JTERA (Jurnal Teknologi Rekayasa), Vol. 9, No. 2, Desember 2024, Hal. 77-84 p-ISSN 2548-737X
Terakreditasi "**Peringkat 3**" oleh Kemenristek/BRIN, Nomor SK: 85/M/KPT/2020 e-ISSN 2548-8678
DOI: 10.31544/jtera.v9.i1.2024.77-84

ANALISIS SENTIMEN OPINI TERHADAP NOVEL PADA WEBSITE GOODREADS MENGGUNAKAN METODE NAIVE BAYES CLASSIFIER

Rahma Zahrani[1], Nita Rosa Damayanti[2], Evi Yulianingsih[3], Muhamad Ariandi[4]
[1,2,3,4]Program Studi Sistem Informasi, Universitas Bina Darma
Jl. Jenderal Ahmad Yani, Kota Palembang, Sumatera Selatan 30111, Indonesia
201410093@student.binadarma.ac.id

Abstrak

Pertumbuhan teknologi yang pesat memudahkan seseorang dalam menyampaikan berbagai opini secara *online*, termasuk mengenai pengalaman membaca buku. Banyak pembaca ingin mengetahui apakah suatu buku menarik menurut pandangan orang lain sebelum memutuskan buku mana yang akan dibaca. Namun, banyaknya komentar dan opini pada suatu novel mengakibatkan seseorang kesulitan untuk membaca satu per satu opini dan mengambil kesimpulan mengenai novel tersebut. Penelitian ini bertujuan untuk mengetahui kecenderungan publik dan memberikan gambaran mengenai novel "Bumi" berdasarkan opini di *website* Goodreads, serta untuk mengetahui performa Naïve Bayes dalam memberikan klasifikasi sentimen. Metode yang digunakan untuk proses analisis sentimen adalah Naïve Bayes dengan menggabungkan metode Random Oversampling dalam proses *resampling* data latih. Klasifikasi Naïve Bayes dalam memberikan sentimen mendapatkan performa yang cukup baik pada nilai akurasi yakni sebesar 79% dan opini pembaca terhadap novel "Bumi" pada *website* Goodreads cenderung positif dengan persentase 85.26% opini positif, serta berdasarkan visualisasi *wordcloud* pada opini novel "Bumi" dapat disimpulkan novel tersebut bertema fantasi dan mempunyai alur yang lambat bagi sebagian orang.

Kata kunci: Goodreads, Naïve Bayes, Opini, Random Oversampling

Abstract

The rapid growth of technology has facilitated the sharing of various opinions online, including experiences with reading books. Many readers want to know if a book is interesting based on others' perspectives before deciding which book to read. However, the vast number of comments and opinions on a novel can make it difficult for someone to read each opinion and draw conclusions about the novel. This study aims to understand public tendencies, provide an overview of the novel "Bumi" based on opinions from the Goodreads website, and evaluate the performance of the Naïve Bayes algorithm in sentiment classification. The method used for the sentiment analysis process is Naïve Bayes, combined with the Random Oversampling method for resampling the training data. The Naïve Bayes classification achieved a solid performance in sentiment analysis, with an accuracy rate of 79%. Reader opinions on the novel "Bumi" on Goodreads tend to be positive, with 85.26% positive opinions. Based on the word cloud visualization of the "Bumi" novel opinions, it can be concluded that the novel is fantasy-themed and has a slow pace for some readers.

Keywords: *Goodreads, Naïve Bayes, Opinion, Random Oversampling*

I. PENDAHULUAN

Pertumbuhan teknologi yang pesat memudahkan seseorang dalam menyampaikan berbagai opini secara *online*, termasuk mengenai pengalaman membaca suatu buku. Opini suatu produk secara *online* adalah sumber informasi penting bagi konsumen, khususnya di dunia yang didorong secara digital saat ini [1]. Banyak pembaca ingin mengetahui menarik tidaknya suatu buku menurut pandangan orang lain sebelum memutuskan buku mana yang akan dibaca [2]. Testimoni dan opini yang muncul di berbagai media sosial dari para pembaca dapat menunjukkan hal ini. Sebuah

Diterima: 28 Mei 2024; Direvisi: 21 September 2024; Disetujui: 25 September 2024
JTERA, Vol. 9, No. 2, Desember 2024

penelitian menunjukkan bahwa *review* konsumen secara *online* memberikan dampak baik pada keputusan pembelian *online* pada produk Erigo di provinsi Bali. Hal ini menunjukkan bahwa tingginya pembelian *online* dapat dipengaruhi oleh *review* konsumen secara *online* yang tinggi. Produk dengan banyak ulasan positif memiliki peluang lebih tinggi untuk diminati konsumen [3]. Melalui hasil penelitian tersebut menunjukkan bahwa opini atau *review* dapat berpengaruh positif terhadap suatu objek tertentu.

