



The Implementation of Naïve Bayes Algorithm for Sentiment Analysis of Shopee Reviews on Google Play Store

Implementasi Algoritma Naive Bayes untuk Analisis Sentimen Ulasan Shopee pada Google Play Store

**Nurhaliza Agustina. C.A^{1*}, Desy Herlina Citra², Wido Purnama³,
Chairun Nisa⁴, Amanda Rozi Kurnia⁵**

^{1,2,3}Information System Department, Faculty of Science and Technology,
Universitas Islam Negeri Sultan Syarif Kasim Riau, Indonesia

⁴Informatics Engineering Department, Faculty of Engineering,
Universitas Riau, Indonesia

⁵Informatics Engineering Department, Faculty of Intelligence Electrical and Informatics Technology,
Institut Teknologi Sepuluh Nopember, Indonesia

E-Mail: ¹11950320632@students.uin-suska.ac.id,

²11950324625@students.uin-suska.ac.id, ³11950311588@students.uin-suska.ac.id,

⁴chairun.nisa4638@students.unri.ac.id, ⁵amanda.19051@mhs.its.ac.id

Received Apr 14th 2022; Revised Apr 19th 2022; Accepted Apr 30th 2022
Corresponding Author: Nurhaliza Agustina

Abstract

The rapid development of technology provides convenience for people in various things in life. Technology that grows continuously certainly produces a very large amount of data that can provide useful and useful information. Google has a service known as the Play Store which has a feature where users can provide reviews of the products they use. In this study, sentiment analysis was carried out on the Shopee application review on the Google Play Store using the Naïve Bayes algorithm using different data distribution techniques. The result showed that the distribution of Hold-Out data with a ratio 80:20 resulted in algorithm accuracy of 83%, which was 1% superior to the average accuracy result using the 10-Cross Fold Validation data sharing technique, which was 82%.

Keywords: Naïve Bayes, Play Store, Reviews, Sentiment Analysis, Shopee

Abstrak

Perkembangan teknologi yang semakin pesat memberikan kemudahan bagi masyarakat dalam berbagai hal dalam kehidupan. Teknologi yang tumbuh secara terus-menerus tentu menghasilkan data dalam jumlah yang sangat besar yang dapat memberikan informasi yang berguna dan bermanfaat. Google memiliki sebuah layanan yang dikenal dengan nama Play Store yang terdapat fitur dimana user dapat memberikan ulasan terhadap produk yang mereka gunakan. Pada penelitian ini melakukan analisis sentimen pada ulasan aplikasi Shopee pada Google Play Store dengan menggunakan algoritma Naïve Bayes dengan menggunakan teknik pembagian data yang berbeda. Hasil penelitian diketahui bahwa pembagian data Hold-Out dengan perbandingan 80:20 menghasilkan akurasi algoritma sebesar 83%, yang lebih unggul 1% dibandingkan dengan hasil akurasi rata-rata dengan menggunakan teknik pembagian data 10-Cross Fold Validation yaitu sebesar 82%.

Kata Kunci: Analisis Sentimen, Naïve Bayes, Play Store, Ulasan, Shopee

1. PENDAHULUAN

Perkembangan teknologi yang semakin pesat dapat memberikan kemudahan bagi masyarakat dalam berbagai hal dalam kehidupan. Teknologi yang tumbuh secara terus-menerus tentu menghasilkan data dalam jumlah yang sangat besar yang mana data tersebut dapat memberikan informasi yang berguna dan bermanfaat apabila dapat diproses dan digunakan [1]. Banyaknya pengguna aktif internet pada saat ini membuat jumlah data yang dapat dihasilkan juga sangat besar (big data). Dengan ada teknologi big data dapat membantu dalam

pengolahan data yang besar, banyak, dan rumit sehingga data-data yang telah diproses tersebut dapat memberikan informasi yang berguna [2].

