

Analisis Sentimen Publik Terhadap Pariwisata Aceh di Media Sosial X Menggunakan Algoritma *Naive Bayes Classifier*

Susilawati*, Sri Wahyuni

Pascasarjana, Magister Teknologi Informasi, Universitas Pembangunan Pancabudi, Medan, Indonesia

Email: ¹susiyahya@gmail.com, ²sriwahyuni@dosen.pancabudi.ac.id

Email Penulis Korenspondensi: susiyahya@gmail.com

ABSTRAK – Aceh identik dengan hal negatif ditelinga masyarakat luar Aceh. Hal ini disebabkan oleh konflik bersenjata yang berkepanjangan dan juga musibah tsunami pada tahun 2004 silam. Aceh menyimpan segudang potensi wisata yang sangat menarik baik itu wisata alam, wisata Sejarah, atraksi seni dan juga wisata religious. Sampai hari ini kita masih sering mendengar hal-hal yang kurang baik yang mempengaruhi keinginan wisatawan untuk berkunjung ke Aceh. Oleh karena itu Penelitian ini bertujuan untuk menganalisa sentiment publik atau pandangan Masyarakat terhadap pariwisata Aceh menggunakan algoritma *Naive Bayes Classifier* pada platform media sosial X (Twitter). Data untuk penelitian ini diambil dari data tweet di media X atau Twitter yang dikumpulkan dengan kata kunci pariwisata Aceh dan kemudian melalui beberapa tahap pre-processing data untuk meningkatkan kualitas data, tahap-tahap yang dilalui yaitu cleaning text, case folding, normalisasi kata, tokenisasi, stopword removal dan stemming. Setelah itu baru diterapkan algoritma naive bayes untuk mengklasifikasikan sentimen tweet kedalam kategori positif dan negatif. Evaluasi model dilakukan menggunakan *confusion matrix*, *accuracy* dan *classification report*. Hasil penelitian menunjukkan bahwa *Naive Bayes* memberikan kinerja yang baik dalam mengklasifikasikan sentimen public dengan akurasi mencapai 81%. Hasil analisis ini menunjukkan bahwa pandangan masyarakat sudah mulai positif terhadap Pariwisata Aceh, sehingga ini menjadi peluang yang baik untuk pengembangan pariwisata Aceh kedepan.

Kata Kunci: Pariwisata Aceh, Text Mining, Analisis Sentimen, Naive Bayes, Media sosial (Twitter)

ABSTRACT - Aceh has been synonymous with negative perceptions among people outside the province. This is due to the prolonged armed conflict and the devastating tsunami in 2004. Despite these challenges, Aceh possesses abundant potential for tourism, including natural attractions, historical sites, cultural arts, and religious tourism. However, negative perceptions continue to influence tourists' decisions to visit Aceh. Therefore, this study aims to analyze public sentiment or public opinion towards Aceh's tourism using the Naive Bayes algorithm on the X (Twitter) social media platform. Data for this study was collected from tweets on X (Twitter) using the keyword "Aceh tourism" and then underwent several data pre-processing stages to improve data quality, including text cleaning, case folding, word normalization, tokenization, stop word removal, and stemming. Afterward, the Naive Bayes algorithm was applied to classify tweet sentiment into positive and negative categories. Model evaluation was conducted using a confusion matrix, accuracy, and classification report. The results showed that Naive Bayes performed well in classifying public sentiment with an accuracy of 81%. This analysis indicates that public perception towards Aceh's tourism has begun to shift positively, presenting a promising opportunity for the future development of Aceh's tourism sector.

Keyword: Text Mining, Sentiment Analysis, Tourism, Aceh, Naive Bayes, tweet

1. PENDAHULUAN

Provinsi Aceh memiliki pesona alam yang indah memukau dan kekayaan budaya yang khas, menyimpan potensi besar untuk menjadi destinasi wisata yang digemari baik wisatawan domestik maupun mancanegara. Aceh yang pernah dilanda konflik bersenjata yang lama dan bencana tsunami dasyat pada tahun 2004. Hal ini memberikan kesan bahwa Aceh tidak aman dan nyaman untuk dikunjungi oleh Masyarakat luar Aceh baik nasional ataupun mancanegara. Ditambah lagi dengan keistimewaan Aceh yang menerapkan Syariat Islam dan hukum cambuk atau qanun jinayat bagi yang melanggar syariat membuat banyak muncul pandangan masyarakat luar yang beragam terhadap aceh.

Provinsi Aceh terletak di ujung paling barat Nusantara yang masyarakatnya terdiri dari berbagai suku dan ras yang beragam, Provinsi Aceh memiliki 13 suku asli yaitu: Suku Aceh, Suku Tamiang, Suku Gayo, Suku Alas, Suku Kluet, Suku Julu, Suku Pakpak, Suku Aneuk Jamee, Suku Sigulai, Suku Lekon, Suku Devayan, Suku Haloban, Suku Nias. Keberagaman ini memberikan daya tarik tersendiri bagi wisatawan. Keragaman suku tersebut juga menghasilkan berbagai ragam kebudayaan dan bahasa yang berbeda.[1]

Keberagaman budaya dan bahasa itu melahirkan destinasi wisata seni, atraksi budaya dan adat istiadat yang menarik bagi wisatawan. Selain itu di Aceh terdapat banyak warisan peninggalan masa lalu, yaitu Naskah/benda kuno, Masjid Raya Baiturrahman, Makam kuno, Kerkhof Peutjoet dan cagar budaya dan lainnya. Ditambah dengan bangunan monumental seperti museum tsunami, situs tsunami kapal PLTD apung, situs tsunami kapal diatas rumah, situs tsunami masjid lampuuk, situs tsunami masjid ulee lheu, kuburan massal korban tsunami.[2]