Salah satu *website* untuk berbagi opini mengenai buku di media sosial adalah Goodreads.com. Pada situs Goodreads ini pengguna dapat mengkategorikan buku apa pun dan mempublikasikan tanggapan mereka terhadap suatu buku seperti pujian, komentar, analisis kritis, narasi pribadi, serta terdapat berbagai fitur menarik seperti pencarian, memberikan ulasan, menyimpan buku yang akan dibaca, bahkan terdapat peringkat suatu buku berdasarkan kategori tertentu [4]-[5]. Salah satu novel yang termasuk dalam jajaran novel Indonesia terbaik versi *website* tersebut adalah novel yang berjudul "Bumi" karya Tere Liye sehingga novel tersebut mendapatkan cukup banyak *review* di *website* Goodreads. Banyaknya komentar serta opini pada novel yang berjudul "Bumi" tersebut mengakibatkan seseorang sulit membaca satu persatu opini untuk mengambil kesimpulan mengenai novel tersebut sehingga dapat dilakukan penelitian mengenai analisis sentimen untuk mengetahui kecenderungan publik serta memberikan gambaran mengenai novel "Bumi".

Metode Naïve Bayes adalah metode klasifikasi teks yang dikenal sangat baik dalam melakukan klasifikasi sentimen karena asumsi kisinya yang efektif dan implementasinya yang cepat dan mudah [6]. Penelitian sebelumnya dalam mengklasifikasi sentimen terhadap suatu film menunjukkan bahwa metode Naïve Bayes menghasilkan tingkat akurasi mencapai 75%, 80% pada nilai *precision*, dan tingkat keberhasilan (*recall*) 79% [7]. Pada penelitian lainnya dalam mengklasifikasikan opini terhadap suatu aplikasi kecantikan, Naïve Bayes mendapatkan akurasi yang lebih tinggi yaitu 90 persen [8]. Penelitian lain dengan menggunakan metode Naïve Bayes serta mengkombinasikan metode *resampling* yaitu teknik Random Oversampling dan Random Undersampling. Hasil yang didapatkan adalah Naïve Bayes mendapatkan akurasi yang lebih tinggi pada teknik Random Oversampling jika dibandingkam dengan teknik Random Undersampling yakni sebesar 85.55% [9].

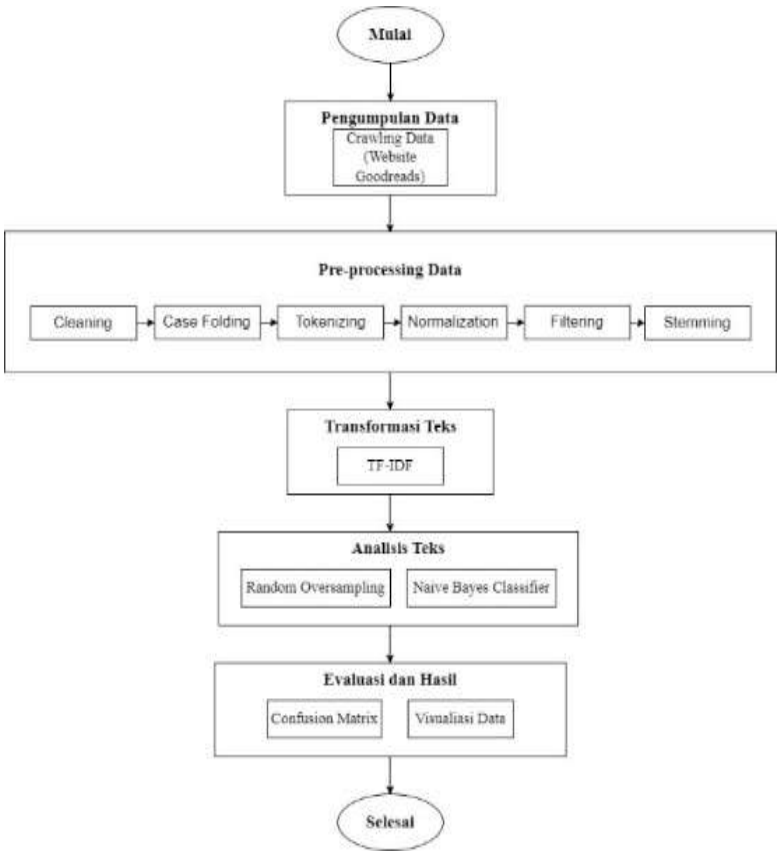
Pada literatur penelitian lain dengan membandingkan teknik pembagian data antara metode Hold-Out dan metode K-Fold Cross Validation didapatkan kinerja Naïve Bayes menggunakan teknik Hold-Out lebih besar 1% yakni 83% dari teknik 10-fold cross validation yang mendapatkan akurasi 82% [10]. Literatur lain yang berorientasi pada analisis sentimen terhadap respon masyarakat di media sosial mengenai lelucon *satire* dengan menggunakan dua metode yakni Naïve Bayes dan KNN. Berdasarkan evaluasi performa Naïve Bayes menunjukkan akurasi yang baik sebesar 90,29%. Sedangkan KNN hanya mencapai akurasi 60,75% yang menunjukkan bahwa Naïve Bayes lebih efektif dalam mengidentifikasi dan mengklasifikasikan sentimen [11].

