

Analisis Sentimen Masyarakat terhadap Budaya Batak di Twitter Menggunakan Metode Naive Bayes

Cristina Situmorang^{1,*}, Amelia Sanna Maria Hutauruk², Sardo Pardingotan Sipayung³, Wilfred Raimond Zebua⁴

^{1,2,3,4} Ilmu Komputer, Teknik Informatika, Universitas Katolik Santo Thomas, Medan, Indonesia

Email: ^{1*}cristinasitumorang26@gmail.com, ²ameliasannamaria@gmail.com, ³pinarsiphom@gmail.com, ⁴wilfredzeb@gmail.com

(* Email Corresponding Author: cristinasitumorang26@gmail.com)

Received: 27 Januari 2026 | Revision: 30 Januari 2026 | Accepted: 30 Januari 2026

Abstrak

Media sosial Twitter telah menjadi salah satu platform utama bagi masyarakat dalam menyampaikan opini dan pandangan terhadap berbagai isu, termasuk kebudayaan lokal. Budaya Batak sebagai salah satu kekayaan budaya Indonesia kerap menjadi topik perbincangan di media sosial, baik dalam bentuk apresiasi maupun kritik. Oleh karena itu, diperlukan suatu pendekatan ilmiah untuk memahami kecenderungan sentimen masyarakat terhadap budaya Batak berdasarkan data yang tersebar di ruang digital. Penelitian ini bertujuan untuk menganalisis sentimen masyarakat di Twitter terhadap budaya Batak menggunakan metode klasifikasi Naive Bayes. Data penelitian dikumpulkan melalui proses pengambilan data secara otomatis menggunakan Twitter API dengan kata kunci yang berkaitan dengan budaya Batak. Data yang diperoleh kemudian melalui tahapan pra-pemrosesan teks yang meliputi case folding, cleansing, tokenisasi, stopword removal, dan stemming guna meningkatkan kualitas data. Selanjutnya, teks diubah ke dalam bentuk numerik menggunakan metode pembobotan TF-IDF sebelum dilakukan proses klasifikasi sentimen. Dalam penelitian ini, sentimen diklasifikasikan ke dalam dua kategori, yaitu sentimen positif dan sentimen negatif. Hasil analisis terhadap 420 tweet menunjukkan bahwa mayoritas sentimen masyarakat terhadap budaya Batak bersifat positif, dengan persentase sebesar 76,67%, yang umumnya mencerminkan kebanggaan terhadap adat istiadat, tradisi, seni, dan identitas budaya Batak. Sementara itu, sentimen negatif sebesar 23,33% berkaitan dengan stereotip budaya, perbedaan pandangan antar generasi, serta anggapan bahwa beberapa praktik budaya kurang relevan dengan perkembangan zaman. Berdasarkan hasil tersebut, dapat disimpulkan bahwa algoritma Naive Bayes cukup efektif digunakan untuk analisis sentimen berbasis teks pada media sosial. Penelitian ini diharapkan dapat menjadi bahan evaluasi serta masukan bagi pemerintah, tokoh adat, dan pelaku budaya dalam upaya pelestarian dan promosi budaya Batak di era digital.

Kata Kunci: Analisis Sentimen, Twitter, Budaya Batak, Naive Bayes, Media Sosial

Abstract

Social media platforms such as Twitter have become important spaces for the public to express opinions and perspectives on various issues, including local culture. Batak culture, as one of Indonesia's rich cultural heritages, is frequently discussed on social media, both in the form of appreciation and criticism. Therefore, an analytical approach is needed to understand public sentiment toward Batak culture based on digital data. This study aims to analyze public sentiment on Twitter regarding Batak culture using the Naive Bayes classification method. The data were collected automatically through the Twitter API using keywords related to Batak culture. The collected tweets were then processed through several text preprocessing stages, including case folding, cleansing, tokenization, stopword removal, and stemming to improve data quality. Furthermore, the textual data were transformed into numerical representations using the TF-IDF weighting method before sentiment classification was performed. In this study, sentiment was classified into two categories, namely positive and negative. The analysis results of 420 tweets indicate that the majority of public sentiment toward Batak culture is positive, accounting for 76.67%, which reflects public pride in Batak traditions, customs, arts, and cultural identity. Meanwhile, 23.33% of tweets express negative sentiment, generally related to cultural stereotypes, generational differences in perspectives, and the perception that certain traditional practices are less relevant in the modern era. These findings demonstrate that the Naive Bayes algorithm is effective for sentiment analysis of social media text data. This research is expected to provide insights for government institutions, cultural practitioners, and community leaders in developing strategies for preserving and promoting Batak culture in the digital age.

