Analisis Sentimen Komentar *Twitter* Terhadap Peristiwa Kebakaran Bromo Menggunakan Metode *Naïve Bayes* dan *Support Vector Machine*

Rahmana Zahara Budi Putra; Khanun Roisatul Ummah, S.Tr.T., M.Tr.Kom.
Teknik Informatika, Fakultas Komunikasi dan Informatika, Universitas Muhammadiyah Surakarta

Abstrak

Kejadian kebakaran Gunung Bromo pada tanggal 6-15 September 2023 mendapatkan perhatian luas di platform *Twitter*, memunculkan beragam opini dan perasaan dari masyarakat. Penelitian ini memiliki tujuan untuk menelaah sentimen komentar di *Twitter* terkait peristiwa tersebut menggunakan metode klasifikasi *Naïve Bayes* dan *Support Vector Machine* (SVM). Penelitian ini berawal dari kebutuhan untuk memahami bagaimana masyarakat menanggapi peristiwa alam serupa di era media sosial. Solusi yang ditawarkan dalam penelitian ini adalah analisis sentimen, yang diharapkan memberikan wawasan yang lebih mendalam mengenai reaksi masyarakat terhadap kebakaran di Bromo. Dalam penelitian ini, *Naïve bayes* memiliki keunggulan dalam kesederhanaan, kecepatan, dan tingkat akurasi yang lebih tinggi. Sementara itu, *hyperplane* yang mengoptimalkan margin antara tiga kelas berbeda dapat ditemukan menggunakan SVM. Tiga tabel berupa positif, negatif, dan netralitas dihasilkan oleh klasifikasi sentimen yang digunakan dalam pengkajian. Analisis dan pengumpulan data komentar Twitter terkait kebakaran Bromo akan dilakukan. Model klasifikasi dilatih menggunakan data pelatihan dalam implementasi pengujian, dan efektivitas kedua pendekatan ini diuji menggunakan data pengujian. Evaluasi menunjukkan bahwa *Naïve Bayes* lebih unggul dalam mengidentifikasi pola dalam data dan memprediksi hasil dengan akurasi 60.09%, *Precision* 60.06%, *recall* 60.09%, dan *F1-Score* 59.90%. Sementara itu, *Support Vector Machine* mencatatkan akurasi 55.92%, *Precision* 55.73%, *recall* 55.92%, dan *F1-Score* 55.80%. Hasil penelitian ini menunjukkan bahwa sentimen negatif mendominasi data dengan persentase 48.2%, diikuti oleh sentimen netral dengan 27.3%, dan sentimen positif dengan 24.5%. Ini mencerminkan bahwa sentimen umum masyarakat cenderung negatif, yang ditunjukkan oleh komentar negatif di *Twitter* selama peristiwa kebakaran Gunung Bromo.

Kata Kunci: Analisis Sentimen, *Twitter*, Gunung Bromo, Naïve Bayes, SVM

Abstract

The Mount Bromo fire incident, which occurred from September 6 to 15, 2023, garnered widespread attention on the Twitter platform, eliciting a variety of opinions and feelings from the public. This study aims to analyze the sentiment of Twitter comments related to the event using the Naïve Bayes and Support Vector Machine (SVM) classification methods. The research stems from the need to understand how the public responds to similar natural events in the era of social media. The solution offered in this study is sentiment analysis, which is expected to provide deeper insights into the public's reaction to the Bromo fire. In this study, Naïve Bayes has advantages in simplicity, speed, and a higher level of accuracy. Meanwhile, the hyperplane that optimizes the margin between three distinct classes can be found using SVM. Three tables, positivity, negativity, and neutrality are produced by the sentiment classification used in this study. Analysis and collection of Twitter comment data pertaining to the Bromo fire will take place. The classification model is trained using training data in the testing implementation, and the effectiveness of these two approaches is tested using testing data. The evaluation shows that Naïve Bayes is superior in identifying patterns in the data and predicting outcomes

with an accuracy of 60.09%, precision of 60.06%, recall of 60.09%, and F1-Score of 59.90%. Meanwhile, the Support Vector Machine recorded an accuracy of 55.92%, precision of 55.73%, recall of 55.92%, and F1-Score of 55.80%. The results of this study indicate that negative sentiment dominates the data with a percentage of 48.2%, followed by neutral sentiment at 27.3%, and positive sentiment at 24.5%. This reflects that the general sentiment of the community tends to be negative, as shown by the negative comments on Twitter during the Mount Bromo fire event.

Keywords: *Sentiment Analysis, Twitter, Mount Bromo, Naïve Bayes, SVM*

1. PENDAHULUAN

Analisis sentimen merupakan teknik untuk menghimpun dan menilai pendapat orang lain tentang suatu isu di media sosial berbasis web. Sebagai bagian dari Proses Bahasa Alami (NLP), analisis sentimen memungkinkan identifikasi konten teks yang berisi opini atau pandangan tentang suatu isu, baik itu positif, negatif, atau netral (Listyarini & Anggoro, 2021). Tingkat analisis ini bisa dilakukan pada dokumen, kalimat, atau kata (Fitri et al., 2019). Karena banyaknya kegunaannya, termasuk perkiraan harga saham, isu politik, kepuasan produk atau layanan, dan analisis reputasi, analisis sentimen semakin populer. Analisis sentimen digunakan pada Twitter yang banyak digunakan di Indonesia.

Di Indonesia, Twitter merupakan jaringan media sosial yang populer (Sasmita et al., 2022). Platform media sosial yang *tweet*-nya, berisi opini dan komentar, dapat digunakan sebagai sumber data untuk menganalisis sentimen publik. Untuk mendapatkan data dari Twitter, kita melakukan proses yang disebut *crawling*, yaitu pengumpulan atau pengunduhan data dari database (Ramadhan & Setiawan, 2019). Data yang dikumpulkan mencakup pengguna dan *tweet* serta atribut-atributnya. Sentimen yang terkandung dalam *tweet*, yang dapat menjadi indikator pandangan masyarakat, dikategorikan menjadi tiga jenis: positif, negatif, dan netral. Mengingat *tweet* sering kali memiliki makna yang ambigu, diperlukan algoritma untuk menentukan sentimen tersebut. Analisis sentimen terhadap komentar Twitter terkait kebakaran Bromo dilakukan untuk pengkajian. Tujuannya untuk memahami cara masyarakat menyampaikan saran dan kritik terkait bencana kebakaran Bromo.

Dalam pengkajian, analisis sentimen dilakukan menggunakan teknik *Naïve Bayes Classifier* dan *Support Vector Machine* pada komentar Twitter terkait kebakaran Bromo. *Naïve Bayes* dikenal karena kemudahan penggunaannya, kecepatan, dan akurasi, meskipun memiliki keterbatasan dalam asumsi independensi atribut (Indrayuni, 2018). Meskipun merupakan metode yang telah lama ada, *Naïve bayes* masih mampu menghasilkan klasifikasi yang akurat (Suasnawa et al., 2023). Di sisi lain, SVM unggul dalam mengklasifikasikan data linier dan non-linier dengan mencari *hyperplane* optimal (Zuraiyah et al., 2023). Meskipun prosesnya lebih lambat, SVM dapat mengklasifikasikan data dengan baik bahkan dengan jumlah data yang sedikit, dan mampu menangani data non-linear dengan

akurasi tinggi (Candrayani et al., 2023). *Naïve bayes* dan SVM sering digunakan dalam analisis sentimen Twitter karena kemampuan mereka dalam mengklasifikasikan polaritas teks dan mengatasi data besar. Penggunaan kombinasi keduanya sering menghasilkan hasil yang lebih baik.

