Building of Informatics, Technology and Science (BITS)
Volume 6, No 3, Desember 2024 Page: 2005−2014
ISSN 2684-8910 (media cetak)
ISSN 2685-3310 (media online)
DOI 10.47065/bits.v6i3.6447

Perbandingan Model Machine Learning dalam Analisis Sentimen Terhadap Komentar Pada Kasus Monkeypox di Media Sosial X

Devi Prasetyoningrum*, Pulung Nurtantio Andono

Fakultas Ilmu Komputer, Teknik Informatika, Universitas Dian Nuswantoro, Semarang, Indonesia
Email: [1,*]111202113378@mhs.dinus.ac.id, [2]pulung@dsn.dinus.ac.id
Email Penulis Korespondensi: 111202113378@mhs.dinus.ac.id
Submitted: **14/12/2024**; Accepted: **29/12/2024**; Published**: 30/12/2024**

Abstrak−Monkeypox atau MPOX, adalah penyakit zoonosis yg ditimbulkan sang virus monkeypox, anggota genus Orthopoxvirus. Monkeypox menjadi perhatian global setelah kasus penularan dilaporkan di berbagai negara, memicu diskusi luas di media sosial X. Platform ini sering digunakan masyarakat untuk menyebarkan informasi dan mengekspresikan kekhawatiran terkait penyakit**.** Penelitian ini bertujuan untuk membandingkan kinerja beberapa model dalam analisis sentimen terkait kasus Monkeypox di media sosial X. Model yang diuji mencakup Support Vector Machine (SVM), K-Nearest Neighbors (KNN), Naïve Bayes, dan Random Forest (RF). Data yang digunakan terdiri dari tweet yang berisi opini atau informasi mengenai Monkeypox, yang kemudian diproses melalui tahap normalisasi teks, remove stopwords, dan stemming. Selanjutnya, dilakukan pembobotan fitur menggunakan teknik TF-IDF dan seleksi fitur dengan metode Chi-Square, menghasilkan jumlah fitur optimal sebanyak 652. Hasil analisis menunjukkan bahwa SVM memberikan akurasi tertinggi sebesar 83%, dengan peningkatan 3% dari jumlah fitur sebelumnya, yaitu 500. Meskipun KNN dan Naïve Bayes menunjukkan peningkatan yang signifikan, Random Forest tidak mengalami perubahan yang signifikan pada performanya. Penelitian ini menyimpulkan bahwa SVM adalah model yang paling efektif dalam menganalisis sentimen terkait Monkeypox di media sosial X. Untuk penelitian selanjutnya, disarankan untuk mengeksplorasi teknik deep learning dan penggunaan dataset yang lebih besar untuk meningkatkan akurasi dan kedalaman analisis sentimen.

Kata Kunci: Analisis sentimen; SVM; KNN; NB; RF

Abstract−Monkeypox or MPOX, is a zoonotic disease caused by the monkeypox virus, a member of the genus Orthopoxvirus. Monkeypox became a global concern after cases of transmission were reported in various countries, sparking widespread discussion on social media X. This platform is often used by the public to disseminate information and express concerns related to the disease. This study aims to compare the performance of several models in sentiment analysis related to the Monkeypox case on social media X. The models tested include Support Vector Machine (SVM), K-Nearest Neighbors (KNN), Naïve Bayes, and Random Forest (RF). The data used consisted of tweets containing opinions or information about Monkeypox, which were then processed through the stages of text normalization, remove stopwords, and stemming. Furthermore, feature weighting was carried out using the TF-IDF technique and feature selection using the Chi-Square method, resulting in an optimal number of features of 652. The results of the analysis show that SVM provides the highest accuracy of 83%, with a 3% increase from the previous number of features, which was 500. Although KNN and Naïve Bayes showed significant improvements, Random Forest did not experience any significant changes in their performance. The study concluded that SVM is the most effective model in analyzing Monkeypox-related sentiment on social media X. For future research, it is recommended to explore deep learning techniques and the use of larger datasets to improve the accuracy and depth of sentiment analysis.

Keywords: Sentiment Analysis; SVM; KNN; NB; RF

1. PENDAHULUAN

Monkeypox atau MPOX, adalah penyakit zoonosis yg ditimbulkan sang virus monkeypox, anggota genus Orthopoxvirus. Penyakit ini pertama kali diidentifikasi dalam tahun 1958 pada kalangan kera laboratorium, ad interim masalah pertama dalam insan dilaporkan dalam tahun 1970 pada Republik Demokratik Kongo[1]. Monkeypox menjadi perhatian global setelah kasus penularan dilaporkan di berbagai negara, memicu diskusi luas di media sosial X. Platform ini sering digunakan masyarakat untuk menyebarkan informasi dan mengekspresikan kekhawatiran terkait penyakit. Sayangnya, informasi yang salah atau tidak akurat dapat memicu ketakutan dan misinformasi.

Kasus tersebut cukup ramai dibahas di dunia media sosial, banyak masyarakat yang menanggapi secara serius menganai kasus Monkeypox ini. Sebagian besar masyarakat ikut serta memberikan komentar melalui media sosial X yang dianggap cukup populer dan cukup berpengarauh dalam setiap topik yang sedang rame. Dari banyaknya komentar yang dituliskan, kasus Monkeypox ini banyak mendapatkan pro dan kontra dari masyarakat sehingga perlu dilakukan analisis sentimen untuk mengetahui apakah topik tersebut mendapat sambutan yang baik atau buruk dari masyarakat.

Analisis sentimen merupakan sebuah proses pemahaman, pengekstrakan dan pengolahan data berupa teks yang bertujuan untuk mendapatkan 3 informasi sentimen yang berasal dari kalimat opini yang diunggah ke publik[2]. Analisis sentimen pada diskusi publik di media sosial menjadi penting untuk memahami emosi masyarakat, khususnya sentimen positif dan negatif terkait isu kesehatan seperti Monkeypox [3].