Di era digital seperti sekarang ini masyarakat berinteraksi dan mendapatkan informasi dengan mudah Internet dan media sosial yang berkembang pesat menjadi penyedia berbagai informasi opini publik yang beragam. Saat ini orang menggunakan internet sebagai wadah mereka mengeluarkan pendapat dan isi pikiran mereka tentang isu yang menarik dan sedang populer. Sebuah ulasan ataupun komentar di media sosial dapat menjadi informasi sangat penting pada saat membuat Keputusan.[3]

Twitter atau X merupakan salah satu *platform* media sosial yang sering digunakan untuk media komunikasi dan mendapatkan informasi serta menyampaikan berbagai opini. Opini mereka disampaikan melalui tweet yang dituliskan

untuk mengomentari satu topik. *Twitter* juga dapat digunakan untuk tujuan survei, karena *tweet* yang dihasilkan dapat memberikan informasi tentang sentimen yang dipandang oleh Masyarakat.[4]

Analisis sentimen merupakan suatu kegiatan bidang studi yang berupa analisis dari suatu pendapat, penilaian, sikap, evaluasi serta emosi seseorang terhadap suatu isu, problem, layanan serta produk dari personal, kelompok atau even tertentu.[5] Analisis Sentimen adalah suatu teknik mengekstrak data teks untuk mendapatkan informasi tentang sentimen bernilai positif, netral maupun negatif. Analisis sentimen diberikan oleh pengguna internet pada media sosial untuk memberikan suatu penilaian atau opini pribadi.[6] Metode analisis sentimen membantu pemerintah atau pengusaha mengidentifikasi sentimen dari konsumen atau masyarakat melalui berbagai media *online* seperti media sosial, survei, dan ulasan situs web *e-commerce*. Dengan metode ini para pemegang kebijakan dapat memperoleh informasi tentang pandangan dan pendapat konsumen terhadap produk atau layanan yang mereka tawarkan.[7]

Penambangan teks mampu mengambil data dalam jumlah besar dan kemudian diproses untuk mendapatkan informasi yang berguna dari kumpulan teks. Aplikasi yang dapat dilakukan dengan text mining berupa analisis sentimen pada media sosial, ekstraksi informasi untuk mengklasifikasi teks, dan peringkasan teks untuk meringkas teks panjang menjadi bentuk 1 atau 2 paragraf terus dikembangkan.[8] *Text mining* bertujuan menganalisis pendapat, sentimen, evaluasi, sikap, penilaian, dan emosi seseorang pembicara atau penulis berkenaan dengan suatu topik, produk, layanan, organisasi, individu dan hal populer lainnya.[9] *Text mining* diartikan sebagai suatu teknik untuk pengambilan informasi dari sejumlah data *unsupervised* dari sebuah topik tertentu yang memiliki kualitas tinggi agar dapat diperoleh data- data permasalahan dalam teks.[10]

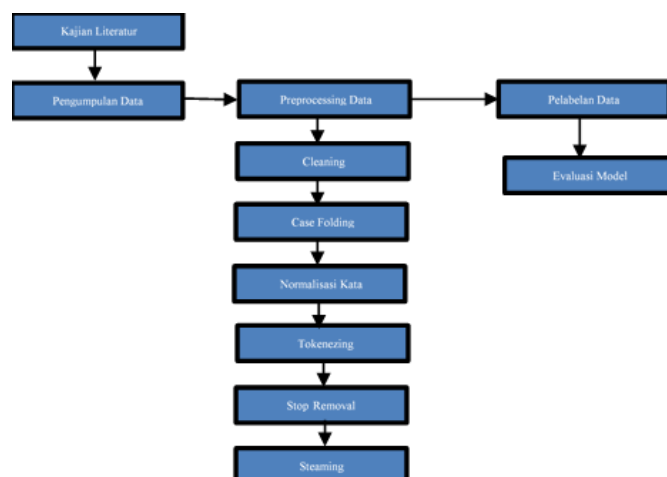
Naïve Bayes Classifier adalah algoritme yang terdapat dalam teknik data mining yang menerapkan teori *Naïve Bayes* dalam klasifikasi. *Naïve Bayes Classifier* yaitu suatu metode pengklasifikasian berakar pada teorema bayes. Keuntungan menggunakan metode *Naïve Bayes Classifier* adalah metode ini hanya memerlukan nilai atau jumlah data pelatihan (Training Data) yang kecil sebagai penentu estimasi parameter yang nantinya diperlukan dalam proses klasifikasi data[11] Pada penelitian sebelumnya yang dilakukan oleh (Fatimah, S., Purwanto, E., & Permatasari, H. ,2023)[4] menunjukkan akurasi sebesar 70,2%%, presisi 67% dan recall 100%, berdasarkan pengujian yang telah dilakukan dengan 500 data menggunakan metode naive bayes didapatkan hasil klasifikasi netral berjumlah 122 data, positif 302 data dan negatif 76 data. Selanjutnya pada penelitian yang dilakukan oleh (Ahmad Adita Shiddiq, Aris Wahyu Murdiyanto, & Arif Himawan,2023)[5] yang menggunakan jumlah data sebanyak 11.282 berhasil mendapatkan akurasi dengan nilai 81,67 % pada data training dan 85 % untuk data testing. Sementara itu pada penelitian yang lainnya yang diteliti oleh (Sari, F. V., & Wibowo, A.,2019)[6]. Hasil penelitian menunjukkan bahwa metode *Naïve Bayes* tanpa penambahan fitur mampu mengklasifikasi sentimen dengan nilai akurasi sebesar 96,44%, sementara jika ditambahkan fitur pembobotan *tf-idf* disertai konversi ikon emosi mampu meningkatkan nilai akurasi menjadi 98%.

