

TEXT MINING UNTUK MENDETEKSI EMOSI PENGGUNA TERHADAP “NUSANTARA” SEBAGAI NAMA IKN

**I Komang Dharmendra¹⁾, Ricky Aurelius Nurtanto Diaz²⁾, Muhamad Samsudin³⁾,
I Gusti Agung Ngurah Rai Semadi⁴⁾ I Made Agus Wirahadi Putra⁵⁾**

Program Studi Sistem Informasi^{1) 2) 3) 4)} Program Studi Manajemen Informatika⁵⁾

Fakultas Informatika dan Komputer, Institut Teknologi Dan Bisnis STIKOM Bali, Denpasar, Bali^{1) 2) 3) 4)}

dharmendra@stikom-bali.ac.id¹⁾ ricky@stikom-bali.ac.id²⁾ samsudin@stikom-bali.ac.id³⁾

raisemadi@stikom-bali.ac.id⁴⁾ wirahadi@stikom-bali.ac.id⁵⁾

ABSTRACT

The capital city plays a crucial role in the governance, serving as the political and economic center of a country. For instance, Indonesia has decided to relocate its capital city from Jakarta to East Kalimantan and construct a new city named Nusantara. The public's reaction to this announcement has been widely discussed on Twitter, a social media platform that allows users to share their opinions and responses on the topic. To analyze the Twitter users' emotions regarding this announcement, Text Mining method was employed to extract information from unstructured text data. In previous research, a dataset of tweets with categories of like, anger, sadness, happiness, and fear was utilized as the training data, while a new dataset consisting of 83,590 tweets was used as the test data. By employing the SVM model with a 'linear' kernel, it can be concluded that the announcement of the new capital city Nusantara has evoked various emotions among the public, including happiness with 39,219 occurrences, anger with 37,594 occurrences, sadness with 5,999 occurrences, liking with 397 occurrences, and fear with 381 occurrences. During the classification process, an imbalance in the number of data points in the emotional classes used as training data to build the model has led to overfitting or underfitting, which subsequently impacted the classification results on the "Nusantara" dataset. This study can serve as a foundation for future research in handling imbalanced datasets in emotion classification.

Keywords: Text Mining, Emotion Detection, Twitter, Nusantara.

ABSTRAK

Ibu kota negara memainkan peran penting dalam pemerintahan, baik sebagai pusat kekuasaan politik maupun perekonomian suatu negara. Indonesia, misalnya, telah memutuskan untuk memindahkan ibu kota negaranya dari Jakarta ke Kalimantan Timur dan membangun kota baru yang diberi nama Nusantara. Reaksi masyarakat terhadap pengumuman ini banyak dibahas di media sosial Twitter, yang memungkinkan pengguna untuk berbagi opini dan tanggapan mereka terhadap topik ini. Untuk menganalisis emosi pengguna Twitter terhadap pengumuman ini, digunakan metode Teks Mining untuk mengekstrak informasi dari data teks yang tidak terstruktur. Dalam penelitian terdahulu, dataset tweet dengan kategori suka, marah, sedih, senang, dan takut digunakan sebagai data latih, dan dataset tweet baru digunakan sebagai data uji berjumlah 83.590 tweet. Menggunakan model SVM dengan kernel 'linear', dapat disimpulkan bahwa pengumuman mengenai ibu kota baru Nusantara menimbulkan emosi yang beragam di kalangan masyarakat, di antaranya senang berjumlah 39.219, marah berjumlah 37.594, sedih berjumlah 5.999, suka berjumlah 397, dan takut berjumlah 381. Pada proses klasifikasi terjadi ketidakseimbangan jumlah data pada kelas emosi yang dijadikan data latih untuk membangun model yang menyebabkan *overfitting* atau *underfitting*, yang pada akhirnya mempengaruhi hasil klasifikasi pada dataset "Nusantara". Penelitian ini bisa menjadi landasan untuk penelitian berikutnya dalam menangani dataset yang tidak seimbang pada klasifikasi emosi.