Penelitian lain yang mengkaji metode serupa dalam melakukan analisis sentimen terhadap opini masyarakat tentang mobil listrik lebih cenderung pada sentimen positif atau negatif. Hasil penelitian tersebut, Naïve Bayes mendapatkan akurasi yang cukup baik yaitu sebesar 77.8% dengan komposisi pembagian dataset 70:30 [12]. Penelitian lainnya dengan menggunakan metode Naïve Bayes pada opini mengenai aplikasi Nanovest mendapatkan akurasi yang sangat baik pada saat melakukan klasifikasi yaitu sebesar 94% [13]. Berdasarkan beberapa penelitian tersebut, Naïve Bayes terbukti cukup akurat dalam melakukan klasifikasi sentimen data teks.

Penelitian ini memiliki fokus pada analisis sentimen terhadap opini pembaca novel "Bumi" karya Tere Liye di *website* Goodreads dengan menggunakan metode Naïve Bayes dan mengkombinasikan metode *resampling* dengan menggunakan teknik Random Oversampling untuk menangani ketidakseimbangan antar kelas sentimen. Proses penelitian dilakukan dengan memanfaatkan bahasa pemrograman Python dan hasil dari penelitian berupa akurasi dari penggunaan Naïve Bayes dalam melakukan klasifikasi sentimen dan mengetahui kecenderungan sentimen publik pada *website* Goodreads mengenai novel "Bumi" karya Tere Liye.

II. METODE PENELITIAN

Penelitian ini mencakup langkah-langkah yang terstruktur dengan menggunakan tahap-tahap umum dalam penelitian analisis sentimen seperti pengumpulan data, pre-processing, transformasi teks, serta evaluasi dan hasil. Tahap-tahap tersebut dapat dilihat pada Gambar.1



Gambar 1. Alur Penelitian

A. Pre-processing Data

Pre-processing data adalah proses dimana data teks dinormalisasikan menjadi data yang dapat lebih mudah diolah saat dilakukan analisis sentimen [14]. Beberapa tahap pra-pemrosesan data ini mencakup *cleaning, case folding, tokenizing, normalization, filtering,* dan *stemming*. Dimulai dari *cleaning* yang digunakan untuk proses pembersihan data seperti menghilangkan tanda baca, angka, simbol, dan *whitespace*. *Case folding* adalah proses transformasi setiap kata yang terdapat di dalam dataset menjadi huruf non-kapital. *Tokenizing* adalah proses kata-kata pada dataset dipecahkan dalam suatu kalimat menjadi bentuk token, dimana setiap kata dalam suatu kalimat dipisahkan dengan spasi. *Filtering* disebut juga dengan tahap *stopword removing*. Tahap ini melakukan penghilangan berbagai kata yang tidak berpengaruh dalam sebuah kalimat ketika dilakukan analisis sentimen seperti kata "tidak", "dengan", "yang", "dan", "ada". *Stemming* untuk menghilangkan prefiks dan sufiks pada setiap token.

B. TF-IDF

Term Frequency-Inverse Document Frequency atau yang biasa disingkat TF-IDF biasa digunakan untuk menghitung tingkat kepentingan sebuah kata di dalam himpunan teks atau dokumen [15]. TF-IDF memiliki landasan konseptual yang memahami pentingnya kata-kata di dalam suatu dokumen atau korpus. Konsep dasar TF menghitung frekuensi kemunculan sebuah kata dalam dokumen [16], dimana jumlah suatu kata yang muncul pada dokumen dibagi berdasarkan jumlah kata yang ada dalam dokumen. Maka dapat dirumuskan sebagai persamaan (1).

$$TF\ t,d = \frac{jumlah\ t\ dalam\ d}{jumlah\ kata\ pada\ d} \quad (1)$$

Sedangkan IDF menilai seberapa unik atau penting sebuah kata yang terdapat di dalam suatu korpus dokumen. [17]. Persamaan IDF dapat disesuaikan dengan rumus (2).