Berdasarkan dari tiga penelitian yang menjadi referensi jurnal penelitian dengan hasil penelitiannya dijelaskan diatas, menunjukkan bahwa algoritma *Naïve Bayes* memiliki akurasi yang cukup tinggi dan mudah dilakukan untuk analisis sentiment dibandingkan algoritma klasifikasi yang lainnya.

Tujuan dari penelitian ini adalah untuk melihat pandangan masyarakat terhadap pariwisata Aceh setelah aceh mengalami berbagai peristiwa seperti konflik senjata, musibah tsunami dan juga pandemi covid. Selain itu penerapan syariat Islam dengan qanun syariahnya yang membuat aceh sering disorot oleh media nasional dan menjadi isu nasional yang memunculkan sentimen yang beragam. Penelitian ini menggunakan algoritma *naive bayes* dan hasilnya nanti bisa memberikan dampak yang baik bagi pariwisata Aceh.

2. METODOLOGI PENELITIAN

Adapun Metodologi penelitian melalui beberapa tahapan seperti yang dijelaskan pada gambar dibawah ini :



Gambar 1. Tahapan Penelitian

1. Kajian Literatur

Tahap ini bertujuan untuk menggali pemahaman tentang penelitian yang akan dilakukan dengan mengumpulkan, dan membaca, berbagai sumber seperti jurnal ilmiah, buku, artikel, dan laporan penelitian sebelumnya yang berkaitan dengan analisis sentiment public di media X dan algoritma *Naïve Bayes* sehingga bisa metode penelitian yang paling sesuai untuk permasalahan yang diteliti.

2. Pengumpulan Data

Crawling merupakan pengunduhan atau pengambilan data dengan tujuan untuk mengumpulkan data dari suatu database misalnya data yang diunduh dari server Twitter berupa user dan tweet termasuk atributnya[12]. Crawling data media X atau Twitter dilakukan melalui google colab dengan menggunakan bahasa pemrograman *python*[13].

Library python yang digunakan untuk *crawling* data ini adalah *library tweepy*. Pustaka *tweepy* di *python* dapat digunakan untuk mengakses API Publik yang disediakan oleh Twitter. Selama proses pengambilan data menggunakan Twitter API dilakukan proses *retweet filter*. Filter *retweet* ini bertujuan untuk menghindari redundansi data karena pada saat pengumpulan data dengan Twitter API terdapat data yang memiliki isi yang sama. API Twitter menganggap *retweet* sebagai tweet biasa. Tweet yang di *Retweet* diawali dengan tulisan RT di depannya. Proses filter *retweet* dilakukan dengan menambahkan "-filter: retweets" saat menggunakan *api.search* dari *tweepy*[8]

Data tweet yang dikumpulkan dalam rentang waktu Januari 2023 sampai dengan Oktober 2024, menggunakan kata kunci "Pariwisata Aceh". Menggunakan perintah seperti terlihat pada gambar 2 dibawah ini.

Gambar 2. Perintah Crawling Data

Hasil
crawling
data
tersebut
disimpan

```
# Crawl Data

filename = 'WisataAceh.csv'
search_keyword = 'Pariwisata Aceh since:2023-01-01 until:2024-10-31 lang:id'
limit = 1000

!npx -y tweet-harvest@2.6.1 -o "{filename}" -s "{search_keyword}" --tab "LATEST" -l {limit} --token {twitter_auth_token}
```

kedalam file csv dengan jumlah data sebanyak 570 tweet.

3. Pre-processing Data

Setelah melakukan pengumpulan data, dilakukan proses pengolahan data atau *preprocessing*, tahapan ini mencakup kegiatan membangun data dan dilanjutkan kegiatan pembersihan data agar dapat dilakukan pengelolaan ketahap selanjutnya.[13] Berikut adalah Langkah-langkah yang dilakukan pada tahap *pre-processing* data.

- *Cleaning*
Data *tweet* melewati tahap *cleaning text* untuk menghilangkan karakter *noise* dengan menghapus angka, simbol, URL, *hashtag*, lalu diseragamkan seluruhnya ke dalam huruf kecil pada tahap *lowering case*. [14]
- *Case folding*
adalah proses mengubah semua huruf dalam teks menjadi huruf kecil atau besar. Tujuannya adalah untuk menyederhanakan data dan menghindari masalah karena perbedaan kapitalisasi. Misalnya, kata "Mobil" dan "mobil" akan dianggap sebagai kata yang sama setelah dilakukan *case folding*.
- *Normalisasi Kata*
Pada tahap ini dilakukan proses mengubah kata-kata yang memiliki ejaan yang tidak sesuai dengan Bahasa Indonesia atau singkatan menjadi bentuk yang lebih baku. Contoh : kata "yg" menjadi "yang", "nggak" menjadi "tidak" sehingga lebih mudah dikenal oleh algoritma.
- *Tokenizing*
Tokenizing adalah proses membagi teks ulasan menjadi kata per kata atau token. Tokenizing memudahkan komputer dalam memproses kata per kata dalam teks ulasan.
- *Stop Word Removal*
Stop word removal adalah proses menghapus kata-kata umum yang tidak memiliki makna yang signifikan untuk analisis seperti "dan", "atau", "yang", "pada", dan sebagainya.
- *Steaming*
steaming adalah proses mengubah setiap kata menjadi bentuk dasar kata atau akarnya kata seperti "menjual" menjadi "jual". Tujuannya adalah untuk menyederhanakan bentuk kata untuk mudah dibaca oleh algoritma.

