

Analisis Sentimen Opini Warga X terhadap Banjir di Sumatera Menggunakan Naive Bayes

Paulina Gorat¹, Frans Steven Pakpahan², Sardo Pardingotan Sipayung³

^{1,2,3}Jurusan Teknik Informatika, Universitas Katolik Santo Thomas Medan

e-mail: [1paulinagorat3@gmail.com](mailto:paulinagorat3@gmail.com), [2franssteven191@gmail.com](mailto:franssteven191@gmail.com),

[3pinsarsiphom@gmail.com](mailto:pinsarsiphom@gmail.com)

Intisari

Penelitian ini bertujuan untuk menganalisis sentimen opini masyarakat terhadap peristiwa banjir di wilayah Sumatera berdasarkan data dari media sosial X. Banjir merupakan bencana alam yang sering terjadi di Sumatera dan menimbulkan berbagai respons masyarakat yang banyak disampaikan melalui media sosial. Media sosial X dipilih sebagai sumber data karena bersifat terbuka dan real-time sehingga dapat merepresentasikan opini publik secara luas.

Data penelitian terdiri dari 1.030 tweet berbahasa Indonesia yang dikumpulkan melalui proses crawling menggunakan API resmi X dengan kata kunci terkait banjir di Sumatera. Setelah dilakukan pembersihan data, diperoleh 873 tweet yang kemudian diproses melalui tahapan text mining, meliputi preprocessing teks, pelabelan sentimen secara manual, serta pembagian data menjadi data latih dan data uji. Data latih berjumlah 650 tweet, sedangkan data uji sebanyak 223 tweet.

Klasifikasi sentimen dilakukan menggunakan algoritma Naive Bayes dengan bantuan perangkat lunak RapidMiner. Evaluasi model dilakukan menggunakan confusion matrix dengan metrik accuracy, precision, recall, dan F1-score. Hasil penelitian menunjukkan bahwa algoritma Naive Bayes mampu memberikan kinerja yang cukup baik dalam mengklasifikasikan sentimen. Selain itu, hasil analisis menunjukkan bahwa opini masyarakat terhadap peristiwa banjir di wilayah Sumatera didominasi oleh sentimen negatif. Penelitian ini diharapkan dapat memberikan gambaran mengenai persepsi masyarakat dan menjadi bahan pertimbangan dalam pengambilan kebijakan penanggulangan bencana.

Kata kunci— Analisis Sentimen, Banjir Sumatera, Media Sosial X, Naive Bayes, Text Mining

Abstract

This study aims to analyze public sentiment regarding flooding in Sumatra based on data from social media platform X. Flooding is a frequent natural disaster in Sumatra and elicits a variety of public responses, many of which are expressed through social media. Social media platform X was chosen as the data source because it is open and real-time, allowing it to broadly represent public opinion.

The research data consists of 1,030 Indonesian-language tweets collected through a crawling process using the official API for X, using keywords related to flooding in Sumatra. After data cleaning, 873 tweets were obtained, which were then processed through text mining stages, including text preprocessing, manual sentiment labeling, and dividing the data into training and test data. The training data consisted of 650 tweets, while the test data consisted of 223 tweets.

Sentiment classification was performed using the Naive Bayes algorithm with the assistance of RapidMiner software. Model evaluation was performed using a

confusion matrix with accuracy, precision, recall, and F1-score metrics. The results showed that the Naive Bayes algorithm performed quite well in sentiment classification. Furthermore, the analysis shows that public opinion regarding the flooding in Sumatra is dominated by negative sentiment. This research is expected to provide insight into public perceptions and inform disaster management policymaking.

Keywords— *Sentiment Analysis, Sumatra Floods, Social Media X, Naive Bayes, Text Mining*

PENDAHULUAN

Banjir adalah suatu peristiwa daratan yang terbenam akibat peningkatan volume air yang melampaui kapasitas suatu wilayah[1]. Peristiwa ini merupakan salah satu bencana alam yang paling sering terjadi di Indonesia dan menimbulkan dampak yang signifikan terhadap kehidupan masyarakat. Banjir tidak hanya menyebabkan kerusakan fisik pada permukiman dan infrastruktur, tetapi juga mengganggu aktivitas ekonomi, sosial, serta menimbulkan risiko terhadap keselamatan dan kesehatan masyarakat.

Pulau Sumatera termasuk wilayah yang memiliki tingkat kerawanan banjir cukup tinggi, terutama pada periode musim hujan. Tingginya curah hujan, meluapnya sungai, serta perubahan tata guna lahan di daerah hulu dan perkotaan menjadi faktor utama yang memicu terjadinya banjir, termasuk banjir bandang yang menimbulkan kerusakan parah. Beberapa peristiwa banjir bandang di Sumatera telah menyebabkan kerugian besar dan menarik perhatian luas dari masyarakat, baik secara lokal maupun nasional.