$$idf\ (t) = log\frac{jumlah\ keseluruhan\ dokumen}{jumlah\ dokumen\ yang\ memuat\ t} \quad (2)$$

Gabungan dari keduanya menciptakan bobot untuk setiap kata sehingga dapat memungkinkan pemahaman yang lebih baik tentang signifikansi relatifnya dalam konteks tertentu [17]. Kemudian pada penilaian TF-IDF dilakukan dengan menghitung perkalian antara persamaan TF dengan persamaan IDF seperti pada rumus persamaan (3).

$$TF\ IDF = TF_{(t,d)}\ \mathrm{X}\ IDF_{(t)} \quad (3)$$

C. Random Oversampling

Data yang telah dikumpulkan perlu mempunyai kelas yang seimbang agar antara kelas dapat menghasilkan timgkat akurasi prediksi yang baik. Random Oversampling adalah salah satu teknik yang menambahkan data pada kelas minoritas secara acak tanpa menambah variasi data kelas sehingga data dari setiap kelas bisa sama atau seimbang. Pendekatan ini membuat replika dari kelas minoritas, replikasi yang dikenal sebagai data sintetis. Setiap data minoritas dibuat dari data sintetis sebanyak persentase duplikasi yang diinginkan [9]-[18].

D. Naive Bayes Classifier

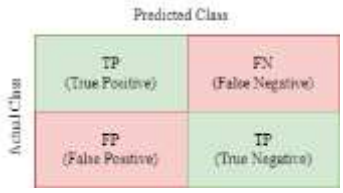
Klasifikasi Naive Bayes merupakan metode pengklasifikasi probabilitas sederhana yang mampu menghitung distribusi probabilitas menggunakan frekuensi dan kombinasi nilai yang ada dalam kumpulan data [19]. Naïve Bayes dapat diformulasikan menjadi persamaan (4) berikut:

$$P(H|X) = \frac{p(X|H)p(H)}{p(X)} \quad (4)$$

Keterangan:
P(H|X) : Nilai probabilitas hipotesa H didasarkan pada kondisi,
P(H) : Nilai probabilitas hipotesa H
P(X|H) : Nilai probabilitas X yang didasarkan pada kondisi dalam hipotesis
P(X) : Probabilitas X

E. Confusion Matrix

Untuk mengukur suatu model agar dapat dikatakan bahwa model tersebut sudah cukup baik, maka perlu dilakukan evaluasi terhadap model tersebut. Mengevaluasi suatu model adalah dengan cara menghitung perbandingan antara nilai aktual dengan nilai prediksi. Confusion matrix dapat digunakan sebagai alat yang memiliki kemampuan untuk menilai seberapa baik model pengklasifikasi dalam memprediksi suatu data [20]. Matriks ini adalah matriks dua dimensi dalam bentuk tabel matriks seperti pada gambar 2.



Gambar 2. Confusion Matrix

Keterangan:
True positive: data pada kelas positif yang tepat diprediksi,
False positive: data pada kelas positif yang tidak tepat diprediksi,
True negative: data pada kelas negatif yang tepat diprediksi,
False negative: data pada kelas negatif yang tidak tepat diprediksi.

Nilai yang telah diperoleh dapat digunakan untuk menemukan parameter evaluasi seperti *Accuracy*, *Precision*, *Recall*, dan *F1-Score*. Rumus yang digunakan untuk mengukur parameter tersebut dapat dilihat pada persamaan (5), (6), (7), dan (8).

$$accuracy = \frac{TP+TN}{TP+TN+FP+FN} \quad (5)$$

$$precision = \frac{TP}{TP+FP} \quad (6)$$

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

$$f1-score = \frac{2(recall\ X\ precision}{recall+precision} \quad (8)$$

III. HASIL DAN PEMBAHASAN

A. Pengumpulan Data

Data yang digunakan sebagai penelitian adalah data opini atau pengguna *website* Goodreads tentang novel ”Bumi”. Pengambilan data opini menggunakan *tools extension* Instant Data Scrapper di Google Chrome dimulai dari Februari 2024 – Maret 2024. Data yang diambil berjumlah 1486 yang dikelompokkan menjadi sentimen negatif dan positif sesuai dengan komentar yang diberikan pengguna secara manual. Tabel 1. menjelaskan beberapa contoh data yang telah diambil dan dilabeli secara manual.

