

Analisis Sentimen Produk Makanan Jepang Di Indonesia Pada *Twitter* Menggunakan *Naïve Bayes*

Alingga Reandito Ikhwan Sumantri^{1*}, Muhamad Fatchan², Tri Ngudi Wiyatno³

Teknik Informatika, Universitas Pelita Bangsa, Bekasi, Indonesia

*e-mail *Corresponding Author*: alinggareandito3@gmail.com

Abstract

This research aims to analyze public sentiment towards Japanese food products in Indonesia using the Naïve Bayes Classifier method. Data was obtained from Twitter through a crawling process, resulting in 214 tweets analyzed for positive and negative sentiment. The methodology used includes data collection with Python and Google Collaboratory, as well as the application of the Naïve Bayes algorithm. The results showed that the algorithm achieved an accuracy of 77.03%, with precision for positive sentiment of 84.87% and negative sentiment of 58.06%, and recall of 83.23%. In conclusion, public sentiment towards Japanese food products in Indonesia tends to be positive, and the Naïve Bayes method proved to be reliable in this analysis.

Keywords: *Data Mining; Sentiment Analysis; Naïve Bayes; Japanese Food Products; Twitter*

Abstrak

Penelitian ini bertujuan untuk menganalisis sentimen masyarakat terhadap produk makanan khas Jepang di Indonesia menggunakan metode *Naïve Bayes Classifier*. Data diperoleh dari *Twitter* melalui proses *crawling*, menghasilkan 214 tweet yang dianalisis untuk sentimen positif dan negatif. Metodologi yang digunakan meliputi pengumpulan data dengan *Python* dan *Google Colaboratory*, serta penerapan algoritma *Naïve Bayes*. Hasil penelitian menunjukkan bahwa algoritma ini mencapai akurasi sebesar 77,03%, dengan presisi untuk sentimen positif 84,87% dan negatif 58,06%, serta recall 83,23%. Kesimpulannya, sentimen masyarakat terhadap produk makanan khas Jepang di Indonesia cenderung positif, dan metode *Naïve Bayes* terbukti andal dalam analisis ini.

Kata kunci: *Data Mining; Analisis Sentimen; Naïve Bayes; Produk Makanan Jepang; Twitter*

1. Pendahuluan

Media sosial telah berkembang menjadi salah satu platform utama bagi pengguna dalam era digital saat ini untuk berbagi pengalaman dan pendapat tentang berbagai hal, termasuk makanan. Sentimen pengguna terhadap produk makanan yang dibagikan di media sosial sangat berharga bagi produsen dalam memahami kepuasan pelanggan, mengidentifikasi masalah produk, dan merumuskan strategi pemasaran [1]. Namun, jumlah data yang sangat besar membuat analisis manual menjadi tidak efisien, sehingga diperlukan pendekatan otomatis. Oleh karena itu, pentingnya penelitian ini terletak pada upaya untuk menemukan metode yang efektif dan efisien dalam mengkategorikan persepsi konsumen terhadap produk makanan di media sosial, terutama *Twitter* [2].

Saat ini, produk makanan Jepang di Indonesia semakin populer dan banyak dibicarakan di media sosial, terutama di *platform Twitter*. Namun, ada masalah signifikan yang muncul, yaitu ketidakseimbangan dalam jumlah sentimen positif dan negatif pada data yang tersedia [3]. Ketidakseimbangan ini dapat mempengaruhi kinerja model klasifikasi yang digunakan, yang dapat menyebabkan bias dalam hasil analisis. Selain itu, banyak opini pengguna mengenai produk makanan Jepang di Indonesia yang belum terklasifikasi dengan baik, sehingga informasi yang diperoleh kurang optimal dalam memberikan wawasan bagi produsen makanan [4].

Untuk mengatasi masalah ini, studi ini menawarkan solusi untuk metode analisis sentimen *Naive Bayes*, yang merupakan algoritma klasifikasi yang telah terbukti efektif dalam banyak aplikasi analisis sentimen, termasuk klasifikasi teks yang bersifat positif, negatif, atau netral [5]. Penelitian sebelumnya mendukung metode ini, yang menunjukkan bahwa itu dapat menyelesaikan masalah klasifikasi dengan data teks yang tidak seimbang. Akibatnya, metode

Naive Bayes dianggap sebagai solusi yang tepat untuk masalah klasifikasi sentimen pada produk makanan Jepang di Indonesia, karena kemampuannya dalam mengelola data dengan tingkat keakuratan yang tinggi [6].

Tujuan dari penelitian ini adalah untuk mengembangkan dan mengevaluasi seberapa efektif metode *Naive Bayes* menggunakan media sosial *Twitter* untuk mengklasifikasikan persepsi pengguna terhadap produk makanan Jepang di Indonesia. Diharapkan manfaat dari penelitian ini dapat membantu produsen makanan meningkatkan kualitas produk dan membuat metode pemasaran yang lebih efisien.

2. Tinjauan Pustaka

Penelitian yang dilakukan Farah Syadza Mufidah, Sri Winarno, Farikh Al Zami, Erika Devi Udayanti, Ramadhan Rakhmat Sani, (2022) yang berjudul “Analisis Sentimen Masyarakat terhadap Layanan *Shopeefood* Melalui Media Sosial *Twitter* dengan Algoritma *Naive Bayes Classifier*” [7]. Hasil penelitian menunjukkan bahwa metode *naive bayes* dapat dengan akurat mengklasifikasikan perasaan positif dan negatif dari data teks; data pelatihan menunjukkan kinerja yang baik dengan tingkat akurasi sebesar 98,17%, tingkat akurasi sebesar 97,68%, dan tingkat akurasi sebesar 93,75%. Nilai perhitungan AUC sebesar 0.97 juga menunjukkan kinerja yang baik.