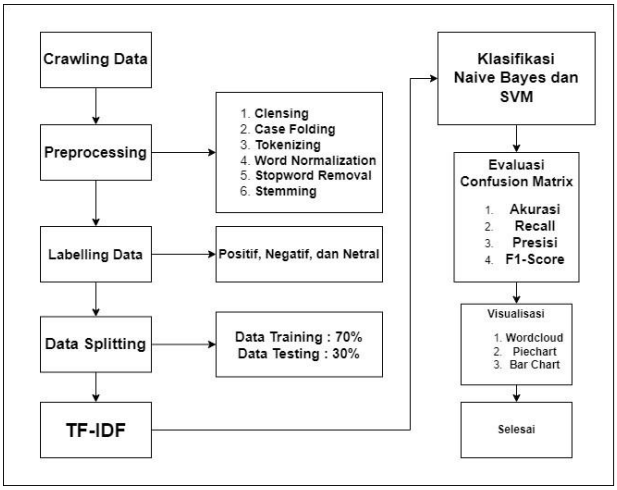
Salah satu jenis metode klasifikasi, *Naïve Bayes*, digunakan untuk melakukan analisis sentimen pada penelitian sebelumnya oleh Alec Go, dkk. (Fikri dkk., 2020). Namun hasil sentimen hanya dibagi menjadi kelompok positif dan negatif. Sementara itu, *tweet* tidak terbatas pada dua klasifikasi tersebut. Perasaan seseorang melebihi senang atau sedih ketika menulis *tweet*. Akibatnya klasifikasi sentimen dalam penelitian ini hanya mencakup satu kategori sentimen yaitu netral. Memiliki tiga kategori sentimen dimaksudkan untuk meningkatkan cara representasi hasil sentimen kumpulan data. Untuk membandingkan hasil, dua teknik klasifikasi, *Naïve Bayes* dan SVM akan dibandingkan.

Analisis sentimen dilakukan dalam pekerjaan dengan menggunakan teknik *Naïve Bayes* dan SVM. *Naïve bayes* menunjukkan akurasi sebesar 79%, sementara SVM mencapai 81%, menunjukkan bahwa SVM memiliki keunggulan dalam hal akurasi (Noviana & Rasal, 2023), Namun, dalam studi lain, *Naïve bayes* mencapai akurasi sekitar 84,73%, sedikit lebih baik dibandingkan SVM yang mencapai akurasi 82,42% (Ginabila & Fauzi, 2023). Pengkajian dari Ketut (Candrayani et al., 2023) menggunakan 953.378 *tweet* dari Januari 2020 hingga Mei 2022, yang diklasifikasikan menjadi tiga kelompok sentimen menggunakan kedua algoritma: positif (23,82%), negatif (32,91%), dan netral (43,28%).

Twitter adalah tempat utama bagi individu untuk mengomunikasikan pendapat dan sentimen mereka mengenai bencana alam, oleh karena itu analisis sentimen terhadap komentar mengenai kebakaran Bromo sangatlah relevan. Hasil yang lebih akurat dalam mengkategorikan sentimen ke dalam kategori positif, negatif, dan netral diharapkan dapat diperoleh dengan menggunakan pendekatan *Naïve Bayes* dan *Support Vector Machine* (SVM). Pengkajian diharapkan memberikan wawasan tentang respons masyarakat terhadap kebakaran Bromo melalui media sosial.

2. METODE

Studi menggunakan teknik penambangan data ilmiah yang disebut Knowledge Discovery in Databases (KDD) untuk mengekstrak pola baru dari kumpulan data yang sangat besar. Dua model kategorisasi digunakan: algoritma *Naïve Bayes* dan SVM. Langkah-langkah dalam proses penelitian meliputi data *scraping, preprocessing, labeling, splitting,* klasifikasi TF-IDF, SVM dan Naïve Bayes, evaluasi (*confusion matriks*), dan visualisasi. Penjelasan setiap langkah pendekatan pengkajian yang digunakan diberikan di bawah.



Gambar 1. Alur Penelitian

Metode pengkajian ditunjukkan pada Gambar 1, dimulai dengan penghimpunan data melalui pengikisan data dan diakhiri dengan pra-pemrosesan data untuk membersihkan dan menata data. Setelah mengkategorikan dataset, dilakukan pelabelan data dan pemisahan data untuk memisahkan data menjadi set pelatihan dan pengujian. Prosedur TF-IDF digunakan untuk mengevaluasi signifikansi suatu kata dalam sebuah dokumen. Selama fase klasifikasi, sentimen teks atau data diklasifikasikan menggunakan *Naïve Bayes* dan *Support Vector Machines*. *Confusion Matrix* digunakan untuk evaluasi, dan plot, diagram, dan visual digunakan untuk menyajikan temuan.

2.1 Scrapping Data

Hal tersebut dilakukan dengan mengumpulkan data review dari Twitter menggunakan *hashtag* "*kebakaran Bromo*" dan mengunduhnya menggunakan *GoogleCollab* di *Google Chrome*. Data tinjauan dikumpulkan antara 6 September 2023 hingga 10 Oktober 2023. *Excel* berisi data kumpulan yang telah disimpan.

2.2 Preprocessing Data

Preprocessing merupakan tahapan untuk mempersiapkan data review sehingga menjadi data berkualitas dan siap digunakan. (Fritama et al., 2023). Dengan menggunakan bahasa pemrograman *Python* dan perpustakaannya, prosedur pra-pemrosesan termasuk pembersihan, pelipatan kasus, tokenisasi, normalisasi kata, penghapusan *stopword*, dan *stemming* dilakukan untuk pengkajian (Danirmala & Nugroho, 2023).

2.2.1 Cleaning

Langkah awal dalam proses *Cleaning* adalah melakukan penghapusan elemen-elemen seperti *url, username, tanda pagar(#), symbol(@),* dan karakter khusus lainnya dari *tweet* yang telah diunduh. Tahap *cleaning* menghasilkan data telah bersih dari simbol-simbol dan *url* (Dewi & Mailoa, 2023).

2.2.2 Case Folding

Proses ini dalaha tahap di mana teks yang memiliki ketidakaturan dalam penggunaan huruf. Tahap ini

seperti komentar dengan huruf yang tidak konsisten akan di modifikasi. Prosedur *Case Folding* bertujuan mengubah huruf dalam teks komentar yang dibersihkan menjadi format standar, yaitu menjadi huruf kecil. Ini memungkinkan analisis teks lebih akurat dan konsisten (Saddam et al., 2023).

2.2.3 Tokenizing

Tokonezing adalah langkah penting dalam pemrosesan teks yang melibatkan sejumlah karakter dalam teks menjadi unit kata. Selama proses ini karakter pembatas seperti spasi, tanda koma, titik, dan karakter khusus lainnya dapat dihapus. Selain itu angka juga dihilangkan menjadi bentuk yang sesuai pada kebutuhan (Buana et al., 2023).

2.2.4 Word Normalization

Normalization adalah proses mengenali dan mengatasi kata-kata yang tidak standar dalam penulisan. Selama proses ini, sistem computer akan mengidentifikasi kata-kata yang mungkin memiliki kesalahan ejaan atau varian lainnya. Selanjutnya, kata-kata tersebut diubah menjadi varian yang lebih sesuai dengan Kamus Besar Bahasa Indonesia (Fitriansyah & Sibaroni, 2023).

