

ANALISIS SENTIMEN ULASAN APLIKASI WATTPAD DI GOOGLE PLAY STORE DENGAN METODE *RANDOM FOREST*

Safira Nur Adhan^{*1}, Gusti Ngurah Adhi Wibawa², Dian Christien Arisona³, Irma Yahya⁴, Agusrawati⁵,
Ruslan⁶

^{1,2,3,4,5,6}Program Studi Statistika, Fakultas Matematika dan Ilmu Pengetahuan Alam, Universitas Halu Oleo

Email: ^{*}safiranutradhan@gmail.com

^{*}Penulis Korespondensi

Abstrak

Wattpad merupakan salah satu aplikasi dan situs komunitas daring yang memungkinkan pengguna menulis atau membaca konten informasi dalam lingkup sastra dengan berbagai genre atau kategori seperti cerita pendek, klasik, aksi, petualangan, romansa, fantasi, humor, spiritual, misteri, horor, puisi, fiksi ilmiah, fiksi sejarah, fiksi remaja, fiksi umum, fiksi penggemar, dan non fiksi. Hingga Desember 2023, 90 juta pengguna menghabiskan lebih dari 23 miliar menit mengakses aplikasi ini setiap bulannya. Penelitian ini bertujuan untuk memberikan gambaran sentimen pengguna sekaligus mengklasifikasikannya sebagai teks sentimen negatif atau positif menggunakan metode *Random Forest* dan *Random Forest* yang di optimasi dengan SMOTE (*Synthetic Minority Oversampling Technique*) pada ulasan pengguna Aplikasi Wattpad yang mengalami ketidakseimbangan kelas. Hasil penelitian menunjukkan bahwa dari 9.975 hasil pengumpulan data, hanya 8.743 data yang dapat digunakan dengan persentase sentimen positif sebesar 64,2% (5.616) dan 35,8% (3.127) sentimen negatif. Metode *Random Forest* tanpa optimasi SMOTE cenderung lebih unggul dalam memprediksi klasifikasi sentimen yang tidak seimbang, ini dapat dilihat dari nilai akurasi yang mencapai 84,05%, presisi 84,71%, *recall* 91,60%, *F1-Score* 88,02%, FPR 8,40%, dan nilai AUC 0,9166 yang dikategorikan sebagai klasifikasi yang sangat baik. Pemodelan SMOTE *Random Forest* mampu meningkatkan kemampuan untuk mengklasifikasikan kelas minoritas yakni sentimen negatif, dapat dilihat dari naiknya nilai presisi dari 84,71% ke 86,70% (1,99%). Sayangnya, penyeimbangan kelas ini berdampak pada penurunan performa akurasi, *recall*, *f1-score* dan nilai AUC. Selain itu, berdasarkan nilai *feature importance*, fitur yang paling berpengaruh pada kedua model adalah atribut kata “kecewa”, “bagus”, dan “baik”.

Kata kunci: *Wattpad, Analisis Sentimen, Random Forest, SMOTE*

Abstract

Wattpad is an application and online community site that allows users to write or read informational content in the literary sphere with various genres or categories such as short stories, classics, action, adventure, romance, fantasy, humor, spiritual, mystery, horror, poetry, science fiction, historical fiction, teen fiction, general fiction, fan fiction, and non-fiction. By December 2023, 90 million users spent more than 23 billion minutes accessing the app each month. This study aims to provide an overview of user sentiment while classifying it as negative or positive sentiment text using Random Forest and Random Forest methods optimized with the SMOTE (Synthetic Minority Oversampling Technique) on Wattpad App user reviews that experience class imbalance. The results showed that out of 9.975 data collection results, only 8.743 data could be used with a percentage of positive sentiment of 64,2% (5.616) and 35,8% (3.127) negative sentiment. The Random Forest method without SMOTE optimization tends to be superior in predicting unbalanced sentiment classification, this can be seen from the accuracy value which reaches 84,05%, precision 84,71%, recall 91,60%, F1-Score 88,02%, FPR 8,40%, and AUC value 0,9166 are categorized as excellent classification. SMOTE Random Forest modeling is able to improve the ability to classify the minority class, negative sentiment, as can be seen from the increase in precision value from 84,71 % to 86,70% (1,99%). Unfortunately, this class balancing resulted in a decrease in the performance of accuracy, recall, f1-score and AUC values. In addition, based on the feature importance values, the most influential features in both models are the word attributes "kecewa", "bagus", and "baik".

Keywords: *Wattpad, Sentiment Analysis, Random Forest, SMOTE*

1. PENDAHULUAN

Perkembangan teknologi yang cepat saat ini secara signifikan memengaruhi sektor penyediaan

produk dan layanan. Perusahaan yang tidak berubah seiring waktu menghadapi tantangan serius dalam mempertahankan keunggulan kompetitif mereka. Apalagi, jumlah pengguna internet aktif di seluruh

dunia mencapai lebih dari 4,6 miliar, dan angka ini diperkirakan akan terus bertambah [1]. Sehingga, perusahaan harus terus berinovasi dalam mengembangkan produk dan layanan yang lebih efisien, responsif, dan sesuai dengan kebutuhan pasar.

Umpan balik (*feedback*) atau ulasan dari publik sebagai konsumen dapat menjadi sumber informasi utama bagi perusahaan. Respon mereka akan memberikan wawasan secara langsung mengenai masalah, kelemahan, atau ketidakpuasan yang dialami, membantu perusahaan untuk mengidentifikasi aspek apa yang perlu ditingkatkan. BrightLocal (2022) menyebutkan bahwa 93% konsumen *online* membaca ulasan sebelum memutuskan pembelian atau pengunduhan dan 85% konsumen mempercayai ulasan tersebut [2]. Hal ini menunjukkan pentingnya memahami sentimen dan kebutuhan konsumen dalam pengembangan produk dan layanan yang sukses.