Google memiliki sebuah layanan yang dikenal dengan nama Play Store yang menyediakan konten-konten digital seperti game, aplikasi, film, musik, dan buku dengan kategori yang beragam. Salah satu fitur yang terdapat pada Play Store adalah fitur rating dan ulasan dimana user produk dari Play Store dapat memberikan opini mereka terhadap produk yang telah mereka gunakan [3]. Salah satu aplikasi e-commerce yang terdapat pada Play Store adalah Shopee. Shopee merupakan sebuah aplikasi yang digunakan untuk melakukan proses jual beli secara online yang dapat digunakan melalui smartphone. Shopee merupakan salah satu aplikasi marketplace yang populer yang mana pada tahun 2020 Shopee merupakan aplikasi e-commerce yang paling banyak di klik di Indonesia [4].

Analisis sentimen merupakan proses yang dilakukan untuk menyediakan informasi yang terdapat pada dataset tidak terstruktur. Proses ini merupakan proses komputasi yang dilakukan dengan cara memahami, mengekstraksi, dan mengolah data yang berbentuk tekstual secara otomatis sehingga mendapatkan informasi yang terdapat dalam opini atau perilaku seseorang [3]. Analisis sentimen berguna untuk mengetahui baik atau tidaknya tanggapan pengguna terhadap suatu produk dengan cara mengekstrak teks dari suatu ulasan untuk mengetahui emosi pengguna [5].

Penelitian yang berjudul “Analisis Sentimen Aplikasi Duolingo Menggunakan Metode Naïve Bayes dan Synthetic Minority Over Sampling Technique” yang menggunakan RapidMiner untuk mencari dan membandingkan dua metode klasifikasi yang berbeda antara dataset yang hanya menggunakan Naïve Bayes Classifier dengan dataset yang menggunakan Naïve Bayes Classifier dan Syntetic Minority Over-Sampling Technique (SMOTE) menghasilkan akurasi tertinggi untuk dataset yang menggunakan Naïve Bayes dan SMOTE dengan akurasi 91,95% dan AVC 0,740 [6]. Selain itu penelitian yang pernah dilakukan untuk menganalisis sentimen pada aplikasi Bibit dan Bareksa yang menggunakan model CRISP-DM dan algoritma KNN dengan perbandingan data training dan data testing 60:40 menghasilkan akurasi 85,14%, precision 91,91%, dan recall 76,44% untuk aplikasi Bibit. Sementara itu, untuk aplikasi Bareksa memiliki akurasi 81,70%, precision 87,15%, dan recall 75,73% [7]. Penelitian tentang analisis sentimen pada aplikasi Zoom Meeting yang membandingkan algoritma Naïve Bayes dan Support Vector Machine dengan menggunakan evaluasi 10-fold cross validation menghasilkan nilai akurasi untuk Naïve Bayes sebesar 74,37% dan nilai AUC 0,659. Sementara itu, untuk Support Vector Machine memiliki akurasi yang lebih unggul daripada Naïve Bayes yaitu sebesar 81,22% dan nilai AUC sebesar 6,85% [3]. Penelitian yang dilakukan dengan cara membandingkan algoritma Naïve Bayes dan Decision Tree untuk analisis sentimen publik terhadap vaksinasi COVID-19 di Indonesia menunjukkan bahwa opini publik cenderung kepada opini negatif dan algoritma yang baik dalam penelitian tersebut adalah algoritma Naïve Bayes dengan akurasi sebesar 100.00%, sedangkan algoritma Decision Tree menghasilkan akurasi sebesar 50.39% [8].

Penelitian ini akan dilakukan dengan menggunakan algoritma Multinomial Naïve Bayes dengan tools Google Colaboratory untuk melakukan analisis sentimen pada salah satu aplikasi e-commerce, yaitu Shopee. Penelitian ini diharapkan dapat membantu memberikan informasi tentang sentimen yang terdapat pada ulasan yang telah diberikan oleh pelanggan terhadap aplikasi tersebut.

2. BAHAN DAN METODE

Metode Penelitian yang secara umum dapat ditunjukkan pada gambar 1.

2.1 *Web Scrapping Data*

Web scrapping data merupakan proses yang dilakukan untuk mengumpulkan data-data ulasan dari Google Play Store dengan menggunakan bahasa pemrograman Python. Hasil dari web scrapping ini didapatkan 1000 record data ulasan pada aplikasi Shopee di Google Play Store.