2.2.5 Stopword Removal

Stopwords adalah langkah dalam pemrosesan teks guna memberikan teks dari kata-kata yang ada dengan frekuensi tinggi namun kurang memberikan kontribusi signifikasi dalam analisis. Termasuk dalam kategori stopwords adalah kata-kata seperti “kok”, “lah”, “dll”. Kata yang sejenis yang seringkali digunakan dalam percakapan sehari-hari namun kurang relevan dalam konteks analisis teks (Hermaliani & Kurniawati, 2023).

2.2.6 Stemming

Stemming adalah suatu proses yang miliki tujuan untuk menyederhankan kata-kata dalam teks ke bentuk dasarnya. Tahapan proses mengaitkan penghapusan imbuhan atau akhiran kata sehingga kata-kata memiliki bentuk bervariasi dapat diubah menjadi bentuk dasarnya. Contoh, kata “berlari” dapat diubah menjadi kata dasar “lari” atau kata “menyanyi” disederhanakan menjadi “nyanyi” (Khoiruddin et al., 2023).

2.3 Labelling Data

Tiga kategori berupa positif, negatif, dan netral, dibuat melalui proses kategorisasi data ini. Secara umum, kebaikan dikaitkan dengan kategori emosi positif. kategori negatif memiliki perasaan jeek atau tidak baik (Oktavia et al., 2023). Sedangkan netral dengan emosi yang tenang, stabil dan tanpa adanya perasaan ekstrem seperti senang, sedih atau marah. Proses dilakukan dengan cara manual dan subjektif.

2.4 Data Splitting

Tiga puluh persen dari keseluruhan dataset yang ada akan menjadi data pengujian, dan tujuh puluh persen sisanya dari dataset review akan dibagi menjadi data pelatihan dan pengujian. Data pengujian digunakan untuk mengevaluasi performa model menggunakan data yang belum pernah dilihat

sebelumnya, sedangkan data pelatihan digunakan untuk melatih model (Anggoro & Kurnia, 2020).

2.5 TF-IDF

Model yang disebut TF-IDF memberikan bobot pada setiap kata dalam dokumen berdasarkan seberapa sering kata tersebut muncul dalam dokumen (TF) dan seberapa umum kata tersebut di antara sekumpulan dokumen (IDF). TF mengukur relevansi kata dalam dokumen, sebagaimana ditunjukan pada rumus (1). Sementara IDF mengukur nilai informasi kata tersebut dalam keseluruhan kumpulan dokumen, sebagaimana ditunjukan pada rumus (2) (Oktavia et al., 2023).

Perhitungan nilai TF secara manual dilakukan dengan persamaan.

$$TF\ (d,t) = \frac{jumlah\ kemunclan\ term\ dalam\ document}{total\ kata\ dalam\ doecument} \quad (1)$$

Dilanjutkan mencari nilai IDF dengan persamaan.

$$IDF\ (t) = Log\frac{total\ dokumen}{jumlah\ total\ \ yangan\ mengandung\ kata\ t} \quad (2)$$

Keterangan:

$t = kata$

$d = dokumen$

Setelah kedua adanya pemerehan, perhitungan perolehan TF-IDF dilaksanakan dengan persamaan, sebagaimana ditunjukan pada rumus (3).

$$TF - IDF\ (d,t) = TF(d,t)\ x\ IDF(t) \quad (3)$$

2.6 Klasifikasi *Naïve bayes* dan *Support Vector Machine (SVM)*

Dalam tahapan klasifikasi, teknik dari *Naïve bayes* dan SVM digunakan. *Naïve bayes* dipilih karena keefisienan serta kemudahannya dalam menghasilkan prediksi sentimen dengan cepat. Di sisi lain, SVM dipilih karena kemampuannya dalam menangani dataset yang non-linier. Kedua metode ini bertujuan untuk mendapatkan nilai sentimen yang positif, negatif, dan netral dari setiap komentar data sehingga percampuran kombinasi dari kedua teknik diharapkan bisa mencapai hasil analisis sentimen yang lebih akurat.

a. *Naïve Bayes*

Dalam klasifikasi ini, dilakukan analisis mendalam terhadap dokumen-dokumen yang sudah dikategorikan. Kata-kata dalam dokumen menjadi pokok pemikiran utama dalam proses pengklasifikasian. Sebuah kelompok kata dengan makna khusus digunakan sebagai pedoman untuk menentukan kategori yang tepat, memungkinkan sistem untuk mengorganisir dokumen dengan efisien dan akurat berdasarkan makna kata-kata tersebut, sebagaimana ditunjukkan pada rumus (4) (Sianipar et al., 2023).

$$P(H|X) = \frac{P(X|H)P(H)}{P(X)} \quad (4)$$

X = data dengan kelas tidak dikenal

H = hipotesis data X adalah kelas khusus

$P(H|X)$ = probabilitas hipotesis H didasarkan pada kondisi X

$P(H)$ = probabilitas hipotesis H

$P(X|H)$ = probavilitas hipotesis X didasarkan pada kondisi H

$P(X)$ = probabilitas X

b. *Support Vector Machine (SVM)*

SVM merupakan teknik relatife baru untuk melakukan prediksi dalam kasus klasifikasi maupun regresi (Oktavia et al., 2023). Metode ini berupaya menemukan *hyperplane* optimal atau batas keputusan yang dapat memisahkan dua kategori dalam ruang input, yaitu memisahkan *tweet* yang bersifat positif dari yang bersifat negatif, sebagaimana ditunjukkan pada rumus (5) (Ilmawan & Mude, 2020).

$$f(x) = sign(w.x + b) \quad (5)$$

$f(x)$ = fungsi keputusan

x = vector bobot

b = bias

$sign$ = fungsi penanda, yang mengembanglikan tanda dari argumen

2.7 Evaluasi (*Confusion Matrix)*

Dalam evaluasi hasil, pengkajian akan memanfaatkan *Confusion Matrix* untuk mengukur kinerja model klasifikasi *Naïve bayes* dan SVM serta membandingkan tingkat akurasinya. *Confusion Matrix* berisi angka mencerminkan efektivitas model dalam mengklasifikasikan data, dengan empat elemen utama: *True Positive* (TP) saat model berhasil mengidentifikasi data positif dengan tepat, *True Negative* (TN) saat model berhasil mengenali data negatif dengan tepat atau betul, *False Positive* (FP) saat model salah dalam mengidentifikasi data negatif sebagai positif, dan *False Negative* (FN) saat model salah dalam mengidentifikasi data positif sebagai tanda negatif (Syahlan et al., 2023).

Accuracy merupakan cara untuk mengukur sejauh mana model kita membuat prediksi yang benar dengan membandingkan dengan data sebenarnya. Untuk menentukan seberapa akurat model kita dalam menghasilkan prediksi, kita harus membandingkan jumlah prediksi yang tepat dengan jumlah total data yang dinilai, seperti yang ditunjukkan oleh rumus (6). Namun, dalam situasi ketidakseimbangan kelas, metrik seperti *Precision*, *Recall*, atau *F1-Score* lebih informatif untuk mengevaluasi kinerja model secara menyeluruh (Halim & Safuwan, 2023).

$$Accuracy = \frac{TP+TN}{TP+TN+FP+FN} \quad (6)$$

Precision mengukur keakuratan model dalam memprediksi kasus positif dengan membandingkan prediksi yang *True Positive* (TP) dengan total prediksi positif, sebagaimana ditunjukkan pada rumus (7). Sementara *Recall* mengukur kemampuan model untuk menemukan

sebagian besar kasus positif dalam data dengan membandingkan TP dengan total kasus positif yang sebenarnya, sebagaimana ditunjukkan pada rumus (8). Baik *Precision* maupun *Recall* adalah faktor penting dalam menilai kinerja model klasifikasi, membantu kita menilai seberapa baik model dalam membuat prediksi yang akurat dan menemukan kasus positif dalam dataset (Edpuganti et al., 2023).