Analisis sentimen atau biasa disebut sebagai penggalian opini (*opinion mining*) adalah studi yang menganalisis opini, sentimen, penilaian, sikap, dan emosi seseorang terhadap suatu produk, layanan, organisasi, individu, peristiwa, isu, atau topik yang diungkapkan dalam teks tertulis [3]. Analisis ini adalah salah satu cabang ilmu komputer yang memuat pemrosesan bahasa alami (*Natural Language Processing*) dan pembelajaran mesin (*Machine Learning*) yang mampu mengidentifikasi dan mengklasifikasikan opini atau emosi yang terkandung dalam teks, seperti ulasan, komentar, cuitan (*tweet*) atau artikel. Sehingga, analisis sentimen hadir sebagai solusi untuk memahami kepuasan pelanggan, mengevaluasi kinerja produk, mendeteksi perubahan tren, dan merekomendasikan konten.

Sebagai salah satu platform distribusi digital layanan aplikasi, permainan, film, musik, buku, dan majalah untuk pengguna perangkat Android, Google Play Store layanan memungkinkan pengguna untuk memberikan ulasan terhadap aplikasi yang diunduh berupa komentar dan peringkat (*rating*) mulai dari 1-5 [4]. Dengan tren baca digital atau novel online yang meningkat setiap tahunnya, aplikasi dalam kategori buku dan literatur berbasis novel *online* mendapat perhatian khusus. Berbagai aplikasi di Google Play Store dalam kategori ini antara lain Noveltoon, Wattpad, Novelplus, GoodNovel, Dreame, Novelme, Fizzo Novel, dan sebagainya [5]

Pada Aplikasi Wattpad, pengguna dapat mengeksplorasi, menulis, dan berbagi cerita, baik karya fiksi maupun nonfiksi. Bahkan, banyak cerita di Wattpad yang berhasil menjadi sumber inspirasi untuk buku *bestseller*, film, atau serial televisi [6]. Similarweb sebagai situs pengukuran peringkat aplikasi berdasarkan jumlah pengunjung mengemukakan bahwa per September 2023, Wattpad menduduki peringkat 208 di tingkat global dan peringkat 2 pada kategori Buku dan Literatur. Ini berarti bahwa Wattpad menjadi salah satu aplikasi

dengan jumlah pengunjung terbesar di dunia, dimana 18,77% pengunjung berasal dari Amerika, diikuti oleh India dengan 11,35% sebagai peringkat kedua, dan Indonesia dengan 6,45% sebagai peringkat ketiga. Selain itu, untuk di Indonesia aplikasi Wattpad menduduki peringkat 135 sebagai aplikasi yang sering digunakan dan peringkat kedua untuk aplikasi berbasis novel [7].

Hasil penelitian milik Nurhafida & Sembiring (2022) yang mengkaji sentimen aplikasi novel online di Google Play Store menggunakan algoritma Support Vector Machine (SVM), dibalik fitur dan program unggulan yang ditawarkan sebagian pengguna merasakan ketidakpuasan. Diantaranya, terdapat kendala terkait update terbaru, ketidakmampuan untuk digunakan secara offline, adanya sistem berbayar untuk membaca buku, serta kesulitan dalam proses registrasi dan saat membuka aplikasi. Pada tahap prosedur, aplikasi ini menghadapi beberapa kendala seperti kewajiban untuk *login* ulang setelah update, waktu yang diperlukan untuk memulai aplikasi, dan masalah saat registrasi. Selain itu, permasalahan harga mencakup ketidakmampuan membaca karya secara offline, beberapa konten yang berbayar, serta kebutuhan untuk berlangganan Wattpad Premium [8].

Dari dua sisi ini dapat dilihat jika pendapat masyarakat sebagai konsumen dibedakan menjadi dua, yakni pro (sentimen positif) dan kontra (sentimen negatif). Sentimen positif dapat menunjukkan keberhasilan suatu produk atau layanan yang dapat menguntungkan perusahaan dan juga pengguna, sedangkan sentimen negatif menunjukkan kekurangan yang harus diperbaiki oleh perusahaan. Namun, dengan basis pengguna Wattpad yang besar, diperlukan metode yang dapat mengklasifikasikan ulasan ke dalam kategori sentimen secara akurat, efisien, serta tidak menghabiskan waktu yang cukup lama dan memerlukan biaya yang besar apabila dilakukan secara manual [9].

Model pembelajaran mesin (*Machine Learning*) dapat mewujudkan dan memberikan kemudahan dalam mengidentifikasi sentimen. Berbagai model yang dapat diterapkan dalam klasifikasi teks, adalah *Random Forest*, *Naive Bayes* (NB), *Support Vector Machine* (SVM), pohon keputusan, *K-Nearest Neighbor*, ataupun metode lainnya. Teknik yang diterapkan dalam penelitian ini adalah metode *Random Forest*. *Random Forest* merupakan sebuah algoritma yang menggabungkan berbagai pohon keputusan untuk tujuan klasifikasi yang mampu mengatasi pencilan dan noise dan lebih mudah digunakan apabila dibandingkan metode lainnya [10]. Dalam metode *Random Forest*, keputusan akhir dicapai melalui suara mayoritas dari seluruh model Pohon Keputusan yang digunakan. Sehingga, *Random Forest* memiliki keunggulan dalam menangani volume data yang besar, serta kemampuannya untuk berperforma baik dalam kasus ketidakseimbangan kelas pada populasi [11]. Jika

dibandingkan dengan *support vector machine* (SVM) dan *Naïve Bayes*, algoritma *Random Forest* terbukti memiliki tingkat akurasi yang lebih tinggi [12].