Penelitian yang dilakukan Muhammad Rafly Al Fattah Zain, Mia Kamayani, (2023) yang berjudul “Analisis Sentimen Ulasan Pelanggan Online Ubi Madu Cilembu Abah Nana Menggunakan Algoritma *Naive Bayes*” [8]. Hasil penelitian ini menunjukkan bahwa algoritma *Naive Bayes* berfungsi dengan baik untuk menganalisis sentimen topik yang dibahas. Dengan 359 data yang dilabeli secara manual, diperoleh hasil Eksperimen 1 dengan nilai akurasi 86,29%, *precision* 78,48%, *recall* 72,58%, dan akurasi negatif 13,71%. Sementara itu, Eksperimen 2 dengan *cross-validation* memberikan akurasi 86,12%, *precision* 78,28%, *recall* 72,26%, dan ulasan negatif 13,88%. Kesimpulannya, Pembagian data manual bekerja lebih baik daripada *cross-validation*.

Pada Penelitian yang dilakukan Hajaroh, Tati Suprpti, Riri Narasati, (2024) yang berjudul “Implementasi Algoritma *Naive Bayes* Untuk Analisis Sentimen Ulasan Produk Makanan Dan Minuman di Tokopedia” [9]. Studi ini menunjukkan bahwa Algoritma *Naive Bayes* memberikan akurasi sebesar 75,90%, presisi sebesar 66,45%, dan *recall* sebesar 97,42% dalam analisis sentimen ulasan produk makanan dan minuman di Tokopedia. Hasil ini menggambarkan mayoritas sentimen positif pada ulasan dan memberikan wawasan penting bagi pemilik bisnis untuk memperbaiki aspek dengan sentimen negatif serta mempertahankan yang positif. Meskipun akurasinya cukup baik, presisi yang lebih rendah menunjukkan perlunya peningkatan dalam klasifikasi ulasan, terutama untuk sentimen negatif dan netral.

Penelitian yang dilakukan Sunneng Sandino Berutu, (2022) yang berjudul “*Text Mining* dan Klasifikasi Sentimen Berbasis *Naive Bayes* Pada Opini Masyarakat terhadap Makanan Tradisional” [10]. Seperti yang ditunjukkan oleh analisis sentimen masyarakat terhadap makanan tradisional internasional dan tradisional yang diambil dari data *Twitter*, gudeg menerima sentuhan positif tertinggi (57,9%), sedangkan rendang menerima sentuhan negatif tertinggi (21,9%). Hamburger juga menerima sentuhan netral tertinggi (60,9%). Studi ini membuat enam model klasifikasi sentimen berbasis *Naive Bayes* untuk setiap jenis makanan, yang diuji menggunakan matriks confusion. Dalam model, kumpulan data hamburger menunjukkan nilai f1 tertinggi untuk variabel ketepatan, ketepatan, dan *recall*. Di sisi lain, kumpulan data rendang dan sate menunjukkan nilai f1-tertinggi untuk variabel ketepatan dan *recall*. Model pizza memiliki nilai f1-terendah pada semua parameter.

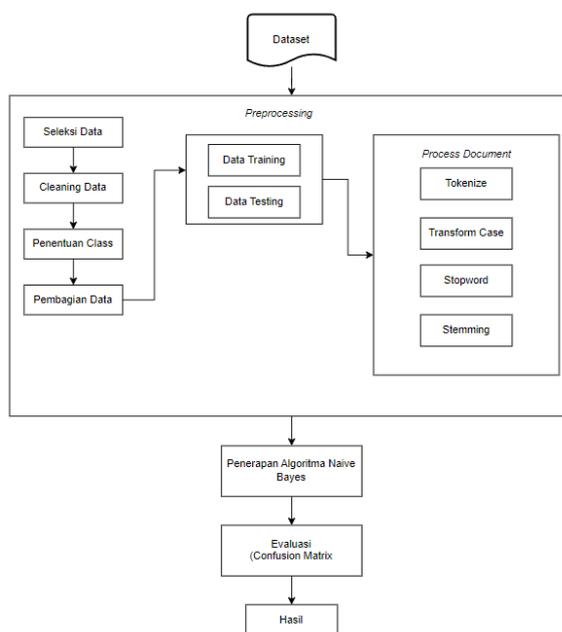
Penelitian yang dilakukan Eka Andrian, Aulia Rahman Isnain (2024) yang berjudul “Analisis Sentimen Masyarakat Terhadap Tiktok Shop di *Twitter* Menggunakan Metode *Naive Bayes Classifier*” [11]. Penelitian menggunakan algoritma *Naive Bayes Classifier* terhadap 4285 data menghasilkan analisis sentimen dengan akurasi 78.22%, presisi 86%, dan *recall* 84%. Dari 3816 data yang diolah, 53.45% menunjukkan sentimen positif dan 46.55% negatif. Hasil ini dapat dijadikan dasar untuk menyusun kebijakan dan strategi yang lebih baik bagi TikTok Shop serta memperbaiki persepsi masyarakat terhadap fitur-fiturnya.

