

Analisis Sentimen Ulasan Pelanggan Online Ubi Madu Cilembu Abah Nana Menggunakan Algoritma Naïve Bayes

Muhammad Rafly Al Fattah Zain, Mia Kamayani*

Fakultas Teknologi Industri dan Informatika, Teknik Informatika, Universitas Muhammadiyah Prof. Dr. Hamka, Jakarta, Indonesia

Email: ¹raflyalfa16@gmail.com, ^{2,*}mia.kamayani@uhamka.ac.id

Email Penulis Korespondensi: mia.kamayani@uhamka.ac.id

Submitted: 11/08/2023; Accepted: 21/09/2023; Published: 25/09/2023

Abstrak—Penelitian ini bertujuan untuk menganalisis sentimen ulasan pelanggan online Ubi Madu Cilembu Abah Nana menggunakan algoritma Naïve Bayes. Penelitian ini mencakup dua tujuan utama, yaitu melakukan klasifikasi analisis sentimen ulasan menjadi positif dan negatif terhadap pelayanan dan produk Ubi Madu Cilembu Abah Nana, serta mengevaluasi tingkat akurasi dari hasil akhir klasifikasi tersebut. Data yang didapatkan dari aplikasi pesan makanan online seperti Gofood, Grabfood, dan Shopeefood. Data yang digunakan dalam penelitian ini berjumlah 259 data, dengan 310 data positif dan 49 data negatif. Setelah dilakukan eksperimen, diperoleh hasil akurasi 86.29% pada eksperimen 1 yang menggunakan operator Split Data, dan akurasi 86.12% pada eksperimen 2 yang menggunakan operator Cross Validation dengan bantuan ahli bahasa. Hasil penelitian ini menunjukkan bahwa algoritma Naïve Bayes dapat digunakan untuk mengklasifikasikan sentimen ulasan pelanggan terhadap pelayanan dan produk Ubi Madu Cilembu Abah Nana dengan tingkat akurasi yang cukup tinggi. Hasil ini dapat berguna bagi Ubi Madu Cilembu Abah Nana dalam meningkatkan kualitas pelayanan dan produk mereka berdasarkan umpan balik pelanggan. Selain itu, penelitian ini juga berkontribusi dalam bidang analisis sentimen dan pengolahan bahasa alami dengan menerapkan algoritma klasifikasi pada data ulasan pelanggan.

Kata Kunci: Analisis Sentimen; Naïve Bayes; RapidMiner; Klasifikasi; Ubi Madu Cilembu Abah Nana

Abstract—This research aims to analyze the sentiment of online customer reviews for Ubi Madu Cilembu Abah Nana using the Naïve Bayes algorithm. The study has two main objectives: to classify the sentiment analysis of reviews into positive and negative categories regarding the service and products of Ubi Madu Cilembu Abah Nana, as well as to evaluate the accuracy level of the final classification results. The data was collected from online food delivery applications such as Gofood, Grabfood, and Shopeefood. The data used in this study amounts to 259 entries, with 310 positive and 49 negative data points. After conducting experiments, an accuracy result of 86.29% was obtained in Experiment 1 using the Split Data operator, and an accuracy of 86.12% was achieved in Experiment 2 utilizing Cross Validation with the assistance of language experts. The findings of this research indicate that the Naïve Bayes algorithm can be employed to classify customer sentiment towards the service and products of Ubi Madu Cilembu Abah Nana with a significantly high accuracy rate. These results can be valuable for Ubi Madu Cilembu Abah Nana in enhancing their service and product quality based on customer feedback. Additionally, this study also contributes to the field of sentiment analysis and natural language processing by applying classification algorithms to customer review data.

Keywords: Sentiment Analysis; Naïve Bayes; RapidMiner; Classification; Ubi Madu Cilembu Abah Nana

1. PENDAHULUAN

Persaingan di pasar Ubi Madu Cilembu telah mendorong pemilik usaha untuk merumuskan strategi guna meningkatkan kualitas pelayanan dan produk dari Ubi Cilembu Abah Nana. Menyadari bahwa pelayanan dan produk yang monoton dapat menyebabkan kebosanan dan ketidakpuasan bagi konsumen, strategi tersebut bertujuan untuk menarik minat konsumen kembali. Prediksi minat konsumen menjadi landasan untuk menciptakan pelayanan yang lebih baik dan produk yang lebih bermutu, sehingga dapat meningkatkan kepuasan dan ketertarikan para konsumen. Produk Ubi Madu Cilembu Abah Nana merupakan karya pendirinya sendiri, Nana Haryana, yang setelah pensiun dari kemiliteran, memiliki niat untuk berkompetisi dengan bisnis Ubi Madu Cilembu lainnya. Berkat dukungan dari masyarakat yang membeli produknya, Ubi Madu Cilembu berhasil berkembang dan berdiri sebagai bisnis kuliner yang ingin menghadirkan yang terbaik bagi konsumennya [1]. Penelitian ini menggunakan data yang diperoleh dari pesan makanan online dan memanfaatkan Algoritma Naïve Bayes untuk mendapatkan ulasan pelanggan dan menjadi acuan dalam meningkatkan kualitas pelayanan dan kuantitas produk [2]. Dalam rangka mengklasifikasikan data mengenai tingkat kepuasan pelanggan terhadap pelayanan dan kuantitas produk, dilakukan analisis sentimen [3]. Analisis sentimen adalah bidang ilmu yang melakukan analisis berdasarkan pendapat, pandangan, opini, sikap, emosi, penilaian, dan evaluasi yang diekspresikan oleh individu terhadap suatu produk, individu, atau organisasi tertentu, serta mengenai isu, peristiwa, atau topik khusus [4]. Terdapat beberapa algoritma yang digunakan untuk melakukan analisis sentimen seperti Support Vector Machine, Naïve Bayes, Decision Tree, dan K-Nearest Neighbor [5]. Peneliti menggunakan algoritma Naïve Bayes karena mudah dipahami, memiliki beberapa karakteristik yang memudahkan implementasinya dan mengurangi kompleksitas perhitungan, dan Naïve Bayes dikenal memiliki kecepatan dan kesederhanaan yang cocok untuk digunakan dalam pengolahan data besar [6]. Berdasarkan uraian tersebut diharapkan dalam penelitian ini dapat meningkatkan dari kualitas pelayanan dan kuantitas produk.