Kata Kunci : Teks Mining, Deteksi Emosi, Twitter, Nusantara

PENDAHULUAN

Ibukota memiliki peran yang penting bagi berbagai aspek pemerintahan, ibu kota mempunyai fungsi sebagai pusat kekuasaan politik maupun perekonomian suatu negara. Tidak hanya itu ibu kota juga mencerminkan sisi kebudayaan dari negara tersebut yang menunjukkan sebuah karakter yang unik dan khas dari negara tersebut. Sebagai identitas dari suatu negara, ibu kota dibangun untuk memajukan negara tersebut agar masyarakatnya menjadi makmur dan berkehidupan yang cukup.

Dalam prosesnya, terkadang kepala pemerintahan sebuah negara memindahkan ibukota negara, baik dipindahkan ke kota yang sudah ada atau membangun kota baru yang dibangun secara khusus untuk menjadi ibu kota negara. Seperti yang dilakukan Indonesia yang berencana memindahkan ibu kota negara yang sebelumnya berada di Jakarta, dipindahkan ke Kalimantan Timur dan membangun kota baru untuk mejadi ibu kota negara dengan nama Nusantara[1]. Pemindahan ibu kota dilakukan untuk membagi pusat ekonomi dan pusat pemerintahan yang sebelumnya terpusat di Jakarta dipisah menjadi pusat ekonomi di Jakarta dan pusat pemerintahan di Nusantara. Pengumuman nama ibu kota baru Indonesia mendapat reaksi dari masyarakat, dengan berbagai opini yang disampaikan di banyak kanal, mulai dari reaksi di berbagai kanal video, hingga tulisan yang dibagikan melalui sosial media. Salah satu kanal sosial media yang banyak digunakan adalah Twitter.

Twitter merupakan sosial media yang memungkinkan penggunaannya untuk mengunggah teks, gambar, atau video pendek[2]. Konten yang diunggah pengguna dapat ditanggapi oleh pengguna lainnya, sehingga memungkinkan untuk sebuah topik yang dibahas oleh seorang pengguna dapat berkembang dan meluas seiring semakin banyak tanggapan dari pengguna lainnya[3].

Untuk mendeteksi emosi dari teks pada layanan media sosial twitter dengan data yang tidak terstruktur maka perlu dilakukan analisis teks salah satunya dengan menggunakan *Text Mining*. *Text Mining* mencoba untuk mengekstrak informasi yang berguna dari sumber data melalui identifikasi dan eksplorasi

dari suatu pola menarik. Sumber data berupa sekumpulan dokumen dan pola menarik yang tidak ditemukan dalam bentuk *database record*, tetapi dalam data teks yang tidak terstruktur.

Dengan banyaknya tanggapan masyarakat melalui media sosial, perlu sebuah proses untuk mengetahui bagaimana respon emosi pengguna twitter atas pengumuman nama ibu kota negara Nusantara. Salah satu proses yang bisa dilakukan adalah melakukan deteksi emosi terhadap tweet yang mengandung kata “Nusantara”.

Untuk penentuan label dan data latih digunakan dataset dari penelitian yang dilakukan oleh Mei Silviana Saputri; Rahmad Mahendra; Mirna Adriani[4]. Dimana pada penelitian tersebut terdapat 4.403 data tweet dengan kategori suka, marah, sedih, senang, dan takut. Untuk data uji yang digunakan adalah data tweet yang dikumpulkan selama 1 bulan, dari 15 Januari 2022 – 18 Februari 2022 yang berjumlah 83590 tweet.

Dengan menggunakan dataset yang telah memiliki label diharapkan bisa mengelompokkan tweet dengan kata “Nusantara”, sehingga mampu diketahui bagaimana reaksi emosi masyarakat terhadap pemilihan nama “Nusantara” sebagai nama ibu kota negara Indonesia yang baru.

TINJAUAN PUSTAKA

Pengertian Nusantara

Nusantara, secara resmi bernama Ibu Kota Nusantara (disingkat IKN) adalah sebuah kota terencana yang akan menggantikan posisi Jakarta sebagai ibu kota negara Indonesia mulai tahun 2024[1]. Nusantara terletak di pesisir timur pulau Kalimantan yang sebelumnya merupakan bagian dari dua kabupaten di Kalimantan Timur, yakni Kabupaten Penajam Paser Utara dan Kabupaten Kutai Kartanegara yang merupakan wilayah Metropolitan Sambo Tenggarong. Pada 17 Januari 2022, Pemerintah Indonesia dan Dewan Perwakilan Rakyat telah resmi mengumumkan nama ibu kota baru yang tertuang dalam undang-undang yang mendapat persetujuan secara aklamasi dalam rapat paripurna ke-13 DPR RI masa sidang 2021–2022 mengenai pemindahan ibu kota ke Kalimantan Timur