4. Algoritma Naïve Bayes Classifier

Naïve Bayes adalah suatu pengklasifikasian probabilitas simpel yang melakukan perhitungan terhadap sekumpulan probabilitas dengan penjumlahan frekuensi dan gabungan nilai dari data set yang didapatkan Keuntungan penggunaan *Naïve Bayes* yaitu metode *Naïve Bayes* hanya membutuhkan jumlah data pelatihan Training Data yang kecil untuk menentukan estimasi parameter yang dibutuhkan dalam proses pengklasifikasian[15]

5. Evaluasi Model.

Tahap ini dilakukan untuk mengetahui akurasi dan kinerja model yang sudah dibangun. Evaluasi ini dilakukan dengan menggunakan tiga metrik utama yaitu *Confusion matrik*, *accurayi* dan *classification report*.

3. HASIL DAN PEMBAHASAN

3.1 Hasil Pengumpulan Data

Pengumpulan data tweet tentang pariwisata Aceh dilakukan melalui *Google Colab* dengan menggunakan Bahasa pemrograman *python* dengan *library tweepy* menggunakan token otentikasi sebagai kunci akses yang digunakan untuk berinteraksi dengan *API Twitter*. Dengan token ini, kita dapat mengakses data *Twitter* secara *programatik*, seperti *tweet*, profil pengguna, dan lainnya. Kemudian mengisi jumlah limit data yang akan diambil menggunakan kata kunci Pariwisata Aceh dengan rentang waktu tanggal, bulan dan tahun dari tweet yang akan diambil. Jumlah tweet yang berhasil dikumpulkan sebanyak 507 tweet yang relevan untuk digunakan pada algoritma *Naïve Bayes*.

Tabel 1 : Hasil *Crawling Data*

index	username	full_text
0	S41dZ41n4I	@AbuYus79147958 @cakimiNOW @EmHasanuddin @zainul_munas @Brigade01Ar Komunitas Pemandu Wisata Aceh Timur merapatkan barisan dukung Gus Imin the ne #GusIminPresiden #PartaiKebangkitanBangsa @cakimiNOW @EmHasanuddin @zai @Mursyidan11 @HasanZulfahmi @AbuYus79147958
1	MaulidaLid45530	Komunitas Pemandu Wisata Aceh Timur merapatkan barisan dukung Gus Imin the ne #GusIminPresiden #PartaiKebangkitanBangsa @cakimiNOW @EmHasanuddin @zai @Mursyidan11 @HasanZulfahmi https://t.co/9ax6V1O8zs
2	AbuYus79147958	Komunitas Pemandu Wisata Aceh Timur merapatkan barisan dukung Gus Imin the ne #GusIminPresiden #PartaiKebangkitanBangsa @cakimiNOW @EmHasanuddin @zai @Mursyidan11 @HasanZulfahmi https://t.co/67TxwKux91
3	KabarBanten72	3 Destinasi Wisata di Banda Aceh yang Populer dan Membuatmu Terpukau https://t.co/9ax6V1O8zs
4	Dinur55908759	3 Destinasi Wisata di Banda Aceh yang Populer dan Membuatmu Terpukau https://t.co/9ax6V1O8zs
5	wisataacehID	Jejak Rimba Raya: Wisata Terbaru di Aceh Selatan untuk Liburan Menyatu dengan Alam https://t.co/w0nGqCZShR
6	AliandaNia	Cafe dan Harga Tiket Wisata Pantai Lampu uk Banda Aceh https://t.co/JtlbliDLw
7	JRKAceh	Vima dan Hajiati Duta Wisata Aceh Singkil 2023 Ini Juara II serta III Pukak dan Nungki https://t.co/WXENH9KUKv

3.2 Tahap Pre-processing data

Tahap awal dari *pre-processing* data adalah data dibersihkan terlebih dahulu untuk menghapus ulasan yang berisikan simbol @, simbol #, RT, URL, HTTP dan emoji. Setelah dibersihkan selanjutnya dilakukan proses *case folding* yaitu mengubah semua huruf besar menjadi huruf kecil. Hasil dari proses *case folding* bisa dilihat pada tabel dibawah ini.

Tabel 2 : Hasil *Case Folding*

Hasil <i>Cleaning</i>	<i>Case Folding</i>
Komunitas pemandu wisata aceh timur merapatkan barisan dukung gus imin the next presiden	komunitas pemandu wisata aceh timur merapatkan barisan dukung gus imin the next presiden
Destinasi wisata di banda aceh yang populer dan membuatmu terpukau	destinasi wisata di banda aceh yang populer dan membuatmu terpukau

Tahap berikutnya adalah normalisasi kata dimana pada tahap ini dilakukan proses mengubah setiap ulasan menjadi kata baku. Hasil proses pada tahap ini bisa dilihat pada tabel berikut.