Penelitian ini pernah dilakukan oleh Whita Parasati [7] dengan menggunakan algoritma Naïve Bayes untuk menganalisis sentimen dari 2.152 data ulasan pelanggan yang dikumpulkan dari tahun 2012 hingga 2019 melalui

teknik Web Scraping. Fokus analisis sentimen adalah aspek kepuasan pelanggan yang terdiri dari Makanan, Layanan, dan Atmosfer di restoran Bakso President Malang. Hasil analisis sentimen menunjukkan tingkat akurasi yang cukup baik. Untuk aspek Makanan, algoritma Naïve Bayes mampu mengenali pandangan positif atau negatif dari ulasan pelanggan terhadap kualitas makanan dengan tingkat akurasi mencapai 88%. Sedangkan pada aspek Layanan, algoritma tersebut dapat memahami pandangan pelanggan tentang kualitas layanan dengan akurasi 76%, yang sedikit lebih rendah dibandingkan dengan aspek Makanan. Untuk aspek Atmosfer, akurasi yang diperoleh mencapai 84%, menandakan bahwa Algoritma Naïve Bayes dapat mengenali pandangan pelanggan terhadap suasana atau atmosfer di restoran dengan baik. Hasil analisis sentimen ini memberikan manfaat penting bagi restoran Bakso President Malang karena memungkinkan mereka untuk lebih memahami perspektif pelanggan terhadap aspek kepuasan pelanggan. Dengan begitu, restoran dapat meningkatkan pelayanan dan kualitas produk mereka sesuai dengan harapan dan kebutuhan pelanggan. Selain itu, visualisasi hasil analisis sentimen dalam bentuk dashboard dengan filter waktu, aspek, dan sentimen juga membantu dalam penggunaan dan pemahaman data dengan lebih efisien.

Penelitian ini juga pernah dilakukan oleh Nia Saurina [8] dengan menggambarkan pandangan pelanggan tentang Ecoprint Batik dan menilai sejauh mana Batik Ecoprint ramah lingkungan dengan menggunakan metode klasifikasi Naïve Bayes dan KNN. Melalui analisis data dari 5.736 tweet dengan sentimen positif dan 1.707 tweet dengan sentimen negatif, dari total 17.197 tweet yang dianalisis, ditemukan bahwa hanya 39 tweet yang menyoroti tentang sifat ramah lingkungan dari Batik Ecoprint, sementara mayoritas lainnya lebih banyak membahas mengenai kualitas Batik Ecoprint.

Penelitian ini juga pernah dilakukan oleh Firman Noor Hasan [9] dengan mengidentifikasi jumlah ulasan positif dan negatif terkait pelayanan Grab Indonesia melalui penggunaan dataset. Selain itu, penelitian ini bertujuan untuk mengevaluasi hasil dari analisis data dan menilai efisiensi layanan Grab Indonesia. Dari hasil output yang dianalisis, terdapat 1000 dataset yang menunjukkan tingkat kepuasan pelanggan terhadap layanan Grab Indonesia. Terdapat 911 dataset dengan sentimen positif dan 89 dataset dengan sentimen negatif. Untuk sentimen negatif, tingkat precision yang dihasilkan adalah 57%, recall sebesar 67%, dan f1-score sebesar 62%. Sementara itu, untuk sentimen positif, tingkat precision adalah 97%, recall sebesar 95%, dan f1-score sebesar 96%.

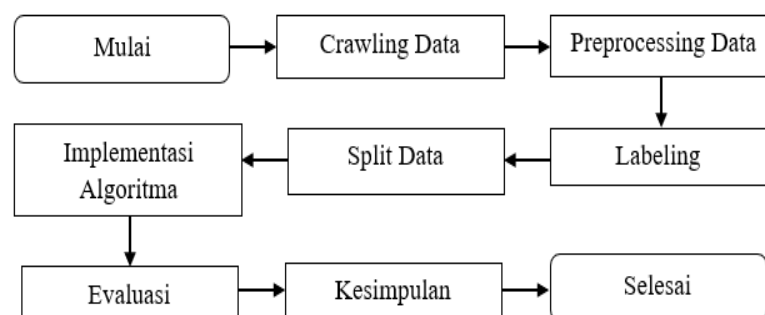
Penelitian serupa juga pernah dilakukan oleh Humairah [10] dengan menilai tingkat kepuasan produk di toko online Rubylicious menggunakan sejumlah besar data ulasan, yang menyebabkan proses analisis menjadi sulit dan kurang efektif. Untuk mengatasi tantangan tersebut, digunakan metode analisis sentimen dengan menggunakan algoritma Naïve Bayes Classifier (NBC) pada ulasan barang. Hasil analisis mencakup sentimen positif, negatif, dan netral. Hasil dari analisis sentimen tersebut menunjukkan akurasi sebesar 82,39%, precision mencapai 91,3%, recall mencapai 51%, dan f1-measure mencapai 51,6%.

Penelitian serupa juga pernah dilakukan oleh Dwi Normawati [11] ini dengan menganalisis sentimen cuitan di Twitter menggunakan metode Naïve Bayes Classifier (NBC). Penelitian ini menggunakan 8 cuitan sebagai studi kasus, dengan 5 cuitan sebagai data latih dan 3 cuitan sebagai data uji. Proses melibatkan crawling data dari Twitter, preprocessing data, klasifikasi menggunakan NBC, dan evaluasi performa dengan confusion matrix. Hasil penelitian menunjukkan akurasi sebesar 82%, presisi 93%, dan recall 52%. Penelitian ini lebih menekankan pada penjelasan proses daripada angka-angka hasilnya. Penelitian berikutnya akan melibatkan dataset yang lebih besar dan sistem aplikasi berbasis komputer.

2. METODOLOGI PENELITIAN

2.1 Tahapan Penelitian

Tahapan yang digunakan dalam penelitian ini antara lain: Crawling Data, Split Data, Preprocessing Data, Labeling, Split Data, implementasi algoritma, hasil evaluasi dengan membandingkan antara Eksperimen 1 dengan menggunakan operator split data dan Eksperimen 2 dengan menggunakan operator Cross Validation, serta penarikan kesimpulan dari perbandingan kedua Eksperimen. Tahapan penelitian dapat diilustrasikan pada Gambar 1 dengan rincian sebagai berikut:



Gambar 1. Tahapan Penelitian

- a. Crawling Data
Tahap ini meliputi pengumpulan data, tetapi penelitian ini tidak sama dengan kebanyakan penelitian lainnya, penelitian ini dilakukan dengan pengambilan ulasan secara Online dari aplikasi GoFood, GrabFood, dan ShopeeFood [12].
- b. Preprocessing Data
Preprocessing adalah proses untuk mengubah istilah-istilah dalam kalimat agar sesuai dengan standar yang diinginkan. Tujuan dari proses ini adalah untuk mempersiapkan data latih yang baik dan fitur-fitur yang diekstrak sesuai dengan kebutuhan, sehingga memudahkan proses pengolahan data [13].
- c. Labeling
Data yang digunakan adalah data yang diambil pada pesan makanan online seperti GoFood, GrabFood, dan ShopeeFood. Lalu dilakukan pelabelan secara manual dengan memasukkan ke dalam kelas positif dan negatif pada sentimen kata pada bagian ulasan [14].
- d. Split Data
Tahap dalam split data dilakukan dengan menggunakan RapidMiner Studio yang menyangkut hasil dari tahapan pada labeling yang dibagi menjadi dua data yaitu data training dan data testing. Pada topik penelitian yang diangkat ini pembagian data dilakukan dengan membagi data training lalu akan diuji pada Naïve Bayes dan data testing yang akan diuji [15].
- e. Implementasi Algoritma
Pada tahap ini, dilakukan implementasi algoritma Naïve Bayes Classifier pada data latih [16]. Algoritma ini merupakan metode klasifikasi yang menggunakan teorema Bayes dan mengasumsikan bahwa semua fitur (kata-kata dalam konteks ini) saling independen. Dengan menggunakan algoritma ini, peneliti dapat memprediksi apakah sentimen dari ulasan pelanggan tersebut bersifat positif atau negatif berdasarkan kata-kata yang muncul dalam ulasan tersebut.
- f. Evaluasi
Pada tahap ini yaitu Evaluasi, untuk menghitung nilai hasil akhir dari analisis klasifikasi dengan accuracy, precision, dan recall [17] menggunakan tabel confusion matrix. Bertujuan untuk membedakan berdasarkan pembagian dataset dimana eksperimen 1 menggunakan operator split data sedangkan eksperimen 2 menggunakan operator cross validation. Hasil dari perhitungan tersebut dapat ditampilkan dalam bentuk tabel confusion matrix [18].
- g. Kesimpulan
Pada kesimpulan ini didapatkan hasil dari perbandingan eksperimen 1 dengan eksperimen 2 apakah hasil yang akan diinterpretasikan oleh analisis sentimen ulasan yang bersifat positif dan negatif dari tingkat kepuasan.

2.2 Preprocessing

Preprocessing merupakan proses yang bertujuan untuk mengurangi kata-kata yang kurang relevan dalam data, meningkatkan kualitas data, dan menghasilkan teks yang lebih optimal dalam dataset. Terdapat beberapa tahapan yang digunakan dalam preprocessing yaitu,

- a. Cleansing : berguna untuk menghapus beberapa kata yang kurang berguna dalam data.
- b. Case Folding : mengubah huruf kapital menjadi huruf kecil.
- c. Tokenizing : membagi kalimat ke beberapa kata dengan jarak yang ada didalam kalimat.
- d. Stemming : mengubah kata yang berimbuhan menjadi kata dasar.
- e. Filter Stopword : melakukan eliminasi terhadap kata-kata yang tidak memiliki makna, seperti kata penghubung dan kata ganti orang ketiga.
- f. Filter token (by length) : mengeliminasi kata yang sudah ditentukan jumlah hurufnya.

2.3 Algoritma Naïve Bayes

Tugas-tugas Data Mining melibatkan klasifikasi data, yaitu membagi atau mengklasifikasikan data ke dalam kelas atau beberapa kelas yang sebelumnya telah ditentukan. Salah satu metode yang digunakan untuk melakukan klasifikasi ini adalah Naïve Bayes. Naïve Bayes merupakan metode machine learning yang memanfaatkan perhitungan probabilitas dan statistik yang diperkenalkan oleh ilmuwan Inggris bernama Thomas Bayes. Metode ini bekerja dengan mengestimasi probabilitas di masa yang akan datang berdasarkan pengalaman di masa yang sudah terjadi [19]. Untuk rumus Naïve Bayes [20] terdapat pada persamaan rumus (1). Rumus Naïve Bayes

$$P(H|X) = \frac{P(X|H) \cdot P(H)}{P(X)} \quad (1)$$

Keterangan

X : Data dengan class yang belum diketahui

H : Hipotesis data merupakan suatu class spesifik

P(H|X) : Probabilitas hipotesis H berdasarkan kondisi X (posteriori probabilitas)

P(H) : Probabilitas hipotesis H (prior probabilitas)

P(X|H) : Probabilitas X berdasarkan kondisi pada hipotesis H

P(X) : Probabilitas X

2.4 Confusion Matrix

Confusion Matrix merupakan sebuah tabel yang digunakan dalam menganalisis seberapa jauh hasil klasifikasi yang mampu memprediksi dengan akurat kelas-kelas data. Untuk rumus Confusion Matrix terdapat pada persamaan rumus

Tabel 1. Confusion Matrix

Prediction Class	Actual Class	
	Pred. Positive	Pred. Negative
	Negative	TP
Positive	FP	TN

Dalam pengukuran performa klasifikasi dengan dilakukan perhitungan dengan beberapa cara, namun cara yang sering digunakan yaitu menggunakan accuracy, precision, dan recall [21]. Accuracy merupakan persentase dari suatu prediksi model yang sudah dibuat. Untuk menghitung nilai Accuracy nya dilakukan dengan menggunakan persamaan rumus (2).

$$\text{Accuracy} = \frac{(TP+TN)}{(TP+TN+FN+FP)} \times 100\% \tag{2}$$

Mengukur performa klasifikasi yang digunakan selain menghitung akurasi adalah menghitung precision dan recall. Precision merupakan tingkat ketepatan antara informasi yang digunakan guna mendapatkan jawaban yang diberikan system. Untuk menghitung nilai precision nya dilakukan dengan menggunakan persamaan rumus (3).

$$\text{Precision} = \frac{TP}{(TP+FN)} \times 100\% \tag{3}$$

Recall merupakan perbandingan antara materi yang relevan dengan jumlah materi yang relevan. Untuk menghitung nilai recall nya dapat dilakukan dengan menggunakan persamaan rumus (4).

$$\text{Recall} = \frac{TP}{(TP+FN)} \times 100\% \tag{4}$$

3. HASIL DAN PEMBAHASAN

3.1 Hasil Crawling Data

Pada tahap awal, telah berhasil mengumpulkan 359 data dari layanan makanan online, seperti Gofood, Grabfood, dan Shopeefood. Proses pengambilan data dilakukan secara manual melalui aplikasi masing-masing platform untuk memastikan kualitas dan keakuratan data yang diambil. Dalam analisis ini, penelitian menggunakan dua kata kunci, yaitu Positif dan Negatif, sebagai acuan untuk menilai sentimen atau tanggapan pelanggan terhadap layanan-layanan tersebut. Bahasa yang digunakan dalam pengambilan data adalah Bahasa Indonesia, sehingga hasil analisis dapat memberikan pemahaman yang lebih baik mengenai ulasan dari pengguna layanan makanan online, seperti yang ditunjukkan pada gambar 2.