Twitter

Twitter adalah jejaring sosial atau *microblog* yang memungkinkan pengguna untuk mengirim dan membaca pesan berbasis teks hingga 140 karakter, yang dikenal dengan sebutan kicauan (tweet)[5]. Twitter didirikan pada bulan maret 2006 oleh Jack Dorsey, dan situs jejaring sosialnya diluncurkan pada bulan juli 2006. Twitter telah menjadi alat yang menarik bagi berbagai kalangan untuk mengikuti keinginan para pengguna terhadap setiap kondisi secara langsung. Hal ini menjadi sumber data yang potensial untuk digunakan oleh jutaan orang. Twitter membuat segala tersedia dalam sebuah *data stream*, yang dapat dimanfaatkan dengan menggunakan teknik stream mining. Secara prinsip hal ini membuat kita mengetahui opini publik secara umum. Micro-blogging adalah sumber informasi baru yang menarik untuk pengolahan data [6].

Emosi

Emosi berasal dari bahasa Latin, yaitu *emovere*, yang berarti bergerak menjauh. Arti kata ini menyiratkan bahwa kecenderungan bertindak merupakan hal mutlak dalam emosi[7]. Daniel Goleman mengatakan bahwa emosi merujuk pada suatu perasaan dan pikiran yang khas, suatu keadaan biologis dan psikologis dan serangkaian kecenderungan untuk bertindak[8]. Emosi merupakan reaksi terhadap rangsangan dari luar dan dalam diri individu, sebagai contoh emosi gembira mendorong perubahan suasana hati seseorang, sehingga secara fisiologi terlihat tertawa, emosi sedih mendorong seseorang berperilaku menangis. Emosi sebagai suatu keadaan yang terangsang dari organisme mencakup perubahan-perubahan yang disadari, yang mendalam sifatnya, dan perubahan perilaku. Emosi sebagai suatu keadaan yang kompleks yang berlangsung tidak lama yang mempunyai komponen pada badan dan pada jiwa individu tersebut. Emosi menunjukkan perubahan organisme yang disertai oleh gejala-gejala kesadaran, keperilakuan dan proses fisiologis. Kesadaran apabila seseorang mengetahui makna situasi yang sedang terjadi[9]. Jantung berdetak lebih cepat, kulit memberikan respon dengan mengeluarkan keringat dan napas terengah-engah termasuk dalam proses

fisiologis dan terakhir apabila orang tersebut melakukan suatu tindakan sebagai suatu akibat yang terjadi.

Support Vector Machines

SVM (*Support Vector Machine*) adalah sebuah metode dalam pembelajaran mesin yang digunakan untuk klasifikasi dan regresi. Metode ini termasuk dalam kelompok algoritma pembelajaran terawasi, yang berarti mereka memerlukan data yang telah diklasifikasikan sebelumnya untuk melakukan pelatihan dan membuat prediksi[10].

Secara umum, SVM berfokus pada pemisahan dua kelas data yang berbeda dengan membangun *hiperplane* (bidang dalam dimensi lebih tinggi) yang memaksimalkan jarak antara kelas-kelas tersebut. *Hiperplane* ini dipilih sedemikian rupa sehingga terdapat margin yang maksimal antara *hiperplane* dan titik-titik data yang paling dekat dari masing-masing kelas. Titik-titik data yang berada di tepi margin ini disebut "vektor pendukung" (*support vectors*), dan mereka memiliki peran penting dalam menentukan posisi dan orientasi *hiperplane*.

Dalam konteks klasifikasi, ketika SVM digunakan, tujuannya adalah untuk menemukan *hiperplane* yang memisahkan dua kelas data dengan margin maksimal. Jarak dari vektor pendukung ke *hiperplane* disebut sebagai "*margin*". SVM bertujuan untuk menemukan *hiperplane* yang memiliki margin maksimal, karena ini dapat meningkatkan kemampuan umum model untuk mengklasifikasikan data yang belum pernah dilihat sebelumnya dengan akurat.

