

Analisis Klasifikasi Sentimen terhadap Penurunan COVID-19 Menggunakan Metode Naïve Bayes dan Support Vector Machine pada Platform X-App

Muhammad Fadhil¹, Iik Muhammad Malik Matin², Sifa Nurpadillah³

^{1,3}Fakultas Teknik Universitas Garut, Garut, 44151, Indonesia

²Departemen Teknik Informatika dan Komputer Politeknik Negeri Jakarta, Jl. Prof. DR. G.A. Siwabessy, Kukusan, Kecamatan Beji, Kota Depok, Jawa Barat, Indonesia

Korespondensi: mhmmdfadhil13@gmail.com

ARTICLE HISTORY

Received: 11-12-2023

Revised: 28-12-2023

Accepted: 28-12-2023

Abstrak

WHO atau biasa disebut *World Health Organization* menetapkan pandemi internasional untuk virus corona, atau Covid-19. Pada penelitian ini penulis ini bertujuan untuk menganalisis klasifikasi *sentiment* masyarakat terhadap penurunan Covid-19 menggunakan metode Algoritma *Naïve Bayes* dan *Support Vector Machine* pada platform X App. Pandemi Covid-19 telah menjadi perhatian seluruh dunia, dan media *social* seperti X App mejadi sumber data yang kaya untuk memahami pandangan dan perasaan masyarakat terkait penurunan kasus Covid-19. Peneliti mengumpulkan data dari X App sebanyak 5000 pada periode bulan Juni 2023, data yang belum melalui tahapan *preprocessing*. Dalam analisis penurunan kasus Covid-19, penelitian ini membandingkan Algoritma *Naïve Bayes* dan *Support Vector Machine*. Berdasarkan hasil penelitian ditemukan bahwasanya algoritma *Support Vector Machine* memiliki kinerja lebih unggul dari akurasi presisi serta recall yaitu 95%, 96%,96% sedangkan kinerja algoritma *Naïve Bayes* menghasilkan Akurasi, presisi serta *recall* yaitu 85%, 94%, 86%.

Kata kunci: Analisis Sentimen, Klasifikasi *Naive Bayes*, Klasifikasi *Support Vector Machine*, X App

Sentiment Classification Analysis of the Reduction of COVID-19 Using the Naïve Bayes Method and Support Vector Machine on the X App Platform

Abstract

The World Health Organization (WHO) declared a global pandemic for the corona virus, or Covid-19. In this research, the author aims to identify the classification of public sentiment regarding the decline in Covid-19 using the Naïve Bayes Algorithm and Support Vector Machine method on the X App platform. The Covid-19 pandemic has attracted worldwide attention, and social media such as X App has become a rich source of data to understand people's views and feelings regarding the decline in Covid-19 cases. Researchers collected 5000 data from X App that had not been processed through preprocessing. In analyzing the decline in Covid-19 cases, this research compares the Naïve Bayes algorithm and the Support Vector Machine. Based on the research results,

it was found that the Support Vector Machine algorithm has better performance in terms of accuracy, precision and recall, namely 95%, 96%, 96%, while the performance of the Naïve algorithm Bayes produces accuracy, precision and recall, namely 85%, 94%, 86%.

Key words: *Sentiment Analysis, Naive Bayes Classification, Support Vector Machine Classification, X App*

1. Pendahuluan

Pandemi COVID-19 telah menjadi salah satu peristiwa internasional yang paling signifikan dan mendalam di abad ke-21. Penyebaran virus ini telah mempengaruhi hampir setiap aspek kehidupan manusia, termasuk kesehatan, ekonomi, pendidikan, dan sosial. Di tengah ketidakpastian dan ketakutan yang ditimbulkan oleh pandemi ini, media sosial, khususnya platform X App, menjadi salah satu sumber informasi penting bagi masyarakat untuk mendapatkan berita terkini dan berbagi perspektif [1].