	A	B	C	D
	Nama Pembeli	Tanggal Pembelian	Tingkat Kepuasan	Ulasan
1				
2	ivicrown	15/05/2023	sangat puas	Ubi nya manis, mantap
3	aditama2995	15/05/2023	sangat puas	rasanya lezat
4	ivicrown	11/05/2023	sangat puas	tidak mengecewakan beli disini
5	ivicrown	09/05/2023	sangat puas	rasa ubinya enak banget
6	endhar19	06/05/2023	sangat puas	mantap ubinya
7	w****0	06/05/2023	sangat puas	rasa enak
8	ivicrown	23/04/2023	sangat puas	pas sampai masih hangat
9	daraudyamila	17/04/2023	sangat puas	tidak gagal, goodjob
10	ivicrown	16/04/2023	sangat puas	mantap lah
11	melisa_nathalie	14/04/2023	sangat puas	puas beli disini
12	ivicrown	14/04/2023	sangat puas	harga terjangkau, enak
13	ivicrown	10/04/2023	sangat puas	enak dan manis
14	ivicrown	08/04/2023	sangat puas	manis ubinya tidak gagal
15	ivi_crown	23/03/2023	sangat puas	masih hangat, makasih
16	fadhilamujahidah	19/02/2023	sangat puas	cocok dinikmati malam hari
17	widiawati220901	08/01/2023	sangat puas	tidak menyesal beli disini
18	atindedi	05/12/2022	sangat puas	enak sekali
19	yussi D.	15/05/2023	sangat puas	mantap betul
20	nandhita	15/05/2023	sangat puas	gurih dan nikmat

Gambar 2. Hasil Crawling

3.2 Preprocessing

Setelah melakukan pengumpulan data sebanyak 359 data, proses selanjutnya dalam penelitian ini adalah

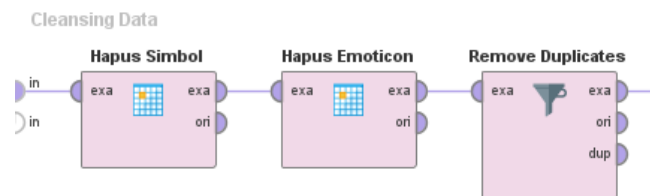
preprocessing data. Preprocessing data dilakukan untuk mengurangi kata-kata yang kurang relevan dalam data, meningkatkan kualitas data, dan menghasilkan teks yang lebih optimal dalam dataset. Proses ini mencakup beberapa tahapan, yaitu:

- a. Dalam proses Cleansing Data, langkah awal yang diambil peneliti adalah menggunakan operator replace. Setelah itu, peneliti mengatur parameter replace sesuai dengan data yang perlu dibersihkan, seperti menghapus data yang ganda (double), data yang mengandung simbol atau tanda baca, dan juga data yang mengandung emoticon seperti yang ditunjukkan pada Tabel 1.

Tabel 1. Hasil Sampel Cleansing Data

Input	Output
<p>udah beberapa kali beli disini. baru kali ini agak kecewa karna dapetnya udah ga panas. bahkan anget aja ngga. padahal jaraknya deket (<1 km) tapi dapetnya gak panas sama sekali. untungya setelah saya angetin sendiri ras tetep enak kayak biasa. respon cepat, ubinya cilembunya asli, manisnya ga lebay... kripik talasnya juga renyah... 🍌🍌 enak banget musim hujan bgini makan ubi cilembu hangat, bajigur hangat.. Duh berasa di Puncak!! Hepi banget 🍌🍌❤️</p> <p>ENAK BANGETTTT SUKAAAKKK LOVEE ❤️</p> <p>Manis, pas nyampe masih anget... Enak banget, suka. Alternatif untuk yang lagi diet 🍌🍌 rasanya enak banget maniss legiitt.. dan kualitasnya oke punyaa 😊😊</p>	<p>udah beberapa kali beli disini baru kali ini agak kecewa karna dapetnya udah ga panas bahkan anget aja ngga padahal jaraknya deket km tapi dapetnya gak panas sama sekali untungya setelah saya angetin sendiri ras tetep enak kayak biasa respon cepat ubinya cilembunya asli manisnya ga lebay kripik talasnya juga renyah enak banget musim hujan bgini makan ubi cilembu hangat bajigur hangat Duh berasa di Puncak Hepi banget</p> <p>ENAK BANGETTTT SUKAAAKKK LOVEE</p> <p>Manis pas nyampe masih anget Enak banget suka Alternatif untuk yang lagi diet rasanya enak banget maniss legiitt dan kualitasnya oke punyaa</p>

Tahapan Clenasing Data yaitu, menghapus data hasil pengambilan data manual pada pesan makanan online menggunakan aplikasi (software) RapidMiner. Seperti data yang terdapat simbol atau tanda baca, data yang terdapat emoticon, dan terdapat data yang ganda (double). Dari operator yang digunakan dalam proses penghapusan data yang bersifat duplicate atau double, juga dilakukan penghapusan emoticon, simbol, dan tanda baca lainnya. Proses Cleansing Data seperti yang ditunjukkan pada gambar 3.



Gambar 3. Proses Cleansing Data

- b. Pada Case Folding ini, proses yang digunakan dalam mengubah suatu kata atau kalimat dengan menggunakan lower case text yang bertujuan mengubah kata atau kalimat dalam bentuk huruf kapital menjadi huruf kecil. Hasil dari proses Case Folding dapat dilihat pada Tabel 2.

Tabel 2. Hasil Case Folding

Input	Output
<p>udah beberapa kali beli disini baru kali ini agak kecewa karna dapetnya udah ga panas bahkan anget aja ngga padahal jaraknya deket km tapi dapetnya gak panas sama sekali untungya setelah saya angetin sendiri ras tetep enak kayak biasa</p> <p>respon cepat ubinya cilembunya asli manisnya ga lebay kripik talasnya juga renyah enak banget musim hujan bgini makan ubi cilembu hangat bajigur hangat duh berasa di Puncak Hepi banget</p> <p>ENAK BANGETTTT SUKAAAKKK LOVEE</p> <p>Manis pas nyampe masih anget Enak banget suka Alternatif untuk yang lagi diet</p>	<p>udah beberapa kali beli disini baru kali ini agak kecewa karna dapetnya udah ga panas bahkan anget aja ngga padahal jaraknya deket km tapi dapetnya gak panas sama sekali untungya setelah saya angetin sendiri ras tetep enak kayak biasa</p> <p>respon cepat ubinya cilembunya asli manisnya ga lebay kripik talasnya juga renyah enak banget musim hujan bgini makan ubi cilembu hangat bajigur hangat duh berasa di puncak hepi banget</p> <p>enak bangetttt sukaaakkk lovee</p> <p>manis pas nyampe masih anget enak banget suka alternatif untuk yang lagi diet</p>

Input	Output
rasanya enak banget maniss legiitt dan kualitasnya oke punyaa	rasanya enak banget maniss legiitt dan kualitasnya oke punyaa

- c. Selanjutnya penghapusan tahap Tokenizing ini bertujuan memecah kata per kata dalam suatu kalimat. Operator yang digunakan yaitu Tokenizing, dan memerlukan bantuan tabel rumus yang terdapat pada web.kaggle.com [22] guna mengerjakan proses seluruh data secara otomatis. Hasil dari Tokenizing dapat dilihat pada Tabel 3.