$$Precision = \frac{TP}{TP+FP} \quad (7)$$

$$Recall = \frac{TP}{TP+FN} \quad (8)$$

F1-Score adalah alat evaluasi dalam model klasifikasi yang menggabungkan *Precision* dan *Recall*. Ini memberikan gambaran keseluruhan tentang seberapa baik model dalam memprediksi dengan tepat kelas positif dan menemukan sebagian besar kasus sebenarnya positif. *F1-Score* membantu kita menilai keseimbangan antara keakuratan dan kemampuan model untuk menemukan kasus positif, terutama dalam situasi dengan ketidakseimbangan kelas. Ini merupakan alat penting bagi praktisi data science untuk mengukur kinerja model klasifikasi, sebagaimana ditunjukkan pada rumus (9) (Edpuganti et al., 2023).

$$F1 - Score = \frac{2\,(Precision \times Recall)}{Precision + Recall} \quad (9)$$

2.8 Visualisasi

Data disederhanakan dan lebih mudah dipahami pembaca melalui penggunaan visualisasi. Kumpulan data ulasan adalah keluaran dari visualisasi, dengan menggunakan *Pie Chart* dan *Bar Plot*. Visualisasi data seperti ini membantu memberikan informasi penting dalam data ulasan dengan cara yang mudah dipahami, memudahkan pengambilan keputusan dan analisis lebih lanjut.

3. HASIL DAN PEMBAHASAN

3.1 Scrapping Data

Dengan *Google Colab*, saya dapat mengikis *tweet* bahasa Indonesia yang berisi istilah Kebakaran Bromo guna mengumpulkan data untuk penelitian ini. ditangkap pada bulan Oktober antara 6 September 2023 dan 10 Oktober 2023. Setelah prosedur pengikisan, 1441 *tweet* dengan kolom tanggal dan *tweet* ditemukan. Gambar 2 menampilkan hasil dari prosedur *Data Scraping*.

index	Tweets,id tweets,date tweets,username,id user,quote count,reply count,retweet count,favorite count,url tweets
0	@duniabulat5 @CNNIndonesia Nk musim kemarau panjang hampir hutan di jatim kebakaran tiap tahunnya. Savana bromo jg kebakar tiap tahun tp pas penghujan bikin mata takjub,1.69924796937845E+018,2023-09-06 02:27:41+00:00,sugikchan,7.60376195493994E+017,0,0,0,2,https://twitter.com/sugikchan/status/1699247969378451458
1	Seng prewed sampek ngarai kebakaran nang bromo Disumpahi wong akeh sampean mbak mas.. Wes kukuse sampek sak kecamatan Ora iso ngeramban..,1.69944624293229E+018,2023-09-06 15:35:33+00:00,orange_puppy331,1.32816744778697E+018,0,0,0,0,https://twitter.com/orange_puppy331/status/1699446242932297957
2	Kakeane kebakaran bromo wes dipadamke seminggu iki malah mau awan diobong wong prewed. Asw,1.69947419273398E+018,2023-09-06 17:26:36+00:00,andriawan98,438039847,0,0,0,0,https://twitter.com/andriawan98/status/1699474192733987044
3	Kenapa kemping di Savana Bromo dilarang? Ya karena ini, rawan kebakaran. Beberapa tahun lalu udah kejadian kebakaran yg disinyalir dari aktivitas kemping. Terus sekarang kejadian karena flare begini. Dan parahnya lagi, Bromo baru dibuka sehari setelah sebelumnya kebakaran.,1.69951756742844E+018,2023-09-06 20:18:58+00:00,brigittasiw,303627559,0,1,0,1,https://twitter.com/brigittasiw/status/1699517567428444210
4	@infomitigasi @RadioElshinta @KementerianLHK Kalau TNBTS keseluruhan berapa min ?? Kan BROMO TENGGER SEMERU jadi 1 kesatuan n pengelolanya sama.. soalnya setau saya kemarin ada pengumuman kebakaran di gunung pananjakan dan semeru juga,1.69956973579572E+018,2023-09-06 23:46:16+00:00,tamiya_mf01x,1.67414005005616E+018,0,0,0,0,https://twitter.com/tamiya_mf01x/status/1699569735795724339

Gambar 2. Hasil *Scrapping Data*

3.2 Preprocessing Data

3.2.1 Cleaning

Langkah *cleaning* dilangsungkan penghapusan username, tanda baca, symbol, tautan, tagar, dan emoji. Tahap ini dilakukan menggunakan *python*. Selanjutnya pada tahap *cleaning* dilakukan penghapusan duplikat data. Tahap ini dilakukan menggunakan *python*. Hasil data *tweet* setelah diproses *cleaning* dipaparkan dalam Gambar 3 berikut.

remove_user	cleaning
Nk musim kemarau panjang hampir hutan di jatim kebakaran tiap tahunnya. Savana bromo jg kebakar tiap tahun tp pas penghujan bikin mata takjub	musim kemarau panjang hampir hutan jatim kebakaran tiap tahunnya Savana bromo kebakar tiap tahun pas penghujan bikin mata takjub
Seng prewed sampek ngarai kebakaran nang bromo Disumpahi wong akeh sampean mbak mas.. Wes kukuse sampek sak kecamatan Ora iso ngeramban...	Seng prewed sampek ngarai kebakaran nang bromo Disumpahi wong akeh sampean mbak mas Wes kukuse sampek sak kecamatan Ora iso ngeramban
Kakeane kebakaran bromo wes dipadamke seminggu iki malah mau awan diobong wong prewed. Asw	Kakeane kebakaran bromo wes dipadamke seminggu iki malah mau awan diobong wong prewed Asw
Kenapa kemping di Savana Bromo dilarang? Ya karena ini, rawan kebakaran. Beberapa tahun lalu udah kejadian kebakaran yg disinyalir dari aktivitas kemping. Terus sekarang kejadian karena flare begini. Dan parahnya lagi, Bromo baru dibuka sehari setelah sebelumnya kebakaran.	Kenapa kemping Savana Bromo dilarang karena ini rawan kebakaran Beberapa tahun lalu udah kejadian kebakaran disinyalir dari aktivitas kemping Terus sekarang kejadian karena flare begini Dan parahnya lagi Bromo baru dibuka sehari setelah sebelumnya kebakaran
Kalau TNBTS keseluruhan berapa min ?? Kan BROMO TENGGER SEMERU jadi 1 kesatuan n pengelolanya sama.. soalnya setau saya kemarin ada pengumuman kebakaran di gunung penanjakan dan semeru juga	Kalau TNBTS keseluruhan berapa min Kan BROMO TENGGER SEMERU jadi kesatuan pengelolanya sama soalnya setau saya kemarin ada pengumuman kebakaran gunung penanjakan dan semeru juga

Gambar 3. Hasil *Cleaning Data*

3.2.2 Case Folding

Tujuan dari langkah ini adalah untuk mengubah huruf besar menjadi huruf kecil. mencoba menyederhanakan langkah berikut. Tabel 1 menampilkan temuan kasus lipat.