Penelitian ini terletak pada penerapan algoritma *Naive Bayes* yang dikombinasikan dengan teknik pra-pemrosesan data yang lebih komprehensif dan spesifik terhadap konteks budaya, yaitu analisis sentimen terhadap produk makanan Jepang di Indonesia melalui platform

Twitter. Berbeda dengan penelitian sebelumnya, studi ini menangani masalah ketidakseimbangan kelas dengan menerapkan metode *resampling* untuk memastikan distribusi data yang lebih seimbang, sehingga meningkatkan akurasi klasifikasi. Selain itu, penelitian ini juga mengintegrasikan *lexicon* khusus bahasa Indonesia yang telah disesuaikan untuk menangani slang dan istilah populer yang sering muncul di media sosial, meningkatkan ketepatan dalam pemahaman konteks dan nuansa bahasa. Implementasi ini diharapkan dapat memberikan kontribusi baru dalam bidang analisis sentimen dengan menyediakan model yang lebih adaptif dan akurat dalam memahami opini publik terhadap produk makanan internasional dalam konteks lokal.

3. Metodologi

Penelitian ini akan melalui beberapa proses, Gambar 1 menunjukkan urutan proses penelitian ini:



Gambar 1. Metodologi Penelitian

3.1. Crawling Data

Crawling merupakan proses pengambilan data yang cepat dan efisien ke dalam penyimpanan lokal dan mengidentifikasi berdasarkan kata kunci yang kita cari [12]. Data merupakan data text berbahasa Indonesia yang diambil dari akun media sosial *twitter* dengan *keyword* "makanan jepang". Setelah memasukkan *keyword* kita memasukkan jumlah *tweet* yang akan kita *crawling*, disini peneliti mengambil sebanyak 492 data *tweet*.

3.2. Preprocessing

Proses *preprocessing* data dimulai pada tahap ini. Ini termasuk perbaikan, normalisasi, *stopwords*, tokenisasi, dan stemming untuk mempersiapkan data untuk analisis. [13].

1) Cleaning Data

Terdapat beberapa komponen yang khas yang biasanya terdapat pada *tweet* antara lain yaitu *Username* (Nama Pengguna), URL (*Uniform Resource Locator*), dan RT (*Retweet*). Setelah melakukan tahapan *select* data kemudian data akan diproses pada tahapan *cleaning* data atau dibersihkan dari data-data yang tidak digunakan untuk mempermudah pada saat proses melakukan sentimen. Adapun tahapan-tahapan *cleaning* data yang dilakukan yaitu menghilangkan RT (*Retweet*), URL (*Uniform Resource Locator*), nama pengguna (*Username*), menghilangkan hastag (#), menghilangkan simbol(!~?.,,:."#*%@%\$&()-) [14].

2) Stopwords

Stopwords dibuat untuk menghapus kata-kata yang tidak memiliki arti atau kata-kata yang tidak terikat dengan kata sifat yang berkaitan dengan data *stopwords* [15].

3) Tokenize

Tokenize merupakan pemotongan kata yang dibuat berdasarkan tiap kata dan menyusunnya menjadi potongan tunggal. Kata-kata dalam dokumen tersebut dipisah oleh spasi [16].

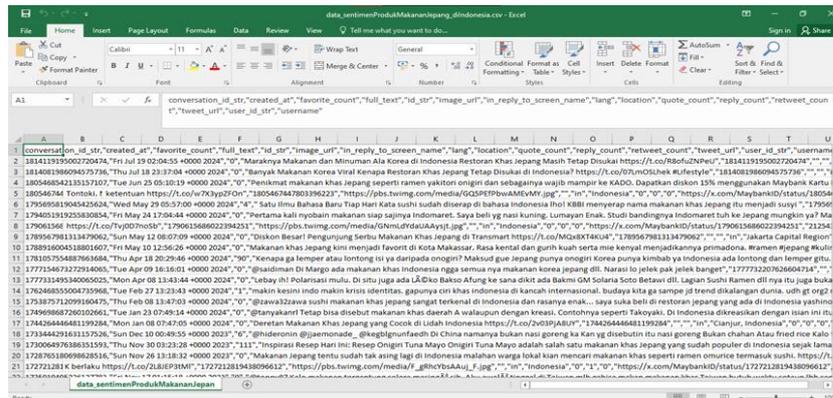
4) Stemming

Proses menghilangkan sufiks, prefix, dan konfiks dari dokumen bahasa Indonesia dikenal sebagai *stemming*. Dalam proses ini, aturan tertentu digunakan untuk mengembalikan kata-kata dalam dokumen ke kata akarnya [17].

4. Hasil dan Pembahasan

4.1. Hasil Crawling Data

Pada titik ini, akan dijelaskan bagaimana crawling data dilakukan di *Twitter*. Dalam penelitian ini, penulis menggunakan *Google Colaboratory* dan bahasa pemrograman *Python* dengan *library Tweet Harvest* untuk melakukan *crawling data*. Alasan penggunaan alat-alat ini adalah karena pada bulan Februari 2023, *Twitter* mengumumkan penghentian akses gratis ke API *Twitter* dan memperkenalkan model berlangganan berbayar. Perubahan ini memengaruhi banyak aplikasi dan layanan pihak ketiga yang menggunakan API *Twitter*, termasuk *RapidMiner*. Akibatnya, *RapidMiner* menghapus operator "*Search Twitter*" yang sangat penting bagi peneliti sentimen. Sebagai solusi, penulis mencari alternatif lain untuk tetap mengakses data dari API *Twitter* dengan menggunakan *library Tweet Harvest* dalam bahasa pemrograman *Python*. Berikut ini adalah gambar yang menunjukkan proses dan hasil penarikan data yang diperoleh.