Keywords: Sentiment Analysis, Twitter, Batak Culture, Naive Bayes, TF-IDF, Social Media

1. PENDAHULUAN

Perkembangan teknologi informasi dan komunikasi telah membawa perubahan signifikan dalam cara masyarakat berinteraksi, berkomunikasi, serta menyampaikan pendapat terhadap berbagai isu yang terjadi di lingkungan sekitarnya[1], [2]. Media sosial menjadi salah satu wujud nyata dari perkembangan tersebut, di mana masyarakat dapat mengekspresikan pandangan, opini, serta sikap secara terbuka dan cepat. Salah satu platform media sosial yang banyak digunakan untuk menyampaikan opini publik adalah Twitter[3], [4]. Karakteristik Twitter yang membatasi jumlah karakter dalam setiap unggahan mendorong pengguna untuk

menyampaikan pendapat secara singkat, spontan, dan langsung pada inti permasalahan, sehingga menjadikannya sumber data yang relevan untuk menganalisis opini masyarakat secara real-time[5].

Dalam konteks kebudayaan, media sosial berperan sebagai ruang publik digital yang memungkinkan terjadinya diskusi, apresiasi, maupun kritik terhadap budaya lokal[6]. Budaya tidak hanya dipandang sebagai warisan masa lalu, tetapi juga sebagai identitas sosial yang terus berkembang seiring perubahan zaman[7], [8]. Salah satu budaya lokal Indonesia yang cukup sering menjadi topik perbincangan di media sosial adalah budaya Batak. Budaya Batak dikenal memiliki kekayaan adat istiadat, bahasa, seni, musik, serta nilai-nilai tradisional yang masih dijunjung tinggi oleh masyarakatnya. Di sisi lain, budaya Batak juga tidak terlepas dari berbagai stereotip, perbedaan sudut pandang antar generasi, serta tantangan dalam pelestarian di tengah arus globalisasi dan modernisasi.

Opini masyarakat terhadap budaya Batak yang tersebar di media sosial mencerminkan persepsi publik yang beragam[9]. Sebagian masyarakat menunjukkan rasa bangga dan apresiasi terhadap kekayaan budaya Batak, sementara sebagian lainnya menyampaikan kritik atau pandangan negatif terhadap praktik budaya tertentu yang dianggap kurang relevan dengan perkembangan zaman. Opini-opini tersebut umumnya disampaikan dalam bentuk teks tidak terstruktur, seperti cuitan di Twitter, sehingga memerlukan metode khusus untuk dapat dianalisis secara sistematis dan objektif[10]. Oleh karena itu, diperlukan pendekatan berbasis teknologi yang mampu mengolah data teks dalam jumlah besar untuk memahami kecenderungan sentimen masyarakat terhadap budaya Batak[10].

Salah satu pendekatan yang banyak digunakan untuk mengolah dan menganalisis opini berbasis teks adalah analisis sentimen[10]. Analisis sentimen merupakan bagian dari bidang Natural Language Processing (NLP) yang bertujuan untuk mengidentifikasi dan mengklasifikasikan sikap atau emosi yang terkandung dalam suatu teks ke dalam kategori tertentu, seperti positif, negatif, atau netral. Dalam penelitian ini, analisis sentimen digunakan untuk mengetahui kecenderungan opini masyarakat terhadap budaya Batak berdasarkan data yang diperoleh dari Twitter[11]. Dengan analisis sentimen, opini-opini subjektif yang sebelumnya bersifat kualitatif dapat diubah menjadi informasi kuantitatif yang lebih mudah dianalisis dan diinterpretasikan[12].

Untuk melakukan klasifikasi sentimen, penelitian ini menggunakan algoritma Naive Bayes, khususnya Multinomial Naive Bayes. Algoritma ini dipilih karena memiliki keunggulan dalam mengklasifikasikan data teks berdimensi tinggi dengan proses komputasi yang relatif sederhana dan cepat[13]. Meskipun Naive Bayes menggunakan asumsi independensi antar fitur, berbagai penelitian menunjukkan bahwa algoritma ini tetap mampu memberikan performa yang baik dalam analisis sentimen, terutama pada data media sosial seperti Twitter. Selain itu, Naive Bayes sangat sesuai dikombinasikan dengan metode pembobotan kata TF-IDF, yang mampu merepresentasikan teks ke dalam bentuk numerik berdasarkan frekuensi dan tingkat kepentingan kata dalam dokumen[14], [15].

Dalam penelitian ini, data dikumpulkan dari Twitter menggunakan Twitter API dengan kata kunci yang berkaitan dengan budaya Batak. Data yang diperoleh kemudian melalui beberapa tahapan pra-pemrosesan, seperti case folding, cleansing, tokenisasi, stopword removal, dan stemming, untuk meningkatkan kualitas data dan mengurangi noise yang dapat memengaruhi hasil klasifikasi. Setelah itu, data direpresentasikan menggunakan metode TF-IDF dan diklasifikasikan ke dalam dua kategori sentimen, yaitu positif dan negatif. Pemilihan dua kategori sentimen ini didasarkan pada karakteristik data yang dianalisis, di mana opini masyarakat cenderung menunjukkan kecenderungan yang jelas terhadap objek yang dibahas.