Namun, ada kalanya suatu dataset yang digunakan mengalami anomali khusus seperti ketidakseimbangan kelas data. Dataset yang memiliki kelas yang tidak seimbang merupakan peristiwa ketika proporsi antar kelas respon tidak ekuivalen [13]. Ketidakseimbangan kelas dalam data pelatihan membuat model pembelajaran mesin mengklasifikasikan kelas yang lebih besar secara berlebihan karena peningkatan probabilitas sebelumnya. Akibatnya, algoritma pembelajaran mesin cenderung salah mengklasifikasikan kelas minoritas. Ini jelas menimbulkan masalah, sebab pemodelan akan menghasilkan akurasi prediksi yang lebih rendah untuk kelas minoritas dan akurasi prediksi yang lebih tinggi untuk kelas mayoritas [14]. Untuk itu, SMOTE (*Synthetic Minority Over-sampling Technique*) hadir untuk membuat data sintesis yang dapat menyeimbangkan distribusi kelas dengan menambah jumlah kelas minoritas untuk mengatasi *oversampling*.

Penelitian sebelumnya dilakukan oleh Fanka Angelina Larasati, Dian Eka Ratnawati, dan Buce Trias Hanggara (2022) yang mengklasifikasikan ulasan aplikasi Dana di Google Play Store menggunakan metode *Random Forest* dengan hasil akurasi model mencapai 84%. Pada penelitian analisis sentimen ulasan film dari situs *Rotten Tomatoes* digunakan metode *Random Forest* yang dimodifikasi seimbang memberikan akurasi 84,15% [15].

Berdasarkan paparan diatas, penelitian ini dilakukan untuk menjawab permasalahan bagaimana gambaran sentimen pengguna terhadap aplikasi Wattpad di Google Play Store berdasarkan ulasan-ulasan yang diberikan dan sejauh mana efektivitas metode *Random Forest* dalam mengklasifikasikan sentimen ulasan pengguna terhadap aplikasi Wattpad di Google Play Store, serta jika dilakukan perbandingan apakah dengan mengoptimalkan data latih pemodelan *Random Forest* dengan teknik SMOTE (*Synthetic Minority Over-sampling Technique*) dapat meningkatkan kinerja model.

2. METODE PENELITIAN

2.1. Populasi dan Sampel Penelitian

Populasi dalam penelitian ini adalah seluruh data ulasan atau komentar aplikasi Wattpad di Google Play Store sebanyak 4.000.000 ulasan. Sedangkan penghitungan jumlah sampel dalam penelitian ini menggunakan rumus slovin yang mampu mengatasi ukuran populasi yang besar dan besaran proporsi populasi yang ideal tidak diketahui [16]. Adapun persamaannya sebagai berikut.

$$n = \frac{N}{1 + Ne^2} \quad (1)$$

Dengan:

- n : jumlah sampel
- N : jumlah populasi
- E : batas dari kesalahan yang diijinkan (*margin of error*)

Pada umumnya, batas kesalahan yang diijinkan (*margin of error*) yang dapat ditolerir adalah 1%, 5% dan 10% karena dianggap masih representatif dalam penentuan sampel. Pada penelitian ini, akan digunakan *margin of error* sebesar 1 %. Dengan demikian berdasarkan persamaan 1, ukuran sampel yang akan digunakan dalam penelitian ini adalah 9.975.

2.2. Pengumpulan Data

Jenis data yang digunakan dalam penelitian ini adalah data sekunder, yaitu jenis data yang tidak dikumpulkan secara langsung oleh peneliti melainkan diperoleh dari sumber-sumber yang telah ada sebelumnya. Metode sampling yang digunakan adalah *cluster one stage* (gerombol satu tahap), yakni metode pengambilan sampel di mana peneliti membuat beberapa kelompok orang dari suatu populasi di mana mereka menunjukkan karakteristik yang homogen dan memiliki kesempatan yang sama untuk menjadi bagian dari sampel.

Pada penelitian ini, terdapat 9.975 data ulasan bahasa Indonesia aplikasi Wattpad di Google Play Store selama 5 tahun terakhir (2019-2023) dalam kategori paling relevan, yang dikumpulkan pada 17 April 2024 melalui metode *scraping*. Metode *scraping* membutuhkan API *Google-Play-Scraper*, yaitu *package* yang tersedia di dalam bahasa pemrograman Python untuk mengambil data dari aplikasi Google Play Store tanpa dependensi eksternal. Adapun informasi yang diambil adalah berikut ini.

- 1) *UserName*, ialah nama pengguna yang digunakan oleh seseorang untuk mengidentifikasi dirinya dalam platform *online*, dalam hal ini adalah Google Play Store.
- 2) *At*, ialah waktu atau tanggal yang menunjukkan kapan pengguna memberikan ulasan terhadap aplikasi.
- 3) *Content*, merupakan komentar, saran, kritik, ataupun masukan lainnya yang ditulis pengguna terkait pengalamannya selama menggunakan aplikasi.

2.3. Peubah Penelitian

Penelitian ini juga menggunakan peubah penelitian, yakni peubah respon dan peubah prediktor. Adapun penjelasannya ditampilkan dalam tabel berikut.