Gambar 2. Tampilan Hasil Crawling Pada Excel

4.2. Hasil Preprocessing

Preprocessing dilakukan sebelum pemberian label sentimen pada data. Pada tahap ini, data dipersiapkan untuk memastikan kualitas dan kebersihannya. Proses ini mencakup pembersihan dari informasi yang tidak relevan, penghapusan *missing value*, dan penghilangan data yang duplikat. Langkah-langkah ini penting untuk meningkatkan akurasi dan efektivitas analisis sentimen. Dengan melakukan *preprocessing* secara menyeluruh, diharapkan data yang dihasilkan akan lebih konsisten dan representatif, sehingga analisis dapat menghasilkan hasil yang lebih akurat dan menguntungkan. Berikut adalah proses *preprocessing* yang diterapkan dalam penelitian ini:

1) Data Selection

Pada tahap ini, dipilih data yang akan digunakan. Proses ini diperlukan karena hasil *crawling* dari *Twitter* menghasilkan sejumlah kolom seperti "*conversation_id_str*", "*created_at*", "*favorite_count*", "*image_url*", dan kolom lainnya yang tidak relevan untuk pengolahan data. Seleksi ini bertujuan untuk menyederhanakan proses pengolahan dan memastikan bahwa hanya data yang penting yang digunakan, sehingga membuat proses menjadi lebih terstruktur dan

efisien. Dalam penelitian ini, hanya kolom yang berisi teks, yaitu kolom "full_text", yang akan digunakan.

2) Data Cleaning

Pada tahap ini, proses pembersihan teks data dilakukan. Proses *cleaning* ini bertujuan untuk menghapus elemen-elemen yang tidak relevan dari isi *tweet*, seperti simbol-simbol, karakter, atau angka yang tidak berguna. Langkah ini penting untuk memastikan bahwa data yang diolah bersih dan siap untuk analisis lebih lanjut. Proses ini dilakukan di *RapidMiner*, di mana berbagai teknik pembersihan diterapkan untuk meningkatkan kualitas data.

3) Remove Duplicates

Tahapan ini bertujuan untuk menghapus data yang duplikat dan *missing values* setelah proses *cleaning* dilakukan. Dalam tahap ini, digunakan dua operator utama: "*Remove Missing Value*" untuk menghapus entri dengan nilai yang hilang, dan "*Remove Duplicates*" untuk menghapus data duplikat. Proses ini sangat penting untuk memastikan integritas dan keakuratan *dataset* sebelum melanjutkan ke analisis berikutnya. Dengan membersihkan data dari duplikasi dan nilai yang hilang, diharapkan analisis yang dilakukan akan lebih valid.

Setelah menyelesaikan semua langkah *preprocessing*, data yang telah diproses kemudian disimpan dalam *file* berformat *.csv dengan tipe *encoding* "utf-8". Penyimpanan ini mempersiapkan data untuk tahap selanjutnya, yaitu labelisasi sentimen menggunakan metode *lexicon-based* yang berdasarkan pada kamus Bahasa Indonesia. Proses ini memastikan bahwa data yang dianalisis telah melalui pembersihan dan pengolahan yang memadai, sehingga analisis sentimen dapat dilakukan dengan akurat.

Row No.	full_text
1	Maraknya Makanan dan Minuman Ala Korea di Indonesia Restoran Khas Jepang Masih Tetap Disukai
2	Penikmat makanan khas Jepang seperti ramen yakitori onigiri dan sebagainya wajib mampir ke KADO. Dapatkan diskon 15 menggunakan Maybank Kartu ...
3	Nikmati makanan khas Jepang favoritmu dengan berkunjung ke Sakana amp Tontoki. Makin hemat dengan diskon 15 menggunakan Maybank Kartu Kredit...
4	Satu Ilmu Bahasa Baru Tiap Hari Kata sushi sudah diserap di bahasa Indonesia lho KBBI menyerap nama makanan khas Jepang itu menjadi susyi
5	Pertama kali nyobain makanan siap sajinya Indomaret. Saya beli yg nasi kuning. Lumayan Enak. Studi bandingnya Indomaret tuh ke Jepang mungkin ya M...
6	Rasakan kenikmatan bersantap makanan khas Jepang di AOKI Japanese bersama rekan bisnis/keluarga. Nikmati diskon 15 menggunakan Maybank Kart...
7	Diskon Besar Pengunjung Serbu Makanan Khas Jepang di Transmart
8	Kenapa ga lempur atau lontong isi ya daripada onogiri Maksud gue Jepang punya onogiri Korea punya kimbab ya Indonesia ada lontong dan lempur gitu ...
9	Di Margo ada makanan khas Indonesia ngga semua nya makanan korea jepang dll. Narasi lo jelek pak jelek banget
10	Lebay ih Potarisasi mulu. Di situ juga ada L&K Bakso Afung ke sana dikit ada Bakmi GM Solaria Soto Betawi dll. Lagian Sushi Ramen dll nya itu juga buk...
11	makin kesini indo makin krisis identitas. gapunya ori khas Indonesia di kancah internasional. budaya kita ga sampe jd trend dikalangan dunia. udh gt org2...
12	sushi makanan khas jepang sangat terkenal di Indonesia dan rasanya enak... saya suka beli di restoran jepang yang ada di Indonesia yashinoya
13	Tetap bisa disebut makanan khas daerah A walaupun dengan kreasi. Contohnya seperti Takoyaki. Di Indonesia dikreasikan dengan isian ini itu bisa dibila...
14	Deretan Makanan Khas Jepang yang Cocok di Lidah Indonesia
15	Di China namanya bukan nasi goreng ka Kan yg disebutin itu nasi goreng Bukan chahan Atau fried rice Kalo fried rice di claim dari Indonesia ya salah. Sus...