Matrix evaluasi

Matrix evaluasi, atau sering disebut juga sebagai *confusion matrix* (matriks kebingungan), adalah alat yang digunakan untuk mengukur kinerja model klasifikasi dengan membandingkan hasil prediksi model dengan label sebenarnya dari data[11]. Matrix evaluasi membantu untuk memahami sejauh mana model berhasil mengklasifikasikan data dengan benar dalam berbagai kategori.

Tabel 1. *Confusion Matrix*

		Nilai Aktual	
		Positif	Negatif
Nilai Prediksi	Positif	TP	FN
	Negatif	FN	TN

Dalam *confusion matrix*, data diklasifikasikan menjadi empat kategori:

- **True Positive (TP):** Data yang benar-benar masuk ke dalam kelas positif dan juga diprediksi dengan benar oleh model sebagai kelas positif.
- **True Negative (TN):** Data yang benar-benar masuk ke dalam kelas negatif dan juga diprediksi dengan benar oleh model sebagai kelas negatif.
- **False Positive (FP):** Data yang sebenarnya masuk ke dalam kelas negatif, tetapi diprediksi oleh model sebagai kelas positif. Juga dikenal sebagai "Type I error" atau kesalahan tipe I.
- **False Negative (FN):** Data yang sebenarnya masuk ke dalam kelas positif, tetapi diprediksi oleh model sebagai kelas negatif. Juga dikenal sebagai "Type II error" atau kesalahan tipe II.

Dengan menggunakan nilai-nilai dari TP, TN, FP, dan FN, dapat dihitung beberapa metrik evaluasi yang berguna:

- **Akurasi (Accuracy):** Mengukur sejauh mana model dapat mengklasifikasikan data secara benar, yang ditunjukkan dalam persamaan 1.

$$Accuracy = \frac{TP+TN}{TP+TN+FP+FN} \quad (1)$$

- **Presisi (Precision):** Mengukur sejauh mana prediksi positif model benar, yang ditunjukkan dalam persamaan 2.

$$Presisi = \frac{TP}{(TP+FP)} \quad (2)$$

- **Recall (Sensitivitas atau True Positive Rate):** Mengukur sejauh mana model dapat mendeteksi semua instance positif yang sebenarnya, yang ditunjukkan dalam persamaan 3

$$Recall = \frac{TP}{(TP+FN)} \quad (3)$$

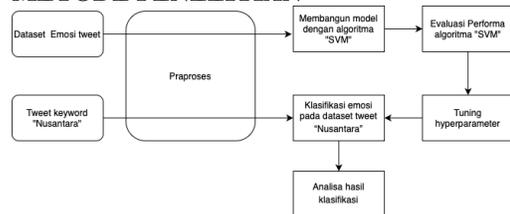
- **F1-Score:** Kombinasi dari *presisi* dan *recall* yang memberikan keseimbangan

antara keduanya, yang ditunjukkan dalam persamaan 4.

$$F1 - Score = \frac{2 \times (Precision \times Recall)}{Precision + Recall} \quad (4)$$

Confusion matrix dan metrik-metrik ini sangat penting dalam mengevaluasi kinerja model klasifikasi, terutama ketika terjadi ketidakseimbangan antara jumlah *instance* dalam setiap kelas. Dengan memahami distribusi hasil prediksi dan kesalahan yang dilakukan oleh model, dapat dibuat keputusan yang lebih baik dalam memperbaiki atau meningkatkan model tersebut.

METODE PENELITIAN



Gambar 1. Metode Penelitian

Teknik Pengumpulan Data

Dalam penelitian dilakukan pengumpulan 2 data, data latih digunakan dataset dari penelitian yang dilakukan oleh Mei Silviana Saputri; Rahmad Mahendra; Mirna Adriani[4]. Dimana pada penelitian tersebut terdapat 4.403 data tweet dengan kategori suka, marah, sedih, senang, dan takut. Untuk data uji yang digunakan adalah data tweet yang mengandung kata nusantara yang dikumpulkan selama 1 bulan, dari 15 Januari 2022 – 18 Februari 2022 yang berjumlah 83590 tweet.

kelas	jumlah
anger	1101
happy	1017
sadness	997
fear	649
love	637

Praproses

Kedua data kemudian diproses untuk untuk menyeleksi data text agar menjadi lebih terstruktur dan menghilangkan nilai yang

tidak diperlukan. Praproses dilakukan untuk menyeleksi data teks agar menjadi lebih terstruktur dan menghilangkan nilai yang tidak diperlukan[12]. Praproses dilakukan pada kedua data, dataset untuk membangun model, dan data tweet yang mengandung kata “Nusantara”. Tabel 2 menunjukkan contoh praproses pada teks tweet.

