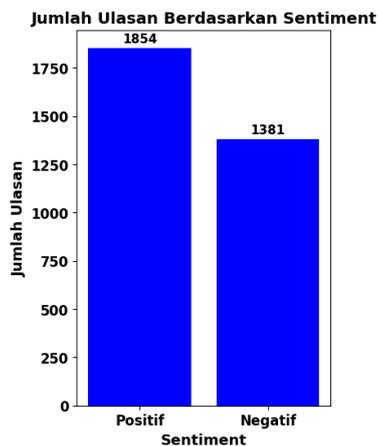
Gambar 2. 10 Kata Frekuensi Tertinggi

Metode TF-IDF digunakan untuk mengevaluasi pentingnya sebuah kata dalam sebuah dokumen terhadap kumpulan dokumen lainnya. Penjelasan untuk setiap kata dalam daftar tersebut adalah sebagai berikut:

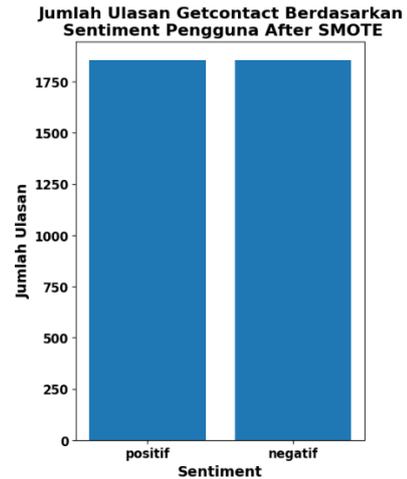
1. 'aplikasi' memiliki bobot TF-IDF sebesar 1090.
2. 'bagus' memiliki bobot TF-IDF sebesar 868.
3. 'nomor' memiliki bobot TF-IDF sebesar 571.
4. 'nama' memiliki bobot TF-IDF sebesar 498.
5. 'kontak' memiliki bobot TF-IDF sebesar 479.
6. 'bantu' memiliki bobot TF-IDF sebesar 463.
7. 'orang' memiliki bobot TF-IDF sebesar 442.
8. 'premium' memiliki bobot TF-IDF sebesar 276.
9. 'login' memiliki bobot TF-IDF sebesar 268.
10. 'masuk' memiliki bobot TF-IDF sebesar 254.

Dengan kata lain, kata-kata ini memiliki tingkat kepentingan yang tinggi dalam *corpus* atau kumpulan dokumen yang dianalisis, berdasarkan perhitungan TF-IDF. Hal ini bisa menunjukkan bahwa kata-kata tersebut merupakan kata kunci yang relevan atau penting dalam konteks dokumen yang dianalisis.

Berdasarkan penelitian yang telah dilakukan, proses SMOTE (*Synthetic Minority Over-sampling Technique*) berhasil meningkatkan jumlah data menjadi 3708 dari jumlah awal sebanyak 3235 data. Data yang telah didapat kemudian dilakukan pembagian data yaitu 80% data *training* sebanyak 2966 data, dan 20% data *testing* sebanyak 742 data.



Gambar 3. Jumlah ulasan pengguna sebelum optimasi SMOTE



Gambar 4. Jumlah ulasan pengguna sesudah optimasi SMOTE

Gambar 3 menunjukkan jumlah ulasan pengguna sebelum penerapan optimasi SMOTE, sedangkan Gambar 4 menunjukkan jumlah ulasan pengguna setelah penerapan optimasi SMOTE. Dengan menggunakan grafik tersebut, kita dapat melihat perbandingan langsung antara volume ulasan pengguna sebelum dan sesudah implementasi teknik SMOTE. Hal ini memungkinkan kita untuk memahami dampak dari penggunaan teknik SMOTE dalam meningkatkan jumlah ulasan pengguna, yang dapat memberikan wawasan penting tentang efektivitas teknik tersebut dalam mengatasi ketidakseimbangan kelas pada data ulasan pengguna.

### b. Hasil Preprocessing

Setelah melalui proses *preprocessing* yang meliputi pembersihan data dari karakter yang tidak relevan, mengubah format, dan mempersiapkan dataset. Dibawah ini merupakan hasil dari tahapan *preprocessing*. Hasil-hasil dari setiap tahapan *preprocessing* data ditampilkan secara rinci dalam Tabel 1 hingga Tabel 6. Tabel-tabel tersebut menguraikan langkah-langkah yang telah dilakukan dalam proses ini, memberikan gambaran lengkap mengenai transformasi data yang dilakukan untuk keperluan analisis lebih lanjut.