Penelitian terkait analisis sentimen telah banyak dilakukan untuk memahami opini publik terhadap isu-isu tertentu. Sebagai contoh, Matheos Sarimole F. Kudrat K. (2024) dalam Jurnal Sains dan Teknologi, melakukan analisis sentimen terhadap aplikasi Satu Sehat di Twitter menggunakan algoritma Naïve Bayes dan Support Vector Machine (SVM) yang memberikan hasil evaluasi model dengan akurasi 87.95 %. Dari 1080 data uji, terprediksi 132 data [4]. Pada penelitian lain Jayanti, T. S., Budiman, B., Habibi, C., & Setiana, E. (2024). Analisis Sentimen

Building of Informatics, Technology and Science (BITS)
Volume 6, No 3, Desember 2024 Page: 2005−2014
ISSN 2684-8910 (media cetak)
ISSN 2685-3310 (media online)
DOI 10.47065/bits.v6i3.6447

Penggunaan Aplikasi Traveloka di Twitter Menggunakan Model Klasifikasi untuk mengetahui cara melakukan analisis sentimen dan melakukan analisis perbandingan serta mendapatkan hasil yang paling baik untuk analisis sentimen Traveloka di Twitter dan hasilnya menunjukkan bahwa SVM memiliki akurasi lebih baik berdasarkan evaluasi metrik dengan nilai sebesar 90%. Namun, melalui uji model menggunakan AUC, XGBOOST memperoleh nilai tertinggi sebesar 71% [5]. Penelitian lain yang mengkomparasi dua algoritma untuk melakukan analisis sentimen terhadap metaverse. Penelitian yang dilakukan oleh Putri Kumalasari dan Ryan Randi S.(2024) menggunakan algoritma Support Vector Machine(SVM) dan juga algoritma Random Forest. Hasil akhir dari penelitian ini berupa nilai akurasi kedua algoritma, algoritma SVM mendapatkan nilai akurasi sebesar 90% sedangkan untuk algoritma Random Forest mendapatkan nilai akurasi sebesar 91% [6].

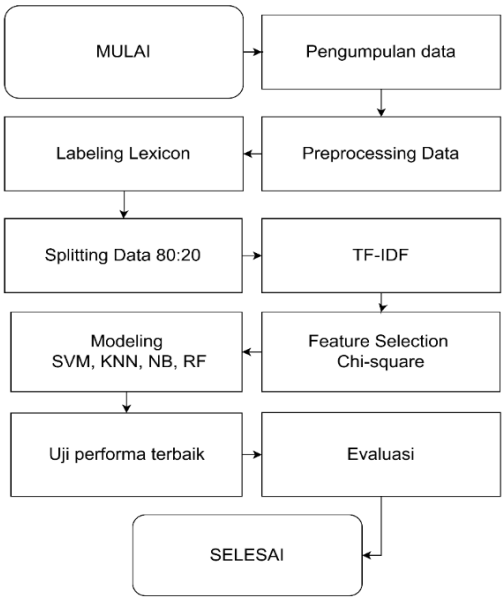
Syahril Dwi Prasetyo, Shofa Shofiah Hilabi, & Fitri Nurapriani. (2023) juga melakukan penelitian untuk mengkomparasi hasil kinerja dari algoritma Naive Bayes dan K-Nearest Neighbor(KNN) dalam melakukan analisis sentimen terhadap relokasi Ibu Kota Negara. Hasil penelitian memberikan kinerja komparatif dari teknik-teknik ini. Dengan kata lain metode Naive Bayes mempunyai akurasi sebesar 82,27%, presisi sebesar 86,36%, dan recall sebesar 76,93% untuk analisis sentimen. Kinerja metode ANN juga memberikan hasil analisis dengan presisi sebesar 88,12%, presisi sebesar 93%, dan recall sebesar 81,53%. Berdasarkan hasil analisis tersebut, proses analisis menggunakan metode KNN lebih baik dibandingkan metode NB dalam mengukur sentimen terhadap pemindahan ibu kota Indonesia [7]. Penelitian lainnya dilakukan oleh Putri Cahyani, Lufty Abdillah (2024) adalah membandingkan Performa Algoritma Naïve Bayes, SVM dan Random Forest dalam studi kasus analisis sentimen penggunaan sosial media X, data terdiri dari 10.000 tweet yang dikumpulkan dengan menggunakan kata kunci terkait IKN. Berdasarkan uji kinerja algoritma, mendapatkan hasil bahwa SVM memiliki kinerja terbaik dibandingkan dengan Naive Bayes dan Random Forest, mencapai presisi 87%, presisi 87%, recall 87%, dan skor f-1 87%. Penelitian ini menggunakan kerangka data mining CRISP-DM untuk memastikan pendekatan terstruktur dan sistematis dalam proses analisis[8].

Dari penelitian yang dilakukan sebelumnya memberikan gambaran bahwa beberapa algoritma yang telah diuji layak untuk digunakan dalam analisis sentimen tentang isu Monkeypox. Dari dasar penelitian sebelumya serta isu yang sedang dibicarakan maka muncul ide untuk melakukan penelitian analisis sentimen terhadap isu MonkeyPox tersebut. Namun, yang berbeda dalam penelitian ini dengan penelitian sebelumnya, pendekatan analisis sentimen dikembangkan lebih lanjut dengan membandingkan performa beberapa algoritma pembelajaran mesin, yaitu Naïve Bayes, Support Vector Machine, Random Forest, dan K-Nearest Neighbors. Penelitian ini bertujuan untuk mengidentifikasi model yang paling optimal dalam mengklasifikasikan sentimen positif dan negatif terkait Monkeypox di media sosial X. Hasil penelitian diharapkan dapat memberikan wawasan yang bermanfaat bagi pengambil kebijakan untuk merespons isu ini secara efektif.

2. METODOLOGI PENELITIAN

2.1 Tahapan Penelitian

Pada penelitian ini memiliki beberapa tahapan yang perlu dilakukan untuk dapat menentukan kesimpulan dari penelitian. Tahapan yang dilakukan dimulai dari proses pengambilan data kemudian melewati tahapan preprocessing data atau persiapan data sebelum masuk ke dalam pengujian model. Tahapan-tahapan yang dilakukan pada penelitian ini dapat dilihat pada Gambar 1 berikut.



Gambar 1. Flowchart Penelitian

Building of Informatics, Technology and Science (BITS)
Volume 6, No 3, Desember 2024 Page: 2005−2014
ISSN 2684-8910 (media cetak)
ISSN 2685-3310 (media online)
DOI 10.47065/bits.v6i3.6447

2.1.1 Pengumpulan Data

Pada tahap awal, dilakukan pengumpulan data melalui teknik *crawling* yang akan mengambil data berupa *tweets* dari platform media sosial X. Teknik crawling ini memungkinkan pengumpulan data dalam jumlah besar yang mencakup beragam komentar dan interaksi terkait topik yang sedang dibahas, dalam hal ini, Monkeypox. Data yang dikumpulkan akan digunakan sebagai dataset untuk analisis sentimen dalam penelitian ini. Proses pengumpulan ini penting untuk mendapatkan sampel data yang representatif dari diskusi publik yang terjadi di media sosial, yang mencerminkan berbagai pendapat dan respons masyarakat terhadap isu kesehatan ini [9].