Tabel 1. Data Opini

No	Opini	Label
1	Alur ceritanya agak lambat bagi saya	0
2	Alurnya lambat dan membosankan. Sifat karakternya mirip dan garing. Banyak hal-hal tidak berkaitan yang dimasukkan ke jalan cerita yang malah bikin tambah membosankan. Plotnya juga biasa aja dan klise, banyak ditemui di cerita kkpk. Akhir cerita bisa ditebak.	0
3	aku suka buku petualangan ini	1
4	Fantasinya keren deh, nggak nyangka bukunya bakal sebagus ini.	1
5	bosan parah sama gaya nulisnya	0

Pada proses pelabelan untuk mempermudah klasifikasi, sentimen positif diberi angka 1 dan sentimen negatif diberi angka 0. Berdasarkan hasil pelabelan tersebut didapatkan data dengan sentimen positif sebanyak 1243 data dan data dengan sentimen negatif sebanyak 242 data tertera pada grafik Gambar 3. berikut:

Gambar 3. Grafik Label Sentimen

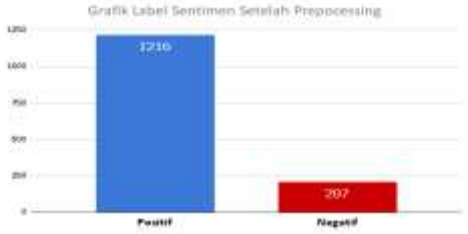
B. Pre-processing Data

Pada tahap ini data teks dinormalisasikan menjadi data yang dapat mudah diolah saat dilakukan proses klasifikasi sentimen. Pre-processing data dilakukan dengan menggunakan bantuan dari beberapa *library* pada bahasa pemrograman Python. Proses ini meliputi tahapan *case folding, cleaning, tokenizing, filtering, normalization* dan *stemming*. Hasil dari tahap pre-processing dapat dilihat pada Tabel 2.

Tabel 2. Tahap Pre-processing

Data mentah	Ceritanya bagus, juga berisi fakta yang bisa aku pelajari.
Proses Cleaning & Case Folding	ceritanya bagus juga berisi fakta yang bisa aku pelajari
Proses Tokenizing	['ceritanya', 'bagus', 'juga', 'berisi', 'fakta', 'yang', 'bisa', 'aku', 'pelajari']
Proses Normalization	['ceritanya', 'bagus', 'juga', 'berisi', 'fakta', 'yang', 'bisa', 'saya', 'pelajari']
Proses Filtering	['ceritanya', 'bagus', 'berisi', 'fakta', 'pelajari']
Proses Stemming	['cerita', 'bagus', 'isi', 'fakta', 'ajar']

Setelah melalui beberapa tahap proses pre-processing, dilakukan penghapusan nilai null dan penghapusan duplikat data yang ada pada data teks. Dari hasil penghapusan tersebut, data berkurang dari 1486 menjadi 1423 baris data, dimana data dengan label sentimen positif sebanyak 1216 dan data dengan label sentimen negatif sebanyak 207 seperti yang ditunjukkan pada Gambar 4. berikut:



Gambar 4. Grafik Label Sentimen Setelah Pre-processing

C. Transformasi Teks

Setelah melalui proses pre-processing data, data akan diubah menjadi bentuk representasi vektor sebelum siap dilakukan klasifikasi menggunakan Naïve Bayes. Pada pengaplikasiannya, proses ini menggunakan metode TF-IDF (*Term Frequency Invers Document Frequency*) dengan memanfaatkan modul *TfidfVectorizer* pada *library scikit-learn* di Python. Tabel 3. menunjukkan hasil dari pengaplikasian TF-IDF pada beberapa kata.

Tabel 3. Tahap TF-IDF

	novel	**alur**	**cerita**	**bagus**	**bosan**	**imajinasi**	**...**	**baca**
Text1	0.63	0.13	0.17	0.12	0.0	0.15	...	0.07
Text2	0.0	0.0	0.0	0.0	0.0	0.0	...	0.0
Text3	0.63	0.22	0.0	0.0	0.26	0.0	...	0.0
Text4	0.0	0.0	0.58	0.80	0.0	0.0	...	0.34
....	...	...	...	...	...	...	...	...
Text1423	0.0	0.0	0.20	0.0	0.0	0.0	...	0.40

D. Analisis Teks Naïve Bayes

Sebelum dilakukan klasifikasi Naive Bayes, data terlebih dahulu dibagi dengan rasio pembagian 80:20, dimana 80% digunakan sebagai data latih dan 20% sebagai data uji secara acak menggunakan modul *train_test_split* dari *scikit-learn*. Seperti yang ditunjukkan pada Gambar 4. sebelumnya, data dengan label sentimen negatif terbilang sedikit dibandingkan dengan data label sentimen positif, maka pada data latih dilakukan pengambilan sampel ulang menggunakan teknik Random Oversampling untuk menangani ketidakseimbangan antar kelas. Random Oversampling bekerja dengan cara menambahkan data pada kelas minoritas secara acak. Melalui proses Random Oversampling tersebut, jumlah data kelas sentimen negatif disamaratakan dengan jumlah data

sentimen positif. Hasil dari pengaplikasian Random Oversampling dapat dilihat pada Tabel 4.

