

ANALISIS SENTIMEN WACANA KENAikan HARGA TIKET CANDI BOROBUDUR MENGGUNAKAN MULTINOMIAL NAÏVE BAYES

Muhammad Rifki¹, Imelda Imelda^{2*}

^{1,2}Teknik Informatika, Fakultas Teknologi Informasi, Universitas Budi Luhur
Email: ¹mrifki028@gmail.com, ^{2*} imelda@budiluhur.ac.id

(Naskah masuk: 23 Agustus 2022, diterima untuk diterbitkan: 24 Agustus 2022)

Abstrak

Wacana kenaikan harga tiket Candi Borobudur ramai diperbincangkan di media sosial. Banyak pendapat yang setuju dengan hal ini, dan tidak sedikit pula yang menolaknya. Analisis sentimen mengungkapkan pernyataan publik tentang isu-isu yang sedang berlangsung di komunitas tertentu. Tujuan penelitian ini untuk menganalisis sentimen masyarakat terhadap wacana kenaikan harga tiket Candi Borobudur menggunakan metode *Multinomial Naïve Bayes*. Alasan dipilihnya *Multinomial Naïve Bayes* karena algoritma ini merupakan salah satu algoritma machine learning yang paling populer digunakan untuk memproses klasifikasi teks, prosesnya sederhana, efisien, dan berkinerja baik. Sumber *dataset* penelitian ini berasal dari media sosial *Twitter*. Kontribusi penelitian ini adalah menganalisis sentimen masyarakat terhadap wacana kenaikan harga tiket Candi Borobudur menggunakan metode *Multinomial Naïve Bayes* dengan dua skenario. Pertama, kelas dibagi menjadi dua kategori: positif dan negatif. Kedua, kelas dibagi menjadi tiga kategori: positif, netral, dan negatif. Kategori netral diberikan untuk data yang tidak termasuk positif atau negatif. Hasil pengujian menunjukkan bahwa nilai akurasi untuk kelas dua kategori lebih tinggi dari 3. Hasil akurasi, presisi, dan *recall* untuk kelas dua kategori adalah 81%, 76%, dan 84%, sedangkan untuk kelas dengan tiga kategori yaitu 57% , 58%, dan 57%.

Kata kunci: *analisis sentimen, Multinomial Naive Bayes, harga tiket, Candi Borobudur, media sosial Twitter*

ANALYSIS OF DISCOURSE SENTIMENT OF BOROBUDUR TEMPLE TICKET PRICE INCREASE USING MULTINOMIAL NAVE BAYES

Abstract

The discourse of increasing the price of Borobudur Temple tickets has been widely discussed on social media. Many opinions agree with this, and not a few reject it. Sentiment analysis reveals public statements about ongoing issues in a particular community. The purpose of this study is to analyze public sentiment on the discourse of increasing ticket prices for Borobudur Temple using the Multinomial Naïve Bayes method. The reason for choosing Multinomial Naïve Bayes is because this algorithm is one of the most popular machine learning algorithms used to process text classification. After all, the process is simple, efficient, and performs well. The source of this research dataset comes from the social media Twitter. The contribution of this study is to analyze public sentiment on the discourse of increasing ticket prices for Borobudur Temple using the Multinomial Naïve Bayes method with two scenarios. First, the class is divided into two categories: positive and negative. Second, the class is divided into three categories: positive, neutral, and negative. The neutral category is given to data that does not include positive or negative. The test results show that the accuracy value for class two categories is higher than 3. The accuracy, precision, and recall results for class two categories are 81%, 76%, and 84%, while for types with three categories, namely 57%, 58%, and 57%.

Keywords: *Sentiment Analysis, Multinomial Naive Bayes, ticket prices, Borobudur Temple, social media Twitter*

1. PENDAHULUAN

Media adalah platform digital yang dapat digunakan orang untuk bersosialisasi, berbagi informasi, dan mengungkapkan pemikiran atau pendapat pengguna tentang suatu masalah yang sedang berlangsung. Orang dapat memperoleh berbagai pandangan tentang suatu isu yang sedang

terjadi. Twitter adalah layanan media sosial yang memungkinkan penggunaannya untuk mengirim dan membaca pesan dengan maksimal 280 karakter, yang disebut dengan kicauan. Menurut data dari KEMKOMINFO, Indonesia menduduki peringkat kelima dalam penggunaan media sosial twitter [1]. Berdasarkan data *tweet*, orang dapat

menggunakannya untuk menghasilkan sentimen yang diperlukan untuk membuat sistem yang dapat mengklasifikasikan analisis sentimen. *Text mining* merupakan salah satu teknik yang digunakan untuk mengklasifikasikan data tekstual [2]. *Text mining theory* adalah teori pemrosesan teks tidak terstruktur untuk menemukan pola unik dengan mengekstraksi pola yang berguna dari kumpulan teks [3].

Berbagai penelitian telah dilakukan untuk menganalisis opini publik dengan menggunakan data dari media sosial Twitter. Penelitian analisis sentimen terhadap pengguna transportasi online [4], pengguna gojek [5], review *steam* [6], review pengguna aplikasi *Shopee* pada hari belanja online nasional 2020 [7], kursi kosong Terawan [8] dan optimasi SVM berbasis PSO account officer ruang guru di Twitter [9].

Analisis sentimen adalah klasifikasi sentimen positif atau negatif dari penggalan pendapat seseorang tentang suatu peristiwa yang sedang atau telah terjadi. Analisis sentimen memberikan gambaran tentang suatu produk, merek, reputasi, ulasan, blog, diskusi, berita, komentar, dan umpan balik [10]. Analisis sentimen digunakan untuk membantu menemukan opini atas komentar pengguna media sosial secara otomatis [11].