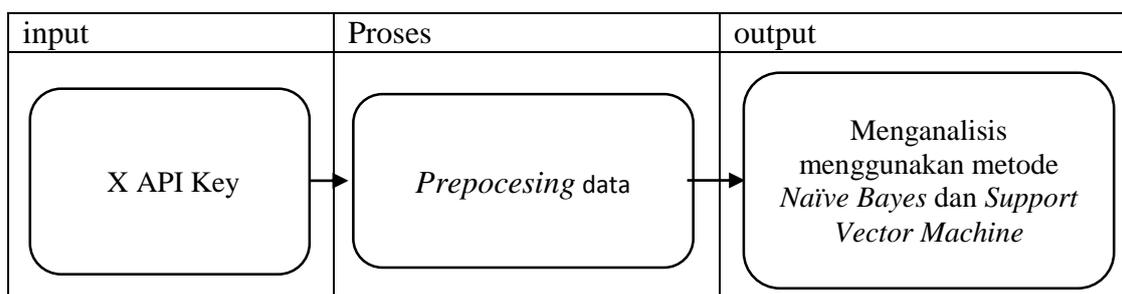
Dalam konteks pandemi COVID-19, masyarakat di penjuru bumi telah mengekspresikan perasaan dan pendapat mereka tentang berbagai aspek penurunan kasus COVID-19 melalui platform X App. *Tweet-tweet* ini mencakup respons terhadap kebijakan pemerintah, protokol kesehatan, perkembangan vaksin, dan lain-lain. Analisis sentimen masyarakat terhadap penurunan COVID-19 dapat memberikan wawasan yang berharga untuk para pengambil kebijakan dan pemangku kepentingan lainnya dalam mengukur efektivitas langkah-langkah penanganan pandemi dan meningkatkan komunikasi dengan masyarakat [2].

Dalam rangka menyikapi kompleksitas volume dan variasi tweet terkait COVID-19 di X App, metode klasifikasi sentimen menjadi penting. Dalam penelitian ini, metode klasifikasi *Naïve Bayes* dan *Support Vector Machine* (SVM) akan digunakan untuk menganalisis sentimen masyarakat terhadap penurunan kasus COVID-19. Kedua metode tersebut telah terbukti efektif dalam melakukan klasifikasi pada data teks dengan akurasi yang tinggi. Melalui analisis klasifikasi sentimen menggunakan kedua metode tersebut, diharapkan dapat ditemukan pola sentiment positif, negatif, dan netral terhadap penurunan COVID-19. Hasil dari penelitian ini dapat memberikan wawasan lebih lanjut tentang pandangan masyarakat terhadap penurunan kasus COVID-19 dan memberikan pemahaman tentang bagaimana opini dan perasaan masyarakat berkembang seiring dengan peristiwa yang berkaitan dengan pandemi [3].

Dengan demikian, penelitian ini bertujuan untuk menganalisis sentimen masyarakat terhadap penurunan covid-19 memakai metode klasifikasi *Naïve Bayes* dan *Support Vector Machine* pada platform X App. Hasil penelitian ini diharapkan dapat memberikan kontribusi bagi pemahaman lebih lanjut tentang pandangan masyarakat terhadap pandemi COVID-19 dan memberikan informasi yang berharga bagi pengambil kebijakan dalam menghadapi tantangan kesehatan global ini [4]

2. Metode

Gambar dibawah merupakan gambar diagram blok input, proses dan output



Gambar 1. Diagram Blok.

Berdasarkan pada Gambar 1 penelitian ini terdiri dari 3 tahapan yaitu input, proses dan output terlihat bahwa Input: Data yang digunakan sebagai masukan dalam pengambilan data yaitu memakai *X API Key*. Proses: Proses pengujian yang dilakukan menggunakan *Preprocessing Data*. Output: Hasil pengujian berupa laporan yang dihasilkan dari pengujian dan dianalisis menggunakan metode Algoritma *Naïve Bayes* dan *Support Vector Machine*.

1. Studi literature

Studi literatur yaitu mempelajari berbagai referensi tentang Algoritma *Naive Bayes* dan *Support Vector Machine* yaitu semua yang berhubungan dengan permasalahan yang telah dijelaskan.

2. Pengumpulan Data

Pengumpulan data cuitan tweet dari X App akan dikumpulkan lalu setelah itu akan diolah

3. Pengolahan Data

Setelah data diklasifikasikan lalu data tersebut akan diolah menggunakan Algoritma *Naive Bayes* dan *Support Vector Machine*.