Gambar 3. Hasil Data Preprocessing

4.3. Sentiment Labeling

Setelah *dataset* mentah dari hasil *crawling* di *Twitter* selesai melalui proses pembersihan, langkah berikutnya adalah melakukan labelisasi sentimen atau penentuan atribut kelas pada data. Dalam penelitian ini, labelisasi sentimen akan dilakukan menggunakan teknik *Lexicon-Based* yang mengacu pada kamus Bahasa Indonesia. Teknik ini memanfaatkan kamus atau daftar kata *lexicon* yang telah diberi label sentimen seperti positif, negatif, atau netral. Setiap kata dalam teks yang dianalisis akan dicocokkan dengan kata-kata dalam *lexicon* tersebut. Untuk mengetahui skor sentimen teks, setiap kata dalam teks dijumlahkan. Misalnya, jika teks mengandung lebih banyak kata dengan label sentimen positif dibandingkan negatif, teks tersebut akan diberi label sentimen positif. Penelitian ini tidak memasukkan label netral karena dianggap tidak memberikan informasi yang berarti. Proses ini akan dilakukan dengan bantuan *Google Colaboratory*, yang akan digunakan sebagai modul untuk menjalankan kode program berbasis *lexicon* ini.

Row No.	full_text	sentiment
1	Maraknya Makanan dan Minuman Ala Korea di Indonesia Restoran Khas Jepang Masih Tetap Disukai	?
2	Penikmat makanan khas Jepang seperti ramen yakitori onigiri dan sebagainya wajib mampir ke KADO Dapatkan diskon menggunakan ...	positive
3	Nikmati makanan khas Jepang favoritmu dengan berkunjung ke Sakana amp Tontoki Makin hemat dengan diskon menggunakan Mayba...	positive
4	Satu Ilmu Bahasa Baru Tiap Hari Kata sushi sudah diserap di bahasa Indonesia lho KBBI menyerap nama makanan khas Jepang itu me...	negative
5	Pertama kali nyobain makanan siap sajinnya Indomaret Saya beli yg nasi kuning Lumayan Enak Studi bandingnya Indomaret tuh ke Jepang...	positive
6	Rasakan kenikmatan bersantap makanan khas Jepang di AOKI Japanese bersama rekan bisniskeluarga Nikmati diskon menggunakan ...	positive
7	Diskon Besar Pengunjung Serbu Makanan Khas Jepang di Transmart	?
8	Kenapa ga lempem atau lontong isi ya daripada onogiri Maksud gue Jepang punya onogiri Korea punya kimbab ya Indonesia ada lontong ...	positive
9	Di Margo ada makanan khas Indonesia ngga semua nya makanan korea jepang di Narasi lo jetek pak jelek banget	?
10	Lebay ih Polarisasi mulu Di situ juga ada L&A&K Bakso Atung ke sana dikit ada Bakmi GM Solaria Soto Betawi dll Lagan Sushi Ramen...	negative
11	makin kesini indo makin kritis identitas gapunya ciri khas Indonesia di kancan internasional budaya kita ga sampe jd trend dikalangan d...	negative
12	sushi makanan khas jepang sangat terkenal di Indonesia dan rasanya enak: saya suka beli di restoran jepang yang ada di Indonesia yas...	?
13	Tetap bisa disebut makanan khas daerah A walaupun dengan kreasi Contohnya seperti Takoyaki Di Indonesia dikreasikan dengan isian ...	positive
14	Deretan Makanan Khas Jepang yang Cocok di Lidah Indonesia	negative
15	Di China namanya bukan nasi goreng ka Kan yg disebut itu nasi goreng Bukan chahan Atau fried rice Kalo fried rice di daim dari Indone...	?

Gambar 4. Hasil Dari Labelisasi Data

4.4. Pembobotan kata dengan TF-IDF

Pada tahapan ini, dilakukan pembobotan kata pada data teks. Pembobotan kata bertujuan untuk memberikan nilai atau bobot pada setiap kata dalam teks, berdasarkan kontribusinya dalam menentukan sentimen keseluruhan. Metode ini memungkinkan identifikasi kata-kata yang memiliki dampak signifikan terhadap sentimen positif atau negatif. Dalam proses ini, pembobotan diterapkan pada data yang telah diproses sebelumnya, tetapi label sentimen tidak disertakan, karena label tersebut tidak mempengaruhi hasil akhir dari pembobotan. Sebelum melanjutkan ke tahap pembobotan kata, *text preprocessing* harus dilakukan terlebih dahulu untuk memastikan bahwa teks yang digunakan bersih dan konsisten. Proses ini juga membantu dalam mengurangi *noise* dan variabilitas dalam data, sehingga hasil pembobotan kata menjadi lebih akurat dan relevan. Berikut adalah proses *text preprocessing* yang diterapkan dalam penelitian ini.

1) Tokenizing

Proses ini adalah bagian dari pemrosesan bahasa alami atau NLP (*Natural Language Processing*), yang melibatkan pemecahan teks menjadi unit-unit kecil yang dikenal sebagai "token". Teks mentah dapat diubah menjadi format yang lebih terorganisir dan dapat diproses lebih lanjut melalui proses ini, yang menjadikannya sangat penting, seperti dalam perhitungan frekuensi kata atau pembobotan kata. Dalam tahapan ini, operator "Tokenize" di *RapidMiner* digunakan untuk memecah teks menjadi token-token individual. Proses ini mempermudah analisis lebih lanjut dengan membuat data teks lebih terstruktur dan mudah dikelola. Berikut adalah hasil dari proses tokenisasi yang dilakukan.