Algoritme yang digunakan pada penelitian ini adalah algoritme *Multinomial Naïve Bayes*. *Multinomial Naïve Bayes* merupakan salah satu teknik *machine learning* dalam *supervised learning* untuk melakukan klasifikasi teks. Caranya, dengan menggunakan nilai probabilitas suatu kelas dalam suatu dokumen yang diproses. Alasan pemilihan *Multinomial Naïve Bayes* karena algoritme ini merupakan salah satu algoritme *machine learning* yang populer digunakan untuk memproses klasifikasi teks karena prosesnya sederhana, efisien, dan berkinerja baik [10].

Tujuan dari penelitian ini menganalisis sentimen masyarakat terhadap wacana kenaikan harga tiket Candi Borobudur dengan menggunakan metode *Multinomial Naïve Bayes*. Hasil akhirnya diperoleh persentase sentimen publik terkait setuju/mendukung wacana kenaikan harga tiket Candi Borobudur atau penolakan wacana kenaikan harga tiket Candi Borobudur. Manfaat penelitian ini agar sistem klasifikasi dengan algoritme *Multinomial Naïve Bayes* dapat membantu memberikan gambaran tentang sentimen masyarakat terkait kenaikan harga tiket Candi Borobudur.

Data yang digunakan dalam penelitian ini adalah data tweet. Data ini diperoleh dari *Twitter* melalui proses *crawling data*. Penelitian ini melakukan 2 skenario kelas. Pertama, kelas dibagi menjadi dua kategori: positif dan negatif. Kedua, kelas dibagi menjadi tiga kategori yaitu positif, netral dan negatif. Kategori netral diberikan untuk data yang tidak termasuk positif atau negatif. Kontribusi penelitian ini menganalisis sentimen masyarakat terhadap wacana kenaikan harga tiket Candi

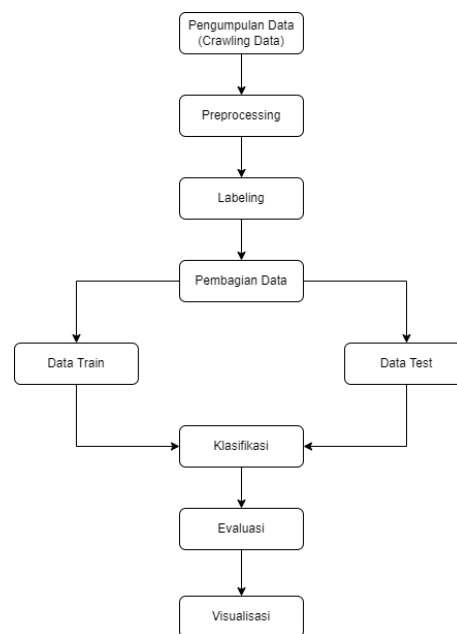
Borobudur menggunakan metode *Multinomial Naïve Bayes* dengan dua skenario kelas.

2. METODE PENELITIAN

Gambar 1 menunjukkan tahapan penelitian. Penelitian ini memiliki tujuh langkah kritis: pengumpulan data (*crawling*), *preprocessing* data, pelabelan, pembagian data, klasifikasi menggunakan *Multinomial Naïve Bayes*, evaluasi menggunakan *confusion matrix*, dan visualisasi. Pembagian data terdiri dari *Data Train* dan *Data Test*. Pembagian data dibagi menjadi 80:20.

2.1. Pengumpulan Data

Pengumpulan data dilakukan melalui proses *crawling* menggunakan *library Tweepy*. *Crawling* merupakan proses mendapatkan dataset berupa tweet, sehingga menghasilkan input pada tahap *text preprocessing* [12]. Proses *crawling* dalam penelitian ini dikumpulkan berdasarkan dua parameter kata kunci, yaitu 'tiket borobudur', dan 'harga tiket candi borobudur'.

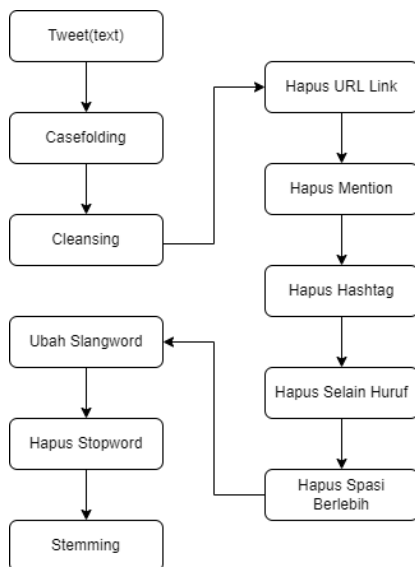


Gambar 1. Tahapan Penelitian

2.2. Text Preprocessing

Text Preprocessing mengolah data mentah sebelum diolah ke tahap selanjutnya dengan menghilangkan data yang tidak perlu atau mengubah data tersebut ke dalam bentuk yang lebih mudah untuk diproses oleh sistem komputasi [13]. Tujuan beberapa proses pada tahap *preprocessing* ini untuk menghasilkan data *tweet* yang bersih sehingga algoritme *Multinomial Naïve Bayes* dapat diuji lebih optimal. Ilustrasi proses *text preprocessing* ditunjukkan pada Gambar 2.