2.1.2 Preprocessing

Tahap ini dilakukan untuk membersihkan data untuk menyesuaikan dengan inti dari tiap kata supaya mudah teridentifikasi. Data dikumpulkan, tahap berikutnya adalah preprocessing, yang bertujuan untuk membersihkan dan mempersiapkan data agar siap untuk analisis lebih lanjut. Proses ini sangat penting untuk memastikan bahwa data yang digunakan tidak mengandung noise atau elemen yang dapat mengganggu akurasi model. Proses preprocessing terdiri dari beberapa tahapan yang harus dilalui secara berurutan untuk memperoleh data yang bersih dan siap untuk dievaluasi.

a. Casefolding
   Pada tahap pertama preprocessing, dilakukan casefolding, yaitu mengubah seluruh teks menjadi huruf kecil untuk menghindari perbedaan interpretasi antara huruf besar dan kecil. Proses ini juga mencakup penghapusan karakter-karakter yang tidak relevan seperti angka, simbol, atau tanda baca yang tidak mendukung analisis sentimen. Selain itu, di tahap ini juga dilakukan pembersihan terhadap karakter *break* yang tidak diperlukan dan penghapusan *double space* yang dapat menyebabkan ketidaktepatan dalam analisis [10].
b. Normalisasi Teks
   Tahap berikutnya adalah normalisasi teks, yang bertujuan untuk mengubah kata-kata yang memiliki variasi atau istilah yang tidak konsisten menjadi bentuk yang lebih umum atau standar. Ini penting agar model dapat memahami kata-kata tersebut dengan lebih baik dan mengklasifikasikannya secara tepat. Misalnya, kata-kata slang, singkatan, atau kata-kata yang memiliki banyak variasi penulisan akan diubah menjadi bentuk yang lebih umum agar mudah diidentifikasi dan diproses lebih lanjut [11].
c. Remove Stopwords
   Tahap ini, dilakukan penghapusan kata-kata yang termasuk dalam daftar kamus stopwords. Stopwords adalah kata-kata yang sering muncul dalam teks tetapi tidak memiliki kontribusi signifikan terhadap analisis sentimen, seperti kata penghubung, artikel, atau kata-kata umum lainnya. Penghapusan stopwords bertujuan untuk mengurangi kompleksitas data dan memungkinkan model untuk fokus pada kata-kata yang lebih penting dan relevan dalam mengidentifikasi sentimen yang ada dalam teks [12].
d. Stemming
   Pada tahap stemming, yang berfungsi untuk mengubah kata-kata turunan menjadi bentuk kata dasarnya. Proses ini penting agar model dapat mengenali kata-kata yang serupa meskipun memiliki variasi bentuk, seperti kata "berlarian" yang akan diubah menjadi "lari". Dengan menggunakan stemming, model dapat mengidentifikasi akar kata yang lebih konsisten dan meningkatkan akurasi analisis sentimen yang dilakukan[13].

2.1.3 Labeling Lexicon

Data dibersihkan melalui tahap preprocessing, langkah selanjutnya adalah melakukan labeling sentimen pada setiap data. Labeling ini dilakukan dengan menggunakan metode lexicon, yang mengandalkan daftar kata dengan bobot positif dan negatif. Setiap kata dalam kalimat akan dihitung bobot positif atau negatifnya berdasarkan kamus yang telah disusun sebelumnya. Dengan demikian, setiap kalimat atau tweet yang dianalisis dapat diberi label sentimen positif, negatif, atau netral. Proses ini sangat penting untuk membantu model dalam mengklasifikasikan sentimen secara otomatis, sehingga hasil analisis sentimen dapat lebih objektif dan akurat[14].

2.1.4 Split Data

Setelah dataset sudah dilakukan preprocessing dan labeling akan dilakukan split data. Tahap ini menggunakan variabel X sebagai komentar dan Y sebagai sentimen yang kemudian data dilakukan pembagian untuk menjadi data training sebagai data untuk melatih model dan data testing sebagai data untuk menguji model. Pembagian data akan menggunakan perbandingan 80:20, dengan 80% dari total kesulurahan data untuk data training dan 20% sisanya untuk digunakan sebagai data testing[15].

2.1.5 TF-IDF

Setelah dilakukan split data yang kemudian data tersebut sudah dipisah jadi data training dan data testing yang masing-masing memiliki variabel X dan Y. Data training X akan dilakukan pembobotan pada tiap kata pada komentar. Pembobotan tiap kata dilakukan untuk mengetahui frekuensi bobot kata tesebut dan disebut sebagai *feature*[13].

Building of Informatics, Technology and Science (BITS)
Volume 6, No 3, Desember 2024 Page: 2005−2014
ISSN 2684-8910 (media cetak)
ISSN 2685-3310 (media online)
DOI 10.47065/bits.v6i3.6447

2.1.6 Feature Selection Chi-Square

Setelah terdapat data feature setelah mengalami pembobotan pada proses TF-IDF kemudian dilakukan *feature selection* dengan metode Chi-Square. Tahap ini dilakukan untuk mencari feature terbaik yang nantinya akan meningkatkan *accuracy* pada model [16].

2.1.7 Modeling

Sebelum dilakukan prediksi sentimen, diperlukan tahap pemodelan. Proses ini menggunakan variabel X dan Y pada data training sebelumnya untuk melatih model yang kemudian diuji menggunakan data training [17].

2.1.8 Uji Performa Terbaik

Setelah dilakukan modeling yang kemudian model tersebut akan dilakukan uji performa terbaik. Pengujian ini tujuannya untuk mencari performa terbaik berdasarkan accuracy yang didapatkan. Metode yang digunakan tetap menggunakan Feature Selection Chi-Square, dengan modifikasi untuk melakukan iterasi berdasarkan jumlah feature yang diinginkan yang nantinya tiap iterasi dilakukan perbandingan pada tiap pengujian terbaik berdasarkan accuracy terkini. Evaluasi dilakukan menggunakan metrik seperti akurasi, precision, recall, dan F1-score pada setiap algoritma yang diuji [18].

2.1.9 Evaluasi

Pada tahap akhir penelitian akan dilakukan evaluasi yang memberikan output berupa classification report dan confusion matrix. Pada hasil tersebut dapat disimpulkan hasil dari classification model pada kasus ini[19]. Selain itu, dilakukan uji statistik Analysis of Variance (ANOVA) untuk menguji apakah terdapat perbedaan signifikan dalam rata-rata akurasi antar algoritma. Pengujian ini bertujuan untuk memastikan bahwa perbedaan kinerja antar model adalah signifikan secara statistik, bukan sekadar hasil acak[20]. Selanjutnya, dilakukan post-hoc Tukey HSD untuk menganalisis pasangan algoritma mana yang memiliki perbedaan signifikan.