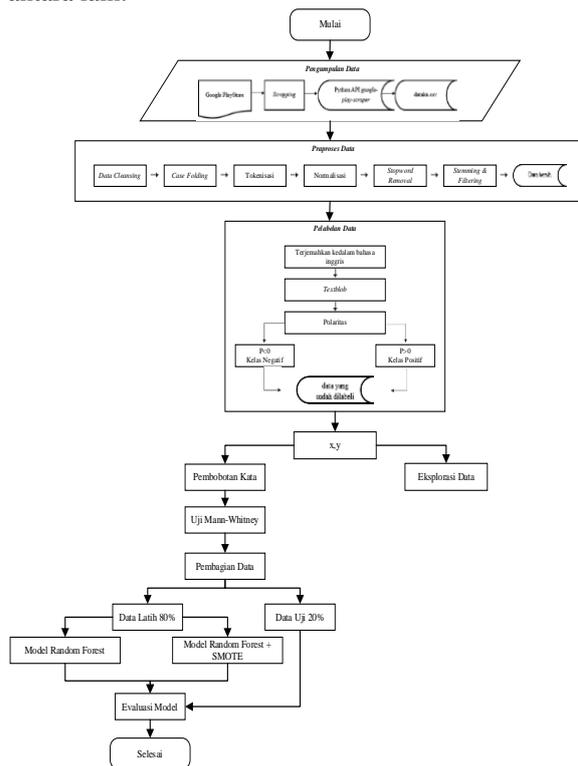
Tabel 1. Peubah Penelitian

Peubah	Definisi	Skala
X	Ulasan yang termuat dalam kolom data <i>content</i> namun sudah melewati proses mulai dari <i>preprocessing</i> sampai pembobotan kata dengan	Rasio

Peubah	Definisi	Skala
	perhitungan TF-IDF, dimana X_1, X_2, \dots, X_n adalah atribut kata.	
Y	Sentimen: 0 = sentimen negatif 1 = sentimen positif	Nominal

2.4. Prosedur Penelitian

Prosedur penelitian ini dilakukan dengan beberapa tahapan sesuai dengan *flow chart* yang terbentuk dalam Gambar 1. Uraian tahapan tersebut antara lain:



Gambar 1. Prosedur Penelitian

- Melakukan pengumpulan data dengan teknik scraping, hasilnya adalah 9.975 data ulasan aplikasi Wattpad dari Google Play Store yang kemudian disimpan dalam penyimpanan lokal berjudul "dataku.csv".
- Praproses data, adalah serangkaian proses agar data yang dianalisis berkualitas tinggi dengan tahapan data *cleansing*, *case folding*, tokenisasi, normalisasi, *stopword removal*, dan *stemming*.
- Pelabelan data menggunakan pustaka dari Python, yakni *Textblob*.
- Eksplorasi data, berisi visualisasi data sentimen menggunakan pie chart dan Awan Kata (*WorldCloud*) dari kata-kata yang sering muncul di tiap sentimen. Sehingga mampu menggambarkan ulasan seperti apa yang paling banyak disebutkan dalam sentimen.
- Pembobotan kata, adalah proses transformasi hasil data ulasan yang telah melewati rangkaian praproses menjadi numerik melalui TF-IDF.

- Untuk mereduksi atribut yang tidak signifikan digunakan Uji Mann-Whitney (*U Test*).
- Pembagian dataset, yakni pembagian keseluruhan data menjadi dua bagian dengan rasio 80% untuk data latih (*training set*) dan 20% untuk data uji (*testing set*).
- Melakukan pemodelan dengan metode klasifikasi algoritma *Random Forest*, dengan melihat informasi kedalaman pohon tertinggi dan terendah, serta jumlah pohon keputusan yang terbentuk.
- Optimasi ketidakseimbangan kelas sentimen menggunakan SMOTE (*Synthetic Minority Over-sampling Technique*).
- Melakukan Evaluasi pemodelan menggunakan *confusion matrix*, yakni 5 parameter pengukuran dari akurasi, presisi, *recall/sensitivity*, *F1-Score*, dan FPR. Selain itu digunakan juga Kurva ROC (*Receiver Operating Characteristic*) dan Nilai AUC serta nilai *feature importance* untuk melihat variabel mana sajakah yang memiliki pengaruh yang besar.

3. HASIL DAN PEMBAHASAN

3.1. Praproses Data

Sebelum data dianalisis lebih lanjut, terlebih dahulu dilakukan proses perubahan data agar lebih terstruktur, terorganisir, dan memudahkan proses analisis. Berikut ini diberikan rincian dari tiap praproses data yang dilakukan.

1) Data Cleansing

Sebagai tahapan awal praproses data, *data cleansing* bertujuan untuk menghapus atau membersihkan ulasan-ulasan yang terduplikasi dan nilai-nilai yang hilang (*missing values*).

Tabel 2. Data yang Terduplikasi

Waktu	Sebelum	Sesudah
7/26/2023 17:17	terlalu banyak iklan	terlalu banyak iklan
10/30/2023 6:30	terlalu banyak iklan	iklan

Ditemukan terdapat 2 ulasan yang terduplikasi dan tidak ada nilai yang hilang, maka menyisakan 9.973 data. Sisa data tersebut kemudian dibersihkan lagi dari *URL*, *mention*, *hashtag*, tanda baca, karakter emoji, dan angka-angka yang tidak mengandung sentimen. Contohnya disajikan dalam Tabel 3 pada 2 data pertama.

Tabel 3. Ulasan Sebelum dan Sesudah *Data Cleansing*

Dok	Sebelum	Sesudah
D1	Ini bagus sih tapi aku masukin tagar kok ngak keluar sih	Ini bagus sih tapi aku masukin tagar kok ngak keluar sih
D2	Gatau, daftar offline gw berubah dari 25 jadi 2.	Gatau daftar offline gw berubah dari jadi

2) *Case Folding*

Proses *case folding* adalah menyeragamkan huruf menjadi bentuk yang standar, dimana semua huruf berubah menjadi huruf kecil (*lower case*).