Tabel 2. Contoh Praproses

Teks Asli
Ngaku paling nusantara tapi pajak aja menunggak
Teks Hasil Praproses
ngaku nusantara pajak tunggak

Membagi Dataset

Pada tahap awal yang dilakukan adalah membagi dataset menjadi 2 kelompok, yaitu data latih, dan data uji dengan ratio 80:20. Dimana pada penelitian ini pembagian yang dilakukan adalah : data latih: 3520 tweet, dan data uji: 881 tweet

Pembobotan

Setelah membagi dataset, dilanjutkan dengan melakukan pembobotan dengan menggunakan TF-IDF (*Term Frequency-Inverse Document Frequency*). Metode ini menghitung bobot kata (term) pada sebuah data latih dengan mempertimbangkan frekuensi kemunculan kata tersebut pada dokumen itu sendiri (TF) dan seberapa umum kata tersebut dalam seluruh dokumen dalam koleksi tersebut (IDF). Dengan cara ini, kata-kata yang sering muncul dalam sebuah dokumen tetapi jarang muncul pada dokumen lainnya akan memiliki bobot yang lebih tinggi. Setelah menggunakan TF-IDF, didapatkan jumlah fitur pada data latih sejumlah 11.909.

Membangun Model SVM

Pada pembangunan model dilakukan dengan menggunakan *Support Vector Machine* (SVM) untuk melakukan klasifikasi pada data teks

dengan menggunakan representasi TF-IDF sebagai pembobotan fitur. Tahap pembangunan model menggunakan SVM dengan *kernel linear*, kemudian model yang dibangun akan diuji menggunakan data uji yang berjumlah 881 tweet.

Evaluasi Performa

Proses dilanjutkan dengan mengevaluasi performa model SVM menggunakan *confusion matrix* dan *classification report*. Confusion matrix digunakan untuk menghitung jumlah prediksi yang benar atau salah dari model, sedangkan *classification report* memberikan informasi mengenai *precision*, *recall*, dan *f1-score* dari model untuk setiap kelas yang ada pada data uji.

Klasifikasi Pada Dataset “Nusantara”

Untuk melakukan klasifikasi emosi pada dataset tweet “Nusantara”, data tweet “Nusantara” yang sudah dipraproses diubah menjadi vektor TF-IDF. Kemudian melakukan klasifikasi menggunakan model SVM yang sudah dibuat sebelumnya, dimana model model SVM tersebut digunakan untuk memprediksi label emosi pada data tweet “Nusantara”.

Hasil & Pembahasan

Pada pengujian model SVM dengan menggunakan dataset tweet dari penelitian Mei Silviana Saputri, dengan menggunakan matrik evaluasi didapatkan tingkat akurasi sebesar 64,13%. Dalam konteks matriks evaluasi, akurasi (*accuracy*) adalah salah satu metrik yang digunakan untuk mengukur kinerja atau performa dari suatu model atau sistem klasifikasi. Akurasi mengukur seberapa tepat model dalam mengklasifikasikan data ke dalam kelas yang benar.

Untuk presisi sebesar 68,02%, presisi (*precision*) adalah salah satu metrik yang digunakan untuk mengukur kinerja atau performa dari suatu model atau sistem klasifikasi. Presisi mengukur seberapa tepat model dalam mengidentifikasi data positif.

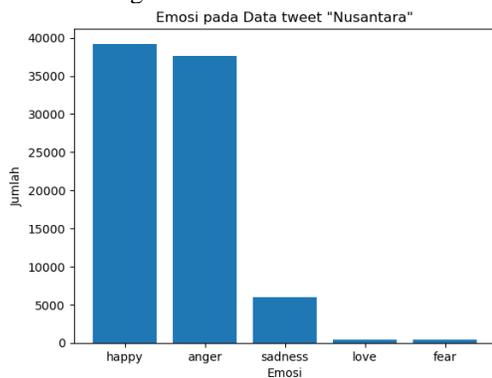
Dalam konteks matriks evaluasi, *recall* (juga dikenal sebagai sensitivitas atau *true positive rate*) adalah salah satu metrik yang digunakan untuk mengukur kinerja atau

performa dari suatu model atau sistem klasifikasi. *Recall* mengukur seberapa baik model dalam menemukan kembali atau mengidentifikasi data positif sebenarnya. Dimana nilai *recall* sebesar 64,29%.