Tabel 3 : Hasil Normalisasi Kata

Normalisasi	Tokenizing
wisata pantan terong di aceh tengah panorama d...	[wisata, pantan, terong, di, aceh, tengah, pan.
pj bupati aceh besar apresiasi program bakti s...	[pj, bupati, aceh, besar, apresiasi, program, .

Setelah melalui tahap normalisasi kata, tahap selanjutnya adalah *tokenizing* yaitu tahap menguraikan ulasan menjadi token-token atau kata-kata. Proses dilakukan menggunakan *library NLTK*. Hasil proses ini bisa dilihat pada table berikut.

Tabel 4 : Hasil Tokenizing

Case Folding	Normalisasi
cafe dan harga tiket wisata pantai lampu uk banda aceh . wisata dan kuliner surga durian unggul di aceh	cafe dan harga tiket wisata pantai lampu uk banda aceh wisata dan kuliner surga durian unggul di aceh

Adapun tahap berikutnya adalah proses penghapusan *stopword removal*. Pada tahap ini dilakukan penghapusan data yang dianggap tidak penting. Proses ini juga dibantu oleh *library NLTK*. Hasil proses dari tahap ini bisa dilihat pada tabel dibawah ini.

Tabel 5 : Hasil Stopword Removal

Tokenizing	stopword removal
[quality, time, ala, negeri, serambi, mekkah, [indrapatra, benteng, peninggalan, kerajaan, h.	quality, time, ala, negeri, serambi, mekkah indrapatra, benteng, peninggalan, kerajaan, h..

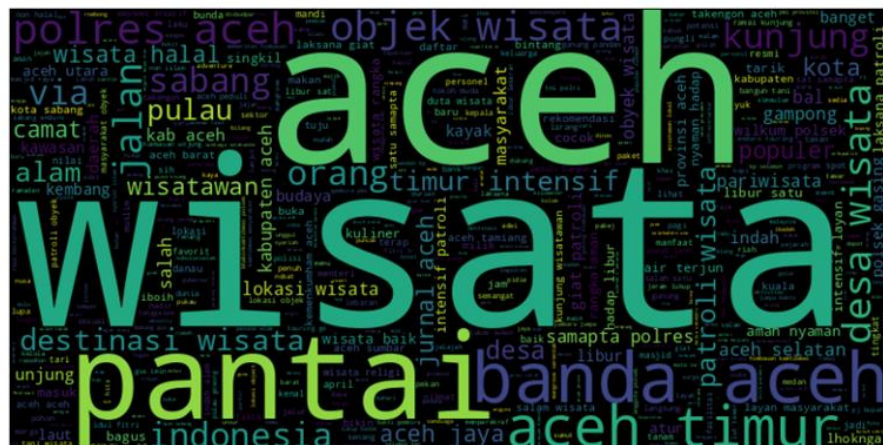
Tahap terakhir dari *pre-processing* data adalah tahap *stemming* data. Pada tahap ini dilakukan proses mengubah kata-kata berimbuhan menjadi kata dasar. Proses ini dilakukan dengan bantuan *library Sastrawi dan swifter*. Hasil dari proses ini dapat dilihat pada tabel dibawah ini.

Tabel 6 : Hasil Stemming

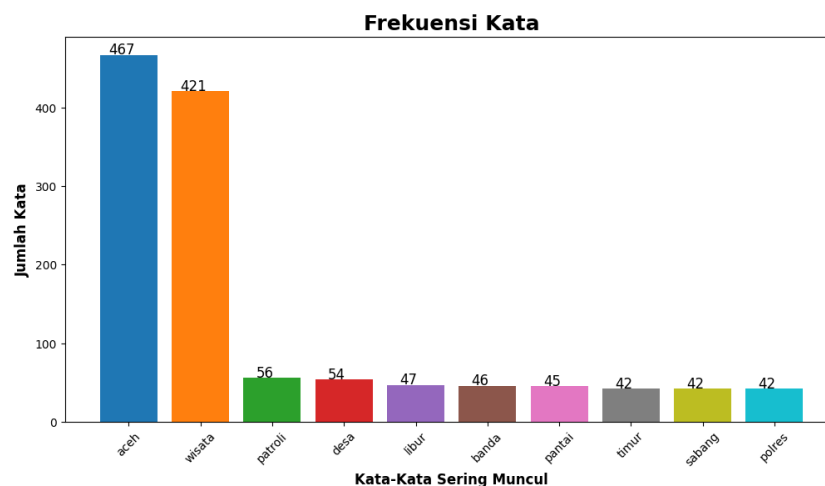
Stopwords	Stemming
aceh, peluang, usaha, bagus, peternakan, p ertanian, dagang, wisata, budaya, mohon, sdm, ya, benahi, sukses, ya, tom	aceh peluang usaha bagus ternak tani dagang wisata budaya mohon sdm ya benah sukses ya tom
indrapatra, benteng, peninggalan, kerajaa n, hindu, aceh, situs, wisata, sejarah, abad, l ampau	indrapatra benteng tinggal raja hindu aceh situs wisata sejarah abad lampau

Setelah data teks melalui proses *pre-processing*, data tersebut telah dibersihkan, dinormalisasi, dan diubah menjadi bentuk yang lebih terstruktur. Hal ini membuat data siap untuk dianalisis lebih lanjut, salah satunya dengan menggunakan *WordCloud*.