Selain dampak fisik, peristiwa banjir juga memunculkan berbagai respons dan opini dari masyarakat. Masyarakat terdampak kerap menyampaikan keluhan, kritik, serta harapan terhadap penanganan banjir yang dilakukan oleh pihak terkait. Seiring dengan perkembangan teknologi informasi, media sosial menjadi sarana utama bagi masyarakat untuk mengekspresikan opini dan pengalaman mereka secara terbuka dan cepat. Media sosial X, dengan karakteristiknya yang bersifat real-time dan terbuka, memungkinkan masyarakat untuk menyampaikan informasi dan pendapat mengenai peristiwa banjir secara langsung.

Opini masyarakat yang tersebar di media sosial X membentuk kumpulan data teks dalam jumlah besar yang sulit dianalisis secara manual. Oleh karena itu, diperlukan pendekatan analisis yang sistematis dan efisien untuk memahami kecenderungan opini publik. Salah satu pendekatan yang dapat digunakan adalah analisis sentimen, yaitu teknik untuk mengklasifikasikan teks berdasarkan muatan emosionalnya ke dalam kategori sentimen positif dan negatif.

Berdasarkan latar belakang tersebut, penelitian ini bertujuan untuk menerapkan algoritma Naive Bayes dalam menganalisis sentimen opini masyarakat terhadap peristiwa banjir di wilayah Sumatera berdasarkan data dari media sosial X. Melalui analisis sentimen ini, diharapkan dapat diperoleh gambaran objektif mengenai persepsi masyarakat terhadap peristiwa banjir serta menjadi bahan pertimbangan bagi pemerintah dan pihak terkait dalam merumuskan kebijakan dan strategi penanggulangan bencana yang lebih tepat dan berbasis data.

LANDASAN TEORI

1.1 Media Sosial X

Media sosial X merupakan platform mikroblog yang memungkinkan pengguna menyampaikan opini, informasi, dan pengalaman secara real-time dalam bentuk teks singkat[2]. Karakteristiknya yang terbuka dan interaktif menjadikan X sebagai sumber data yang relevan untuk merepresentasikan opini publik terhadap peristiwa aktual, termasuk bencana banjir. Data yang dihasilkan bersifat tidak terstruktur dan beragam secara linguistik, sehingga memerlukan proses pengolahan teks sebelum dianalisis menggunakan metode pembelajaran mesin.

1.2 Analisis Sentimen

Analisis sentimen merupakan cabang dari pemrosesan bahasa alami (Natural Language Processing) yang bertujuan mengidentifikasi kecenderungan opini atau sikap seseorang terhadap suatu topik atau peristiwa[3]. Analisis ini mengklasifikasikan teks ke dalam kategori sentimen, seperti positif dan negatif, berdasarkan muatan emosional yang terkandung di dalamnya.

Dalam konteks media sosial, analisis sentimen memungkinkan pengolahan opini masyarakat secara otomatis dan berskala besar. Pendekatan ini efektif untuk memahami respons publik terhadap isu sosial dan bencana, termasuk banjir, karena data yang dihasilkan bersifat aktual dan mencerminkan kondisi lapangan.

1.3 Text Mining

Text mining merupakan pengembangan dari data mining yang berfokus pada pengolahan dan analisis data berbentuk teks tidak terstruktur[4]. Tujuan utama text mining adalah mengekstraksi informasi dan pola yang berguna dari kumpulan data teks dalam jumlah besar sehingga data dapat diolah menjadi informasi yang bermakna.

Dalam penelitian analisis sentimen, text mining berperan penting karena mampu menangani karakteristik data media sosial yang singkat dan beragam. Proses text mining umumnya meliputi pembersihan data, preprocessing teks, ekstraksi fitur, dan penerapan algoritma klasifikasi.

1.4 Preprocessing Data Teks

Preprocessing data teks merupakan tahap awal yang bertujuan membersihkan dan menyiapkan data agar dapat diproses secara optimal oleh algoritma klasifikasi[5]. Data teks dari media sosial sering mengandung noise seperti simbol, singkatan, dan bahasa tidak baku, sehingga perlu dilakukan pembersihan dan normalisasi.

Tahapan preprocessing yang diterapkan dalam penelitian ini meliputi case folding, tokenisasi, stopword removal, dan stemming. Proses ini bertujuan mengurangi fitur yang tidak relevan dan meningkatkan kualitas data sehingga klasifikasi sentimen menggunakan algoritma Naive Bayes dapat dilakukan secara lebih efektif.