Tabel 4. Ulasan Sebelum dan Sesudah *Case folding*

Dok	Sebelum	Sesudah
D1	Ini bagus sih tapi aku masukin tagar kok ngak keluar sih	ini bagus sih tapi aku masukin tagar kok ngak keluar sih
D2	Gatau daftar offline gw berubah dari jadi	gatau daftar offline gw berubah dari jadi

3) Tokenisasi

Tahap tokenisasi dilakukan untuk memecah kalimat dalam bentuk kata per kata. Tanda spasi dalam tiap data komentar akan berperan sebagai pembentuk token sehingga dalam satu komentar terdiri dari beberapa token.

Tabel 5. Ulasan Sebelum dan Sesudah Tokenisasi

Dok	Sebelum	Sesudah
D1	ini bagus sih tapi aku masukin tagar kok ngak keluar sih	['ini', 'bagus', 'sih', 'tapi', 'aku', 'masukin', 'tagar', 'kok', 'ngak', 'keluar', 'sih']
D2	gatau daftar offline gw berubah dari jadi	['gatau', 'daftar', 'offline', 'gw', 'berubah', 'dari', 'jadi']

4) Normalisasi

Proses perbaikan kesalahan ketik, singkatan, dan kata tidak baku atau kata-kata gaul (*slang words*) dari token yang didapatkan hasil proses tokenisasi.

Tabel 6. Ulasan Sebelum dan Sesudah Normalisasi

Dok	Sebelum	Sesudah
D1	['ini', 'bagus', 'sih', 'tapi', 'aku', 'masukin', 'tagar', 'kok', 'ngak', 'keluar', 'sih']	['ini', 'bagus', 'sih', 'tapi', 'aku', 'masukkan', 'tagar', 'kok', 'tidak', 'keluar', 'sih']
D2	['gatau', 'daftar', 'offline', 'gw', 'berubah', 'dari', 'jadi']	['tidak tahu', 'daftar', 'offline', 'saya', 'berubah', 'dari', 'jadi']

5) *Stopword Removal*

Tahapan ini bertujuan untuk menghapus kata-kata yang tidak penting atau tidak berpengaruh terhadap makna dari teks.

Tabel 7. Ulasan Sebelum dan Sesudah *Stopword Removal*

Dok	Sebelum	Sesudah
D1	['ini', 'bagus', 'sih', 'tapi', 'aku', 'masukin', 'tagar', 'kok', 'tidak', 'keluar', 'sih']	['bagus', 'masukkan', 'tagar', 'keluar']
D2	['tidak tahu', 'daftar', 'offline', 'saya', 'berubah', 'dari', 'jadi']	['daftar', 'offline', 'berubah']

6) *Stemming dan Filtering*

Stemming mengubah kata ke bentuk dasarnya seperti menghilangkan imbuhan atau konjugasi yang tidak diperlukan. Sedangkan pemilahan data (*filtering*) dimaksudkan untuk mengecek apakah terdapat hasil praproses yang tidak mengandung sentimen ataupun lainnya.

Tabel 8. Ulasan Sebelum dan Sesudah *Stemming dan Filtering*

Dok	Sebelum	Sesudah
D1	['bagus', 'masukin', 'tagar', 'keluar']	bagus masuk tagar keluar
D2	['daftar', 'offline', 'berubah']	daftar offline ubah

3.2. Pelabelan Sentimen

Pelabelan sentimen merupakan proses pemberian nilai atau kategori sentimen terhadap komentar apakah termasuk dalam kategori negatif ataupun positif. Salah satu cara untuk melabeli sentimen dengan metode otomatis adalah dengan *Textblob*. Sebagai pustaka pemrosesan bahasa alami dalam Python, *Textblob* bekerja dengan cara menghitung polaritas dan subjektivitas teks [17]. Polaritas (*polarity*) adalah fungsi kecenderungan sentimen sebuah teks, yang menilai sebuah teks dalam rentang [-1,0 , 1,0]. Jika menerima skor mendekati -1,0, maka dikategorikan sebagai sentimen negatif dan jika skornya mendekati 1,0 maka dianggap mengandung sentimen positif.

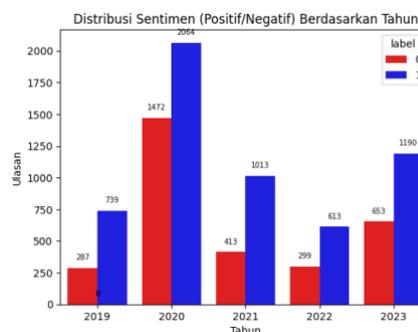
Data yang telah dipraproses diterjemahkan ke dalam bahasa Inggris menggunakan pustaka *googletrans* versi 3.1.0, kemudian dilabeli menggunakan *TextBlob* berdasarkan nilai polaritas. Polaritas kurang dari 0 dilabeli sebagai negatif, polaritas lebih dari 0 dilabeli sebagai positif, sementara polaritas sama dengan 0 (netral) dihapus untuk menghindari bias. Dengan demikian, dataset akhir hanya mencakup ulasan yang pro dan kontra terhadap aplikasi Wattpad. Hasil pelabelan sentimen pada setiap kelas klasifikasi disajikan dalam Tabel 9.

Tabel 9. Jumlah Pelabelan Sentimen di setiap Kelas

Klasifikasi Sentimen	Jumlah
Positif	5.616
Negatif	3.127
Total	8.743

3.3. Eksplorasi Data

Visualisasi dari sebaran sentimen ini ditunjukkan pada Gambar 2 dan 3.