#### 1. Data Cleaning

Untuk memahami hasil dari proses cleansing data, berikut ini disajikan hasil rinci dalam Tabel 1. Tabel ini menggambarkan tahap awal dari *preprocessing*, di mana data mentah telah dibersihkan dari noise dan informasi yang tidak relevan, sehingga siap untuk analisis selanjutnya.

<i>Tweet</i>	<i>Cleaning Data</i>
Aplikasi ini baik dan bagus, tapi sepertinya masih harus ditingkatkan membaca nama-nama yang tersimpan dalam 1 kontak atau no telpon	Aplikasi ini baik dan bagus tapi masih harus membaca namanama yang tersimpan dalam kontak atau no telpon
Dari awal masuk main klik oke oke lanjutkan gataunya jadi berlangganan langsung ke	Dari awal masuk main klik oke oke lanjutkan gataunya jadi langsung ke

premium. Dan udah di batalkan lewat googleplay juga. Apakah udh otomatis batal? Kok bacaan nya malah dibatalkan pas masa gratis nya habis. Sama aja bohong dong?	premium Dan udah di batalkan lewat juga Apakah udh otomatis batal Kok bacaan nya malah pas masa gratis nya habis Sama aja bohong dong
--	---

2. Case Folding

Langkah kedua dalam proses preprocessing adalah melakukan case folding untuk menyamakan format teks. Hasil dari proses ini dapat dilihat pada Tabel 2.

Tabel 2. Case Folding

Tweet	Case Folding
Aplikasi ini baik dan bagus tapi masih harus membaca namanama yang tersimpan dalam kontak atau no telpon	aplikasi ini baik dan bagus tapi masih harus membaca namanama yang tersimpan dalam kontak atau no telpon
Dari awal masuk main klik oke oke lanjutkan gataunya jadi langsung ke premium Dan udah di batalkan lewat juga Apakah udh otomatis batal Kok bacaan nya malah pas masa gratis nya habis Sama aja bohong dong	dari awal masuk main klik oke oke lanjutkan gataunya jadi langsung ke premium dan udah di batalkan lewat juga apakah udh otomatis batal kok bacaan nya malah pas masa gratis nya habis sama aja bohong dong

3. Tokenizing

Setelah proses awal penyiapan data, langkah berikutnya adalah melakukan tokenisasi. Tabel 3 menyajikan hasil dari proses tokenisasi ini, yang merupakan tahapan kunci dalam mempersiapkan data untuk analisis lebih lanjut.

Tabel 3. Tokenizing

Tweet	Tokenize
aplikasi ini baik dan bagus tapi masih harus membaca namanama yang tersimpan dalam kontak atau no telpon	['aplikasi', 'ini', 'baik', 'dan', 'bagus', 'tapi', 'masih', 'harus', 'membaca', 'namanama', 'yang', 'tersimpan', 'dalam', 'kontak', 'atau', 'no', 'telpon']
dari awal masuk main klik oke oke lanjutkan gataunya jadi langsung ke premium dan udah di batalkan lewat juga apakah udh otomatis batal kok bacaan nya malah pas masa gratis nya habis sama aja bohong dong	['dari', 'awal', 'masuk', 'main', 'klik', 'oke', 'oke', 'lanjutkan', 'gataunya', 'jadi', 'langsung', 'ke', 'premium', 'dan', 'udah', 'di', 'batalkan', 'lewat', 'juga', 'apakah', 'udh', 'otomatis', 'batal', 'kok', 'bacaan', 'nya', 'malah', 'pas', 'masa', 'gratis', 'nya', 'habis', 'sama', 'aja', 'bohong', 'dong']

4. Stopword

Langkah selanjutnya dalam analisis adalah tahapan penghapusan stopwords, yang hasilnya dapat dilihat pada Tabel 4.