Text preprocessing memiliki 5 tahap, meliputi: *Case Folding*, *Cleansing*, *Slang word*, *Stop word*, dan *Stemming*. Tahap pertama *text preprocessing* dimulai dengan proses *case folding*. *Case Folding* adalah teknik mengubah semua huruf kapital dalam dokumen akan diubah menjadi huruf kecil [14]. Tahap kedua *text preprocessing* dilanjutkan dengan proses *cleansing*. *Cleansing* merupakan proses pembersihan data bertujuan agar *dataset* tidak mengandung *noise* yang dapat mempengaruhi hasil klasifikasi [15]. Tahap ketiga *text preprocessing* dilakukan proses mengubah *slang word*. *Slang word* merupakan penggunaan bahasa gaul yang mengakibatkan penggunaan bahasa Indonesia tidak baku. Proses mengubah *slang word* yaitu dengan melakukan *convert word*. Caranya, ganti *slang word* menjadi kata yang baku sesuai ejaan Bahasa Indonesia (EYD). Proses ini juga termasuk mendapatkan informasi sebanyak mungkin dari teks, kata-kata tidak baku, baik kata gaul, singkatan atau yang lainnya, untuk ditambahkan ke dalam kamus *slang word* [16].



Gambar 2. Text Preprocessing

Tahap keempat *text preprocessing* menghilangkan *stopword*. Proses menghilangkan *stopword* merupakan proses menghilangkan kata. Proses ini menghapus kata-kata yang sering muncul, bersifat umum dan tidak berpengaruh seperti kata “ini”, “apa”, “lagi” dengan menggunakan daftar kata *stopword* atau daftar kata (*stoplist*) berbahasa Indonesia [17]. Penambahan kata secara manual karena terdapat beberapa kata yang belum ada di daftar kata *stopword* [18]. Tahap kelima *text preprocessing* melakukan proses *stemming*. *Stemming* merupakan proses untuk mendapatkan *root/stem* atau kata dasar dari suatu kata dalam kalimat dengan cara memisahkan masing-masing kata dari kata dasar dan imbuhan baik awalan (*prefix*) maupun akhiran (*suffix*) [19]. Penelitian ini menggunakan pustaka Sastrawi untuk proses

stemming yang menerapkan algoritme *stemming* Nazief dan Adriani [20]. Proses *stemming* dilakukan dengan menghilangkan *inflection suffixes* (“-lah”, “-kah”, “-ku”, “-mu”, atau “-nya”), *possessive pronouns* (“-ku”, “-mu”, atau “-nya”), *derivation suffixes* (“-i”, “-an” atau “-kan”) dan *derivation prefixes* (“di-”, “ke-”, “se-”, “te-”, “be-”, “me-”, atau “pe-”) [21].

2.3. Labeling

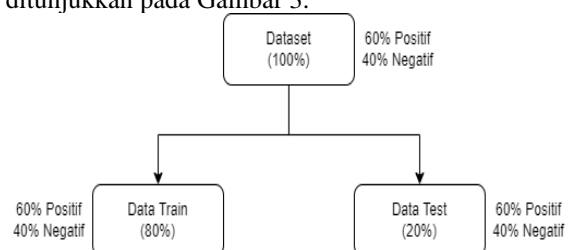
Labeling atau pelabelan merupakan proses pemberian kelas berdasarkan ciri atau karakteristik yang terkandung dalam sebuah dokumen atau kalimat [12]. Proses ini dilakukan pemberian label/kelas pada setiap data berdasarkan karakteristik dari kalimat yang terdapat pada dokumen. Pelabelan 2 kelas dan 3 kelas dilakukan manual secara subjektif.

Skenario pertama, data tweet yang telah dilakukan proses *preprocessing* diberi label kelas :positif dan negatif. Kelas positif diberikan pada data *tweet* yang mengandung pernyataan setuju/mendukung wacana kenaikan harga tiket Candi Borobudur. Kelas negatif diberikan pada data *tweet* yang mengandung pernyataan menyangkal/menolak wacana kenaikan harga tiket Candi Borobudur.

Skenario kedua, data tweet yang telah dilakukan proses *preprocessing* diberi label kelas: positif, netral dan negatif. Label positif diberikan pada setiap tweet yang mengandung opini setuju, mendukung dan menerima soal wacana kenaikan harga tiket candi borobudur. Label negatif diberikan pada setiap tweet yang mengandung opini menolak, mengkritik dan menyangkal soal wacana kenaikan harga tiket candi borobudur. Label netral diberikan pada setiap tweet yang tidak memiliki sentimen positif maupun negatif.

2.4. Pembagian Data

Pada tahapan ini *data tweet* yang telah diberikan label/kelas akan dibagi menjadi data *train* dan data *test* dengan perbandingan rasio 80:20 secara proporsional dengan menggunakan teknik *Stratified Random Sampling*. Metode *Stratified Random Sampling* merupakan metode penarikan sampel yang dilakukan dengan cara membagi populasi menjadi populasi yang lebih kecil, pembentukan harus sedemikian rupa sehingga setiap *stratum homogeny* berdasarkan suatu atau beberapa kriteria tertentu, kemudian dari setiap *stratum* diambil sampel secara acak [22]. Ilustrasi tahapan pembagian data ditunjukkan pada Gambar 3.



Gambar 1 Pembagian Data

2.5. Klasifikasi *Multinomial Naïve Bayes*

Algoritme *Multinomial Naïve Bayes* merupakan salah satu metode pembelajaran probabilistik yang didasarkan pada teorema *Bayes*. Algoritme ini bekerja dengan konsep *term frequency* yang menghitung berapa kali kata tersebut muncul dalam dokumen. Model ini juga menjelaskan kata tersebut muncul atau tidak serta frekuensi kemunculannya dalam suatu dokumen [14]. *Multinomial Naïve Bayes* terdiri dari 3 tahap perhitungan. Formula perhitungan nilai prior ditunjukkan pada persamaan 1.

$$P(c_j) = \frac{\text{count}(c_j)}{N} \quad (1)$$

dengan $P(c_j)$ merupakan probabilitas dari sebuah kelas, $\text{count}(c_j)$ merupakan banyaknya kelas c_j yang ada pada dokumen dan N merupakan total dokumen. Formula perhitungan nilai likelihood ditunjukkan pada persamaan 2.