Gambar 2. Distribusi Sentimen Berdasarkan Tahun

Berdasarkan Gambar 2, komentar positif mendominasi setiap tahunnya dengan puncaknya pada tahun 2020. Namun, terjadi lonjakan drastis

$$W_{ij} = tf_{i,j} \times df_i = \frac{n_{i,j}}{\sum_k n_{k,i}} \times \log \frac{N}{df_i} \quad (2)$$

Keterangan:

$tf_{i,j}$: term frequency (term ke- i dalam dokumen ke- j)

$n_{i,j}$: jumlah kemunculan term ke- i pada dokumen ke- j

$\sum_k n_{k,i}$: jumlah kemunculan seluruh term pada dokumen ke- j

idf_i : inverse document frequency untuk (term ke- i)

N : jumlah seluruh dokumen

df_i : jumlah dokumen yang mengandung term ke- i

Pada penelitian ini, 8.743 dataset, sebanyak 5.724 fitur atau atribut kata yang berhasil di representasikan vektor TF-IDF, yang digunakan sebagai peubah bebas penelitian dan selanjutnya memasuki tahap pemodelan

3.5. Uji Beda Nyata Mann-Whitney (U Test)

Uji Mann-Whitney (U Test) merupakan uji statistik nonparametrik yang digunakan untuk menilai perbedaan antara dua kelompok independen ketika data bersifat nominal (sentimen positif dan negatif) namun tidak mengikuti distribusi normal. Oleh karena itu sebelum dilakukan pengujian, dilakukan pengecekan asumsi normalitas pada tiap atribut atau fitur seperti berikut.

Tabel 10. Uji Normalitas tiap Atribut

Atribut/Fitur	p -value	Keterangan
abad	0	Tidak Berdistribusi Normal
abai	0	Tidak Berdistribusi Normal
abang	0	Tidak Berdistribusi Normal
abdi	0	Tidak Berdistribusi Normal
⋮	⋮	⋮
zoom	0	Tidak Berdistribusi Normal

Berdasarkan Tabel 10, hasil uji normalitas menggunakan *Kolmogorov-Smirnov* Test ialah:

- Hipotesis
 H_0 : Data berdistribusi normal
 H_1 : Data tidak berdistribusi normal
- Selang Kepercayaan, yang digunakan sebesar 95%, maka $\alpha=0,05$
- Kriteria Pengujian
 H_0 ditolak apabila p -value < α
 H_1 diterima apabila p -value $\geq \alpha$
- Keputusan

Nilai p -value adalah 0 pada semua atribut yang berarti p -value < α (0,05) maka H_0 ditolak, atau tidak cukup bukti untuk mengatakan bahwa data berdistribusi dengan normal. Maka selanjutnya dapat dilakukan pengujian Mann-Whitney.

Berikut ini ditampilkan uraian pengujian Mann-Whitney pada 5.724 Fitur atau Atribut.

- Hipotesis
 H_0 : Tidak ada perbedaan yang signifikan antara kelas sentimen positif dan negatif.
 H_1 : Terdapat perbedaan yang signifikan antara kelas sentimen positif dan negatif.
- Selang Kepercayaan, yang digunakan sebesar 95%, maka $\alpha=0,05$
- Kriteria Pengujian
 H_0 ditolak apabila p -value < α
 H_1 diterima apabila p -value $\geq \alpha$
- Keputusan
 Hasil pengujian dilampirkan dalam Tabel 11 berikut.

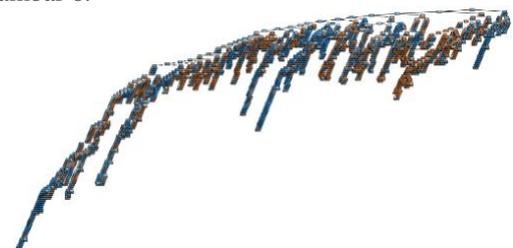
Tabel 11. Uji Mann-Whitney

Atribut	U -statistic	p -value	Keterangan
abad	8781860	0,674726395	Tidak Signifikan
abai	8775604	0,328709201	Tidak Signifikan
abang	8784669,5	0,264032831	Tidak Signifikan
abdi	8779052,5	0,455695788	Tidak Signifikan
abrek	8779052,5	0,455695788	Tidak Signifikan
absen	8777489	0,291349422	Tidak Signifikan
absurd	8783424	0,180277975	Tidak Signifikan
abu	8775609	0,329191997	Tidak Signifikan
acak	8847366	1,48E-06	Signifikan
⋮	⋮	⋮	⋮
zoom	8783424	0,180278	Tidak Signifikan

Sebanyak 277 fitur atau atribut dari 5.724 saja yang memiliki p -value < α (0,05) sehingga 277 fitur atau atribut itu kemudian dilanjutkan dalam pemodelan klasifikasi dengan *Random Forest*.

3.6. Klasifikasi *Random Forest*

Proses klasifikasi *Random Forest* diawali dengan pembagian dataset, dimana 80% data akan menjadi data latih sebanyak 6.994, dan 20% sisanya menjadi data uji sebanyak 1.749 data dari 277 atribut atau fitur yang tersisa. Dari 100 pohon keputusan yang terbentuk, kedalaman pohon tertinggi adalah 149 dan terendah ialah 83 tingkat atau lapisan, salah satu contohnya seperti yang ditunjukkan dalam Gambar 6.