Tabel 4. Stopword

Tweet	Tokenize
['aplikasi', 'ini', 'baik', 'dan', 'bagus', 'tapi', 'masih', 'harus', 'membaca', 'namanama', 'yang', 'tersimpan', 'dalam', 'kontak', 'atau', 'no', 'telpon']	['aplikasi', 'bagus', 'membaca', 'namanama', 'yang', 'tersimpan', 'kontak', 'telpon']
['dari', 'awal', 'masuk', 'main', 'klik', 'oke', 'oke', 'lanjutkan', 'gataunya', 'jadi', 'langsung', 'ke', 'premium', 'dan', 'udah', 'di', 'batalkan', 'lewat', 'juga', 'apakah', 'udh', 'otomatis', 'batal', 'kok', 'bacaan', 'nya', 'malah', 'pas', 'masa', 'gratis', 'nya', 'habis', 'sama', 'aja', 'bohong', 'dong']	['masuk', 'main', 'klik', 'oke', 'oke', 'lanjutkan', 'gataunya', 'langsung', 'premium', 'udah']

'batalkan', 'lewat', 'juga', 'apakah', 'udh', 'otomatis', 'batal', 'kok', 'bacaan', 'nya', 'malah', 'pas', 'nya', 'pas', 'gratis', 'nya', 'masa', 'gratis', 'nya', 'habis', 'habis', 'aja', 'bohong']	'batalkan', 'udh', 'otomatis', 'batal', 'bacaan', 'nya', 'pas', 'gratis', 'nya', 'habis', 'aja', 'bohong', 'dong']
---	--

5. Stemming

Sebagai kelanjutan dari tahapan preprocessing data, tabel-tabel sebelumnya telah memberikan gambaran tentang langkah-langkah yang dilakukan dalam mempersiapkan data. Sekarang, perhatian kita beralih ke Tabel 5, yang menyajikan hasil dari proses stemming yang telah dilakukan.

Tabel 5. Stemming

Tweet	Stemming
['aplikasi', 'bagus', 'membaca', 'namanama', 'yang', 'tersimpan', 'kontak', 'telpon']	aplikasi bagus baca namanama yang simpan kontak telpon
['masuk', 'main', 'klik', 'oke', 'oke', 'lanjutkan', 'gataunya', 'langsung', 'premium', 'udah', 'batalkan', 'udh', 'otomatis', 'batal', 'bacaan', 'nya', 'pas', 'gratis', 'nya', 'habis', 'aja', 'bohong']	masuk main klik oke oke lanjut gataunya langsung premium udah batal udh otomatis batal baca nya pas gratis nya habis aja bohong

6. Labeling Data

Dari tahapan preprocessing yang telah dijelaskan sebelumnya, langkah berikutnya adalah proses labeling positif dan negatif. Ini dilakukan untuk memberikan identifikasi pada data yang telah dipersiapkan sebelumnya agar dapat digunakan dalam analisis selanjutnya. Hal ini tercermin dalam Tabel 6, di mana hasil dari proses labeling tersebut disajikan secara detail.

Tabel 6. Labeling

Tweet	Labelling
Aplikasi ini baik dan bagus, tapi sepertinya masih harus ditingkatkan membaca nama-nama yang tersimpan dalam 1 kontak atau no telpon	Positif
Dari awal masuk main klik oke oke lanjutkan gataunya jadi berlangganan langsung ke premium. Dan udah di batalkan lewat googleplay juga. Apakah udh otomatis batal? Kok bacaan nya malah dibatalkan pas masa gratis nya habis. Sama aja bohong dong?	Negatif

c. Tahap Visualisasi WordCloud

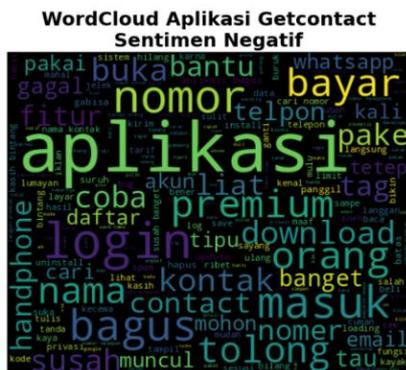
Penggunaan WordCloud bertujuan untuk memberikan visualisasi tentang dataset, hal ini memungkinkan pengidentifikasian kata-kata yang paling sering muncul dalam dokumen. Proses pembuatan WordCloud dilakukan dengan menggunakan library matplotlib dalam bahasa pemrograman Python.

Gambar 5 merupakan hasil visualisasi WordCloud aplikasi GetContact sentimen positif, kata "bagus" mendominasi dengan ukuran yang paling besar, menunjukkan bahwa mayoritas pengguna menganggap aplikasi ini memiliki nilai positif yang signifikan, karena kemudahan penggunaan atau manfaat yang diberikan. Kata "aplikasi" dan "nomor" juga muncul cukup besar,

menunjukkan bahwa pengguna banyak membicarakan aplikasi ini secara umum dan menghargai kemampuannya untuk memperoleh nomor kontak dengan efisien. Selain itu, kata-kata seperti "bantu", "nama", dan "orang" menunjukkan bahwa pengguna merasa aplikasi ini memberikan bantuan dalam menemukan informasi yang relevan. Gambar 6 menampilkan WordCloud untuk Sentimen Negatif.