		Data Sebenarnya	
		Positif	Negatif
Hasil Prediksi	Positif	*True Positive* (TP) *correct result*	*False Positive* (FP) *unexpected result/false alarm*
	Negatif	*False Negative* (FN) *missing result*	*True Negative* (TN) *correct rejection*

Gambar 2. Confusion Matrix

Empat istilah pada Gambar 2 dijelaskan sebagai berikut:
1. True Positives (TP) adalah jumlah data kelas positif yang tepat diprediksi sebagai kelas positif.
2. True Negatives (TN) adalah jumlah data kelas negatif yang tepat diprediksi sebagai kelas negatif.
3. False Positives (FP) adalah jumlah data kelas negatif yang salah diprediksi menjadi kelas positif.
4. False Negatives (FN) adalah jumlah data kelas positif yang salah diprediksi menjadi kelas negatif.

3. HASIL DAN PEMBAHASAN

Pada penelitian ini, setelah metode penelitian yang telah direncanakan diterapkan, penelitian akan dilaksanakan sesuai dengan tahapan yang telah ditentukan. Setiap tahapan penelitian tersebut akan dijelaskan secara rinci untuk memberikan gambaran yang jelas mengenai proses yang dilakukan.

3.1 Pengumpulan Data

Pengumpulan data dilakukan menggunakan metode Crawling. Jumlah data yang dikumpulkan dengan keyword yang masih berhubungan dengan Monkeypox diperoleh sebanyak 1380 data. Data tweets yang diambil pada sosial media X memiliki rentang tanggal 1 Mei 2024 Hingga 15 Oktober 2024.

3.2 Preprocessing

Dataset yang diperoleh dilakukan preprocessing dengan beberapa tahapan secara berurutan. Proses preprocessing ini dilakukan untuk membersihkan dataset dari media sosial X agar siap digunakan dalam penelitian.

Building of Informatics, Technology and Science (BITS)
Volume 6, No 3, Desember 2024 Page: 2005−2014
ISSN 2684-8910 (media cetak)
ISSN 2685-3310 (media online)
DOI 10.47065/bits.v6i3.6447

3.2.1 Casefolding

Tahap pertama adalah tahapan casefolding yang merupakan tahapan penghilangan atribut yang tidak diperlukan seperti hastag dan username pengirim komentar. Bisa dilihat pada Tabel 1 merupakan perbandingan komentar sebelum dibersihkan dan sesudah dibersihkan.

Tabel 1. Hasil Casefolding

Text	Casefolding
@AdibNorudin @Marchfoward Boleh baca research mpox atau dgr podcast Dr Rolland. Monkey pox ni dah lama sbenarnya.plus mpox ni symptom nmpk kt bdn. dah tahu org tu ada jangan la kau pergi cium. Kalau xpuas ati meh kita bukak paper apa kita ada.	boleh baca research mpox atau dgr podcast dr rolland dgr ni dah lama sbenarnya plus mpox ni symptom nmpk kt bdn dah tahu org tu ada jangan la kau pergi cium kalau xpuas ati meh kita bukak paper apa kita ada
Kapolresta Balikpapan Gelar Sosialisasi Kesehatan dan Screening Monkeypox bagi Personel Balikpapan Kapolresta Balikpapan Komisaris Besar Polisi Anton Firmanto SH SIK MSI membuka kegiatan sosialisasi kesehatan mengenai virus Monkeypox dan mengadakan screening kesehatan https://t.co/SZSf37Ws7E	kapolresta balikpapan gelar sosialisasi kesehatan dan screening monkeypox bagi personel balikpapan kapolresta balikpapan komisaris besar polisi anton firmanto sh sik msi membuka kegiatan sosialisasi kesehatan mengenai virus monkeypox dan mengadakan screening kesehatan
@Wandadu_ Indo kena pandemi boti ngalahin monkey pox	indo kena pandemi boti ngalahin monkey pox

3.2.2 Normalisasi Teks

Pada tahap normalisasi teks, serangkaian proses dilakukan untuk mengubah teks menjadi bentuk yang lebih konsisten dan seragam, sehingga lebih mudah untuk dianalisis. Proses ini mencakup beberapa langkah, antara lain pemisahan kata (tokenization), pemeriksaan dan perbaikan ejaan, penghapusan tanda baca dan karakter khusus, serta penggantian kata atau sinonim dengan bentuk standar yang lebih mudah dikenali. Selain itu, teknik stemming dan lemmatization juga diterapkan untuk mengubah kata-kata turunan menjadi bentuk dasarnya. Hasil dari semua langkah normalisasi ini dapat dilihat pada Tabel 2.

Tabel 2. Hasil Normalisasi Teks

Casefolding	Normalisasi Teks
boleh baca research mpox atau dgr podcast dr rolland monkey pox ni dah lama sbenarnya plus mpox ni symptom nmpk kt bdn dah tahu org tu ada jangan la kau pergi cium kalau xpuas ati meh kita bukak paper apa kita ada	boleh baca penelitian monkey pox atau dengar podcast dari rolland monyet pox ini sudah lama sebenarnya plus monkey pox ini symptom tampak kita bdn sudah tahu orang itu ada jangan la kau pergi cium kalau xpuas ati meh kita bukak paper apa kita ada
kapolresta balikpapan gelar sosialisasi kesehatan dan screening monkeypox bagi personel balikpapan kapolresta balikpapan komisaris besar polisi anton firmanto sh sik msi membuka kegiatan sosialisasi kesehatan mengenai virus monkeypox dan mengadakan screening kesehatan	kapolresta balikpapan gelar sosialisasi kesehatan dan screening monkeypox bagi personel balikpapan kapolresta balikpapan komisaris besar polisi anton firmanto sh sik msi membuka kegiatan sosialisasi kesehatan mengenai virus monkeypox dan mengadakan screening kesehatan
indo kena pandemi boti ngalahin monkey pox	indonesia kena pandemi boti mengalahkan monyet pox

3.2.3 Remove Stopwords

Pada tahap penghapusan stopwords, dilakukan proses penghilangan kata-kata tertentu yang tergolong dalam daftar kamus stopwords. Kata-kata ini sering muncul dalam teks namun tidak memberikan kontribusi yang signifikan terhadap makna atau analisis sentimen. Stopwords biasanya terdiri dari kata sambung, preposisi, atau artikel, seperti "dan", "atau", "di", "ke", "yang", dan lain-lain. Tujuan penghapusan kata-kata ini adalah untuk menyederhanakan teks dengan mengeliminasi elemen-elemen yang tidak menyajikan informasi relevan, sehingga proses analisis dapat lebih terfokus pada kata-kata yang memiliki makna lebih dalam konteks sentimen. Hasil dari penghapusan stopwords ini dapat dilihat pada Tabel 3.