1.5 Naive Bayes

Naive Bayes merupakan algoritma klasifikasi probabilistik yang bekerja berdasarkan Teorema Bayes dengan asumsi bahwa setiap fitur bersifat

independen[6]. Meskipun asumsi ini sederhana, Naive Bayes terbukti efektif dalam pengolahan data teks berdimensi tinggi.

Dalam analisis sentimen, algoritma ini menghitung probabilitas kemunculan kata pada setiap kelas sentimen dan menentukan kelas dengan nilai probabilitas terbesar sebagai hasil klasifikasi. Keunggulan Naive Bayes terletak pada kecepatan komputasi dan kemampuannya mengolah teks pendek seperti tweet, sehingga sesuai digunakan dalam penelitian ini.

Secara matematis, klasifikasi Naive Bayes didasarkan pada Teorema Bayes sebagai berikut:

$$P(C | X) = \frac{P(X | C) \times P(C)}{P(X)}$$

dengan:

- $P(C | X)$ adalah probabilitas data X termasuk ke dalam kelas C ,
- $P(X | C)$ adalah probabilitas kemunculan data X pada kelas C ,
- $P(C)$ adalah probabilitas awal (prior) dari kelas C ,
- $P(X)$ adalah probabilitas kemunculan data X .

Dalam klasifikasi teks, data X terdiri dari sekumpulan fitur kata (x_1, x_2, \dots, x_n). Dengan asumsi independensi antar fitur, perhitungan probabilitas dapat disederhanakan menjadi:

$$P(C | X) = P(C) \prod_{i=1}^n P(x_i | C)$$

Kelas dengan nilai probabilitas terbesar akan dipilih sebagai hasil klasifikasi. Dalam penelitian ini, algoritma Naive Bayes digunakan untuk mengklasifikasikan tweet ke dalam dua kelas sentimen, yaitu positif dan negatif, berdasarkan probabilitas kemunculan kata-kata dalam masing-masing kelas.

METODE PENELITIAN

2.1. Tahapan Penelitian

Penelitian ini dilakukan melalui beberapa tahapan yang disusun secara sistematis untuk mencapai tujuan penelitian. Tahapan tersebut meliputi pengumpulan data, preprocessing teks, pelabelan sentimen, pembagian dataset, pemodelan menggunakan algoritma Naive Bayes, serta evaluasi kinerja model. Alur penelitian ini bertujuan memastikan bahwa data yang dianalisis telah diolah dengan baik sehingga hasil klasifikasi sentimen dapat dipertanggungjawabkan secara ilmiah.



Gambar 2.1 Diagram Alur Tahapan Penelitian Sentimen Banjir Sumatera Menggunakan Naïve Bayes

2.2. Pengumpulan Data

Pengumpulan data pada penelitian ini menggunakan media sosial X sebagai sumber data utama. Data diperoleh melalui API resmi X yang memungkinkan pengambilan data secara terstruktur dengan menggunakan kata kunci “Banjir Sumatera”. Kata kunci tersebut dipilih karena merepresentasikan peristiwa banjir di wilayah Sumatera dan banyak digunakan oleh masyarakat dalam menyampaikan opini serta informasi terkait kejadian tersebut.

Dari proses pengambilan data, diperoleh sebanyak 1.030 tweet berbahasa Indonesia. Data yang dikumpulkan masih berupa data mentah yang mengandung elemen tidak relevan, seperti duplikasi dan karakter khusus, sehingga dilakukan proses pembersihan sebelum dianalisis. Data hasil pembersihan selanjutnya digunakan pada tahapan text mining dan klasifikasi sentimen menggunakan algoritma Naive Bayes.

```

filename = "Banjir sumatera.csv"
search_keyword = " Banjir Sumatera lang:id"
limit = 1000
    
```

Gambar 2.2 Proses Pengambilan Data Dari Twitter

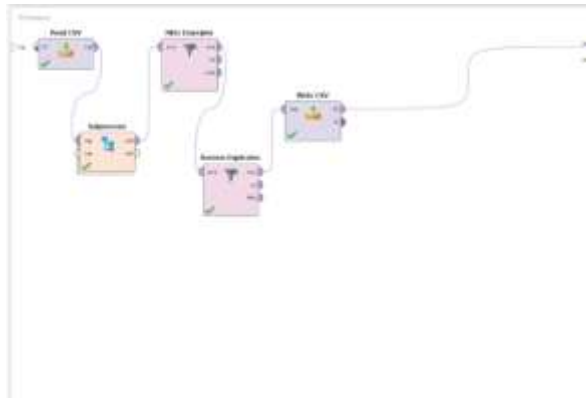
2.3. Cleaning Data

Data yang bersumber dari media sosial X umumnya masih mengandung elemen non-teks dan informasi tambahan yang tidak relevan terhadap analisis sentimen, sehingga perlu dilakukan proses pembersihan data sebelum dianalisis lebih lanjut. Pembersihan data bertujuan meningkatkan kualitas data serta mengurangi noise yang dapat memengaruhi validitas hasil penelitian[7].