2.2 *Text Preprocessing*

Text preprocessing merupakan proses yang dilakukan untuk mempersiapkan data sebelum dilakukan pemodelan [6]. *Preprocessing* merupakan salah satu Teknik dalam *data mining* untuk melakukan transformasi pada data mentah agar menjadi format yang bisa lebih mudah dimengerti. Tahapannya dilakukan untuk menyelesaikan masalah-masalah, seperti redundansi data, *noisy data*, data yang hilang [2]. Langkah-langkah dalam *text preprocessing* adalah sebagai berikut.

a. *Case Folding*

Case folding merupakan suatu proses yang dilakukan untuk merubah semua huruf menjadi huruf kecil (*lower case*) [2].

b. *Tokenization*

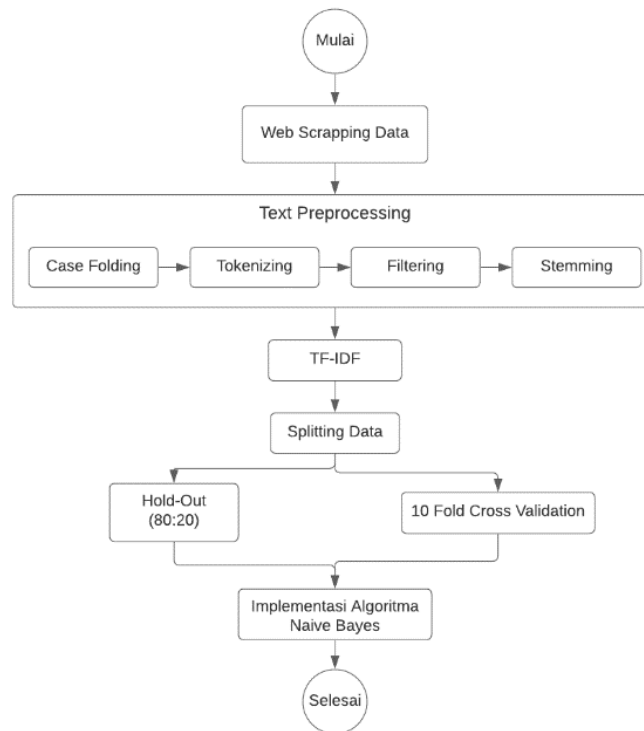
Proses *tokenization* merupakan proses yang dilakukan untuk memecah teks menjadi potongan kata yang lebih kecil yang dipisahkan berdasarkan karakter spasi dan juga dapat dilakukan penghapusan tanda baca [9] [2].

c. *Filtering*

Filtering merupakan suatu proses yang dilakukan untuk menghapus data yang kurang lengkap, data yang memiliki kesalahan [10]. Selain itu, *filtering* juga melakukan proses *stopword removal* yaitu menghapus kata yang tidak relevan atau tidak memiliki arti [2][11].

d. *Stemming*

Stemming merupakan proses yang dilakukan untuk menemukan kata dasar dan juga menghilangkan semua imbuhan yang terdapat pada suatu kata tersebut [10]. *Stemming* dilakukan bertujuan untuk menghilangkan kata yang memiliki pengejaan yang kurang tepat. Setiap Bahasa memiliki algoritma stemming yang berbeda-beda [3]. *Stemming* juga dilakukan untuk membuat data menjadi lebih seragam dan mengurangi daftar kata yang terdapat pada data latih [2].



Gambar 1. Metodologi Penelitian

2.3 TF-IDF

Setelah dataset melalui proses *text pre-processing*, selanjutnya adalah melakukan pembobotan pada teks tersebut untuk mendapatkan nilai agar bisa diklasifikasikan [10]. Pembobotan kata dilakukan untuk memberikan bobot pada setiap kata yang terkandung di dalam sebuah dokumen dengan menggunakan salah satu metode pembobotan kata yang paling populer, yaitu metode TF-IDF (*Term Frequency – Inversed Document Frequency*). *Term-Frequency* memiliki fokus pada seberapa seringnya suatu istilah yang muncul dalam suatu dokumen. Sementara itu, *Inversed Document Frequency* memiliki fokus pada memberikan bobot terendah terhadap istilah yang muncul dalam banyak dokumen [2].