Tabel 3. Hasil Remove Stopwords

Normalisasi Teks	*Remove Stopwords*
boleh baca penelitian monkey pox atau dengar podcast dari rolland monyet pox ini sudah lama sebenarnya plus monkey pox ini symptom tampak kita bdn sudah tahu orang itu ada jangan la kau pergi cium kalau xpuas ati meh kita bukak paper apa kita ada	baca penelitian monkey pox dengar podcast rolland monyet pox plus monkey pox symptom bdn orang la kau pergi cium xpuas ati meh bukak paper

Building of Informatics, Technology and Science (BITS)
Volume 6, No 3, Desember 2024 Page: 2005−2014
ISSN 2684-8910 (media cetak)
ISSN 2685-3310 (media online)
DOI 10.47065/bits.v6i3.6447

kapolresta balikpapan gelar sosialisasi kesehatan dan screening monkeypox bagi personel balikpapan kapolresta balikpapan komisaris besar polisi anton firmanto sh sik msi membuka kegiatan sosialisasi kesehatan mengenai virus monkeypox dan mengadakan screening kesehatan	kapolresta balikpapan gelar sosialisasi kesehatan screening monkeypox personel balikpapan kapolresta balikpapan komisaris polisi anton firmanto sh sik msi membuka kegiatan sosialisasi kesehatan virus monkeypox mengadakan screening kesehatan
indonesia kena pandemi boti mengalahkan monyet pox	indonesia kena pandemi boti mengalahkan monyet pox

3.2.4 Stemming

Pada tahap stemming, kata-kata yang berbentuk turunan atau berimbuhan akan diubah menjadi bentuk dasar atau akar kata. Proses ini bertujuan untuk mengidentifikasi bentuk dasar dari suatu kata, sehingga kata-kata yang memiliki makna serupa dapat digabungkan dan dianalisis dengan lebih efektif. Sebagai contoh, kata "berlarian" akan disederhanakan menjadi "lari," yang merupakan bentuk dasarnya. Dengan melakukan stemming, variasi kata turunan yang memiliki arti yang sama dapat disatukan, sehingga memudahkan dalam analisis lebih lanjut. Hasil dari proses stemming ini dapat dilihat pada Tabel 4.

Tabel 4. Hasil Stemming

Remove Stopwords	*Stemming*
baca penelitian monkey pox dengar podcast rolland monyet pox plus monkey pox symptom bdn orang la kau pergi cium xpuas ati meh bukak paper	baca teliti monkey pox dengar podcast rolland monyet pox plus monkey pox symptom bdn orang la kau pergi cium xpuas ati meh bukak paper
kapolresta balikpapan gelar sosialisasi kesehatan screening monkeypox personel balikpapan kapolresta balikpapan komisaris polisi anton firmanto sh sik msi membuka kegiatan sosialisasi kesehatan virus monkeypox mengadakan screening kesehatan	kapolresta balikpapan gelar sosialisasi sehat screening monkeypox personel balikpapan kapolresta balikpapan komisaris polisi anton firmanto sh sik msi buka giat sosialisasi sehat virus monkeypox ada screening sehat
indonesia kena pandemi boti mengalahkan monyet pox	indonesia kena pandemi bot kalah monyet pox

3.3 Labeling Lexicon

Pada tahap ini, akan dilakukan penentuan sentimen menggunakan metode Lexicon. Proses ini melibatkan perbandingan kata-kata dalam ulasan dengan daftar kata berpolaritas yang telah ditentukan sebelumnya. Kata-kata tersebut dikelompokkan berdasarkan nilai sentimen, yaitu positif, negatif, atau netral. Setiap kata dalam teks ulasan akan diberikan nilai sentimen yang sesuai dengan daftar lexicon yang telah dibuat atau diadopsi. Prosesnya dimulai dengan membandingkan setiap kata dalam teks ulasan dengan daftar dalam lexicon. Apabila ditemukan kata yang bernilai positif, nilai sentimen positif akan ditambahkan; sebaliknya, untuk kata-kata negatif, nilai sentimen negatif yang akan ditambahkan. Kata-kata yang tidak terdapat dalam daftar lexicon dianggap sebagai netral dan tidak memengaruhi nilai sentimen keseluruhan.

Setelah semua kata dalam ulasan dianalisis, nilai sentimen keseluruhan ditentukan berdasarkan jumlah nilai positif dan negatif. Jika nilai positif lebih tinggi, maka sentimen dianggap positif; jika nilai negatif lebih tinggi, sentimen dianggap negatif. Apabila keduanya seimbang, sentimen akan dikategorikan sebagai netral.

Penggunaan metode Lexicon ini memungkinkan untuk mengidentifikasi sentimen secara otomatis pada tweet mengenai Monkeypox, yang sangat bermanfaat dalam mendukung analisis sentimen dalam penelitian ini. Hasil pelabelan menggunakan metode Lexicon dapat dilihat pada Tabel 5.

Tabel 5. Hasil Labeling Lexicon

Clean Text	Sentimen
baca teliti monkey pox dengar podcast rolland monyet pox plus monkey pox symptom bdn orang la kau pergi cium xpuas ati meh bukak paper	positif
kapolresta balikpapan gelar sosialisasi sehat screening monkeypox personel balikpapan kapolresta balikpapan komisaris polisi anton firmanto sh sik msi buka giat sosialisasi sehat virus monkeypox ada screening sehat	positif
indonesia kena pandemi bot kalah monyet pox	negatif

3.4 Splitting Data

Pada tahap splitting data dalam analisis sentimen, data dibagi menjadi dua bagian: 80% untuk data training dan 20% untuk data testing. Data training digunakan untuk melatih model, sementara data testing digunakan untuk menguji kinerjanya. Dalam konteks ini, variabel X berisi teks yang telah diproses (setelah normalisasi, penghapusan stopwords, dan stemming), sementara variabel Y berisi label sentimen, seperti positif, negatif, atau netral. Model dilatih untuk memprediksi label sentimen (Y) berdasarkan clean text (X).