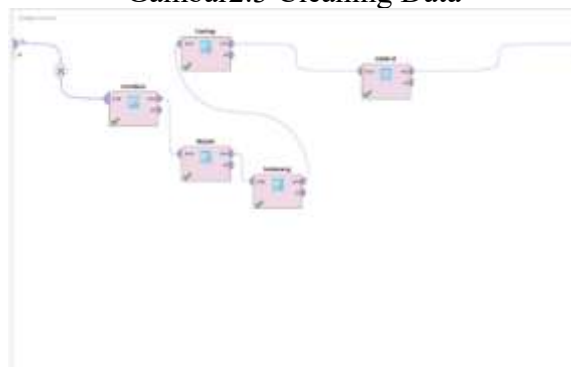
Pada penelitian ini, data mentah hasil crawling masih mengandung elemen seperti mention akun, URL, hashtag, tanda baca, simbol khusus, angka, serta kata atau frasa yang tidak mencerminkan sentimen. Oleh karena itu, dilakukan proses data cleaning menggunakan RapidMiner dengan menghapus elemen-

elemen tersebut, termasuk karakter non-alfabet, spasi berlebih, dan tweet duplikat atau yang tidak mengandung opini.

Setelah proses pembersihan, jumlah data berkurang dari 1.030 tweet menjadi 873 tweet yang bersih dan relevan. Data hasil cleaning selanjutnya digunakan pada tahap pelabelan sentimen dan preprocessing teks agar proses klasifikasi sentimen menggunakan algoritma Naive Bayes dapat berjalan secara lebih optimal.



Gambar2.3 Cleaning Data



Gambar2.5 Penghapusan Elemen-Elemen Teks

2.4. Pelabelan Sentimen

Pelabelan sentimen merupakan proses pemberian label pada setiap data teks berdasarkan makna dan konteks yang terkandung di dalamnya untuk keperluan klasifikasi sentimen [8]. Pada penelitian ini, pelabelan sentimen dilakukan secara manual untuk memastikan ketepatan penentuan kelas sentimen, mengingat variasi bahasa informal dan singkatan yang banyak digunakan pada media sosial X.

Tweet yang telah melalui proses pembersihan data kemudian dikelompokkan ke dalam dua kategori sentimen, yaitu positif dan negatif. Sentimen positif diberikan pada tweet yang berisi dukungan atau penilaian positif terhadap peristiwa banjir, sedangkan sentimen negatif diberikan pada tweet yang mengandung keluhan, kritik, atau ketidakpuasan.

Dari total 873 tweet, sebanyak 650 tweet digunakan sebagai data latih yang terdiri dari 410 tweet bersentimen negatif dan 240 tweet bersentimen positif, sedangkan 223 tweet lainnya digunakan sebagai data uji untuk mengevaluasi kinerja model. Kualitas pelabelan sentimen berpengaruh langsung terhadap kinerja dan akurasi algoritma Naive Bayes dalam proses klasifikasi.

2.5. Preprocessing Data

Text preprocessing merupakan tahap lanjutan setelah pelabelan sentimen yang bertujuan menyiapkan data teks agar dapat diproses secara optimal oleh algoritma klasifikasi. Data dari media sosial X bersifat tidak terstruktur dan mengandung noise, sehingga diperlukan preprocessing untuk meningkatkan kualitas data sebelum analisis sentimen dilakukan. Pada penelitian ini, text preprocessing dilakukan menggunakan RapidMiner untuk menghasilkan data teks yang lebih konsisten dan siap digunakan pada tahap ekstraksi fitur dan klasifikasi.

Tahapan text preprocessing yang diterapkan meliputi:

a. Case Folding

Case folding dilakukan dengan mengubah seluruh teks menjadi huruf kecil untuk menyeragamkan penulisan kata. Tahapan ini bertujuan agar perbedaan penggunaan huruf besar dan kecil tidak dianggap sebagai kata yang berbeda, sehingga perhitungan frekuensi kata menjadi lebih akurat.

b. Tokenisasi

Tokenisasi merupakan proses pemecahan teks menjadi kata-kata (token) agar setiap kata dapat dianalisis secara individual. Tahapan ini memudahkan pembentukan fitur dan mendukung proses klasifikasi sentimen[9].