Tabel 3. Hasil Tokenizing

Input	Output
udah beberapa kali beli disini baru kali ini agak kecewa karna dapetnya udah ga panas bahkan anget aja ngga padahal jaraknya deket km tapi dapetnya gak panas sama sekali untungya setelah saya angetin sendiri ras tetep enak kayak biasa	udah, beberapa, kali, beli, disini, baru, kali, ini, agak, kecewa, karna, dapetnya, udah, ga, panas, bahkan, anget, aja, ngga, padahal, jaraknya, deket, km, tapi, dapetnya, gak, panas, sama, sekali, untungya, setelah, saya, angetin, sendiri, ras, tetep, enak, kayak, biasa,
respon cepat ubinya cilembunya asli manisnya ga lebay kripik talasnya juga renyah	respon, cepat, ubinya, cilembunya, asli, manisnya, ga, lebay, kripik, talasnya, juga, renyah,
enak banget musim hujan bgini makan ubi cilembu hangat bajigur hangat duh berasa di puncak hepi banget	enak, banget, musim, hujan, bgini, makan, ubi, cilembu, hangat, bajigur, hangat, duh, berasa, di, puncak, hepi, banget,
enak bangetttt sukaaaakkk lovee	enak, bangetttt, sukaaaakkk, lovee,
manis pas nyampe masih anget enak banget suka alternatif untuk yang lagi diet	manis, pas, nyampe, masih, anget, enak, banget, suka, alternatif, untuk, yang, lagi, diet,
rasanya enak banget maniss legiitt dan kualitasnya oke punyaa	rasanya, enak, banget, maniss, legiitt, dan, kualitasnya, oke, punyaa,

- d. Selanjutnya, tahap Stemming bertujuan mengubah kata-kata berimbuhan, kata sambung, serta kata-kata yang bersifat kekinian atau bahasa gaul menjadi kata dasar agar dapat digunakan dalam proses stemming, seperti yang ditunjukkan pada Tabel 4.

Tabel 4. Kata untuk Stemming

Input	Output
sdh	sudah
yg	yang
karna, krn	karena
ga, gak, nggak	enggak
tp	tapi
sm	sama
kl, klo	kalau
spt	seperti
bs	bisa
lsng	langsung
dtng	datang
lg	juga
lg	lagi
d	di
bgt	banget
sy, syaa	saya
bru	baru
aga	agak
n	dan
bbrp	beberapa
kmrn	kemarin
tdk	tidak
gitu	begitu
ky	kayak
alsn	alasan
kluar	keluar

Hasil dari stemming, didapatkan hasil proses stemming dengan mengubah kata dari kata kekinian atau bahasa gaul menjadi kata dasar dan membagi nya berdasarkan parameter yang sudah ditentukan. Hasil dari Stemming dapat dilihat pada Tabel 5.

Tabel 5. Hasil Stemming

Input	Output
<p>udah, beberapa, kali, beli, disini, baru, kali, ini, agak, kecewa, karna, dapetnya, udah, ga, panas, bahkan, anget, aja, nggak, padahal, jaraknya, dekat, km, tapi, dapetnya, gak, panas, sama, sekali, untungnya, setelah, saya, angetin, sendiri, ras, tetep, enak, kayak, biasa,</p> <p>respon, cepat, ubinya, cilembunya, asli, manisnya, ga, lebay, kripik, talasnya, juga, renyah, enak, banget, musim, hujan, bgini, makan, ubi, cilembu, hangat, bajigur, hangat, duh, berasa, di, puncak, hepi, banget, enak, bangetttt, sukaaakkk, lovee,</p> <p>manis, pas, nyampe, masih, anget, enak, banget, suka, alternatif, untuk, yang, lagi, diet, rasanya, enak, banget, maniss, legiitt, dan, kualitasnya, oke, punyaa,</p>	<p>udah, beberapa, kali, beli, disini, baru, kali, ini, agak, kecewa, karna, dapetnya, udah, tidak, panas, bahkan, anget, aja, enggak, padahal, jaraknya, dekat, kamu, tapi, dapetnya, tidak, panas, sama, sekali, untungnya, setelah, saya, angetin, sendiri, ras, tetep, enak, kayak, biasa,</p> <p>respon, cepat, ubinya, cilembunya, asli, manisnya, tidak, lebay, kripik, talasnya, juga, renyah, enak, banget, musim, hujan, begini, makan, ubi, cilembu, hangat, bajigur, hangat, aduh, berasa, di, puncak, hepi, banget, enak, bangetttt, sukaaakkk, lovee,</p> <p>manis, pas, nyampe, masih, anget, enak, banget, suka, alternatif, untuk, yang, lagi, diet, rasanya, enak, banget, maniss, legiitt, dan, kualitasnya, oke, punyaa,</p>

e. Lalu, pada tahap ini yaitu hasil dari tokenizing yang bertujuan mengambil kata-kata penting. Tahap ini digunakan library Corpus stopwords (Indonesia) dan rumus dari web.kaggle.com [23] yang bertujuan mengambil kata-kata yang penting pada kamus corpus stopwords yang sudah ada dalam bentuk bahasa Indonesia. Hasil dari filter stopwords dapat dilihat pada Tabel 6.

Tabel 6. Hasil Filter Stopword

Input	Output
<p>udah, beberapa, kali, beli, disini, baru, kali, ini, agak, kecewa, karna, dapetnya, udah, tidak, panas, bahkan, anget, aja, enggak, padahal, jaraknya, dekat, kamu, tapi, dapetnya, tidak, panas, sama, sekali, untungnya, setelah, saya, angetin, sendiri, ras, tetep, enak, kayak, biasa,</p> <p>respon, cepat, ubinya, cilembunya, asli, manisnya, tidak, lebay, kripik, talasnya, juga, renyah, enak, banget, musim, hujan, begini, makan, ubi, cilembu, hangat, bajigur, hangat, aduh, berasa, di, puncak, hepi, banget, enak, bangetttt, sukaaakkk, lovee,</p> <p>manis, pas, nyampe, masih, anget, enak, banget, suka, alternatif, untuk, yang, lagi, diet, rasanya, enak, banget, maniss, legiitt, dan, kualitasnya, oke, punyaa,</p>	<p>udah, beberapa, kali, beli, disini, baru, kali, ini, agak, kecewa, karna, dapetnya, udah, tidak, panas, bahkan, anget, aja, enggak, padahal, jaraknya, dekat, kamu, tapi, dapetnya, tidak, panas, sama, sekali, untungnya, setelah, saya, angetin, sendiri, ras, tetep, enak, kayak, biasa,</p> <p>respon, cepat, ubinya, cilembunya, asli, manisnya, tidak, lebay, kripik, talasnya, juga, renyah, enak, banget, musim, hujan, begini, makan, ubi, cilembu, hangat, bajigur, hangat, aduh, berasa, puncak, hepi, banget, enak, bangetttt, sukaaakkk, lovee,</p> <p>manis, pas, nyampe, masih, anget, enak, banget, suka, alternatif, lagi, diet, rasanya, enak, banget, maniss, legiitt, kualitasnya, oke, punyaa,</p>

f. Pada tahap ini, Filter Token (By Length) proses membuang kata yang sudah ditentukan yaitu kalimat yang terdiri dari minimal 3 huruf. Hasil dari filter token (by length) dapat dilihat pada Tabel 7.