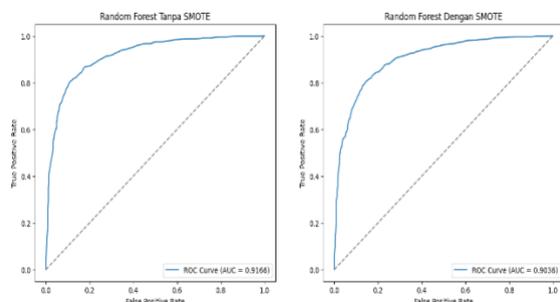
Gambar 6. Contoh Pohon Keputusan dengan *Random Forest*

Untuk melihat lebih dekat, telah diatur jika tingkat kedalaman (*max depth*) yang ditampilkan=3,

Pada Tabel 12, ditunjukkan jika *Random Forest* tanpa SMOTE cenderung lebih unggul. *Random Forest* dengan SMOTE memang mampu menaikkan nilai presisi sebesar 1,99% tetapi ini juga berdampak langsung pada penurunan performa *classification report* yang lain, yakni akurasi, *recall*, *F1-Score*, dan naiknya 3,66% nilai FPR. Hal ini menunjukkan bahwa penggunaan SMOTE dapat menghasilkan peningkatan dalam identifikasi positif palsu, namun secara keseluruhan *Random Forest* tanpa SMOTE lebih baik dalam mengidentifikasi kasus kesalahan prediksi secara keseluruhan, karena mampu menjaga keseimbangan yang lebih baik antara metrik-metrik evaluasi utama.

2) Kurva ROC dan Nilai AUC

Dalam upaya melihat seberapa baik model dapat membedakan dua kelas sentimen, disajikan dalam Kurva ROC (*Receiver Operating Characteristic*) serta nilai AUC (*Area Under the Curve*) berikut.



Gambar 9. Perbandingan ROC-AUC Pemodelan

Berdasarkan Gambar 9, dua kurva ROC menunjukkan kinerja model *Random Forest* dalam klasifikasi sentimen, satu menggunakan SMOTE dan satu tanpa SMOTE. Meskipun bentuk kurva ROC serupa, terdapat perbedaan pada nilai AUC. Model tanpa SMOTE memiliki nilai AUC 0,9166, sedangkan model dengan SMOTE memiliki nilai AUC 0,9036. Jika dilihat berdasarkan pengkategorian AUC [19], nilai AUC 0,9166 dikategorikan sebagai klasifikasi sangat baik, sementara nilai 0,9036 dikategorikan sebagai klasifikasi baik. Sehingga, model *Random Forest* tanpa SMOTE menunjukkan kinerja yang sedikit lebih baik dalam membedakan antara sentimen positif dan negatif.

3) Feature Importance

Dalam *Random Forest*, pentingnya fitur diukur berdasarkan frekuensi penggunaan fitur untuk membagi node dan besarnya penurunan kriteria (seperti *Gini impurity*) yang dihasilkan. Semakin tinggi skor fitur, semakin besar pengaruhnya dalam memprediksi variabel tertentu. Untuk itu, dari 277 fitur atau variabel yang digunakan dalam penelitian ini, ditampilkan 10 fitur yang paling berpengaruh pada kedua model.

Tabel 13. Nilai *Feature Importance* dari 10 Fitur Utama

No,	Random Forest		Random Forest + SMOTE	
	Fitur	Nilai FI	Fitur	Nilai FI
1	kecewa	0,115837	kecewa	0,112788
2	bagus	0,089163	bagus	0,106109
3	baik	0,040974	baik	0,047040
4	ganggu	0,028049	aplikasi	0,027311
5	cerita	0,025198	ganggu	0,027198
6	aplikasi	0,024784	cerita	0,026309
7	iklan	0,024673	iklan	0,025934
8	tolong	0,023492	tolong	0,025389
9	susah	0,022396	suka	0,023569
10	suka	0,021852	susah	0,019626

Pada Tabel 13, ditunjukkan jika terdapat 10 fitur atau variabel yang memberikan sumbangsih pengaruh yang paling besar dalam memprediksi kelas sentimen diantara 277 fitur lainnya. Kata "kecewa" mendominasi dengan nilai *feature importance* tertinggi di kedua model, menunjukkan bahwa kemunculan kata tersebut sangat kuat dalam memprediksi sentimen negatif. Sementara itu, kata-kata seperti "bagus" dan "baik" juga memiliki kontribusi besar, mengindikasikan bahwa keduanya memainkan peran penting dalam membedakan sentimen positif. Secara keseluruhan, keduanya sama hanya berbeda penempatan atau urutan saja.

4. KESIMPULAN

Berdasarkan hasil penelitian, diperoleh kesimpulan bahwa:

- 1) Dari 9.975 hasil pengumpulan data ulasan aplikasi Wattpad di Google Play Store, terdapat 8.743 data yang dianggap bersih dan siap untuk di analisis. Sebanyak 64,2% (5.616) sentimen positif diberikan pengguna yang menyatakan menyukai dan mengapresiasi aplikasi Wattpad ditandai dengan berbagai ulasan seperti "suka", "baik", dan "aplikasi bagus". Sisanya, yakni 35,8% (3.127) sentimen negatif yang tidak menyukai atau merasa dirugikan telah mengunduh aplikasi Wattpad yang ditandai dengan komentar seperti "kecewa", "iklan", dan "loading".
- 2) Pada kasus ini, pemodelan *Random Forest* sudah mampu mengklasifikasikan sentimen dengan baik. Nilai akurasi mencapai 84,05%, presisi 84,71%, *recall* 91,60%, *F1-Score* 88,02%, dan nilai FPR mencapai 8,40%. Kemudian apabila dilihat dari nilai AUC, pemisahan antar kelas juga diklasifikasikan sangat baik yakni 0,9166 dengan fitur yang paling berpengaruh adalah atribut kata "kecewa", "bagus", "baik", dan sebagainya. Secara keseluruhan, tanpa di optimasi *Random Forest* sudah mampu mengklasifikasikan dua kelas sentimen yang tidak seimbang dan bervolume besar dari Aplikasi Wattpad.