Row No.	A	AKURATNE...	AOKI	AYAM	Abon	Ada	Adakah	Adalah	Adha
1	0	0	0	0	0	0	0	0	0
2	0	0	0	0	0	0	0	0	0
3	0	0	0	0	0	0	0	0	0
4	0	0	0	0	0	0	0	0	0
5	0	0	0	0	0	0	0	0	0
6	0	0	0.245	0	0	0	0	0	0
7	0	0	0	0	0	0	0	0	0
8	0	0	0	0	0	0	0	0	0
9	0	0	0	0	0	0	0	0	0
10	0	0	0	0	0	0	0	0	0
11	0	0	0	0	0	0	0	0	0
12	0	0	0	0	0	0	0	0	0
13	0.216	0	0	0	0	0	0	0	0
14	0	0	0	0	0	0	0	0	0

Gambar 5. Hasil Proses Tokenizing

2) Case Folding

Tujuan dari proses ini adalah untuk mengubah setiap huruf dalam teks menjadi huruf kecil, sehingga menghilangkan perbedaan antara huruf kapital dan huruf kecil. Tujuan utama dari proses ini adalah untuk menyederhanakan analisis teks dengan memastikan bahwa kata-kata yang sama diidentifikasi secara konsisten tanpa memperhatikan perbedaan kasus. Sebagai contoh, kata "Jepang" dan "jepang" akan dianggap berbeda dalam proses tokenisasi jika perbedaan kasus tidak dihapus. Dengan melakukan *case folding*, kedua kata tersebut akan

dianggap sebagai entitas yang sama, sehingga analisis menjadi lebih akurat. Proses ini dilakukan menggunakan operator “*Transform Cases*” pada *tools RapidMiner*, yang memudahkan penghapusan perbedaan huruf kapital dan huruf kecil. Ini adalah hasil dari proses penyiapan case yang digunakan.

Row No.	a	abon	acara	ada	adakah	adalah	adanya	adha	afdol
1	0	0	0	0	0	0	0	0	0
2	0	0	0	0	0	0	0	0	0
3	0	0	0	0	0	0	0	0	0
4	0	0	0	0	0	0	0	0	0
5	0	0	0	0	0	0	0	0	0
6	0	0	0	0	0	0	0	0	0
7	0	0	0	0	0	0	0	0	0
8	0	0	0	0.059	0	0	0	0	0
9	0	0	0	0.093	0	0	0	0	0
10	0	0	0	0.125	0	0	0	0	0
11	0	0	0	0	0	0	0	0	0
12	0	0	0	0.147	0	0	0	0	0
13	0.216	0	0	0	0	0.114	0	0	0
14	0	0	0	0	0	0	0	0	0

Gambar 6. Hasil Proses *Case Folding*

3) *Stopword Removal*

Penghapusan kata-kata umum atau "*stopwords*" seperti "dan", "atau", "adalah", dan sebagainya diperlukan dalam tahapan ini karena mereka tidak memberikan informasi yang penting untuk analisis sentimen. Tujuan dari proses ini adalah untuk mengurangi jumlah kata yang perlu dianalisis serta mengurangi kebisingan dalam data, sehingga analisis dapat difokuskan pada kata-kata yang lebih relevan dan signifikan. Karena *RapidMiner* tidak menyediakan operator *Filter Stopword* untuk bahasa Indonesia secara *default*, penulis menggunakan operator *Filter Stopword (Dictionary)* dengan referensi dari file Indonesian Stoplist <https://www.kaggle.com/datasets/oswinrh/indonesian-stoplist>. Proses ini membantu memastikan bahwa kata-kata yang tidak relevan tidak mempengaruhi hasil analisis.

Word	Attribut...	Total Occurences	Document Occurences
a	a	2	2
abon	abon	1	1
acara	acara	19	19
adakah	adakah	2	2
adha	adha	2	2
afdol	afdol	1	1
afung	afung	1	1
agustus	agustus	1	1
ahli	ahli	1	1
ahmad	ahmad	1	1
aja	aja	3	3
ajak	ajak	2	2
ajang	ajang	3	2
akrab	akrab	4	4

Gambar 7. Setelah Melakukan *Stopword*

4) *Filter Tokens*

Dalam proses preprocessing teks penelitian ini, operator tambahan "*Filter Tokens (by Length)*" digunakan. Berdasarkan panjangnya, operator ini berguna untuk menghapus kata-kata, dengan menentukan batas minimum dan maksimum jumlah karakter yang diizinkan. Tujuan dari proses ini adalah untuk menghilangkan kata-kata singkatan seperti “yg”, “dri”, dan “sbg”, serta kata-kata lain yang dianggap terlalu pendek atau panjang. Dengan menetapkan batas minimal 4 karakter dan batas maksimal 25 karakter, proses ini memastikan bahwa hanya kata-kata yang cukup informatif dan baik yang tersisa untuk analisis. Hal ini membantu meningkatkan kualitas

data dengan menghilangkan elemen yang tidak signifikan yang dapat mengganggu hasil analisis. Dengan demikian, teks yang tersisa akan lebih bersih dan relevan untuk analisis lebih lanjut. Berikut adalah pengaturan parameter yang diterapkan pada operator "*Filter Tokens (by Length)*" serta hasil yang diperoleh dari proses ini.