$$P(w_i, c_j) = \frac{\text{count}(w_i, c_j) + 1}{\sum \text{count}(w_i, c_j) + |V|} \quad (2)$$

dengan $P(w_i, c_j)$ merupakan probabilitas setiap kata w_i pada kelas c_j , $\text{count}(w_i, c_j)$ merupakan jumlah kemunculan kata w_i pada kelas c_j , $\sum \text{count}(w_i, c_j)$ merupakan jumlah total kata w_i dalam kelas c_j dan $|V|$ merupakan jumlah kata unik atau token yang ada pada seluruh kelas. Penambahan nilai 1 disebut sebagai pendekatan parameter laplace smoothing yang berfungsi untuk menghindari probabilitas menghasilkan nilai 0. Formula perhitungan nilai posterior ditunjukkan pada persamaan 3.

$$P(c_j|w_i) = P(c_j) \times P(w_1, c_j) \times \dots \times P(w_n, c_n) \quad (3)$$

dengan $P(c_j|w_i)$ merupakan peluang kelas c_j jika terdapat kemunculan kata w_i , $P(c_j)$ merupakan hasil nilai probabilitas prior dan $P(w_i, c_j)$ merupakan hasil nilai probabilitas likelihood untuk setiap kata.

2.6. Evaluasi

Evaluasi dilakukan untuk mengetahui nilai akurasi, presisi dan *recall* dari pengujian model *Multinomial Naïve Bayes*. Evaluasi ini dilakukan dengan cara membandingkan data hasil prediksi yang didapatkan dari model hasil klasifikasi dengan *data labeling* aktual dari proses *labeling*. *Confusion Matrix* merupakan tahapan untuk menganalisis dan menilai kinerja dari sistem yang dirancang. Kinerja diukur dari segi akurasi, presisi, dan recall [13]. Formula menghitung nilai akurasi, presisi dan recall ditunjukkan pada persamaan 4, 5, dan 6.

$$\text{Akurasi} = \frac{TP+TN}{TP+TN+FP+FN} \quad (4)$$

$$\text{Presisi} = \frac{TP}{TP+FP} \quad (5)$$

$$\text{Recall} = \frac{TP}{TP+FN} \quad (6)$$

Persamaan 4,5, dan 6 menunjukkan bahwa *True Positive* (TP) merupakan data aktual positif dan diprediksi sebagai data positif oleh sistem. *True*

Negative (TN) merupakan data aktual negatif dan diprediksi sebagai data negatif oleh sistem. *False Positive* (FP) merupakan data aktual negatif namun diprediksi sebagai data positif oleh sistem. *False Negative* (FN) merupakan data aktual positif namun diprediksi sebagai data negatif oleh sistem.

3. HASIL DAN PEMBAHASAN

3.1. Pengumpulan Data

Dataset penelitian bersumber dari media sosial *Twitter* berupa data *tweet*. *Dataset* tersebut diperoleh melalui proses *crawling* data menggunakan pustaka *tweepy*. Data dikumpulkan mulai tanggal 7 Juni 2022 hingga 15 Juni 2022. Jumlah data *tweet* yang didapat sebanyak 740 data. Parameter kata kunci yang digunakan yaitu : ‘tiket borobudur’ dan harga tiket candi borobudur’. *Dataset* yang diambil melalui proses *crawling data* yaitu : *id*, *text(tweet)*, *username* dan *created_at*. Kemudian, hasilnya disimpan ke dalam file excel (.xlsx).

3.2. Text Preprocessing

Tahap *preprocessing* merupakan tahapan yang dapat dilakukan setelah tersedianya *dataset* dari hasil tahap pengumpulan data pada database. Tahap *text preprocessing* meliputi: *Case Folding*, *Cleansing*, *Slang word*, *Stop word*, dan *Steeming*.

Tahap pertama yaitu *Case Folding*. Tahap ini mengubah semua kata yang mengandung huruf besar/kapital pada dokumen menjadi huruf kecil. Proses *Case Folding* dapat ditunjukkan pada Tabel 1.

Tabel 1. Tabel *Case Folding*

Sebelum <i>Case Folding</i>	Setelah <i>Case Folding</i>
Kalo tiketnya di naikin 750k sbenrnya gapapa sih, toh itukan tiket buat naik ke candinya, kalo masuk komplek candi masih bisa di jangkau semua kalangan, dan poin plusnya candi borobudur juga bisa jadi lebih sakral buat umat hindu. https://t.co/ym1kd5vVQ5	kalo tiketnya di naikin 750k sbenrnya gapapa sih, toh itukan tiket buat naik ke candinya, kalo masuk komplek candi masih bisa di jangkau semua kalangan, dan poin plusnya candi borobudur juga bisa jadi lebih sakral buat umat hindu. https://t.co/ym1kd5vvq5

Tahap kedua yaitu *Cleansing*. Tahap ini mengubah karakter yang tidak diperlukan. Karakter seperti *mention*, *hashtag*, *link URL*, karakter selain huruf a-z diubah menjadi spasi. Proses *Cleansing* juga termasuk menghapus spasi berlebih diakhir kata dan menghapus spasi berlebih diantara kata. Proses *Cleansing* ditunjukkan pada Tabel 2.