Hasil penelitian menunjukkan bahwa mayoritas opini masyarakat di Twitter terhadap budaya Batak bersifat positif, yang mencerminkan adanya apresiasi dan kebanggaan terhadap nilai-nilai budaya, adat istiadat, serta identitas budaya Batak. Namun demikian, masih ditemukan sejumlah sentimen negatif yang berkaitan dengan stereotip budaya, perbedaan pandangan antar generasi, serta tantangan pelestarian budaya di era modern. Temuan ini menunjukkan bahwa meskipun budaya Batak masih dipandang positif oleh sebagian besar masyarakat, tetap diperlukan upaya berkelanjutan untuk menjaga relevansi dan keberlanjutan budaya tersebut di tengah perubahan sosial.

Berdasarkan latar belakang tersebut, penelitian ini bertujuan untuk menganalisis sentimen masyarakat terhadap budaya Batak di Twitter menggunakan metode Naive Bayes. Penelitian ini diharapkan dapat memberikan gambaran yang lebih objektif mengenai persepsi masyarakat terhadap budaya Batak di ruang digital, serta menjadi bahan masukan bagi pemerintah, tokoh adat, dan pelaku budaya dalam merumuskan strategi pelestarian dan promosi budaya Batak yang lebih adaptif di era digital.

2. METODOLOGI PENELITIAN

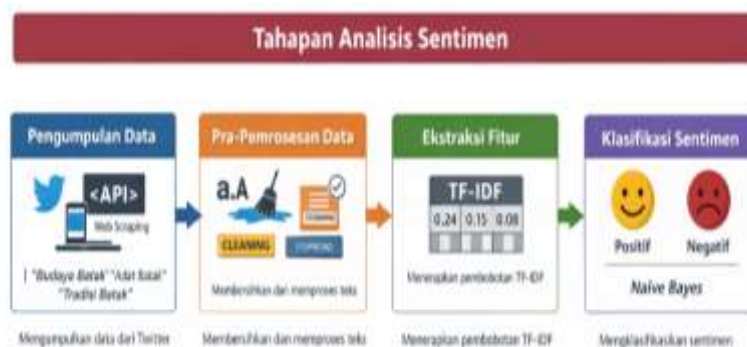
2.1 Analisis Sentimen

Sentimen analisis (sentiment analysis) atau opinion mining merupakan salah satu cabang dari text mining yang bertujuan untuk mengidentifikasi, mengekstraksi, serta mengklasifikasikan opini, persepsi, dan sikap subjektif pengguna terhadap suatu objek atau topik tertentu ke dalam kategori sentimen seperti positif, negatif, maupun netral. Teknik ini memungkinkan peneliti untuk memahami kecenderungan emosi publik melalui data teks tidak terstruktur yang dihasilkan dari berbagai platform digital. Liu (2020) menyatakan bahwa analisis sentimen berperan penting dalam pengambilan keputusan berbasis data karena mampu mengubah opini tekstual menjadi informasi kuantitatif yang dapat dianalisis secara sistematis.

Seiring meningkatnya penggunaan media sosial, analisis sentimen menjadi semakin relevan karena masyarakat cenderung mengekspresikan pendapatnya secara terbuka melalui platform daring. Feldman (2016) menjelaskan bahwa media sosial menyediakan sumber data yang besar, dinamis, serta bersifat real-time sehingga sangat cocok digunakan sebagai bahan penelitian opini publik. Dalam penelitian ini, Twitter dipilih sebagai sumber data utama karena karakteristiknya yang memungkinkan pengguna menyampaikan pendapat secara singkat, spontan, dan aktual mengenai isu-isu budaya, termasuk budaya Batak. Dengan memanfaatkan data Twitter, penelitian ini berupaya menggambarkan persepsi masyarakat terhadap budaya Batak secara objektif berdasarkan data aktual yang beredar di ruang digital.

Untuk melakukan klasifikasi sentimen, penelitian ini menerapkan algoritma Naive Bayes. Algoritma tersebut dipilih karena memiliki proses komputasi yang relatif cepat, sederhana, serta efektif dalam menangani data teks berdimensi tinggi yang umumnya mengandung noise. Kowsari et al. (2019) menegaskan bahwa Naive Bayes merupakan salah satu metode yang paling stabil dan efisien untuk tugas klasifikasi teks berskala besar, terutama pada data media sosial. Secara keseluruhan, proses sentimen analisis dalam penelitian ini dilaksanakan melalui beberapa tahapan terstruktur yang mencakup pengumpulan data, pra-pemrosesan, ekstraksi fitur, klasifikasi sentimen, serta evaluasi dan visualisasi hasil.