Gambar 5. WordCloud Sentimen Positif

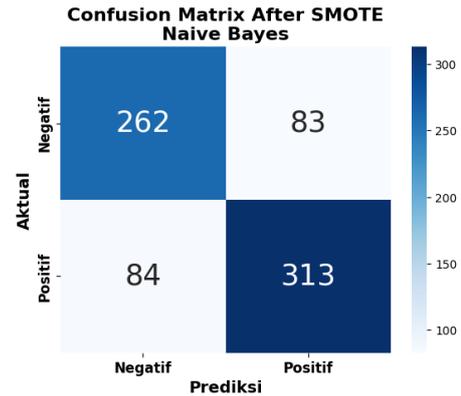


Gambar 6. WordCloud Sentimen Negatif

Gambar 6 merupakan hasil *WordCloud* aplikasi *GetContact* sentimen negatif menyoroti beberapa masalah yang dihadapi pengguna dengan aplikasi *Get Contact*. Kata "aplikasi" menjadi fokus utama, menunjukkan adanya ketidakpuasan atau masalah yang signifikan terkait dengan penggunaan aplikasi secara keseluruhan. Kata-kata seperti "bayar" dan "premium" menunjukkan bahwa sebagian pengguna merasa tidak puas dengan aspek pembayaran, seperti biaya langganan atau fitur premium. Masalah teknis seperti "login" dan "susah" juga mencuat, menunjukkan bahwa pengguna mungkin mengalami kesulitan dalam mengakses atau menggunakan aplikasi.

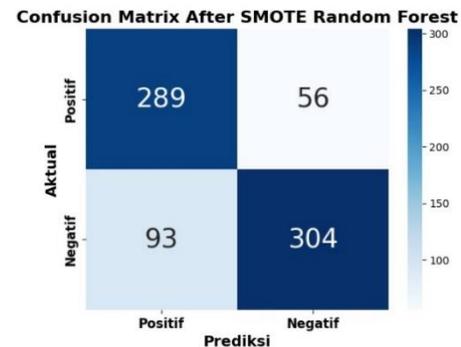
**d. Tahap Pengujian**

Setelah melalui tahapan preprocessing yang teliti, langkah selanjutnya adalah melakukan pengujian untuk mengevaluasi kinerja model. Hasil pengujian ini dapat dilihat pada Gambar 7, yang memberikan gambaran menyeluruh tentang performa model yang telah dikembangkan.



Gambar 7. Confusion Matrix Naive Bayes

Pada algoritma *Naive Bayes*, dari sisi *True Positive* (TP), jumlahnya mencapai 313 data. Hal ini menunjukkan bahwa model mampu dengan baik dalam mengidentifikasi data yang sebenarnya positif. Selanjutnya, terdapat 262 data yang berhasil diklasifikasikan sebagai *True Negative* (TN). Kinerja yang baik dalam mengidentifikasi data negatif penting untuk memastikan bahwa model tidak salah mengklasifikasikan data yang seharusnya negatif. Namun, perlu diperhatikan bahwa jumlah *False Positive* (FP) sebanyak 84 data mengindikasikan adanya kecenderungan model untuk salah mengklasifikasikan data yang sebenarnya negatif sebagai positif. Ini dapat menjadi perhatian serius karena dapat mengakibatkan kesalahan interpretasi dan pengambilan keputusan yang salah dalam aplikasi dari model ini. Selain itu, adanya *False Negative* (FN) sebanyak 83 data juga menandakan bahwa model memiliki kelemahan dalam mengidentifikasi data yang sebenarnya positif sebagai negatif. Ini juga merupakan hal yang perlu diperbaiki, karena dapat mengakibatkan pengabaian terhadap data yang sebenarnya penting. Dalam Gambar 8, kita dapat melihat *Confusion Matrix* untuk model *Random Forest* yang dihasilkan, yang akan memberikan wawasan yang berharga tentang kinerja model dalam mengklasifikasikan data.