2.4 Splitting Data

Pada penelitian ini, pembagian data dilakukan dengan dua teknik pembagian yang berbeda, yaitu sebagai berikut.

a. *Hold-Out*

Salah satu metode yang dilakukan untuk melakukan pemecahan data dengan membagi data menjadi data latih dan data uji disebut dengan metode *Hold-Out*. Pada penerapan metode ini, satu bagian data digunakan sebagai data latih dan sebagian lainnya digunakan sebagai data uji [12].

b. *K-Fold Cross Validation*

Cross Validation merupakan Teknik yang dilakukan untuk dapat menilai atau dapat memvalidasi tingkat keakuratan dari sebuah model berdasarkan dataset tertentu yang mana model tersebut digunakan untuk melakukan prediksi atau klasifikasi [7]. Dalam metode K-Fold Cross Validation, data dibagi menjadi K bagian setelah dilakukan proses pembobotan *term* sebelumnya. Percobaan akan dilakukan sebanyak K kali dengan menggunakan satu bagian sebagai data uji [10].

2.5 Multinomial Naïve Bayes

Multinomial Naïve Bayes merupakan metode klasifikasi yang difokuskan untuk melakukan klasifikasi pada data tekstual. Multinomial Naïve Bayes memiliki fitur dimana hasil yang didapat dari masing-masing kelas bersifat independen karena dokumen yang satu dengan dokumen yang berikutnya tidak memiliki keterkaitan sehingga mendapatkan hasil yang murni hanya diolah dari dokumen itu sendiri. Persamaannya dapat dilihat sebagai berikut [9].

$$P(c|d) \propto P(c) \prod_{i=1}^{n_d} P(w_i|c) \quad (3)$$

Keterangan:

- $P(c|d)$ = Probabilitas suatu kelas c pada dokumen/teks d
 $P(c)$ = Probabilitas prior c
 $P(w_i|c)$ = Probabilitas suatu kata pada kelas c

3. HASIL DAN ANALISIS

3.1 Web Scrapping

Web Scrapping yang dilakukan dengan tools Google Colaboratory dan bahasa pemrograman Python pada link Shopee di Google Play Store didapatkan sebanyak 1000 data ulasan terbaru pada tahun 2021. Proses ini dilakukan dengan melakukan instalasi *library google-play-scraper* yang digunakan untuk dapat melakukan *scrapping* data-data ulasan pada *Google Play Store* dengan hanya memasukkan id dari aplikasi yang ingin diambil ulasannya yang terdapat pada link aplikasi tersebut. Data hasil *scrapping* lalu diekspor dalam format excel (.xlsx). Contoh dokumen yang dihasilkan dari *web scrapping* adalah sebagai berikut.

Tabel 1. Contoh Dokumen Hasil *Scrapping*

Dokumen	Ulasan
Dokumen 1	Semenjak ada shopee belanja lancar dan hemat
Dokumen 2	Tolong dioptimalkan lagi saat membuka halamannya
Dokumen 3	Ga nyesel deh belanja di Shopee
...	...
Dokumen x	Keren sekali pelayanannya dan gak pake lama

3.2. Text Preprocessing

Preprocessing merupakan tahapan penting yang harus dilakukan sebelum dapat mengimplementasikan algoritma klasifikasi pada dokumen. Tahapan dalam preprocessing adalah sebagai berikut.

a. Case Folding

Proses ini dilakukan untuk mentransformasikan semua huruf menjadi huruf kecil guna menyeragamkan hasil teks. Dalam proses ini juga dilakukan pembersihan kata-kata dari tanda baca atau karakter-karakter tertentu yang tidak dibutuhkan. Contoh hasil dari penerapan *case folding* dapat dilihat pada tabel 2.