Building of Informatics, Technology and Science (BITS)
Volume 6, No 3, Desember 2024 Page: 2005−2014
ISSN 2684-8910 (media cetak)
ISSN 2685-3310 (media online)
DOI 10.47065/bits.v6i3.6447

3.5 TF-IDF

Setelah data dibagi menjadi data training dan testing, selanjutnya data pada variabel X, yang berisi teks yang telah dibersihkan, dilakukan proses pembobotan menggunakan metode TF-IDF (Term Frequency-Inverse Document Frequency). Proses ini bertujuan untuk memberi bobot pada setiap kata dalam teks berdasarkan frekuensinya di dalam dokumen dan seberapa jarang kata tersebut muncul di seluruh korpus. Kata-kata yang sering muncul dalam dokumen tetapi jarang muncul di seluruh korpus akan diberikan bobot tinggi, sehingga lebih mencerminkan kata-kata yang relevan dan signifikan dalam analisis. Pembobotan TF-IDF ini membantu model untuk fokus pada kata-kata yang memiliki informasi penting dan mengurangi pengaruh kata-kata umum yang tidak bermakna dalam konteks analisis sentimen. Hasil pembobotan ini kemudian digunakan untuk melatih model dalam memprediksi sentimen pada data testing.

3.6 Feature Selection Chi-Square

Setelah dilakukan pembobotan pada variabel X di data training yang menghasilkan fitur berupa kata-kata dengan bobot tertentu, langkah selanjutnya adalah melakukan feature selection menggunakan metode Chi-Square. Pada tahap ini, dilakukan seleksi fitur untuk memilih kata-kata yang paling relevan dan memiliki hubungan yang signifikan dengan label sentimen (variabel Y). Metode Chi-Square digunakan untuk mengukur ketergantungan antara setiap fitur (kata) dengan label target, dengan cara menguji apakah distribusi kata tersebut dalam berbagai kelas sentimen lebih besar atau lebih kecil dari yang diharapkan secara acak. Hanya fitur yang menunjukkan ketergantungan yang kuat dengan label sentimen yang akan dipilih. Sebagai langkah awal, ditargetkan untuk memilih 500 fitur teratas berdasarkan hasil uji Chi-Square ini, yang dianggap paling berkontribusi dalam membedakan kategori sentimen. Fitur yang terpilih selanjutnya digunakan untuk melatih model, sehingga proses analisis menjadi lebih efisien dan akurat.

3.7 Modeling

Pada tahap modeling, sejumlah algoritma machine learning diterapkan untuk membangun model analisis sentimen. Di antara algoritma yang digunakan dalam penelitian ini terdapat Support Vector Machine (SVM), K-Nearest Neighbors (KNN), Random Forest (RF), dan Naive Bayes. Masing-masing algoritma tersebut memiliki keunggulan dan pendekatan yang berbeda dalam mengklasifikasikan teks berdasarkan fitur yang telah dipilih.

SVM berfungsi untuk mencari hyperplane terbaik yang dapat memisahkan kelas sentimen, sementara KNN beroperasi dengan mencari tetangga terdekat dari data yang belum diketahui untuk meramalkan kelasnya. Random Forest menggabungkan beberapa pohon keputusan untuk meningkatkan akurasi prediksi, dan Naive Bayes mengasumsikan bahwa fitur-fitur bersifat independen serta menggunakan probabilitas untuk melakukan klasifikasi.

Keempat algoritma ini dilatih menggunakan data pelatihan yang telah melalui tahap pembobotan serta seleksi fitur, dan kemudian dievaluasi dengan menggunakan data pengujian. Proses ini bertujuan untuk membandingkan kinerja masing-masing algoritma dan memilih model terbaik dalam mengklasifikasikan sentimen teks. Hasil perbandingan dengan 500 fitur terbaik menunjukkan tingkat akurasi antara 67% hingga 81%, seperti yang dapat dilihat pada Tabel 6.

Tabel 6. Perbandingan Hasil Akurasi

Model	Feature	Akurasi
Support Vector Machine	500	81%
K-Nearest Neighbors	500	67%
Naïve Bayes	500	68%
Random Forest	500	78%

3.8 Uji Performa Terbaik

Pada tahap selanjutnya, dilakukan uji performa terbaik dengan membandingkan kinerja model menggunakan berbagai jumlah fitur yang dipilih, mulai dari 10 hingga 1000 fitur, melalui teknik iterasi. Proses ini bertujuan untuk mengetahui jumlah fitur yang optimal yang memberikan kinerja terbaik dalam klasifikasi sentimen. Setiap jumlah fitur yang diuji akan dievaluasi menggunakan metrik performa seperti akurasi, presisi, recall, dan F1-score untuk setiap model yang diterapkan, baik itu SVM, KNN, Naive Bayes, maupun Random Forest. Dengan menggunakan teknik iterasi, kita dapat melihat bagaimana peningkatan jumlah fitur memengaruhi hasil klasifikasi dan menentukan titik optimal di mana model mencapai kinerja terbaik.

Hasil uji performa menunjukkan bahwa Support Vector Machine (SVM) unggul dalam seluruh metrik evaluasi, dengan nilai akurasi sebesar 0.83, precision 0.79, recall 0.89, dan F1-score 0.84. Hal ini menjadikan SVM sebagai model terbaik untuk analisis sentimen. Model Random Forest menempati posisi kedua, dengan nilai akurasi sebesar 0.78, precision 0.76, recall 0.85, dan F1-score 0.80, menunjukkan performa yang konsisten meskipun berada di bawah SVM. Sebaliknya, algoritma Naive Bayes dan KNN memiliki performa yang lebih rendah, dengan nilai F1-score masing-masing sebesar 0.72 dan 0.75. Hasil perbandingan performa masing-masing algoritma pada metrik akurasi, precision, recall, dan F1-score beserta jumlah fitur optimal ditampilkan pada Tabel 7.

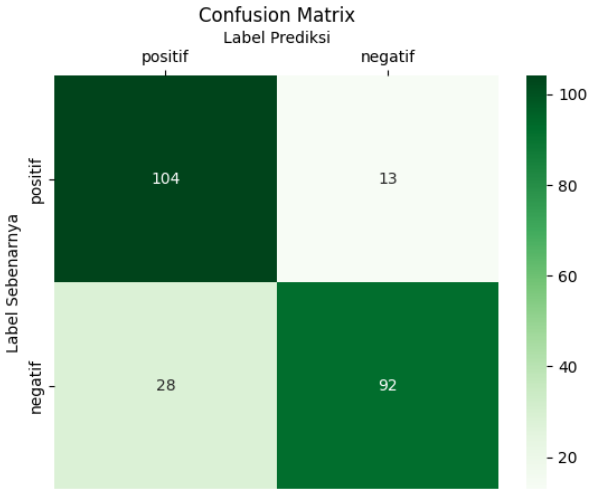
Building of Informatics, Technology and Science (BITS)
Volume 6, No 3, Desember 2024 Page: 2005−2014
ISSN 2684-8910 (media cetak)
ISSN 2685-3310 (media online)
DOI 10.47065/bits.v6i3.6447