2.2 Tahapan Analisis Sentimen



Gambar 2. Tahapan penelitian

Tahap awal penelitian difokuskan pada proses pengumpulan data yang bertujuan memperoleh kumpulan teks yang relevan sebagai bahan analisis. Data dikumpulkan dari Twitter menggunakan teknik web scraping melalui API Twitter. Proses ini memungkinkan pengambilan data secara otomatis dalam jumlah besar dengan memanfaatkan kata kunci tertentu, seperti "Budaya Batak", "Adat Batak", dan "Tradisi Batak". Pendekatan otomatis ini dinilai lebih efisien dibandingkan metode manual karena mampu mempercepat proses akuisisi data sekaligus meningkatkan cakupan data yang diperoleh. Sari dan Yulianti (2021) menjelaskan bahwa teknik web scraping sangat efektif untuk penelitian berbasis big data karena mampu mengakses informasi secara kontinu dan terstruktur. Setelah data berhasil dikumpulkan, tahap selanjutnya adalah pra-pemrosesan data. Tahap ini memiliki peranan penting karena data mentah dari media sosial umumnya mengandung banyak elemen yang tidak relevan, seperti simbol, singkatan, tautan, maupun karakter khusus yang dapat mengganggu proses analisis. Nurfadillah dan Wibowo (2020) menyatakan bahwa pra-pemrosesan bertujuan meningkatkan kualitas data dengan menghilangkan noise sehingga informasi utama dapat diidentifikasi secara lebih jelas. Indrayani et al. (2021) juga menambahkan bahwa normalisasi kata serta reduksi fitur sangat memengaruhi akurasi model klasifikasi sentimen.

Proses pra-pemrosesan dalam penelitian ini mencakup beberapa teknik pembersihan dan normalisasi teks yang dirangkum pada Tabel 1 berikut.

Tabel 1. Pre-pemrosesan Data

No	Tahapan	Deskripsi
	Pra-Pemrosesan	
1	<i>Case Folding</i>	Mengubah semua teks menjadi huruf kecil (<i>lowercase</i>) untuk menghindari perbedaan karena kapitalisasi. Contoh: “Budaya” → “budaya”
2	<i>Cleansing</i>	Menghapus karakter yang tidak relevan seperti simbol (@, #, !), angka, tautan URL, mention, dan emoji. Contoh: @user, #adat, https://... dihapus dari teks.
3	<i>Tokenisasi</i>	Memecah kalimat menjadi potongan kata. Contoh: “budaya Batak kaya” → (budaya, Batak, kaya).
4	<i>Stopword Removal</i>	Menghapus kata-kata umum yang tidak bermakna penting. Contoh: “yang”, “itu”, “adalah”.
5	<i>Stemming</i>	Mengubah kata ke bentuk dasarnya (<i>root word</i>). Contoh: “pelestarian”, “melestarikan” → “lestar”.

Dengan diterapkannya rangkaian pra-pemrosesan tersebut, data teks menjadi lebih bersih, terstruktur, dan representatif sehingga mampu meningkatkan performa model klasifikasi secara signifikan. Tahap ini juga membantu mengurangi dimensi fitur serta mempercepat proses komputasi.

Setelah data dibersihkan, dilakukan tahap ekstraksi fitur. Pada tahap ini, teks diubah menjadi representasi numerik agar dapat diproses oleh algoritma pembelajaran mesin. Representasi numerik diperlukan karena model klasifikasi tidak dapat mengolah data berbentuk teks secara langsung. Beberapa metode yang umum digunakan antara lain Bag-of-Words, TF-IDF, serta teknik embedding seperti Word2Vec dan BERT. Goldberg (2017) menyebutkan bahwa embedding modern mampu menangkap makna semantik secara mendalam, namun dalam penelitian ini dipilih metode TF-IDF karena lebih sederhana, ringan secara komputasi, serta sesuai dengan karakteristik algoritma Naive Bayes. TF-IDF bekerja dengan memberikan bobot pada kata berdasarkan frekuensi kemunculan dalam dokumen dan tingkat keunikannya di seluruh korpus.

Tahap berikutnya adalah klasifikasi sentimen. Pada tahap ini, model pembelajaran mesin dilatih menggunakan data berlabel untuk mempelajari pola distribusi kata pada setiap kategori sentimen. Model kemudian digunakan untuk memprediksi sentimen dari data baru yang belum pernah dilihat sebelumnya. Algoritma Naive Bayes diterapkan karena kemampuannya dalam menangani data dengan jumlah fitur yang besar dan sifat sparsity yang tinggi, yang umum terjadi pada teks Twitter. Zhang et al. (2010) menunjukkan bahwa Naive Bayes tetap memberikan kinerja yang kompetitif meskipun menggunakan asumsi independensi antar fitur.