Tabel 2. Proses *Case Folding*

semenjak ada shopee belanja lancar dan hemat
--

b. Tokenizing

Proses ini dilakukan untuk memecah atau memotong suatu kalimat menjadi term-term yang dipisahkan berdasarkan spasi. Contoh hasil proses *tokenizing* dapat dilihat pada tabel 3.

Tabel 3. Proses *Tokenizing*

['semenjak', 'ada', 'shopee', 'belanja', 'lancar', 'dan', 'hemat']
--

c. Filtering

Pada proses *filtering* ini dilakukan penghapusan *stopwords* yang tidak memiliki kaitan dengan analisis sentiment. Hasil dari proses *filtering* ini dapat dilihat pada tabel 4.

Tabel 4. Proses *Filtering*

['semenjak', 'shopee', 'belanja', 'lancar', 'hemat']
--

d. *Stemming*

Stemming merupakan proses yang dilakukan untuk menghilangkan kata imbuhan, kata depan, kata ganti, dan kata belakang yang sesuai dengan KBBI. Pada penelitian ini, proses *Stemming* dilakukan dengan menggunakan *library* Sastrawi yang dapat digunakan untuk melakukan *stemming* pada teks berbahasa Indonesia. Hasil dari proses *stemming* dapat dilihat pada tabel berikut.

Tabel 5. Proses *Stemming*

['semenjak', 'shopee', 'belanja', 'lancar', 'hemat']
--

3.3. TF-IDF

Proses pembobotan yang dilakukan pada penelitian ini menggunakan metode TF-IDF yang mana nantinya dapat menghasilkan sebuah *vector* yang memiliki banyak *vector* sehingga dapat dikenali tiap-tiap katanya.

a. *Term Frequency (TF)*

Proses ini dilakukan untuk menghitung banyaknya kata yang muncul dalam sebuah dataset. Menghitung bobot TF dapat dilakukan dengan menggunakan persamaan berikut.

$$tf = 0.5 + 0.5 \frac{tf}{\max(tf)} \quad (4)$$

Keterangan:

tf = banyak data yang muncul dalam sebuah dokumen

max(tf) = panjang kata dari sebuah dokumen

Sebagai contoh melakukan perhitungan nilai TF berdasarkan dokumen-1 yaitu:

['semenjak', 'shopee', 'belanja', 'lancar', 'hemat'].

Jika dilakukan perhitungan dengan menggunakan persamaan di atas maka didapatkan hasil sebagai berikut.

$$semenjak = 0.5 + 0.5 \frac{1}{5} = 0.2$$

Karena setiap kata yang muncul dalam dokumen tersebut berjumlah satu kali, maka setiap kata tersebut memiliki nilai bobot TF yang sama yaitu sebagai berikut:

['semenjak'=0.2, 'shopee'=0.2, 'belanja'=0.2, 'lancar'=0.2, 'hemat'=0.2]

b. *Inverse Document Frequency (IDF)*

Proses ini dilakukan untuk mengetahui jumlah dokumen yang mengandung *term* yang dicari yang dapat dihitung dengan menggunakan persamaan sebagai berikut.

$$idf = \ln \frac{N}{df} + 1 \quad (5)$$

Keterangan:

ln = logaritma natural

N = jumlah semua dokumen

df = jumlah kata pada dokumen

Contoh penerapan persamaan IDF untuk dokumen-1 ['semenjak', 'shopee', 'belanja', 'lancar', 'hemat']. Melalui proses pengkodean, didapatkan banyaknya kata-kata tersebut yang muncul dalam satu dataset seperti pada tabel 6.

Tabel 6. Frekuensi Kata yang Muncul

Kata	Frekuensi
semenjak	1
shopee	321
belanja	150
lancar	7
hemat	6

Jika hasil dari dokumen tersebut dimasukkan ke dalam persamaan (5), maka hasilnya adalah sebagai berikut.

$$\text{semenjak} = \ln \frac{969}{1} + 1 = 7.876265$$

$$\text{shopee} = \ln \frac{969}{321} + 1 = 2.104594$$

$$\text{belanja} = \ln \frac{969}{150} + 1 = 2.865629$$

$$\text{lancar} = \ln \frac{969}{7} + 1 = 5.901029$$

$$\text{hemat} = \ln \frac{969}{6} + 1 = 6.084505$$

Sehingga nilai bobot IDF untuk setiap *term* pada dokumen-1 dapat dilihat sebagai berikut.