Tabel 2. Tabel *Cleansing*

Sebelum <i>Cleansing</i>	Setelah <i>Cleansing</i>
kalo tiketnya di naikin 750k sbenrnya gapapa sih, toh itukan tiket buat naik ke candinya, kalo masuk komplek candi masih bisa di jangkau semua kalangan, dan poin plusnya candi borobudur juga bisa jadi lebih sakral buat umat hindu. https://t.co/ym1kd5vvq5	kalo tiketnya di naikin k sbenrnya gapapa sih toh itukan tiket buat naik ke candinya kalo masuk komplek candi masih bisa di jangkau semua kalangan dan poin plusnya candi borobudur juga bisa jadi lebih sakral buat umat hindu https://t.co/ym1kd5vvq5

Tahap ketiga yaitu mengubah *slang word*. Tahap ini mengganti kata tidak baku menjadi kata baku sesuai kaidah KBBI yang berlaku. Proses mengubah *slang word* ditunjukkan pada Tabel 3.

Tabel 3. Tabel Mengubah *Slang word*

Sebelum <i>Slang word</i>	Setelah <i>Slang word</i>
kalo tiketnya di naikin k sbenrnya gapapa sih toh itukan tiket buat naik ke candinya kalo masuk kompleks candi masih bisa di jangkau semua kalangan dan poin plusnya candi borobudur juga bisa jadi lebih sakral buat umat hindu	kalau tiketnya di naik ke sebenarnya gapapa sih toh itukan tiket buat naik ke candinya kalau masuk kompleks candi masih bisa di jangkau semua kalangan dan poin plusnya candi borobudur juga bisa jadi lebih sakral buat umat hindu

Tahap keempat yaitu menghilangkan *stop word*. Tahap ini menghapus atau menghilangkan setiap kata yang tidak memiliki makna yang sering ada pada teks. Proses menghilangkan *stop word* ditunjukkan pada Tabel 4.

Tabel 4. Tabel Menghilangkan *Stop word*

Sebelum <i>Stop word</i>	Setelah <i>Stop word</i>
kalau tiketnya di naik ke sebenarnya gapapa sih toh itukan tiket buat naik ke candinya kalau masuk kompleks candi masih bisa di jangkau semua kalangan dan poin plusnya candi borobudur juga bisa jadi lebih sakral buat umat hindu	tiketnya naik gapapa itukan tiket naik candinya masuk kompleks candi jangkau kalangan poin plusnya candi borobudur sakral umat hindu

Tahap kelima yaitu *stemming*. Tahap ini mengubah setiap kata dalam dokumen menjadi kata dasarnya. Proses *stemming* ditunjukkan pada Tabel 5.

Tabel 5. Tabel *Stemming*

Sebelum <i>Stemming</i>	Setelah <i>Stemming</i>
tiketnya naik gapapa itukan tiket naik candinya masuk kompleks candi jangkau kalangan poin plusnya candi borobudur sakral umat hindu	tiket naik gapapa itu tiket naik candi masuk kompleks candi jangkau kalang poin plus candi borobudur sakral umat hindu

3.3. Labeling

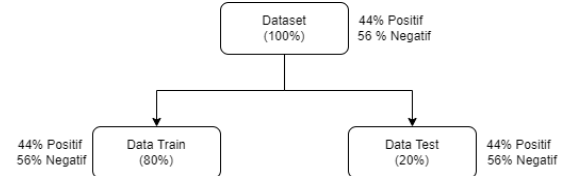
Pada tahap ini dilakukan proses *Labeling*. Pelabelan dilakukan secara manual terhadap dataset yang sudah melalui proses *preprocessing*.

Skenario pertama, label yang diberikan adalah label kelas positif dan negatif. Label positif diberikan pada setiap tweet yang mengandung opini setuju, mendukung dan menerima soal wacana kenaikan harga tiket Candi Borobudur. Label negatif diberikan pada setiap tweet yang mengandung opini menolak, mengkritik dan menyangkal soal wacana kenaikan harga tiket Candi Borobudur. Data skenario pertama hanya diseleksi untuk kelas positif dan negatif. Data yang tidak termasuk kelas positif dan negatif maka data itu diabaikan.

Skenario kedua, label yang diberikan adalah label kelas: positif, netral dan negatif. Label positif diberikan pada setiap tweet yang mengandung opini setuju, mendukung dan menerima soal wacana kenaikan harga tiket candi borobudur. Label negatif diberikan pada setiap *tweet* yang mengandung opini menolak, mengkritik dan menyangkal soal wacana kenaikan harga tiket Candi Borobudur. Label netral diberikan pada setiap *tweet* yang tidak memiliki sentimen positif maupun negatif.

3.4. Pembagian Data

Tahapan pembagian data dilakukan proses membagi dataset menjadi data *train* dan data *test* dengan menggunakan teknik *Stratified Random Sampling*. Teknik *Stratified Random Sampling* bertujuan untuk membuat pembagian data *train* dan data *test* berdasarkan *dataset* menjadi proporsional dan tidak bias. Total data tweet adalah 740. Skenario pertama, total data berlabel adalah 213 untuk *dataset* yang terdiri dari 44% data positif dan 56% data negatif. Data di seleksi secara manual. Data sebanyak 717 diabaikan karena tidak berlabel positif atau negatif. Sistem ini menggunakan pembagian data dengan perbandingan 80 : 20. Data ini menghasilkan 170 data *train* dan 43 data *test*.



Gambar 4 Pembagian Data

Gambar 4 menunjukkan teknik pembagian data. Pembagian strata dilakukan terlebih dahulu pada label positif dan negatif yang ada dalam *dataset*. Jika terdapat 44% label positif dan 56% label negatif pada dataset, maka data dibagi 80% data *train* dan 20% data *test*. Data *train* dan data *test* memiliki perbandingan 44% label positif dan 56% label negatif pada tiap data (data *train* dan data *test*).