Setelah proses klasifikasi selesai, dilakukan evaluasi model untuk mengukur tingkat keberhasilan prediksi. Evaluasi dilakukan menggunakan metrik akurasi, precision, recall, dan F1-score guna memperoleh gambaran menyeluruh mengenai performa model. Selain evaluasi numerik, hasil analisis juga divisualisasikan dalam bentuk grafik batang, diagram lingkaran, word cloud, serta grafik tren waktu. Visualisasi ini bertujuan mempermudah interpretasi data sehingga pola opini masyarakat terhadap budaya Batak dapat dipahami secara intuitif. Liu (2012) menekankan bahwa visualisasi memainkan peran penting dalam menyampaikan hasil analisis data kompleks agar lebih informatif dan komunikatif bagi pembaca.

2.3 Algoritma Naive Bayes

Algoritma Naive Bayes merupakan metode klasifikasi berbasis probabilistik yang didasarkan pada Teorema Bayes dengan asumsi bahwa setiap fitur dalam data bersifat independen satu sama lain. Meskipun asumsi independensi ini sering kali tidak sepenuhnya terpenuhi pada data nyata, pendekatan tersebut tetap mampu memberikan performa yang baik, khususnya pada klasifikasi teks. Manning, Raghavan, dan Schütze (2008) menyatakan bahwa Naive Bayes termasuk algoritma yang sangat efisien karena hanya memerlukan informasi distribusi kata dan kelas tanpa parameter kompleks, sehingga proses pelatihan maupun prediksi dapat dilakukan dengan cepat.

Secara matematis, probabilitas suatu dokumen termasuk ke dalam kelas tertentu dihitung menggunakan rumus Teorema Bayes, yaitu:

$$P(C|X) = \frac{P(X|C).P(C)}{P(X)} \tag{a}$$

Dimana $P(C | X)$ menyatakan probabilitas posterior suatu dokumen termasuk kelas sentimen tertentu, $P(X | C)$ merupakan likelihood kemunculan fitur pada kelas tersebut, $P(C)$ adalah probabilitas prior, dan $P(X)$ merupakan evidence atau probabilitas keseluruhan fitur. Dalam konteks analisis sentimen, fitur berupa kumpulan kata dalam tweet, sedangkan kelas berupa label sentimen seperti positif, negatif, atau netral.

Dalam penerapannya pada data teks, tipe Naive Bayes yang paling umum digunakan adalah Multinomial Naive Bayes karena mampu memperhitungkan frekuensi kemunculan kata dalam dokumen. Pendekatan ini dinilai paling sesuai untuk representasi TF-IDF. Beberapa penelitian menunjukkan bahwa Multinomial Naive Bayes mampu menghasilkan akurasi tinggi meskipun dengan kompleksitas perhitungan yang relatif rendah. Oleh karena itu, algoritma ini sangat cocok digunakan dalam penelitian analisis sentimen berbasis media sosial yang melibatkan data dalam jumlah besar.

Dengan penerapan metode Naive Bayes yang dikombinasikan dengan pra-pemrosesan dan ekstraksi fitur yang optimal, penelitian ini diharapkan mampu menghasilkan model klasifikasi sentimen yang akurat, efisien, serta mampu menggambarkan opini masyarakat terhadap budaya Batak secara komprehensif.

3. HASIL DAN PEMBAHASAN

3.1 Pengumpulan Data

Data diperoleh melalui proses web crawling dengan memanfaatkan Twitter API sebagai sumber utama. Pengambilan data difokuskan pada cuitan yang memuat kata kunci tertentu yang relevan dengan topik, seperti: budaya Batak, adat Batak, tradisi Batak, dan kebudayaan Batak.

3.2 Filtering Data

Tahapan penyaringan dilakukan secara sistematis untuk menjamin bahwa data yang dianalisis memiliki relevansi dan kualitas yang memadai.

Tabel berikut menunjukkan tahapan filtering yang diterapkan:

Tabel 2. Filtering Data

No	Tahap Filtering	Deskripsi	Jumlah Tweet Tersisa
1.	Data awal hasil crawling	Data mentah hasil crawling dengan kata kunci Budaya Batak	775
2.	Penghapusan duplikasi	Menghapus tweet yang sama (duplikat) atau retweet	669
3.	Penghapusan tweet non-bahasa indonesia	Menggunakan filter lang='id' dan pemeriksaan manual untuk menyaring bahasa lain	670
4.	Penghapusan spam	Menghapus tweet yang berisi iklan otomatis atau promosi massal	540
5.	Filtering berdasarkan relevansi topik	Menghapus tweet yang mengandung kata kunci tetapi tidak relevan dengan topik penelitian	518
6.	Filtering tweet kosong atau yang tidak bermakna	Menghapus tweet yang terlalu pendek (< 3 kata) atau tidak memiliki makna signifikan	420