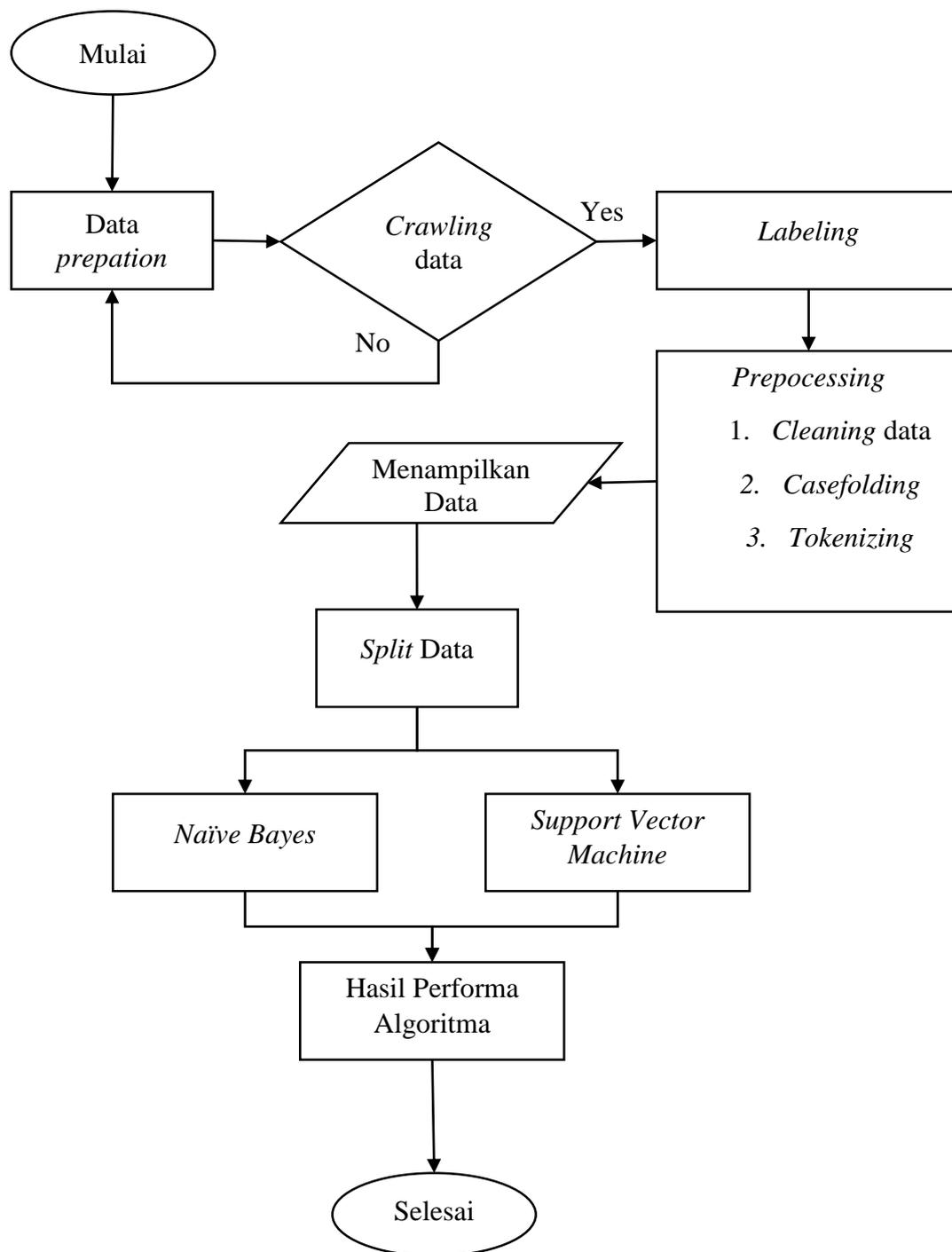
4. Analisis Data

Setelah mendapatkan data yang sudah diolah menjadi beberapa bagian maka akan dibuat sebuah Analisa.