[‘semenjak’=7.876265 ‘shopee’=2.104594, ‘belanja’=2.865629, ‘lancar’=5.901029, ‘hemat’=6.084505]

Selanjutnya proses perhitungan nilai bobot TF-IDF dapat dilakukan dengan cara mengalikan nilai TF dengan IDF sehingga mendapatkan hasil seperti pada tabel 7.

Tabel 7. Hasil TF-IDF

Dokumen	TF IDF	Label
[‘semenjak’, ‘shopee’, ‘belanja’, ‘lancar’, ‘hemat’]	[1.57525, 0.420919, 0.573126, 1.180206, 1.216901]	Positif
[‘tolong’, ‘optimal’, ‘buka’, ‘halaman’]	[3.050333, 4.922666, 3.319501, 4.4889448]	Positif

3.4. *Splitting Data dan Implementasi Algoritma Naïve Bayes*

Proses pembagian data pada penelitian ini dibagi menjadi dua bagian untuk melakukan percobaan klasifikasi, yaitu menggunakan *hold-out* dan *cross validation*.

a. Hold-Out

Proses pembagian *data training* dan *data testing* yang dilakukan dengan menggunakan metode *Hold-Out* membagi *data training* sebesar 80% dan *data testing* sebesar 20%. Proses pembagian data ini dilakukan dengan menggunakan bahasa pemrograman Python dengan program seperti berikut.

Pembagian data

```
from sklearn.model_selection import train_test_split
X_train, X_test, y_train, y_test = train_test_split(text_tf, data_clean['kategori'], test_size=0.2, random_state = 42)
```

Pengujian menggunakan *confusion matrix* dilakukan untuk menguji model yang diimplementasikan pada data training dan data testing sehingga menghasilkan matriks berukuran 3x3. Tabel *confusion matrix* dapat dilihat pada tabel 8.

Tabel 8. *Confusion Matrix*

Data Prediksi	Data Aktual	
	Positif	Negatif
Positif	TP	FP
Negatif	FN	TN

Keterangan:

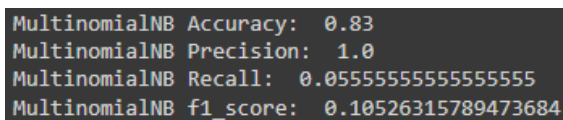
TP = memprediksi kata positif dan itu benar
 TN = memprediksi kata negatif dan itu benar
 FP = memprediksi kata positif dan itu salah
 FN = memprediksi kata itu negatif dan itu salah

Berdasarkan hasil pengujian menggunakan *confusion matrix* didapatkan hasil pada tabel 9.

Tabel 9. Hasil *Confusion Matrix*

Data Prediksi	Data Aktual	
	Positif	Negatif
Positif	2	34
Negatif	0	164

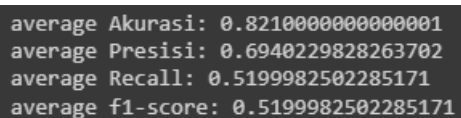
Hasil performa algoritma Naïve Bayes dengan menggunakan teknik pembagian data *Hold-Out* menghasilkan akurasi sebesar 83% serta nilai *precision*, *recall*, dan *f1-score* yang dapat dilihat pada gambar 2.