3.3 Pra-pemrosesan Data

Pra-pemrosesan data merupakan langkah awal yang krusial dalam analisis teks, terutama ketika data berasal dari media sosial seperti Twitter yang cenderung tidak terstruktur dan mengandung berbagai elemen tidak relevan. Tahapan ini bertujuan untuk menyaring dan menyederhanakan teks agar lebih mudah diproses oleh algoritma klasifikasi seperti Naive Bayes. Adapun tahapan-tahapan pra-pemrosesan data yang digunakan dalam penelitian ini meliputi:

- a. **CaseFolding**
Langkah awal ini bertujuan untuk menyamakan format penulisan dengan mengubah seluruh huruf dalam teks menjadi huruf kecil. Dengan begitu, kata-kata seperti “Budaya” dan “budaya” akan dianggap sama dan tidak dibedakan oleh sistem.
- b. **Cleansing**
Pada tahap ini, berbagai elemen yang tidak dibutuhkan dalam analisis seperti angka, simbol khusus, alamat URL, mention (@nama), hashtag (#tema), emoji, dan tanda baca akan dibersihkan dari teks agar tidak mengganggu proses pengolahan data selanjutnya.
- c. **Tokenisasi**
Proses ini memecah teks panjang menjadi bagian-bagian kata tunggal yang disebut token. Contohnya, kalimat “pelestarian budaya Batak penting” akan dipisah menjadi token-token seperti (pelestarian, budaya, Batak, penting).
- d. **Stopword Removal**
Kata-kata umum yang sering digunakan namun tidak memberikan informasi penting, seperti “yang”, “dari”, “untuk”, akan disingkirkan dari teks. Tujuannya adalah mengurangi kata yang tidak berkontribusi terhadap makna sentimen.
- e. **Stemming**
Stemming dilakukan untuk menyederhanakan kata ke bentuk dasarnya. Kata-kata seperti “melestarikan”, “pelestarian”, dan “dilestarikan” akan diubah menjadi satu bentuk dasar, misalnya “lestar”. Untuk Bahasa Indonesia proses ini dibantu dengan pustaka seperti *Sastrawi* agar hasilnya lebih relevan.

3.4 Penerapan Algoritma Naïve Bayes

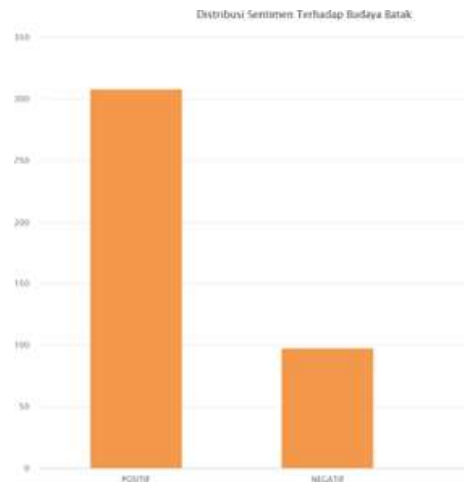
Setelah tahapan pra-pemrosesan selesai dilakukan, proses selanjutnya adalah membangun model klasifikasi dengan memanfaatkan algoritma **Multinomial Naïve Bayes**, yang dikenal efektif dalam menangani data teks berbasis frekuensi kata seperti tweet.

Langkah-langkah dalam proses pelatihan model adalah sebagai berikut:

- a. **Konversi teks menjadi vektor dengan metode TF-IDF**
Teks yang telah dibersihkan diubah ke dalam bentuk numerik menggunakan teknik **TF-IDF (Term Frequency-Inverse Document Frequency)**, yaitu metode yang memberikan bobot pada kata berdasarkan seberapa sering kata tersebut muncul dalam suatu dokumen dan seberapa jarang kata tersebut muncul di seluruh korpus.
- b. **Pembagian dataset menjadi data pelatihan dan data pengujian (80:20)**
Dataset dibagi menjadi dua bagian, yaitu 80% data digunakan untuk proses pelatihan model (**training**), dan 20% sisanya digunakan untuk menguji kinerja model (**testing**) agar hasil klasifikasi dapat divalidasi.
- c. **Pelatihan model dengan data training**
Model Multinomial Naïve Bayes kemudian dilatih menggunakan data latih untuk mengenali hubungan antara fitur kata dan kelas sentimen (positif, negatif) dalam konteks opini masyarakat terhadap budaya Batak.
- d. **Pengujian model dan evaluasi hasil klasifikasi**
Setelah pelatihan, model digunakan untuk melakukan prediksi pada data uji. Hasil prediksi tersebut kemudian dievaluasi menggunakan metrik seperti **akurasi, precision, recall, dan F1-score**, untuk mengetahui seberapa baik model dalam mengklasifikasikan sentimen yang terkandung dalam tweet mengenai budaya Batak.

3.5 Visualisasi Hasil

Visualisasi dilakukan dalam bentuk Bar Chart.