Tabel 7. Perbandingan Kinerja Model Terbaik

Model	Feature	Akurasi	Presisi	Recall	F1-Score
Support Vector Machine	652	83%	79%	89%	84%
K-Nearest Neighbors	92	71%	65%	88%	75%
Naïve Bayes	163	75%	76%	68%	72%
Random Forest	248	78%	76%	85%	80%

3.9 Evaluasi

Pada penelitian ini, yang menggunakan model SVM, KNN, Naïve Bayes, dan Random Forest, didapatkan hasil akurasi tertinggi sebesar 83% dengan menggunakan model SVM pada jumlah fitur sebanyak 652. Model SVM ini menunjukkan peningkatan 3% dari jumlah fitur sebelumnya, yaitu 500, yang mengindikasikan bahwa penambahan fitur yang relevan dapat memberikan dampak positif terhadap performa model. Sementara itu, pada model lain, KNN dan Naïve Bayes menunjukkan kenaikan yang signifikan dalam akurasi ketika jumlah fitur ditingkatkan, namun perubahan ini tidak sebesar yang terjadi pada SVM. Di sisi lain, model Random Forest menunjukkan sedikit perubahan dalam kinerjanya meskipun jumlah fitur yang digunakan telah ditambah. Hal ini menunjukkan bahwa meskipun Random Forest cukup kuat dalam menangani data, penambahan jumlah fitur tidak memberikan peningkatan signifikan dalam akurasi model tersebut. Hasil evaluasi ini memberikan gambaran bahwa SVM dengan jumlah fitur 652 adalah model yang paling optimal dalam penelitian ini untuk tugas analisis sentimen. Hasil classification report dapat dilihat pada Gambar 3.



Gambar 3. Matriks Evaluasi

Untuk membuktikan perbedaan signifikan antar model, dilakukan pengujian statistik menggunakan Analysis of Variance (ANOVA) dan Honestly Significant Difference (Tukey HSD). Hasil uji ANOVA menunjukkan adanya perbedaan signifikan pada rata-rata akurasi antar algoritma (F-Statistic: 96.99, P-Value: 1.25e-06). Ini mengindikasikan bahwa perbedaan kinerja antar algoritma bukan terjadi secara kebetulan, dan pemilihan model yang tepat sangat berpengaruh pada hasil analisis sentimen.

Tabel 8. Post-Hoc

Group 1	Group 2	Mean Diff	P-Adj	Lower	Upper	Reject
KNN	Naive Bayes	0.04	0.0052	0.0139	0.0661	True
KNN	Random Forest	0.09	0.0000	0.0639	0.1161	True
KNN	SVM	0.13	0.0000	0.1039	0.1561	True
Naive Bayes	Random Forest	0.05	0.0013	0.0239	0.0761	True
Naive Bayes	SVM	0.09	0.0000	0.0639	0.1161	True
Random Forest	SVM	0.04	0.0052	0.0139	0.0661	True

Berdasarkan uji Tukey HSD, hasil perbandingan akurasi antar algoritma menunjukkan perbedaan yang signifikan pada semua pasangan algoritma dengan tingkat kepercayaan 95% (p-value < 0.05). SVM memiliki rata-rata akurasi tertinggi dibanding algoritma lainnya, sedangkan KNN memiliki rata-rata akurasi terendah. Perbedaan signifikan ini mengindikasikan bahwa pemilihan algoritma berpengaruh besar terhadap kinerja model pada tugas klasifikasi sentimen.

Building of Informatics, Technology and Science (BITS)
Volume 6, No 3, Desember 2024 Page: 2005−2014
ISSN 2684-8910 (media cetak)
ISSN 2685-3310 (media online)
DOI 10.47065/bits.v6i3.6447

4. KESIMPULAN

Berdasarkan hasil penelitian ini, dapat disimpulkan bahwa model Support Vector Machine (SVM) memberikan performa terbaik dalam analisis sentimen terhadap kasus Monkeypox di media sosial X, dengan akurasi mencapai 83% pada jumlah fitur sebanyak 652. Peningkatan jumlah fitur dari 500 ke 652 memberikan kontribusi positif terhadap peningkatan akurasi model, menunjukkan pentingnya pemilihan fitur yang tepat dalam meningkatkan kinerja klasifikasi sentimen. Meskipun model lain seperti K-Nearest Neighbors (KNN) dan Naïve Bayes juga menunjukkan peningkatan yang signifikan dalam akurasi, perbedaan kinerjanya tidak sebesar yang terlihat pada SVM. Di sisi lain, Random Forest menunjukkan performa yang stabil, namun tidak mengalami peningkatan signifikan meskipun jumlah fitur ditambah. Selanjutnya, uji ANOVA yang dilakukan untuk mengevaluasi perbedaan akurasi antar model menunjukkan bahwa terdapat perbedaan yang signifikan di antara model-model yang diuji, dengan SVM yang secara konsisten menunjukkan kinerja superior dibandingkan dengan model lainnya. Hasil ANOVA mendukung temuan bahwa pemilihan model yang tepat, bersama dengan penyesuaian jumlah fitur, dapat memberikan dampak yang signifikan terhadap akurasi klasifikasi sentimen. Secara keseluruhan, SVM dengan jumlah fitur 652 terbukti menjadi model yang paling efektif dalam mengklasifikasikan sentimen terkait topik Monkeypox di media sosial X, dan hasil uji ANOVA memperkuat validitas temuan ini.

REFERENCES

[1] I Ketut Suarayasa, Zulkifli, and O. mazmur Kristoper, "Mekanisme penyebaran Cacar Monyet Dan Faktor-Faktor Yang Mempengaruhinya," *SEHATMAS: Jurnal Ilmiah Kesehatan Masyarakat*, vol. 2, no. 1, pp. 28–34, Jan. 2023. doi:10.55123/sehatmas.v2i1.980

[2] T. T. Widowati and M. Sadikin, "Analisis sentimen Twitter Terhadap tokoh publik Dengan Algoritma naive Bayes dan support Vector Machine," *Simetris: Jurnal Teknik Mesin, Elektro dan Ilmu Komputer*, vol. 11, no. 2, pp. 626–636, Oct. 2021. doi:10.24176/simet.v11i2.4568

[3] D. Ahmad Dzulhijjah, H. Sanjaya, A. Said Wahyudi Hidayat, A. Yulistia Alwanda, and E. Utami, "Perbandingan metode random forest Dan Knn Pada Analisis Sentimen twitter," *Smart Comp: Jurnalnya Orang Pintar Komputer*, vol. 12, no. 3, pp. 767–772, Jul. 2023. doi:10.30591/smartcomp.v12i3.5106