5. Kesimpulan

Setelah data di Analisa maka tahapan selanjutnya adalah membuat kesimpulan.

2.1 Prosedur



Gambar 2. Flowchart.

Berdasarkan diagram alur penelitian yang ditunjukkan pada Gambar 2, penjelasan berikut menjelaskan setiap tahap penelitian yang dilakukan.

1. Mulai
2. *Crawling* data X App menggunakan API

Crawling dilakukan untuk mengambil data yang akan penulis pakai untuk bahan analisis sentimen. Karena lebih praktis dan efektif dibandingkan dengan pengumpulan data manual, teknik web scraping data digunakan untuk mengumpulkan data. Pada proses *scraping* menggunakan aplikasi *web scraping* yaitu, *google colab*. Pengambilan data dilakukan secara online, dan *web scraping* data digunakan untuk mengumpulkan data.

3. *Labeling Data*

Dilakukannya *Labeling Data* untuk pemberian label untuk data positif atau netral dan *negative*.

4. *Preprocessing Data*

Tujuan dari text *preprocessing* adalah untuk membersihkan dan mengubah data menjadi lebih terstruktur. Dalam penelitian ini, tahapan text *preprocessing* dilakukan dalam empat tahap, dan menggunakan *tools python 3.11*. Tahapan-tahap ini terdiri dari :

a. *Cleaning*

Langkah ini melibatkan pembersihan teks dari karakter-karakter yang tidak relevan atau mengganggu, seperti tanda baca, angka, *URL*, atau karakter khusus lainnya. Pembersihan juga mencakup normalisasi teks seperti menghapus spasi berlebih, mengatasi masalah khusus seperti tautan atau emotikon.

b. *Case Folding*

Langkah ini bertujuan mengonversi semua karakter dalam sebuah teks menjadi huruf kecil atau huruf besar, tergantung kebutuhan. Dan fungsi utama *casefolding* adalah normalisasi teks.

c. *Tokenizing*

Proses ini merupakan pemecah teks menjadi potongan yang lebih kecil, guna untuk mempermudah pemrosesan teks lebih lanjut.

5. Klasifikasi dengan Metode *Naïve Bayes* dan *Support Vector Machine*

Pada tahapan ini menggunakan metode *machine learning*. Dengan menggunakan metode *Naïve Bayes* dan *Support Vector Machine* untuk melakukan klasifikasi opini masyarakat terhadap kasus Covid-19 baik positif, netral maupun *negative*. Pada tahapan ini digunakan *python 3.11* untuk proses menklasifikasi.

6. Selesai

Data yang diambil menggunakan data kuantitatif untuk menganalisis data dari teks berbahasa Indonesia dan Inggris serta bahasa lainnya berupa opini tentang Covid-19 yang didapat dari *X App*.

3. Hasil dan Pembahasan

Tahapan pertama ialah *crawling* data, setelah data di *crawling* data disimpan dalam format csv. Setelah itu dilakukan proses data di *Labeling* bertujuan untuk menilai data tersebut dan diklasifikasikan ke dalam kelas *negative*, *positif*, dan *netral*. Kemudian proses selanjutnya dilakukanlah tahapan *Preprocessing Data* terdapat 3 tahapan dalam *Preprocessing Data* ialah *Cleaning Data* yang bertujuan membersihkan teks data, kemudian *casefolding* yang bertujuan penormalisasian kata dan *tokenizing* memecah kata. selanjut tahapan Pengklasifikasian menggunakan metode *Naïve bayes* dan *Support Vector Machine*.

3.1 Klasifikasi Naïve Bayes

Proses Klasifikasi Algoritma Naïve Bayes ialah sebuah proses dimana terdapat fitur ekstraksi dan pada proses klasifikasi akan berjalan dengan mengaplikasikan *Libray* pada bahasa pemrograman *python*. Yang terdiri dari *library* *sckit-learn* untuk proses klasifikasi, serta terdapat *library* *numpy* dan *pandas* untuk pembaca data dan pembagi data menjadi *data head* atau *data frame*. Dan untuk *library* *sckit-learn* yang digunakan oleh penulis adalah *CountVectorizer*, *Pipeline*, *Naïve Bayes*, *TfidfTransformer*, *Confussion matrix*, *MultinomialNB* dan *f1 Score*. Langkah selanjutnya adalah proses pembagian data testing dan data training, proses ini dilakukan agar memudahkan proses mencari nilai akurasi dan menghasilkan model yang baik.