Skenario kedua, kami menggunakan seluruh data yang ada. Total data berjumlah 740 data *tweet*. Jumlah data yang diberi label netral sebanyak 717 data. Data berlabel negatif sebanyak 119 data. Data berlabel positif sebanyak 94 data. Tahap selanjutnya dilakukan pembagian data seperti yang ditunjukkan pada Gambar 4.

3.4. Klasifikasi Multinomial Naive Bayes

Tahapan klasifikasi *Multinomial Naive Bayes* dilakukan setelah tahap *Labeling* dan pembagian data. Data *train* digunakan untuk membangun model *Multinomial Naive Bayes* dan data *test* digunakan untuk mengevaluasi performa model yang diperoleh dari proses *training*. Pada proses *training model* diawali dengan menghitung nilai probabilitas *prior*

dari setiap kelas yaitu positif dan negatif. Kemudian dilanjutkan dengan menghitung nilai probabilitas *likelihood* dengan menghitung setiap kata pada kelas. Tahapan terakhir yaitu menghitung nilai probabilitas *posterior* dengan menghitung peluang sebuah dokumen terhadap suatu kelas serta menentukan kelas dokumen dengan memilih nilai probabilitas tertinggi.

3.6. Pengujian

Pengujian merupakan salah satu hal yang sangat penting dalam pengembangan sistem. Tujuan pengujian untuk mengevaluasi dan menganalisis tingkat akurasi yang telah dicapai oleh sistem yang telah dibuat. Pada pengujian ini dilakukan pengujian terhadap akurasi, presisi dan *recall* terhadap metode *Multinomial Naïve Bayes*.

Tabel 6 menunjukkan sampel hasil data prediksi dengan 2 kelas. Label aktual didapatkan dari proses *labeling* manual. Label prediksi didapatkan dari proses klasifikasi dari *Multinomial Naïve Bayes*. Keseluruhan hasil prediksi ditunjukkan pada Tabel 6.

Tabel 6. Tabel sampel data hasil prediksi dengan 2 kelas

No	Tweet	Label aktual	Label prediksi
1	@Dina4rtya Berapa waktu lalu kita ke Borobudur, ngak bisa naik keatas, akhirnya kita batalin masuk ke @BorobudurJogja, ngapain jauh jauh datang ngak bisa melihat langsung, sekarang tiket udah naik 750k mending uang nya di tabung...boros amat ... #BOROBUDUR #WISATAJOGJA	Negatif	Negatif
2	@RANGERmounts @alisyarief @jokowi Pada belum pernah ke Hagia Sofia di Turki? Itu lebih tua dari Borobudur dan menunjukkan bahwa constantinopel udah jauh lebih maju pada abad segitu. Berapa tiket masuk Hagia Sofia? Gratis.	Negatif	Negatif
3	@susipudjiastuti Bu @susipudjiastuti sdh baca belum ttg tiket 750rb itu? Tiket masuk kawasan Borobudur ttp Bu. Yg 750m itu baik ke Candi nya. Utk melindungi candi dr kerusakan harusnya jangan dinaikkan harganya tp tutup akses naik candi. Kayaknya Ibu sdh masuk barisan 'Baca Judul Abaikan Isi'	Negatif	Positif
4	Kalau tiket masuk Borobudur jadi 750 ribu auto anak muda domisili radius 30 km pacarannya cari lokasi lain. Dulu kan emang sengaja sambil terangin relief dinding candi biar si dia terpesona. Kembali rumah langsung putus, soale ga dijanin dawet. Panas kan itu tempat.	Negatif	Positif

5	Memang enggak gampang. Tp kudu diusahakan. Soal tiket Candi Borobudur ini jadi momen untuk mulai memikirkan Cara menjaga situs sejarah lainnya	Positif	Positif
43	@convomf Nah gitu dong, dikira warga Indo pada sultan2 semua kali ya jd tiket candi borobudur dibikin mahal. Minat pengunjung lokal udah minim ditambah naikin harga tiket masuk, makin gak ada yg mau kesana.	Negatif	Negatif

Tabel 7 menunjukkan sampel hasil data prediksi dengan 2 kelas. Label aktual didapatkan dari proses *labeling* manual. Label prediksi didapatkan dari proses klasifikasi dari *Multinomial Naïve Bayes*. Keseluruhan hasil prediksi ditunjukkan pada Tabel 7.

Tabel 7. Hasil pengujian confusion matrix 2x2

Prediksi	Aktual	
	Positif	Negatif
Positif	16	5
Negatif	3	19

Hasil pengujian dengan 2 kelas menunjukkan bahwa algoritme *Multinomial Naïve Bayes* mampu memperoleh nilai pengujian dengan akurasi sebesar 81%, presisi sebesar 76%, dan *recall* sebesar 84%. Tabel 8 menunjukkan sampel hasil data prediksi dengan 3 kelas. Label aktual didapatkan dari proses *labeling* manual. Label prediksi didapatkan dari proses klasifikasi dari *Multinomial Naïve Bayes*. Keseluruhan hasil prediksi ditunjukkan pada Tabel 8.

Tabel 9 menunjukkan sampel hasil data prediksi dengan 3 kelas. Label aktual didapatkan dari proses *labeling* manual. Label prediksi didapatkan dari proses klasifikasi dari *Multinomial Naïve Bayes*. Keseluruhan hasil prediksi ditunjukkan pada Tabel 9 dalam *confusion matrix*.