Gambar 2. Hasil Akurasi, *Precision*, *Recall*, dan *f1-score* Metode *Hold-Out*

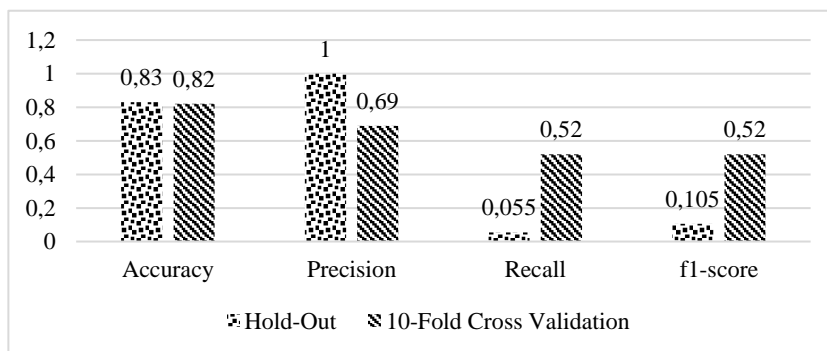
b. *K-Fold Cross Validation*

Pembagian data dengan menggunakan metode *K-fold cross validation* dimana data akan diproses sebanyak K kali. Dalam penelitian ini nilai K yang digunakan adalah $K = 10$. Hasil *precision*, *recall*, *f1-score*, dan akurasi dari algoritma Naïve Bayes dengan metode pembagian data menggunakan *10-fold cross validation* dapat dilihat pada gambar 3.



Gambar 3. Hasil Akurasi, *Precision*, *Recall*, dan *f1-score* Metode *10-fold cross validation*

Perbandingan nilai akurasi, *precision*, *recall*, dan *f1-score* dari kedua metode pembagian data tersebut dapat dilihat pada gambar 4.



Gambar 4. Perbandingan Akurasi, *Precision*, *Recall*, dan *f1-score*

Berdasarkan gambar tersebut dapat diketahui bahwa pembagian data dengan metode *Hold-Out* menghasilkan akurasi terbaik yaitu sebesar 83%. Sementara itu hasil akurasi yang didapatkan dengan teknik pembagian data *10-fold cross validation* adalah sebesar 82%. Kecenderungan sentimen pada ulasan aplikasi Shopee di Google Play Store dengan teknik pembagian data *Hold-Out* dapat dilihat pada gambar 5.

	precision	recall	f1-score	support
Negatif	1.00	0.06	0.11	36
Positif	0.83	1.00	0.91	164
accuracy			0.83	200
macro avg	0.91	0.53	0.51	200
weighted avg	0.86	0.83	0.76	200

Gambar 5. Hasil Sentimen

Berdasarkan gambar tersebut dapat diketahui bahwa sentimen pengguna aplikasi Shopee yang memberikan ulasan pada *Google Play Store* cenderung positif dengan nilai *precision* sebesar 83%, *recall* 100%, dan *f1-score* 91%.

4. KESIMPULAN

Berdasarkan hasil dari pembahasan disertai hasil analisis yang telah dipaparkan menghasilkan kesimpulan diantaranya adalah Proses pengumpulan data ulasan dengan melakukan teknik web scrapping pada *Google Play Store* didapatkan sebanyak 1000 data ulasan yang selanjutnya melalui proses preprocessing yang menghasilkan data bersih sebanyak 969 data. Performa algoritma Naïve Bayes dengan menggunakan teknik pembagian data Hold-Out menghasilkan akurasi yang lebih baik yaitu sebesar 83% dibandingkan dengan menggunakan teknik pembagian data 10-fold cross validation yang menghasilkan akurasi sebesar 82%. Berdasarkan hasil pengujian yang telah dilakukan dimana hasil akurasi dengan metode Hold-Out menghasilkan akurasi terbaik, dapat disimpulkan bahwa sentimen pengguna Shopee yang memberikan ulasan pada *Google Play Store* cenderung positif dengan nilai *precision* sebesar 83%, *recall* 100%, dan *f1-score* 91%.