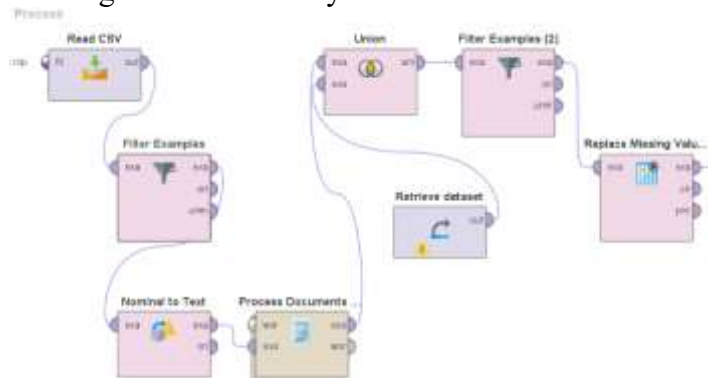
c. Stopword Removal

Stopword removal bertujuan menghapus kata-kata umum yang tidak berpengaruh terhadap penentuan sentimen. Penghapusan stopwords membantu mengurangi fitur yang tidak relevan dan meningkatkan efisiensi proses klasifikasi.

d. Stemming

Stemming dilakukan untuk mengubah kata berimbuhan menjadi kata dasar sehingga variasi kata dengan makna yang sama dapat disatukan. Tahapan ini membantu menghasilkan fitur yang lebih konsisten dan meningkatkan efektivitas klasifikasi sentimen[10].

Melalui tahapan text preprocessing ini, data tweet menjadi lebih terstruktur dan siap digunakan dalam proses ekstraksi fitur serta klasifikasi sentimen menggunakan algoritma Naive Bayes.



Gambar2.4 Tahap Preprocessing Data

2.6. Representasi Fitur

Ekstraksi fitur merupakan tahap penting dalam analisis sentimen yang bertujuan mengubah data teks hasil preprocessing menjadi representasi numerik agar dapat diproses oleh algoritma klasifikasi. Algoritma Naive Bayes tidak dapat mengolah teks mentah secara langsung, sehingga diperlukan representasi fitur yang bersifat kuantitatif.

Pada penelitian ini, metode Bag of Words (BoW) digunakan untuk ekstraksi fitur dengan merepresentasikan teks berdasarkan frekuensi kemunculan kata tanpa memperhatikan urutan kata. Setiap kata diperlakukan sebagai fitur dengan nilai yang ditentukan dari jumlah kemunculannya dalam tweet.

Metode Bag of Words dipilih karena sesuai untuk data tweet yang bersifat singkat dan memiliki struktur bahasa sederhana. Selain itu, metode ini bersifat sederhana, mudah diterapkan, dan umum digunakan dalam klasifikasi teks[11]. Proses ekstraksi fitur dilakukan menggunakan RapidMiner dengan mengubah data teks menjadi matriks dokumen-term, sehingga setiap tweet direpresentasikan dalam bentuk vektor numerik yang dapat digunakan oleh algoritma Naive Bayes dalam proses klasifikasi sentimen.

2.7. Klasifikasi Metode Naive Bayes

Klasifikasi sentimen pada penelitian ini dilakukan menggunakan algoritma Naive Bayes untuk mengelompokkan tweet ke dalam dua kelas sentimen, yaitu positif dan negatif. Algoritma ini dipilih karena efektif dalam mengolah data teks berdimensi tinggi dan sesuai untuk data teks pendek seperti tweet.

Proses klasifikasi dilakukan setelah tahap ekstraksi fitur menggunakan metode Bag of Words, di mana setiap tweet direpresentasikan dalam bentuk vektor numerik berdasarkan frekuensi kata. Model Naive Bayes menghitung probabilitas setiap kelas sentimen berdasarkan data latih yang telah diberi label dengan asumsi independensi antar kata.

Model dilatih menggunakan 650 data latih yang terdiri dari 410 tweet bersentimen negatif dan 240 tweet bersentimen positif, kemudian diuji menggunakan 223 data uji. Hasil klasifikasi menunjukkan proporsi sentimen negatif sebesar 0,631 dan sentimen positif sebesar 0,369, yang menunjukkan bahwa opini masyarakat terhadap peristiwa banjir di wilayah Sumatera didominasi oleh sentimen negatif.

2.8. Evaluasi Model

Evaluasi model dilakukan untuk menilai kinerja algoritma Naive Bayes dalam mengklasifikasikan sentimen opini masyarakat terhadap peristiwa banjir di wilayah Sumatera. Evaluasi bertujuan mengukur kemampuan model dalam memprediksi kelas sentimen secara tepat pada data yang tidak digunakan pada tahap pelatihan.