Tabel 8. Tabel sampel data hasil prediksi dengan 3 kelas

No	Tweet	Label aktual	Label prediksi
1	@chandra_lucky @susipudjiastuti betul ka, makanya pada nolak ni https://t.co/Z8INF3t6rN	Negatif	Netral
2	@cokkyturnip @Wisata_IDN betul sih ka, lebih efisien gitu ya. kasian orang yang mau berkunjung beneran jadi ga bisa. https://t.co/Z8INF3t6rN	Positif	Negatif
3	Kalau tiket masuk Borobudur jadi 750 ribu auto anak muda domisili radius 30 km pacarannya cari lokasi lain. Dulu kan emang sengaja sambil terangin relief dinding candi biar si dia terpesona. Kembali rumah langsung	Negatif	Positif

	putus, soale ga dijajani dawet. Panas kan itu tempat.		
4	Konservasi juga jadi salah satu alasan menaikkan harga tiket. https://t.co/Gf7ls4rbkn #CNNIndonesia https://t.co/86kBAZxrjs	Negatif	Netral
:	:	:	:
148	???????? Mau nurunin minyak goreng eeeehhh yg naik harga tiket masuk Candi Borobudur. ???????? Super tolol sekali. https://t.co/nM10HgibmX	Negatif	Negatif

Tabel 9. Hasil pengujian confusion matrix 3x3

Prediksi	Aktual		
	Negatif	Netral	Positif
Negatif	29	9	14
Netral	10	27	17
Positif	9	5	28

Hasil pengujian dengan 3 kelas dari Tabel 9 menunjukkan bahwa algoritme *Multinomial Naïve Bayes* mampu memperoleh nilai pengujian dengan akurasi sebesar 57%, presisi sebesar 58%, dan *recall* sebesar 57%.

4. KESIMPULAN

Penelitian ini berhasil melakukan analisis sentimen menggunakan algoritme *Multinomial Naïve Bayes* dengan topik wacana kenaikan harga tiket Candi Borobudur dengan 2 skenario kelas. Pertama, kelas dibagi menjadi dua kategori: positif dan negatif. Kedua, kelas dibagi menjadi tiga kategori: positif, netral, dan negatif. Kategori netral diberikan untuk data yang tidak termasuk positif atau negatif. Sentimen masyarakat Indonesia terhadap wacana kenaikan harga tiket Candi Borobudur yang netral sebesar 36% dari total 740 data *tweet*. Sentimen positif sebesar 44% dan sentimen negatif sebesar 56% dari 213 data *tweet*. Hasil pengujian menunjukkan bahwa nilai akurasi untuk kelas dua kategori lebih tinggi dari 3. Hasil akurasi, presisi, dan *recall* untuk kelas dua kategori adalah 81%, 76%, dan 84%, sedangkan untuk kelas dengan tiga kategori yaitu 57%, 58%, dan 57%. Tahap *preprocessing* yang baik menjadi penentu dalam terbentuknya hasil yang optimal untuk tahap selanjutnya. Penelitian selanjutnya diharapkan pelabelan divalidasi oleh pakar. Akurasi sistem masih dapat ditingkatkan pada penelitian selanjutnya dengan menerapkan beberapa algoritma misalnya *Particle Swarm Optimization (PSO)*, *Ant Colony Optimization (ACO)* atau *Genetic Algorithm (GA)*. Penelitian selanjutnya juga dapat menambahkan seleksi fitur misalnya *Information Gain* sebagai usulan *novelty* berikutnya.

5. DAFTAR PUSTAKA

- [1] U. S. A. Rahman, Y. Wibisono, dan E. P. Nugroho, 2020, "Implementasi Multinomial Naive Bayes Untuk Klasifikasi Ujaran Kebencian Pada Dataset Kicauan (Twitter) Bahasa Indonesia," *Jatikom J. Apl. dan Teor. Ilmu Komput.*, vol. 3, no. 2, hal. 78–84.
- [2] N. S. Wardani, A. Prahutama, dan P. Kartikasari, 2020, "Analisis Sentimen Pemindahan Ibu Kota Negara Dengan Klasifikasi Naïve Bayes Untuk Model Bernoulli Dan Multinomial," *J. Gaussian*, vol. 9, no. 3, hal. 237–246, doi: 10.14710/j.gauss.v9i3.27963.
- [3] A. Sabrani, I. G. W. Wedashwara W., dan F. Bimantoro, 2020, "Multinomial Naïve Bayes untuk Klasifikasi Artikel Online tentang Gempa di Indonesia," *J. Teknol. Informasi, Komputer, dan Apl. (JTika)*, vol. 2, no. 1, hal. 89–100, doi: 10.29303/jtika.v2i1.87.
- [4] S. Mandasari, B. H. Hayadi, dan R. Gunawan, 2022, "Jurnal Teknologi Sistem Informasi dan Sistem Komputer TGD Analisis Sentimen Pengguna Transportasi Online Terhadap Layanan Grab Indonesia Menggunakan Multinomial Naive Bayes Classifier Jurnal Teknologi Sistem Informasi dan Sistem Komputer TGD," vol. 5, hal. 118–126.
- [5] F. R. Irawan, A. Jazuli, dan T. Khotimah, 2022, "ANALISIS SENTIMEN TERHADAP PENGGUNA GOJEK MENGGUNAKAN METODE K-NEAREST NEIGHBORS SENTIMENT ANALYSIS OF GOJEK USERS USING K-NEAREST NEIGHBOR," *JIKO (Jurnal Inform. dan Komputer)*, vol. 5, no. 1, hal. 62–68, doi: 10.33387/jiko.
- [6] R. Haditira, D. T. Murdiansyah, dan W. Astuti, 2022, "Analisis Sentimen Pada Steam Review Menggunakan Metode Multinomial Naïve Bayes dengan Seleksi Fitur Gini Index Text," *e-Proceeding Eng.*, vol. 9, no. 3, hal. 1793–1799.
- [7] L. M. Siniwi, A. Prahutama, dan A. R. Hakim, 2021, "QUERY EXPANSION RANKING PADA ANALISIS SENTIMEN MENGGUNAKAN KLASIFIKASI MULTINOMIAL NAÏVE BAYES," *J. Gaussian*, vol. 10, no. 3, hal. 377–387.
- [8] P. Morgan, B. Show, dan N. Media, 2022, "ANALISIS SENTIMEN PENGGUNA YOUTUBE TERHADAP TAYANGAN #MATANAJWAMENANTITERAWAN DENGAN METODE NAÏVE BAYES SENTIMENT ANALYSIS OF USER COMMENTS ON YOUTUBE VIDEO #MATANAJWAMENANTITERAWAN WITH NAIVE BAYES CLASSIFIER METHOD," *JIKO (Jurnal Inform. dan Komputer)*, vol. 5, no.