Gambar 8. Confusion Matrix Random Forest

Hasil evaluasi performa model algoritma *Random Forest* dalam penelitian ini menunjukkan adanya 289 kasus yang berhasil diklasifikasikan secara tepat sebagai positif (*True Positive*), sementara

terdapat 56 kasus yang salah diklasifikasikan sebagai negatif padahal seharusnya positif (*False Positive*). Selain itu, model juga berhasil mengidentifikasi 304 kasus negatif dengan benar (*True Negative*), namun terdapat 93 kasus yang seharusnya negatif namun keliru diklasifikasikan sebagai positif (*False Negative*). Nilai *True Positive* yang cukup tinggi menunjukkan kemampuan model dalam mengenali pola atau fitur yang terkait dengan kelas positif, namun perlu perhatian terhadap jumlah *False Positive* yang bisa mempengaruhi kehandalan model.

Sementara itu, meskipun *True Negative* memiliki nilai yang baik, jumlah *False Negative* tetap perlu diperhatikan untuk mencegah kesalahan dalam mengidentifikasi kasus positif. Analisis ini memberikan wawasan penting untuk mengoptimalkan model guna meningkatkan akurasi. Dalam penelitian ini komparasi dilakukan menggunakan dua metode, yaitu *Naive Bayes* dan *Random Forest*. Dalam tahap pengujian berdasarkan hasil klasifikasi model menggunakan SMOTE, nilai *accuracy*, *F1-score*, *precision*, *recall*, akan dibandingkan antara kedua metode untuk dilakukan evaluasi. Langkah selanjutnya adalah melihat hasil klasifikasi menggunakan SMOTE. Ini ditampilkan dalam Tabel 7, yang memberikan gambaran tentang efektivitas metode SMOTE dalam meningkatkan kinerja model klasifikasi.

Tabel 7. Hasil Klasifikasi Menggunakan SMOTE

Model		<i>Accuracy</i>	<i>Precision</i>	<i>Recall</i>	<i>f1-score</i>
<i>Random Forest</i>	Positif	80%	84%	77%	80%
	Negatif		76%	84%	80%
<i>Naive Bayes</i>	Positif	77%	79%	79%	79%
	Negatif		76%	76%	76%

Dalam hasil klasifikasi, terlihat bahwa model *Random Forest* menunjukkan performa yang lebih baik dibandingkan dengan model *Naive Bayes*. Model *Random Forest* mencapai akurasi sebesar 80%, untuk klasifikasi sentimen positif, nilai *precision* mencapai 84%, *recall* sebesar 77%, dan *F1 score* 80%. Sementara itu, model *Naive Bayes* mencapai akurasi sebesar 77%, *precision* 79%, *recall* 79%, dan *F1 score* 79% untuk klasifikasi sentimen positif. Dalam hal ini, meskipun keduanya memberikan hasil yang cukup baik, namun model *Random Forest* memiliki tendensi untuk memberikan prediksi yang lebih baik dalam mengidentifikasi sentimen positif dengan presisi yang lebih tinggi, meskipun *recall* dan *F1 score*-nya sedikit lebih rendah daripada *Naive Bayes*. Ketika berbicara mengenai klasifikasi sentimen negatif, kedua model menunjukkan performa yang serupa. Model *Random Forest* memiliki *precision* sebesar 76%, *recall* 84%, dan *F1 score* 80%, sedangkan model *Naive Bayes* memiliki *precision* 76%, *recall* 76%, dan *F1 score* 76%. Meskipun ada perbedaan dalam *precision* dan *recall* antara kedua model ini, namun secara keseluruhan, keduanya

memberikan hasil yang hampir setara dalam mengklasifikasikan sentimen negatif.

## 4. DISKUSI

### 4.1. Penelitian Sebelumnya

Dalam penelitian terdahulu yang telah dilakukan sebelumnya. Menurut penelitian yang dilakukan oleh Leonardo & Pratama, 2020, *Random Forest* memiliki performa yang lebih baik dalam memprediksi keberhasilan klien telemarketing dibandingkan dengan *Naive Bayes*. Dalam penelitian tersebut, *Random Forest* mencapai akurasi sebesar 90% dan AUC sebesar 0.97, sedangkan *Naive Bayes* hanya mencapai akurasi sebesar 85% dan AUC sebesar 0.79. Hasil ini menunjukkan bahwa *Random Forest* lebih unggul dalam melakukan prediksi keberhasilan klien telemarketing [28].