F1-Score adalah salah satu metrik yang digunakan untuk mengukur kinerja atau performa dari suatu model atau sistem klasifikasi. *F1-Score* adalah *harmonic mean* (rata-rata harmonis) dari presisi (*precision*) dan *recall*. Metrik ini memberikan keseimbangan antara presisi dan recall, dan sangat berguna ketika kelas dalam dataset tidak seimbang. Dan pada penelitian didapatkan nilai *F1-score* sebesar 65,62%

Classification Report:				
	precision	recall	f1-score	support
anger	0.64	0.74	0.68	229
fear	0.88	0.62	0.73	119
happy	0.62	0.64	0.63	214
love	0.76	0.71	0.74	119
sadness	0.50	0.51	0.50	200
accuracy			0.64	881
macro avg	0.68	0.64	0.66	881
weighted avg	0.65	0.64	0.64	881

Dari hasil matrik evaluasi didapatkan hasil bahwa klasifikasi emosi menggunakan SVM cukup baik dengan nilai akurasi yang bagus, selanjutnya dilakukan Klasifikasi pada dataset tweet yang mengandung kata "Nusantara". Dataset tweet "Nusantara" sebelumnya telah dipraproses dan dilakukan pembobotan kemudian diklasifikasi dengan model yang sudah dibangun.



Dari hasil klasifikasi menunjukkan bahwa klasifikasi emosi tertinggi terdapat pada emosi *Happy*, sebesar 46,92%, kemudian emosi *anger* sebesar 44,97%, sementara emosi

sadness hanya 7,18%, *sadness* 0,47% dan *fear* sebesar 0,46%. Tabel hasil klasifikasi dapat dilihat pada tabel 3,

Kelas	Jumlah
<i>happy</i>	39219
<i>anger</i>	37594
<i>sadness</i>	5999
<i>love</i>	397
<i>fear</i>	381

Pada grafik perbandingan hasil klasifikasi pada gambar 3 terlihat bahwa hasil klasifikasi emosi *happy* dan *anger* berada cukup jauh dibandingkan dengan kelas emosi *sadness*, *love* dan *fear*. Dalam dataset awal yang digunakan untuk melatih model klasifikasi, jumlah data untuk emosi *happy* dan *anger* jauh lebih tinggi dibandingkan dengan kelas emosi yang lain. Jumlah data yang seimbang antara kelas emosi diharapkan untuk menghindari bias dalam model yang menyebabkan ketidakseimbangan pada hasil klasifikasi. Ketidakseimbangan jumlah dataset dapat menyebabkan *overfitting* ataupun *underfitting*, *Overfitting* dapat terjadi jika model terlalu kompleks untuk ukuran dataset yang digunakan, sementara *underfitting* terjadi jika model terlalu sederhana.

SIMPULAN

model klasifikasi emosi yang dibangun menggunakan SVM memberikan performa yang cukup baik dengan akurasi sebesar 64,13%, presisi sebesar 68,02%, nilai *recall* sebesar 64,29%, dan *F1-score* sebesar 65,62%. Hasil klasifikasi emosi pada dataset tweet yang mengandung kata "Nusantara" menunjukkan bahwa emosi *happy* dan *anger* memiliki persentase yang jauh lebih tinggi dibandingkan dengan emosi lainnya. Hal ini disebabkan oleh ketidakseimbangan jumlah data pada dataset awal yang digunakan untuk melatih model, dimana data untuk emosi *happy* dan *anger* jauh lebih banyak dibandingkan dengan emosi lainnya. Ketidakseimbangan dataset dapat menyebabkan *overfitting* atau *underfitting*, yang pada akhirnya

mempengaruhi hasil klasifikasi pada dataset "Nusantara".