Tabel 4. Hasil Random Oversampling

Sentimen	Sebelum Oversampling	Setelah Oversampling
Positif (1)	973	973
Negatif (0)	165	973

Data pelatihan yang telah diproses melalui tahapan Random Oversampling kemudian digunakan untuk melatih model klasifikasi Naive Bayes dengan menerapkan modul *MultinomialNB* pada *library scikit-learn*. Selanjutnya data uji dapat digunakan untuk memprediksi sentimen berdasarkan data yang telah dilatih. Setelah melalui tahap prediksi maka diperoleh hasil prediksi sentimen. Data uji sebanyak 285 data diklasifikasikan menjadi sentimen positif dan negatif. Hasil klasifikasi ditunjukkan pada Tabel 5.

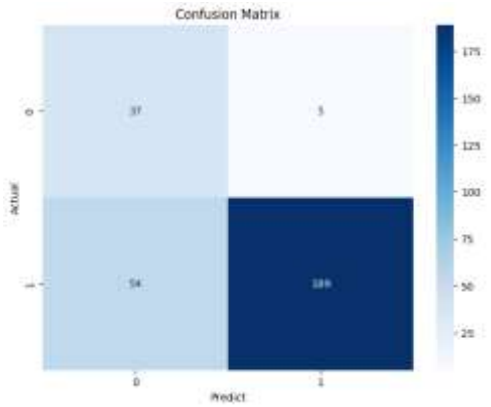
Tabel 5. Hasil Prediksi Sentimen Naïve Bayes

Sentimen	Jumlah Prediksi	Persentase
Positif (1)	243	85.26%
Negatif (0)	42	14.73%

Melalui Tabel 9. tersebut didapatkan hasil bahwa opini pembaca terhadap novel berjudul ”Bumi” karya Tere Liye pada *website* Goodreads cenderung positif sebanyak 85% sedangkan opini dengan sentimen negatif sebanyak 15%.

E. Evaluation

Mengevaluasi suatu model adalah dengan cara menghitung perbandingan antara nilai aktual dengan nilai prediksi. Penelitian ini menerapkan Confusion matrix untuk menilai seberapa baik model pengklasifikasi Naive Bayes yang telah dilakukan dalam memprediksi suatu data. Hasil yang didapatkan dari klasifikasi menggunakan Naive Bayes dapat dilihat pada Gambar 5.



Gambar 5. Hasil Confusion Matrix

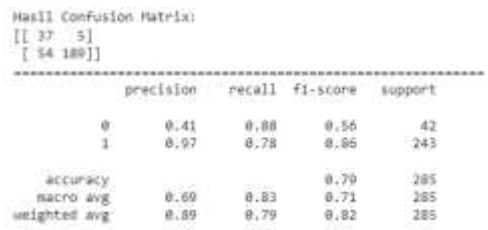
Melalui gambar tersebut, nilai *True Negative, False Negative, True Positive* dan *False Positive* dapat dijelaskan sebagai berikut:

1) *True Negative* (TN), data sentimen negatif yang tepat diprediksi: 37 data
2) *False Negative* (FN), data sentimen negatif yang tidak tepat diprediksi: 54 data
3) *True Positive* (TP), data sentimen positif yang tepat diprediksi: 189 data
4) *False Positive* (FP), data sentimen positif yang tidak tepat diprediksi: 5 data

Performa klasifikasi Naive Bayes dapat dihitung berdasarkan nilai dari visualisasi confusion matrix tersebut dengan menggunakan rumus yang telah dijelaskan sebelumnya. Tabel 6. menunjukkan hasil evaluasi keseluruhan data dan Gambar 6. menampilkan hasil perhitungan evaluasi pada setiap kelas.

Tabel 6. Evaluasi Naïve Bayes

Accuracy	79%
Precision	97%
Recall	77%
F1-Score	86%



Gambar 6. Evaluasi Naïve Bayes Setiap Kelas

Tabel 10. menunjukkan nilai akurasi Naive Bayes dalam memberikan prediksi sentimen mengenai novel ”Bumi” sebesar 79%, nilai presisi 97%, nilai *recall* 77% dan *f1-score* 86%.

Gambar 6. menunjukkan hasil dari klasifikasi menggunakan Naive Bayes pada opini pembaca novel ”Bumi” pada setiap kelas, dimana nilai presisi negatif sebesar 41%, presisi positif 97%, *recall* negatif 88%, *recall* positif 78%, *f1-score* negatif 56%, dan *f1-score* positif 86%.