Tabel 7. Hasil Filter Token (By Length)

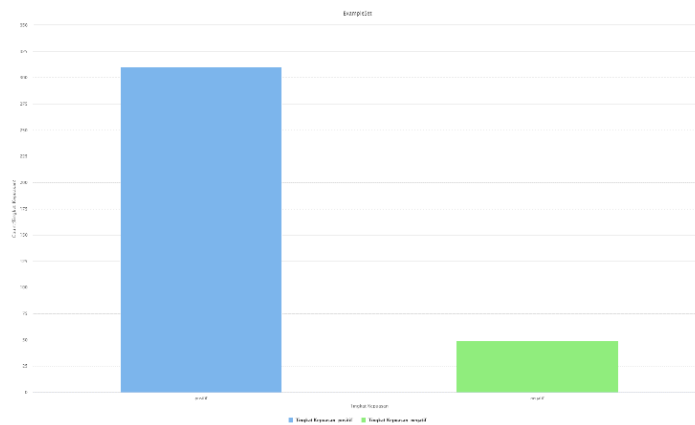
Input	Output
<p>udah, beberapa, kali, beli, disini, baru, kali, ini, agak, kecewa, karna, dapetnya, udah, tidak, panas, bahkan, anget, aja, enggak, padahal, jaraknya, dekat, kamu, tapi, dapetnya, tidak, panas, sama, sekali, untungnya, setelah, saya, angetin, sendiri, ras, tetep, enak, kayak, biasa,</p> <p>respon, cepat, ubinya, cilembunya, asli, manisnya, tidak, lebay, kripik, talasnya, juga, renyah, enak, banget, musim, hujan, begini, makan, ubi, cilembu, hangat, bajigur, hangat, aduh, berasa, puncak, hepi, banget, enak, bangetttt, sukaaakkk, lovee,</p> <p>manis, pas, nyampe, masih, anget, enak, banget, suka, alternatif, lagi, diet,</p>	<p>udah beberapa kali beli disini baru kali agak, kecewa karna dapetnya udah tidak panas bahkan anget enggak padahal jaraknya dekat kamu tapi dapetnya tidak panas sama sekali untungnya setelah saya angetin sendiri tetep enak kayak biasa</p> <p>respon cepat ubinya cilembunya asli manisnya tidak lebay kripik talasnya juga renyah enak banget musim hujan begini makan cilembu hangat bajigur hangat aduh berasa puncak hepi banget enak bangetttt sukaaakkk lovee manis nyampe masih anget enak banget suka alternatif lagi diet</p>

Input	Output
rasanya, enak, banget, maniss, legiitt, kualitasnya, oke, punyaa,	rasanya enak banget maniss legiitt kualitasnya punyaa

3.2 Labeling

Pada tahap ini memiliki empat proses dalam pengklasifikasian data. Dengan Labeling untuk melabelkan data yang akan dianalisis, lalu Split Data untuk membagi data menjadi dua yaitu Data Training (data latih) dan Data Testing (data uji), implementasi algoritma dengan menggunakan operator Naïve Bayes, dan terakhir yaitu Evaluasi dengan menambahkan grafik dan wordcloud sebagai hasil akhir dari uji data. Tahap mengevaluasi peneliti menggunakan dua Eksperimen untuk dianalisis yaitu dengan menggunakan operator Split Data dan operator Cross Validation.

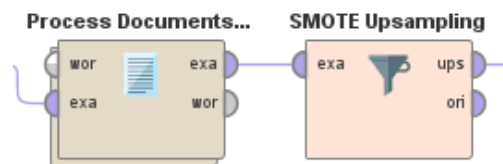
- Peneliti menggunakan sampel labeling data, gambar 4 merupakan gambar grafik dari hasil Labeling manual yang telah didapatkan dalam bentuk kelas positif dan negatif. Grafik dari labeling dapat dilihat pada gambar 5.



Gambar 4. Grafik Hasil Labeling

Didapatkan dari jumlah data class positif yang didapatkan sebanyak 248 data dan data class negatif didapatkan sebanyak 39 data.

- Tahap SMOTE bertujuan agar kelas yang tidak seimbang dimodifikasi menjadi seimbang guna mendapatkan sintetik kelas baru untuk kinerja dari klasifikasi menjadi lebih baik dari kelas tersebut sebelum masuk ke tahap split data maupun Cross Validation, guna mencegah terjadinya overfitting maupun underfitting maka peneliti menggunakan operator SMOTE, operator ini berguna untuk menyeimbangkan dataset. Tahap SMOTE dapat dilihat pada gambar 5.



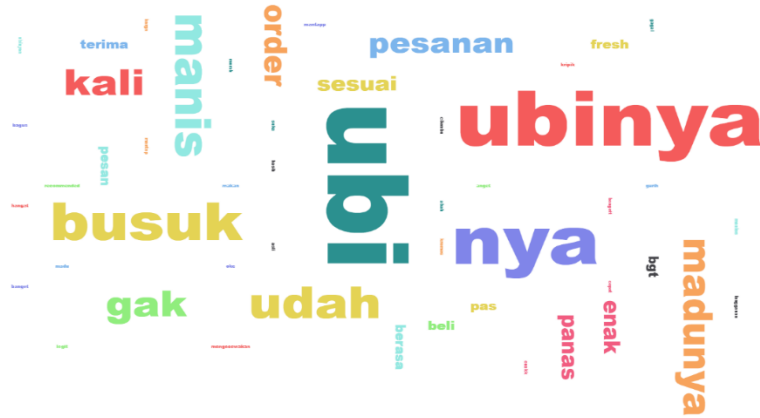
Gambar 5. Proses SMOTE

Visualisasi wordcloud digunakan untuk melihat kata yang sering muncul pada ulasan positif yang ditunjukkan pada gambar 6. Contoh dari kata-katanya yaitu: 'enak', 'ubinya', 'manis', 'banget', 'ubi', 'banget', 'nya', 'terima', 'kasih', 'sesuai', 'beli', 'mantap', 'gurih', 'enakk', 'langganan', 'suka'.



Gambar 6. Tampilan wordcloud ulasan positif

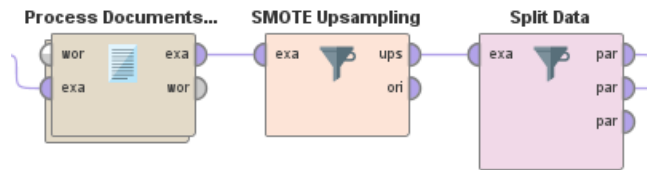
Visualisasi wordcloud digunakan untuk melihat kata yang sering muncul pada ulasan negatif yang ditunjukkan pada gambar 7. Dari contoh kata-kata yang sering muncul yaitu sebagai berikut: 'ubi', 'ubinya', 'nya', 'busuk', 'manis', 'gak', 'udah', 'madunya', 'order', 'pesanan', 'sesuai', 'enak', 'panas', 'terima', 'fresh', 'bgt', 'terima', 'pesan'.