REFERENSI

- [1] S. Wahyu Handani, D. Intan Surya Saputra, Hasirun, R. Mega Arino, and G. Fiza Asyroti Ramadhan, "Sentiment analysis for go-jek on google play store," *J. Phys. Conf. Ser.*, vol. 1196, no. 1, 2019, doi: 10.1088/1742-6596/1196/1/012032.
- [2] B. Gunawan, H. S. Pratiwi, and E. E. Pratama, "Sistem Analisis Sentimen pada Ulasan Produk Menggunakan Metode Naive Bayes," *J. Edukasi dan Penelit. Inform.*, vol. 4, no. 2, p. 113, 2018, doi: 10.26418/jp.v4i2.27526.
- [3] N. Herlinawati, Y. Yuliani, S. Faizah, W. Gata, and S. Samudi, "Analisis Sentimen Zoom Cloud Meetings di Play Store Menggunakan Naïve Bayes dan Support Vector Machine," *CESS (Journal Comput. Eng. Syst. Sci.)*, vol. 5, no. 2, p. 293, 2020, doi: 10.24114/cess.v5i2.18186.
- [4] D. Pratmanto, R. Rousyati, F. F. Wati, A. E. Widodo, S. Suleman, and R. Wijianto, "App Review Sentiment Analysis Shopee Application in Google Play Store Using Naive Bayes Algorithm," *J. Phys. Conf. Ser.*, vol. 1641, no. 1, 2020, doi: 10.1088/1742-6596/1641/1/012043.
- [5] A. Rahman, E. Utami, and S. Sudarmawan, "Sentimen Analisis Terhadap Aplikasi pada Google Playstore Menggunakan Algoritma Naïve Bayes dan Algoritma Genetika," *J. Komtika (Komputasi dan Inform.)*, vol. 5, no. 1, pp. 60–71, 2021, doi: 10.31603/komtika.v5i1.5188.
- [6] S. Chohan, A. Nugroho, A. M. B. Aji, and W. Gata, "Analisis Sentimen Pengguna Aplikasi Duolingo Menggunakan Metode Naïve Bayes dan Synthetic Minority Over Sampling Technique," *Paradig. - J. Komput. dan Inform.*, vol. 22, no. 2, pp. 139–144, 2020, doi: 10.31294/p.v22i2.8251.
- [7] A. D. Adhi Putra, "Analisis Sentimen pada Ulasan pengguna Aplikasi Bibit Dan Bareksa dengan Algoritma KNN," *JATISI (Jurnal Tek. Inform. dan Sist. Informasi)*, vol. 8, no. 2, pp. 636–646, 2021, doi: 10.35957/jatisi.v8i2.962.
- [8] A. Harun and D. P. Ananda, "Analysis of Public Opinion Sentiment About Covid-19 Vaccination in Indonesia Using Naïve Bayes and Decission Tree Analisa Sentimen Opini Publik Tentang Vaksinasi Covid-19 di Indonesia Menggunakan Naïve Bayes dan Decission Tree," *Indones. J. Mach. Learn. Comput. Sci.*, vol. 1, no. April, pp. 58–63, 2021.
- [9] D. P. Daryfayi Edyt and I. Asror, "Sentimen Analisis pada Ulasan Google Play Store Menggunakan Metode Naïve Bayes," *Sentimen Anal. pada Ulas. Google Play Store Menggunakan Metod. Naïve Bayes*, vol. 7, no. Ulasan Pada Google Play Store, p. 11, 2020.
- [10] F. F. Irfani, "Analisis Sentimen Review Aplikasi Ruangguru Menggunakan Algoritma Support Vector Machine," *JBMI (Jurnal Bisnis, Manajemen, dan Inform.)*, vol. 16, no. 3, pp. 258–266, 2020, doi: 10.26487/jbmi.v16i3.8607.
- [11] D. Musfiroh et al., "Analisis Sentimen terhadap Perkuliahan Daring di Indonesia dari Twitter Dataset Menggunakan InSet Lexicon," *MALCOM Indones. J. Mach. Learn. Comput. Sci.*, vol. 1, no. 1, pp. 24–33, 2021.
- [12] S. T. Rizaldi and M. Mustakim, "Perbandingan Teknik Pembagian Data untuk Klasifikasi Sarana Akses Air pada Algoritma K- Nearest Neighbor dan Naïve Bayes Classifier," *Semin. Nas. Teknol. Informasi, Komun. dan Ind.* 12, pp. 130–137, 2020.