Sehingga diperlukan langkah-langkah untuk mengatasi ketidakseimbangan dataset, seperti menggunakan teknik *oversampling* atau *undersampling*, memilih fitur-fitur yang lebih representatif, dan memastikan model tidak terlalu kompleks atau sederhana. Dengan demikian, model dapat lebih umum mengklasifikasikan emosi dengan lebih seimbang dan menghasilkan hasil klasifikasi yang lebih baik pada dataset baru yang belum pernah dilihat sebelumnya

DAFTAR PUSTAKA

- [1] U.-U. Kotanegara, "PRESIDEN REPUBLIK INDONESIA MEMUTUSI(AN:," p. 54.
- [2] G. A. Buntoro, "Sentiments Analysis for Governor of East Java 2018 in Twitter," *Sinkron*, vol. 3, no. 2, p. 49, Dec. 2019, doi: 10.33395/sinkron.v3i2.10025.
- [3] K. Dharmendra, N. N. U. Januhari, R. A. N. Diaz, I. P. Ramayasa, and I. M. A. W. Putra, "Visualisasi Data Opini Publik pada Media Sosial Twitter (Studi Kasus : Nusantara Sebagai IKN Indonesia)," *Jurnal Teknik Informatika UNIKA Santo Thomas*, pp. 214–222, Dec. 2022.
- [4] M. S. Saputri, R. Mahendra, and M. Adriani, "Emotion Classification on Indonesian Twitter Dataset," in *2018 International Conference on Asian Language Processing (IALP)*, Nov. 2018, pp. 90–95. doi: 10.1109/IALP.2018.8629262.
- [5] F. M. J. Mehedi Shamrat *et al.*, "Sentiment analysis on twitter tweets about COVID-19 vaccines using NLP and supervised KNN classification algorithm," *IJECS*, vol. 23, no. 1, p. 463, Jul. 2021, doi: 10.11591/ijeecs.v23.i1.pp463-470.
- [6] E. Alomari, I. Katib, A. Albeshri, and R. Mehmood, "COVID-19: Detecting Government Pandemic Measures and Public Concerns from Twitter Arabic Data Using Distributed Machine Learning," *International Journal of Environmental Research and Public Health*, vol. 18, no. 1, Art. no. 1, Jan. 2021, doi: 10.3390/ijerph18010282.
- [7] I. M. D. Ardiada, M. Sudarma, and D. Giriantari, "Text Mining pada Sosial Media untuk Mendeteksi Emosi Pengguna Menggunakan Metode Support Vector Machine dan K-Nearest Neighbour," *Majalah Ilmiah Teknologi Elektro*, vol. 18, no. 1, p. 55, May 2019, doi: 10.24843/mite.2019.v18i01.p08.
- [8] L.-A.-M. Bostan and R. Klinger, "An Analysis of Annotated Corpora for Emotion Classification in Text," in *Proceedings of the 27th International Conference on Computational Linguistics*, Santa Fe, New Mexico, USA: Association for Computational Linguistics, Aug. 2018, pp. 2104–2119. Accessed: Dec. 04, 2022. [Online]. Available: <https://aclanthology.org/C18-1179>
- [9] K. Sailunaz and R. Alhaji, "Emotion and sentiment analysis from Twitter text," *Journal of Computational Science*, vol. 36, p. 101003, Sep. 2019, doi: 10.1016/j.jocs.2019.05.009.
- [10] I. K. Dharmendra, N. N. U. Januhari, I. P. Ramayasa, and I. M. A. W. Putra, "Uji Komparasi Sentiment Analysis Pada Opini Alumni Terhadap Perguruan Tinggi," *Jurnal Teknik Informatika UNIKA Santo Thomas*, pp. 1–6, May 2022, doi: 10.54367/jtiust.v7i1.1748.
- [11] M. G. Pradana, "PENGUNAAN FITUR WORDCLOUD DAN DOCUMENT TERM MATRIX DALAM TEXT MINING," *JURNAL ILMIAH INFORMATIKA*, vol. 8, no. 01, Art. no. 01, Mar. 2020, doi: 10.33884/jif.v8i01.1838.
- [12] I. K. Dharmendra, I. G. N. A. Kusuma, I. A. M. C. Dewi, and Edwar, "IMPLEMENTASI TEXT MINING UNTUK KLASIFIKASI OPINI ALUMNI PADA PERGURUAN TINGGI," *Jurnal Teknologi Informasi dan Komputer*, vol. 9, no. 3, Art. no. 3, Jul. 2023, Accessed: Aug. 18, 2023. [Online]. Available: <https://jurnal.undhirabali.ac.id/index.php/jutik/article/view/2504>