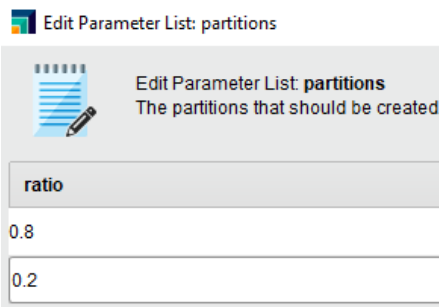
Gambar 7. Tampilan wordcloud ulasan negatif

3.3 Split Data

tahap selanjutnya yaitu split data dimana data dibagi menjadi dua yaitu data training (data latih) dan data testing (data uji). Pada topik penelitian yang diangkat ini pembagian data dilakukan dengan membagi data training (data latih) sebesar 80% atau 0.8 sedangkan data testing (data uji) sebesar 20% atau 0.2 menggunakan data uji, seperti yang ditunjukkan pada gambar 8 dan 9.



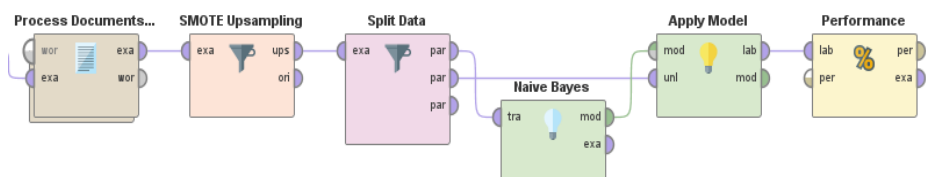
Gambar 8. Proses Split Data



Gambar 9. Operator Split Data

3.4 Implementasi Algoritma Naïve Bayes

Tahapan selanjutnya implementasi algoritma naïve bayes tahap implementasi klasifikasi algoritma Naïve Bayes, dalam tahap ini data training akan dianalisis dan diklasifikasikan menggunakan algoritma Naïve Bayes, sedangkan data testing dimasukkan ke dalam bentuk pemodelan. Proses dari penerapan algoritma naïve bayes dapat dilihat pada gambar 10.



Gambar 10. Proses penerapan algoritma naïve bayes

Naïve bayes digunakan untuk klasifikasi data, yaitu membagi atau mengklasifikasi data ke dalam kelas atau beberapa kelas yang sebelumnya sudah didefinisikan.

3.5 Evaluasi

Tahapan terakhir yaitu Evaluasi, Peneliti membandingkan 2 eksperimen hasil analisis klasifikasi yang berbeda dengan menggunakan Naïve Bayes dengan operator Split Data dan Naïve Bayes dengan Cross Validation. Hasil dari perhitungan tersebut dapat ditampilkan dalam bentuk tabel confusion matrix. Berikut adalah perbandingan 2 eksperimen dengan hasil yang dapat ditampilkan menggunakan tabel confusion matrix.

1. Eksperimen 1

Hasil pada eksperimen 1 merupakan hasil yang didapatkan dari analisis dengan menggunakan Naïve Bayes dengan menggunakan operator Split Data mendapatkan score confusion matrix. Tabel hasil dari eksperimen 1 dapat dilihat pada tabel 7.

Tabel 7. Hasil Eksperimen 1

	True Positive	True Negative	Class Precision
Pred. positive	45	0	100.00%
Pred. negative	17	62	78.48%
Class recall	72.58%	100.00%	86.29%
Accuracy			

Didapatkan dari eksperimen 1 dengan menggunakan operator Split Data hasil dari accuracy sebesar 86.29%. untuk kelas positif hasil Precision sebesar 100%, Recall 72.58%, dan kelas negatif hasil Precision sebesar 78.48%, Recall 100%.

2. Eksperimen 2

Hasil pada eksperimen 2 merupakan hasil yang didapatkan dari analisis dengan menggunakan Naïve Bayes dengan Cross Validation untuk mendapatkan score confusion matrix. Tabel hasil dari eksperimen 2 dapat dilihat pada tabel 8.

Tabel 8. Hasil Eksperimen 2

	True Positive	True Negative	Class Precision
Pred. positive	224	0	100.00%
Pred. negative	86	310	78.28%
Class recall	72.26%	100.00%	86.12%
Accuracy			

Didapatkan dari eksperimen 2 dengan menggunakan operator Cross Validation hasil dari accuracy sebesar 86.23%. untuk kelas positif hasil Precision sebesar 100%, Recall 78.28%, dan kelas negatif hasil Precision sebesar 72.26%, Recall 100%.

4. KESIMPULAN

Berdasarkan hasil pengujian pada penelitian ini, dapat disimpulkan bahwa Dari jumlah data yang diperoleh, yaitu sebanyak 359 dengan labeling manual dibantu oleh ahli bahasa bahwa algoritma Naïve Bayes cocok digunakan dalam analisis sentimen pada topik yang peneliti angkat. Didapatkan nilai accuracy, precision, dan recall yang telah didapatkan pada penelitian kali ini seperti beberapa poin yaitu, Nilai accuracy dari Naïve Bayes Eksperimen 1 sebesar 86.29%, didapatkan untuk nilai precision sebesar 78.48%, nilai recall sebesar 72.58% dan untuk accuracy negatif sebesar 13.71%. Nilai accuracy dari Naïve Bayes dengan menggunakan cross validation Eksperimen 2 sebesar 86.12%, didapatkan untuk nilai precision sebesar 78.28%, dan nilai recall sebesar 72.26% dan untuk ulasan negatif sebesar 13.88%. Hasil penelitian menunjukkan bahwa Eksperimen 1 memberikan hasil yang lebih baik daripada Eksperimen 2, sehingga dapat disimpulkan bahwa pembagian data secara manual menghasilkan performa yang lebih baik dibandingkan dengan cross validation. Sebelumnya, peneliti juga menemukan kesamaan dengan penelitian yang dilakukan oleh Humairah [10] mengenai tingkat kepuasan produk di toko online Rubylicious, yang menunjukkan tingkat akurasi sebesar 82.93%. Namun, dalam penelitian ini Eksperimen 1 dan Eksperimen 2 berhasil mencapai tingkat akurasi yang lebih tinggi, yaitu masing-masing sebesar 86.29% dan 86.12%. Secara keseluruhan, hasil penelitian menegaskan bahwa algoritma Naïve Bayes merupakan pilihan yang tepat untuk analisis sentimen pada topik yang diteliti, dan hasilnya menunjukkan tingkat akurasi yang memuaskan.