- 1, hal. 1–6, doi: 10.33387/jiko.
- [9] R. Darmawan, I. Indra, dan A. Surahmat, 2022, “Optimalisasi Support Vector Machine (SVM) Berbasis Particle Swarm Optimization (PSO) Pada Analisis Sentimen Terhadap Official Account Ruang Guru Di Twitter,” *J. Kaji. Ilm.*, vol. 22, no. 2, hal. 143–152.
- [10] L. Lesmana, F. Nabyala, dan Mukrodin, 2020, “Analisis Sentimen Pengguna Twitter Terhadap Kebijakan Sistem Zonasi Ppdb Menggunakan Algoritma Multinomial Naive Bayes,” *J. Sist. Inf. dan Teknol. Perad.*, vol. 1, no. 1.
- [11] F. Prasetyawan, S. Widiyanesti, dan Tri Widarmanti, 2022, “Analisis Sentimen Mengenai Kualitas Layanan Jasa Ekspedisi Barang Sicepat Di Media Sosial Twitter,” in *eProceedings of Management*.
- [12] M. Priandi dan P. Painem, 2021, “Analisis Sentimen Masyarakat Terhadap Pembelajaran Daring di Era Pandemi Covid-19 pada Media Sosial Twitter Menggunakan Ekstraksi Fitur Countvectorizer dan Algoritma K-Nearest Neighbor,” in *Senamika (Seminar Nasional Mahasiswa Ilmu Komputer dan Aplikasinya)*, no. September, hal. 311–319.
- [13] C. S. Sriyano dan E. B. Setiawan, 2021, “Pendeteksian Berita Hoax Menggunakan Naive Bayes Multinomial Pada Twitter dengan Fitur Pembobotan TF-IDF,” in *e-Proceeding of Engineering*, vol. 8, no. 2, hal. 3396–3405.
- [14] Yuyun, Nurul Hidayah, dan Supriadi Sahibu, 2021, “Algoritma Multinomial Naive Bayes Untuk Klasifikasi Sentimen Pemerintah Terhadap Penanganan Covid-19 Menggunakan Data Twitter,” *J. RESTI (Rekayasa Sist. dan Teknol. Informasi)*, vol. 5, no. 4, hal. 820–826, doi: 10.29207/resti.v5i4.3146.
- [15] N. L. P. M. Putu, Ahmad Zuli Amrullah, dan Ismarmiaty, 2021, “Analisis Sentimen dan Pemodelan Topik Pariwisata Lombok Menggunakan Algoritma Naive Bayes dan Latent Dirichlet Allocation,” *J. RESTI (Rekayasa Sist. dan Teknol. Informasi)*, vol. 5, no. 1, hal. 123–131, doi: 10.29207/resti.v5i1.2587.
- [16] M. A. Nurrohmat dan A. SN, 2019, “Sentiment Analysis of Novel Review Using Long Short-Term Memory Method,” *IJCCS (Indonesian J. Comput. Cybern. Syst.)*, vol. 13, no. 3, hal. 209–218.
- [17] H. Dwiharyono dan S. Suyanto, 2022, “Stemming for Better Indonesian Text-to-Phoneme ☆,” *Ampersand*, vol. 9, hal. 100083, doi: 10.1016/j.amper.2022.100083.
- [18] R. Rinandyaswara, Y. A. Sari, dan M. T. Furqon, 2022, “PEMBENTUKAN DAFTAR STOPWORD MENGGUNAKAN TERM BASED RANDOM SAMPLING PADA ANALISIS SENTIMEN DENGAN METODE NAÏVE BAYES (STUDI KASUS: KULIAH DARING DI MASA PANDEMI),” *J. Teknol. Inf. dan Ilmu Komput.*, vol. 9, no. 4, hal. 717–724, doi: 10.25126/jtiik.202294707.
- [19] M. T. A. Bangsa, S. Priyanta, dan Y. Suyanto, 2020, “Aspect-Based Sentiment Analysis of Online Marketplace Reviews Using Convolutional Neural Network,” *IJCCS (Indonesian J. Comput. Cybern. Syst.)*, vol. 14, no. 2, hal. 123–134, doi: 10.22146/ijccs.51646.
- [20] D. Mustikasari, I. Widaningrum, R. Arifin, dan W. H. E. Putri, 2021, “Comparison of Effectiveness of Stemming Algorithms in Indonesian Documents,” in *Advances in Engineering Research*, vol. 203, hal. 154–158.
- [21] W. A. Rifai dan E. Winarko, 2019, “Modification of Stemming Algorithm Using A Non Deterministic Approach To Indonesian Text,” *IJCCS (Indonesian J. Comput. Cybern. Syst.)*, vol. 13, no. 4, hal. 379–388.
- [22] S. F. Ulya, Y. Sukestiyarno, dan P. Hendikawati, 2018, “Analisis Prediksi Quick Count dengan Metode Stratified Random Sampling dan Estimasi Confidence Interval Menggunakan Metode Maksimum Likelihood,” *UNNES J. Math.*, vol. 7, no. 1, hal. 108–119.