1. Uji Model

Tabel 1. Model Confussion Matrix

		<i>Positive</i>	<i>Negative</i>
<i>Actual Class</i>	<i>Positive</i>	AA	BB
	<i>Negative</i>	BA	AB

$$\text{Akurasi} = \frac{TP+TN}{TP+TN+FP+FN} * 100\% \quad (1)$$

2. Evaluasi Model

Proses pengujian model menggunakan bantuan *libray confusion matrix*. Adapun hasil pada table 4.2 dibawah. Hasil dari evaluasi model terdapat di table 4.2 dengan menunjukkan hasil *confusion matrix*. Lalu selanjutnya proses pencarian atau penentuan nilai test dan nilai prediksi atau nilai yang sebenarnya nantinya akan dilakukan proses penentuan akurasi.

Tabel 2. Hasil Confussion Matrix

		<i>Positive</i>	<i>Negative</i>
<i>Actual Class</i>	<i>Positive</i>	377	95
	<i>Negative</i>	95	37

3. Hasil

Berdasarkan hasil pengujian evaluasi model pada table 2 dengan hasil *confusion matrix*, maka langkah selanjutnya dilakukan proses penentuan nilai nilai prediksi atau nilai test lalu dilakukan proses penentuan akurasi seperti pada gambar 3 di bawah ini.

```
[ ] accuracy = accuracy_score(y_test, y_pred)
confusion = confusion_matrix(y_test, y_pred)
report = classification_report(y_test, y_pred)

print("Accuracy:", accuracy)
print("\nConfusion Matrix:\n", confusion)
print("\nClassification Report:\n", report)
```

Accuracy: 0.8631578947368421

Confusion Matrix:

```
[[ 5  0  3]
 [ 2 72  8]
 [ 0  0  5]]
```

Classification Report:

	precision	recall	f1-score	support
negatif	0.71	0.62	0.67	8
netral	1.00	0.88	0.94	82
positif	0.31	1.00	0.48	5
accuracy			0.86	95
macro avg	0.68	0.83	0.69	95
weighted avg	0.94	0.86	0.89	95

Gambar 3. Hasil Akurasi *Naïve Bayes Classification*.

Nilai akurasi yang diperoleh dari proses pengujian model ialah sebesar 86,32% dengan proses perhitungan berdasarkan jumlah nilai dari *confusion matrix* lalu dibagi dengan seluruh data.

3.2 Klasifikasi *Support Vector Machine*

Proses Klasifikasi Algoritma *Support Vector Machine* ialah sebuah proses dimana terdapat fitur ekstraksi dan pada proses klasifikasi akan berjalan dengan mengaplikasikan *Libray* pada bahasa pemrograman *python*. *Support Vector Machine* adalah fungsi yang mempunyai dimensi yang lebih tinggi dan menggunakan algoritma latih pembelajaran yang berdasarkan teori optimasi, dan ruang virtual yang berupa *system linear*. Dan *Support Vector Machine* merupakan algoritma yang mempunyai teknologi terbaru dibandingkan algoritma yang lainnya, dan pemilihan fungsi kernel yang tepat seperti linear dan nonlinear sangat penting. Karena itu merupakan fungsi kernel menentukan ruang fitur dimana fungsi pengklasifikasian dicari. Kemudian tahapan awal untuk pengerjaan proses ekstrasi fitur serta klarifikasi adalah instalasi library yang diperlukan dan dipakai. Lalu setelah proses instalasi *library* dilakukan proses pendeklarasian *library* yang diinstall.