Tabel 1. Hasil *Case Folding*

Sebelum	Sesudah
Kenapa kemping Savana Bromo dilarang karena ini rawan kebakaran Beberapa tahun lalu udah kejadian kebakaran disinyalir dari aktivitas kemping Terus sekarang kejadian karena flare begini Dan parahnya lagi Bromo baru dibuka sehari setelah sebelumnya kebakaran	kenapa kemping savana bromo dilarang karena ini rawan kebakaran beberapa tahun lalu udah kejadian kebakaran disinyalir dari aktivitas kemping terus sekarang kejadian karena flare begini dan parahnya lagi bromo baru dibuka sehari setelah sebelumnya kebakaran

3.2.3 Tokenizing

Pelipatan kasus diikuti dengan tokenisasi, yang membagi pernyataan menjadi beberapa kata. pemisahan menggunakan fungsi *word_tokenize* dengan *Python* dan modul *nltk tokenize*. Tabel 2 menampilkan temuan tokenisasi.

Tabel 2. Hasil *Tokenizing*

Sebelum	Sesudah
kenapa kemping savana bromo dilarang karena ini rawan kebakaran beberapa	'kenapa', 'kemping', 'savana', 'bromo', 'dilarang', 'karena', 'ini', 'rawan', 'kebakaran', 'beberapa',

tahun lalu udah kejadian kebakaran disinyalir dari aktivitas kemping terus sekarang kejadian karena flare begini dan parahnya lagi bromo baru dibuka sehari setelah sebelumnya kebakaran	'tahun', 'lalu', 'udah', 'kejadian', 'kebakaran', 'disinyalir', 'dari', 'aktivitas', 'kemping', 'terus', 'sekarang', 'kejadian', 'karena', 'flare', 'begini', 'dan', 'parahnya', 'lagi', 'bromo', 'baru', 'dibuka', 'sehari', 'setelah', 'sebelumnya', 'kebakaran'

3.2.4 Word Normalization

Normalisasi kata digunakan mengubah kata-kata yang salah eja atau disingkat menjadi bentuk yang tepat. Tujuannya adalah untuk membuat kata-kata berubah dengan arti sama namun salah eja menjadi bentuk benar ejaannya. Sebagai contoh, kata "dgn" dinormalisasi menjadi "dengan.". Hasil pada *word normalization* pada Tabel 2 berikut.

Tabel 2. Hasil *Word Normalization*

Sebelum	**Sesudah**
'kenapa', 'kemping', 'savana', 'bromo', 'dilarang', 'karena', 'ini', 'rawan', 'kebakaran', 'beberapa', 'tahun', 'lalu', 'udah', 'kejadian', 'kebakaran', 'disinyalir', 'dari', 'aktivitas', 'kemping', 'terus', 'sekarang', 'kejadian', 'karena', 'flare', 'begini', 'dan', 'parahnya', 'lagi', 'bromo', 'baru', 'dibuka', 'sehari', 'setelah', 'sebelumnya', 'kebakaran'	'kenapa', 'kemping', 'savana', 'bromo', 'dilarang', 'karena', 'ini', 'rawan', 'kebakaran', 'beberapa', 'tahun', 'lalu', 'sudah', 'kejadian', 'kebakaran', 'disinyalir', 'dari', 'aktivitas', 'kemping', 'terus', 'sekarang', 'kejadian', 'karena', 'flare', 'begini', 'dan', 'parahnya', 'lagi', 'bromo', 'baru', 'dibuka', 'sehari', 'setelah', 'sebelumnya', 'kebakaran'

3.2.5 Stopword Removal

Prosedur *Stopword Removal* membantu menghilangkan istilah-istilah seperti kata penghubung ('itu", "yang", "ko", dsb.) yang tidak memiliki arti apa pun dari teks ulasan. Sebuah perpustakaan bernama NLTK (*Natural Language Toolkit*) memiliki daftar *stopword* bahasa Indonesia yang dapat digunakan untuk melakukan penghapusan ini. Tabel 3 menampilkan hasil dari prosedur *Stopword Removal*.

Tabel 3. Hasil *Stopword Removal*

Sebelum	Sesudah
'kenapa', 'kemping', 'savana', 'bromo', 'dilarang', 'karena', 'ini', 'rawan', 'kebakaran', 'beberapa', 'tahun', 'lalu', 'sudah', 'kejadian', 'kebakaran', 'disinyalir', 'dari', 'aktivitas', 'kemping', 'terus', 'sekarang', 'kejadian', 'karena', 'flare', 'begini', 'dan', 'parahnya', 'lagi', 'bromo', 'baru', 'dibuka', 'sehari', 'setelah', 'sebelumnya', 'kebakaran'	'kemping', 'savana', 'bromo', 'dilarang', 'rawan', 'kebakaran', 'kejadian', 'kebakaran', 'disinyalir', 'aktivitas', 'kemping', 'kejadian', 'flare', 'parahnya', 'bromo', 'dibuka', 'sehari', 'kebakaran'

3.2.6 Stemming

Stemming dilakukan untuk membentuk kata dasar mengubah maupun menghilangkan morfem atau imbuhan semacam 'mem', 'per', 'ber', 'an'. Proses ini dapat dilakukan dengan memanfaatkan *library Sastrawi*, sebuah perpustakaan yang menyediakan alat untuk melakukan *stemming* pada teks. Perolehan tahap *stemming* disajikan dalam Tabel 4 berikut.

Tabel 4. Hasil *Stemming*

Sebelum	Sesudah
'kemping', 'savana', 'bromo', 'dilarang', 'rawan', 'kebakaran', 'kejadian', 'kebakaran', 'disinyalir', 'aktivitas', 'kemping', 'kejadian', 'flare', 'parahnya', 'bromo', 'dibuka', 'sehari', 'kebakaran'	'kemping', 'savana', 'bromo', 'larang', 'rawan', 'bakar', 'jadi', 'bakar', 'sinyalir', 'aktivitas', 'kemping', 'jadi', 'flare', 'parah', 'bromo', 'buka', 'hari', 'bakar'

3.3 Labelling Data

Hasil *scrapping* yang sudah diproses pada tahap *Preprocessing* selanjutnya dilakukan *labelling*. Proses ini menjadi tigas kelas sentimen, meliputi positif, negatif, dan netral. Dilakukan secara manual, dapat menggunakan bahasa pemrograman yang sudah disediakan. Hasil *labelling* tampak dalam Tabel 5.

Tabel 5. Hasil *Labelling*

Tweet	Tanggal	Sentimen
sabana gunung bromo hijau pasca bakar kondisi cantik	17/10/2023	Positif
udah bikin savana bromo kebakaran hasil fotonya jelek pula	09/07/2023	Negatif
bromo kebakaran ternyata ulah yg prewedding?! WHAT?!	09/07/2023	Netral

3.4 Data Splitting

Langkah selanjutnya, data dipisahkan menjadi dua kategori: data pelatihan dan data pengujian. Model pembelajaran mesin dilatih pada data pelatihan, dan performanya diuji pada data pengujian. Fungsi *train_test_split* dari *Scikit-lear* digunakan untuk menyelesaikan tahap pembagian data. Mengikuti rasio data 70:30, dihasilkan 432 data pengujian dan 1005 data pelatihan. Gambar 4 menampilkan perolehan pemisahan data.

```
Jumlah baris data latih (X_train): 1005
Jumlah baris data uji (X_test): 432
Jumlah baris label latih (y_train): 1005
Jumlah baris label uji (y_test): 432
```

Gambar 4. Hasil *Data Splitting*

3.5 TF-IDF

Selanjutnya melakukan pembobotan (TF), tujuan pembobotan ini adalah untuk mengukur frekuensi munculnya setiap kata dalam teks, sehingga kata-kata sering muncul memiliki nilai besarnya melebihi perolehan normal. Pembobotan kata ini bermanfaat unutk membantu memproses dan memahami informasi, terutama dalam konteks analisis teks dan dokumen. Bisa dilihat pada Tabel 6 menunjukan hasil dari TF.