Penelitian membandingkan metode *Naive Bayes* dan *Random Forest* dalam analisis sentimen ulasan pengguna aplikasi *BCA Mobile*. Data terdiri dari 2,453 ulasan Januari 2023. Hasilnya menunjukkan *Random Forest* memiliki akurasi lebih baik: 93,93% akurasi, 93,02% presisi, 89,89% *recall*, dan 91,43% *F1-score*. Beberapa jurnal membahas analisis sentimen dengan metode *Naive Bayes*, dan *Random Forest* pada berbagai aplikasi dan topik. Penelitian menggunakan data ulasan *BCA Mobile* dari *Google Play Store* dengan metode web *scraping*, menunjukkan *Random Forest* lebih baik daripada *Naive Bayes* [29].

### 4.2. Interpretasi Hasil

Penelitian terdahulu menunjukkan bahwa *Random Forest* secara konsisten mengungguli *Naive Bayes* dalam prediksi keberhasilan klien telemarketing dan analisis sentimen ulasan aplikasi *BCA Mobile*. Penelitian ini berfokus pada penggunaan *dataset* ulasan pengguna dari *Google Play Store* dengan teknik *oversampling* menggunakan SMOTE. Data yang semula 3235 berhasil diperluas menjadi 3708 dengan SMOTE. Dua model, *Naive Bayes* dan *Random Forest*, digunakan untuk mengklasifikasikan sentimen. Hasilnya menunjukkan bahwa *Random Forest* memiliki akurasi lebih tinggi (80%) daripada *Naive Bayes* (77%) untuk klasifikasi sentimen positif. Meskipun *Random Forest* memiliki *recall* yang sedikit lebih rendah, presisinya lebih tinggi. Namun, keduanya memiliki performa serupa dalam mengklasifikasikan sentimen negatif, meskipun dengan sedikit perbedaan dalam *precision* dan *recall*. Kesimpulannya, penelitian ini menegaskan superioritas *Random Forest* dalam mengklasifikasikan sentimen positif, sementara keduanya memiliki kinerja yang sebanding dalam mengklasifikasikan sentimen negatif.

## 5. KESIMPULAN

Berdasarkan penelitian yang telah dilakukan, penerapan teknik klasifikasi seperti *Naive Bayes* dan *Random Forest* berhasil menginvestigasi serta mengevaluasi sentimen ulasan pengguna dengan efektif. Hasil penelitian menunjukkan bahwa penggunaan teknik SMOTE untuk menangani ketidakseimbangan kelas data telah memberikan peningkatan yang signifikan dalam klasifikasi sentimen. Dengan demikian, didapat kesimpulan bahwa klasifikasi lebih baik menggunakan metode *Random Forest*, yang mencapai akurasi sebesar 80%, dibandingkan dengan *Naive Bayes* yang hanya mencapai akurasi sebesar 77%. Oleh karena itu, *Random Forest* menunjukkan potensi yang lebih tinggi dalam melakukan analisis sentimen pada aplikasi *GetContact*.