Setelah proses pendeklarasian dan instalasi *libray* kemudian dilakukan proses klasifikasi menggunakan bahasa pemrograman *python* dengan menggunakan kode pemrograman untuk *support vector machine*. Dan penulis juga menggunakan metode fungsi kernel linear untuk pengklasifikasian algoritma *support vector machine*. Lalu hasil akurasi juga sudah didapatkan ketika kode program berjalan.

```
[ ] clf = svm.SVC(kernel='linear') # Gunakan kernel linear untuk data ini

# Latih model SVM pada data latih
clf.fit(X_train, y_train)

# Lakukan prediksi pada data uji
y_pred = clf.predict(X_test)

accuracy = accuracy_score(y_test, y_pred)
confusion = confusion_matrix(y_test, y_pred)
report = classification_report(y_test, y_pred)
```

Gambar 4. Kode Pemrograman Klasifikasi

```
print("Accuracy:", accuracy)
print("\nConfusion Matrix:\n", confusion)
print("\nClassification Report:\n", report)
```

Accuracy: 0.9578947368421052

Confusion Matrix:

```
[[ 5  1  2]
 [ 0 82  0]
 [ 0  1  4]]
```

Classification Report:

	precision	recall	f1-score	support
negatif	1.00	0.62	0.77	8
netral	0.98	1.00	0.99	82
positif	0.67	0.80	0.73	5
accuracy			0.96	95
macro avg	0.88	0.81	0.83	95
weighted avg	0.96	0.96	0.96	95

Gambar 5. Hasil Akurasi Support Vector Machine.

Nilai akurasi yang diperoleh dari proses pengklasifikasian data menggunakan algoritma *Support Vector Machine* adalah sebesar 95,79% dengan proses perhitungan dan berdasarkan jumlah dari nilai *confusion matrix*.

4. Kesimpulan

Berdasarkan klasifikasi dengan metode *naïve bayes* dan *support vector machine*, maka diketahui bahwa data latih dan data uji itu persentasenya 80:20 lalu menghasilkan tingkat akurasi untuk metode algoritma *Naïve Bayes* didapatkan sebesar 86,32% dan untuk metode algoritma *Support Vector Machine* didapatkan akurasi sebesar 95,79%. Dan untuk akurasi, presisi, *recall* sebesar 85%, 94%, 86% dan untuk metode *Naïve Bayes* Akurasi, presisi, *recall* 95%, 96%, 96%. *Factor* mendapatkan akurasi diatas 85% ialah *factor* keberhasilan dalam proses split data dengan rasio 80:20, kemudian penggunaan Algoritma *Naïve Bayes* dan *Support Vector Machine*, lalu pada tahapan data *preprocessing* berjalan mengolah data sehingga data bisa menjadi bahan acuan akurasi bisa diatas 85%. Metode *Support Vector Machine* lebih baik dibandingkan metode algoritma *Naïve Bayes* karena dari nilai akurasinya lebih tinggi dan bagus.

Ucapan Terima Kasih

Penulis mengucapkan terimakasih kepada Program Studi di Teknik Elektro Fakultas Teknik Universitas Garut dan pihak lain yang telah memberikan kerjasama dalam penelitian ini.

Daftar Pustaka

- [1] V. C. M. ., N. J. P. Andre Ertanto, "ANALISA TOPIK TERHADAP KOMENTAR MENGENAI METAVERSE MENGGUNAKAN METODE CLUSTERING K-MEANS," *Jurnal Ilmu Komputer dan Sitem Informasi*, vol. 11, 2023.
- [2] R. I. P. H. U. W. A. R. Muhammad Iqbal Aditama, "Journal Information Engineering and Educational Technology," *Analisis Klasifikasi Sentimen Pengguna Media Sosial Twitter Terhadap Pengadaan Vaksin COVID-19*, p. 90, 2020.
- [3] ., A. W. Rani Puspita, "Analisis Sentimen terhadap Layanan Indihome di Twitter dengan Metode Machine Learning," *Jurnal Informatika Universitas Pamulang*, vol. 6, pp. 759-766, 2021.
- [4] S. W. S. Muhamad Anton Permana, "ANALISIS SENTIMEN PENGGUNAAN APLIKASI VIDEO CONFERENCE PADA ULASAN GOOGLE PLAY STORE MENGGUNAKAN METODE NBC (NAIVE BAYES CLASSIFIER)," *Jurnal Sistem Informasi dan Teknologi Informasi*, vol. 5, pp. 178-191, 2023.