Word	Attribut...	Total Occurrences	Document Occurrences
abon	abon	1	1
acara	acara	19	19
adakah	adakah	2	2
adha	adha	2	2
afdol	afdol	1	1
atung	atung	1	1
agustus	agustus	1	1
ahli	ahli	1	1
ahmad	ahmad	1	1
ajak	ajak	2	2
ajang	ajang	3	2
akrab	akrab	4	4
akultura...	akultura...	1	1
akun	akun	3	2

Gambar 8. Hasil Dari Proses *Filter Tokens (by length)*

4.5. Implementasi dan Evaluasi Algoritma *Naive Bayes Classifier*

Tahapan ini adalah bagian inti dari penelitian, yaitu tahapan *deployment* atau implementasi metode klasifikasi *Naive Bayes*. Proses ini bertujuan untuk mengaplikasikan metode tersebut guna menarik kesimpulan atau memberikan wawasan dari data yang telah dianalisis. Fokus utama dari tahapan ini adalah untuk membangun model *Machine Learning* yang dapat menentukan apakah sentimen dalam data bersifat positif atau negatif. Selain itu, dalam langkah ini, perhitungan untuk menilai kinerja model yang dikembangkan termasuk metrik akurasi, presisi, dan *recall*. Penilaian ini dilakukan menggunakan teknik *K-Fold Cross Validation*, yang membantu memastikan bahwa model tidak hanya berkinerja baik pada data pelatihan tetapi juga pada data yang tidak terlihat sebelumnya. Dengan begitu, model yang dibangun diharapkan dapat memberikan hasil yang lebih andal dan *generalizable*. Proses ini juga memungkinkan penyesuaian dan perbaikan model berdasarkan hasil evaluasi untuk mencapai performa yang optimal.

1) *Accuracy*

Accuracy mengacu pada seberapa dekat hasil prediksi dengan hasil fakta. Prediksi mengacu pada seberapa tepat informasi pengguna dengan respons sistem. Nilai prediksi dihitung setelah data dibagi. Hasil pengujian prediksi data adalah sebagai berikut:

accuracy: 77.03% +/- 7.34% (micro average: 77.10%)

	true positive	true negative	class precision
pred. positive	129	23	84.87%
pred. negative	26	36	58.06%
class recall	83.23%	61.02%	

Gambar 9. Hasil Evaluasi Model Menggunakan *K-Fold Cross Validation*

Prediksi untuk data sentimen positif adalah 129 dan yang salah adalah 23, sedangkan prediksi untuk data sentimen negatif adalah 26 dan yang salah adalah 36. Dengan menggunakan pengujian model *Naive Bayes Classification*, hasil rata-rata keakuratan adalah 77,03%. Ini diperoleh dengan membagi jumlah data *True Positive* (TP) dan *True Negative* (TN) dengan total keseluruhan data. Berikut adalah cara menghitung keakuratan secara manual:

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN} \times 100\%$$

$$\begin{aligned}
 &= \frac{129 + 36}{129 + 36 + 23 + 26} \times 100\% \\
 &= \frac{165}{214} \times 100\% \\
 &= 77.03\%
 \end{aligned}$$

2) Precision

Precision adalah ukuran tingkat ketepatan informasi yang diberikan oleh sistem. Untuk menghasilkan hasil yang presisi, pengujian ini menghitung jumlah data Positif Benar (TP) dibagi dengan total data Positif Benar (TP) dan ditambah dengan *False Positive* (FP). Metode perhitungan precision manual adalah sebagai berikut:

$$\begin{aligned}
 Precision &= \frac{TP}{TP + FP} \times 100\% \\
 &= \frac{129}{129 + 23} \times 100\% \\
 &= \frac{129}{152} \times 100\% \\
 &= 84.87\%
 \end{aligned}$$

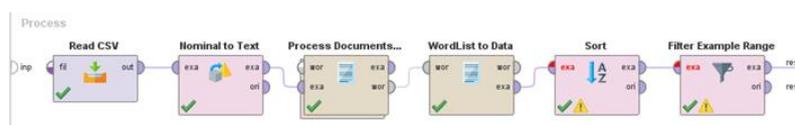
3) Recall

Recall adalah ukuran seberapa efektif pengumpulan data yang relevan. Nilai recall dari data tes akan dihitung pada pengujian ini. Data Positif Positif (TP) dibagi dengan total data Positif Positif (TP) ditambah dengan *False Negative* (FN) untuk mendapatkan hasil recall. Perhitungan recall manual dilakukan dengan cara ini.

$$\begin{aligned}
 Recall &= \frac{TP}{TP + FN} \times 100\% \\
 &= \frac{129}{129 + 26} \times 100\% \\
 &= \frac{129}{155} \times 100\% \\
 &= 83.23\%
 \end{aligned}$$

4.6. Visualisasi Word Cloud

Setelah mendapatkan hasil yang akurat dari pengujian model Klasifikasi *Naive Bayes*, Proses pembentukan *word cloud* dimulai untuk menghitung berapa kali setiap kata digunakan, visualisasi *cloud* kata menunjukkan variasi kata yang muncul pada data. Kata yang lebih banyak digunakan dalam teks yang lebih besar, dan kata yang lebih sedikit digunakan dalam teks yang lebih kecil. Untuk menampilkan visualisasi *word cloud* dapat dilakukan setelah melakukan *text preprocessing*, berikut adalah operator-operator yang dibutuhkan:



Gambar 10. Operator Untuk Visualisasi Wordcloud

Gambar diatas, memerlukan beberapa operator untuk menampilkan visualisasi *wordcloud*. Daftar kata dapat diubah menjadi format data yang dapat digunakan untuk analisis