Gambar 3. Distribusi Sentimen terhadap Budaya Batak

Grafik di atas menggambarkan hasil klasifikasi sentimen masyarakat Twitter terhadap topik budaya Batak, menggunakan algoritma Naïve Bayes. Dari total 420 tweet yang dianalisis, diperoleh distribusi sentimen sebagai berikut:

- Sentimen Positif sebanyak 322 tweet (76,67%), menunjukkan bahwa mayoritas masyarakat memiliki pandangan yang positif terhadap budaya Batak. Sentimen ini umumnya mencerminkan kebanggaan terhadap kekayaan adat, tradisi, bahasa, dan seni budaya Batak yang masih dilestarikan hingga kini.
- Sentimen Negatif sebanyak 98 tweet (23,33%), menunjukkan masih adanya opini atau kritik negatif terkait budaya Batak. Sentimen ini bisa disebabkan oleh stereotip, konflik antar budaya, atau pandangan bahwa beberapa aspek tradisional dianggap sudah tidak relevan di era modern.
- Pada dataset ini tidak ditemukan tweet dengan sentimen netral, sehingga seluruh data yang dianalisis dikategorikan secara jelas ke dalam opini positif atau negatif.

Distribusi sentimen ini memberikan gambaran bahwa secara umum masyarakat memandang budaya Batak secara positif, namun tetap ada sebagian kecil yang memberikan kritik atau menyuarakan kekhawatiran. Hasil ini dapat menjadi bahan pertimbangan bagi pemerintah daerah, tokoh adat, dan pelaku budaya dalam upaya pelestarian sekaligus pembaruan budaya secara adaptif di era digital.

4. KESIMPULAN

Berdasarkan hasil penelitian yang telah dilakukan, dapat disimpulkan bahwa analisis sentimen terhadap budaya Batak dengan menggunakan algoritma Naïve Bayes mampu mengelompokkan opini masyarakat yang diperoleh dari media sosial Twitter ke dalam dua kategori utama, yaitu sentimen positif dan sentimen negatif, dengan tingkat performa klasifikasi yang cukup efektif. Proses klasifikasi ini menunjukkan bahwa algoritma Naïve Bayes dapat bekerja dengan baik dalam mengolah data teks yang bersifat tidak terstruktur, khususnya pada konten media sosial yang memiliki karakteristik bahasa yang beragam dan dinamis. Hasil pengujian terhadap 420 tweet yang dianalisis memperlihatkan bahwa sebanyak 322 tweet atau sekitar 76,67% mengandung sentimen positif, sedangkan 98 tweet atau sekitar 23,33% mengandung sentimen negatif. Dominasi sentimen positif ini mengindikasikan bahwa mayoritas masyarakat memiliki pandangan yang mendukung serta rasa bangga terhadap budaya Batak, terutama yang berkaitan dengan adat istiadat, warisan leluhur, nilai-nilai tradisional, serta identitas budaya yang masih dijunjung tinggi hingga saat ini. Meskipun demikian, keberadaan sentimen negatif yang ditemukan tidak dapat diabaikan, karena mencerminkan adanya isu-isu sosial dan budaya yang masih menjadi bahan perdebatan di tengah masyarakat. Beberapa opini bernada negatif menyoroti adanya stereotip terhadap budaya Batak, perbedaan sudut pandang antara generasi muda dan generasi tua, serta kritik terhadap praktik budaya tertentu yang dianggap kurang relevan dengan perkembangan zaman modern. Hal ini menunjukkan perlunya pendekatan yang lebih adaptif dan komunikatif dari para pemangku kepentingan budaya dalam upaya pelestarian budaya Batak, agar nilai-nilai budaya dapat

tetap lestari tanpa mengabaikan dinamika sosial yang terus berkembang. Selain itu, hasil penelitian ini juga membuktikan bahwa algoritma Naïve Bayes merupakan metode yang efisien dan andal dalam menganalisis data teks dari media sosial, khususnya Twitter. Sifat algoritma yang sederhana, cepat, serta memiliki tingkat akurasi yang cukup baik menjadikannya sangat sesuai untuk diterapkan dalam analisis sentimen berbasis opini publik terhadap isu-isu budaya lokal, termasuk budaya Batak.