Tabel 6. Hasil TF

TF
'"musim"': 0.07142857142857142, '"kemarau"': 0.07142857142857142, '"hutan"': 0.07142857142857142, '"jatim"': 0.07142857142857142, '"bakar"': 0.14285714285714285, '"tahun"': 0.07142857142857142, '"savana"': 0.07142857142857142, '"bromo"': 0.07142857142857142, '"pas"': 0.07142857142857142, '"hujan"': 0.07142857142857142, '"bikin"': 0.07142857142857142, '"mata"': 0.07142857142857142, '"takjub"': 0.07142857142857142

Untuk mengukur relevansi suatu kata dalam dokumen, kita perlu menghitung frekuensi dokumen (DF) dari kata tersebut, yaitu jumlah dokumen yang mengandung kata tersebut. Selanjutnya, kita dapat membagi jumlah total dokumen dalam koleksi dengan DF untuk mendapatkan nilai dokumen/DF. Tampilan perolehan dalam Tabel 7 menunjukkan data dari DF.

Tabel 7. Hasil DF

Term	**DF**
“musim	28
“Kemarau”	37
“Hutan”	64

Dengan menentukan frekuensi invers dari dokumen yang mengandung sebuah kata (DF) dan frekuensi kemunculan sebuah kata dalam sebuah dokumen (TF), TF-IDF tercapai. Temuan IDF ditampilkan pada Tabel 8.

Tabel 8. Hasil TF-IDF

TF-IDF
'"musim"': 0.2787372085444068, '"kemarau"': 0.2594307564201274, '"hutan"': 0.22108781997946658, '"jatim"': 0.3802647571823468, '"bakar"': 0.000199104064086269, '"tahun"': 0.3802647571823468, '"savana"': 0.2002649022109704, '"bromo"': 0.0, '"pas"': 0.2147812051833467, '"hujan"': 0.28124373138806896, '"bikin"': 0.14225014931268945, '"mata"': 0.3360476708676166, '"takjub"': 0.46974782636058743

3.6 Klasifiasi

3.6.1 Naïve Bayes

Pada tahap ini menerapkan klasifikasi *Naïve bayes* dengan fungsi *MultinomialNB*. Fungsi ini sesuai untuk klasifikasi *Naïve bayes* yang dapat memproses data dalam bentuk teks. Pada proses testing menunjukkan tingkat akurasi sebesar 0.60, diikut dengan nilai *Precision* Negatif 0.65, *recall* 0.68, *F1-Score* 0.67, nilai *Precesion* Netral 0.58, *recall* 0.49, *F1-Score* 0.53, serta *Precision* Positif 0.53, *recall* 0.59, *F1-Score* 0.56. Hasil klasifikasi *Naïve bayes* dapat dilihat pada Gambar 5.

Classification Report untuk Data Test:

	precision	recall	f1-score	support
Negatif	0.65	0.68	0.67	201
Netral	0.58	0.49	0.53	135
Positif	0.53	0.59	0.56	95
accuracy			0.60	431
macro avg	0.59	0.59	0.58	431
weighted avg	0.60	0.60	0.60	431

Gambar 5. Hasil Klasifikasi *Naïve bayes*

3.6.2 Support Vector Machine

Angkah berikutnya ialah proses klasifikasi SVM menggunakan pendekatan kernel linier untuk mencari perolehan akurasi yang diperoleh. Pada proses testing menunjukan tingkat akurasi sebesar 0.56, diikut dengan nilai *Precision* Negatif 0.62, *recall* 0.64, *F1-Score* 0.63, nilai *Precesion* Netral 0.50, *recall* 0.47, *F1-Score* 0.49, dan *Precision* Positif 0.50, *recall* 0.51, *F1-Score* 0.50. Hasil klasifikasi SVM ditampikan dalam bentuk Gambar 6 berikut.

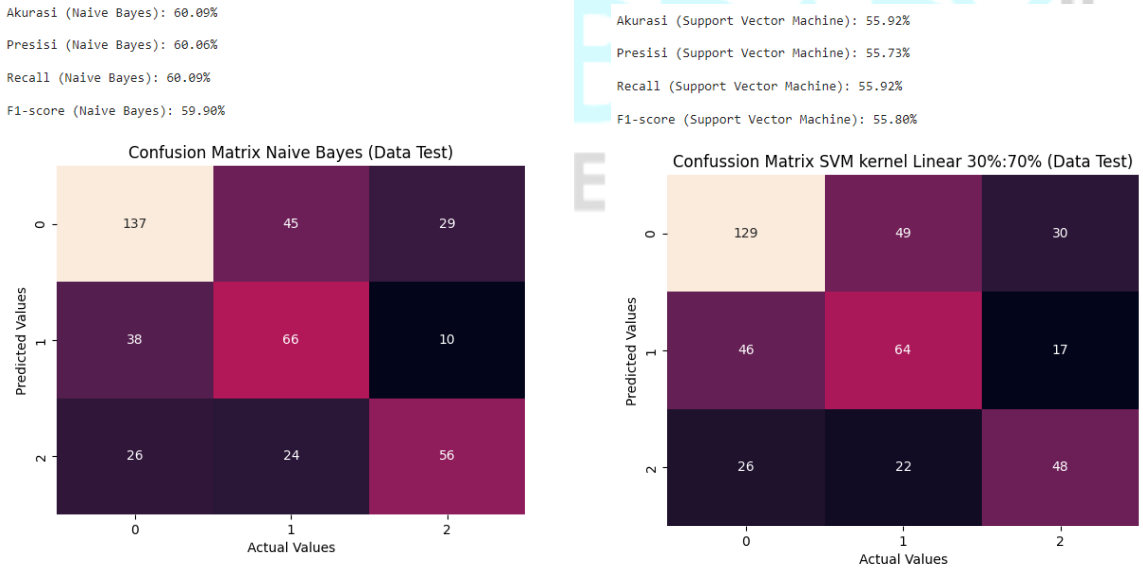
Classification Report untuk Data Test:

	precision	recall	f1-score	support
Negatif	0.62	0.64	0.63	201
Netral	0.50	0.47	0.49	135
Positif	0.50	0.51	0.50	95
accuracy			0.56	431
macro avg	0.54	0.54	0.54	431
weighted avg	0.56	0.56	0.56	431

Gambar 6. Hasil Klafisikasi SVM

3.7 Evaluasi

Untuk menilai kinerja dari metode klasifikasi yang digunakan, kita dapat menggunakan *Confusion Matrix* yang menunjukan jumlah prediksi yang benar dan salah dari model. Modul *scikit-learn* bahasa komputer *Python* memungkinkan kita mengimplementasikan fungsi *confusion_matrix* untuk menghasilkan matriks *Confusion* dan menentukan akurasi, *presisi*, *recall*, dan *F1score* model. Anda dapat membandingkan dua sistem kategorisasi menggunakan *Confusion Matrix*. *Confusion Matrix* yang diperoleh dari dua klasifikasi *Naïve Bayes* dan *Support Vector Machine* ditampilkan pada Gambar 7a dan 7b.



a b

Gambar 7. *Confusion Matrix* (a) *Naïve bayes*, (b) SVM

Pada Gambar 7a dengan menggunakan *Confusion Matrix* untuk klasifikasi *Naïve bayes* dapat melihat bahwa ada 259 data prediksi dengan benar dan 172 data prediksi dengan salah. Nilai *True Negative* (TN) 137, *True Positive* (TP) 56, *False Negative* (FN) 74, *False Positive* (FP) 50. Model *Naïve bayes* memiliki akurasi 60.09%, *Precision* 60.06%, *recall* 60.09%, dan *F1-Score* 59.90%.