Pada penelitian ini, evaluasi dilakukan menggunakan 223 data uji yang telah dipisahkan dari data latih. Penggunaan data uji bertujuan untuk memastikan kemampuan generalisasi model terhadap data baru.

2.8.1 Confusion Matrix

Metode evaluasi yang digunakan adalah confusion matrix, yaitu tabel yang membandingkan hasil prediksi model dengan label sentimen yang sebenarnya[12]. Confusion matrix memberikan informasi mengenai jumlah prediksi yang benar dan salah, yang terdiri dari empat komponen,

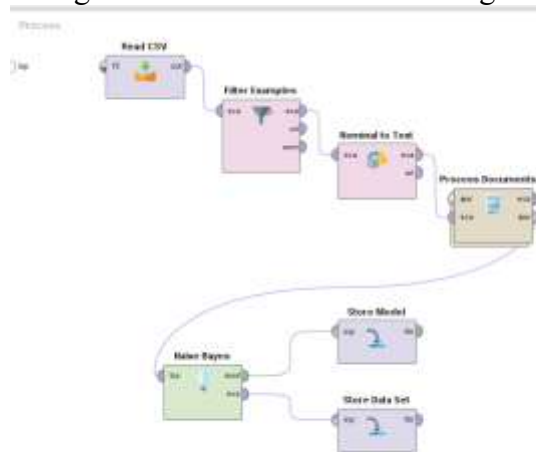
yaitu True Positive (TP), True Negative (TN), False Positive (FP), dan False Negative (FN).

Confusion matrix digunakan sebagai dasar untuk menghitung metrik evaluasi kinerja model, yaitu accuracy, precision, recall, dan F1-score.

2.8.2 Metode Pengukuran Kinerja

Berdasarkan nilai confusion matrix, kinerja model dievaluasi menggunakan beberapa metrik, yaitu accuracy, precision, recall, dan F1-score. Accuracy menunjukkan tingkat ketepatan model dalam mengklasifikasikan data uji secara keseluruhan. Precision mengukur ketepatan prediksi model pada suatu kelas sentimen, sedangkan recall menunjukkan kemampuan model dalam mengenali seluruh data yang termasuk dalam kelas sentimen tersebut. F1-score digunakan untuk memberikan penilaian yang seimbang terhadap precision dan recall, terutama pada kondisi distribusi kelas yang tidak seimbang.

Hasil evaluasi digunakan untuk menilai efektivitas algoritma Naive Bayes dalam analisis sentimen data tweet. Evaluasi ini menunjukkan bahwa model mampu mengklasifikasikan sentimen opini masyarakat terhadap peristiwa banjir di wilayah Sumatera dengan cukup baik serta mendukung temuan dominasi sentimen negatif.



Gambar3.5 Proses Pelatihan Model Naïve Bayes

HASIL DAN PEMBAHASAN

3.1. Hasil Penelitian

Penelitian ini bertujuan menganalisis sentimen opini masyarakat terhadap peristiwa banjir di wilayah Sumatera berdasarkan data media sosial X menggunakan algoritma Naive Bayes. Data dikumpulkan melalui API X dengan kata kunci “Banjir Sumatera” dan menghasilkan 1.030 tweet. Setelah proses pembersihan data, diperoleh 873 tweet yang digunakan dalam penelitian.

Dari total data tersebut, 650 tweet digunakan sebagai data latih dan 223 tweet sebagai data uji. Pelabelan sentimen dilakukan secara manual ke dalam dua kategori, yaitu positif dan negatif, dengan distribusi sentimen yang didominasi oleh sentimen negatif.

Proses klasifikasi dilakukan melalui tahapan preprocessing teks dan ekstraksi fitur menggunakan metode Bag of Words dengan bantuan

RapidMiner. Hasil klasifikasi menunjukkan proporsi sentimen negatif sebesar 0,631 dan sentimen positif sebesar 0,369, yang mengindikasikan dominasi opini negatif terhadap peristiwa banjir.

Kinerja model dievaluasi menggunakan confusion matrix dengan metrik accuracy, precision, recall, dan F1-score untuk menilai efektivitas algoritma Naive Bayes dalam mengklasifikasikan sentimen.

3.2. Confusion Matrix

Confusion matrix digunakan untuk melihat kinerja model Naive Bayes dalam mengklasifikasikan sentimen opini masyarakat terhadap peristiwa banjir di wilayah Sumatera. Evaluasi dilakukan menggunakan data uji sebanyak 223 tweet yang terdiri dari dua kelas sentimen, yaitu positif dan negatif. Hasil confusion matrix disajikan pada Tabel 1.