Daftar Referensi

- [1] Nabila Zahra, "Analisis Media Monitoring: Sentimen Publik Terhadap Inovasi Produk Brand Indomie Pada Periode Maret 2024," *Harmoni: Jurnal Ilmu Komunikasi dan Sosial*, vol. 2, no. 2, pp. 157–171, May 2024, doi: 10.59581/harmoni-widyakarya.v2i2.3175.
- [2] T. Agustiranti, A. Khalfani Izzati Kurdiana, B. Al Ghiffari, E. Dwi Juniar, and D. Gita Purnama, "Penerapan Naive Bayes Terhadap Sentimen Analisis Media Sosial Twitter Pengguna Kereta Cepat Jakarta-Bandung (Whoosh)," *Jurnal Ilmu Komputer dan Sistem Informasi (JIKOMSI)*, vol. 7, no. 1, pp. 297–305, 2024.
- [3] M. Ulum, "Permintaan Produk Makanan Asal Jepang Meningkat," Feb. 2024. [Online]. Available: <https://surabaya.bisnis.com/read/20240205/532/1738390/permintaan-produk-makanan-asal-jepang-meningkat>
- [4] M. A. Rahman, H. Budianto, and E. I. Setiawan, "Aspect Based Sentimen Analysis Opini Publik Pada Instagram dengan Convolutional Neural Network," *INSYST: Journal of Intelligent System and Computation*, vol. 1, no. 2, pp. 50–57, 2019.
- [5] F. V. Sari and A. Wibowo, "Analisis sentimen pelanggan toko online Jd. Id menggunakan metode Naive Bayes Classifier berbasis konversi ikon emosi," *Simetris: Jurnal Teknik Mesin, Elektro dan Ilmu Komputer*, vol. 10, no. 2, pp. 681–686, 2019.
- [6] Y. A. Singgalen, "Analisis Sentimen dan Sistem Pendukung Keputusan Menginap di Hotel Menggunakan Metode CRISP-DM dan SAW," *Journal of Information System Research (JOSH)*, vol. 4, no. 4, pp. 1343–1353, Jul. 2023, doi: 10.47065/josh.v4i4.3917.
- [7] F. S. Mufidah, S. Winarno, F. Alzami, E. D. Udayanti, and R. R. Sani, "Analisis Sentimen Masyarakat Terhadap Layanan ShopeeFood Melalui Media Sosial Twitter Dengan Algoritma Naive Bayes Classifier," *JOINS (Journal of Information System)*, vol. 7, no. 1, pp. 14–25, May 2022, doi: 10.33633/joins.v7i1.5883.
- [8] M. R. A. F. Zain and M. Kamayani, "Analisis Sentimen Ulasan Pelanggan Online Ubi Madu Cilembu Abah Nana Menggunakan Algoritma Naive Bayes," *Jurnal Sistem Komputer dan Informatika (JSON)*, vol. 5, no. 1, pp. 11-21, Sep. 2023, doi: 10.30865/json.v5i1.6646.
- [9] H. Hajaroh, T. Suprpti, and R. Narasati, "Implementasi Algoritma Naive Bayes Untuk Analisis Sentimen Ulasan Produk Makanan Dan Minuman Di Tokopedia," *JATI (Jurnal Mahasiswa Teknik Informatika)*, vol. 8, no. 1, pp. 111–118, 2024.
- [10] S. S. Berutu, "Text Mining dan Klasifikasi Sentimen Berbasis Naive Bayes Pada Opini Masyarakat terhadap Makanan Tradisional," *Jurnal Sistem Komputer dan Informatika (JSON)*, vol. 4, no. 2, p. 254, Dec. 2022, doi: 10.30865/json.v4i2.5138.
- [11] E. Andrian and A. R. Isnain, "Analisis Sentimen Masyarakat Terhadap Tiktok Shop di Twitter Menggunakan Metode Naive Bayes Classifier," *JURNAL MEDIA INFORMATIKA BUDIDARMA*, vol. 8, no. 2, p. 788, Apr. 2024, doi: 10.30865/mib.v8i2.7530.
- [12] D. Darwis, N. Siskawati, and Z. Abidin, "Penerapan Algoritma Naive Bayes Untuk Analisis Sentimen Review Data Twitter Bmkg Nasional," *Jurnal Tekno Kompak*, vol. 15, no. 1, pp. 131–145, 2021.
- [13] S. Khairunnisa, A. Adiwijaya, and S. Al Faraby, "Pengaruh Text Preprocessing terhadap Analisis Sentimen Komentar Masyarakat pada Media Sosial Twitter (Studi Kasus Pandemi COVID-19)," *Jurnal Media Informatika Budidarma*, vol. 5, no. 2, p. 406, Apr. 2021, doi: 10.30865/mib.v5i2.2835.
- [14] D. Oktavia and Y. R. Ramadhan, "Analisis Sentimen Terhadap Penerapan Sistem E-Tilang Pada Media Sosial Twitter Menggunakan Algoritma Support Vector Machine (SVM)," *Media Online*, vol. 4, no. 1, pp. 407–417, 2023, doi: 10.30865/klik.v4i1.1040.
- [15] Z. Alhaq, A. Mustopa, S. Mulyatun, and J. D. Santoso, "Penerapan Metode Support Vector Machine Untuk Analisis Sentimen Pengguna Twitter," *Journal of Information System Management (JOISM)*, vol. 3, no. 1, pp. 16–21, 2021.
- [16] A. Halim and A. Safuwani, "Analisis Sentimen Opini Warganet Twitter Terhadap Tes Screening Genose Pendeteksi Virus Covid-19 Menggunakan Metode Naive Bayes Berbasis Particle Swarm Optimization," *Jurnal Informatika Teknologi dan Sains (Jinteks)*, vol. 5, no. 1, pp. 170–178, 2023.
- [17] L. D. Andhika, D. R. Cahyani, D. Saputra, T. Herawati, M. Khoiruddinsyah, and D. D. Saputra, "Analisis Sentimen Kosumen KFC Berdasarkan Pendekatan Naive Bayes dan Ada Boost Berbasis Data Twitter," *Jurnal INSAN Journal of Information System Management Innovation*, vol. 3, no. 1, pp. 55–61, 2023.