WordCloud adalah representasi visual dari kata-kata yang paling sering muncul dalam suatu teks. Kata-kata yang lebih sering muncul akan ditampilkan dengan ukuran font yang lebih besar, sehingga memberikan gambaran cepat tentang topik utama dan kata kunci yang dominan dalam teks tersebut.[16] Berikut hasil *wordcloud* bisa dilihat pada gambar dibawah ini.



Gambar 3 : Hasil dari wordcloud



Gambar 4 : Grafik Kata-kata yang paling sering muncul

3.3. Pelabelan Data

Tahap berikutnya adalah menentukan hasil output untuk mengidentifikasi opini positif atau negatif. Proses mengidentifikasi hasil positif dan negatif menggunakan *lexicon-base*. Kode lexicon-basenya bisa dilihat pada gambar dibawah ini.

```
import pandas as pd

# Fungsi untuk menentukan sentimen
def determine_sentiment(text):
    positive_count = sum(1 for word in text.split() if word in positive_lexicon)
    negative_count = sum(1 for word in text.split() if word in negative_lexicon)
    if positive_count > negative_count:
        return "Positif"
    elif positive_count < negative_count:
        return "Negatif"

# Baca kamus leksikon positif dan negatif
positive_lexicon = set(pd.read_csv("positive.tsv", sep="\t", header=None)[0])
negative_lexicon = set(pd.read_csv("negative.tsv", sep="\t", header=None)[0])

def replace_none_sentiment(sentiments):
    replace_flag = "Positif"
    for i in range(len(sentiments)):
        if sentiments[i] is None:
            sentiments[i] = replace_flag
            replace_flag = "Negatif" if replace_flag == "Positif" else "Positif"
    return sentiments
```

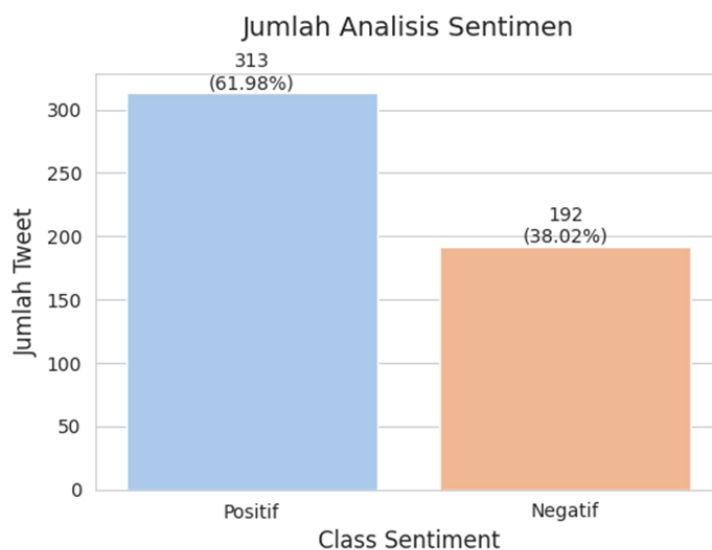
Gambar 5 : kode *lexicon-based*

Fungsi dari kode *lexicon-based* diatas adalah untuk melakukan analisis sentiment yang cepat dan akurat pada teks. Metode ini bekerja dengan membandingkan kata-kata dalam teks dengan sebuah kamus sentimen (lexicon) yang telah ditentukan sebelumnya.[17] Pendekatan metode lexicon base efektif digunakan untuk menganalisis teks dari sosial media. Kemampuannya dalam mengenali sentiment positif dan negative serta mengukur intensitas dalam teks. Hasil dari pelabelan sentiment bisa dilihat pada tabel dibawah ini.

Tabel 7 : Hasil Pelabelan Data

Stemming Data	Sentimen
berburu durian musangking lhoong	Positif
pulau beting aceh destinasi wisata daerah pulau babi duduk tinggal sana kurang nyaman karena penambangan pasir	Negatif

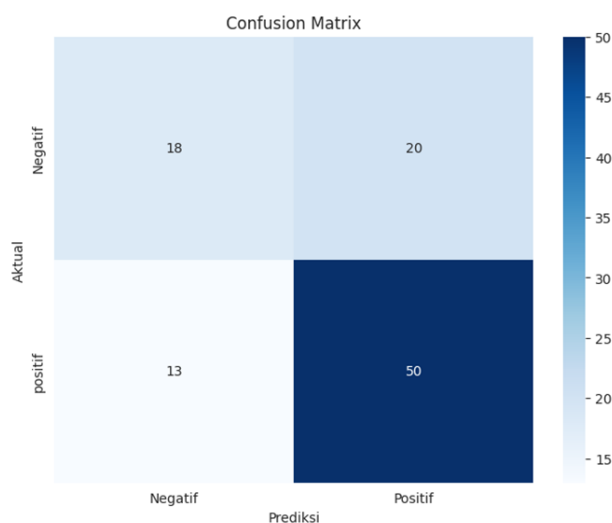
Total dari keseluruhan dari setiap kelas sentiment (*positif dan negative*) dalam dataset yang digunakan divisualisasikan menggunakan diagram batang dengan bantuan *library seaborn*. Hasil visualisasinya bisa dilihat pada gambar dibawah ini.