	Aktual Negatif	Aktual Positif
Prediksi Negatif	110 (TP)	32 (FP)
Prediksi Positif	31 (FN)	50 (FN)

Tabel 1. Hasil Confusion Matrix

Berdasarkan Tabel X, model Naive Bayes mampu mengklasifikasikan 110 tweet bersentimen negatif dengan benar, sementara 31 tweet negatif masih salah diklasifikasikan sebagai sentimen positif. Pada kelas sentimen positif, model berhasil mengklasifikasikan 50 tweet dengan benar, namun 32 tweet positif masih salah diprediksi sebagai sentimen negatif. Hasil ini menunjukkan bahwa model memiliki kecenderungan lebih baik dalam mengenali sentimen negatif dibandingkan sentimen positif, yang sejalan dengan dominasi kelas negatif pada dataset penelitian.

3.3. Pembahasan

a. Analisis Kinerja Model Naive Bayes

Berdasarkan hasil evaluasi, algoritma Naive Bayes menunjukkan kinerja yang cukup baik dalam mengklasifikasikan sentimen masyarakat terhadap peristiwa banjir di wilayah Sumatera berdasarkan data tweet media sosial X. Setelah melalui tahapan preprocessing dan ekstraksi fitur menggunakan metode Bag of Words, model mampu mengenali pola kata pada sentimen positif dan negatif.

Nilai akurasi menunjukkan bahwa sebagian besar tweet dapat diklasifikasikan dengan benar. Namun, karena distribusi data sentimen tidak seimbang, dengan jumlah sentimen negatif lebih dominan, akurasi perlu ditinjau bersama metrik evaluasi lainnya. Model cenderung lebih baik dalam mengenali sentimen negatif yang sering muncul dalam pembahasan banjir.

b. Analisis Precision, Recall, dan Pengaruh Ketidakseimbangan Data

Hasil evaluasi menunjukkan bahwa nilai precision pada kelas sentimen dominan cukup tinggi, yang menandakan bahwa prediksi model pada kelas tersebut relatif akurat. Model mampu mengklasifikasikan tweet yang berisi keluhan dan dampak negatif banjir dengan baik.

Sebaliknya, nilai recall pada kelas sentimen minoritas masih rendah, yang menunjukkan bahwa sebagian tweet pada kelas tersebut belum berhasil dikenali. Kondisi ini dipengaruhi oleh ketidakseimbangan data,

sehingga model lebih fokus mengenali sentimen negatif dibandingkan sentimen positif. Oleh karena itu, penggunaan precision dan recall penting untuk menilai kinerja model secara lebih menyeluruh.

c. Analisis F1-Score

Nilai F1-score menunjukkan keseimbangan antara precision dan recall dalam kinerja model. Nilai F1-score yang belum optimal menandakan bahwa kemampuan model dalam mengenali seluruh variasi sentimen masih terbatas, terutama pada kelas sentimen minoritas.

Keterbatasan ini dipengaruhi oleh penggunaan metode Bag of Words yang tidak mempertimbangkan konteks dan urutan kata. Oleh karena itu, meskipun Naive Bayes memberikan hasil yang cukup baik, pengembangan metode lebih lanjut masih diperlukan untuk meningkatkan keseimbangan kinerja model.

d. Interpretasi Sentimen Masyarakat terhadap Peristiwa Banjir di Sumatera

Hasil analisis sentimen menunjukkan bahwa opini masyarakat terhadap peristiwa banjir di wilayah Sumatera didominasi oleh sentimen negatif. Hal ini mencerminkan besarnya dampak banjir terhadap kehidupan masyarakat, seperti kerusakan lingkungan dan terganggunya aktivitas sehari-hari.

Media sosial X menjadi sarana bagi masyarakat untuk menyampaikan keluhan dan kritik terkait kondisi banjir dan penanganannya. Meskipun demikian, sentimen positif juga muncul dalam bentuk empati dan dukungan terhadap upaya penanggulangan banjir. Hasil penelitian ini memberikan gambaran umum mengenai respons masyarakat terhadap banjir di Sumatera yang dapat dimanfaatkan sebagai bahan pertimbangan dalam penyusunan kebijakan penanggulangan bencana.

KESIMPULAN

Berdasarkan hasil penelitian, dapat disimpulkan bahwa algoritma Naive Bayes dapat diterapkan secara efektif dalam analisis sentimen opini masyarakat terhadap peristiwa banjir di wilayah Sumatera menggunakan data media sosial X. Tahapan text mining yang meliputi data cleaning, preprocessing teks, pelabelan sentimen, dan ekstraksi fitur menggunakan metode Bag of Words berperan penting dalam mendukung kinerja model klasifikasi.