[4] F. Matheos Sarimole and K. Kudrat, "Analisis Sentimen Terhadap aplikasi Satu Sehat pada twitter Menggunakan algoritma naive Bayes Dan Support Vector Machine," *Jurnal Sains dan Teknologi*, vol. 5, no. 3, pp. 783–790, Feb. 2024. doi:10.55338/saintek.v5i3.2702

[5] D. Atmajaya, A. Febrianti, and H. Darwis, "Metode SVM Dan Naive Bayes Untuk Analisis Sentimen chatgpt di twitter," *The Indonesian Journal of Computer Science*, vol. 12, no. 4, Aug. 2023. doi:10.33022/ijcs.v12i4.3341

[6] P. K. Sari and R. R. Suryono, "Komparasi Algoritma Support vector machine dan random forest Untuk Analisis Sentimen metaverse," *Jurnal Mnemonic*, vol. 7, no. 1, pp. 31–39, Feb. 2024. doi:10.36040/mnemonic.v7i1.8977

[7] A. C. Prasetyo, M. P. Arnandi, H. S. Hudnanto, and B. Setiaji, "Perbandingan Algoritma astar Dan Dijkistra dalam menentukan Rute Terdekat," *SISFOTENIKA*, vol. 9, no. 1, p. 36, Feb. 2019. doi:10.30700/jst.v9i1.456

[8] T. A.M and A. Yaqin, "Perbandingan algoritma naïve Bayes, K-nearest neighbors Dan Random Forest Untuk Klasifikasi Sentimen TERHADAP BPJS kesehatan pada media twitter," *InComTech : Jurnal Telekomunikasi dan Komputer*, vol. 12, no. 1, p. 01, Apr. 2022. doi:10.22441/incomtech.v12i1.13642

[9] M. Yasir, Marissa Grace Haque, Robertus Suraji, and Istianingsih, "Analisis Sentimen Terhadap kontroversi fatwa Mui Nomor 83 Tahun 2023 Tentang pemboikotan produk Yang Terafiliasi Israel," *Jurnal Ekonomi Manajemen Sistem Informasi*, vol. 5, no. 4, pp. 409–422, Mar. 2024. doi:10.31933/jemsi.v5i4.1845

[10] W. Yulita, "Analisis Sentimen Terhadap opini Masyarakat Tentang Vaksin covid-19 Menggunakan algoritma naïve Bayes classifier," *Jurnal Data Mining dan Sistem Informasi*, vol. 2, no. 2, p. 1, Aug. 2021. doi:10.33365/jdmsi.v2i2.1344

[11] A. C. Najib, A. Irsyad, G. A. Qandi, and N. A. Rakhmawati, "Perbandingan metode lexicon-based Dan SVM Untuk Analisis Sentimen Berbasis Ontologi Pada Kampanye pilpres Indonesia tahun 2019 di twitter," *Fountain of Informatics Journal*, vol. 4, no. 2, p. 41, Nov. 2019. doi:10.21111/fij.v4i2.3573

[12] M. Samantri and Afiyati, "Perbandingan Algoritma support vector machine dan random forest Untuk Analisis Sentimen TERHADAP kebijakan Pemerintah Indonesia Terkait Kenaikan Harga BBM Tahun 2022," *Jurnal JTIK (Jurnal Teknologi Informasi dan Komunikasi)*, vol. 8, no. 1, pp. 1–9, Jan. 2024. doi:10.35870/jtik.v8i1.1202

[13] N. Sidauruk, N. Riza, and Rd. N. Siti Fatonah, "Penggunaan metode SVM Dan Random Forest Untuk Analisis Sentimen Ulasan Pengguna Terhadap Kai Access di Google Playstore," *JATI (Jurnal Mahasiswa Teknik Informatika)*, vol. 7, no. 3, pp. 1901–1906, Nov. 2023. doi:10.36040/jati.v7i3.6899

[14] A. N. Syafia, M. F. Hidayattullah, and W. Suteddy, "Studi Komparasi Algoritma SVM Dan Random Forest Pada Analisis Sentimen Komentar YouTube BTS," *Jurnal Informatika: Jurnal Pengembangan IT*, vol. 8, no. 3, pp. 207–212, Sep. 2023. doi:10.30591/jpit.v8i3.5064

[15] S. S. Salim and J. Mayary, "Analisis Sentimen Pengguna Twitter Terhadap Dompet Elektronik Dengan METODE lexicon based Dan K – Nearest Neighbor," *Jurnal Ilmiah Informatika Komputer*, vol. 25, no. 1, pp. 1–17, 2020. doi:10.35760/ik.2020.v25i1.2411

[16] K. T. Putra, M. A. Hariyadi, and C. Crysdian, "Perbandingan Feature extraction TF-IDF dan BOW Untuk Analisis Sentimen Berbasis SVM," *Jurnal Cahaya Mandalika*, vol. 3, no. 2, pp. 1449-1463, Nov. 2023.

[17] D. E. Sondakh, S.Kom, M.T, Ph.D, S. W. Taju, M. G. Tene, and A. E. Pangaila, "Sistem Analisis Sentimen Ulasan Aplikasi belanja online menggunakan metode ensemble learning," *CogITo Smart Journal*, vol. 9, no. 2, pp. 280–291, Dec. 2023. doi:10.31154/cogito.v9i2.525.280-291

Building of Informatics, Technology and Science (BITS)
Volume 6, No 3, Desember 2024 Page: 2005−2014
ISSN 2684-8910 (media cetak)
ISSN 2685-3310 (media online)
DOI 10.47065/bits.v6i3.6447

[18] E. Hokijuliandy, H. Napitupulu, and F. Firdaniza, "Analisis Sentimen menggunakan metode Klasifikasi support vector machine (SVM) Dan Seleksi Fitur Chi-Square," *SisInfo : Jurnal Sistem Informasi dan Informatika*, vol. 5, no. 2, pp. 40–49, Aug. 2023. doi:10.37278/sisinfo.v5i2.670

[19] A. A. Saputro, "Sistem Pendukung keputusan Penerimaan bantuan sosial program Keluarga Harapan (PKH) Dengan Menggunakan metode naïve Bayes classifier (Studi Kasus di Balai Desa Bendungan kraton pasuruan)," *Jurnal Ilmiah Edutic : Pendidikan dan Informatika*, vol. 9, no. 1, pp. 40–48, Nov. 2022. doi:10.21107/edutic.v9i1.12232

[20] I. Hendrawan Rifky, E. Utami, and A. Hartanto Dwi, "Analisis Perbandingan metode tf-IDF dan Word2vec Pada klasifikasi teks sentimen masyarakat TERHADAP Produk Lokal di Indonesia," *Smart Comp: Jurnalnya Orang Pintar Komputer*, vol. 11, no. 3, Jul. 2022. doi:10.30591/smartcomp.v11i3.3902