## DAFTAR PUSTAKA

- [1] F. Nufairi, N. Pratiwi, And F. Herlando, "Analisis Sentimen Pada Ulasan Aplikasi Threads Di Google Play Store Menggunakan Algoritma Support Vector Machine," *Jipi (Jurnal Ilmiah Penelitian Dan Pembelajaran Informatika)*, Vol. 9, No. 1, Pp. 339–348, Feb. 2024, Doi: 10.29100/Jipi.V9i1.4929.
- [2] T. M. Siregar, A. S. Margareta, B. Berliana, D. Simanungkalit, L. H. Gultom, and Y. S. Pasaribu, "Analisis Perbandingan Pertumbuhan Pengguna Internet Di Sumatera: Faktor-Faktor Yang Mempengaruhinya", *Innovative*, vol. 3, no. 6, pp. 10672–10680, Dec. 2023.
- [3] P. : Jurnal, K. Masyarakat, A. Yani, And B. P. Kesehatan, "Pemanfaatan Teknologi Dalam Bidang Kesehatan Masyarakat Utilization Of Technology In The Health Of Community Health," *Artikel Xii*, Vol. 8, No.1 2018.
- [4] S. Fitri, "Dampak Positif Dan Negatif Sosial Media Terhadap Perubahan Sosial Anak," *Naturalistic : Jurnal Kajian Penelitian Pendidikan Dan Pembelajaran*, Vol. 1, No. 2, Pp. 118–123, Apr. 2017, Doi: 10.35568/Naturalistic.V1i2.5.
- [5] M. Wicaksono Aji Bawono, S. Dwi Utomo, And S. Kasman, "Machine Learning Sentiment Analysis In Cyber Threat Intelligence Recommendation System," *International Journal On Information And Communication Technology (Ijoict)*, Vol. 9, No. 2, Pp. 75–85, 2023, Doi: 10.21108/Ijoict.V9i2.849.
- [6] C. K. Widada, "Mengambil Manfaat Media Sosial Dalam Pengembangan Layanan," *Journal Of Documentation And Information Science*, Vol. 2, No. 1, Pp. 23–30, Feb. 2018, Doi: 10.33505/Jodis.V2i1.130.
- [7] J. Handoko, K. Kevin, P. Paulus, And Z. Salsabila, "Sistem Deteksi Nomor Telepon Dan Rekening Bank Terindikasi Penipuan Berbasis Aplikasi Android Dan Web," *J. Sifo Mikroskil*, Vol. 23, No. 2, Pp. 183–196, 2022.
- [8] N. E. Sakinah, "Analisis User Experience Pada Aplikasi," *Jurnal Politeknik Negeri Jakarta*, Vol. 28, No. 3, P. 16, 2019.
- [9] N. Dwi Santri, T. Riza Zarzani, And S. Asmi Hasibuan, "Kajian Yuridis Hukum Aplikasi Get Contact Berdasarkan Undang-Undang Perlindungan Konsumen Dan Peraturan Nomor 20 Tahun 2016," *Jurnal Rectum: Tinjauan Yuridis Penanganan Tindak Pidana*, Vol. 4, No. 2, Pp. 480–487, 2022.
- [10] D. G. Nugroho, Y. H. Chrisnanto, and A. Wahana, "ANALISIS SENTIMEN PADA JASA OJEK ONLINE MENGGUNAKAN METODE NAÏVE BAYES," Nugroho | Prosiding Sains Nasional dan Teknologi, Vol.1,No1,2016.<https://doi.org/10.36499/psnst.v1i1.1526>
- [11] S. Rita, D. Indrayana, and A. Pambudi, "Penggunaan Support Vector Machine Untuk Analisis Sentimen Ulasan Aplikasi Truecaller Dan Getcontact," *Bit*, Vol. 20, No. 2, 2023.<https://journal.budiluhur.ac.id/index.php/bit/article/view/2493>
- [12] R. Julianto, E. D. Bintari, And I. Indrianti, "Analisis Sentimen Layanan Provider Telepon Seluler Pada Twitter Menggunakan Metode Naïve Bayesian Classification," *Journal Of Big Data Analytic And Artificial Intelligence*, Vol. 3, No. 1, Pp. 23–30, 2017.
- [13] D. Science, I. Technology, And D. Analytics, "Journal Of Dinda," Vol. 4, No. 1, Pp. 8–13, 2024.
- [14] A. Nugroho And Y. Religia, "Analisis Optimasi Algoritma Klasifikasi Naive Bayes Menggunakan Genetic Algorithm Dan Bagging," *Jurnal Resti (Rekayasa Sistem Dan Teknologi Informasi)*, Vol. 5, No. 3, Pp. 504–510, Jun. 2021, Doi: 10.29207/Resti.V5i3.3067.
- [15] H. Pramoedyo, D. Ariyanto, And N. N. Aini, "Comparison Of Random Forest And Naïve Bayes Methods For Classifying And Forecasting Soil Texture In The Area Around Das Kalikonto, East Java," *Barekeng: Jurnal Ilmu Matematika Dan Terapan*, Vol. 16, No. 4, Pp. 1411–1422, Dec. 2022, Doi: 10.30598/Barekengvol16iss4pp1411-1422.
- [16] S. Rita, D. Indrayana, And A. Pambudi, "Penggunaan Support Vector Machine Untuk Analisis Sentimen Ulasan Aplikasi Truecaller Dan Getcontact," *Bit (Fakultas Teknologi Informasi Universitas Budi Luhur)*, Vol. 20, No. 2, P. 131, 2023, Doi:

- 10.36080/Bit.V20i2.2493.
- [17] Y. Ansori And K. F. H. Holle, "Perbandingan Metode Machine Learning Dalam Analisis Sentimen Twitter," *Jurnal Sistem Dan Teknologi Informasi (Justin)*, Vol. 10, No. 4, P. 429, 2022, Doi: 10.26418/Justin.V10i4.51784.
- [18] F. A. Larasati, D. E. Ratnawati, And B. T. Hanggara, "Analisis Sentimen Ulasan Aplikasi Dana Dengan Metode Random Forest," ... *Teknologi Informasi Dan ...*, Vol. 6, No. 9, Pp. 4305–4313, 2022.
- [19] N. Ramadhani And N. Fajarianto, "Sistem Informasi Evaluasi Perkuliahan Dengan Sentimen Analisis Menggunakan Naïve Bayes Dan Smoothing Laplace," *Jurnal Sistem Informasi Bisnis*, Vol. 10, No. 2, Pp. 228–234, 2020, Doi: 10.21456/Vol10iss2pp228-234.
- [20] N. M. Farhan And B. Setiaji, "Indonesian Journal Of Computer Science," *Indonesian Journal Of Computer Science*, Vol. 12, No. 2, Pp. 284–301, 2023.
- [21] D. Faroek, M. Yusuf, And G. Syatauw, "Sentiment Analysis Of The Popularity Of Parties Supporting The 2024 Presidential Candidates On Twitter Using The Naive Bayes Classifier Algorithm," *Antivirus : Jurnal Ilmiah Teknik Informatika*, Vol. 17, No. 2, Pp. 216–227, 2023, Doi: 10.35457/Antivirus.V17i2.3261.
- [22] K. S. Putri, I. R. Setiawan, And A. Pambudi, "Analisis Sentimen Terhadap Brand Skincare Lokal Menggunakan Naïve Bayes Classifier," *Technologia : Jurnal Ilmiah*, Vol. 14, No. 3, P. 227, 2023, Doi: 10.31602/Tji.V14i3.11259.
- [23] D. Ahmad Dzulhijjah, H. Sanjaya, A. Said Wahyudi Hidayat, A. Yulistia Alwanda, and E. Utami, "Perbandingan Metode Random Forest dan KNN pada Analisis Sentimen Twitter," *Smart Comp: Jurnalnya Orang Pintar Komputer*, vol. 12, no. 3, pp. 767–772, Jul. 2023, doi: 10.30591/smartcomp.v12i3.5106.
- [24] A. S. Rahayu, A. Fauzi, And R. Rahmat, "Komparasi Algoritma Naïve Bayes Dan Support Vector Machine (Svm) Pada Analisis Sentimen Spotify," *Jurnal Sistem Komputer Dan Informatika (Jsn)*, Vol. 4, No. 2, P. 349, 2022, Doi: 10.30865/Jsn.V4i2.5398.
- [25] R. Azhar And M. F. Wijayanto, "Analisis Sentimen Di Twitter: Mengungkap Persepsi Dan Emosi Publik Seputar Konflik Palestina-Israel," *Stains (Seminar Nasional ...)*, Vol. 3, Pp. 118–124, 2024.
- [26] F. M. Fathoni, C. A. Putra, And A. L. Nurlaili, "Klasifikasi Penyakit Daun Anggur Menggunakan Metode K-Nearest Neighbor Berdasarkan Gray Level Co-Occurrence Matrix," *Biner: Jurnal Ilmiah Informatika Dan Komputer*, Vol. 3, No. 1, Pp. 8–15, 2024.
- [27] A. Fatunnisa And H. Marcos, "Prediksi Kelulusan Tepat Waktu Siswa Smk Teknik Komputer Menggunakan Algoritma Random Forest Prediction Of On-Time Graduation For Computer Engineering Vocational School Students Using The Random Forest Algorithm," Vol. 14, No. April, Pp. 101–111, 2024.
- [28] R. Leonardo And J. Pratama, "Perbandingan Metode Random Forest Dan Naïvebayes Dalam Prediksi Keberhasilan Klien Telemarketing," *Jurnal Penelitian Teknik Informatika*, Vol. 3, No. 123, Pp. 455–459, 2020.
- [29] A. Miftahusalam, A. F. Nuraini, A. A. Khoirunisa, And H. Pratiwi, "Perbandingan Algoritma Random Forest, Naïve Bayes, Dan Support Vector Machine Pada Analisis Sentimen Twitter Mengenai Opini Masyarakat Terhadap Penghapusan Tenaga Honorer," *Seminar Nasional Official Statistics*, Vol. 2022, No. 1, Pp. 563–572, 2022, Doi: 10.34123/Semnasoffstat.V2022i1.1410.