Gambar 6 : Class Sentiment

3.4 Algoritma Naïve Bayes Classifier

Algoritma yang digunakan dalam mengklasifikasikan opini adalah algoritma *Naïve Bayes*. [18]. Teknik klasifikasi yang dijalankan adalah menggunakan pembagian dataset uji dan dataset latih dengan rasio 80% : 20% dengan jumlah data latih sebanyak 404 dan jumlah data uji sebanyak 101. Hasil *confusion Matrik* dari Algoritma Naïve Bayes ini bisa dilihat pada gambar dibawah ini.



Gambar 7 : Confusion Matrik

Hasil dari *classification report* yang diperoleh melalui algoritma naïve bayes memberikan informasi mengenai kinerja model dalam mengklasifikasi data ulasan publik. *classification report* ini terdiri dari precision, recall dan F1-score untuk masing-masing sentiment yaitu positif dan negative. Adapun hasil lengkap dari *classification report* Naïve Bayes pada data latih dan pada data uji bisa dilihat pada gambar dibawah ini.

Tabel 8 : Classification Report Data Uji

Accuracy Naïve Bayes pada data Uji :				0,67
Classification Report:				
	precision	recall	f1-score	support
Negatif	0,58	0,47	0,52	38
Positif	0,71	0,79	0,57	63
Accuracy			0,67	101
Macro avg	0,65	0,63	0,64	101
Weighted avg	0,66	0,67	0,67	101

Pada tabel 8 terlihat hasil evaluasi model Naïve Bayes yang digunakan untuk analisis sentiment public tentang pariwisata Aceh di media X. Model ini mencapai akurasi 67% atau 0,67. Pada *Classification report* terdapat dua kategori sentiment yaitu positif dan negatif. Untuk sentiment negatif model menggambarkan nilai preciosion sebesar 0,58, recall 0,47 dan f1-score 0,52 dengan jumlah data pendukung sebanyak 38 data. Sementara untuk kategori sentiment positif precion call mencapai 0,71, recall 0,79 dan f1-score 0,57 dengan dukungan data sebanyak 63 data dari total sampel sebesar 101 data sampel.

Tabel 9 : Classification Report Data Latih

Accuracy Naïve Bayes pada data Latih :				0,95
Classification Report:				
	precision	recall	f1-score	support
Negatif	0,95	0,67	0,78	154
Positif	0,72	0,96	0,82	250
Accuracy			0,95	404
Macro avg	0,95	0,95	0,95	404
Weighted avg	0,95	0,95	0,95	404

Pada tabel 9 terlihat hasil evaluasi model Naïve Bayes yang dihasilkan dari data latih yang mencapai akurasi 95% atau 0,95. Hasil *Classification report* untuk dua kategori sentiment yaitu positif dan negatif menggambarkan nilai precision sebesar 0,95, recall 0,67 dan f1-score 0,78 dengan jumlah data pendukung sebanyak 154 data. Sementara untuk kategori sentiment positif precision mencapai 0,72, recall 0,96 dan f1-score 0,82 dengan dukungan data sebanyak 250 data dari total sampel sebesar 250 data sampel.

4. KESIMPULAN

Algoritma Naive Bayes berhasil mengklasifikasikan sentimen publik terhadap pariwisata Aceh dengan akurasi yang cukup baik, yaitu sebesar 67% pada data uji. Meskipun akurasi pada data latih lebih tinggi 95%, hal ini adalah wajar karena model telah dilatih secara khusus pada data latih. Hasil analisis menunjukkan adanya dominasi sentimen positif terhadap pariwisata Aceh. Hal ini menunjukkan bahwa pandangan Masyarakat luar terhadap Aceh semakin baik. Hal ini tentu sangat menguntungkan bagi pengembangan pariwisata Aceh. Model masih mengalami kesulitan dalam mengklasifikasikan sentimen negatif dengan akurasi yang tinggi. Hal ini terlihat dari nilai *precision*, *recall*, dan *F1-score* yang lebih rendah untuk kelas negatif dibandingkan dengan kelas positif. Dampak dari penelitian ini adalah kita bisa mengetahui perkembangan pandangan atau opini publik terhadap Aceh atau khususnya pariwisata Aceh semakin positif. Hal ini bisa menjadi input bagi pelaku usaha dibidang pariwisata dalam mengembangkan usahanya. Analisis sentimen secara berkala juga dapat membantu pemerintah dan stakeholder pariwisata untuk memantau perubahan persepsi publik terhadap pariwisata Aceh.