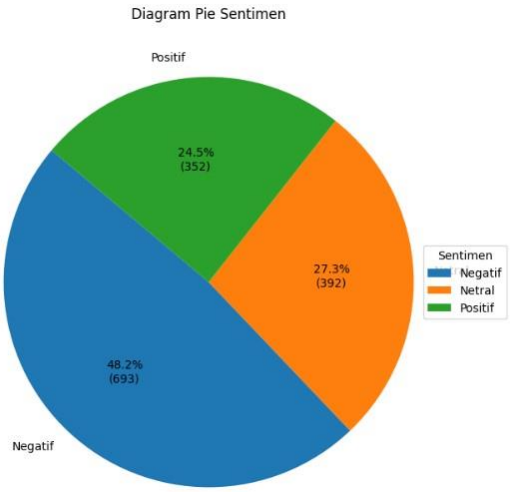
Pada Gambar 7b dengan menggunakan *Confusion Matrix* untuk klasifikasi SVM, dapat melihat bahwa ada 241 data yang diprediksi dengan benar dan 190 data yang diprediksi dengan salah. Nilai

True Negative (TN) adalah 129, *True Positive* (TP) adalah 48, *False Negative* (FN) adalah 79, dan *False Positive* (FP) adalah 48. Model SVM memiliki akurasi 55.92%, *Precision* 55.73%, *recall* 55.92%, dan *F1-Score* 55.80%.

3.8 Visualisasi

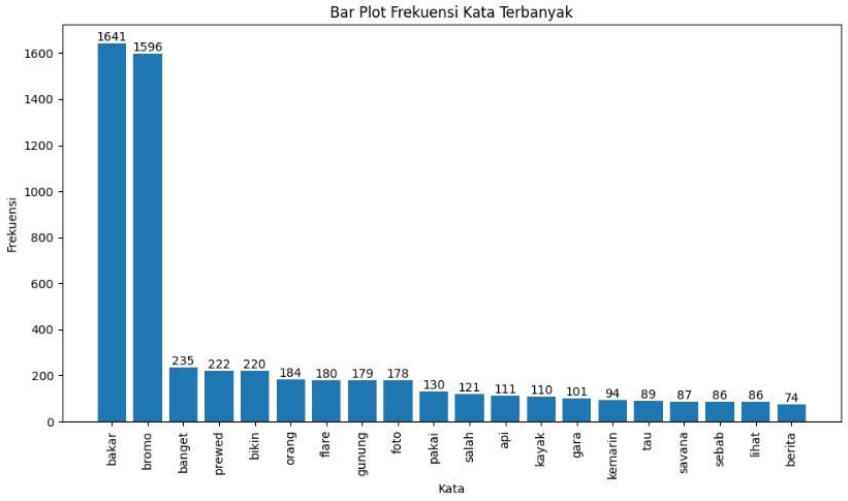
Gambar 8. *Wordcloud* (a) Sentimen Positif, (b) Sentimen Negatif, (c) Sentimen Netral

Gambar 8a, 8b, dan 8c merupakan representasi visual dalam bentuk *wordcloud* yang menggambarkan sebaran sentimen dari analisis *Twitter* yang berkaitan dengan peristiwa Gunung Bromo. Gambar 8a memperlihatkan sentimen positif dengan kata-kata yang menyoroti aspek positif dari data. Gambar 8b menampilkan sentimen netral dengan kata-kata yang bersifat informatif dan tidak mengandung emosi tertentu. Gambar 8c berfokus pada sentimen negatif, dengan kata-kata yang mencerminkan aspek negatif yang lebih dominan dibicarakan.



Gambar 9. *Pie Chart* Total Sentimen.

Pada Gambar 9 menampilkan sebuah *pie chart* yang menggambarkan hasil analisis sentimen dari sebuah kumpulan data. Dari *pie chart* tersebut, dapat dilihat bahwa sebagian besar data memiliki sentimen negatif, yaitu sekitar 48.2% yang berjumlah 693 data. Sentimen negatif ini mungkin menunjukan adanya ketidakpuasa, kekecewaan, atau kritik terhadap suatu hal. Selain itu, ada juga data yang bersifat netral, yaitu sekitar 27.3% yang berjumlah 392 data. Sentimen netral ini mungkin menunjukan adanya sikap objektif, netral, atau tidak berpihak terhadap suatu hal. Terakhir, ada juga data yang memiliki sentimen positif, yaitu sekitar 24.5% yang berjumlah 352 data. Sentimen positif ini mungkin menunjukan adanya kepuasan, penghargaan, atau pujian terhadap suatu hal. Dengan demikian, *pie chart* ini memberikan gambaran umum tentang persepsi atau opini yang terdapat dalam data.



Gambar 10. *Bar Plot* Frekuensi Kata Terbanyak

Pada Gambar 10 menampilkan sebuah *bar plot* yang menggambarkan frekuensi kata terbanyak yang muncul dalam data ulasan. Kata “bakar” mendominasi dengan muncul sebanyak 1.641 kali, disusul oleh “bromo” dengan 1.596 kali. Kata lain seperti “banget” yang muncul 235 kali dan “prewed”

sebanyak 222 kali juga sering terlihat, sampai ke kata dengan frekuensi paling sedikit yaitu “berita” yang tercatat 74 kali. *Bar Plot* ini mengungkapkan tren umum dalam data ulasan melalui frekuensi penggunaan kata-kata tertentu.

4. PENUTUP

4.1 Kesimpulan

Penelitian ini telah membandingkan efektivitas algoritma *Naïve Bayes dan Support Vector Machine* dalam menganalisis sentimen dari *tweet*. Dari 1441 *tweet* yang dikumpulkan selama periode 6 September hingga 10 Oktober 2023, 1437 di antaranya digunakan dalam analisis setelah proses Preprocessing. Dari hasil analisis, ditemukan bahwa 24.5% sebanyak 352 data bersentimen positif, 27.3% sebanyak 392 data netral, dan 48.2% sebanyak 693 data negatif. Evaluasi menunjukkan bahwa *Naïve Bayes* lebih unggul dalam mengidentifikasi pola dalam data dan memprediksi hasil dengan akurasi 60.09%, *Precision* 60.06%, *recall* 60.09%, dan *F1-Score* 59.90%. Sementara itu, *Support Vector Machine* mencatatkan akurasi 55.92%, *Precision* 55.73%, *recall* 55.92%, dan *F1-Score* 55.80%. Analisis ini juga menemukan bahwa sentimen negatif lebih dominan daripada sentimen positif dalam komentar tentang kebakaran Bromo di dataset Twitter yang dianalisis.

4.2 Saran

Penulis menyadari bahwa penelitian ini masih memiliki ruang untuk peningkatan dan perbaikan di masa mendatang. Dalam upaya untuk meningkatkan kualitas hasil, penulis berharap penelitian berikutnya dapat memperluas dataset yang digunakan, sehingga analisis sentimen yang dihasilkan dapat lebih kaya informasi. Selain itu, penelitian ini juga dapat diperkaya dengan penggunaan algoritma yang berbeda untuk tujuan perbandingan, serta penambahan algoritma lainnya untuk dibandingkan dengan *Naïve bayes* dan *Support Vector Machine,* dengan harapan dapat meningkatkan hasil yang lebih baik.

DAFTAR PUSTAKA

Anggoro, D. A., & Kurnia, N. D. (2020). Comparison of accuracy level of support vector machine (SVM) and K-nearest neighbors (KNN) algorithms in predicting heart disease. *International Journal*, *8*(5), 1689–1694.