REFERENCES

- [1] Viona Francesca Purba, Rouli Br Lumban Batu, Reh Bungana Beru Perangin-Angin, and Maulana Ibrahim, "Dampak Perubahan Teknologi Komunikasi Terhadap Peningkatan Kejahatan Sosial: Studi Kasus Penggunaan Media Sosial Di Era Digital," *Indones. J. Islam. Jurisprudence, Econ. Leg. Theory*, vol. 1, no. 3, pp. 477–485, 2023, doi: 10.62976/ijjel.v1i3.79.
- [2] Ilham Lucky Alamsyah, Nur Aulya, and Siti Handayani Satriya, "Transformasi Media Dan Dinamika Komunikasi Dalam Era Digital : Tantangan Dan Peluang Ilmu Komunikasi," *J. Ilm. Res. Student*, vol. 1, no. 3, pp. 168–181, 2024.
- [3] T. Fajriah, "Pengaruh Teknologi Komunikasi Terhadap Interaksi Sosial Di Era Digital," *Merdeka Indones. J. Int.*, vol. 4, no. 1, pp. 149–158, 2024.
- [4] M. R. Saputra and F. Hidayat, "Dinamika Komunikasi Persuasif dalam Media Massa: Teknik, Strategi, dan Pengaruh terhadap Perilaku Masyarakat," *J. Al-Nahyan J. Komun. dan Penyiaran Islam*, vol. 2, no. 1, pp. 62–73, 2025.
- [5] Mawaaddah, "Analisis Opini Publik tagar #JanganJadiDosen pada media sosial X (Twitter)," *Ikraith-Humaniora*, vol. 9, no. 2, pp. 408–416, 2025, [Online]. Available: <https://doi.org/10.37817/ikraith-humaniora.v9i2>
- [6] Agustinus Gulo, "Revitalisasi Budaya Di Era Digital Dan Eksplorasi Dampak Media Sosial Terhadap Dinamika Sosial-Budaya Di Tengah Masyarakat," *J. Pendidik. dan Kebud.*, vol. 3, no. 3, pp. 172–184, 2023, doi: 10.55606/juridibud.v3i3.2655.
- [7] A. R. Astari, I. Shilfa, O. A. Aulia, T. Hanafiah, and E. Purwanto, "Strategi Branding Budaya dalam Media Digital oleh Pemerintah Daerah," *J. Technol. Syst. Inf.*, vol. 2, no. 3, p. 12, 2025, doi: 10.47134/jtsi.v2i3.4334.
- [8] F. Fahma and D. Safitri, "... kesempatan media sosial terhadap budaya masyarakat lokal: Dynamics of cultural identity in the era of globalization: Challenges and opportunities for social media ...," *J. Intelek Dan Cendekiawan Nusant.*, vol. 1, no. 3, pp. 3675–3682, 2024, [Online]. Available: <https://jicnusantara.com/index.php/jicn/article/view/423%0Ahttps://jicnusantara.com/index.php/jicn/article/download/423/496>
- [9] S. Atfianti, "Pengaruh Media Sosial Terhadap Persepsi Publik dalam Seni dan Budaya," *Imajin. J. Seni*, vol. 18, no. 2, pp. 64–71, 2024.
- [10] G. E. Sagala, E. Damayanti, M. Harap, and N. Dora, "Mangaranto dalam Budaya Batak Toba: Tantangan, Dampak dan Perkembangan di Era Digital," *Aksi Kolekt. J. Pengabd.*, vol. 1, no. 01, pp. 1–9, 2024.
- [11] N. Raisa and N. Riza, "Sentimen Analisis Terhadap Opini Masyarakat Mengenai Drama Korea Pada Twitter Menggunakan Metode Naïve Bayes," *JATI (Jurnal Mhs. Tek. Inform.)*, vol. 7, no. 2, pp. 1312–1320, 2023, doi: 10.36040/jati.v7i2.6765.
- [12] R. Zulcharnain, G. Abdurrahman, and D. Daryanto, "Analisis Sentimen Ulasan Duolingo dengan Metode Algoritma Multinomial Naïve Bayes," *J. Inform. dan Teknol. Pendidik.*, vol. 5, no. 1, pp. 1–15, 2025, doi: 10.59395/jitp.v5i1.113.
- [13] N. Fibriyanti Arminda, N. Sulistiyowati, and T. Nur Padilah, "Implementasi Algoritma Multinomial Naive Bayes Pada Analisis Sentimen Terhadap Ulasan Pengguna Aplikasi Brimo," *JATI (Jurnal Mhs. Tek. Inform.)*, vol. 7, no. 3, pp. 1817–1822, 2023, doi: 10.36040/jati.v7i3.7012.
- [14] K. F. H. Holle, R. Alfianita, and H. M. Putri, "Evaluasi Teknik Preprocessing terhadap Kinerja Multinomial Naive Bayes dalam Klasifikasi Pertanyaan Insincere," *JUSTIN (Jurnal Sist. dan Teknol. Informasi)*, vol. 12, no. 4, pp. 707–715, 2024.
- [15] T. Taslim, S. Handayani, and F. Fajrizal, "Kinerja Komparatif Optimasi Algoritma Naive Bayes dalam Klasifikasi Teks untuk Uji Klinis Kanker," *J. Eksplora Inform.*, vol. 13, no. 1, pp. 113–123, 2023, doi: 10.30864/eksplora.v13i1.994.