DAFTAR PUSTAKA

- [1] Asgha, A. Y., & Novita, S. (2023), Konsep Pengembangan Kawasan Pariwisata sebagai Upaya Meningkatkan Promosi Pariwisata di Provinsi Aceh. *ArtComm: Jurnal Komunikasi Dan Desain*, 6(1), 1–8. <https://doi.org/10.37278/artcomm.v6i1.612>
- [2] Herizal, H., Rasanjani, S., & Muhkrijal, M. (2021). Kebijakan Kepariwisata di Provinsi Aceh: Peluang dan Tantangan. *Jurnal Public Policy*, 7(1), 21. <https://doi.org/10.35308/jpp.v7i1.3341>.
- [3] Hayuningtyas, R. Y., & Sari, R. (2019). ANALISIS SENTIMEN OPINI PUBLIK BAHASA INDONESIA TERHADAP WISATA TMII MENGGUNAKAN NAÏVE BAYES DAN PSO. *Jurnal Techno Nusa Mandiri*, 16(1), 37–42. <https://doi.org/10.33480/techno.v16i1.115>
- [4] Fatimah, S., Purwanto, E., & Permatasari, H. (2023). Analisis Sentimen Twitter Tentang Wisata di Kota Solo. *Techno.Com*, 22(4), 854–869. <https://doi.org/10.33633/tc.v22i4.9154>.
- [5] Ahmad Adita Shiddiq, Aris Wahyu Murdiyanto, & Arif Himawan. (2023). Analisis Sentimen Transfer Pemain Klub La Liga Spanyol Pada Bursa Transfer Musim Dingin Eropa Di Twitter. *INDONESIAN JOURNAL ON DATA SCIENCE*, 1(1), 1–19. <https://doi.org/10.30989/ijds.v1i1.859>.
- [6] Sari, F. V., & Wibowo, A. (2019). Analisis Sentimen Pelanggan Toko Online Jd.Id Menggunakan Metode Naïve Bayes Classifier Berbasis Konversi Ikon Emosi. *Jurnal SIMETRIS*, 10(2), 681–686.
- [7] Ahmad, S., Ridwan, A. M., & Setyawan, G. D. (2023). ANALISIS SENTIMEN PRODUCT TOOLS & HOME MENGGUNAKAN METODE CNN DAN LSTM. *TEKNOKOM*, 6(2), 133–140. <https://doi.org/10.31943/teknokom.v6i2.154>
- [8] Hermawan, A., Jowensen, I., Junaedi, J., & Edy. (2023). Implementasi Text-Mining untuk Analisis Sentimen pada Twitter dengan Algoritma Support Vector Machine. *JST (Jurnal Sains Dan Teknologi)*, 12(1), 129–137. <https://doi.org/10.23887/jstundiksha.v12i1.52358>.
- [9] Mailoa, F. F. (2021). Analisis sentimen data twitter menggunakan metode text mining tentang masalah obesitas di indonesia. *Journal of Information Systems for Public Health*, 6(1), 44. <https://doi.org/10.22146/jisph.44455>
- [10] Afdal, M., & Elita, L. R. (2022). PENERAPAN TEXT MINING PADA APLIKASI TOKOPEDIA MENGGUNAKAN ALGORITMA K-NEAREST NEIGHBOR. *Jurnal Ilmiah Rekayasa Dan Manajemen Sistem Informasi*, 8(1), 78. <https://doi.org/10.24014/rmsi.v8i1.16595>
- [11] Surapati, U., & Zulkarnain, A. Y. (2023). Implementasi Metode Naïve Bayes Untuk Mendeteksi Hate Speech Pada Twitter. *INTECOMS: Journal of Information Technology and Computer Science*, 6(2), 830–837. <https://doi.org/10.31539/intecom.v6i2.7678>
- [12] Rendy Rian Chrisna Putra, Eza Budi Perkasa, Tri Sugihartono, Ahmad Pakih Alkayess, Irsad Dwi Sandro, & Royhan Indallah. (2023). Komparasi Naive Bayes dengan Support Vector Machine dalam Analisis Sentimen Aplikasi MyPertamina. *SATIN - Sains Dan Teknologi Informasi*, 9(2), 90–99. <https://doi.org/10.33372/stn.v9i2.1042>
- [13] Aryanti, P. G., & Santoso, I. (2023). Analisis Sentimen Pada Twitter Terhadap Mobil Listrik Menggunakan Algoritma Naive Bayes. *IKRA-ITH Informatika: Jurnal Komputer Dan Informatika*, 7(2), 133–137. Retrieved from <https://journals.upi-yai.ac.id/index.php/ikraith-informatika/article/view/2821>

- [14] Shevira, S., Suarjaya, I. M. A. D., & Buana, P. W. (2022). Pengaruh Kombinasi dan Urutan Pre-Processing pada Tweets Bahasa Indonesia. *JITTER : Jurnal Ilmiah Teknologi Dan Komputer*, 3(2), 1074. <https://doi.org/10.24843/jtrti.2022.v03.i02.p06>
- [15] Wiwik Sri Rahayu Ginantra, N. L., Arifah, F. N., Wijaya, A. H., Septarini, R. S., Dewa, N. A., Yudhi Ardiana, P., Effendy, F., Hazriani, A. I., Sari, I. Y., Gustiana, Z., Prianto, C., Gustian, D., & Negara, E. S. (2021). *Data Mining dan Penerapan Algoritma*. Yayasan Kita Menulis.
- [16] Hanafiah, A., Nasution, A. H., Arta, Y., Wandri, R., Nasution, H. O., & Mardafora, J. (2023). Sentimen Analisis Terhadap Customer Review Produk Shopee Berbasis Wordcloud Dengan Algoritma Naïve Bayes Classifier. *INTECOMS: Journal of Information Technology and Computer Science*, 6(1), 230–236. <https://doi.org/10.31539/intecom.v6i1.5845>
- [17] Prayuda, M. W., & Permana, A. A. (2022). Penerapan Metode Lexicon Based untuk Menganalisis Sentimen Terhadap Mudik Lebaran. *Jurnal Minfo Polgan*, 11(2), 137–143. Retrieved from <https://doi.org/10.33395/jmp.v11i2.12348>
- [18] Krisdiyanto, T. (2021). Analisis Sentimen Opini Masyarakat Indonesia Terhadap Kebijakan PPKM pada Media Sosial Twitter Menggunakan Naïve Bayes Clasifiers. *Jurnal CoreIT: Jurnal Hasil Penelitian Ilmu Komputer Dan Teknologi Informasi*, 7(1), 32. <https://doi.org/10.24014/coreit.v7i1.12945>