Buana, R. S., Gata, W., Widodo, A. Z. P., Setiawan, H., & Hilyati, K. (2023). Analisis Sentimen pada Komen Twitter Pawang Hujan Mandalika dengan Support Vector Machine (SVM) dan Naïve Bayes. *Jurnal JTIK (Jurnal Teknologi Informasi Dan Komunikasi)*, *7*(2), 194–200.

Candrayani, K. M. A., Suarjaya, I. M. A. D., & Wiranatha, A. A. K. A. C. (2023). Analisis Sentimen Pembelajaran Daring Era Pandemi COVID-19 Menggunakan Naive Bayes Dan SVM. *TEMATIK*, *10*(1), 47–53.

Danirmala, T., & Nugroho, Y. S. (2023). Analisis Sentimen Terhadap Topik Kenaikan Harga Bahan Bakar Minyak (BBM) pada Media Sosial Twitter. *Indonesian Journal of Computer Science*, *12*(3).

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.

Dewi, T. A., & Mailoa, E. (2023). Perbandingan Implementasi Metode Smote Pada Algoritma Support Vector Machine (Svm) Dalam Analisis Sentimen Opini Masyarakat Tentang Mixue. *Jurnal Indonesia: Manajemen Informatika Dan Komunikasi*, *4*(3), 849–855.

Edpuganti, A., Akshaya, P., Gouthami, J., Sajith Variyar, V. V, Sowmya, V., & Sivanpillai, R. (2023). Effect of data quality on water body segmentation with deeplabv3+ algorithm. *The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, *48*, 81–85.

Fikri, M. I., Sabrila, T. S., & Azhar, Y. (2020). Perbandingan metode naïve bayes dan support vector machine pada analisis sentimen twitter. *SMATIKA Jurnal: STIKI Informatika Jurnal*, *10*(02), 71–76.

Fitri, V. A., Andreswari, R., & Hasibuan, M. A. (2019). Sentiment analysis of social media Twitter with case of Anti-LGBT campaign in Indonesia using Naïve Bayes, decision tree, and random forest algorithm. *Procedia Computer Science*, *161*, 765–772.

Fitriansyah, A. R., & Sibaroni, Y. (2023). Analisis Sentimen Terhadap Pembangunan Kereta Cepat Jakarta-Bandung Pada Media Sosial Twitter Menggunakan Metode SVM dan GloVe Word Embedding. *E-Proceeding of Engineering*, *10*(2), 1713.

Fritama, S. D., Ramadhan, Y. R., & Komara, M. A. (2023). Analisis Sentimen Review Produk Acne Spot Treatment di Female Daily Menggunakan Algoritma K-Nearest Neighbor. *KLIK: Kajian Ilmiah Informatika Dan Komputer*, *4*(1), 134–143.

Ginabila, G., & Fauzi, A. (2023). Analisis Sentimen Terhadap Pemutar Musik Online Spotify Dengan Algoritma Naive Bayes dan Support Vector Machine. *Jurnal Ilmiah ILKOMINFO-Ilmu Komputer & Informatika*, *6*(2), 111–122.

Halim, A., & Safuwan, A. (2023). Analisis Sentimen opini Warganet Twitter Terhadap Tes Screening Genose Pendeteksi Virus Covid-19 Menggunakan Naive Bayes Berbasis Particle Swarm Opimization. *JINTEKS*, *5*(1), 170–178.

Hermaliani, E. H., & Kurniawati, L. (2023). Analisis Klasifikasi Sentimen Pengguna Media Sosial Twitter Terhadap Penundaan Pemilu Presiden Tahun 2024. *Jurnal Indonesia: Manajemen Informatika Dan Komunikasi*, *4*(2), 570–579.

Ilmawan, L. B., & Mude, M. A. (2020). Perbandingan metode klasifikasi Support Vector Machine dan Naïve Bayes untuk analisis sentimen pada ulasan tekstual di Google Play Store. *Ilk. J. Ilm*, *12*(2), 154–161.

Indrayuni, E. (2018). Komparasi Algoritma Naive Bayes Dan Support Vector Machine Untuk Analisa Sentimen Review Film. *Jurnal Pilar Nusa Mandiri*, *14*(2), 175–180.

Khoiruddin, Y., Fauzi, A., & Siregar, A. M. (2023). Analisis Sentimen Gojek Indonesia Pada Twitter Menggunakan Algoritme Naïve Bayes Dan Support Vector Machine. *Progresif: Jurnal Ilmiah Komputer*, *19*(1), 391–400.

Listyarini, S. N., & Anggoro, D. A. (2021). Analisis Sentimen Pilkada di Tengah Pandemi Covid-19 Menggunakan Convolution Neural Network (CNN). *Jurnal Pendidikan Dan Teknologi Indonesia*, *1*(7), 261–268.

Noviana, R., & Rasal, I. (2023). Penerapan Algoritma Naive Bayes dan SVM untuk Analisis Sentimen Boy Band BTS pada Media Sosial Twitter. *JTS*, *2*(2), 51–60.

Oktavia, D., Ramadahan, Y. R., & Minarto, M. (2023). Analisis Sentimen Terhadap Penerapan Sistem E-Tilang Pada Media Sosial Twitter Menggunakan Algoritma Support Vector Machine (SVM). *KLIK: Kajian Ilmiah Informatika Dan Komputer*, *4*(1), 407–417.

Ramadhan, D. A., & Setiawan, E. B. (2019). Analisis Sentimen Program Acara di SCTV pada Twitter Menggunakan Metode Naive Bayes dan Support Vector. *E-Proceeding of Engineering*, *6*(2), 9736–9743.

Saddam, M. A., Kurniawan, E., & Indra, I. (2023). Analisis Sentimen Fenomena PHK Massal Menggunakan Naive Bayes dan Support Vector Machine. *Jurnal Informatika: Jurnal Pengembangan IT*, *8*(3), 226–233.

Sasmita, A. B., Rahayudi, B., & Muflikhah, L. (2022). Analisis Sentimen Komentar pada Media Sosial Twitter tentang PPKM Covid-19 di Indonesia dengan Metode Naïve Bayes. *Jurnal Pengembangan Teknologi Informasi Dan Ilmu Komputer*, *6*(3), 1208–1214.

Sianipar, J. F., Ramadhan, Y. R., & Jaelani, I. (2023). Analisis Sentimen Pembangunan Kereta Cepat Jakarta-Bandung di Media Sosial Twitter Menggunakan Metode Naive Bayes. *KLIK: Kajian Ilmiah Informatika Dan Komputer*, *4*(1), 360–367.

Suasnawa, I. W., Caturbawa, I. G. N. B., Widharma, I. G. S., Sapteka, A. A. N. G., Indrayana, I. N. E., & Sunaya, I. G. A. M. (2023). *Twitter Sentiment Analysis on the Implementation of Online Learning during the Pandemic using Naive Bayes and Support Vector Machine*.

Syahlan, M. S., Irmayanti, D., & Alam, S. (2023). Analisis Sentimen Terhadap Tempat Wisata Dari Komentar Pengunjung Dengan Menggunakan Metode Support Vector Machine (Svm). *Simtek: Jurnal Sistem Informasi Dan Teknik Komputer*, *8*(2), 315–319.

Zuraiyah, T. A., Mulyati, M. M., & Harahap, G. H. F. (2023). Perbandingan Metode Naive Bayes, Support Vector Machine dan Recurrent Neural Network Pada Analisis Sentimen Ulasan Produk E-Commerce. *MULTITEK INDONESIA*, *17*(1), 27–43.