Hasil klasifikasi menunjukkan bahwa opini masyarakat terhadap peristiwa banjir di wilayah Sumatera didominasi oleh sentimen negatif, yang mencerminkan besarnya dampak banjir terhadap kehidupan masyarakat. Meskipun demikian, sentimen positif tetap muncul dalam bentuk empati dan dukungan terhadap upaya penanggulangan banjir.

Evaluasi kinerja model menggunakan confusion matrix serta metrik accuracy, precision, recall, dan F1-score menunjukkan bahwa algoritma Naive Bayes memiliki kinerja yang cukup baik, terutama dalam mengenali sentimen negatif sebagai kelas dominan. Namun, ketidakseimbangan data menyebabkan model kurang optimal dalam mendeteksi sentimen minoritas.

Secara keseluruhan, penelitian ini menunjukkan bahwa analisis sentimen berbasis Naive Bayes dapat digunakan sebagai pendekatan yang efektif untuk memahami persepsi publik terhadap peristiwa banjir melalui media sosial dan dapat

menjadi bahan pertimbangan awal bagi pihak terkait dalam mendukung pengambilan keputusan berbasis data.

DAFTAR PUSTAKA

- [1] Husniawati, N., & Herawati, T. M. (2023). Pengaruh Pengetahuan dan Peran Individu terhadap Kesiapsiagaan Bencana Banjir pada Masyarakat. *Jurnal Ilmu Kesehatan Masyarakat*, 12(01), 11-19.
- [2] Shidiq, M. F. A., & Alita, D. (2025). ANALISIS SENTIMEN MASYARAKAT TERHADAP KASUS JUDI ONLINE MENGGUNAKAN DATA DARI MEDIA SOSIAL X PENDEKATAN NAIVE BAYES DAN SVM. *Jurnal Sistem Informasi dan Informatika (Simika)*, 8(1), 24-35.
- [3] Manullang, O., & Prianto, C. (2023). Analisis Sentimen dalam Memprediksi Hasil Pemilu Presiden dan Wakil Presiden: Systematic Literature Review. *Jurnal Informatika dan Teknologi Komputer (J-ICOM)*, 4(2), 104-113.
- [4] Fathiarahma, A., Voutama, A., Ridwan, T., & Heryana, N. (2023). Analisis Text Mining Klasifikasi Kegiatan Keluarga menggunakan Orange dengan Metode Naive Bayes. *Jurnal Teknologi Terpadu*, 9(1), 35-41.
- [5] Khairunnisa, S., Adiwijaya, A., & Al Faraby, S. (2021). Pengaruh Text Preprocessing terhadap Analisis Sentimen Komentar Masyarakat pada Media Sosial Twitter (Studi Kasus Pandemi COVID-19). *J. Media Inform. Budidarma*, 5(2), 406.
- [6] Lutfi, M., Surorejo, S., & Septiana, P. (2022). SYSTEMATIC literature review: penerapan algoritma naives bayes dalam sistem pakar. *Jurnal Minfo Polgan*, 11(2), 7-13.
- [7] Santoso, L., & Priyadi, P. (2024). Mengoptimalkan Proses Pembersihan Data dalam Analisis Big Data Menggunakan Pipeline Berbasis AI. *Elkom: Jurnal Elektronika dan Komputer*, 17(2), 657-666.
- [10] Pane, S. F., & Ramdan, J. (2022). Pemodelan Machine Learning: Analisis Sentimen Masyarakat Terhadap Kebijakan PPKM Menggunakan Data Twitter. *Jurnal Sistem Cerdas*, 5(1), 12-20.
- [11] Darwis, D., Siskawati, N., & Abidin, Z. (2021). Penerapan Algoritma Naive Bayes Untuk Analisis Sentimen Review Data Twitter Bmkg Nasional. *Jurnal Tekno Kompak*, 15(1), 131-145.
- [12] Wibowo, S. H., Toyib, R., Muntahanah, M., & Darnita, Y. (2022). Time complexity in rejang language stemming. *JURNAL INFOTEL*, 14(3), 174-179.
- [13] Krisnandi, D., Ambarwati, R. N., Asih, A. Y., Ardiansyah, A., & Pardede, H. F. (2023). Analisis Komentar Cyberbullying Terhadap Kata Yang Mengandung Toksisitas Dan Agresi Menggunakan Bag of Words dan TF-IDF Dengan Klasifikasi SVM. *Jurnal Linguistik Komputasional (JLK)*, 6(2), 36-41.
- [14] Valerian, F. R., Syarief, M., & Fatah, D. A. (2025). Klasifikasi tingkat obesitas menggunakan metode gbm dan confusion matrix. *JATI (Jurnal Mahasiswa Teknik Informatika)*, 9(2), 2242-2249.