F. Visualisasi Data

Untuk mendapatkan gambaran visual dari frekuensi kemunculan kata pada opini novel ’Bumi’, visualisasi teks ke dalam gambar dilakukan dengan menggunakan WordCloud. Gambar 7. dan Gambar 8. menunjukkan visualisasi *wordcloud* dari teks yang memiliki sentimen positif dan negatif. Semakin sering kata tersebut muncul, maka ukuran kata tersebut pada *wordcloud* juga akan semakin besar.

Gambar 7. Visualisasi Wordcloud Positif

Gambar 7. menunjukkan kata yang sering muncul dengan sentimen positf yaitu kata ”buku”, ”baca”, ”novel”, ”cerita”, ”bagus”, ”seru”, ”fantasi”, dan lain-lain.

Gambar 8. Visualisasi Wordcloud Negatif

Ganbar 8. menunjukkan kata yang sering muncul dengan sentimen negatif yaitu kata ”buku”, ”cerita”, ”baca”, ”kurang”, ”alur”, ”lam bat”, ”biasa”, dan lain-lain.

IV. KESIMPULAN

Berdasarkan penelitian analisis sentimen pada *website* Goodreads terhadap opini pembaca novel “Bumi” dengan data sebanyak 1486 data saat dilakukan proses *crawling* data. Tetapi setelah melalui tahap *preprocessing* data berkurang menjadi 1423 data, dimana data yang mempunyai sentimen positif sebanyak 1216 dan sentimen negatif sebanyak 207. Klasifikasi Naïve Bayes dilakukan dengan pembagian data 80:20 serta jumlah data latih kelas sentimen negatif disamaratakan dengan jumlah data latih sentimen positif dengan melakukan *resampling* menggunakan pendekatan Random Oversampling. Hasil klasifikasi menunjukkan bahwa opini pembaca novel “Bumi” cenderung positif dengan persentase sebesar 85.26% opini positif dan opini negatif sebesar 14.73%. Performa Naïve Bayes dalam melakukan klasifikasi sentimen dengan menggabungkan teknik pengambilan sampel ulang menggunakan Random Oversampling menghasilkan akurasi yang cukup baik saat dilakukan uji confusion matrix yakni sebesar 79% dengan nilai presisi sebesar 97%, nilai *recall* 77%, dan *f1-score* 86%. Selain itu pada visualisasi menggunakan *wordcloud* juga menunjukkan frekuensi kata dari data opini positif didominasi dengan kata-kata “buku”, “baca”, “novel”, “cerita”, “bagus”, “seru”, “fantasi”, dan lain sebagainya. Sedangkan frekuensi kemunculan kata dari data opini negatif didominasi dengan kata-kata “buku”, “cerita”, “baca”, “kurang”, “alur”, “lambat”, “biasa”, dan lain-lain. Dapat disimpulkan bahwa novel “Bumi” bertema fantasi dan memiliki alur yang lambat bagi sebagian orang. Selain itu pembaca lain dapat mempertimbangkan untuk membaca novel tersebut dari beberapa kata yang ada dalam visualisasi *wordcloud* sebelumnya. Pada penelitian selanjutnya dapat menggabungkan metode klasifikasi lainnya seperti KNN, SVM, Random Forest dan lain sebagainya dengan teknik resampling lain seperti SMOTE atau Random Undersampling.

REFERENSI

[1] S. Bhatia, M. Sharma, and K. K. Bhatia, “Sentiment Analysis and Mining of Opinions,” *Stud. Big Data*, vol. 30, no. May, pp. 503–523, 2018, doi: 10.1007/978-3-319-60435-0_20.

[2] D. Ikasari and Widiastuti, “Sentiment Analysis Review Novel ‘Goodreads’ Berbahasa Indonesia Menggunakan Naive Bayes Classifier,” *Semin. Nas. Ris. dan Inov. Teknol. (SEMNAS RISTEK) 2021*, vol. 5, pp. 760–765, 2021, [Online]. Available: http://www.proceeding.unindra.ac.id/index.php/semnasristek/article/view/5040

[3] N. P. N. Anggraini, Y. A. Restuningtyas, and W. G. A. S. Jodi, “Pengaruh Citra Merk, Desain Produk, Dan Online Customer Review Pada Keputusan Pembelian Online,” *IDEI J. Ekon. Bisnis*, vol. 4, no. 2, pp. 87–96, 2023, doi: 10.38076/ideijeb.v4i2.196.

[4] M. Walsh and M. Antoniak, “The goodreads ‘classics’: A computational study of readers, amazon, and crowdsourced amateur criticism,” *J. Cult. Anal.*, vol. 6, no. 2, pp. 243–287, 2021, doi: 10.22148/001C.22221.