- 3) Pemodelan *Random Forest* yang di optimasikan dengan SMOTE mampu meningkatkan kemampuan untuk mengklasifikasikan kelas minoritas yakni sentimen negatif. Sebelum menggunakan SMOTE, model lebih cenderung mengidentifikasi kelas mayoritas yaitu sentimen positif. Ini dapat dilihat dari naiknya nilai presisi dari 84,71% ke 86,70% (1,99%). Sayangnya, penyeimbangan kelas ini berdampak pada penurunan performa seperti, akurasi, recall, f1-score dan nilai AUC. Adapun fitur yang paling berpengaruh juga adalah atribut kata “kecewa”, “bagus”, “baik”, dan seterusnya.

5. DAFTAR PUSTAKA

- [1] Statista, "Web Developers: primary source of improvement feedback," accessed December 12, 2023. [Online]. Available: <https://www.statista.com/statistics/271544/web-developer-survey-most-important-sources-for-identifying-site>.
- [2] BrightLocal, "93% of online consumers read reviews before making a purchase or downloading an app," accessed December 12, 2023. [Online]. Available: <https://www.brightlocal.com/research/local-consumer-review-survey/>
- [3] B. Liu, *Sentiment Analysis: Mining Opinions, Sentiments, and Emotions*. Cambridge, UK: Cambridge University Press, 2015.
- [4] M. R. Pribadi, D. Manongga, H. D. Purnomo, I. Setyawan, and Hendry, "Sentiment Analysis of the PeduliLindungi on Google Play using the Random Forest Algorithm with SMOTE," *ISITIA 2022-Proceeding*, pp. 115-119, 2022.
- [5] Media Indonesia, "7 Rekomendasi Aplikasi Novel Gratis yang Bisa Bikin Cuan," accessed April 28, 2024. [Online]. Available: <https://mediaindonesia.com/humaniora/633350/7-rekomendasi-aplikasi-novel-gratis-bisa-bikin-cuan>.
- [6] N. Wulansari, "Pengaruh Penggunaan Aplikasi Wattpad Terhadap Kemampuan Menulis Naskah Drama," in *Seminar Nasional Bahasa dan Sastra Indonesia*, 2018.
- [7] SimilarWeb, "Wattpad," accessed December 12, 2023. [Online]. Available: <https://www.similarweb.com/website/wattpad.com>.
- [8] S. I. Nurhafida and F. Sembiring, "Analisis Sentimen Aplikasi Novel Online Di Google Play Store Menggunakan Algoritma Support Vector Machine (SVM)," *J. Sains Komput. & Informatika (J-SAKTI)*, vol. 6, no. 1, pp. 317-327, 2022.
- [9] F. A. Larasati, D. E. Ratnawati, and B. T. Hanggara, "Analisis Sentimen Ulasan Aplikasi Dana dengan Metode Random Forest," *J. Pengemb. Teknol. Inf. Dan Ilmu Komput.*, vol. 6, no. 9, pp. 4305-4313, 2022.
- [10] T. Bouabana-Tebibel and S. H. Rubin, *Theoretical Information Reuse and Integration*. Switzerland: Springer, 2016.
- [11] E. Renata and M. Ayub, "Penerapan Metode Random Forest untuk Analisis Risiko pada Dataset Peer to Peer Lending," *J. Tek. Inform. Dan Sist. Inf.*, vol. 6, no. 3, pp. 462-474, 2020.
- [12] E. Fitri, Y. Yuliani, and W. Gata, "Analisis Sentimen Terhadap Aplikasi Ruangguru Menggunakan Algoritma Naive Bayes, Random Forest, dan Support Vector Machine," *J. Transformatika*, vol. 88, no. 1, pp. 71-80, 2020.
- [13] M. Syukron, R. Santoso, and T. Widiharih, "Perbandingan Metode SMOTE Random Forest dan SMOTE XGBoost untuk Klasifikasi Tingkat Penyakit Hepatitis C pada Imbalance Class Data," *J. Gaussian*, vol. 9, no. 3, pp. 227-236, 2020.
- [14] E. Erlin, Y. Desnelita, N. Nasution, L. Suryati, and F. Zoromi, "Dampak SMOTE terhadap Kinerja Random Forest Classifier berdasarkan Data Tidak Seimbang," *J. Manaj., Tek. Inform. Dan Rekayasa Komput.*, vol. 21, no. 3, pp. 677-690, 2022.
- [15] [M. R. Nugraha, M. D. Purbolaksono, and W. Astuti, "Sentiment Analysis on Movie Review from Rotten Tomatoes Using Modified Balanced Random Forest Method and Word2Vec," *Building of Informatics, Technology and Science (BITS)*, vol. 5, no. 1, pp. 153-161, 2023.
- [16] [16] A. Santoso, "Rumus Slovin: Panacea Masalah Ukuran Sampel?" *Suksma J. Psikol. Univ. Sanata Dharma*, vol. 4, no. 2, pp. 24-43, 2023.
- [17] A. Baita and N. Cahyono, "Analisis Sentimen Mengenai Vaksin Sinovac Menggunakan Algoritma Support Vector Machine (SVM) dan K-Nearest Neighbor (KNN)," *Infos*, vol. 4, no. 2, pp. 42-42, 2021.
- [18] Wattpad, "Pusat Bantuan," accessed March 8, 2024. [Online]. Available: <https://support.wattpad.com/hc/id>.
- [19] D. G. Kleinbaum and M. Klein, *Statistics for Biology and Health: Logistical Regression*. New York, NY, USA: Springer, 2010.