REFERENCES

- [1] A. I. Tanggraeni and M. N. N. Sitokdana, "Analisis Sentimen Aplikasi E-Government pada Google Play Menggunakan Algoritma Naïve Bayes," *JATISI (Jurnal Tek. Inform. dan Sist. Informasi)*, vol. 9, no. 2, pp. 785–795, 2022, doi: 10.35957/jatisi.v9i2.1835.
- [2] A. A. Arifiyanti and E. D. Wahyuni, "Smote: Metode Penyeimbang Kelas Pada Klasifikasi Data Mining," *SCAN - J.*

- Teknol. Inf. dan Komun., vol. 15, no. 1, pp. 34–39, 2020, doi: 10.33005/scan.v15i1.1850.
- [3] C. Cahyaningtyas, Y. Nataliani, and I. R. Widiyari, “Analisis Sentimen Pada Rating Aplikasi Shopee Menggunakan Metode Decision Tree Berbasis SMOTE,” *Aiti*, vol. 18, no. 2, pp. 173–184, 2021, doi: 10.24246/aiti.v18i2.173-184.
- [4] P. S. Dewi, C. K. Sastradipraja, and D. Gustian, “Sistem Pendukung Keputusan Kenaikan Jabatan Menggunakan Metode Algoritma Naïve Bayes Classifier,” *J. Teknol. dan Inf.*, vol. 11, no. 1, pp. 66–80, 2021, doi: 10.34010/jati.v11i1.3593.
- [5] S. Febriani and H. Sulistiani, “Analisis Data Hasil Diagnosa Untuk Klasifikasi Gangguan Kepribadian Menggunakan Algoritma C4.5,” *89Jurnal Teknol. dan Sist. Inf.*, vol. 2, no. 4, pp. 89–95, 2021.
- [6] I. H. Witten, E. Frank, M. A. Hall, and C. J. Pal, “Data Mining: Practical Machine Learning Tools and Techniques,” *Data Min. Pract. Mach. Learn. Tools Tech.*, pp. 1–621, 2016.
- [7] W. Parasati, F. A. Bachtiar, and N. Y. Setiawan, “Analisis Sentimen Berbasis Aspek pada Ulasan Pelanggan Restoran Bakso President Malang dengan Metode Naïve Bayes Classifier,” vol. 4, no. 4, 2020.
- [8] N. Saurina, T. Rahayuningsih, and L. Retnawati, “Analisis Sentimen Ulasan Pelanggan Batik Ecoprint Menggunakan Naïve Bayes Dan KNN Classifier,” vol. 9, no. 2, 2022.
- [9] F. N. Hasan and M. Dwijayanti, “Analisis Sentimen Ulasan Pelanggan Terhadap Layanan Grab Indonesia Menggunakan Multinomial Naïve Bayes Classifier,” *J. Linguist. Komputasional*, vol. 4, no. 2, pp. 52–58, 2021, doi: <https://doi.org/10.26418/jlk.v4i2.61>.
- [10] I. Darmawan, O. N. Pratiwi, F. R. Industri, and U. Telkom, “Analisis Sentimen Ulasan Produk Toko Online Rubylicious Untuk,” vol. 7, no. 2, pp. 7026–7034, 2020.
- [11] D. Normawati and S. A. Prayogi, “Implementasi Naïve Bayes Classifier Dan Confusion Matrix Pada Analisis Sentimen Berbasis Teks Pada Twitter,” *J. Sains Komput. Inform. (J-SAKTI)*, vol. 5, no. 2, pp. 697–711, 2021.
- [12] A. Surahman, A. F. Octaviansyah, and D. Darwis, “Ekstraksi Data Produk E-Marketplace Sebagai Strategi Pengolahan Segmentasi Pasar Menggunakan Web Crawler,” *Sistemasi*, vol. 9, no. 1, p. 73, 2020, doi: 10.32520/stmsi.v9i1.580.
- [13] O. I. Gifari, M. Adha, F. Freddy, and F. F. S. Durrand, “Analisis Sentimen Review Film Menggunakan TF-IDF dan Support Vector Machine,” *J. Inf. Technol.*, vol. 2, no. 1, pp. 36–40, 2022, doi: 10.46229/jifotech.v2i1.330.
- [14] A. Aziz, “Analisis Sentimen Identifikasi Opini Terhadap Produk, Layanan dan Kebijakan Perusahaan Menggunakan Algoritma TF-IDF dan SentiStrength,” *J. Sains Komput. Inform. (J-SAKTI)*, vol. 6, no. 1, p. 115, 2022.
- [15] 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.
- [16] F. Ratnawati, “Implementasi Algoritma Naive Bayes Terhadap Analisis Sentimen Opini Film Pada Twitter,” *INOVTEK Polbeng - Seri Inform.*, vol. 3, no. 1, p. 50, 2018, doi: 10.35314/isi.v3i1.335.
- [17] S. Mujahidin, B. Prasetio, and M. C. C. Utomo, “Implementasi Analisis Sentimen Masyarakat Mengenai Kenaikan Harga BBM Pada Komentar Youtube Dengan Metode Gaussian naïve bayes,” *Voteteknika (Vocational Tek. Elektron. dan Inform.)*, vol. 10, no. 3, p. 17, 2022, doi: 10.24036/voteteknika.v10i3.118299.
- [18] H. F. Putro, R. T. Vuldari, and W. L. Y. Saptomo, “Penerapan Metode Naive Bayes Untuk Klasifikasi Pelanggan,” *J. Teknol. Inf. dan Komun.*, vol. 8, no. 2, 2020, doi: 10.30646/tikomsin.v8i2.500.
- [19] D. Darwis, N. Siskawati, and Z. Abidin, “Penerapan Algoritma Naive Bayes Untuk Analisis Sentimen Review Data Twitter Bmkg Nasional,” *J. Tekno Kompak*, vol. 15, no. 1, p. 131, 2021, doi: 10.33365/jtk.v15i1.744.
- [20] M. Hofmann and R. Klinkenberg, *Data Mining and Knowledge Discovery Series Edited by*. 2014.
- [21] T. Wiratama Putra, A. Triayudi, and A. Andrianingsih, “Analisis Sentimen Pembelajaran Daring Menggunakan Metode Naïve Bayes, KNN, dan Decision Tree,” *J. JTIK (Jurnal Teknol. Inf. dan Komunikasi)*, vol. 6, no. 1, pp. 20–26, 2022, doi: 10.35870/jtik.v6i1.368.
- [22] RAVI SHAH, “Guide to Tokenizers and Preprocessing,” Kaggle, 2022. <https://www.kaggle.com/competitions/feedback-prize-english-language-learning/overview> (accessed Aug. 05, 2023).
- [23] AMED, “Indonesian stemmer / stopwords,” Kaggle, 2021. <https://www.kaggle.com/competitions/shopee-product-matching/overview> (accessed Aug. 22, 2023).