[5] M. Thelwall and K. Kousha, “Goodreads: A social network site for book readers,” *J. Assoc. Inf. Sci. Technol.*, vol. 68, no. 4, pp. 972–983, 2017, doi: 10.1002/asi.23733.

[6] M. Abbas, K. Ali, A. Jamali, K. Ali Memon, and A. Aleem Jamali, “Multinomial Naive Bayes Classification Model for Sentiment Analysis,” *IJCSNS Int. J. Comput. Sci. Netw. Secur.*, vol. 19, no. 3, p. 62, 2019, doi: 10.13140/RG.2.2.30021.40169.

[7] Y. Nurtikasari, S. Alam, T. I. Hermanto, T. Informatika, S. Tinggi, and T. Wastukancana, “Analisis Sentimen Opini Masyarakat Terhadap Film Pada Platform Twitter Menggunakan

Algoritma Naive Bayes," vol. 1, no. 4, pp. 411–423, 2022, doi: 10.55123/insologi.v1i4.770.

[8] T. S. Rambe, M. Nirmala, S. Hasibuan, and M. H. Dar, "Sentiment Analysis of Beauty Product Applications using the Naïve Bayes Method," vol. 8, no. 2, pp. 980–989, 2023.

[9] P. A. Perwira and N. I. Widiastuti, "Imbalance Dataset in Aspect-Based Sentiment Analysis on Game Genshin Impact Review," *J. Infotel*, vol. 16, no. 1, pp. 71–81, 2024, doi: 10.20895/infotel.v16i1.984.

[10] N. Agustina, D. H. Citra, W. Purnama, C. Nisa, and A. R. Kurnia, "Implementasi Algoritma Naive Bayes untuk Analisis Sentimen Ulasan Shopee pada Google Play Store," *MALCOM Indones. J. Mach. Learn. Comput. Sci.*, vol. 2, no. 1, pp. 47–54, 2022, doi: 10.57152/malcom.v2i1.195.

[11] R. Ihsan, P. Selian, A. V. Vitianingsih, S. Kacung, and A. Lidya, "Sentiment Analysis of Public Responses on Social Media to Satire Joke Using Naive Bayes and KNN," vol. 8, no. 3, pp. 1443–1451, 2024.

[12] NURUL AFIFAH, Dony Permana, Dodi Vionanda, and Dina Fitria, "Sentiment Analysis of Electric Cars Using Naive Bayes Classifier Method," *UNP J. Stat. Data Sci.*, vol. 1, no. 4, pp. 289–296, 2023, doi: 10.24036/ujsds/vol1-iss4/68.

[13] L. E. Pradana and Y. Ruldeviyani, "Sentiment Analysis of Nanovest Investment Application Using Naive Bayes Algorithm," *J. Nas. Pendidik. Tek. Inform.*, vol. 12, no. 2, pp. 283–293, 2023, doi: 10.23887/janapati.v12i2.62302.

[14] T. Aksoy, S. Gülseçen, and S. Çelik, "Data Pre-processing in Text Mining," *Who Runs World Data*, no. December, pp. 123–144, 2020, doi: 10.26650/b/et06.2020.011.07.

[15] F. Lan, "Research on Text Similarity Measurement Hybrid Algorithm with Term Semantic Information and TF-IDF Method," *Adv. Multimed.*, vol. 2022, 2022, doi: 10.1155/2022/7923262.

[16] M. Dauda Abubakar, Haisal dan Umar, "Sentiment Classification: Review of Text Vectorization Methods: Bag of Words, Tf-Idf, Word2vec and Doc2vec," *SLU J. Sci. Technol.*, vol. 4, pp. 27–33, 2022, doi: 10.56471/slujst.v4i.266.

[17] A. Hanani, "Term Frequency Inverse Document Frequency (TF-IDF)," *Visit. Anal.*, no. December, 2023, [Online]. Available: https://www.visitor-analytics.io/es/glosario/t/term-frequency-inverse-document-frequency-tf-idf

[18] M. Hayaty, S. Muthmainah, and S. M. Ghufran, "Random and Synthetic Over-Sampling Approach to Resolve Data Imbalance in Classification," *Int. J. Artif. Intell. Res.*, vol. 4, no. 2, p. 86, 2021, doi: 10.29099/ijair.v4i2.152.

[19] M. M. Saritas and A. Yasar, "Performance Analysis of ANN and Naive Bayes Classification Algorithm for Data Classification," *Int. J. Intell. Syst. Appl. 7*, vol. 7, no. January 2019, 2019.

[20] Z. Dong, X. Guo, S. Rajana, and B. Chen, "Understanding 21st century bordeaux wines from wine reviews using naïve bayes classifier," *Beverages*, vol. 6, no. 1, pp. 1–16, 2020, doi: 10.3390/beverages6010005.